

ISRIC Report 2002/02c

**SOIL PARAMETER ESTIMATES FOR THE SOIL TYPES OF
THE WORLD FOR USE IN GLOBAL AND REGIONAL
MODELLING**

(Version 2.1)

N.H. Batjes
(July 2002)



INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE



INTERNATIONAL SOIL REFERENCE AND INFORMATION CENTRE

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise, without the prior permission of the copyright owner. Applications for such permission, with a statement of the purpose and extent of the reproduction, should be addressed to the Director, International Soil Reference and Information Centre, P.O. Box 353, 6700 AJ Wageningen, The Netherlands.

Copyright © 2002, International Soil Reference and Information Centre (ISRIC)

Disclaimer

While every effort has been made to ensure that the data are accurate and reliable, ISRIC cannot assume liability for damages caused by in-accuracies in the data or as a result of the failure of the data to function on a particular system. ISRIC provides no warranty, expressed or implied, nor does an authorized distribution of the data set constitute such a warranty. ISRIC reserves the right to modify any information in this document and related data sets without notice.

Full acknowledgement and referencing must be included in any documentation using any of the material contained in this data set.

Correct citation:

Batjes, N.H., 2002. Soil parameter estimates for the soil types of the world for use in global and regional modelling (Version 2.1; July 2002). ISRIC Report 2002/02c [Available on-line at <http://www.isric.org>], International Food Policy Research Institute (IFPRI) and International Soil Reference and Information Centre (ISRIC), Wageningen.

Inquiries:

Director, ISRIC
P.O. Box 353
6700 AJ Wageningen
The Netherlands
Fax: +31-317-471700
E-mail: soil@isric.nl
Web: www.isric.org

TABLE OF CONTENTS

FOREWORD.....	i
ABSTRACT.....	ii
1 INTRODUCTION	1
1.1 Rationale	1
1.2 Objectives	2
1.3 Structure of report.....	2
2 METHODOLOGY.....	5
2.1 Source of soil data.....	5
2.2 Geographic and taxonomic distribution of soil profile data	6
2.3 Comparability of soil analytical data	6
2.4 Outlier rejection scheme	11
2.5 List of derived soil parameters	11
2.6 Data processing and analysis	13
2.6.1 Depth-weighting by attribute	13
2.6.2 Stratification by topsoil textural class.....	13
2.6.3 Application of and outlier-rejection procedure.....	14
2.6.4 Data analyses	14
2.7 Computation of water retention by profile	14
2.8 Development of taxotransfer rules.....	15
3 RESULTS AND DISCUSSION.....	19
3.1 Data aggregation.....	19
3.2 Overview of data available for developing taxotransfer rules.....	19
3.3 Application of taxotransfer rules.....	19
3.4 Files of derived soil parameters	19
3.5 Possible limitations	21
4 CONCLUSIONS.....	25
ACKNOWLEDGEMENTS	26
REFERENCES	27
APPENDICES	33

List of Figures

<i>Figure 1. Representation of Major Soil Groups in profile data set relative to their extents on the 1:5M Soil Map of the World.....</i>	<i>7</i>
<i>Figure 2. Representation of Major Soil Groups in profile data set relative to their extents on the 1:25M World Soil Resources map.....</i>	<i>8</i>

List of Tables

<i>Table 1. Main digital sources of soil profiles used</i>	<i>5</i>
<i>Table 2. Geographic distribution of soil profiles by broad region.....</i>	<i>6</i>
<i>Table 3. Possible sources of variation in soil analytical data (Pleijzier, 1986).....</i>	<i>9</i>
<i>Table 4. List of soil parameters derived from WISE-2 profile data.....</i>	<i>12</i>
<i>Table 5. Criteria for defining ‘confidence’ in the derived data</i>	<i>14</i>
<i>Table 6. Summary of derived data for FAO-Unesco (1974) Legend</i>	<i>20</i>
<i>Table 7. Summary of derived data for FAO-Unesco (1988) Legend</i>	<i>20</i>
<i>Table 8. Overview of type of TTRs used and their frequency (1974 Legend).....</i>	<i>22</i>
<i>Table 9. Overview of type of TTRs used and their frequency (1988 Legend).....</i>	<i>23</i>

List of Appendices

<i>Figure 1. Representation of Major Soil Groups in profile data set relative to their extents on the 1:5M Soil Map of the World.....</i>	<i>7</i>
<i>Figure 2. Representation of Major Soil Groups in profile data set relative to their extents on the 1:25M World Soil Resources map.....</i>	<i>8</i>

Addendum

<i>Median soil depth by FAO soil unit (March 2002).....</i>	<i>44</i>
<i>Main changes in Version 2.1 (July 2002).....</i>	<i>45</i>

FOREWORD

There are likely many who will be surprised to find the logo of the International Food Policy Research Institute (IFPRI), the home of many an economist, on the cover of a report dealing exclusively with soil data. But such surprise would be misplaced. The quality of soil and water resources is becoming ever more critical to the sustainable productive capacity of agroecosystems and consequently to the livelihoods of the rural poor. Thus, soil is an increasingly strategic development resource in many developing countries, where soil productivity is often in sharp decline just as the demand for food is growing at unprecedented rates. This brings soil quality concerns to the attention of enlightened policymakers, and hence to the policy analysis community.

IFPRI has recently established a *Spatial Analysis Research Group*, one of whose goals is to foster the development and dissemination of databases and analytical tools to support policy-relevant, spatially explicit, decision support systems. ISRIC has developed, in close collaboration with FAO and UNEP, land resources information systems for assessment and monitoring (SOTER) at global, regional and national scales, and has worked with FAO and IIASA to initiate the development of a globally comprehensive and consistent dataset of derived soil attributes for all the major soil types of the world. Since that initial effort, ISRIC and its partners have compiled new data from many more soil profiles, and opportunities also existed to improve the analytical approach.

The provision of a globally comprehensive and consistent database of soil profile attributes for all the major soils of the world is seen by IFPRI as an important prerequisite to undertaking, for example, strategic global, regional, and sub-regional analyses of commodity and technology performance. This study, commissioned by IFPRI, using funds from the European Economic Commission, and undertaken by ISRIC, has generated a significantly improved global assessment of soil properties. Linking this data to the Digital Soil Map of the World, for example, provides place-based insights into 28 key soil attributes, for any area on earth where agriculture is practiced.

This is the second direct collaborative link between IFPRI and ISRIC, who together with other partners produced the *Pilot Analysis of Global Ecosystems: Agroecosystems*, and we remain firmly convinced that such cross-disciplinary efforts are both stimulating and increasingly necessary given the food security, poverty alleviation and environmental challenges we seek to address.

Stanley Wood
Senior Scientist
IFPRI, Washington D.C.

Roel Oldeman
Director
ISRIC, Wageningen

ABSTRACT

In 1997, the International Institute for Applied Systems Analysis (IIASA), Food and Agriculture Organization (FAO), and the International Soil Reference and Information Centre (ISRIC) identified the need for refinement of the agro-edaphic element in the revision of FAO's Agro-Ecological Zones (AEZ) methodology and in IIASA's Land Use Change and Land Cover project for Europe and Northern Eurasia (LUC). The resulting study was based on the analysis of the 4353 soil profiles held in version 1.0 of ISRIC's World Inventory of Soil Emission Potential (WISE) database. It led to the development of a set of files holding 'derived soil data for use in global and regional AEZ studies'. The study further identified several geographic, taxonomic and soil physico-chemical gaps, showing the persisting need for expanding the set of soil profile data available for this type of analyses.

Important gaps in the soil profile data were filled and the initial methodology was reviewed and refined in order to generate a revised set of physical and chemical parameter estimates for the soils units of the world, as described by the two FAO soil legends (version 1974 and 1988).

The current study uses all profiles held in WISE, version 1.0, to which have been added profiles extracted from currently available soil and terrain (SOTER) databases as well as new data from the literature. The corresponding data set holds physical and chemical attributes for over 9600 soil profiles.

Soil unit, topsoil textural class, and depth zone (0-30 cm and 30-100 cm) were used to cluster the horizon data. Criteria used are in accordance with conventions developed by FAO for use with the Soil Map of the World. Upon a screening on data 'integrity' and application of a statistical outlier rejection scheme, derived statistics were generated for 28 soil chemical and physical attributes. These attributes were identified as being useful for AEZ studies, crop growth simulation, and analyses of global environmental change.

Derived parameters are presented for: organic carbon; total nitrogen; C/N ratio; pH(H₂O), pH(KCl), pH(CaCl₂); sum of exchangeable Ca, Mg, Na and K (TEB); the ratio of exchangeable Ca/Mg and exchangeable (Ca + Mg)/K; the cation exchange capacity of the fine earth fraction (CEC_{soil}), the effective cation exchange capacity (ECEC), apparent cation exchange capacity of the clay fraction (CEC_{ap}), and cation exchange capacity of the clay fraction corrected for the contribution of organic matter (CEC_{clay}); base saturation; aluminium saturation; CaCO₃ content; gypsum content; exchangeable sodium percentage (ESP); electrical conductivity (ECe); bulk density; total porosity; weight percent of sand, silt and clay; gravel content; and, available water held between -5 kPa and -1500 kPa, -10 kPa and -1500 kPa, and -33kPa and -1500 kPa, respectively. A number of statistics is presented for each of these parameters, including number of observations, means, coefficients of variation, 95% confidence intervals, medians, and medians of absolute deviations. Simple taxotransfer rules are introduced to fill gaps that remained in the derived data, notably where sufficient measured data were lacking for particular attributes and soil units.

The attached set of derived soil parameters should be seen as the best possible estimates, based on the present selection of soil profiles and adopted data clustering procedure. The derived data set is considered appropriate for use in studies at a regional to global scale (< 1:250,000). Correlation of soil analytical data, however, must be done more accurately when more precise scientific research is considered.

Keywords: derived soil parameters; environmental modelling; FAO soil legend; WISE soil database; taxotransfer rules

1 INTRODUCTION

1.1 Rationale

People depend on the earth's ecosystems, which include soils, to sustain them. Soils are necessary for the growth of commercial crops of food and fibre. They also play an essential role in biogeochemical cycles, acting as sinks and sources of greenhouse gases, and in the hydrological cycle.

The soil is one of the essential components of the terrestrial biosphere, necessitating policies for management, evaluation and conservation. To implement such policies, it is necessary to have information that is harmonised both in time and space. Globally compatible soil databases are essential for establishing the current status of the earth's components and processes, and for monitoring rates of change in the status of these resources. A second important need for globally compatible soil databases is for guidance in the transfer of known technology and research observations from one area of the earth to any other area with similar soils and climate conditions (Baumgardner, 1999).

Requirements for soil information include the need for an up-to-date geographical coverage, access to secondary soil information obtained via transfer functions or models from the primary (measured) soil data, and monitoring of changes in soil characteristics as associated, for example, with changes in land use systems and processes of global change.

The Soil Map of the World, produced by FAO-Unesco (1974-1981), is still the only available, globally consistent soil database. Yet, it has been compiled using data collected prior to the 1970s and parts of it are now out-of-date. Hence the global update of the information on the world's soil and terrain resources in the World Soils and Terrain (SOTER) programme. Upon its completion, SOTER is to replace the digital version of the Soil Map of the World (Nachtergaele, 1999).

Until recently, many of the soil data sets available to global modellers were based on limited profile data and coarsely aggregated soil geographic data (Webb *et al.*, 1991; Zabler, 1986). With the completion of the Digital Soil Map of the World (SMW) at FAO (1995) and the World Inventory of Soil Emission Potentials (WISE) database at ISRIC (Batjes, 1997; Batjes and Bridges, 1994), it became possible to present updated, derived parameters for a number of soil characteristics for each of the soil units considered on the Soil Map of the World. As a sequel to these initial studies, staff at IIASA, ISRIC and FAO jointly presented a preliminary set of derived parameters for 20 soil attributes identified as being important for land evaluation in the context of AEZ studies (Batjes *et al.*, 1997). This collaborative activity further pointed at several geographic and taxonomic gaps that remained in the profile data set used.

Examples of applications that used soil parameters derived from WISE, version 1.0, include global modelling of environmental change (Alcamo *et al.*, 1998; Bouwman and Van Vuuren, 1999; Cindery *et al.*, 1998; Ganzenveld *et al.*, 1998; Hootsmans *et al.*, 2001), analyses of global ecosystems (Wood *et al.*, 2000), up-scaling and down-scaling of greenhouse gas emissions (Bouwman *et al.*, 2002; Denier van der Gon *et al.*, 2000), and crop simulation and agro-ecological zoning (Fischer *et al.*, 2000; Fischer *et al.*, 2001; Knox *et al.*, 2000). In addition, a subset of 1125 profiles from WISE 1.0

provided the soil attribute basis for the activities of the Global Soil Data Task Force of IGBP-DIS (2000).

1.2 Objectives

The aims of the current work are to review and complete the study by Batjes *et al.* (1997). Like for the preceding study, an important criterion for developing the derived data set is that results must allow linkage with the digital Soil Map of the World (FAO, 1995), or with more recent spatial databases that are based on the Revised Legend (FAO, 1988). The Revised Legend, for example, has been used for developing SOTER databases¹ at various scales, ranging from 1:5M to 1:50,000 (e.g., FAO and ISRIC, 2000; FAO *et al.*, 1998; Graef, 1999; Mantel *et al.*, 1999).

Topsoil textural class is an important criterion in the SMW for further characterization of the dominant soil of a soil association or mapping unit (FAO, 1995). Topsoil texture classes of associated soils and inclusions, and for dominant soils when these are not indicated in the mapping unit, are derived from composition rules. The soil association composition rules used for the SMW have been established in the context of the Agro-ecological Zones Project (FAO, 1978-1981). Soil units within a soil association are shown as dominant soil, associated soils (more than 20% of the area), or as inclusions (more than 5% of total area of the mapping unit). Soil phases, topsoil texture, and slope indicators as shown on the map reflect properties of dominant soil units only. For associated and included soils, rules were developed that specify topsoil texture and slope based on the most common occurrence of each soil unit. Further, if only the dominant soil group is indicated in a mapping unit then it is assumed that this represents the most common soil unit of this group. A full explanation of the composition rules is presented in FAO (1995).

All profile data are aggregated by soil unit, topsoil textural class and depth zone (i.e., topsoil (0-30 cm) and subsoil (30-100cm) respectively). The topsoil textural class is considered as a differentiating criterion, or flag, for the properties of the underlying profile in accordance with FAO conventions (FAO, 1995) and data requirements of FAO-IIASA (e.g., Fischer *et al.*, 2000; Fischer *et al.*, 2001). An important issue, in this respect, concerns the level-of-detail that can be considered justifiable in analysing the data and presenting results, keeping in mind the size and representativeness of the current profile data set, the type of attributes considered, and the generalizations (composition rules) applied.

1.3 Structure of report

The methodology is discussed in Chapter 2, which includes a description of the data sources (2.1); information on the geographic and taxonomic distribution of the soils profile (2.2); a discussion on the comparability of soil analytical data, as obtained from disparate data sources (2.3); the application of a statistical data outlier rejection scheme to homogenize the data sets by attribute (2.4); a listing of the set of derived soil parameters under consideration (2.5); data processing and analysis (2.6); the procedure for estimating water retention (2.7), and the development of a system of taxotransfer rules for generating 'surrogate' values in case of missing measured data (2.8).

¹ Subsequent to the approval of the World Reference Base for Soil Resources (WRB, 1988) by IUSSS at the 16th World Congress of Soil Science in Montpellier (1988), all newly developed SOTER databases follow the WRB classification.

Chapter 3 discusses aspects of data aggregation (3.1), the derived data (3.2), presents a summary of the derived data (3.3), and results of the application of the taxotransfer rules (3.4). Concluding remarks are drawn in Chapter 4.

Only selected summary tables are presented in this report due to limitations of space. The complete set of derived soil parameters, for all units considered in the original (1974) and Revised (1988) Legend of FAO, is available on-line at <http://www.isric.org>.

The classification in the 1974 and 1988 Legends is given in Appendix 1 and 2, respectively. Field names and percentage of records filled in the primary database, from now on referred to as WISE-2, are presented in Appendix 3. Appendix 4 presents a list of derived soil parameters for haplic Acrisols, as an example of data output. Finally, the structure of the digital data files, and coding conventions, are given in Appendix 5.

2 METHODOLOGY

2.1 Source of soil data

The core of the current soil profile collection is formed by the quality-controlled WISE data set (Batjes, 1997). In the systematic compilation of the soil profile data, special attention was given to the compatibility of laboratory methods by which the various analytical results were obtained (Batjes and Bridges, 1994). The 1974 and 1988 FAO-Unesco classification of the 4,353 profiles has been documented elsewhere (Batjes *et al.*, 1997), indicating several taxonomic and geographic gaps in the data set.

The initial soil analytical data set was complemented with profiles from the literature and with profiles extracted from various soil digital databases (Table 1). The laboratory of origin and source of the various profile data have been carefully recorded (see Appendix 3) to allow for comparability studies of soil analytical data.

