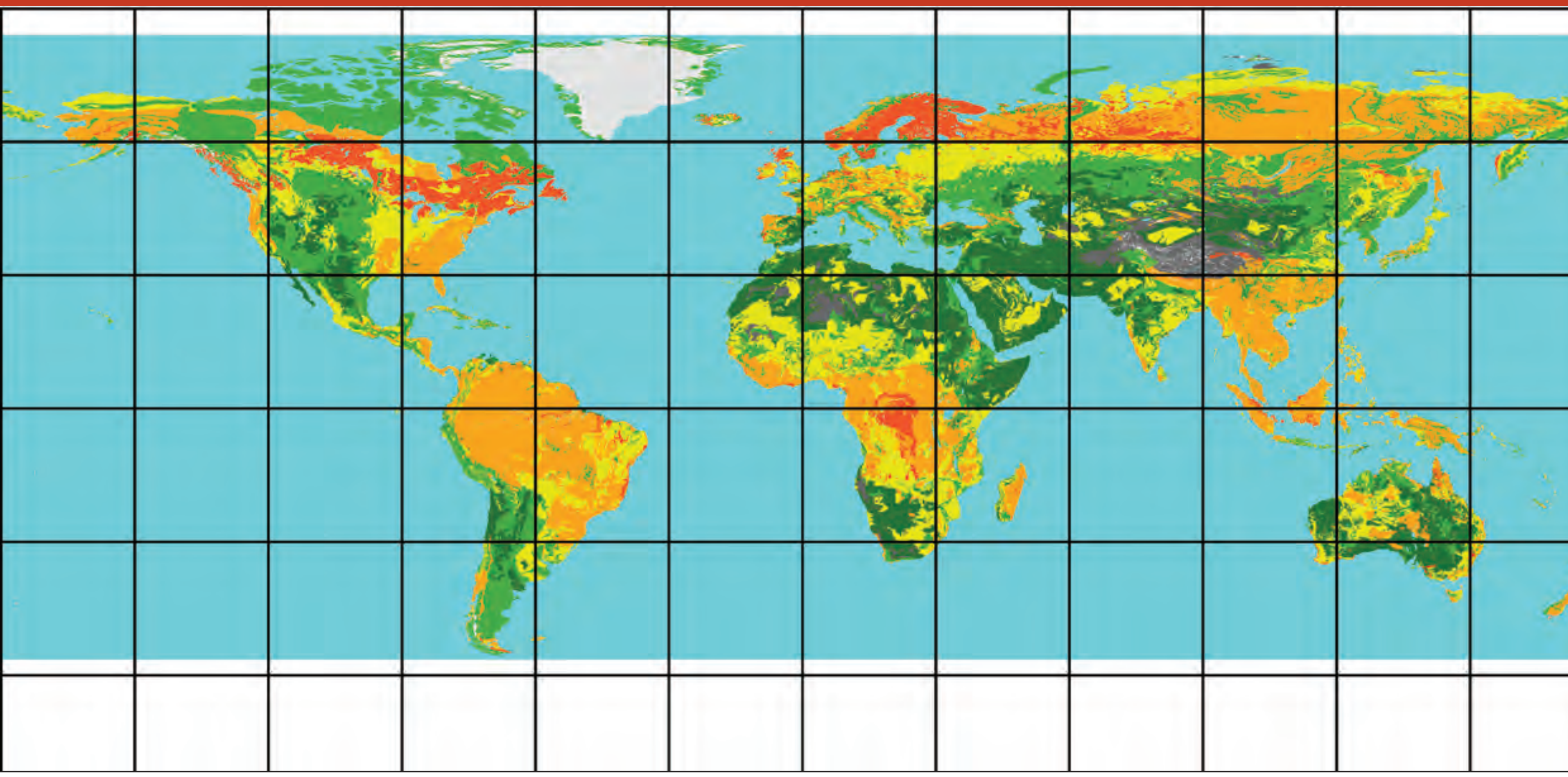
ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.2)

World Soil Information

ISRIC Report 2012/01



Niels H. Batjes

ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.2)

Niels H. Batjes

ISRIC Report 2012/01

Wageningen, 2012

Director, ISRIC – World Soil Information
PO BOX 353
6700 AJ Wageningen
The Netherlands
E-mail: soil.isric@wur.nl

Additional information on ISRIC – World Soil Information can be accessed through http://www.isric.org

Citation
Batjes N.H. 2012. *ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.2)*. Report 2012/01, ISRIC—World Soil Information, Wageningen (with data set, available at www.isric.org) Report 2012/01. 52 pp.; 9 fig.; 6 tab.; 48 ref.

Front cover: Soil pH classes for the World (0-20 cm depth; see Appendix 7 for legend).

ISRIC Report 2012/01

Contents

List of Tables

List of Figures

Preface

ISRIC – World Soil Information has the mandate to create and increase the awareness and understanding of the role of soils in major global issues. As an international institution, we inform a wide audience about the multiple roles of soils in our daily lives; this requires scientific analysis of sound soil information.

This study presents soil property estimates for the world. It draws on two large databases. The spatial data are derived from the 5 by 5 arcminutes version of the 1:5 million Digital Soil Map of the World (DSMW). Component soil units, characterised according to the FAO 1974 Legend, of the various DSMW mapping units are described using so-called virtual profiles for which there are no measured data. Derived soil property values, for selected variables such as pH, bulk density and clay content, are calculated for each virtual profile using a scheme of taxotransfer procedures. These take into consideration the type of FAO soil unit, depth layer, soil textural class, and soil property under consideration. The taxotransfer procedure was developed for general-purpose applications and draws heavily on soil analytical data held in the ISRIC-WISE soil profile database, version 3.1.

The present dataset may be used for broad, exploratory assessments at global scale (< 1:5 million; 5 by 5 minutes and coarser), keeping in mind the generalisations, assumptions and limitations for use that are described in this report. For more detailed assessments, global soil databases at a finer spatial resolution will need to be developed that take into account updated soil geographical and measured soil profile data for the world as collated during the on-going world soil and terrain database programme and similar, or using new digital mapping techniques for predicting soil properties.

In order to consolidate its world soil information services, ISRIC – World Soil Information is seeking collaboration with national institutes with a mandate for soil resource inventories.

Dr ir Prem Bindraban
Director, ISRIC – World Soil Information

Summary

This report describes a harmonized dataset of derived soil properties for the world. It was created using the soil distribution shown on the 1:5 million scale FAO-Unesco Soil Map of the World (DSMW) and soil property estimates derived from the ISRIC-WISE soil profile database, version 3.1.

The dataset considers 19 soil variables that are commonly required for agro-ecological zoning, land evaluation, crop growth simulation, modelling of soil gaseous emissions, and analyses of global environmental change. It presents 'best' estimates for: soil drainage class, organic carbon content, total nitrogen, C/N ratio, pH (H_2O), CEC_{soil}, CEC_{clay}, effective CEC, base saturation, aluminium saturation, calcium carbonate content, gypsum content, exchangeable sodium percentage (ESP), electrical conductivity, particle size distribution (i.e. content of sand, silt and clay), content of coarse fragments (> 2 mm), bulk density, and available water capacity (-33 to -1500 kPa). These estimates are presented by FAO soil unit for fixed depth intervals of 20 cm up to 100 cm depth (or less when appropriate) for so-called virtual profiles. The associated soil property values were derived from analyses of some 10,250 profiles held in WISE using a scheme of taxonomy-based taxotransfer rules complemented with expert-rules. The type of rules used to derive the various soil property values have been flagged in the database to provide an indication of the possible confidence in the derived data.

Most map units on the DSMW are complex, comprising up to eight different soil units. Assessments and model applications that use the derived soil properties therefore should consider the full map unit composition and depth range.

The soil property values presented here should be seen as best estimates based on the current selection of soil profiles in WISE, the procedure for clustering the measured data, taxotransfer scheme used for deriving soil, properties, and the spatial data of the digital Soil Map of the World. The derived information may be used for exploratory assessments at a broad scale (< 1:5 million; 5 by 5 arcminutes and coarser), pending the global update of the information on world soil resources at more detailed scales, upon due consideration of the underlying generalisations and assumptions.

Limitations for use are described in detail, explicitly stating why this type of generalised database may not be used to calculate soil organic carbon stocks, for example, which involves the analysis of co-varying soil properties.

Keywords: soils, derived properties, environmental modelling, ISRIC-WISE database, Digital Soil Map of the World (FAO-Unesco), appropriate data use

1 Introduction

This report describes the sources and procedures used to develop a generalised data set of derived soil properties, with a spatial resolution of 5 by 5 arc-minutes. The spatial data were derived from the Digital Soil Map of the World (DSWM, see FAO 1995), originally published at a scale of 1:5 million. Soil analytical data to characterize the main soil units of the DSMW were taken from the ISRIC-WISE database (from now on referred to as WISE). Version 3.1 of WISE holds some 10,250 globally distributed soil profiles, collated from disparate sources (Batjes 2009).

The DSMW was derived from the printed version of the Soil Map of the World (FAO-Unesco 1971-1981, 1974), with minor corrections and updates (FAO 1995). Compilation of the Soil Map of the World was a huge task involving collection and correlation of soil information from all regions of the world using the 1974 Legend. Most of this information was collated prior to the 1970s; the reliability thereof is known to vary considerably between different areas. Further, international soil classification systems have since then evolved through the Revised Legend (FAO 1988) to the World Reference Base for Soil Resources (IUSS Working Group WRB 2006).

The present, derived dataset supersedes version 1.1 that was based on analyses of a smaller selection of globally distributed soil profiles (Batjes 2006). However, it still uses the mapping unit information of the 1:5M scale DSMW; part of this information is outdated. Hence the on-going update of the information on world soil resources, at increasingly detailed scales, using various approaches (Sanchez *et al.* 2009; Nachtergaele *et al.* 2011; Van Engelen 2011; FAO/IIASA/ISRIC/ISSCAS/JRC 2012).

The materials and methods are described in Section 2. Results are presented and discussed in Section 3, while conclusions are drawn in Section 4. The structure of the various output tables, GIS-legend files, and installation procedure are presented in the Appendices.

2 Materials and methods

2.1 Primary soil data

2.1.1 Spatial data

The spatial distribution of soil units, per 5 by 5 arc-minute grid cell, was taken from the 1:5 million scale, digital Soil Map of the World. The base map of the DSMW is based on ETOPO5[1] (Earth Topography - 5 arc minute). ETOPO5 was assembled from several uniformly gridded databases into a worldwide data set with a cell size of 5 minutes of latitude by 5 minutes of longitude. Cells are written row-wise from the West to the East, starting at the Northwest corner. The cell size corresponds with about 0.08333 decimal degrees (see FAO 1995) or some 9 by 9 km at the equator.

The ERDAS files (WORLD.GIS and WORLD.TRL), prepared by FAO (1995), were first converted to ArcGis® GRID format. The spatial data are bound by longitudes -180°W and +180°E and latitudes +84°N and -56.50°S. This corresponds to 4320 columns and 1686 rows or 7,283,520 grid cells in total. The Arctic, Antarctica and some islands are not included on the DSMW (see FAO 1995).

The DSMW legend is comprised of 4931 different map units. These consist of soil units or associations thereof. Complex map units are comprised of one dominant soil unit and up to seven component soils. According to FAO's composition rules, the latter include associated soils (>20% of the map unit) and inclusions (<20% of the map unit).

Prior to starting with the analyses, the expansion files where checked for possible inconsistencies, also *vis a vis* the spatial, raster data. Several map units, while described in the expansion file, did not occur on the raster GIS map (SNUM 699, 1213, 1874, 1875, 1876, 1877, 1881, 1890, 1898, 1912, 1914, 1915, 1918, 1931, 3587, 6208, 6209, 6232, 6243, 6269, 6272, 6283, and 6324). In 35 cases, the dominant soil unit as given in the DSMW expansion files proved to be incorrect. For example, in map unit number (SNUM) 1823, coded 'Yh10-a', the fifth soil unit ($soil_5$; DS= dune sands) was estimated to cover 50% of the map unit and the so-called dominant soil ($soil_1$, Yh) only 10%. In all instances, this related to miscellaneous soil units with more than 50% coverage and these were always listed as the last component soil for the given map unit. These minor inconsistencies have been corrected. The original map unit codes (SNUM), however, were maintained to preserve consistency with the original codes used on the DSMW.

In seven instances (SNUM 3075, 3076, 3663, 4205, 5018, 5101, and 5211), the same soil unit was listed twice in FAO's expansion files. In such cases, the area of identical soil units within the map unit was summed and the expansion files were updated accordingly.

Some map units have been described as comprising say 50% of soil unit 1 and 50% of soil unit 2. In such cases, it will be difficult to select the so-called dominant soil unit when spatially aggregating the derived data (see Section 3.4).

---

1 http://www.ngdc.noaa.gov/mgg/global/etopo5.HTML

Statistics for the proportion of the dominant and component soils – based on the updated expansion files – are given in Table 1. The median relative area of the dominant soil unit ($soil_1$) within a map unit is 60%, with lower and upper quartiles of 50 and 70% respectively. The median, relative area for $soil_1$, $soil_2$, and $soil_3$ combined is 100% with lower quartile of 90% and a minimum of 60%.

Table 1
Relative area of dominant and component soils within map units of the Soil Map of the World.

Descriptive statistics	Relative area[a] of dominant and component soils (%)							
	Soil1[b]	Soil2	Soil3	Soil4	Soil5	Soil6	Soil7	Soil8
Minimum	24	0	0	0	0	0	0	0
1st Quartile	50	20	0	0	0	0	0	0
Median	60	20	10	0	0	0	0	0
3rd Quartile	70	30	20	10	0	0	0	0
Maximum	100	50	34	25	10	10	5	4

[a] The actual area within a 5 by 5 arc-minutes grid will vary with latitude – the grid cell size is some 9 by 9 km at the equator and will decrease gradually to the poles according to a cosine function of latitude. [b] $Soil_1$ is the dominant soil and $Soil_2$ to $Soil_8$ are the component or associated soils.

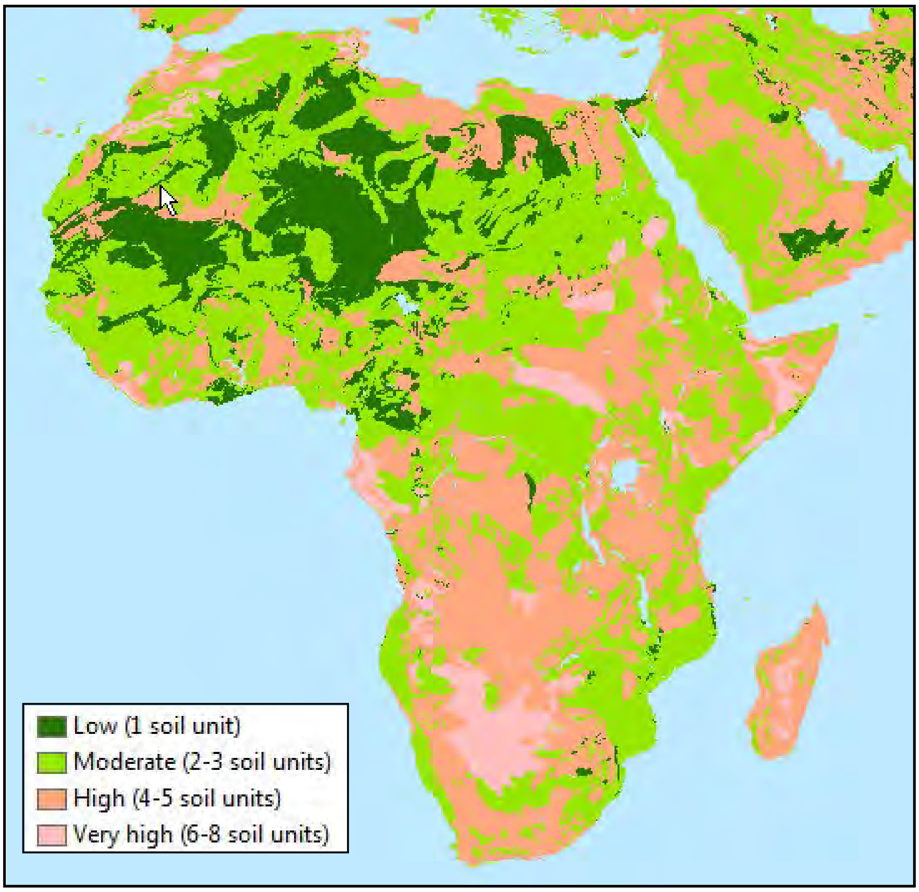


Figure 1
Complexity of grid cells for Africa according to the Digital Soil Map of the World.

Consideration of the dominant soil unit only will ignore the inherent complexity of many map units (Figure 1). Therefore, the full map unit composition should be considered when using the derived data in model applications, while keeping in the mind the 'suitability for use' of the present data set (see Section 3.5).

2.1.2 Profile data

Component soils of individual DSMW map units are not characterised by typical (real) profiles, as is the case with SOTER (van Engelen 1999), only by their classification (1974 FAO Legend). Therefore, the classification code was used to define a *virtual* profile for each of the 106 soil units of the 1974 FAO Legend, for example 'WD-Fr' for Rhodic Ferralsols. Sometimes, only the major soil group is known (e.g. Ferralsols) which implies a higher level of generalisation for the corresponding virtual profile (e.g. WD-F) and greater uncertainty.

In the absence of detailed textural data, all virtual profiles were assumed to have the modal textural class of the corresponding soil unit (as determined from statistical analyses of the corresponding selection of soil profile data in WISE). Typically, as discussed by FAO-Unesco (1974), marked changes in texture within the soil that result from profile formation are expressed in the definitions of the soil units (for example, the occurrence of an abrupt textural change or argillic horizon). Subsequently, selected soil property values were derived for each soil unit using a system of taxonomy-based pedotransfer and expert-rules (see Section 2.2). In accordance with FAO (1995) criteria, the procedure assumes that similarly classified soil units will have the same modal properties irrespective of their geographic occurrence in the world. This is a simplification, as it does not explicitly consider regional differences in climate, parent material, topography, natural vegetation and land use history and management. Based on the above generalisations, the range in soil conditions for the world is described here using 132 different soil classes.

Attribute data for the virtual profiles were derived using taxotransfer rules. These rules were developed based on statistical analyses of some 10,250 profiles held in WISE. Most profiles are from Africa (41 %), followed by Asia (18 %), South America (18 %), and Europe (13%) (Table 2). The corresponding soil descriptions and analytical data were largely derived from field surveys carried out between 1960 and 2000 (see Batjes 2009).

Table 2
Number of profiles in WISE by continent and their description status.

Continent	Profile description status[a]				Total
	1	2	3	4	
Africa	421	1337	2392	23	4173
Asia	441	970	426	10	1847
Antarctica	4	6	0	0	10
Europe	225	712	359	20	1316
North America	495	222	127	11	855
Oceania	50	49	106	4	209
South America	149	1380	313	1	1843
Total	*1785*	*4676*	*3723*	*69*	*10253*

[a] The number code under 'profile description status' refers to the completeness and apparent reliability of the soil profile descriptions and accompanying analytical data for the specified profile in the original source. The status is highest for '1' and lowest for '4' (see text). Details by country are given in a Technical Report (Batjes 2008).

The description status provides a *coarse* indicator for the inferred quality and completeness of profile data stored in WISE; it is indicative of the reliability of soil information entered into a database (FAO 2006). Essentially, the description status is determined by the confidence the various data compilers had in the various sources of soil profiles that are globally available and accessible. Analytical data for profiles having a description status of 1 or 2 may be considered as most reliable. Nonetheless, profiles flagged as having a description status of 1 or 2 will commonly show gaps in the measured data, in particular soil physical attributes. Often, even the so-called mandatory analytical attributes required by SOTER (van Engelen and Wen 1995, p. 17) are simply not available – abundance of fragments > 2 mm, particle size distribution (sand, wt% 2.0-0.05 mm; silt, wt% 0.05-0.002 mm; and clay, wt% < 0.002 mm); bulk density; pH-H_2O; exchangeable Ca^{2+}, Mg^{2+}, K^+, Na^+, and Al^{3+}; CEC_{soil}; and, organic carbon content. Similarly, descriptive information on essential site data, such as climate, parent material, and land use or natural vegetation as well as detailed location, is not provided in many source materials.

Almost every country has its own analytical methods and these methods may vary from one laboratory to the next within one country. Consequently, issues of quality and comparability of analytical data, collated from disparate sources, are critical in any analysis of soil profiles. Yet, there are no straightforward solutions for harmonising the data (Pleijsier 1989; van Reeuwijk 1998; Batjes 1999; Dobos *et al.* 2006). Consequently, these issues have been addressed pragmatically in this study similar to what has been the case with the 1:5 million scale Soil Map of the World (FAO-Unesco 1981, p. 91; FAO 1995) and similar. Correlation of soil analytical data, however, should be done more rigorously when more detailed scientific work, at a finer spatial resolution, is considered.

2.1.3 Taxonomic coverage

The relative number of soil profiles in WISE, available for each major soil group of the 1974 FAO Legend is shown in Figure 2. Yermosols (Y), for example, account for about 8% of the total extent of world soils and are represented by some 1% of the total number of profiles in WISE. Conversely, Regosols (R) also cover some 8% of DSMW yet are represented by some 5% of the profiles in WISE. Several soil units remain under-represented in WISE; details may be found elsewhere (Batjes 2008). Profile representation in WISE is not based on an area-weighted basis, but mainly on the availability of sufficiently detailed legacy data. This will introduce regional and taxonomic bias.

Typically, in order to present robust property estimates for a given soil property, at least 30 profiles – with complete and comparable sets of soil analytical data – would be needed for each soil unit. This ideal situation, however, will seldom occur as reflected by on-going discussions on the comparability of soil analytical data from disparate sources (Pleijsier 1989; van Reeuwijk and Houba 1998; Dobos *et al.* 2006; Hartemink *et al.* 2008).

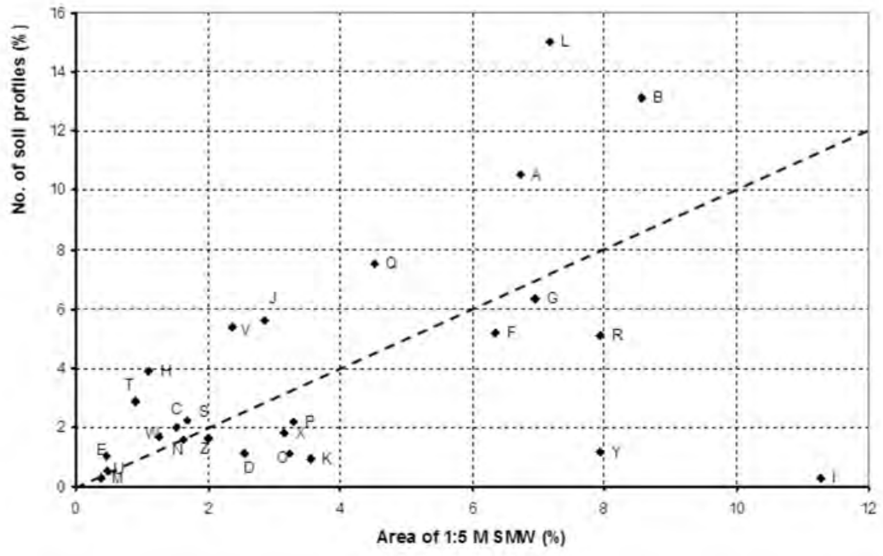


Figure 2
Representation of major soil groups in WISE3.1 relative to their relative extent on the 1:5M digital Soil Map of the World[2]*.*

2.2 Derived soil data

2.2.1 Soil properties

This study considers nineteen soil attributes (Table 4), commonly required in studies of land suitability and environmental change, for which primary or measured data are collated in WISE.

2.2.2 Taxotransfer procedure

Derived values for each soil property —by soil unit, depth range of 20 cm up to 100 cm depth, and five textural classes (CEC 1985; FAO 1988)— were derived from statistical analyses of the measured data present in WISE. The procedure, which includes an outlier-rejection scheme to screen the available data, is similar to the one used to generate secondary SOTER databases (Batjes *et al.* 2007). However, it has been adapted to run with soils characterised according to the 1974 FAO Legend, virtual profiles, and spatial data of the DSMW. Additional information is provided in Section 3.3.