Table 1. Main digital sources of soil profiles used

Region	Source type	Reference
Global	WISE soil profile data set (version 1.1)	Batjes (2002)
Uganda	Digital Soil Map of Uganda (1:250,000)	ICRAF (1995)
Tanzania	SOTER database (1:2M)	NSS (1997)
Mozambique	SOTER database for Nampulo region (1:1 M)	Unpublished data from INIA, Maputo
Ghana	Ghana Soil Information System 1-3	Velthuizen (1998)
Namibia	SOTER database (1:1 M)	Unpublished data from Agricultural Laboratories, Windhoek
Kenya	SOTER database (1:1 M)	Kenya Soil Survey, Nairobi (data set available via: http://www.isric.org)
Latin America and the Caribbean	SOTER database (1:5 M)	FAO et al. (1998)
Central and eastern Europe	SOTER database (1:2.5 M)	FAO and ISRIC (2000)
Syria	SOTER database (1:500,000)	Unpublished data from Ministry of Agriculture and Agrarian Reform, Damascus
Jordan	SOTER database (1:500,000)	Unpublished data from Soil Survey Section of Forestry and Management Department, Ministry of Agriculture, Amman
ISIS5	ISRIC Soil Information System	Kauffman (1996), Van Waveren and Bos (1988)
Zimbabwe	SOTER database (1:1M)	Unpublished data from Chemistry and Soil Research Institute, Harare
Niger and Benin	SOTER database (1:1M)	DFG (1999)
Nepal	SOTER database (1:50,000)	Unpublished data from ICIMOD, Kathmandu (A. Rotmans, <i>pers. comm.</i>)

Prior to being included, all potential additions were thoroughly screened for their 'acceptability' into the quality-controlled WISE format. After having passed this first threshold, all remaining profiles were re-classified according to the Legend and Revised Legend where necessary. Subsequently, they were entered or transferred to the common WISE-2 data base format. Thereafter all new additions were subjected to the automated and rigorous routine WISE-checks on data integrity (see Batjes, 1995).

2.2 Geographic and taxonomic distribution of soil profile data

Table 2 shows the distribution of the available soil profiles, clustered by major world region. Profiles from over 135 countries are represented in the data set. The soil classification, according to the Legend and Revised Legend of FAO, is shown in Appendix 1 and 2, respectively. Complete data sets are not always available for each sample or horizon for all soil attributes that can be accommodated in the WISE database (Appendix 3).

Table 2. Geographic distribution of soil profiles by broad region

Region	Number of profiles	
	WISE-1 [♦]	WISE-2
Africa	1799	3998
Australia and Pacific Islands	122	147
China, India, Indonesia & Philippines	553	628
Europe	492	1204
North America	266	326
South America and the Caribbean	599	2115
South west and Northern Asia (incl. Siberia)	522	1113
<i>Total</i>	<i>4353</i>	<i>9607</i>

[♦] See: Batjes et al. (1997)

The relative number of soil profiles, available for each major soil group of the Legend and Revised Legend respectively, is shown in Figure 1 and 2. For instance, Luvisols (see 'L' in Fig. 1) account for about seven percent of the total extent of soils in the SMW, and for about 14 percent of the total number of soil profiles. On the other hand, Lithosols (I) represent some eleven percent of the soils in the SMW and less than one per cent of the soil profiles under consideration. To a certain extent, the distribution of profiles in Figure 1 and 2 is a reflection of the fact that 'past' soil surveys mainly focussed on agriculturally promising areas.

2.3 Comparability of soil analytical data

International soil classification, correlation and interpretation imply international exchange of analytical data, and this requires international comparability of analytical data. The necessity of exchange of such data was already recognized during the 1936 ISSS Congress at Oxford (see ISSS Bulletin No. 69, 1986/1). In the 1960s, the use of different soil survey methods and classification systems throughout the world formed a problem for global correlation. As a result an international soil legend was developed for the Soil Map of the World (FAO-Unesco, 1974). Chemical and

physical characteristics necessary for classification according to this Legend followed the procedures of the US Soil Conservation Service (SCS, 1967). Similarly, the Revised Legend (FAO, 1988) and the World Reference Base for Soil Resources (WRB, 1998) list procedures for soil analysis deemed necessary for proper classification according to these systems.

In the 1980's, ISRIC initiated the International Laboratory Methods and Data Exchange (LABEX) programme in recognition of the large variability in analytical results worldwide (Sombroek, 1986). The variation in analytical data can have various sources, some of which are given in Table 1.

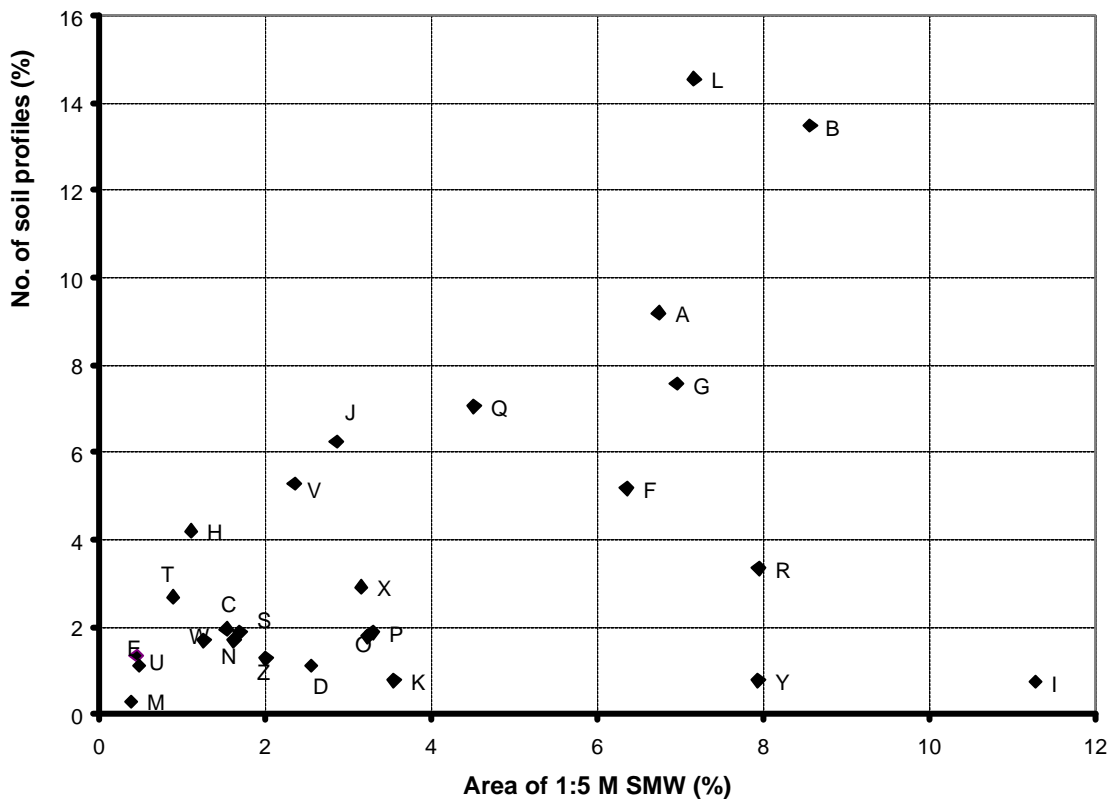


Figure 1. Representation of Major Soil Groups in profile data set relative to their extents on the 1:5M Soil Map of the World².

² Relative area is expressed as percentage of total area of all major soil groups considered, that is excluding all miscellaneous units shown on the 1:5 M scale Soil Map of the World (FAO, 1995). For soil codes see FAO-Unesco (1974).

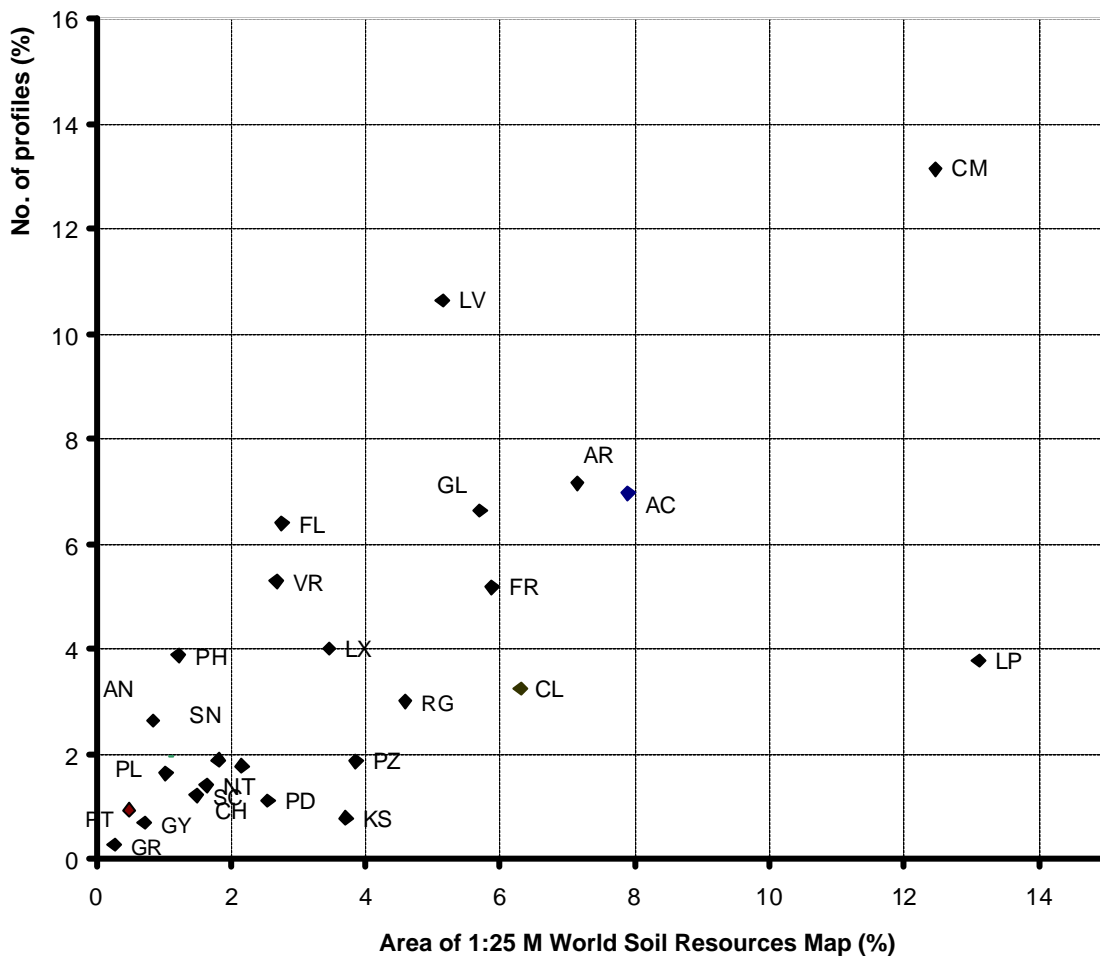


Figure 2. Representation of Major Soil Groups in profile data set relative to their extents on the 1:25M World Soil Resources map³.

Table 3 is illustrative of the difficulties in comparing analytical data collected in world-encompassing soil data compilation activities that, *of necessity*, are based on available 'historic data'. This aspect is already reflected by the fairly heterogeneous soil profile descriptions and analytical datasheets provided with the Soil Map of the World (FAO-Unesco, 1974-1981). Similarly, for the Soil Map of the European Communities (CEC, 1985, p. 80) and Soil Map of Middle Europe (ISSS, 1983), it 'has not been possible to give the method of analysis because every country has its own methods'. In databases such as WISE (Batjes, 1995; Batjes, 1997) and SOTER (Van Engelen and Wen, 1995), unique codes are given for the types of analytical procedures used, and the source laboratory has been documented. The Soil Database for Europe (Finke *et al.*, 1998) indicates that analytical data must have been determined according to acceptable methodologies such as those referred to by Madsen and Jones (1995), yet of necessity again mostly used available 'historic data'.

³ Relative area is expressed as percentage of total area of all major soil groups considered on the 1:25 M scale World Soil Resources, which excludes areas of Anthrosoles and Alisols (FAO, 1993). For soil codes see FAO (1988).

Table 3. Possible sources of variation in soil analytical data (Pleijsier, 1986).

Source of variation	Within laboratory	Between laboratory
Definitions		X
Procedures		X
Execution of procedure	X	
Instruments	X	
Operator	X	
Random error	X	
Calculations	X	X

Very seldom have all profile data in a single, trans-national or global database been determined according to a uniform set of analytical procedures, and in a single laboratory, as has been the case with ISRIC's ISIS database and the International Profiles analysed by NRCS-USDA (Lincoln). Thus, in practice, soil profiles held in trans-national databases often had to be classified with reference to the original methods in use at the national level (see CEC, 1985; FAO and ISRIC, 2000; FAO *et al.*, 1998; FAO-Unesco, 1974). Thus serious problems are prone to arise with the comparison of analytical data originating from disparate datasets (Pleijsier, 1989), especially when they cover a wide range of countries.

According to Vogel (1994), comparison of results determined by different methods shows that conversion of data can be accomplished in two ways. Either by merely adding or subtracting the average difference between methods or by using appropriate regression equations. The latter method is more accurate, but generally it can only be applied occasionally for specified soil types and for a limited number of analytical methods, and certainly not for each of the 106 soil units of the original Legend (FAO-Unesco, 1974) or the 153 soil units of the Revised Legend (FAO, 1988).

A desk study of the comparability of soil analytical methods, commonly in use in Central and Eastern European laboratories, with those of the ISRIC and NRCS reference laboratories provided no usable procedures for converting the various attribute data to the later standards (A.J.M. Van Oostrum, 1998, *unpubl. data*). Similarly, results of a study towards standardization of methods for gathering soil data and its application for computer modelling, which included a 'compatibility analysis of European soil data', have never been published (see European Soil Bureau, ESB Bulletin No. 1, March 2000, p. 6). Conversion from the Katschinski scheme to the USDA textural scheme, for example, requires data for more (> 4) particle-size fractions than are routinely considered in databases such as SOTER and WISE (see Rousseva, 1997; Wösten *et al.*, 1998).

In generating a list of derived soil properties for the 106 soil units considered in the Legend of the Soil Map of the World using the WISE database (Batjes, 1997), a broad screening by soil analytical methods was performed. Similarly, in a follow up study jointly carried out by ISRIC, IIASA and FAO soil data held in WISE were screened according to the broad type of soil analytical methods used (Batjes *et al.*, 1997, p. 6-7). A practical implication of such a screening is, however, that the actual size of the data set available for analyses is greatly reduced; only those profiles that have been analysed according to NRCS-USDA or ISRIC procedures could be maintained for the actual statistical analyses. In a SOTER-based study aimed at generating derived soil parameters for Central and Eastern Europe, however, a similar screening on the comparability of analytical methods proved unrealistic (Batjes, 2000b), similar to what was the case earlier for the Soil Map of the World (FAO, 1995) and the European Soil Database (ESB, 1999), despite the fact that the original methods and laboratories have been documented in these databases.

According to Van Reeuwijk (1982), a certain minimum level of variation will always occur even when analytical methods are standardized:

- ± 11% for clay content determination
- ± 20% for CEC measurements
- ± 25% for CEC of the clay fraction
- ± 10% for base saturation
- ± 0.2 units for pH measurements.

The above figures indicate that laboratory data cannot be taken as absolute figures and that therefore interpretations should be done very cautiously, especially for soil classification (Breimer *et al.*, 1986; Van Reeuwijk, 1983) and data interpretation (Nettleton *et al.*, 1996). In this context, it would be worthwhile to compare the relative importance of spatial variability for a specific group of soil profiles, for example belonging to one classification unit in a particular agro-ecological zone (Vogel, 1994), and uncertainties associated with mapped variations in soil conditions (Batjes, 2000a) versus variations caused by differences in soil analytical procedures.

Although observational data are almost always subject to error, spatial data seem to suffer more from imperfect quality than do other kinds of data (Nettleton *et al.*, 1996). Conventionally, soil mapping units are supposed to be about 85 per cent pure, but in practice the figure is likely to be about 50 to 65 per cent (see Landon, 1991). Therefore, differences in soil analytical methods may not need to be considered explicitly in data analyses for regional and global modelling (scale < 1:250,000). Common sources of error in spatial data have been reviewed elsewhere (Burrough, 1986; Goodchild, 1994).

It can be argued, for example, that differences in CEC values obtained with different analytical approaches for a deeply weathered Acrisol will be typically different from those obtained for a montmorillonitic Vertisol. For example, median CEC values of 7 $\text{cmol}_c \text{kg}^{-1}$ versus 48 $\text{cmol}_c \text{kg}^{-1}$ in 1 M NH_4OAc in the subsoil (see Batjes, 1997). Alternatively, differences in CEC values obtained with different analytical methods for 'low activity clay' soils, for example, would be in the order of 'CEC < 4 $\text{cmol}_c \text{kg}^{-1}$ ' for 'sum of bases plus exchangeable acidity' versus 'CEC < 7 $\text{cmol}_c \text{kg}^{-1}$ for sum of cations at pH 7' (Kamprath, 1986). In case of non-variable clay charge soils such differences should be smaller.

Even when analytical methods are standardized, the minimum level of variation for CEC measurements is ± 20% (Van Reeuwijk, 1982). Similarly, it can be argued that differences in results of particle size analyses obtained with say the pipette and hydrometer and different dispersion methods for sandy soils should be typically different from those obtained for clay soils (having a similar clay mineralogy and similar 'cementing' agents). Here, again, the minimum level of variation is ± 11% for clay content determination, if analytical methods are standardized (Van Reeuwijk, 1984).

Based on the above considerations, a pragmatic approach to the comparability of soil analytical data has been adopted for the present study pending the availability of more elaborate 'comparability' studies at the Pan-European and Global level.

2.4 Outlier rejection scheme

The analytical data sets for each combination of soil unit, texture class and depth zone was screened using a robust data outlier rejection scheme, by attribute, similar to the one applied during the initial rounds of the LABEX programme. As the data distributions are skewed, the median is a better estimate of the centre of the data than the mean. For each attribute the median value (MED) and the median of the absolute deviations (MAD) are calculated. The MAD is the median of the differences between each observation and the median. Thus, like the standard deviation, MAD is a measure for the spread of the data. Outliers, flagged as values falling outside the range $MED \pm 2 \cdot F \cdot MAD$, are rejected. The factor F is such that, had the data been normally distributed, 5% of the data would have been marked for rejection. Inherently, the size of F is determined by the sample size; methodological details may be found elsewhere (Pleijssier, 1986; Van Reeuwijk, 1998, p. 123).

Each sample set (e.g., weighted topsoil data for CEC_{soil} for medium textured Ferric Acrisol profiles) was subjected to the above outlier rejection scheme. The rejection scheme has been skipped, however, when the original sample population was small ($n < 5$). In such cases, the derived soil data have been flagged as having a 'low' confidence level (see Table 4).

As has been indicated earlier, the current approach is considered appropriate for soil data applications at an observational scale ($< 1:250,000$). Correlation of soil analytical data, however, must be done more accurately when more precise scientific research is considered. In such cases, regressions of results obtained with different analytical results, for example CEC in an un-buffered 1 M KCl solution and 1 M NH_4Oac solution buffered at pH7, should be carried out for each of the soil types and functional groupings under consideration. Often, however, the primary datasets necessary to perform such comparative analyses will be lacking or only be available for a limited number of well-defined soil types (Pleijssier *et al.*, 1986; Vogel, 1994).

Clustering of soil data according to the USDA family differentiae for mineral soils (Soil Survey Staff, 1992) and inferred clay activity might also be worthwhile considering at more detailed scales, where soil classification would be at a level of detail equivalent to or below that of the subgroup. This would probably allow for a better functional grouping of soils than the necessarily coarse textural groupings adopted for the 1:5 M scale Soil Map of the World (FAO, 1995; FAO-Unesco, 1974).

2.5 List of derived soil parameters

Twenty soil chemical and physical attributes were identified as being required for Agro-ecological Zones (AEZ) and Land Use Change (LUC) studies (see Batjes *et al.*, 1997). Many of these attributes are also needed for global change research (Batjes *et al.*, 1995; Ingram and Gregory, 1996; Scholes *et al.*, 1995). This list has been expanded to include 8 new attributes.

Table 4. List of soil parameters derived from WISE-2 profile data

Profile identifier

FAO-Unesco soil unit (in 1974 and 1988 Legend, respectively)

Topsoil textural class

Measured data to be analysed (for topsoil and subsoil, respectively)

Organic carbon

Total Nitrogen[#]C/N ratio^{# \diamond} pH(H₂O)pH(CaCl₂)[#]pH(KCl)[#]Sum of exchangeable Ca²⁺, Mg²⁺, Na⁺ and K⁺ (TEB) ^{\diamond} Ratio of exchangeable Ca²⁺/Mg²⁺ ^{\diamond #}Ratio of exchangeable (Ca²⁺+Mg²⁺)/K⁺ ^{\diamond} Effective CEC[†]CEC_{soil}CEC_{clay} (corrected for the contribution of organic matter) ^{\diamond ∇}Apparent CEC (not corrected for the contribution of organic matter)^{# \diamond} Base saturation (as % of CEC_{soil}) ^{\diamond} Aluminium saturation (Al³⁺ as % of ECEC)[#]CaCO₃ content

Gypsum content

Exch. sodium percentage (ESP) ^{\diamond} Electrical conductivity (ECe)[#]

Bulk density

Total porosity (as derived from bulk density) ^{\diamond ♦}

% sand

% silt

% clay

Available Water (AWC₁; from -5 to -1500 kPa or pF 1.7 to pF 4.2)^{#♦}Available Water (AWC₂; from -10 to -1500 kPa or pF 2.0 to pF 4.2)[♦]Available Water (AWC₃; from -33 to -1500 kPa or pF 2.5 to pF 4.2)[♦]Gravel[#][#] New parameter. ^{\diamond} Calculated from other measured soil properties.[†] EC is defined as sum of exchangeable (Ca²⁺ + Mg²⁺ + K⁺ + Na⁺) plus exchangeable H⁺ + Al³⁺ (Van Reeuwijk, 1993).[∇] EC_{clay}, the CEC of the clay size minerals was calculated from CEC_{soil} by assuming a mean contribution of 350 cmol_c kg⁻¹ OC, the common range being from 150 to over 750 cmol_c kg⁻¹ OC (Klamt and Sombroek, 1988).♦ Total porosity was calculated from bulk density, assuming an average particle density of 2.65 g cm⁻³.♦ The soil moisture ranges considered in determining water retention are: pF 1.7 to pF 4.2 (-5 to -1500 kPa; AWC₁), pF 2.0 to pF 4.2 (-10 to -1500 kPa; AWC₂), and pF 2.5 to pF 4.2 (-33 to -1500 kPa; AWC₃). The suction limits for AWC₃ conform to USDA standards (Soil Survey Staff, 1983), whereas AWC₂ is commonly used by FAO in its AEZ studies (Doorenbos and Kassam, 1978).

Table 4 shows those attributes for which quantitative data are considered desirable, and which are to be derived from the primary (measured) soil data. The current set includes the basic soil physical and chemical attributes needed by crop simulation models (Knox *et al.*, 2000; DSSAT/IDSS⁴ interface under development at IFDC/IFPRI), and can be used for exploring land quality effects on world food supply (Bouma *et al.*, 1998).

Derived parameters for several other attributes, such as surface slope, permeability, soil drainage and depth of soil, can be inferred from information embedded in the codes of the Soil Map of the World. Expert rules for deriving a selection of these attributes are presented on the digital Soil Map of the World (FAO, 1995), and have since then be refined in the context of the Global Agro-Ecological Zones project (Fischer *et al.*, 2000). Refinement of such procedures falls outside the scope of the current study.

2.6 Data processing and analysis

Data screening and analysis by combination of soil unit, textural class and depth zone involved 6 stages: (1) screening by profile on apparent reliability/completeness of the measured physical and chemical data (see Section 2.1); (2) allocation of individual samples of a profile to the topsoil and subsoil, respectively, and depth-weighting by attribute; (3) allocation of each profile to the Coarse, Medium or Fine topsoil textural class; (4) an outlier-rejection analysis, followed by (5) the actual statistical analysis of the data; and, (6) filling of gaps in the derived data sets using taxotransfer rules.

2.6.1 Depth-weighting by attribute

Depth weighed values for all attributes under consideration (see Table 1) were computed for both the *topsoil* (0 to 30 cm) and *subsoil* (30 to 100 cm), based on their depth of occurrence in a profile. In case no measured data were available, this was flagged by a '-1'; such data were excluded from the subsequent analyses. Soil analytical data for horizons below 100 cm, as well as litter layers, were not included in this study.

2.6.2 Stratification by topsoil textural class

Topsoil textural class was determined according to the definitions of the Legend (FAO-Unesco, 1974) and Revised Legend (FAO, 1988), respectively. The three topsoil textural classes considered are: Coarse (C), Medium (M), and Fine (F). In addition, the symbol '#' is used for analytical data analysed by soil unit, attribute and depth zone only, that is without further stratification by topsoil textural class. In accordance with FAO (1995) conventions, the code for the topsoil textural class was used as a *flag* (i.e., a clustering criterion) for the corresponding subsoil. For example, a subsoil may be flagged as 'C' based on the topsoil textural class while in fact it is medium textured (see Explanatory Note, FAO, 1995).

⁴ http://www.ifdc.org/Global_Presence/Research_and_Development/decision_support_systems.html

2.6.3 Application of and outlier-rejection procedure

Although all profile descriptions, corresponding soil classifications, and soil physical and analytical attributes have been subjected to an intensive screening prior to their acceptance into the WISE-2 database, a number of outliers are likely to remain. In order to reduce the influence of such outliers, use of the median is generally preferred to the average (Snedecor and Cochran, 1980). Values of each attribute were tested for departure from the median at the 95% level-of-confidence according to Pleijsier (1989), using the procedure described in Section 2.4.

Application of this procedure led to a median rejection percentage of ~ 8 % with lower and upper quartiles of 0 % and ~ 14 %, respectively (for both legends). The highest (median) rejection percentages were found for data relating to calcium carbonate content, electrical conductivity, exchangeable sodium percentage, gravel content, and nutrient ratios. The reduced sample population that remained after exclusion of the outliers was used for the actual statistical analyses.

2.6.4 Data analyses

The statistical parameters generated in this study include sample size, means, medians and 95%-confidence intervals (see Appendix 4). In addition, an *indicator* for the level of possible 'confidence' (*CONF*) in the derived medians has been introduced (Table 5). The underlying assumption is that the 'confidence' in the results shown should increase with size of the sample populations. Since the current analyses are based on a still relatively small and not necessarily representative selection of soil profile descriptions, consideration of *CONF* in conjunction with expert knowledge will be essential when developing taxotransfer rules to fill gaps in the derived data (see Section 2.8).

Table 5. Criteria for defining 'confidence' in the derived data

CONF	Description	NUM
V	Very high	>30
H	High	15 - 29
M	Moderate	5 - 14
L	Low	1 - 4
-	No data	0

- NUM is the sample size remaining after the data screening and outlier rejection procedure. CONF is the inferred confidence level in the derived data.

2.7 Computation of water retention by profile

In spite of the limitations that are associated with the concept of available water capacity (AWC), it remains useful in regional studies of soil moisture deficits and irrigation intervals (Bregt and Beemster, 1989; Kern, 1995), agro-ecological zoning and assessments of crop production potentials (FAO, 1978-1981; Fischer *et al.*, 2001; Luyten, 1995), and in simulation of global land cover changes as affected by economic factors and climatic change (Prentice *et al.*, 1992; Zuidema *et al.*, 1994).

Available Water Capacity (AWC_i), computed by soil unit and depth zone, has been used to estimate the 'water retention by profile' or so-called Total Available Water Capacity ($TAWC_i$):

- Shallow soils (i.e., Lithosols, Rendzinas and Rankers):

$$TAWC_i = d * AWC_{it}$$

- Other soils:

$$TAWC_i = 3*AWC_{it} + 7*AWC_{ib}$$

where:

i is the tension range considered for *AWC*

AWC is the median water retention (obtained after 2.6.4)

t and *b* refer to the topsoil and subsoil, respectively

d is the maximum depth range (*maxdep*, in dm).

A maximum depth of 100 cm has been adopted for all soil units, except for Lithosols (and lithic Leptosols, LPq; *maxdep*= 10 cm), Rendzinas and Rankers (*maxdep*= 30 cm).

The number of observations used for the original pF measurements, the variability of these measurements, and the method by which the original pF values were measured and then averaged to produce single data are seldom traceable, making pF and AWC data prone to errors (Shouse *et al.*, 1995; Vogel, 1994). Values measured for $AWC_{(pF1.7-pF4.2)}$ will be larger than those for $AWC_{(pF2.0-pF4.2)}$ and $AWC_{(pF2.5-pF4.2)}$, and the 'pF range' used should always be specified. Commonly used intervals for defining TAWC include -5 kPa to -1500 kPa (i.e., pF 1.7 to 4.2) in the United Kingdom, -10 kPa to -1500 kPa (i.e., pF 2.0 to 4.2) in the Netherlands and by FAO (Doorenbos and Kassam, 1978), and -33 kPa to -1500 kPa (i.e., pF 2.5 to 4.2) in the USA (Soil Survey Staff, 1983).

2.8 Development of taxotransfer rules

The total number of soil profiles in WISE-2 has been increased with 5254 profiles, or almost doubled, *vis a vis* WISE-1 (Table 2). Nonetheless there still are gaps both in term of soil units represented (Appendix 1 and 2) and in terms of attributes for which measured data are commonly available. Soil physical attributes, such as soil moisture retention and bulk density, in particular are under-represented as these data are seldom collected during routine soil surveys (Appendix 3). Some derived attributes such as aluminium saturation and exchangeable sodium percentage are only measured for specific soil types (see Van Reeuwijk, 1993).

Medians rather than means should be used in defining taxotransfer rules, as this will reduce the effect of outliers. In view of the very large number of combinations of soil unit, textural class, depth zone, and attributes under consideration (i.e., over 28000 in total for the 1974 Legend alone, including major groups) a pragmatic approach had to be adopted for filling gaps in the derived data. These generalized procedures, which expand on those developed during the preceding study by researchers from IIASA-ISRIC-FAO, are presented in the next paragraphs.

Such procedures, referred to as taxotransfer rules⁵ (TTR), will be used whenever the confidence in a certain derived attribute is considered to be 'low' (defined as $NUM < 5$). In order to keep track of

⁵ A *taxotransfer* function is the estimation of soil parameters based on modal soil characteristics of soil units, as derived from a combination of their classification name or taxon (which by definition often implies a certain range for a number of properties), expert knowledge and empirical rules, and a statistical analysis of a large number of soil profiles belonging to the same taxon. A *pedotransfer* function is a mathematical relationship between two or more soil parameters which shows a reasonable high level of statistical confidence. This relationship is used to facilitate the estimation of a non-measured soil parameter from one or more measured ones (see Batjes *et al.*, 1997).

the rules adopted, they have been documented in the TTR-derived data sets together with information on the (original) confidence in the substituted-data, the number of samples considered, and the substituted median (see Appendix 4).

The taxotransfer rules are applied sequentially, as needed:

Rule 0:

If $NUM \geq 5$ for the considered combination of soil unit, attribute, depth zone and topsoil textural class, then use the median (*MED*) for the corresponding population. That is, the derived parameter remains as is and this is flagged as rule 'R0' in the corresponding data file. (Note: If the attribute either is sand, silt or clay the TTR procedure is skipped, so that the sum of (sand + silt + clay) remains 100 %, and the rule is flagged as 'T0').

Rule 1:

If there is only a limited number of measured data ($NUM < 5$) for a specific combination of soil unit, soil attribute and topsoil textural class but $NUM \geq 5$ for the corresponding combination of major soil group, soil attribute and topsoil textural class, then the median for this major group, topsoil textural class, depth interval and soil attribute is substituted in the derived data set. (Note: in case of textural data, i.e. for sand, silt and clay, the rule is flagged as 'T1' should it apply)

Rule 2:

If median pH(H₂O) for the considered combination of soil unit, depth zone and topsoil textural class is less than 6.5 (rule 'r0')⁶, then the CaCO₃ content is set at 0 percent (see Brunt and Van Reeuwijk, 1997, p. 6-5) if $NUM_{CaCO_3} < 5$. The rule is flagged with 'R2' in the derived data set.

Rule 3:

The CaCO₃ content is tentatively set at 3 % if pH(H₂O) > 7.3 and at 1 % when $7.0 < \text{pH}(\text{H}_2\text{O}) < 7.3$, provided $NUM_{CaCO_3} < 5$. These rules are flagged as 'R3' and 'r3' respectively in the derived data set (CONF is then flagged as 'ee' for expert estimate).

Rule 4:

If median pH(H₂O) for the considered combination of soil unit, depth zone and topsoil textural class is less than 7.0 and $NUM_{gypsum} < 5$, then the gypsum content is tentatively set at 0 percent.

Rule 5:

If median pH(H₂O) for the considered combination of soil unit, depth zone and topsoil textural class is less than 6.5 and $NUM_{ESP} < 5$ then the exchangeable sodium percentage (ESP) is set at 0 percent.

Rule 6:

If median pH(H₂O) for the considered combination of soil unit, depth zone and topsoil textural class is over 5.4 and $NUM_{ALSAT} < 5$ then the exchangeable aluminium percentage is set at 0 percent (see Brunt and Van Reeuwijk, 1997, p. 6-5).

⁶ Whenever a pedotransfer rule is used that requires soil reaction data, the rule for PHH2O is recoded to 'r0' (this applies for rules R2 to R7).

Rule 7:

Based on preliminary regression analyses of aluminium saturation (ALSAT) versus soil reaction (PHH₂O), ALSAT is set at 30 % if $4.5 < \text{pH}(\text{H}_2\text{O}) < 5.4$ and at 50 % if $\text{pH}(\text{H}_2\text{O}) < 4.5$ plus $\text{NUM}_{\text{ALSAT}} < 5$. In such cases, the rule is flagged as 'r7' and 'R7', respectively.

Rule 8:

The available water capacity (TAWC_i) is calculated from AWC_1 , AWC_2 and AWC_3 data as appropriate (see Section 2.7). If the rule for AWC_i for both the topsoil and subsoil is 'R0' (i.e., $\text{NUM} \Rightarrow 5$ in both cases) then the rule for TAWC_i is also flagged 'R0'. If there are not enough measured data to compute water retention for the upper 1 m of soil (i.e., $\text{NUM} < 5$ for either the topsoil or subsoil) then TAWC_i is computed using results of the preceding TTRs applied to AWC_1 , AWC_2 , and AWC_3 as appropriate (e.g., Lithosols are at most 10 cm deep by definition; Rendzinas and Rankers are less than 30 cm deep). If the rule for the topsoil and subsoil is either 'R0' or 'R1' then the final rule for is flagged 'R1'. Otherwise, the estimated value for TAWC_i is flagged as 'R8' (see Tables 8 and 9).

Rule 9:

If there are no measured data ($\text{NUM} = 0$) for a certain combination of attribute, textural class, depth interval and soil unit, and $\text{NUM} < 5$ for the corresponding combination of major group, soil attribute and textural class, then no data substitution is made and the rule is originally flagged as 'R?'. Ideally, in these cases, a group of experts must recommend the proper substitution-procedure and agree on the value obtained through this procedure (see Batjes *et al.*, 1997). However, in order avoid 'gaps' in the set of derived soil parameters the following approach has been introduced where the rule is originally flagged as 'R?'. The median for the major group, topsoil textural class, depth interval and soil attribute for those mineral soil units with 'over 5% coverage on SMW' (see Figure 1) is substituted in the derived data set as our current best guess; a differentiation is made for soil units that are essentially acid versus those that are essentially basic in nature. These cases are flagged as 'R9' for soil units with a predominantly basic reaction, and as 'r9' for those that are predominantly acid. Inherently, all taxotransfer-derived data that have been flagged as 'R9' or 'r9' must be viewed with additional care. (Note: Certain combinations of soil unit and topsoil texture are unlikely to occur in 'reality' when classification criteria are strictly applied: that is without allowing for any 'fuzziness' in differentiating criteria used. For example, Vertisols with a coarse topsoil texture or Arenosols with a fine topsoil texture. Yet, such combinations may also have been flagged as 'R9' or 'r9' with the present procedure).