[2] Relative area is expressed as percentage of total area of all major soil groups considered on the 1:5M scale, digital Soil Map of the World (FAO 1995); codes follow the original Legend (FAO-Unesco 1974): A, Acrisols; B, Cambisols; C, Chernozems; D, Podzoluvisols; E, Rendzinas; F, Ferralsols; G, Gleysols; H, Phaeozems; I, Lithosols; J, Fluvisols; K, Kastanozems; L, Luvisols; M, Greyzems; N, Nitosols; O, Histosols; P, Podzols; Q, Arenosols; R, Regoso ls; S, Solonetz; T, Andosols; U, Rankers; V, Vertisols; W, Planosols; X, Xerosols; Y, Yermosols; Z, Solonchaks.

Table 3
List of soil variables for which property estimates are presented.

Organic carbon
Total nitrogen
C/N ratio‡
Soil reaction (pH_{H2O})
Cation exchange capacity (CEC_{soil})
Cation exchange capacity of clay size fraction (CEC_{clay}) • ‡
Base saturation (as % of CEC_{soil}) ‡
Effective cation exchange capacity (ECEC) † ‡
Aluminium saturation (as % of ECEC) ‡
$CaCO_3$ content
Gypsum content
Exchangeable sodium percentage (ESP) ‡
Electrical conductivity (ECe)
Bulk density
Coarse fragments (volume %)
Sand (mass %)
Silt (mass %)
Clay (mass %)
Available water capacity (AWC; from -33 to -1500 kPa; cm m^{-1}) ‡ ▫

‡ Calculated from other measured soil properties.
† ECEC is defined as exchangeable (Ca^{++}+Mg^{++}+K^{+}+Na^{+}) + exchangeable (H^{+}+Al^{+++}) (van Reeuwijk 2002).
• CEC_{clay} 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} (Klamt and Sombroek 1988).
▫ The soil water potential limits for AWC conform to USDA standards (Soil Survey Staff 1983). Values shown have not been corrected for the presence of fragments > 2 mm.

2.2.3 Miscellaneous units

There are six miscellaneous units on the Soil Map of the World for which the following assumptions are used:

- Dune sands (DS): Soil property estimates for cambic Arenosols (Qb) were used as default. However, organic carbon content was set at 1 g C kg^{-1} for the topsoil (0-20 cm) and 0.5 g C kg^{-1} for the subsoil to reflect the arid conditions; the content of sand was set at 98%; available water capacity was set at 5 cm m^{-1} for the topsoil; and soil drainage at excessively drained.
- Salt Flats (ST): Gaps were filled using soil property estimates for orthic Solonchaks (Z) as the default.
- Not Determined (NA): Soil property estimates for Cambisols (B) were used as the default; Cambisols are the last major group in the hierarchical FAO legend.
- Oceans and Inland Waters (WR): All property estimates were set at -1 to facilitate visualisation using GIS (SNUM= 6997 and 1972 for Africa).
- Glaciers and snowcaps (GL): All property estimates were set at -2 (SNUM= 6998).
- Rock outcrops (RK): Being non-soil units, all soil property estimates were set at -7. Similarly, for shallow soils, including Lithosols, Rendzinas and Rankers, all property estimates for the rocky subsoil were set at -7.

3 Results and discussion

3.1 General

The compilation of the DSMW, at scale 1:5 million, encompassed a marked degree of data integration, the aim being to simplify the geographical distribution of soil types to a regionally representative pattern. Inherently, all soil mapping units will include a number of impurities, often in excess of 15% (see Landon 1991), which cannot be mapped at the given scale. Further, the exact location of the component soil units within a given map unit is not known, only their estimated proportion (see FAO 1995).

3.2 Soil unit composition

Each grid cell may contain up to eight component soils. The relative extent thereof — estimated according to FAO's expert-based composition rules — has been clustered in five classes to arrive at a compact map unit code: 1 – from 80 to 100 per cent; 2 – from 60 to 80 per cent; 3 – from 40 to 60 per cent; 4 – from 20 to 40 per cent, and 5 – less than 20 per cent (Appendix 2). The original FAO map unit code has also been maintained in the data set for ease of reference.

3.3 Derived soil properties

3.3.1 Tabular data

Results of the statistical analyses of the clustered data are stored in several MS Access® tables the structure of which is presented in Appendix 3. This information provided the basis for the taxotransfer procedure the results of which are presented in table *WISEparameterEstimates* (Appendix 4). The types of taxotransfer rules (*TTR*) that have been applied for each virtual profile/layer/attribute are documented in a separate table (*WISEflagRules*, see Appendix 5). Data listed under *TTRsub* indicate that the data substitution for a given attribute and depth layer was based on WISE-derived property estimates (means) for similar soil units, as illustrated in Figure 3. Otherwise, if the corresponding population in WISE was too small ($n_{WISE} < 5$) for the substitution to be considered meaningful, the rules were flagged under *TTRmain*.

Each flag consists of a sequence of letters followed by a numeral (see under *TTRsub* and *TTRmain*). The letters indicate soil attributes for which a TTR has been applied (Figure 3). The number code reflects the size of the sample population, after outlier-rejection, used for the statistical analyses (Table 4).

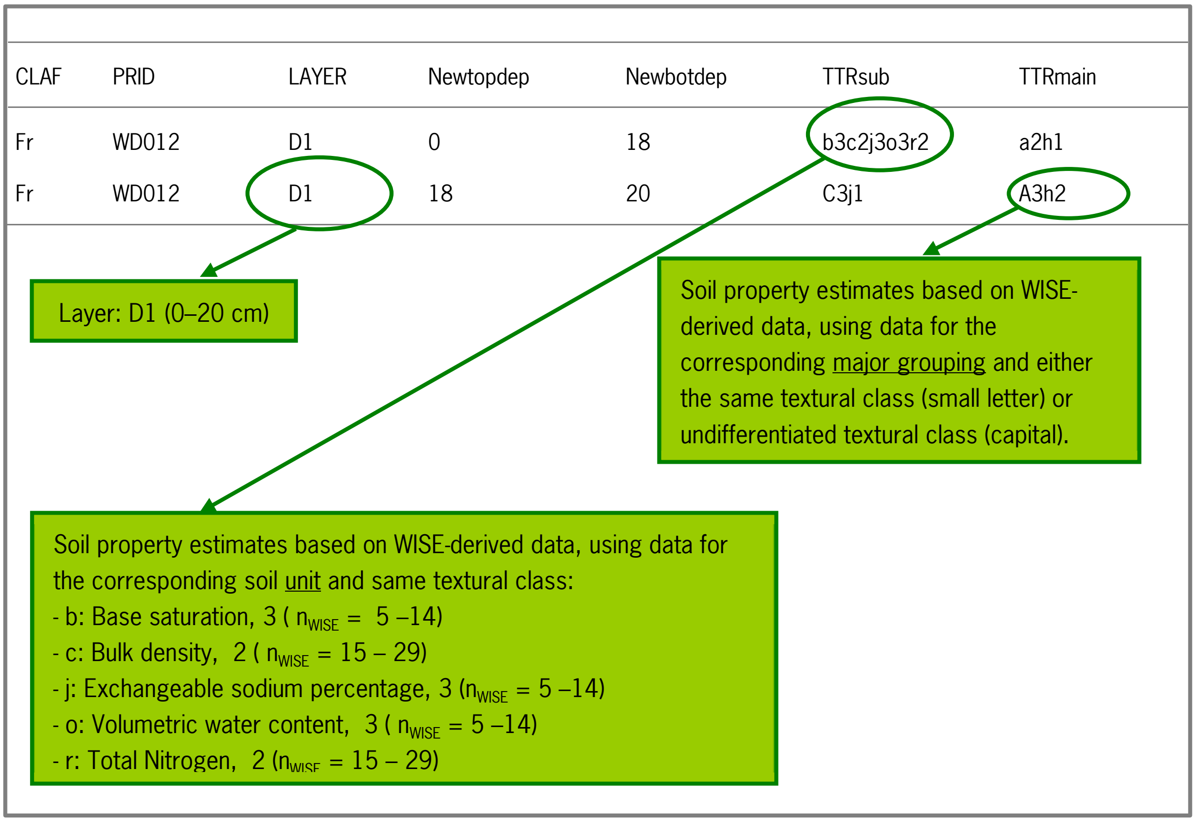


Figure 3
Flagging of taxotransfer rules by profile, depth zone and attribute.

Table 4
Criteria for defining confidence in the derived data.

Code	Confidence level	n_{WISE}[a]
1	High	> 30
2	Moderate	15-29
3	Low[b]	5-14
4	Very low	1-4
5	No data	0

[a] n_{WISE} is the sample size after outlier rejection.
[b] The cut-off point for applying any TTR is $n_{WISE} < 5$.

When a small letter is used, the substitution considered derived data for the corresponding textural class (for example, Fine). Otherwise, when a capital is used, this indicates that the substitution was based on the whole set for the corresponding soil unit and depth layer, irrespective of soil texture (i.e. undifferentiated or 'u' , which will always be the case for the present virtual profiles). The same coding conventions apply for *TTRmain*. The overall scheme is illustrated in Figure 3 for a virtual rhodic Ferralsol (Fr), coded *WD012*.

Finally, derived soil property values for each component soil of the mapping units considered on the DSMW are stored in table *WISEsummaryFile* (Appendix 6). Figure 4 shows an excerpt for map unit *WD6* as an example. It comprises 80% of orthic Acrisols (Ao), 10% of dystric Cambisols (Bd) and 10% of dystric Fluvisols (Jd). The top layer (D1, 0-20 cm) of the main soil unit (Ao) has a derived bulk density (B*ULK*) value of 1.36 kg m^{-3}. Similarly, the proportion of fragments > 2 mm (*CFRAG*) has been estimated at 24% at 80-100 cm (D5) depth for component soil number 3, here dystric Fluvisols.

WISEsummaryFile

ISO	SUII	Map unit	Laye	TCII	SC	PR(	CL	PRID	Dra	Top	Botl	CFRA	SDT	STF	CLF	PSC	BUL	TAW(	CE	BSAT	CEC	PHA	TCE
WD	6	WD6	D1	1	1	80	Ao	WD-Ao	W	0	20	11	59	20	21	M	1.36	13	7	43	19	5.1	0
WD	6	WD6	D2	1	1	80	Ao	WD-Ao	W	20	40	12	53	19	28	M	1.39	12	6	31	16	5.0	0
WD	6	WD6	D3	1	1	80	Ao	WD-Ao	W	40	60	14	49	18	33	M	1.40	12	6	28	15	5.0	0
WD	6	WD6	D4	1	1	80	Ao	WD-Ao	W	60	80	18	47	17	36	F	1.40	13	6	26	15	5.0	0
WD	6	WD6	D5	1	1	80	Ao	WD-Ao	W	80	100	16	46	18	36	F	1.38	13	6	24	15	5.0	0
WD	6	WD6	D1	1	2	10	Bd	WD-Bd	W	0	20	15	43	35	22	M	1.22	19	15	32	34	5.1	0
WD	6	WD6	D2	1	2	10	Bd	WD-Bd	W	20	40	19	44	32	24	M	1.29	17	11	26	37	5.1	0
WD	6	WD6	D3	1	2	10	Bd	WD-Bd	W	40	60	21	45	32	23	M	1.34	15	11	27	40	5.2	0
WD	6	WD6	D4	1	2	10	Bd	WD-Bd	W	60	80	25	46	31	23	M	1.36	16	11	32	42	5.2	0
WD	6	WD6	D5	1	2	10	Bd	WD-Bd	W	80	100	26	47	31	22	M	1.40	14	10	34	41	5.3	0
WD	6	WD6	D1	1	3	10	Jd	WD-Jd	I	0	20	10	51	28	21	M	1.25	13	8	34	29	4.9	0
WD	6	WD6	D2	1	3	10	Jd	WD-Jd	I	20	40	8	52	26	22	M	1.33	13	8	27	26	4.9	0
WD	6	WD6	D3	1	3	10	Jd	WD-Jd	I	40	60	15	55	24	21	M	1.42	14	6	25	28	5.0	0
WD	6	WD6	D4	1	3	10	Jd	WD-Jd	I	60	80	20	56	24	20	M	1.48	12	6	22	27	5.0	0
WD	6	WD6	D5	1	3	10	Jd	WD-Jd	I	80	100	24	58	22	20	M	1.55	14	6	25	31	5.1	0

Figure 4
Excerpt from summary file for mapping unit WD6.