Rule 'Or' and rule 'An':

Rule number 9 has not been applied to Histosols and Andosols in view of their rather specific behaviour. In such cases, if $1 \leq \text{NUM} < 5$ (that is when the rule is flagged as 'R?' after application of Rule 1) the best available estimate for the corresponding combination of major soil group, attribute and textural class is used, irrespective of NUM . In the database, this is flagged as rule 'Or' for Histosol units and as rule 'An' for Andosols units.

Clearly, the application of taxotransfer rules remains fraught with uncertainty. Especially when such rules are applied to soil properties that are highly variable in space and time (e.g., ECe levels) or prone to change with management practices (e.g., pH, BSAT and ALSAT subsequent to liming; topsoil bulk density subsequent to tillage).

The fact that a certain derived parameter has been assigned a high 'confidence' rating does not necessarily imply that this derived parameter will be representative for the soil unit under consideration. Profile selection for this study, like for all other global profile databases, is not probabilistic but based on available survey data. Also, several of the attributes considered in this study are not diagnostic in the FAO-Unesco Legend. Therefore, it is recommended that soil experts and other users of the derived parameters also review estimates with a so-called high 'confidence' rating, prior to their use in auxiliary models.

3 RESULTS AND DISCUSSION

3.1 Data aggregation

Generalization of measured soil (profile) data by soil unit and topsoil textural class — to permit linkage with the units shown on the digital Soil Map of the World — for use in regional and global models, involves the transformation of variables that show a marked spatial and temporal variability, and that have been determined in many laboratories according to various methods. No attempt was made in this study to establish the location of individual profiles, because each profile description was assumed to be representative for a particular soil unit in the Legend and Revised Legend, respectively. As such, differences in landform, parent material, land use history, natural vegetation, and time of sampling and laboratory analyses are not considered explicitly.

3.2 Overview of data available for developing taxotransfer rules

All attributes have been analysed by soil unit, topsoil textural class and depth zone. A summary of the frequencies of occurrence is shown in Table 5 and 6. Taxotransfer rules will generally have to be applied when the frequency of occurrence reads 'N(one)' or 'F(ew)'. This is particularly the case for soil physical attributes such as water retention where the frequency of occurrence of categories 'N' and 'F' often exceeds 60 per cent.

Summarizing, Table 5 and 6 show that there generally exists an adequate basis ('CMA' > 70%) for filling gaps in the derived data using taxotransfer rules, except for water retention.

3.3 Application of taxotransfer rules

The scheme of taxotransfer rules described in section 2.8 has been applied to the various attributes. Tables 8 and 9 give an overview of the type of taxotransfer rules that have been used to fill gaps in the derived data for both the 1974 and 1988 Legend. They also show how often a particular rule has been used for each attribute under consideration.

On the whole, over 75% of all the derived parameters could be generated using either rule 'R0' (i.e., no changes as enough measured data exist) or rule 'R1' (i.e., derived soil parameter based on measured data available for the corresponding combination of major soil group, textural class, and depth zone). Notable exceptions are the derived parameters for soil water retention per depth zone (AWC_i) amounting to ~ 30 % of the cases for tensions in the -5 to -1500 kPa range (pF1.7 to pF4.2), about 50% for the -10 kPa to - kPa range (pF2.0 to pF4.2), and ~ 75% for the -33 kPa to - 1500 kPa range (pF2.5 to pF4.2). In case of profile available water retention ($TAWC_i$), these values are ~ 30 %, ~ 50 %, and ~ 70 %, respectively.

3.4 Files of derived soil parameters

Results of the various analyses have been retained as dbf-files, which can readily be imported by a range of software packages. Listings of derived parameters for Haplic Acrisols are attached in a self-explanatory table (Appendix 4), as an example. The complete set of derived soil data set is available via <http://www.isric.org>.

Table 6. Summary of derived data for FAO-
Unesco (1974) Legend

Attrib	Depth	Frequency of occurrence					
		N	F	C	M	A	CMA
ALSAT	A	2	5	13	18	62	93
ALSAT	B	2	5	15	17	61	93
APARCEC	A	1	8	12	20	59	91
APARCEC	B	1	9	12	22	56	90
AWC1	A	45	39	16	0	0	16
AWC1	B	49	38	14	0	0	14
AWC2	A	25	42	25	8	0	33
AWC2	B	26	40	26	8	0	34
AWC3	A	12	20	36	17	15	68
AWC3	B	13	22	35	15	16	66
BSAT	A	1	12	26	25	36	87
BSAT	B	1	17	26	22	34	82
BULKDENS	A	4	8	25	25	38	88
BULKDENS	B	4	8	29	24	35	88
CACO3	A	1	6	12	22	59	93
CACO3	B	1	6	17	19	57	93
CECCLAY	A	2	8	17	19	55	91
CECCLAY	B	3	7	15	22	53	90
CECSOIL	A	1	5	12	21	61	94
CECSOIL	B	1	7	11	23	58	92
CLAY	A	1	6	10	14	69	93
CLAY	B	1	7	11	17	65	93
C_N	A	1	8	20	23	48	91
C_N	B	1	10	27	22	40	89
ECE	A	3	5	16	23	54	93
ECE	B	3	6	17	22	51	90
ECEC	A	7	16	34	20	24	78
ECEC	B	7	17	34	17	24	75
ESP	A	1	9	24	27	39	90
ESP	B	1	13	21	28	37	86
GRAVEL	A	7	26	30	24	13	67
GRAVEL	B	7	20	37	23	13	73
GYPSUM	A	3	12	20	21	44	85
GYPSUM	B	2	15	20	21	42	83
ORGC	A	1	6	14	18	61	93
ORGC	B	1	6	14	22	57	93
PHACL2	A	7	26	37	9	21	67
PHACL2	B	6	28	38	14	15	67
PHH2O	A	2	5	10	19	64	93
PHH2O	B	2	5	13	18	62	93
PHKCL	A	5	10	27	25	33	85
PHKCL	B	5	11	30	22	32	84
R_CAMG_K	A	1	14	15	31	39	85
R_CAMG_K	B	3	17	20	28	31	79
R_CA_K	A	1	14	17	29	39	85
R_CA_K	B	3	17	22	23	34	79
R_CA_MG	A	1	6	20	25	49	94
R_CA_MG	B	1	11	19	27	42	88
SAND	A	1	6	10	14	69	93
SAND	B	1	7	11	15	67	93
SILT	A	1	6	10	14	69	93
SILT	B	1	7	11	15	67	93
TEB	A	1	12	21	28	38	87
TEB	B	1	17	22	24	36	82
TOTN	A	1	7	19	22	52	93
TOTN	B	1	10	22	20	47	89
TOTPORES	A	4	8	26	24	38	88
TOTPORES	B	4	8	28	25	35	88

Table 7. Summary of derived data for FAO-
Unesco (1988) Legend

Attrib	Depth	Frequency of occurrence					
		N	F	C	M	A	CMA
ALSAT	A	5	12	17	22	44	83
ALSAT	B	5	12	19	22	41	82
APARCEC	A	4	13	18	20	46	84
APARCEC	B	5	12	20	20	43	83
AWC1	A	56	35	10	0	0	10
AWC1	B	58	34	9	0	0	9
AWC2	A	39	39	18	5	0	23
AWC2	B	41	38	17	5	0	22
AWC3	A	20	28	35	9	8	52
AWC3	B	19	30	32	11	8	51
BSAT	A	7	17	31	18	27	76
BSAT	B	7	22	30	17	24	71
BULKDENS	A	7	18	29	20	25	74
BULKDENS	B	7	19	30	21	23	74
CACO3	A	5	12	17	22	44	83
CACO3	B	5	13	20	20	42	82
CECCLAY	A	6	12	20	22	40	82
CECCLAY	B	6	13	21	23	38	82
CECSOIL	A	4	10	17	24	44	85
CECSOIL	B	5	11	17	23	44	84
CLAY	A	4	12	16	17	52	85
CLAY	B	5	12	18	18	47	83
C_N	A	5	15	22	23	35	80
C_N	B	5	17	31	20	28	79
ECE	A	6	12	22	21	39	82
ECE	B	6	13	24	20	37	81
ECEC	A	11	24	33	15	16	64
ECEC	B	14	26	30	14	15	59
ESP	A	7	16	27	24	26	77
ESP	B	6	20	27	21	26	74
GRAVEL	A	18	29	25	18	8	51
GRAVEL	B	18	25	35	14	9	58
GYPSUM	A	8	14	25	19	33	77
GYPSUM	B	7	20	24	18	31	73
ORGC	A	5	10	20	20	45	85
ORGC	B	5	12	20	21	43	84
PHACL2	A	12	37	27	10	13	50
PHACL2	B	15	39	27	9	11	47
PHH2O	A	5	12	16	22	45	83
PHH2O	B	5	12	17	24	42	83
PHKCL	A	10	16	30	21	23	74
PHKCL	B	9	18	34	16	23	73
R_CAMG_K	A	6	20	25	20	28	73
R_CAMG_K	B	8	24	26	16	26	68
R_CA_K	A	6	20	24	22	27	73
R_CA_K	B	7	25	27	15	26	68
R_CA_MG	A	4	16	25	20	35	80
R_CA_MG	B	5	19	25	20	31	76
SAND	A	4	12	16	17	52	85
SAND	B	5	13	17	18	47	82
SILT	A	4	12	16	17	52	85
SILT	B	5	12	18	18	47	83
TEB	A	7	16	27	22	27	76
TEB	B	7	22	28	18	26	72
TOTN	A	5	13	21	24	38	83
TOTN	B	5	14	27	22	32	81
TOTPORES	A	7	18	29	20	25	74
TOTPORES	B	7	19	30	21	23	74

See top of next page for Table 'footnotes'.

Notes for Table 6 and 7. Sample populations by attribute are stratified by soil unit and depth zone, i.e. for all available textural classes (category #). 'A' stands for topsoil and 'B' for subsoil. Frequency of occurrence refers to the number of cases for which either No (0), Few (1-4), Common (5-14), Many (15-29) and Abundant (>30) observations are available for the specified attribute, and is expressed as a percentage of the total number of soil units (i.e. 106 in 1974-Legend and 153 for the 1988-Legend). CMA stands for sum of frequencies for categories C, M and A. Totals for N+F+C+M+A may differ from 100 in places due to rounding. Taxotransfer rules will be used in cases the frequency of occurrence is either 'N' or 'F' for any given combination of soil unit, textural class, depth zone and soil chemical or physical attribute.

Summary files are also available that present values for the 28 soil parameters, for each unique combination of soil unit, depth zone and topsoil textural class, in one 'single' line or record. These files have been created to facilitate input into model applications and linkage to the digital Soil Map of the World or SOTER databases. Summary files (i.e., SUMTAB74.dbf and SUMTAB90.dbf) give the actual number of decimal places considered justified for presenting results of the analyses, by attribute. All other files, however, give two decimal places by default when presenting interim results (see Appendix 5).

Overall confidence in the results shown should be highest where the degrees of freedom are highest, under the assumption that all available profiles are equally representative for the corresponding combination of soil unit, topsoil textural class, and depth zone.

3.5 Possible limitations

In view of the default depth ranges adopted for the topsoil (0-30 cm) and the subsoil (30-100 cm) the occurrence of several chemical properties — such as a high aluminium saturation in the deeper subsoil only or a high exchangeable sodium percentage in parts of the topsoil — may be 'levelled out' during depth-weighting for some profiles. When occurring, this will also be reflected in derived values obtained by taxotransfer (when NUM < 5).

Results may be biased for those attributes that have not been standardly recorded as being "not observed" or "nil" in the original source, for example gravel content. In such cases, the means and medians computed here may well give a biased impression of 'modal' conditions for certain soil units as, probably, only the limiting gravel contents may have been documented in field surveys.

In some cases medians for TAWC₂ are larger than for TAWC₁ for a certain combination of soil unit, topsoil textural class and depth zone, especially when the number of profiles available for the analyses is small. Clearly, this is in apparent contradiction with pedological reality! The discrepancy, however, can be explained by the fact that there are only few profiles for which volumetric water content was measured both at - 5 kPa, -10 kPa, -33 kPa and -1500 kPa. Also the FAO textural classes are fairly wide and it was not possible to account for differences in composition of the sand fraction, for example. Further, the adopted segregation into two broad depth zones may not permit to account adequately for the contribution of organic matter and differences in clay activity to water retention.

An alternative framework for deriving water-holding capacity was tested that considers particle size distribution, organic matter content and inferred clay activity, as initially proposed by the Global Soil Data Task Force (see Scholes, 1994). Although preliminary results were promising, this approach proved not very useful in the current context. In case of missing water retention data, for example, measured or TTR-derived data on soil texture, inferred clay activity (which in turns requires data on CEC_{soil}, clay %, organic matter content, and inferred activity of the organic matter fraction to be available) are required to implement the TTR framework. This makes application of such a scheme much more complex, data demanding, and less reliable than the one currently in use.

Table 8. Overview of type of TTRs used and their frequency (1974 Legend)

ATTRIB	DEPTH	R0	R1	R2	R3	R4	R5	R6	R7	R8	R9	T0	T1	AN	OR
ALSAT	A	75	5	0	0	0	0	10	4	0	4	0	0	0	1
ALSAT	B	73	5	0	0	0	0	12	4	0	4	0	0	0	1
APARCEC	A	73	20	0	0	0	0	0	0	0	5	0	0	0	1
APARCEC	B	72	21	0	0	0	0	0	0	0	6	0	0	0	1
AWC1	A	6	25	0	0	0	0	0	0	0	63	0	0	4	3
AWC1	B	5	26	0	0	0	0	0	0	0	62	0	0	4	3
AWC2	A	15	40	0	0	0	0	0	0	0	41	0	0	2	3
AWC2	B	17	39	0	0	0	0	0	0	0	40	0	0	2	3
AWC3	A	39	38	0	0	0	0	0	0	0	22	0	0	0	1
AWC3	B	37	40	0	0	0	0	0	0	0	22	0	0	0	1
BSAT	A	61	28	0	0	0	0	0	0	0	9	0	0	0	1
BSAT	B	58	30	0	0	0	0	0	0	0	10	0	0	0	1
BULKDENS	A	63	26	0	0	0	0	0	0	0	10	0	0	0	1
BULKDENS	B	62	26	0	0	0	0	0	0	0	11	0	0	0	1
CACO3	A	74	7	9	6	0	0	0	0	0	4	0	0	0	1
CACO3	B	72	6	8	8	0	0	0	0	0	4	0	0	0	1
CECCLAY	A	71	21	0	0	0	0	0	0	0	7	0	0	0	1
CECCLAY	B	70	22	0	0	0	0	0	0	0	7	0	0	0	1
CECSOIL	A	74	20	0	0	0	0	0	0	0	4	0	0	0	1
CECSOIL	B	73	21	0	0	0	0	0	0	0	5	0	0	0	1
CLAY	A	0	0	0	0	0	0	0	0	0	0	75	25	0	0
CLAY	B	0	0	0	0	0	0	0	0	0	0	74	26	0	0
C_N	A	68	23	0	0	0	0	0	0	0	8	0	0	0	1
C_N	B	63	26	0	0	0	0	0	0	0	10	0	0	0	1
ECE	A	71	22	0	0	0	0	0	0	0	5	0	0	0	1
ECE	B	69	23	0	0	0	0	0	0	0	7	0	0	0	1
ECEC	A	50	34	0	0	0	0	0	0	0	14	0	0	0	1
ECEC	B	49	35	0	0	0	0	0	0	0	14	0	0	0	1
ESP	A	65	13	0	0	0	14	0	0	0	7	0	0	0	1
ESP	B	63	14	0	0	0	15	0	0	0	7	0	0	0	1
GRAVEL	A	39	41	0	0	0	0	0	0	0	16	0	0	1	3
GRAVEL	B	41	40	0	0	0	0	0	0	0	15	0	0	1	3
GYPSSUM	A	62	17	0	0	11	0	0	0	0	9	0	0	0	1
GYPSSUM	B	60	19	0	0	10	0	0	0	0	9	0	0	0	1
ORGC	A	73	20	0	0	0	0	0	0	0	6	0	0	0	1
ORGC	B	72	21	0	0	0	0	0	0	0	6	0	0	0	1
PHCACL2	A	40	38	0	0	0	0	0	0	0	20	0	0	1	1
PHCACL2	B	38	40	0	0	0	0	0	0	0	20	0	0	1	1
PHH2O	A	90	5	0	0	0	0	0	0	0	4	0	0	0	1
PHH2O	B	89	5	0	0	0	0	0	0	0	4	0	0	0	1
PHKCL	A	57	29	0	0	0	0	0	0	0	13	0	0	0	1
PHKCL	B	55	30	0	0	0	0	0	0	0	13	0	0	0	1
R_CAMG_K	A	63	26	0	0	0	0	0	0	0	9	0	0	0	1
R_CAMG_K	B	55	33	0	0	0	0	0	0	0	10	0	0	0	1
R_CA_K	A	63	26	0	0	0	0	0	0	0	9	0	0	0	1
R_CA_K	B	56	33	0	0	0	0	0	0	0	9	0	0	0	1
R_CA_MG	A	69	21	0	0	0	0	0	0	0	8	0	0	0	1
R_CA_MG	B	66	25	0	0	0	0	0	0	0	8	0	0	0	1
SAND	A	0	0	0	0	0	0	0	0	0	0	75	25	0	0
SAND	B	0	0	0	0	0	0	0	0	0	0	74	26	0	0
SILT	A	0	0	0	0	0	0	0	0	0	0	75	25	0	0
SILT	B	0	0	0	0	0	0	0	0	0	0	74	26	0	0
TEB	A	62	27	0	0	0	0	0	0	0	9	0	0	0	1
TEB	B	60	29	0	0	0	0	0	0	0	9	0	0	0	1
TOTN	A	70	21	0	0	0	0	0	0	0	8	0	0	0	1
TOTN	B	65	25	0	0	0	0	0	0	0	9	0	0	0	1
TOTPORES	A	63	25	0	0	0	0	0	0	0	10	0	0	0	1
TOTPORES	B	62	25	0	0	0	0	0	0	0	11	0	0	0	1
TAWC1	P	8	20	0	0	0	0	0	0	72	0	0	0	0	0
TAWC2	P	21	52	0	0	0	0	0	0	47	0	0	0	0	0
TAWC3	P	42	31	0	0	0	0	0	0	27	0	0	0	0	0

Note: 'A' stands for 0-30 cm, except for Lithosols where 'A' is from 0 to 10 cm, and 'B' for 30-100 cm. 'P' stands for maximum profile depth (see section 2.7).