3.3.2 Sources of uncertainty

Inherently, the calculation of derived soil properties for the 106 FAO soil units and 28 major soil groups of the Soil Map of the World remains fraught with uncertainty. Possible effects of regional variation in climate, relief and parent material on specific soil properties are not considered explicitly on the DSMW (FAO-Unesco 1971-1981; FAO 1995). By implication, the same property estimates had to be used irrespective of the occurrence say eutric Cambisols in the world, which is a simplification. Alternatively, climatic effects are expressed in the classification for Xerosols and Yermosols that are defined in terms of their aridic moisture regime. Similarly, humic, gleyic and gelic soil units generally occur under conditions that are suited for accumulation of soil organic matter. The variation in soil properties observed within each soil unit, for the present selection of soil profiles, is quantified in an attribute table; see Appendix 3 and the corresponding data set for details.

The above type of regional differences, however, can be accounted for in SOTER and similar. In SOTER, for example, each component soil of a mapping unit is characterized by a soil profile selected as being regionally representative by regional soil experts (e.g., Dijkshoorn *et al.* 2005). Inherently, such profiles may show gaps in the measured data. In the case of SOTER, taxotransfer rules may be applied with respect to the measured textural class of the selected profiles as opposed to the present forced use of the modal textural class with the gridded DSMW. Further, profiles in WISE that occur under similar climatic conditions (e.g., Köppen climate) may be selected to underpin the taxotransfer procedures for specific SOTER data sets (e.g., Batjes 2010). The present study uses a data model that is similar to that of FAO (1995), albeit using different soil textural classes and depth ranges as well as a much larger selection of globally distributed soil profiles.

It should be stressed that the natural range in measured chemical and physical properties can be considerable at soil unit level, with coefficients of variation (CV) often exceeding 50% (Beckett and Webster 1971; Spain *et al.* 1983; Landon 1991). Similarly, the median of absolute deviations (MADs) —the median of the differences between each observation and the median— are often large, pointing at a large spread of the data (see Appendix 3).

The overall assumption in this assessment has been that the confidence in a *TTR*-based soil property estimate should increase with the size of the sample populations present in WISE after outlier-rejection. In addition, in principle, the confidence in soil property estimates listed under *TTRsub* should be higher than for those listed under *TTRmain*. Nonetheless, a high confidence rating does not necessarily imply that the soil property estimates shown will be representative for the soil component under consideration. Profile selection for WISE, as for any other large database, is not probabilistic but based on available data and expert knowledge.

Results may also be biased for those soil properties that were recorded as 'not observed' or 'nil' in the original surveys, for example volumetric gravel content (*CFRAG*). In such cases, derived properties computed using the TTRs may well give a biased impression of 'modal' conditions for some soil units. For example, for soil units that are generally devoid of coarse fragments only the limiting gravel contents may have been documented in field surveys.

Several soil chemical properties —such as a high aluminium saturation in parts of the subsoil or a high exchangeable sodium percentage in parts of the topsoil— may be 'levelled out' during depth-weighting. When occurring, this will also be reflected in derived values obtained through pedotransfer.

Some of the soil attributes under consideration, such as the presence of fragments > 2 mm and water holding capacity, are not diagnostic in the FAO Legend. Many soil processes and properties are readily modified by changes in land use and management. For example, soil pH and aluminium saturation upon liming; salinity and electrical conductivity upon irrigation or soil drainage; and soil organic matter quantity and quality upon changes in tillage practices, deforestation or climate change. Such effects, however, cannot be considered explicitly here when analysing the available primary data.

3.4 Linkage to GIS

Soil property values derived from the *TTR*-scheme can be joined to the rasterized Soil Map of the World using GIS. This linkage will be through the unique map unit identifier (*SUID*), which corresponds to FAO's soil map unit number (*SNUM*) (Appendix 4). Similarly, the derived soil data can be linked to individual, continental *vector* sets provided with the digital Soil Map of the World.

Various options exist to display or spatially aggregate the derived soil data, each of these having their strengths and limitations (e.g., FAO 1995; Kern 1995; Carter and Scholes 2002; Batjes 2006). The type of research purpose will determine which soil property estimates, and depth layers, or class intervals will be required for a specific application. Generally, the necessary data selections can best be made with tailor-made programs designed to meet the scope of these applications; these should consider the full map unit composition and five depth intervals. Alternatively, for ease of visualization using GIS, other approaches (generalisations) may need to be used.

For example, Figure 5 shows the pH_{H2O} for the dominant soil unit and upper soil layer (D1, 0-20 cm) as an example of GIS output with special attention for map unit *WD5475* (see Appendix 6, method A). The dominant soil unit (Fx) only covers 25% of map unit 5475 with the other 75% being comprised of orthic Acrisols (20), plinthic Acrisols (20%), plinthic Ferralsols (20%), dystric Gleysols (5%), plinthic Gleysols (5%) and albic Arenosols (5%). The pH for 'Fx/D1' for soil unit 1 (*SCID*) is given as 4.45 as shown by the 'excerpt' of the attribute table in Figure 5. Similar maps may be created for the other soil attributes and depth layers using the appropriate data tables.

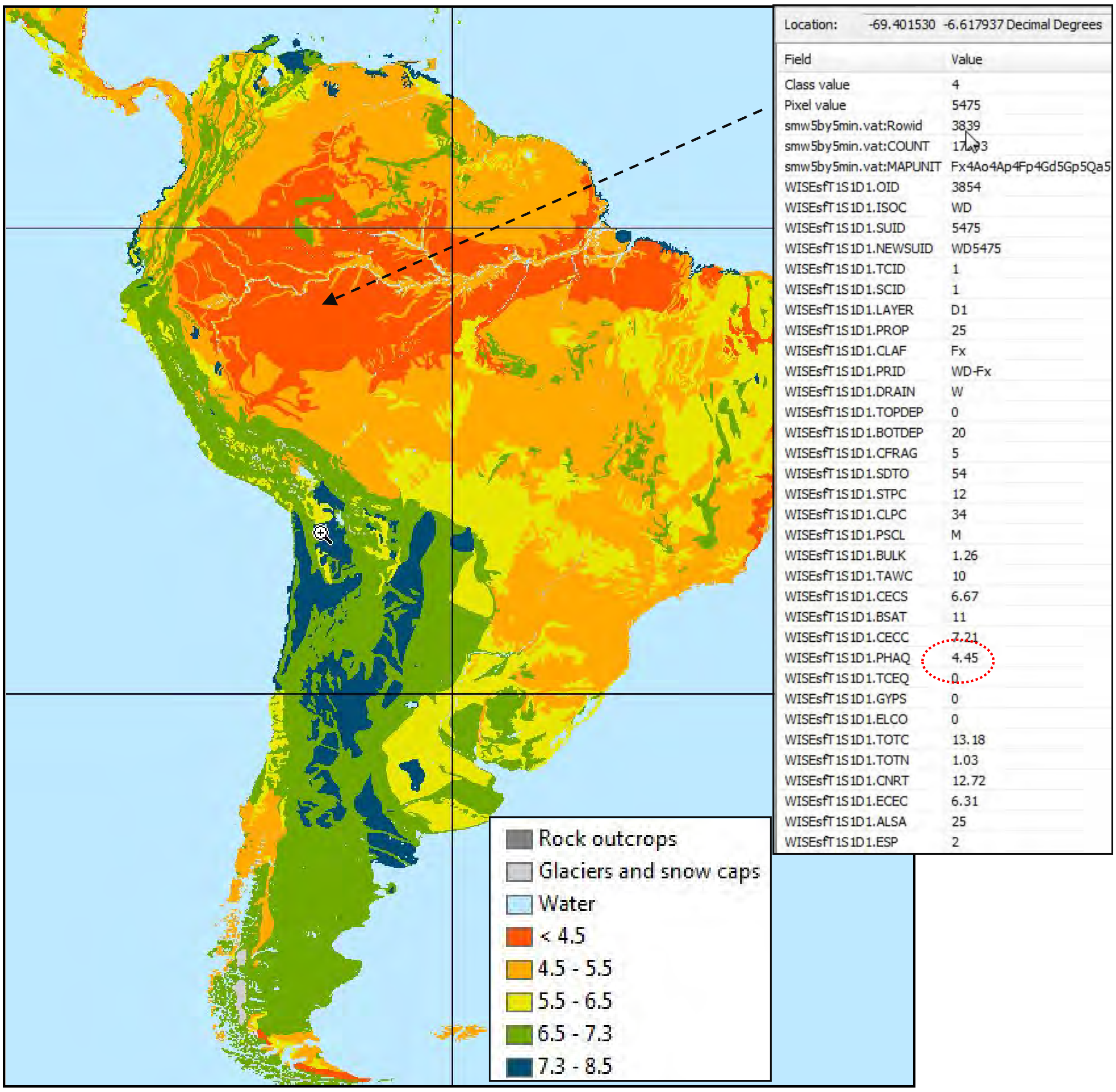
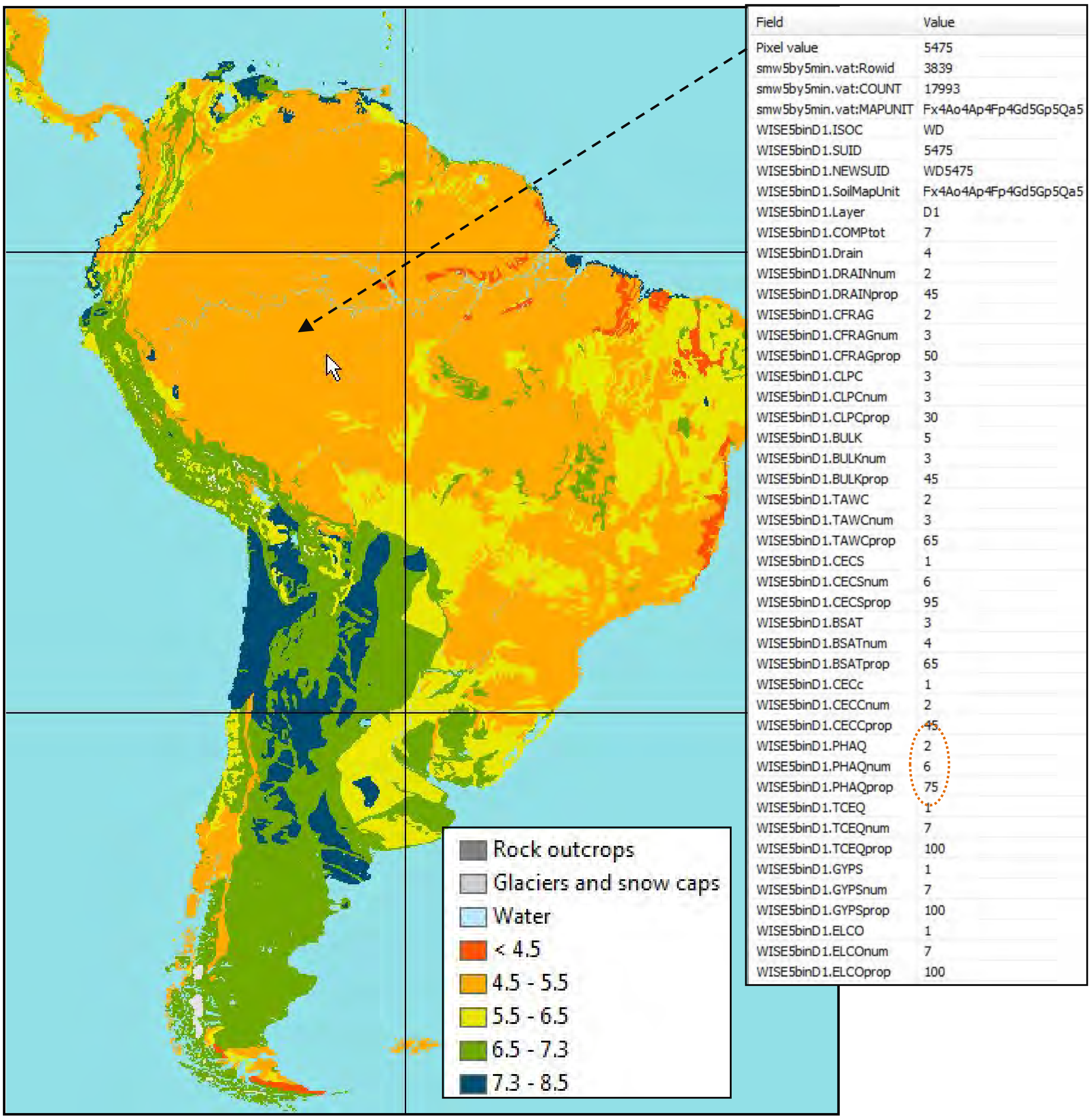


Figure 5
Soil pH (0-20 cm) of the dominant soil unit of mapping units for South America with details for map unit 5475.

Alternatively, when the full mapping unit composition is considered a different map is obtained for the 0-20 cm layer (see Appendix 6, method B). Based on the derived information for the 7 component soil units, the dominant pH_{H2O} class is 2 ($4.5 \leq pH < 5.5$). It covers 75% of the mapping unit and 6 out of the 7 component soil units have a pH_{H2O} that falls within this range. Similar maps may be created for the other soil attributes by depth layer.



Field	Value
Pixel value	5475
smw5by5min.vat:Rowid	3839
smw5by5min.vat:COUNT	17993
smw5by5min.vat:MAPUNIT	Fx4Ao4Ap4Fp4Gd5Gp5Qa5
WISE5binD1.ISOC	WD
WISE5binD1.SUID	5475
WISE5binD1.NEWSUID	WD5475
WISE5binD1.SoilMapUnit	Fx4Ao4Ap4Fp4Gd5Gp5Qa5
WISE5binD1.Layer	D1
WISE5binD1.COMPtot	7
WISE5binD1.Drain	4
WISE5binD1.DRAINnum	2
WISE5binD1.DRAINprop	45
WISE5binD1.CFRAG	2
WISE5binD1.CFRAGnum	3
WISE5binD1.CFRAGprop	50
WISE5binD1.CLPC	3
WISE5binD1.CLPCnum	3
WISE5binD1.CLPCprop	30
WISE5binD1.BULK	5
WISE5binD1.BULKnum	3
WISE5binD1.BULKprop	45
WISE5binD1.TAWC	2
WISE5binD1.TAWCnum	3
WISE5binD1.TAWCprop	65
WISE5binD1.CECS	1
WISE5binD1.CECSnum	6
WISE5binD1.CECSprop	95
WISE5binD1.BSAT	3
WISE5binD1.BSATnum	4
WISE5binD1.BSATprop	65
WISE5binD1.CECc	1
WISE5binD1.CECCnum	2
WISE5binD1.CECCprop	45
WISE5binD1.PHAQ	2
WISE5binD1.PHAQnum	6
WISE5binD1.PHAQprop	75
WISE5binD1.TCEQ	1
WISE5binD1.TCEQnum	7
WISE5binD1.TCEQprop	100
WISE5binD1.GYPS	1
WISE5binD1.GYPSnum	7
WISE5binD1.GYPSprop	100
WISE5binD1.ELCO	1
WISE5binD1.ELCOnum	7
WISE5binD1.ELCOprop	100

Figure 6
Dominant soil pH class (0-20 cm) for South America with details for map unit 5475.

Table *WISstat9_c_AtoF*, described in Appendix 3, shows that the pH_{H20} value for 0-20 cm for xanthic Ferralsols (Fx), in Figure 5, was based on the analysis of 97 observations, after outlier rejection. Alternatively, the estimate for TAWC is based on just 11 observations in this case, which points at a 'low' inferred confidence in the derived data based on the scheme in Table 5. For soil organic carbon, 101 observations have been used in this case giving a mean value of 13.2 g C kg^{-1} for 0-20 cm and reported range of 2.9 to 28.3 g C kg^{-1}. For bulk density this is 1.26 kg m^{-3} with an observed range of 1.01-1.55 kg m^{-3} (n= 22), while for the proportion of coarse fragments this is 5% (range: 1-28%; n= 29). Soil property estimates presented here for organic carbon, bulk density and proportion of coarse fragments thus need not have been derived from the same profiles as gaps are current in the underpinning soil profile dataset.

3.5 Appropriate use of the derived data

The assessment of the accuracy and applicability of any data set for a specific purpose remains a user responsibility. The issue of scale is particularly important in this respect (e.g., Finke 2006). The base map for the DSMW is at scale 1:5 000 000. By implication, DSMW's gridding at 5 by 5 arc minutes will result in multiple grid cells with identical soil units occurring in individual mapping units (polygons) as presented on the original vector map. However, the actual location of the component soil units in a given mapping unit is not known, only their relative proportion. Similar limitations will apply for any gridded information on topsoil textural class, which applies to the dominant soil of a mapping unit (FAO-Unesco 1974, p. 5).

For some studies that consider co-varying soil attributes, it will be necessary to go back to the primary profile data, do the necessary calculations for the relevant attributes by soil horizon, aggregate, and subsequently link results back to the spatial data. As explicitly indicated for an earlier version of the database (Batjes 2006, p. 19), such an approach will be necessary, for example, when computing soil carbon stocks as this involves analysing co-varying data on SOC content and bulk density, as well as the content of coarse fragments, for each soil horizon in a given profile to a pre-defined depth (Eswaran *et al.* 1993; Sombroek *et al.* 1993; Batjes 1996; Jobbagy and Jackson 2000).

Effects of the above 'constraint for use' are illustrated in Table 5 as an example only. It shows the magnitude of SOC stock estimates that can be obtained using various assumptions for the proportion of coarse fragments:

- $CFRAG_{TTR}$: Uses the proportion of coarse fragments as derived from the TTR scheme.
- $CFRAG_{EXR}$: Constrains results for $CFRAG_{TTR}$ using simple expert rules to reduce the possible bias discussed in Section 3.3.2.
- $CFRAG_0$: Assumes that there are no coarse fragments (case only used here to set a possible upper boundary for SOC stocks).

Table 5
Estimates of global soil carbon stocks obtained using three assumptions for the proportion of coarse fragments (Pg C, Pg= 10^{15} g).

	Global SOC stocks (Pg C) [a]		
Depth (m)	$CFRAG_{TTR}$ [b]	$CFRAG_{EXR}$	$CFRAG_0$
0-0.3	601	610	673
0-0.5	818	837	927
0-1	1178	1224	1366

[a] See text for observations concerning the 'appropriate use of the data'; the above values for SOC stocks thus should not be quoted out of context.

[b] $CFRAG_{TTR}$ stands for values for CFRAG derived from the TTR scheme, as is. $CFRAG_{EXR}$ are values for *CFRAG* based on expert rules (EXR) for the proportion of coarse fragments: a) Rendzinas (Ex), Rankers (Ux) or Lithosols (Ix): If $CFRAG_{TTR}$> 50 then $CFRAG_{EXR}$= 50 else $CFRAG_{EXR}$= $CFRAG_{TTR}$ (i.e. no change); b) Regosols (R): $CFRAG_{TTR}$> 40 then $CFRAG_{EXR}$= 40 else $CFRAG_{EXR}$= $CFRAG_{TTR}$; c) Xerosols (X) and Yermosols (Y): if $CFRAG_{TTR}$> 15 then $CFRAG_{EXR}$= 15 else $CFRAG_{EXR}$= $CFRAG_{TTR}$); d) other soil units: If $CFRAG_{TTR}$> 10 then $CFRAG_{EXR}$= 10 Else $CFRAG_{EXR}$= $CFRAG_{TTR}$. Case $CFRAG_0$ assumes a proportion of coarse fragments of 0 per cent for all soil units (to set an upper boundary for the estimates).

Estimates of global SOC stocks to 1 m obtained with the rules for $CFRAG_{EXR}$ (1224 Pg C) are some 4% higher than those that would have been obtained using the unfiltered $CFRAG_{TTR}$ scheme (1178 Pg C). Alternatively, larger estimates will be obtained when a more restrictive expert-scheme is applied to the *CFRAG* data. For example, if *CFRAG* is arbitrarily set at zero ($CFRAG_0$) for all soil units the maximum SOC stock obtainable with the current approach would be 673 and 1366 Pg C for 0-0.3 and 0-1 m, respectively. The latter estimate, however, is still much lower than the value that was obtained using analyses of real soil profiles for each soil unit, with gap-filling, the results of which were then coupled to the spatial data (1462–1548 Pg C to 1 m depth, see Batjes 1996), *albeit* using much less (i.e., 4353) profiles than currently available in WISE 3.1 and a coarser (i.e., ½ by ½ arc-degree) version of the DSMW.

By comparison, if the gap-filling procedure for the virtual profiles had been based on median values even lower estimates of global SOC stocks would have been obtained *in casu*: 536 Pg C (0-0.3 m) and 1050 Pg C (0-1 m), when using the $CFRAG_{TTR}$ values, and 538 Pg C (0-0.3 cm) and 1058 Pg C (0-1 m) when values for $CFRAG_{EXR}$ are used.

The above cases are discussed here for illustrative purposes only. As explicitly stipulated by Batjes (2006 p. 19), such derived data should not be used for the calculation of global SOC stocks, and similar, indicating that such calculations should be based on more demanding analyses of the actual profile data. By implication, the above figures for SOC stocks should not be quoted out of context.

Finally, it should be noted that it is better to aggregate model results — for a given spatial or temporal unit — than to aggregate the spatial data before modelling (Bouwman *et al.* 1999; Middelburg *et al.* 1999).

4 Conclusions

Joining WISE-derived soil property estimates to the spatial component of the DSMW required generalisation of measured soil (profile) data by soil unit and depth zone. This involved the transformation of soil variables that show a marked spatial and temporal variation. These variables have been determined in a range of laboratories, according to various analytical methods, and over a range of years. Other sources and types of uncertainty that are associated with the spatial data and aggregation procedures, and possible implications for modelling have been discussed elsewhere (Cramer and Fischer 1997; Batjes 1999; Bouwman *et al.* 1999; Middelburg *et al.* 1999; Dobos *et al.* 2006).

There are often gaps or omissions in the information provided in the soil literature or in available auxiliary databases with respect to several of the input variables required for WISE. The derived soil property values presented here should be seen as 'best possible' estimates derived from the present selection of measured soil profile data, scheme of taxotransfer and expert-rules, and spatial data of the DSMW. The overall assumption here is that the inferred confidence in the soil property estimates derived for a given combination of 'soil unit-variable-depth zone-soil textural class' should increase with the size of the corresponding sample population. Nonetheless, the uncertainty attached to individual soil property estimates can be large —the types of taxotransfer and expert rules used have been documented in the data set to provide an indication of this uncertainty. Further, some of the soil variables under consideration are only diagnostic for specific soil units (see FAO-Unesco, 1974).

Changes in the number, spatial distribution and type of profiles analysed as well as differences in data clustering and analysis procedures, used for the attribute and spatial data, will lead to different property estimates and binned maps for any given soil variable. Binned maps, developed using broad class limits, are presented here as examples only; typically, users should define their own set of criteria for defining such classes depending on the research questions being asked.

Limitations for use of the derived soil data are described in detail and explicitly state why this type of generalised database may not be used to calculate soil carbon stocks, for example.

The dataset described in this report may be used for broad, exploratory assessments at global scale (< 1:5 million; 5 by 5 minutes and coarser), keeping in mind the generalisations, assumptions, uncertainties and 'appropriate data use' described in this report. Future WISE-derived products at 5 by 5 arcminutes resolution should consider the updated soil geographic information of the Harmonised World Soil Database.