Table 9. Overview of type of TTRs used and their frequency (1988 Legend)

ATTRIB	DEPTH	R0	R1	R2	R3	R4	R5	R6	R7	R8	R9	T0	T1	AN	OR
ALSAT	A	63	10	0	0	0	0	14	7	0	4	0	0	0	2
ALSAT	B	61	10	0	0	0	0	16	7	0	4	0	0	0	2
APARCEC	A	63	28	0	0	0	0	0	0	0	7	0	0	0	2
APARCEC	B	62	31	0	0	0	0	0	0	0	6	0	0	0	2
AWC1	A	4	24	0	0	0	0	0	0	0	66	0	0	4	3
AWC1	B	3	21	0	0	0	0	0	0	0	68	0	0	4	3
AWC2	A	11	38	0	0	0	0	0	0	0	46	0	0	2	3
AWC2	B	12	37	0	0	0	0	0	0	0	46	0	0	2	3
AWC3	A	28	47	0	0	0	0	0	0	0	24	0	0	0	2
AWC3	B	27	49	0	0	0	0	0	0	0	23	0	0	0	2
BSAT	A	49	39	0	0	0	0	0	0	0	10	0	0	0	2
BSAT	B	46	39	0	0	0	0	0	0	0	13	0	0	0	2
BULKDENS	A	48	40	0	0	0	0	0	0	0	10	0	0	0	2
BULKDENS	B	47	40	0	0	0	0	0	0	0	12	0	0	0	2
CACO3	A	62	12	14	7	0	0	0	0	0	4	0	0	0	2
CACO3	B	60	12	13	9	0	0	0	0	0	4	0	0	0	2
CECCLAY	A	60	31	0	0	0	0	0	0	0	7	0	0	0	2
CECCLAY	B	59	33	0	0	0	0	0	0	0	7	0	0	0	2
CECSOIL	A	64	30	0	0	0	0	0	0	0	5	0	0	0	2
CECSOIL	B	63	30	0	0	0	0	0	0	0	5	0	0	0	2
CLAY	A	0	0	0	0	0	0	0	0	0	0	66	34	0	0
CLAY	B	0	0	0	0	0	0	0	0	0	0	64	36	0	0
C_N	A	57	33	0	0	0	0	0	0	0	8	0	0	0	2
C_N	B	51	39	0	0	0	0	0	0	0	9	0	0	0	2
ECE	A	60	33	0	0	0	0	0	0	0	5	0	0	0	2
ECE	B	57	36	0	0	0	0	0	0	0	5	0	0	0	2
ECEC	A	39	46	0	0	0	0	0	0	0	14	0	0	0	2
ECEC	B	36	44	0	0	0	0	0	0	0	18	0	0	0	2
ESP	A	51	20	0	0	0	20	0	0	0	7	0	0	0	2
ESP	B	50	20	0	0	0	21	0	0	0	7	0	0	0	2
GRAVEL	A	30	47	0	0	0	0	0	0	0	19	0	0	1	3
GRAVEL	B	31	47	0	0	0	0	0	0	0	18	0	0	1	3
GYP SUM	A	52	23	0	0	16	0	0	0	0	8	0	0	0	2
GYP SUM	B	49	23	0	0	14	0	0	0	0	12	0	0	0	2
ORGC	A	64	30	0	0	0	0	0	0	0	5	0	0	0	2
ORGC	B	61	32	0	0	0	0	0	0	0	6	0	0	0	2
PHCACL2	A	30	46	0	0	0	0	0	0	0	21	0	0	1	2
PHCACL2	B	26	46	0	0	0	0	0	0	0	25	0	0	1	2
PHH2O	A	84	10	0	0	0	0	0	0	0	4	0	0	0	2
PHH2O	B	84	10	0	0	0	0	0	0	0	4	0	0	0	2
PHKCL	A	45	43	0	0	0	0	0	0	0	10	0	0	0	2
PHKCL	B	44	44	0	0	0	0	0	0	0	11	0	0	0	2
R_CAMG_K	A	50	39	0	0	0	0	0	0	0	9	0	0	0	2
R_CAMG_K	B	44	43	0	0	0	0	0	0	0	12	0	0	0	2
R_CA_K	A	51	38	0	0	0	0	0	0	0	9	0	0	0	2
R_CA_K	B	44	43	0	0	0	0	0	0	0	11	0	0	0	2
R_CA_MG	A	56	35	0	0	0	0	0	0	0	7	0	0	0	2
R_CA_MG	B	53	35	0	0	0	0	0	0	0	10	0	0	0	2
SAND	A	0	0	0	0	0	0	0	0	0	0	66	34	0	0
SAND	B	0	0	0	0	0	0	0	0	0	0	63	37	0	0
SILT	A	0	0	0	0	0	0	0	0	0	0	66	34	0	0
SILT	B	0	0	0	0	0	0	0	0	0	0	64	36	0	0
TEB	A	50	40	0	0	0	0	0	0	0	8	0	0	0	2
TEB	B	47	41	0	0	0	0	0	0	0	10	0	0	0	2
TOTN	A	59	31	0	0	0	0	0	0	0	8	0	0	0	2
TOTN	B	55	35	0	0	0	0	0	0	0	9	0	0	0	2
TOTPORES	A	49	40	0	0	0	0	0	0	0	10	0	0	0	2
TOTPORES	B	47	40	0	0	0	0	0	0	0	12	0	0	0	2
TAWC1	P	5	17	0	0	0	0	0	0	78	0	0	0	0	0
TAWC2	P	15	30	0	0	0	0	0	0	55	0	0	0	0	0
TAWC3	P	33	39	0	0	0	0	0	0	28	0	0	0	0	0

Note: 'A' stands for 0-30 cm, except for Lithic Leptosols where 'A' is from 0 to 10 cm, and 'B' for 30-100 cm. 'P' stands for maximum profile depth (see section 2.7).

4 CONCLUSIONS

This paper presents revised statistics for 28 physical and chemical properties for each soil unit considered in the FAO-Unesco '1974 Legend' and '1988 Revised Legend'. The analysis is based on over 9600 profile descriptions. By comparison the initial assessment of derived soil properties for the Soil Map of the World was based on only 1700 profiles (FAO, 1995), most of which have been included in WISE-1.

The compilation of representative soil profiles for the major soils of the world, as described by the Legend and Revised Legend, is in essence an open-ended activity. In the context of the current IFPRI-funded project, the geographical focus in expanding the profile data set has been on soils from developing countries. In addition, ISRIC's focus is mainly on soils of tropical and subtropical regions. Consequently, there are still relatively few profile observations for several sections of the world, including Australia, South East Asia, and North America, for which continental SOTER databases have not yet been developed (Nachtergaele, 1999; Van Engelen, 1999).

From a taxonomic perspective there still are no or only a limited number of observations – defined here as less than 10 profiles – for the following units of the original Legend: gleyic Chernozems; gleyic Greyzems; gelic Histosols; ferric Podzols; humic and gelic Planosols; takyric Yermosols; and mollic and takyric Solonchaks. With respect to the Revised Legend the following units are under-represented: plinthic Alisols; gleyic and gelic Andosols; aric, fimic and urbic Anthrosols; gleyic and glossic Chernozems; andic and thionic Gleysols; gleyic Greyzems; luvic Gypsisols; gelic, folic and thionic Histosols; gypsic Kastanozems; gelic Leptosols; stagnic Lixisols; gelic and stagnic Podzoluvisols; gelic and umbric Planosols; humic Plinthosols; ferric Podzols; gelic and gypsic Regosols; gelic and mollic Solonchaks; gypsic Solonetz; and gypsic Vertisols.

There are often omissions or gaps in the information provided in the soil literature or in available auxiliary soil digital databases with respect to several of the input variables required by the WISE database. As very few new soil surveys are being undertaken, it is unlikely that gaps for the under-represented attributes can easily be filled in data compilation activities that do not include new soil surveys. In particular, there are very few measured data on soil water retention.

When few measured data were available (or remained after the rigorous data integrity checks and subsequent outlier rejection scheme) a system of taxotransfer rules was applied. Inherently, such rules are fraught with uncertainty. The taxotransfer rules and the results obtained through their application must be viewed as the 'best possible' estimates based on the currently available selection of soil profiles and data clustering system used. Although a particular derived attribute may get a high 'confidence' rating, this does not necessarily imply that this derived value will be representative for the soil unit under consideration. Profile selection for WISE, like for other global profile databases, is not probabilistic but based on available data. Also, several of the attributes considered in this study are not diagnostic in the FAO-Unesco Legend. Therefore, it is recommended that users carefully evaluate such TTR-based estimates prior to using them for specific model applications.

Analysis of soil data is sensitive to the classification of the data. Clustering of the data in this study is in accordance with conventions of FAO (1995), and data requirements for global and regional AEZ studies by IIASA and FAO (Fischer *et al.*, 2000; Fischer *et al.*, 2001). Current results give a straightforward summary of the available data, using classes for the factors that can be linked to the spatial units shown on the digital Soil Map of the World. The dataset of derived soil parameters presented for use with the 1988 Legend can be linked to all spatial data sets that use this legend (e.g., FAO and ISRIC, 2000; FAO *et al.*, 1998). Thereby the summary files can be used in various

GIS-based studies, for example soil gaseous emission potentials, soil vulnerability to pollution, and crop productivity.

The current set of derived soil parameters is considered appropriate for making analyses at a regional to global scale (< 1:250,000). Correlation of soil analytical data, however, must be done more accurately when more precise scientific research is considered

5 ACKNOWLEDGEMENTS

This study was commissioned by the International Food Policy Research Institute (IFPRI) as Research Agreement Contract No. 2001X020.ISR, using funds from the European Economic Commission. Constructive comments by Stanley Wood (IFPRI) and Vincent van Engelen (ISRIC) are gratefully acknowledged.

REFERENCES

- Alcamo, J., Kreileman, E., Krol, M., Leemans, R., Bollen, J., van Minnen, J., Schaeffer, M., Toet, S. and de Vries, B., 1998. Global modelling of environmental change: an overview of IMAGE 2.1. In: J. Alcamo, R. Leemans and E. Kreileman (Editors), *Global Change Scenarios of the 21st Century. Results from the IMAGE 2.1 Model*. Elsevier, Amsterdam, pp. 19-21.
- Batjes, N.H., 1995. *World Inventory of Soil Emission Potentials: WISE 2.1 - Database user's manual and coding protocols*. Tech. Pap. 26, ISRIC, Wageningen.
- Batjes, N.H., 1997. A world data set of derived soil properties by FAO-UNESCO soil unit for global modelling. *Soil Use and Management*, 13: 9-16.
- Batjes, N.H., 2000a. Effects of mapped variation in soil conditions on estimates of soil carbon and nitrogen stocks for South America. *Geoderma*, 97(1-2): 135-144.
- Batjes, N.H., 2000b. SOTER summary file for Central and Eastern Europe (SOVEUR Project; *ver. 1.0*). Report 2000/06, ISRIC, Wageningen.
- Batjes, N.H., 2002. A homogenized soil profile data set for global and regional environmental research. ISRIC Report 2002/01 [Available online, with derived data sets, at <http://www.isric.org>], International Soil Reference and Information Centre (ISRIC), Wageningen.
- Batjes, N.H. and Bridges, E.M., 1994. Potential emissions of radiatively active gases from soil to atmosphere with special reference to methane: development of a global database (WISE). *Journal of Geophysical Research*, 99(D8): 16479-16489.
- Batjes, N.H., Bridges, E.M. and Nachtergaele, F.O., 1995. World Inventory of Soil Emission Potentials: development of a global soil data base of process-controlling factors. In: S. Peng, K.T. Ingram, H.U. Neue and L.H. Ziska (Editors), *Climate Change and Rice*. Springer-Verlag, Heidelberg, pp. 102-115.
- Batjes, N.H., Fischer, G., Nachtergaele, F.O., Stolbovoy, V.S. and van Velthuisen, H.T., 1997. Soil data derived from WISE for use in global and regional AEZ studies (*ver. 1.0*) [Available on-line via <http://www.iiasa.ac.at>]. Interim Report IR-97-025, FAO/ IIASA/ ISRIC, Laxenburg.
- Baumgardner, M.F., 1999. Soil databases. In: M.E. Sumner (Editor), *Handbook of Soil Science*. CRC Press, Boca Raton, pp. H:1-40.
- Bouma, J., Batjes, N.H. and Groot, J.J.R., 1998. Exploring land quality effects on world food supply. *Geoderma*, 86: 43-59.
- Bouwman, A.F., Boumans, L.J.M. and Batjes, N.H., 2002. Modeling global annual N₂O and NO emissions from fertilized fields. *Global Biogeochemical Cycles* (in press).
- Bouwman, A.F. and Van Vuuren, D., 1999. Global assessment of acidification and eutrophication of natural ecosystems. Environment Information and Assessment Technical Report 6 (UNEP/DEIA&EW/TR.99-6; RIVM/4002001012), UNEP and RIVM, Bilthoven.
- Bregt, A.K. and Beemster, J.H.G., 1989. Accuracy in predicting moisture deficits and changes in yield from soil maps. *Geoderma*, 43: 301-310.
- Breimer, R.F., Van Kekem, A.J. and Van Reuler, H., 1986. Guidelines for soil survey and land evaluation in ecological research. MAB Series 17, UNESCO, Paris.
- Brunt, J. and Van Reeuwijk, L.P., 1997. SOILIMS - Laboratory Information System for Soil and Plant Laboratories: Manual and Tutor (*ver. 1.4*). Technical Paper 24, ISRIC, Wageningen.
- Burrough, P.A., 1986. *Principles of Geographical Information Systems*. Oxford, Oxford University press.
- CEC, 1985. *Soil Map of the European Communities (1:1,000,000)*. Report EUR 8982, Office for Official Publications of the European Communities, Luxembourg.
- Cindery, S., Cambridge, H.M., Herrera, R., Hicks, W.K., Kuylenstierna, J.C.I., Murray, F. and Olbrich, K., 1998. Global assessment of ecosystem sensitivity to acidic deposition. Stockholm Environmental Institute, York, pp. 19.

- Denier van der Gon, H.A.C., Bodegom, P.M.v., Houweling, S., Verburg, P.H. and Van Breemen van, 2000. Combining upscaling and downscaling of methane emissions from rice fields: methodologies and preliminary results. *Nutrient Cycling in Agroecosystems*, 58: 285-301.
- DFG, 1999. Atlas of agronomic resources of Niger and Benin. (CD ROM), Deutsche Forschungsgemeinschaft and University of Hohenheim, Hohenheim.
- Doorenbos, J. and Kassam, A.H., 1978. Yield response to water. *Irrigation and Drainage Paper 33*, FAO, Rome.
- ESB, 1999. European Soil Database (scale 1:1,000,000; ver. 1.0), European Soil Bureau (on behalf of the Contributing Organisations), Ispra.
- FAO, 1977. Guidelines for the description of soils. FAO, Rome.
- FAO, 1978-1981. Reports of the Agro-ecological Zones Project, Vol. 1, FAO, Rome.
- FAO, 1988. FAO/Unesco Soil Map of the World, Revised Legend (with corrections and updates), FAO World Soil Resources Report 60 (reprinted with updates as ISRIC Technical Paper 20 in 1997). ISRIC, Wageningen, 140 pp.
- FAO, 1995. Digital Soil Map of the World and Derived Soil Properties. Food and Agriculture Organization of the United Nations, FAO.
- FAO and ISRIC, 2000. Soil and terrain database, soil degradation status, and soil vulnerability assessments for Central and Eastern Europe (scale 1:2.5 million; ver. 1.0). Land and Water Digital Media Series 10, FAO, Rome.
- FAO, ISRIC, UNEP and CIP, 1998. Soil and terrain digital database for Latin America and the Caribbean at 1:5 million scale. Land and Water Digital Media Series No. 5, Food and Agriculture Organization of the United Nations, Rome.
- FAO-Unesco, 1974. Soil Map of the World, 1:5,000,000. Vol. 1 - Legend. United Nations Educational, Scientific, and Cultural Organization, Paris.
- FAO-Unesco, 1974-1981. Soil Map of the World, 1:5,000,000. Vol. 1 to 10. United Nations Educational, Scientific, and Cultural Organization, Paris.
- Finke, P., Hartwich, R., Dudal, R., Ibanez, J., Jamagne, M., King, D., Montanarella, L. and Yassoglou, N., 1998. Georeferenced soil database for Europe. Manual of procedures (ver. 1.0). Report EUR 18092, European Soil Bureau, Joint Research Centre, Ispra.
- Fischer, G., Van Velthuizen, H.T. and Nachtergaele, F.O., 2000. Global Agro-ecological Zones Assessment: Methodology and results. IR-00-064, Food and Agriculture Organization of the United Nations (FAO) and International Institute for Applied Systems Analysis (IIASA), Laxenburg.
- Fischer, G., Van Velthuizen, H.T. and Prieler, S., 2001. Assessment of potential productivity of tree species in China, Mongolia and the Former Soviet Union: Methodology and Results. IR-01-015, International Institute for Applied Systems Analysis (IIASA), Laxenburg.
- Ganzenveld, L., Lelieveld, J. and Roelofs, G.-J., 1998. A dry deposition parameterization for sulfur oxides in a chemistry and general circulation model. *Journal of Geophysical Research*, 103(D5): 5679-5694.
- Goodchild, M.F., 1994. Sharing imperfect data. In: A. Singh (Editor), Proceedings UNEP and IUFPRO international workshop in cooperation with FAO in developing large environmental databases for sustainable development. UNEP/GRID, Sioux Falls, pp. 102-110.
- Graef, F., 1999. Evaluation of agricultural potentials in semi-arid SW-Niger: a soil and terrain (NiSOTER) study. Heft 54 Thesis, Universität Hohenheim, Hohenheim, 217 pp.
- Hootsmans, R.M., Bouwman, A.F., Leemans, R. and Kreileman, G.J.J., 2001. Modelling land degradation in IMAGE 2. RIVM Report 481508009, National Institute of Public Health and the Environment, Bilthoven.
- ICRAF, 1995. Uganda Digital Soil Map (1:250,000), International Council for Research in Agroforestry (ICRAF), Entebbe.

- IGBP-DIS, 2000. Global Soil Data Products CD-ROM (IGBP-DIS). IGBP Data and Information Services, Potsdam.
- Ingram, J. and Gregory, P. (Editors), 1996. Effects of Global Change on Soils: Implementation Plan. International Geosphere-Biosphere Programme (IGBP), Wallingford, 55 pp.
- ISSS, 1983. Soil Map of the Middle Europe (1:1,000,000), ECSC-EEC-EAEC and International Society of Soil Science, Wageningen.
- Kamprath, E.J. (Editor), 1986. Soil fertility evaluation. Proceedings of an International Workshop on the laboratory Methods and Data Exchange Programme (LABEX), Technical Paper 13. ISRIC, Wageningen, 128-138 pp.
- Kauffman, J.H., Mantel, S. and Spaargaren, O.C., 1996. Soils of the humid and seasonally dry (sub)tropics: A correlation of reference soil data and their assessment for agriculture using the ISIS database. In: J.H. Kauffman (Editor), National Soil Reference Collections and Databases (NASREC). ISRIC, Wageningen, pp. 1-41.
- Klamt, E. and Sombroek, W.G., 1988. Contribution of organic matter to exchange properties of Oxisols. In: F.H. Beinroth, M.N. Camargo and H. Eswaran (Editors), Classification, characterization and utilization of Oxisols. Proc. of the 8th International Soil Classification Workshop (Brazil, 12 to 23 May 1986). Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA), Soil Management Support Services (SMSS) and University of Puerto Rico (UPR), Rio de Janeiro, pp. 64-70.
- Kern, J.S., 1995. Geographic patterns of soil-water holding capacity in the contiguous United States. *Journal of the American Society for Horticultural Science*, 59(1126-1133).
- Knox, J.W., Matthews, R.B. and Wassmann, R., 2000. Using a crop/soil simulation model and GIS techniques to assess methane emissions from rice fields in Asia. III: Databases. *Nutrient Cycling in Agroecosystems*, 58: 179-199.
- Landon, J.R., 1991. *Booker Tropical Soil Manual*. Longman Scientific & Technical, New York.
- Luyten, J.C., 1995. Sustainable world food production and environment. Internal Report No. 37, DLO Institute for Agrobiological and Soil Fertility Research, Wageningen.
- Madsen, B.H. and Jones, R.J.A., 1995. The establishment of a soil profile analytical database for the European Union. In: D. King, R.J.A. KJones and A.J. Thomasson (Editors), *European Land Information for Agro-environmental Monitoring*. Office for Official Publications of the European Community, Luxembourg, pp. 55-63.
- Mantel, S., Samsudin, S. and Tyrie, G., 1999. Inventory of site qualities for forest management planning. In: Y. Laumonier, B. King, C. Legg and K. Rennolls (Editors), *Data management and modelling using remote sensing and GIS for tropical forest land inventory*. Rodeo International Publishers, Jakarta, pp. 383-395.
- Nachtergaele, F.O., 1999. From the Soil Map of the World to the Digital Global Soil and Terrain Database: 1960-2002. In: M.E. Sumner (Editor), *Handbook of Soil Science*. CRC Press, Boca Raton, pp. H5-17.
- Nettleton, W.D., Hornsby, A.G., Brown, R.B. and Coleman, T.L. (Editors), 1996. Data reliability and risk assessment in soil interpretations, SSSA Special Publication No. 47. Soil Science Society of America, Madison WI.
- NSS, 1997. Soil and Terrain (SOTER) database for Tanzania (1:2 million scale; version 1.0), National Soil Service (NSSS) of Tanzania & AGLS and GIS unit of FAO, Dar Es Salaam.
- Pleijzier, K., 1989. Variability in soil data. In: J. Bouma and A.K. Bregt (Editors), *Land Qualities in Space and Time*. PUDOC, Wageningen, pp. 89-98.
- Pleijzier, L.K. (Editor), 1986. The LABEX programme and its results so far. Proceedings of an International Workshop on the Laboratory Methods and Data Exchange Programme (LABEX), Technical Paper 13. ISRIC, Wageningen, 6-15 pp.
- Pleijzier, J.L., Janssens, J. and Cremers, A. (Editors), 1986. Extraction of cations from some kaolinitic soils of the tropics. Proceedings of an International Workshop on the laboratory

- Methods and Data Exchange Programme (LABEX), Technical Paper 13. ISRIC, Wageningen, 51-66 pp.
- Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Momserud, R.A. and Solomon, A.M., 1992. A global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.*, 19: 117-134.
- Rousseva, S.S., 1997. Data transformations between soil texture schemes. *European Journal of Soil Science*, 48: 749-758.
- Scholes, R.J., Skole, D. and Ingram, J.S., 1995. A global database of soil properties: proposal for implementation. IGBP-DIS Working Paper 10, International Geosphere Biosphere Program, Data & Information System, Paris.
- Scheffer, F. and Schachtschabel, P., 1984. *Lehrbuch der Bodenkunde*. Ferdinand Enke Verlag, Stuttgart.
- SCS, 1967. Soil survey laboratory methods and procedures for collecting soil samples. Soil Survey investigation No. 1, US Soil Conservation Service, Washington, DC.
- Shouse, P.J., Russel, W.B., Romberger, J.S. and Papendick, R.I., 1995. Spatial variability of soil water retention functions in a silt loam. *Soil Science*, 159: 1-12.
- Snedecor, G.W. and Cochran, W.G., 1980. *Statistical Methods* (7th. ed). The Iowa State University Press, Iowa, 507 pp.
- Soil Survey Staff, 1983. *Soil Survey Manual* (rev. ed.). United States Agriculture Handbook 18, USDA, Washington.
- Soil Survey Staff, 1992. *Keys to Soil Taxonomy* (5th. ed.), SMSS Technical Monograph No. 19. Agency for International Development, United States Department of Agriculture, Soil Conservation Service, Soil Management Support Services. Pocahontas Press, Inc., Blacksburg (Virginia).
- Sombroek, W.G. (Editor), 1986. Introduction. Proceedings of an International Workshop on the laboratory Methods and Data Exchange Programme (LABEX), Technical Paper 13. ISRIC, Wageningen, 5 pp.
- Van Engelen, V.W.P., 1999. SOTER: The World Soils and Terrain Database. In: M.E. Sumner (Editor), *Handbook of Soil Science*. CRC Press, Boca Raton, pp. H19-28.
- Van Engelen, V.W.P. and Wen, T.T., 1995. *Global and National Soils and Terrain Digital Databases (SOTER): Procedures Manual*. (Published also as FAO World Soil Resources Report No. 74), UNEP, IUSS, ISRIC, FAO, Wageningen.
- Van Reeuwijk, L.P., 1982. Laboratory methods and exchange programme for soil characterization: A report on the pilot round, Part I: CEC and texture. Technical Paper No. 6, ISRIC, Wageningen.
- Van Reeuwijk, L.P., 1983. On the way to improve international soil classification and correlation: the variability of soil analytical data. Annual Report 1983, ISRIC, Wageningen.
- Van Reeuwijk, L.P., 1984. Laboratory methods and exchange programme for soil characterization: A report on the pilot round, Part II: Exchangeable bases, base saturation and pH. Technical Paper No. 8, ISRIC, Wageningen.
- Van Reeuwijk, L.P., 1993. *Procedures for soil analysis* (4th ed.). Technical Paper 9, ISRIC, Wageningen.
- Van Reeuwijk, L.P., 1998. Guidelines for Quality management in Soil and Plant Laboratories, *Soils Bulletin* 74. FAO, Rome, 143 pp.
- Van Waveren, E.J. and Bos, A.B., 1988. ISRIC Soil Information System - User manual and technical manual. Technical Paper 15 (rev. ed.), ISRIC, Wageningen.
- Velthuisen, H.v., 1998. Environmental Information System Development Project: Ghana Soil Information System 1-3, Environmental Protection Agency (EPA), Remote Sensing Application Unit (RSAU), Soil Research Institute (SRI), Meteorological Services Lands Commission (MSLC), and Survey Department (SD), Kumasi.

- Vogel, A.W., 1994. Comparability of soil analytical data: determinations of cation exchange capacity, organic carbon, soil reaction, bulk density, and volume percentage of water at selected pF values by different methods. Work. Pap. 94/07, ISRIC, Wageningen.
- Webb, S.R., Rosenzweig, C.E. and Levine, E.R., 1991. A global data set of soil particle size properties. Technical Memorandum 4286, NASA, New York.
- Wood, S., Sebastian, K. and Scherr, S.J., 2000. Pilot Analysis of Global Ecosystems: Agroecosystems. World Resources Institute (WRI) and International Food Policy Research Institute (IFPRI), Washington, 110 pp.
- Wösten, J.H.M., Lilly, A., Nemes, A. and Le Bas, C., 1998. Using existing soil data to derive hydraulic parameters for simulation models in environmental studies and in land use planning. Report 156, DLO-Staring Centre, Wageningen.
- WRB, 1998. World Reference Base for Soil Resources. World Soil Resources Report 84, ISSS, ISRIC, and FAO, Rome.
- Zobler, L., 1986. A world soil file for global climate modeling. Technical Memorandum 87802, NASA Goddard Institute for Space Studies (GISS), New York, NY.
- Zuidema, G., Van den Born, G.J., Alcamo, J. and Kreileman, G.J.J., 1994. Simulation of global land cover changes as affected by economic factors and climate. *Water, Air, and Soil Pollution*, 76: 163-198.

APPENDICES

Appendix 1. Number and classification of profiles used for the analyses (1974 Legend)

A: Acrisols (882)	Af= 302 Ag= 42 Ah= 99 Ao= 348 Ap= 91
B: Cambisols (1293)	Bc= 90 Bd= 218 Be= 361 Bf= 113 Bg= 118 Bh= 106 Bk= 187 Bv= 76 Bx= 24
C: Chernozems (188)	Cg= 2 Ch= 84 Ck= 63 Cl= 39
D: Podzoluvisols (107)	Dd= 11 De= 77 Dg= 19
E: Rendzinas (129)	E = 129
F: Ferralsols (496)	Fa= 30 Fh= 70 Fo= 173 Fp= 14 Fr= 94 Fx= 115
G: Gleysols (726)	Gc= 24 Gd= 138 Ge= 309 Gh= 62 Gm= 110 Gp= 69 Gx= 14
H: Phaeozems (403)	Hc= 37 Hg= 33 Hh= 150 Hl= 183
I: Lithosols (71)	I = 71
J: Fluvisols (599)	Jc= 166 Jd= 104 Je= 292 Jt= 37
K: Kastanozems (74)	Kh= 34 Kk= 25 Kl= 15
L: Luvisols (1396)	La= 35 Lc= 228 Lf= 296 Lg= 185 Lk= 174 Lo= 419 Lp= 20 Lv= 39
M: Greyzems (28)	Mg= 4 Mo= 24
N: Nitosols (162)	Nd= 65 Ne= 62 Nh= 35
O: Histosols (171)	Od= 83 Oe= 81 Ox= 6
P: Podzols (179)	Pf= 5 Pg= 36 Ph= 40 Pl= 17 Po= 67 Pp= 14
Q: Arenosols (677)	Qa= 35 Qc= 366 Qf= 148 Ql= 128
R: Regosols (321)	Rc= 52 Rd= 83 Re= 178 Rx= 8
S: Solonetz (180)	Sg= 55 Sm= 24 So= 101
T: Andosols (258)	Th= 135 Tm= 48 To= 30 Tv= 45
U: Rankers (107)	U = 107
V: Vertisols (507)	Vc= 192 Vp= 315
W: Planosols (162)	Wd= 31 We= 83 Wh= 6 Wm= 15 Ws= 27 Wx= 0
X: Xerosols (280)	Xh= 42 Xk= 109 Xl= 91 Xy= 38
Y: Yermosols (74)	Yh= 19 Yk= 15 Yl= 20 Yt= 1 Yy= 19
Z: Solonchaks (123)	Zg= 36 Zm= 7 Zo= 74 Zt= 6

Total number of profiles is 9607, of which 9593 can be classified according to the FAO-Unesco (1974) Legend.

Appendix 2. Number and classification of soil profiles used for the analyses (1988 Legend)

AC: Acrisols (669)
 ACf= 175 ACg= 21 ACh= 337 ACp= 81 ACu= 55
 AL: Alisols (193)
 ALf= 61 ALg= 14 ALh= 81 ALj= 12 ALp= 4 ALu= 21
 AN: Andosols (254)
 ANg= 8 ANh= 31 ANi= 0 ANm= 42 ANu= 107 ANz= 66
 AR: Arenosols (688)
 ARa= 30 ARb= 60 ARc= 29 ARg= 73 ARh= 206 ARl= 163 ARo= 127
 AT: Anthrosols (44)
 ATa= 2 ATc= 32 ATf= 9 ATu= 1
 CH: Chernozems (182)
 CHg= 7 CHh= 72 CHk= 65 CHl= 37 CHw= 1
 CL: Calcisols (312)
 CLh= 208 CLl= 44 CLp= 60
 CM: Cambisols (1262)
 CMc= 174 CMD= 210 CMe= 352 CMg= 115 CMi= 22 CMo= 133 CMu= 99 CMv= 78
 CMx= 79
 FL: Fluvisols (614)
 FLc= 163 FLd= 74 FLe= 267 FLm= 37 FLs= 14 FLt= 33 FLu= 26
 FR: Ferralsols (498)
 FRg= 31 FRh= 192 FRp= 10 FRr= 98 FRu= 43 FRx= 124
 GL: Gleysols (638)
 GLa= 0 GLd= 133 GLe= 310 GLi= 13 GLk= 20 GLm= 106 GLt= 5 GLu= 51
 GR: Greyzems (27)
 GRg= 3 GRh= 24
 GY: Gypsisols (67)
 GYh= 34 GYk= 16 GYl= 2 GYp= 15
 HS: Histosols (171)
 HSf= 62 HSi= 5 HSl= 4 HSs= 90 HSt= 8
 KS: Kastanozems (75)
 KSh= 30 KSk= 29 KSl= 16 KSy= 0
 LP: Leptosols (363)
 LPd= 94 LPe= 82 LPi= 4 LPk= 54 LPm= 24 LPq= 69 LPu= 36
 LV: Luvisols (1022)
 LVa= 31 LVf= 136 LVg= 99 LVh= 255 LVj= 81 LVk= 122 LVv= 41 LVx= 256
 LX: Lixisols (386)
 LXa= 3 LXf= 106 LXg= 13 LXh= 252 LXj= 1 LXp= 10
 NT: Nitisols (135)
 NTh= 55 NTr= 44 NTu= 36
 PD: Podzoluvisols (107)
 PDd= 11 PDe= 77 PDg= 13 PDi= 0 PDj= 6
 PH: Phaeozems (373)
 PHc= 36 PHg= 27 PHh= 114 PHj= 10 PHl= 186
 PL: Planosols (157)
 PLd= 34 PLe= 98 PLi= 0 PLm= 19 PLu= 5
 PT: Plinthosols (91)
 PTa= 23 PTd= 30 PTe= 35 PTu= 3
 PZ: Podzols (179)
 PZb= 16 PZc= 31 PZf= 6 PZg= 45 PZh= 65 PZi= 16
 RG: Regosols (289)
 RGc= 47 RGd= 56 RGe= 168 RGi= 4 RGu= 13 Rgy= 1
 SC: Solonchaks (117)
 SCg= 35 Sch= 22 SCi= 0 Sck= 14 SCm= 6 SCn= 23 SCy= 17
 SN: Solonetz (186)
 SNg= 43 SNh= 69 SNj= 22 SNk= 26 SNm= 23 SNy= 3
 VR: Vertisols (508)
 VRd= 20 VRe= 328 VRk= 156 VRy= 4

Total number of profiles classified according to Revised Legend (9607).