5 References

Anon. 1984. *Memento de l' agronome (troisieme edition)*. Collection "Techniques rurales en Afrique". Ministere des Relations Exterieures - Cooperation et Developpement -, Republique Francaise, Paris, 1604 p

Batjes NH 1996. Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science* 47, 151-163

Batjes NH 1999. Soil vulnerability mapping in Central and Eastern Europe: Issues of data acquisition, quality control and sharing. In: Naff T (editor), *Data Sharing for International Water Resource Management: Eastern Europe, Russia and the CIS*. NATO Science Series 2: Environmental Security (Vol. 61). Kluwer Academic Publishers, Dordrecht, pp 187-206

Batjes NH 2006. *ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.0)*. Report 2006/02, ISRIC - World Soil Information, Wageningen, 50 p http://www.isric.org/isric/webdocs/Docs/ISRIC_Report_2006_02.pdf

Batjes NH 2008. *ISRIC-WISE Harmonized Global Soil Profile Dataset (Version 3.1)*. Report 2008/02, ISRIC - World Soil Information, Wageningen, 47 p http://www.isric.org/isric/webdocs/Docs/ISRIC_Report_2008_02.pdf

Batjes NH 2009. Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. *Soil Use and Management* 25, 124-127

Batjes NH 2010. Soil property estimates for Tunisia derived from SOTER and WISE (SOTWIS-Tunisia, ver. 1.0). ISRIC - World Soil Information, Wageningen, pp 41 http://www.isric.org/isric/Webdocs/Docs/ISRIC_Report_2010_01.pdf

Batjes NH, Al-Adamat R, Bhattacharyya T, Bernoux M, Cerri CEP, Gicheru P, Kamoni P, Milne E, Pal DK and Rawajfih Z 2007. Preparation of consistent soil data sets for SOC modelling purposes: secondary SOTER data sets for four case study areas. *Agriculture, Ecosystems and Environment* 122, 26-34

Beckett PHT and Webster R 1971. Soil variability - a review. *Soils and Fertilizers* 34, 1-15

Bouwman AF, Derwent RG and Dentener FJ 1999. Towards reliable global bottom-up estimates of temporal and spatial patterns of emissions of trace gases and aerosols from land-use related and natural resources. In: Bouwman AF (editor), *Approaches to Scaling of Trace Gas Fluxes in Ecosystems*. Elsevier, Amsterdam, pp 3-26

Burrough PA 1989. Fuzzy mathematical methods for soil surveys and land evaluation. *Journal of Soil Science* 40, 477-492

Carter AJ and Scholes RJ 2002. *Generating a global database of soil properties (Available on: Global Soil Data Products, CD-ROM (IGBP-DIS), USDA, CSIR, IGBP-DIS, FAO and ISRIC)*, IGBP Data and Information Services, Potsdam

CEC 1985. *Soil Map of the European Communities (1:1,000,000)*. Report EUR 8982, Office for Official Publications of the European Communities, Luxembourg

Chadwick MJ and Kuylenstierna JCI 1990. *The relative sensitivity of ecosystems in Europe to acidic deposition - A preliminary assessment of the sensitivity of aquatic and terrestrial ecosystems*, Stockholm Environmental Institute, University of York, York

Cramer W and Fischer A 1997. Data requirements for global terrestrial ecosystem modelling. In: Walker B and W Steffen (editors), *Global Change and Terrestrial Ecosystems*. Cambridge University Press, Cambridge, pp 529-565

Dijkshoorn JA, Huting JRM and Tempel P 2005. *Update of the 1:5 million Soil and Terrain Database for Latin America and the Caribbean (SOTERLAC, ver. 2.0)*. Report 2005/01, ISRIC - World Soil Information, Wageningenhttp://www.isric.org/sites/default/files/ISRIC_Report_2005_01.pdf

Dobos E, F. C, Hengl T, Reuter HI and Toth G 2006. *Digital soil mapping as a support to production of functional maps*. EUR 22123 EN, Office for Official Publications of the European Communities, Luxembourg, 68 p

Eswaran H, van den Berg E and Reich P 1993. Organic carbon in soils of the world. *Soil Science Society of America Journal* 57, 192-194

FAO-Unesco 1971-1981. *FAO-Unesco Soil Map of the World, 1:5,000,000 (Vol. 1 to 10)*. United Nations Educational, Scientific, and Cultural Organization, Paris

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 1981. *Soil Map of the World, 1:5,000,000. Vol. 5 - Europe*. United Nations Educational, Scientific, and Cultural Organization, Paris

FAO 1983. *Guidelines: land evaluation for rainfed agriculture*. FAO Soils Bulletin 52, Food and Agriculture Organization of the United Nations, Rome, 237 p

FAO 1988. *FAO-Unesco Soil Map of the World, Revised Legend, with corrections and updates*. World Soil Resources Report 60, FAO, Rome; reprinted with updates as Technical Paper 20 by ISRIC, Wageningen,1997, 140 p

FAO 1995. *Digital Soil Map of the World and derived properties (ver. 3.5)*. FAO Land and Water Digital Media Series # 1, FAO, Rome, 23 p

FAO 2006. *Guidelines for soil description (Fourth ed.)*, FAO, Rome, 97 p ftp://ftp.fao.org/agl/agll/docs/guidel_soil_descr.pdf

FAO/IIASA/ISRIC/ISSCAS/JRC 2012. *Harmonized World Soil Database (version 1.2)*, Food and Agriculture Organization of the United Nations, International Institute for Applied Systems Analysis, ISRIC - World Soil Information, Institute of Soil Science - Chinese Academy of Sciences, Joint Research Centre of the European Commission, Laxenburg

Finke P 2006. Quality assessment of digital soil maps: producers and users perspectives. In: Lagacherie P, A McBratney and M Voltz (editors), *Digital soil mapping: An introductory perspective*. Elsevier, Amsterdam, pp 523-541

Hartemink AE, McBratney A and Mendoca-Santos ML (editors) 2008. Digital soil mapping with limited data. Springer, 445 p

IUSS Working Group WRB 2006. *World Reference Base for Soil Resources (2nd ed.)*. World Soil Resources Report 103, FAO, Rome, 145 p http://www.fao.org/ag/agl/agll/wrb/doc/wrb2006final.pdf

Jobbagy E and Jackson R 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications* 10, 423 - 436

Kern JS 1995. Geographic patterns of soil-water holding capacity in the contiguous United States. *Soil Sci. Soc. Am. J.* 59

Klamt E and Sombroek WG 1988. Contribution of organic matter to exchange properties of Oxisols. In: Beinroth FH, MN 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

Landon JR 1991. *Booker Tropical Soil Manual*. Longman Scientific & Technical, New York, 474 p

McBratney AB and Odeh IOA 1997. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma* 77, 85-113

Middelburg JJ, Liss PS, Dentener FJ, Taminski T, Kroeze C, Malingreau J-P, Noväl M, Panikov NS, Plant R, Starink M and Wanninkhof R 1999. Relations between scale, model approach and model parameters. In: Bouwman AF (editor), *Approaches to Scaling of Trace Gas Fluxes in Ecosystems*. Elsevier, Amsterdam, pp 219-232

Nachtergaele FO, Van Engelen VWP and Batjes NH 2011. Qualitative and quantitative aspects of world and regional soil databases and maps. In: Pan Ming Huang, Yuncong Li and ME Sumner (editors), *Handbook of Soil Sciences: Resource management and environmental aspects (2nd ed.)*. CRC Press, Boca Raton, pp 27-1/10

Pleijsier K 1989. Variability in soil data. In: Bouma J and AK Bregt (editors), *Land Qualities in Space and Time*. PUDOC, Wageningen, pp 89-98

Sanchez PA, Ahamed S, Carre F, Hartemink AE, Hempel J, Huising J, Lagacherie P, McBratney AB, McKenzie NJ, Mendonca-Santos MdL, Minasny B, Montanarella L, Okoth P, Palm CA, Sachs JD, Shepherd KD, Vagen T-G, Vanlauwe B, Walsh MG, Winowiecki LA and Zhang G-L 2009. Digital Soil Map of the World. *Science* 325, 680-681

Soil Survey Staff 1983. *Soil Survey Manual (rev. ed.)*. United States Agriculture Handbook 18, USDA, Washington

Sombroek WG, Nachtergaele FO and Hebel A 1993. Amounts, dynamics and sequestering of carbon in tropical and subtropical soils. *Ambio* 22, 417-426

Spain AV, Isbell RF and Probert ME 1983. Soil organic matter, *Soils: An Australian Viewpoint*. CSIRO Melbourne Academic Press, London, pp 552-563

Sys IC, van Ranst E, Debaveye IJ and Beenaert F 1993. *Land evaluation (Part I - III)*, Agricultural Publications, General Administration for Development Cooperation, Brussels, 199 p

Van Engelen VWP 1999. SOTER: The World Soils and Terrain Database. In: Sumner ME (editor), *Handbook of Soil Science*. CRC Press, Boca Raton, pp H19-28

Van Engelen VWP 2011. Standardizing soil data (*e-SOTER* regional pilot platform as EU contribition to a Global Soil Information System). *International Innovation* June, 48-49

Van Engelen VWP and Wen TT 1995. *Global and National Soils and Terrain Digital Databases (SOTER): Procedures Manual (rev. ed.)*. (Published also as FAO World Soil Resources Report No. 74), UNEP, IUSS, ISRIC and FAO, Wageningen, 125 p http://www.isric.org/isric/webdocs/docs/SOTER_Procedures_Manual_1995.pdf

Van Reeuwijk LP 1998. *Guidelines for quality management in soil and plant laboratories*, FAO, Rome, 143 p http://www.fao.org/docrep/W7295E/W7295E00.htm

Van Reeuwijk LP 2002. *Procedures for soil analysis (6th ed.)*. Technical Paper 9, ISRIC, Wageningen, 1.1 to 19.5 p http://www.isric.org/Isric/Webdocs/Docs/ISRIC_TechPap09_2002.pdf

Van Reeuwijk LP and Houba VJG 1998. External quality control of data. In: Van Reeuwijk LP (editor), *Guidelines for quality management in soil and plant laboratories*. ISRIC and FAO, Rome, pp 121-127

Appendix 1. Number of profiles by FAO soil unit in WISE3.1

A: Acrisols (1081)[a]
Af= 390 Ag= 86 Ah= 84 Ao= 392 Ap= 129
B: Cambisols (1343)
Bc= 72 Bd= 200 Be= 336 Bf= 156 Bg= 137 Bh= 131 Bk= 192 Bv= 81 Bx= 38
C: Chernozems (206)
Cg= 6 Ch= 75 Ck= 61 Cl= 64
D: Podzoluvisols (115)
Dd= 22 De= 70 Dg= 23
E: Rendzinas (103)
F: Ferralsols (533)
Fa= 34 Fh= 68 Fo= 210 Fp= 15 Fr= 98 Fx= 108
G: Gleysols (649)
Gc= 16 Gd= 108 Ge= 280 Gh= 73 Gm= 107 Gp= 48 Gx= 17
H: Phaeozems (400)
Hc= 39 Hg= 35 Hh= 134 Hl= 192
I: Lithosols (27)
J: Fluvisols (575)
Jc= 146 Jd= 92 Je= 283 Jt= 54
K: Kastanozems (97)
Kh= 27 Kk= 44 Kl= 26
L: Luvisols (1542)
La= 37 Lc= 227 Lf= 391 Lg= 289 Lk= 119 Lo= 404 Lp= 39 Lv= 36
M: Greyzems (27)
Mg= 3 Mo= 24
N: Nitosols (163)
Nd= 55 Ne= 75 Nh= 33
O: Histosols (112)
Od= 63 Oe= 38 Ox= 11
P: Podzols (222)
Pf= 7 Pg= 43 Ph= 49 Pl= 20 Po= 89 Pp= 14
Q: Arenosols (771)
Qa= 38 Qc= 423 Qf= 163 Ql= 147
R: Regosols (521)
Rc= 139 Rd= 144 Re= 224 Rx= 14
S: Solonetz (226)
Sg= 62 Sm= 30 So= 134
T: Andosols (293)
Th= 148 Tm= 50 To= 35 Tv= 60
U: Rankers (53)
V: Vertisols (549)
Vc= 226 Vp= 323
W: Planosols (171)
Wd= 36 We= 85 Wh= 7 Wm= 15 Ws= 28 Wx= 0
X: Xerosols (185)
Xh= 33 Xk= 76 Xl= 62 Xy= 14
Y: Yermosols (118)
Yh= 20 Yk= 34 Yl= 44 Yt= 1 Yy= 19
Z: Solonchaks (167)
Zg= 44 Zm= 9 Zo= 111 Zt= 3

[a] For details see Legend (FAO-Unesco 1974). For computational reasons, the codes for Lithosols (I), Rendzinas (E), and Rankers (U) have been changed to Ix, Ex and Ux in the database.