Appendix 3. Field names and percentage of records filled with measured data

a) Site data (total number of records= 9607)

Attribute	Filled	% of Total
WISE_ID	9607	100%
COUN	9607	100%
LAB_ID	9607	100%
SOURCE_ID	9607	100%
HORNUM	9607	100%
FAO_90	9607	100%
PHA_90	419	4%
FAO_74	9593	100%
PHA_74	414	4%
DRAIN	8065	84%
SOLDEP	9607	100%

Notes: WISE_ID is the unique profile reference number; COUN is the country ISO-code; LAB_ID a unique laboratory reference number, which provides the key to the type of analytical methods used; SOURCE_ID a unique reference number for the source of profile data; FAO_74 the FAO-Unesco (1974) classification as code, and PHA_74 the code for the (main) phase; FAO_90 is the FAO-Unesco (1988), classification as code, and PHA_90 the code for the code for (main) phase; DRAIN is the code for drainage condition (FAO, 1977) and SOLDEP the soil depth. The full complement of site and horizon data considered in WISE, version 1.0, is described elsewhere (Batjes, 1995).

b) Horizon data (total number of records= 46064)

Attribute	Filled	% of Total
WISE_ID	46064	100%
HORIZ	46064	100%
DESIG	40410	88%
TOPDEP	46064	100%
BOTDEP	46064	100%
ORGC	39869	87%
TOTN	26177	57%
C_N	26007	56%
CACO3	33062	72%
GYP SUM	26451	57%
PHH2O	41427	90%
PHKCL	16503	36%
PHCA CL2	27306	59%
ECE	32339	70%
EXCA	30227	66%
EXMG	29842	65%
EXNA	23270	51%
EXK	27608	60%
EXACID	22584	49%
EXALUM	21256	46%
CECSOIL	40922	89%
ECEC	3860	8%
BSAT	16730	36%
ALSAT	2199	5%
SAND	41929	91%
SILT	41955	91%
CLAY	41698	91%
GRAVEL	7446	16%
BULKDENS	16061	35%

PF17	958	2%
PF20	2133	5%
PF25	7188	16%
PF27	634	1%
PF37	779	2%
PF42	9916	22%

Appendix 4. Derived soil parameters for Haplic Acrisols before and after application of taxotransfer rules¹

FAO_90	ATTRIB	DEPZONE	TOPTX	NUM	MEA	CVA	MED	MAD	MIN	MAX	LLI	ULI	CONF	R_MED	R_CONF	RULE
ACH	AWC1	A	1	0	-1.00	-1	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-	16.00	a	R9
ACH	AWC1	A	2	1	25.00	-1	25.00	0.00	-1.00	-1.00	-1.00	-1.00	L	27.00	a	R9
ACH	AWC1	A	3	0	-1.00	-1	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-	23.00	a	R9
ACH	AWC1	A	#	1	25.00	-1	25.00	0.00	-1.00	-1.00	-1.00	-1.00	L	25.00	a	R9
ACH	AWC1	B	1	0	-1.00	-1	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-	14.00	a	R9
ACH	AWC1	B	2	1	18.00	-1	18.00	0.00	-1.00	-1.00	-1.00	-1.00	L	24.00	a	R9
ACH	AWC1	B	3	0	-1.00	-1	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-	25.50	a	R9
ACH	AWC1	B	#	1	18.00	-1	18.00	0.00	-1.00	-1.00	-1.00	-1.00	L	23.00	a	R9
ACH	AWC2	A	1	4	7.75	46	6.50	1.00	5.00	13.00	2.03	13.47	L	14.00	M'	R1
ACH	AWC2	A	2	8	13.63	40	12.00	4.50	7.00	22.00	9.07	18.18	M	12.00	M	R0
ACH	AWC2	A	3	5	12.40	24	12.00	1.00	9.00	17.00	8.72	16.08	M	12.00	M	R0
ACH	AWC2	A	#	17	11.88	41	11.00	3.00	5.00	22.00	9.39	14.37	H	11.00	H	R0
ACH	AWC2	B	1	4	7.75	30	7.00	1.00	6.00	11.00	3.99	11.51	L	11.00	M	R1
ACH	AWC2	B	2	8	11.63	34	12.50	2.50	5.00	16.00	8.28	14.97	M	12.50	M	R0
ACH	AWC2	B	3	5	12.60	35	14.00	3.00	7.00	17.00	7.15	18.05	M	14.00	M	R0
ACH	AWC2	B	#	17	11.00	37	11.00	4.00	5.00	17.00	8.91	13.09	H	11.00	H	R0
ACH	AWC3	A	1	7	6.43	63	5.00	3.00	2.00	12.00	2.70	10.16	M	5.00	M	R0
ACH	AWC3	A	2	29	9.00	54	8.00	3.00	2.00	20.00	7.15	10.85	H	8.00	H	R0
ACH	AWC3	A	3	14	11.00	38	10.50	3.50	6.00	19.00	8.60	13.40	M	10.50	M	R0
ACH	AWC3	A	#	50	9.20	51	8.00	3.00	2.00	20.00	7.86	10.54	V	8.00	V	R0
ACH	AWC3	B	1	6	5.33	48	4.50	1.50	3.00	9.00	2.62	8.04	M	4.50	M	R0
ACH	AWC3	B	2	24	8.13	39	8.00	2.00	3.00	16.00	6.79	9.46	H	8.00	H	R0
ACH	AWC3	B	3	13	10.85	53	11.00	5.00	5.00	25.00	7.36	14.33	M	11.00	M	R0
ACH	AWC3	B	#	43	8.56	51	8.00	3.00	3.00	25.00	7.22	9.90	V	8.00	V	R0
ACH	BULKDENS	A	1	42	1.45	11	1.44	0.14	1.10	1.72	1.40	1.50	V	1.44	V	R0
ACH	BULKDENS	A	2	42	1.43	10	1.43	0.09	1.13	1.66	1.39	1.47	V	1.43	V	R0
ACH	BULKDENS	A	3	23	1.25	12	1.25	0.10	1.01	1.61	1.18	1.31	H	1.25	H	R0
ACH	BULKDENS	A	#	107	1.40	12	1.40	0.13	1.01	1.72	1.36	1.43	V	1.40	V	R0
ACH	BULKDENS	B	1	41	1.48	9	1.50	0.10	1.27	1.71	1.44	1.52	V	1.50	V	R0
ACH	BULKDENS	B	2	44	1.42	10	1.40	0.10	1.10	1.71	1.37	1.46	V	1.40	V	R0
ACH	BULKDENS	B	3	21	1.23	9	1.25	0.08	1.05	1.47	1.18	1.28	H	1.25	H	R0
ACH	BULKDENS	B	#	106	1.40	11	1.40	0.11	1.05	1.71	1.37	1.43	V	1.40	V	R0

¹ This example shows only a small selection of the soil chemical and physical parameters under consideration.

FAO_90	ATTRIB	DEPZONE	TOPTX	NUM	MEA	CVA	MED	MAD	MIN	MAX	LLI	ULI	CONF	R_MED	R_CONF	RULE
ACh	CACO3	A	1	91	0.00	-1	0.00	0.00	0.00	0.00	0.00	0.00	V	0.00	V	R0
ACh	CACO3	A	2	124	0.00	-1	0.00	0.00	0.00	0.00	0.00	0.00	V	0.00	V	R0
ACh	CACO3	A	3	59	0.00	-1	0.00	0.00	0.00	0.00	0.00	0.00	V	0.00	V	R0
ACh	CACO3	A	#	274	0.00	-1	0.00	0.00	0.00	0.00	0.00	0.00	V	0.00	V	R0
ACh	CACO3	B	1	101	0.00	-1	0.00	0.00	0.00	0.00	0.00	0.00	V	0.00	V	R0
ACh	CACO3	B	2	121	0.00	-1	0.00	0.00	0.00	0.00	0.00	0.00	V	0.00	V	R0
ACh	CACO3	B	3	57	0.00	-1	0.00	0.00	0.00	0.00	0.00	0.00	V	0.00	V	R0
ACh	CACO3	B	#	279	0.00	-1	0.00	0.00	0.00	0.00	0.00	0.00	V	0.00	V	R0
ACh	CECCLAY	A	1	85	18.57	58	18.90	8.20	1.00	45.50	16.25	20.90	V	18.90	V	R0
ACh	CECCLAY	A	2	110	13.21	51	12.70	4.50	1.20	29.90	11.94	14.47	V	12.70	V	R0
ACh	CECCLAY	A	3	55	12.01	55	12.60	5.00	1.50	25.90	10.23	13.80	V	12.60	V	R0
ACh	CECCLAY	A	#	250	14.77	59	13.55	5.95	1.00	45.50	13.69	15.85	V	13.35	V	R0
ACh	CECCLAY	B	1	84	12.93	49	12.80	4.80	2.20	29.50	11.57	14.29	V	12.80	V	R0
ACh	CECCLAY	B	2	117	12.45	52	12.20	5.30	1.40	28.80	11.27	13.63	V	12.20	V	R0
ACh	CECCLAY	B	3	54	12.04	58	10.85	5.05	1.70	27.80	10.15	13.94	V	10.85	V	R0
ACh	CECCLAY	B	#	255	12.52	52	12.10	4.90	1.40	29.50	11.72	13.32	V	12.10	V	R0
ACh	CECSOIL	A	1	101	3.45	45	3.00	1.00	1.00	7.10	3.14	3.75	V	3.00	V	R0
ACh	CECSOIL	A	2	121	6.42	34	6.40	1.70	1.70	11.60	6.02	6.81	V	6.40	V	R0
ACh	CECSOIL	A	3	61	10.87	42	10.40	3.70	2.80	21.70	9.71	12.04	V	10.40	V	R0
ACh	CECSOIL	A	#	283	6.32	61	5.60	2.20	1.00	21.70	5.87	6.77	V	5.60	V	R0
ACh	CECSOIL	B	1	101	3.58	45	3.40	1.30	0.70	7.30	3.27	3.90	V	3.40	V	R0
ACh	CECSOIL	B	2	121	5.79	41	5.50	1.80	0.90	11.30	5.36	6.22	V	5.50	V	R0
ACh	CECSOIL	B	3	60	8.61	49	8.30	2.75	1.90	17.50	7.52	9.70	V	8.30	V	R0
ACh	CECSOIL	B	#	282	5.60	58	5.00	1.90	0.70	17.50	5.22	5.98	V	5.00	V	R0
ACh	C_N	A	1	74	11.67	21	11.41	1.41	7.50	17.50	11.10	12.25	V	11.41	V	R0
ACh	C_N	A	2	102	11.97	22	11.98	1.89	7.09	18.43	11.44	12.49	V	11.98	V	R0
ACh	C_N	A	3	48	11.61	24	11.33	2.27	7.64	17.26	10.78	12.43	V	11.33	V	R0
ACh	C_N	A	#	224	11.79	22	11.67	1.67	7.09	18.43	11.45	12.14	V	11.67	V	R0
ACh	C_N	B	1	57	9.64	21	10.00	1.77	7.14	15.07	9.11	10.17	V	10.00	V	R0
ACh	C_N	B	2	84	11.10	22	10.75	1.95	7.00	17.66	10.56	11.63	V	10.75	V	R0
ACh	C_N	B	3	36	10.85	23	10.23	1.61	7.40	16.06	9.99	11.71	V	10.23	V	R0
ACh	C_N	B	#	177	10.58	23	10.00	1.79	7.00	17.66	10.22	10.93	V	10.00	V	R0

FAO_90	ATTRIB	DEPZONE	TOPTX	NUM	MEA	CVA	MED	MAD	MIN	MAX	LLI	ULI	CONF	R_MED	R_CONF	RULE
ACh	ORGC	A	1	92	0.54	38	0.50	0.16	0.23	1.03	0.50	0.59	V	0.50	V	R0
ACh	ORGC	A	2	118	1.05	40	1.00	0.30	0.35	2.10	0.97	1.13	V	1.00	V	R0
ACh	ORGC	A	3	54	1.31	40	1.24	0.33	0.30	2.72	1.16	1.45	V	1.24	V	R0
ACh	ORGC	A	#	264	0.93	53	0.83	0.30	0.23	2.72	0.87	0.99	V	0.83	V	R0
ACh	ORGC	B	1	94	0.29	34	0.30	0.08	0.12	0.54	0.27	0.31	V	0.30	V	R0
ACh	ORGC	B	2	120	0.44	41	0.43	0.14	0.16	0.87	0.41	0.48	V	0.43	V	R0
ACh	ORGC	B	3	54	0.49	39	0.45	0.12	0.13	0.99	0.44	0.54	V	0.45	V	R0
ACh	ORGC	B	#	268	0.40	44	0.35	0.11	0.12	0.99	0.38	0.42	V	0.35	V	R0
ACh	PHH2O	A	1	105	5.50	12	5.50	0.50	4.10	7.00	5.37	5.63	V	5.50	V	r0
ACh	PHH2O	A	2	125	4.90	13	4.90	0.50	3.60	6.40	4.79	5.01	V	4.90	V	r0
ACh	PHH2O	A	3	57	4.85	13	4.80	0.50	3.80	6.20	4.68	5.03	V	4.80	V	r0
ACh	PHH2O	A	#	287	5.11	14	5.10	0.50	3.60	7.00	5.03	5.19	V	5.10	V	r0
ACh	PHH2O	B	1	105	5.17	10	5.20	0.30	4.20	6.40	5.07	5.27	V	5.20	V	r0
ACh	PHH2O	B	2	124	4.92	10	4.90	0.40	3.90	6.10	4.83	5.00	V	4.90	V	r0
ACh	PHH2O	B	3	57	4.94	10	5.00	0.30	3.80	6.00	4.81	5.07	V	5.00	V	r0
ACh	PHH2O	B	#	286	5.02	10	5.00	0.35	3.80	6.40	4.96	5.07	V	5.00	V	r0
ACh	PHKCL	A	1	31	4.30	9	4.20	0.30	3.70	5.20	4.15	4.45	V	4.20	V	R0
ACh	PHKCL	A	2	39	4.21	8	4.10	0.30	3.60	4.90	4.11	4.32	V	4.10	V	R0
ACh	PHKCL	A	3	28	4.37	12	4.25	0.40	3.50	5.40	4.17	4.56	H	4.25	H	R0
ACh	PHKCL	A	#	98	4.28	9	4.20	0.30	3.50	5.40	4.20	4.37	V	4.20	V	R0
ACh	PHKCL	B	1	36	4.12	7	4.05	0.25	3.70	4.80	4.02	4.22	V	4.05	V	R0
ACh	PHKCL	B	2	39	4.17	9	4.10	0.20	3.50	4.90	4.05	4.29	V	4.10	V	R0
ACh	PHKCL	B	3	25	4.21	10	4.10	0.30	3.70	5.20	4.04	4.39	H	4.10	H	R0
ACh	PHKCL	B	#	100	4.16	9	4.10	0.30	3.50	5.20	4.09	4.23	V	4.10	V	R0
ACh	TOTN	A	1	87	0.05	45	0.05	0.01	0.02	0.11	0.04	0.05	V	0.05	V	R0
ACh	TOTN	A	2	101	0.09	40	0.08	0.02	0.02	0.18	0.08	0.10	V	0.08	V	R0
ACh	TOTN	A	3	45	0.13	38	0.12	0.03	0.04	0.24	0.11	0.14	V	0.12	V	R0
ACh	TOTN	A	#	233	0.08	54	0.08	0.03	0.02	0.24	0.08	0.09	V	0.08	V	R0
ACh	TOTN	B	1	78	0.04	34	0.04	0.01	0.02	0.07	0.03	0.04	V	0.04	V	R0
ACh	TOTN	B	2	92	0.05	32	0.05	0.02	0.02	0.08	0.04	0.05	V	0.05	V	R0
ACh	TOTN	B	3	46	0.06	33	0.06	0.01	0.02	0.10	0.05	0.06	V	0.06	V	R0
ACh	TOTN	B	#	216	0.05	38	0.04	0.01	0.02	0.10	0.04	0.05	V	0.04	V	R0

Notes:

- 1) 'A', under the heading 'DEPZONE', stands for topsoil and 'B' for subsoil (i.e. the 0-30 cm and 30-100 cm depth zone respectively).
- 2) 'TOPTX' is the abbreviation for topsoil textural class, i.e. Coarse (1), Medium (2) and Fine (3), while '#' refers to all classes combined, as appropriate. In case of subsoils, TOPTX is used as a flag to permit linkage with map units with the corresponding topsoil textural class as shown on the Soil Map of the World (FAO, 1995).

- 3) For explanation of soil parameter codes and units of measurements used, see Appendix 5.
- 4) *NUM* is the population size for the considered attribute upon screening on data integrity and application of an outlier-rejection based on median test at 95% confidence level (see Pleijsier, 1989) (see Section 2.4; note: *NUM*= *REJ2*).
- 5) The mean (*MEA*), coefficient of variation (*CVA*), median (*MED*), median of the absolute deviations from median (*MAD*), minimum (*MIN*), maximum (*MAX*), and lower and upper 95% confidence limits (*LLI* and *ULI*) are shown by sample population. *R_MED* is the value of the derived soil parameter after application of taxotransfer rules and *RULE* the code for the rule used (see Section 2.8). *R_CONF*, refers to the (original) *CONF* for the data substituted with the taxotransfer rule. In case 'R9' has been applied extra data collection would be useful; the symbol 'a' applies to rules based on default parameters acid soil units and 'b' for those based on default parameters for basic soil units; see section 2.8). The abbreviation 'ee' stands for an expert estimate based on pH-related rules.
- 6) Confidence in results shown should increase with sample size, and be lowest where *NUM* is 1 (*CONF*: -, *NUM*= 0; Low, 0< *NUM* #5; Moderate, 5< *NUM* #15; High, 15< *NUM* # 30; Very High, 30 < *NUM*). Table shows all results both as 'is', i.e. before the application of taxotransfer rules, and after the application of taxotransfer rules. *R_MED* is the median for the considered attribute after application of the taxotransfer rules.
- 7) Similar data for the other soil units and attributes are presented in file *FAO_74S8.DBF*, while the full ASCII-table for individual parameters is available also as files (see *74_xxxx.txt*, where 'xxxx' is the parameter under consideration (see App. 6).

Appendix 5. Structure of digital data files and coding conventions

a) Data database files

File Name	Brief description
Database files:	
- FAO_XXS8.DBF	Initial data file with statistics by soil unit, attribute, textural class and depth zone for 19xx FAO-Unesco Legend (with xx=74 and 88, respectively), both before and after application of taxotransfer rules (see section 2.8).
- FAO_XXST.DBF	Data file with all statistics by soil unit, attribute, textural class and depth zone for all soil units from major groups that cover at least 5 percent of the surface of the Soil map of the World. Such data used are as 'last resort' default values for mineral soils, excluding Andosols, when all other taxo transfer rules do not apply (see Rule 9 in section 2.8; abbreviation 'a' is for acid soil units and 'b' for basic soil units).
Summary files:	
- SUMTABxx.DBF	Compact summary file showing median values for all soil parameters considered with additional information on the type of taxotransfer rules used (see section 2.8). (Note: this file shows the number of decimal places that should be used, while all other files use 2 decimal places by default for pragmatic reasons).
-xx_attri.txt	Summary ASCII text-files by attribute (e.g., CaCO ₃) listing medians by soil unit, topsoil textural class and depth zone, and documenting the taxotransfer rule used (see section 2.8), for the specified legend (xx). Excerpt from file FAO_XXS8.DBF.

Note: The files holding the derived soil parameters, and documentation, can be downloaded via: <http://www.isric.org>.

b) Structure of data-files

The structures of the various database files are described below, using the 1974 Legend as example. The ASCII text-files use the same coding conventions as described below for the dbf-files.

File: FAO_74S8.DBF

Field Name	Type	Width	Dec	Description
FAO_74	Character	2		Classification in FAO-Unesco (1974) Legend
ATTRIB	Character	8		Name of attribute; for abbreviations see Table at end of current Appendix.
DEPZONE	Character	3		Depth interval
TOPTX	Character	1		Topsoil textural class (or flag in case of subsoil)
NUM	Numeric	4		Number of observations after outlier rejection (NUM= REJ2)
MED	Numeric	6	2	Median for specified attribute
MAD	Numeric	6	2	Median of the absolute deviations from median
CVA	Numeric	3		Coefficient of variation (in per cent)
MIN	Numeric	6	2	Minimum
MAX	Numeric	6	2	Maximum
LLI	Numeric	6	2	Confidence interval, 95%, lower limit
ULI	Numeric	6	2	Confidence interval, 95%, lower limit
CONF	Character	1		Indicator for confidence in the derived data.
R_MED	Numeric	8	2	Median after use of taxotransfer rules (see Section 2.8)
R_CONF	Character	2		Confidence in above
RULE	Character	2		Code for taxotransfer rule see section 2.8)

File: SUMTAB74.DBF

Field Name	Type	Width	Dec	Description
FAO_74	Character	2		Classification in FAO-Unesco (1974) Legend
TOPTX	Character	1		Code for topsoil textural class (i.e., C, M, F or #)
ORGC_TM	Numeric	6	2	Organic matter content, median for the topsoil (TM)
ORGC_TR	Character	2		Number of taxotransfer rule used for topsoil (TR)
ORGC_BM	Numeric	6	2	Organic matter content, median for the subsoil (SM)
ORGC_BR	Character	2		Number of taxotransfer rule used for subsoil (SR)
TOTN_TM	Numeric	6	2	As above, but for TOTN
TOTN_TR	Character	2		As above, but for TOTN
TOTN_BM	Numeric	6		As above, but for TOTN
TOTN_BR	Character	2		As above, but for TOTN
CN_TM	Numeric	3	0	As above, but for C_N
CN_TR	Character	2		As above, but for C_N
CN_BM	Numeric	3	0	As above, but for C_N
CN_BR	Character	2		As above, but for C_N
PHH2O_TM	Numeric	4	1	As above, but for pH(H ₂ O)
PHH2O_TR	Character	2		As above, but for pH(H ₂ O)
PHH2O_BM	Numeric	4	1	As above, but for pH(H ₂ O)
PHH2O_BR	Character	2		As above, but for pH(H ₂ O)
PHKCL_TM	Numeric	4	1	As above, but for pH(KCl)
PHKCL_TR	Character	2		As above, but for pH(KCl)
PHKCL_BM	Numeric	4	1	As above, but for pH(KCl)
PHKCL_BR	Character	2		As above, but for pH(KCl)
PHCaCl2_TM	Numeric	4	1	As above, but for pH(CaCl ₂)
PHCaCl2_TR	Character	2		As above, but for pH(CaCl ₂)
PHCaCl2_BM	Numeric	4	1	As above, but for pH(CaCl ₂)
PHCaCl2_BR	Character	2		As above, but for pH(CaCl ₂)
CECSOIL_TM	Numeric	6	0	As above, but for CEC _{soil}
CECSOIL_TR	Character	2		As above, but for CEC _{soil}
CECSOIL_BM	Numeric	6	0	As above, but for CEC _{soil}
CECSOIL_BR	Character	2		As above, but for CEC _{soil}
CECCLAY_TM	Numeric	6	0	As above, but for CEC _{clay}
CECCLAY_TR	Character	2		As above, but for CEC _{clay}
CECCLAY_BM	Numeric	6	0	As above, but for CEC _{clay}
CECCLAY_BR	Character	2		As above, but for CEC _{clay}
APCEC_TM	Numeric	6	0	As above, but for APARCEC
APCEC_TR	Character	2		As above, but for APARCEC
APCEC_BM	Numeric	6	0	As above, but for APARCEC
APCEC_BR	Character	2		As above, but for APARCEC
TEB_TM	Numeric	6	0	As above, but for Total Exchangeable Bases
TEB_TR	Character	2		As above, but for Total Exchangeable Bases
TEB_BM	Numeric	6	0	As above, but for Total Exchangeable Bases
TEB_BR	Character	2		As above, but for Total Exchangeable Bases
BSAT_TM	Numeric	6	0	As above, but for base saturation
BSAT_TR	Character	2		As above, but for base saturation
BSAT_BM	Numeric	6	0	As above, but for base saturation
BSAT_BR	Character	2		As above, but for base saturation
ALSAT_TM	Numeric	6	0	As above, but for aluminium saturation
ALSAT_TR	Character	2		As above, but for aluminium saturation
ALSAT_BM	Numeric	6	0	As above, but for aluminium saturation
ALSAT_BR	Character	2		As above, but for aluminium saturation
ESP_TM	Numeric	6	0	As above, but for Exch. Sodium Percentage
ESP_TR	Character	2		As above, but for Exch. Sodium Percentage
ESP_BM	Numeric	6	0	As above, but for Exch. Sodium Percentage

Field Name	Type	Width	Dec	Description
ESP_BR	Character	2		As above, but for Exch. Sodium Percentage
ECE_TM	Numeric	6	1	As above, but for Electrical conductivity
ECE_TR	Character	2		As above, but for Electrical conductivity
ECE_BM	Numeric	6	1	As above, but for Electrical conductivity
ECE_BR	Character	2		As above, but for Electrical conductivity
CAMG_K_TM	Numeric	6	1	As above, but for ratio of Exch.Ca ²⁺ / Exch.K ⁺
CAMG_K_TR	Character	2		As above, but for ratio of Exch.Ca ²⁺ / Exch.K ⁺
CAMG_K_BM	Numeric	6	1	As above, but for ratio of Exch.Ca ²⁺ / Exch.K ⁺
CAMG_K_BR	Character	2		As above, but for ratio of Exch.Ca ²⁺ / Exch.K ⁺
CA_K_TM	Numeric	6	1	As above, but for ratio of Exch.Ca ²⁺ / Exch.K ⁺
CA_K_TR	Character	2		As above, but for ratio of Exch.Ca ²⁺ / Exch.K ⁺
CA_K_BM	Numeric	6	1	As above, but for ratio of Exch.Ca ²⁺ / Exch.K ⁺
CA_K_BR	Character	2		As above, but for ratio of Exch.Ca ²⁺ / Exch.K ⁺
CA_MG_TM	Numeric	6	1	As above, but for ratio of Exch.Ca ²⁺ / Exch.Mg ²⁺
CA_MG_TR	Character	2		As above, but for ratio of Exch.Ca ²⁺ / Exch.Mg ²⁺
CA_MG_BM	Numeric	6	1	As above, but for ratio of Exch.Ca ²⁺ / Exch.Mg ²⁺
CA_MG_BR	Character	2		As above, but for ratio of Exch.Ca ²⁺ / Exch.Mg ²⁺
ECEC_TM	Numeric	6	0	As above, but for Effective CEC
ECEC_TR	Character	2		As above, but for Effective CEC
ECEC_BM	Numeric	6	0	As above, but for Effective CEC
ECEC_BR	Character	2		As above, but for Effective CEC
CACO3_TM	Numeric	6	1	As above, but for calcium carbonate content
CACO3_TR	Character	2		As above, but for calcium carbonate content
CACO3_BM	Numeric	6	1	As above, but for calcium carbonate content
CACO3_BR	Character	2		As above, but for calcium carbonate content
GYPSUM_TM	Numeric	6	0	As above, but for gypsum content
GYPSUM_TR	Character	2		As above, but for gypsum content
GYPSUM_BM	Numeric	6	0	As above, but for gypsum content
GYPSUM_BR	Character	2		As above, but for gypsum content
BULK_TM	Numeric	6	2	As above, but for bulk density
BULK_TR	Character	2		As above, but for bulk density
BULK_BM	Numeric	6	2	As above, but for bulk density
BULK_BR	Character	2		As above, but for bulk density
TPOR_TM	Numeric	6	0	As above, but for total porosity
TPOR_TR	Character	2		As above, but for total porosity
TPOR_BM	Numeric	6	0	As above, but for total porosity
TPOR_BR	Character	2		As above, but for total porosity
AWC1_TM	Numeric	6	0	As above, but for Av. Water Capacity (AWC1)
AWC1_TR	Character	2		As above, but for Av. Water Capacity (AWC1)
AWC1_BM	Numeric	6	0	As above, but for Av. Water Capacity (AWC1)
AWC1_BR	Character	2		As above, but for Av. Water Capacity (AWC1)
AWC2_TM	Numeric	6	0	As above, but for Av. Water Capacity (AWC2)
AWC2_TR	Character	2		As above, but for Av. Water Capacity (AWC2)
AWC2_BM	Numeric	6	0	As above, but for Av. Water Capacity (AWC2)
AWC2_BR	Character	2		As above, but for Av. Water Capacity (AWC2)
AWC3_TM	Numeric	6	0	As above, but for Av. Water Capacity (AWC3)
AWC3_TR	Character	2		As above, but for Av. Water Capacity (AWC3)
AWC3_BM	Numeric	6	0	As above, but for Av. Water Capacity (AWC3)
AWC3_BR	Character	2		As above, but for Av. Water Capacity (AWC3)
TAWC1_M	Numeric	6	0	As above, but for Total Av. Water Cap. (TAWC1)
TAWC1_R	Character	2		As above, but for Total Av. Water Cap. (TAWC1)
TAWC2_M	Numeric	6	0	As above, but for Total Av. Water Cap. (TAWC2)
TAWC2_R	Character	2		As above, but for Total Av. Water Cap. (TAWC2)
TAWC3_M	Numeric	6	0	As above, but for Total Av. Water Cap. (TAWC3)
TAWC3_R	Character	2		As above, but for Total Av. Water Cap. (TAWC3)
SAND_TR	Character	2		As above, but for sand
SAND_BM	Numeric	6	0	As above, but for sand
SAND_BR	Character	2		As above, but for sand
SAND_TM	Numeric	6	0	As above, but for sand

Field Name	Type	Width	Dec	Description
SILT_TM	Numeric	6	0	As above, but for silt
SILT_TR	Character	2		As above, but for silt
SILT_BM	Numeric	6	0	As above, but for silt
SILT_BM	Numeric	6	0	As above, but for silt
CLAY_TM	Numeric	6	0	As above, but for clay
CLAY_TR	Character	2		As above, but for clay
CLAY_BM	Numeric	6	0	As above, but for clay
CLAY_BR	Character	2		As above, but for clay
GRAVEL_TM	Numeric	6	0	As above, but for gravel
GRAVEL_TR	Character	2		As above, but for gravel
GRAVEL_BM	Numeric	6	0	As above, but for gravel
GRAVEL_BR	Character	2		As above, but for gravel

Note:

- This summary files lists all the considered soil parameters, for each combination of soil unit, depth zone and topsoil textural class, in one 'single' line or record to facilitate linkage with the digital Soil Map of the World and various model applications. The distribution of FAO-Unesco (1974) soil units by topsoil textural classes as occurring on the 1:5 M scale Soil Map of the World is presented elsewhere (see App. 5 in Batjes *et al.*, 1997); several combinations such as Vertisols with a coarse topsoil or Arenosols with a fine textured topsoil do not occur.
- Additional details may be found in the 'underlying' files from which the summary file has been derived (e.g., FAO_74s8.dbf).
- Contrary to the preceding files, SUMTAB74 and SUMTAB90 give the *actual number of decimal places considered justified for presenting results of the analyses*, by attribute. All other files use 2 decimal places by default for presenting interim results.
- The summary file can easily be expanded to include parameters derived from FAO's CD-ROM, such as slope class, soil drainage class and soil depth class, and other auxiliary information held in soil phase descriptors.

c) Codes for physical and chemical attributes

Attribute	Explanation
ALSAT	Aluminium saturation (% of ECEC)
APCEC	Apparent CEC (CEC of clay fraction, not corrected for organic matter)
AWC1	Water retention (for - 5 to - 1500 kPa; % v/v)
AWC2	Water retention (for -10 to - 1500 kPa; % v/v)
AWC3	Water retention (for -33 to - 1500 kPa; % v/v)
BSAT	Base saturation (% of CEC _{soil})
BULKDENS	Bulk density (g cm ⁻³)
CACO3	Calcium carbonate (% by weight)
CECCLAY	CEC of clay fraction (corrected for organic matter; cmol _c kg ⁻¹)
CECSOIL	CEC of soil fraction (1 M NH ₄ OAc at pH 7; cmol _c kg ⁻¹)
C_N	C/N ratio
ECE	Electrical conductivity (mS cm ⁻¹ or dS m ⁻¹ or mmho cm ⁻¹)
ECEC	Effective CEC (cmol _c kg ⁻¹)
ESP	Exchangeable sodium percentage (as % of CEC _{soil})
GYPSUM	Total gypsum, as CaSO ₄ .2H ₂ O (% by weight)
ORGC	Organic carbon content (% by weight)
PHH2O	pH water
PHKCL	pH measured in KCl
PHCACL2	pH measured in CaCl ₂
R_CAMG_K	Ratio of exchangeable (Ca ²⁺ + Mg ²⁺) over K ⁺
R_CA_K	Ratio of exchangeable Ca ²⁺ over K ⁺
R_CA_MG	Ratio of exchangeable Ca ²⁺ over Mg ²⁺
TAWC1	Available water to 1 m depth (for - 5 to - 1500 kPa; % v/v)
TAWC2	Available water to 1 m depth (for -10 to - 1500 kPa; % v/v)
TAWC3	Available water to 1 m depth (for -33 to - 1500 kPa; % v/v)
R_CA_MG	Ratio of exchangeable Ca ²⁺ over Mg ²⁺
TEB	Total exchangeable bases (Ca ²⁺ + Mg ²⁺ + K ⁺ + Na ⁺)
TOTN	Total nitrogen (% by weight)
TOTPORES	Total porosity (% v/v)
GRAVEL	Gravel percentage (% v/v)

Information on recommended soil analytical procedures may be found elsewhere (Batjes, 1995; Van Engelen and Wen, 1995; Van Reeuwijk, 1993).

ADDENDUM

Median soil depth (March 2002)

Files SUMDEP74.DBF and SUMDEP90.DBF present information about the median soil depth by FAO soil unit, determined both according to the original Legend (FAO-Unesco, 1974) and the Revised Legend (FAO, 1988) to the Soil Map of the World. It serves to complement the list of soil parameters for the soil types of world presented earlier by Batjes (2002).

Soil depth in the current context corresponds with the lowermost depth considered during pit description (this is generally the maximum depth without strong physical limitations or to rock). In combination with information on soil chemical constraints, such as toxic levels of aluminium or salts, the current data set can be used to approximate effective soil depth for specific crops.

Structure for database: SUMDEP74.DBF

Field Name	Type	Width	Dec	description
FAO_74	Char.	2		Soil classification (FAO-Unesco, 1974)
NUM	Num.	4	0	Number of profiles
QUAR1	Num.	3	0	1 st quartile
MEDIAN	Num.	3	0	Median, maximum soil depth described (cm)
QUAR3	Num.	3	0	3 rd quartile
MAD	Num.	3	0	Median of the absolute deviation from the median
RULE	Char.	2		Number of pedotransfer rule used (see text)

Note: File SUMDEP90.DBF has a similar structure, but present results for soil units of the Revised Legend.

Changes in Version 2.1 of the Soil Parameter Data Set (July 2002)

The main changes *vis a vis* version 2.0 modifications are:

- 1) Integrity checks for several soil parameters were refined, notably for CaCO₃ and Aluminium saturation with reference to soil pH_{water} values (see Brunt and Van Reeuwijk, 1997).
- 2) A more stringent exclusion procedure has been applied for soils with very low contents of organic carbon (< 0.1%) as, in retrospect, these very low values were found to have a marked effect on the median during application of the statistical outlier rejection procedure.
- 3) Use of 350 cmol_c kg⁻¹ as default CEC value for organic carbon when computing CEC_{clay} (Klamt and Sombroek, 1988). In the preceding version, the default value of 240 cmol_c kg⁻¹ OC was used based on Scheffer and Schachtschabel (1984; see Batjes *et al.*, 1997).
- 4) Organic carbon and total nitrogen for about 25 profiles from Australia were erroneously given as g kg⁻¹ instead of the units (wt %) required for WISE. This has now been corrected.
- 5) Instead of applying the outlier rejection scheme to data for each of the “C”, “M”, “F” and “#” topsoil textural classes, the procedure has only been applied to categories “C”, “M” and “F” in version 2.1. The sample set used for analysis of category “#” now consists of the amalgamation of the data sets that remained after applying the data outlier procedure to categories “C”, “M” and “F” (by soil unit, depth zone and attribute).

- 6) Note: In case of organic soils, the abbreviation “C” for topsoil textural class stands for ‘organiC’. The “C” was used instead of the more logical abbreviation of “O”, for organic soils, as the character “C” is also a flow-driving variable in one of the TTR programmes. Taxotransfer rules applied to organic soils, however, are flagged as “Or” (see Section 2.8).