Appendix 2. Soil map unit composition file

Table *WISEunitComposition* gives the full composition of each DSMW map unit in terms of its main soil units (FAO-Unesco, 1974), their relative extent, and the identifier for the corresponding virtual soil profile. The contents of this table can be joined to the raster data (*VALUE* field) using the *SUID* field .

Structure of table WISEunitComposition[a]

Name	Type	Description
ISOC	Text	ISO-3166 country code (1994) or WD for World
SUID	Integer	The identification code of a DSMW unit on the map and in the database (corresponds with SNUM on the DSMW)
NEWSUID	Text	Globally unique code, comprising fields ISOC plus SUID (e.g. WD1234)
SoilMapunit	Text	Aggregated code for map unit summarizing the overall composition (see text)
SOIL1	Text	Characterization of the first (main) soil unit according to The FAO-Unesco Legend
PROP1	Integer	Proportion, as a percentage, that the main soil unit occupies within the DSMW unit
PRID1	Text	Unique code for the corresponding virtual soil profile (e.g WD-Fr)
SOIL2	Text	As above but for the next soil unit
PROP2	Integer	As above
PRID2	Text	As above
SOIL3	Text	As above but for the next soil unit
PROP3	Integer	As above
PRID3	Text	As above
SOIL4	Text	As above but for the next soil unit
PROP4	Integer	As above
PRID4	Text	As above
SOIL5	Text	As above but for the next soil unit
PROP5	Integer	As above
PRID5	Text	As above
SOIL6	Text	As above but for the next soil unit
PROP6	Integer	As above
PRID6	Text	As above
SOIL7	Text	As above but for the next soil unit
PROP7	Integer	As above
PRID7	Text	As above
SOIL8	Text	As above but for the next soil unit
PROP8	Integer	As above
PRID8	Text	As above

[a] For the sake of consistency, table structure conventions used for secondary SOTER databases have been retained here -table WISExxx has the same structure as table SOTERxxx.

Property estimates for any given virtual profile can also be linked to the spatial data through the unique profile identifier (PRID, see Appendix 3 and Figure 3).

Appendix 3. Statistical output files

Tables with names like *WISstat9_x_YtoZ* show results of the statistical analyses of measured soil data in WISE, by FAO-Unesco (1974) unit, soil attribute, depth layer, and soil textural class. The corresponding information has been stored in 16 tables —in view of their length, these tables are only available in digital format (via www.isric.org):

- *WISstat9_a_AtoF*: This table holds statistical data by *major soil groupings* (FAO 1974), ranging from Acrisols (A) to Ferralsols (F), considering five soil textural classes: **C**oarse, **M**edium, Medium Fine (**Z**), **F**ine, and **V**ery Fine (Figure 8). Where there is no measured particle size analysis for a given layer this has been flagged by '-'. The textural class of Histosols is coded 'O' to differentiate these organic soils from mineral soil types.

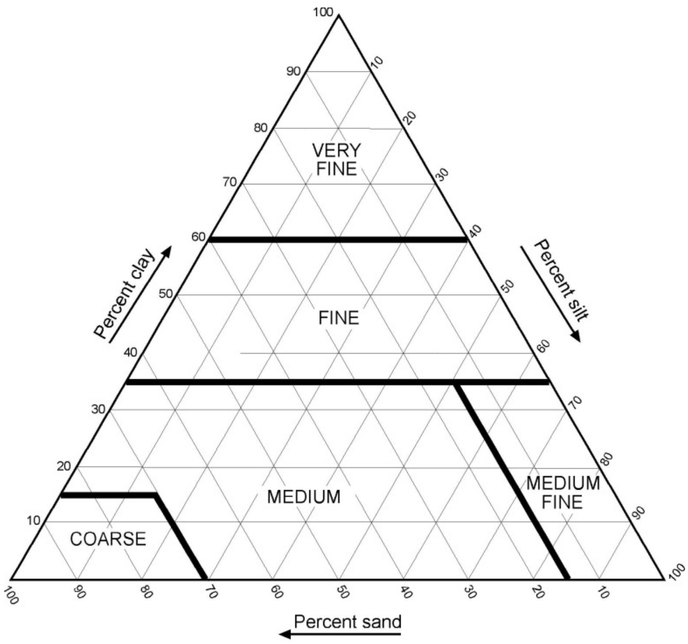


Figure 7
Soil textural classes.

- *WISstat9_a_GtoM*: Similar to above but for *major soil groupings* ranging from Gleysols (G) to Greyzems (M)
- *WISstat9_a_NtoS*: Similar to above but for *major soil groupings* ranging from Nitosols (N) to Solonetz (S)
- *WISstat9_a_TtoZ*: Similar to above but for *major soil groupings* ranging from Andosols (T) to Solonchaks (Z)
- *WISstat9_b_AtoF*: This table holds statistical data by *soil unit* (FAO 1988), ranging from Acrisols (A) to Ferralsols (F), considering all five soil textural classes
- *WISstat9_b_GtoM*: Similar to above but for *soil units* ranging from Gleysols (G) to Greyzems (M)
- *WISstat9_b_NtoS*: Similar to above but for *soil units* ranging from Nitosols (N) to Solonetz (S)
- *WISstat9_b_TtoZ*: Similar to above but for *soil units* ranging from Andosols (T) to Solonchaks (Z)

Structure of statistical output tables like WISstat9_d_YtoZ[*].

Name	Type	Description
Short_Id	Text	Code comprising abbreviation for FAO major soil Grouping (resp. soil unit), attribute, depth layer, and soil textural class (e.g., ABSATD1M)
Num0	Integer	Number of observations (before outlier rejection)
Num	Integer	Number of observations (after outlier rejection)
Mean	Single	Mean
STD	Single	Standard deviation
CV	Single	Coefficient of variation
Median	Single	Median
MAD	Single	Median of absolute deviations
Min	Single	Minimum
Max	Single	Maximum
Var	Single	Variance
Fao_74	Text	FAO Legend code (this field is intentionally left blank)

Notes:
1) Applies to tables: _WISstat9_a_AtoF, _WISstat9_a_GtoM, _WISstat9_c_NtoS, and _WISstat9_c_TtoZ.
2) The structure of tables _WISstat9_b_AtoF, _WISstat9_b_GtoM, _WISstat9_d_NtoS, and _WISstat9_d_TtoZ, is similar to the one listed above except that the first field is called Long_ID. This field differs from Short_ID in that the first two letters refer to the FAO soil unit code (e.g., AfBSATD1M for the sample set that relates to base saturation data, BSAT) for Ferric Acrisols (Af) that have medium (M) texture and belong to layer D1 (i.e. 0 to 20 cm).
3) Descriptive statistics are for depth-weighted data, per layer (from D1 to D5, see text)
4) These tables list results for all analyses after outlier rejection. The taxotransfer scheme, however, will only consider means from the corresponding tables when Num > 5 (see n_{WISE} in text; Appendix 4).

– *WISstat9_c_AtoF*: This table holds statistical data by *soil unit*, ranging from Acrisols (A) to Ferralsols (F), irrespective of soil texture. This has been flagged as class 'u', for undifferentiated, which comprises soil textural classes C, M, Z, F and V as well as '–'.
– *WISstat9_c_GtoM*: Similar to above but for *soil units* ranging from Gleysols (G) to Greyzems (M)
– *WISstat9_c_NtoS*: Similar to above but for *soil units* ranging from Nitosols (N) to Solonetz (S)
– *WISstat9_c_TtoZ*: Similar to above but for *soil units* ranging from Andosols (T) to Solonchaks (Z)
– *WISstat9_d_AtoF*: This table holds statistical data by *soil unit*, ranging from Acrisols (A) to Ferralsols (F), irrespective of soil texture. This has been flagged as class 'u', for undifferentiated.
– *WISstat9_d_GtoM*: Similar to above but for *major soil groupings* ranging from Gleysols (G) to Greyzems (M)
– *WISstat9_d_NtoS*: Similar to above but for *major soil groupings* ranging from Nitosols (N) to Solonetz (S)
– *WISstat9_d_TtoZ*: Similar to above but for *major soil groupings* ranging from Andosols (T) to Solonchaks (Z)

Appendix 4. Taxotransfer rule-based soil property estimates

Table *WISEparameterEstimates* lists soil propertys estimates for all virtual profiles considered in the derived data set . This information can be linked to the geographical component of the DSMW —in a GIS— through the unique profile code (PRID, see Appendix 2).

Structure of table WISEparameterEstimates.

Name	Type	Description
CLAF	Text	FAO-Unesco (1974) Legend code
PRID	Text	Profile ID (as documented in table WISEunitCo*mposition*)
Drain	Text	FAO soil drainage class
Layer	Text	Code for depth layer (from D1 to D5; e.g. D1 is from 0 to 20 cm)
TopDep	Integer	Depth of top of layer (cm)
BotDep	Integer	Depth of bottom of layer (cm)
CFRAG	Integer	Coarse fragments (vol% > 2 mm)
SDTO	Integer	Sand (mass %)
STPC	Integer	Silt (mass %)
CLPC	Integer	Clay (mass %)
PSCL	Text	FAO texture class
BULK	Single	Bulk density (kg dm^{-3})
TAWC	Integer	Available water capacity (cm m^{-1}, -33 to -1500 kPa conform to USDA standards)
CECs	Single	Cation exchange capacity ($cmol_c$ kg^{-1}) for fine earth fraction
BSAT	Integer	Base saturation as percentage of CEC_{soil}
CECc	Single	CEC_{clay}, corrected for contribution of organic matter ($cmol_c$ kg^{-1})
PHAQ	Single	pH measured in water
TCEQ	Single	Total carbonate equivalent (g C kg^{-1})
GYPS	Single	Gypsum content (g kg^{-1})
ELCO	Single	Electrical conductivity (dS m^{-1})
TOTC	Single	Organic carbon content (g C kg^{-1})
TOTN	Single	Total nitrogen (g kg^{-1})
CNrt	Single	C/N ratio
ECEC	Single	Effective CEC ($cmol_c$ kg^{-1})

Notes:
A minus 3 indicates that no meaningful substitution was possible for the specified soil unit and attribute using the present selection of soil profiles, -1 is used for Oceans and inland waters, -2 for Glaciers and snow caps, -7 for rock outcrops (or shallow subsoils) to facilitate visualization using GIS.

The above table should be consulted in conjunction with table *WISEflagTTRrules* that documents the taxotransfer and expert rules that have been applied (see Appendix 5). Table 6 lists conventions for coding attributes used in the taxotransfer scheme.

Table 6
Conventions for coding attributes in the taxotransfer scheme.

SOTnam	WISnam	SoilVariable	TTRflag	Comments[a]
ALSA	ALSA	ALSAT	A	Exchangeable Aluminium percentage (% of ECEC)
BSAT	BSAT	BSAT	B	Base saturation (% of CECs)
BULK	BULK	BULKDENS	C	Bulk density
CECC	CECC	CECCLAY	D	Cation exchange capacity of clay fraction (corrected for organic C)
CECS	CECS	CECSOIL	E	Cation exchange capacity
CFRAG	GRAV	GRAVEL	F	Coarse fragments % (> 2 mm)
CLPC	CLAY	CLAY	G	Clay %
ECEC	ECEC	ECEC	H	Effective CEC
ELCO	ECE	ECE	I	Electrical conductivity
ESP	ESP	ESP	J	Exchangeable Na percentage (as % of CECs)
GYPS	GYPS	GYPSUM	K	Gypsum content
PHAQ	PHH2	PHH2O	L	PH in water
SDTO	SAND	SAND	M	Sand %
STPC	SILT	SILT	N	Silt %
	TAWC	TAWC	O	Volumetric water content (-33 to -1500 kPa, cm m^{-1})
TCEQ	CACO	CACO3	P	Carbonate content
TOTC	ORGC	ORGC	Q	Organic carbon content
TOTN	TOTN	TOTN	R	Total nitrogen content
CN	CNrt	CN	Z	C/N ratio

[a] See Table 3 for units of measurement. C/N ratios have been calculated 'as is' from the measured data, not as the ratio of the derived values for C and N; ditto for CEC_{clay}.

Appendix 5. Flagging taxotransfer rules

The type of taxotransfer and expert rules that have been used when creating table *WISEparameterEstimates* (see Appendix 2) is documented in table *WISEflagTTRrules*. Further details on coding conventions may be found in the text (Section 3.3).

Structure of table WISEflagTTRrules.

Name	Type	Description
CLAF	Text	FAO Legend code
PRID	Text	Unique identifier for representative profile
Topdep	Integer	Depth of top of layer (cm)
Botdep	Integer	Depth of bottom of layer (cm)
TTRsub	Text	Codes showing the type of taxotransfer rule used (based on data for soil *units*; see text)
TTRmain	Text	Codes showing the type of taxotransfer rule used (based on data for *major units*; see text)
EXR	Text	Codes for expert rules

Note:
The exchangeable aluminium percentage (ALSA) has been set at zero when pH_{water} is higher than 5.5. Similarly, the electrical conductivity (ELCO), content of gypsum (GYPS) and content of carbonates (TCEQ) have been set at zero when pH_{water} is less than 6.5.

Appendix 6. Summary files of derived soil properties

To facilitate access to the derived data, a so-called *summary file* has been created. This file summarizes un-binned derived soil properties for all component soil units in a given grid cell (e.g. SNUM or SUID) and layer (i.e., D1 to D5), per soil variable.

Structure of table WISEsummaryFile[3].

Name	Type	Description
ISOC	Text	ISO-3166 country code (1994); WD stands for World
SUID	Integer	The identification code of a DSMW unit on the map and in the database (i.e., SNUM)
NEWSUID	Text	Globally unique map unit code, comprising fields ISOC plus SUID (or SNUM)
TCID	Integer	Number of terrain component in given map unit (Relevant only for SOTER databases; set at 1 by default for the DSMW)
SCID	Integer	Number of soil unit within the given DSMW unit (ranges from 1 to 8 for DSMW)
Layer	Text	Code for depth layer (from D1 to D5; e.g. D1 is from 0 to 20 cm and D5 from 80 to 100 cm)
PROP	Integer	Relative proportion of above in given DSMW unit
CLAF	Text	FAO-Unesco Legend code
PRID	Text	Profile ID (as documented in table *WISE- unitComposition*)
Drain	Text	FAO soil drainage class
TopDep	Integer	Upper depth of layer (cm)
BotDep	Integer	Lower depth of layer (cm)
CFRAG	Integer	Coarse fragments (vol. % > 2mm)
SDTO	Integer	Sand (mass %)
STPC	Integer	Silt (mass %)
CLPC	Integer	Clay (mass %)
PSCL	Text	FAO texture class (see Figure 8)
BULK	Single	Bulk density (kg dm^{-3})
TAWC	Integer	Available water capacity (cm m^{-1}, -33 to -1500 kPa, USDA standards)
CECS	Single	Cation exchange capacity ($cmol_c$ kg^{-1}) of fine earth fraction
BSAT	Integer	Base saturation as percentage of CECsoil
CECc	Single	CECclay, corrected for contribution of organic matter
PHAQ	Single	pH measured in water
TCEQ	Single	Total carbonate equivalent (g C kg^{-1})
GYPS	Single	Gypsum content (g kg^{-1})
ELCO	Single	Electrical conductivity (dS m^{-1})
TOTC	Single	Organic carbon content (g kg^{-1})
TOTN	Single	Total nitrogen (g kg^{-1})
CNrt	Single	C/N ratio
ECEC	Single	Effective CEC ($cmol_c$ kg^{-1})

Notes:
These are depth-weighted values, per 20 cm layer.
Components within a given DSMW unit are numbered starting with the spatially dominant one. The sum of the relative proportions of all soil units within a DSMW unit is always 100 per cent.
Tables having the same structure have also been prepared for the DOMINANT soil unit only, by depth layer (i.e., for layer D1 to D5, see files: *WISEsummaryFile_T1S1D1* etc. to facilitate linkage to GIS).

3 The structure of this table is indentical to that used for secondary SOTER databases (see: SOTER*summaryFile).*

In view of the map unit complexity (see Figure 1), additional operations will often be needed before results can be visualized (meaningfully) using GIS. As indicated in the report, two types of simplified and generalised output files were generated to facilitate the linkage:

<u>Method A</u>:
Yields simplified tables showing un-binned soil property estimates for the dominant soil unit only, for each 20 cm layer up to 100 cm depth (see Figure 5). The structure of this table is similar to that of table WISE*summaryFile* (with as extra selection TCID=1 and SCID=1, and Layer= D1 to resp. D5); clearly, such files give a highly simplified view.

In view of the length, and visibility in the ArcMAP 'symbology attribute frame' the corresponding table names have been shortened to *WISEsfT1S1Di*, where *i*=1 to 5 and *sf* stands for summary file.

<u>Method B</u>:
Gives generalised tables showing binned or classified data for each grid cell and 20 cm layer that take into account the full map unit composition (see Figure 6). The corresponding tables are called *WISE5binDi,* where *Di* stands for the 20 cm layer under consideration. Criteria for allocating each grid cell to a given legend unit, per soil variable, are presented in Appendix 7 to allow for delineation of a limited number of broadly defined classes, commensurate with the 1:5M sale of the DSMW.

Structure of tables WISE5binDi.

Name	Type	Description
ISOC	Text	ISO-3166 country code (1994), WD for WORLD
SNUM	Integer	The identification code of a DSMW unit, or grid cell, on the map and in the database
MAPUNIT	Text	Globally unique map unit code, comprising fields ISOC plus SNUM, e.g WD123
Layer	Text	Code for depth layer *Di* (from D1 to D5; e.g. D1 is from 0 to 20 cm)
COMPtot	Integer	Total number of soil units or miscellaneous units in grid cell or map unit
SOLprop	Integer	Relative (total) proportion of above soil units
DRAIN	Integer	FAO soil drainage class
DRAINnum	Integer	Number of soil units considered when rating above class
DRAINrop	Integer	Relative area of above soil units (%)
CFRAG	Integer	Class number for Coarse fragments (see Appendix 7 for Legend)
CFRAGnum	Integer	Number of soil units considered when rating above class
CFRAGprop	Integer	Relative area of above soil units (%)
CLPC	Integer	Class for Clay %
CLPCnum	Integer	Number of soil units considered when rating above class
CLPCprop	Integer	Relative area of above soil units (%)
BULK	Integer	Class for Bulk density (kg dm^{-3})
BULKnum	Integer	Number of soil units considered when rating above class
BULKprop	Integer	Relative area of above soil units (%)
TAWC	Integer	Class for Available water capacity (cm m^{-1}, -33 to -1500 kPa, USDA standards)
TAWCnum	Integer	Number of soil units considered when rating above class
TAWCprop	Integer	Relative area of above soil units (%)
CECS	Single	Class for cation exchange capacity ($cmol_c$ kg^{-1}) of fine earth fraction
CECSnum	Integer	Number of soil units considered when rating above class
CECSprop	Integer	Relative area of above soil units (%)
BSAT	Integer	Class for base saturation as percentage of CECsoil
BSATnum	Integer	Number of soil units considered when rating above class
BSATprop	Integer	Relative area of above soil units (%)
CECc	Single	Class for CECclay, corrected for contribution of organic matter ($cmol_c$ kg^{-1})

Name	Type	Description
CECCnum	Integer	Number of soil units considered when rating above class
CECCprop	Integer	Relative area of above soil units (%)
PHAQ	Single	Class for pH, measured in water
PHAQnum	Integer	Number of soil units considered when rating above class
PHAQprop	Integer	Relative area of above soil units (%)
TCEQ	Single	Class for total carbonate equivalent (g kg^{-1})
TCEQnum	Integer	Number of soil units considered when rating above class
TCEQprop	Integer	Relative area of above soil units (%)
GYPS	Single	Class for gypsum content (g kg^{-1})
GYPSnum	Integer	Number of soil units considered when rating above class
GYPSprop	Integer	Relative area of above soil units (%)
ELCO	Single	Class for electrical conductivity (dS m^{-1})
ELCOnum	Integer	Number of soil units considered when rating above class
ELCOprop	Integer	Relative area of above soil units (%)
TOTC	Single	Class for organic carbon content (g kg^{-1})
TOTCnum	Integer	Number of soil units considered when rating above class
TOTCprop	Integer	Relative area of above soil units (%)
TOTN	Single	Class for Total nitrogen (g kg^{-1})
TOTNnum	Integer	Number of soil units considered when rating above class
TOTNprop	Integer	Relative area of above soil units (%)
CNrt	Single	Class for C/N ratio
CNRTnum	Integer	Number of soil units considered when rating above class
CNRTprop	Integer	Relative area of above soil units (%)

Note:
The abbreviation *TOTC*, a SOTER code, is only used here as 'field code'. However, the associated data are for organic carbon, or *ORGC* as used in WISE.

Appendix 7. Legends for binned data sets

Binned data sets, called *WISE5binDi (see Appendix 6),* have been prepared for each depth layer and soil property. These take into account the full mapping unit composition and spatially dominant class for the selected soil properties, unlike files like WISE*sfT1S1Di* in Appendix 6 that only consider derived soil properties for the dominant soil unit in a mapping unit for each 20 cm layer to 100 cm depth.

Selection of suitable class limits for the binned or classified data will depend on the uses to which the data are to be put (see FAO 1983; Anon. 1984; Chadwick and Kuylenstierna 1990; Landon 1991; Sys *et al.* 1993). Such class limits are inherently fuzzy (Burrough 1989; McBratney and Odeh 1997) and the spread of the soil property estimates is often large (see Appendix 3). In view of the 1:5 million scale, and related present generalisations, the legends for the binned GIS files comprise from 5 to 6 classes —depending on the soil variable under consideration— plus three miscellaneous classes (RK, GL and WR).

The class boundaries below are broadly defined and mainly serve as an example; other class limits may be required for specific applications.

Class number:		1	2	3	4	5	6	Max.
		<——>	<——>	<——>	<——>	<——>	<——>	
Class limits:	A0	A1	A2	A3	A4	A5	A6	
Soil variable:								
ALSA	0	20	40	60	80	10000	10000	(83)
BSAT	0	20	40	60	80	10000	10000	(83)
BULK	0	0.4	0.8	1.2	1.4	1.6	10000	(1.79)
CECC	0	8	16	24	48	72	10000	(149)
CECS	0	10	20	30	40	60	10000	(103)
CFRAG	0	5	10	15	30	45	10000	(53)
CLPC	0	10	20	30	40	50	10000	(67)
CNrt	0	10	15	20	25	10000	10000	(30)
DRAIN	V, P	I	M	W	E, SE	-	-	-
ELCO	0	5	10	15	20	25	10000	(31)
ESP	0	5	15	30	45	60	10000	(98)
GYPS	0	20	40	80	160	320	10000	(416)
PHAQ	0	4.5	5.5	6.5	7.3	8.5	10000	(8.8)
TAWC	0	8	12	16	24	32	10000	(52)
TCEQ	0	20	40	80	160	320	10000	(505)
TOTC	0	10	30	60	120	240	10000	(407)
TOTN	0	1	3	6	12	24	10000	(27)

Notes:

1) Soil drainage was aggregated into five classes using: class 1 for Very poorly and Poorly drained soils; class 2 for Imperfectly drained soils; class 3 for Moderately well drained soils; class 4 for Well-drained soils; and, class 5 for Somewhat excessively and Excessively drained soils. Rock outcrops were assigned to the class Excessively drained.

2) The value of 10000 has been used as an artificial upper limit to define the highest-class (*CL*). For example, class 6 for bulk density encompasses values ≥ 1.6 and < 10000. The number of classes considered (excluding classes -1, -2 and -7) ranges from 5 to 6. In addition, should there be no estimates for a given soil property this has been flagged as class -3, for no data. Values in the column Max are the greatest (derived) values reported for the soil variable, irrespective of soil layer. The full range in soil properties may be found in the tables that are described in Appendix 3; units of measurement are given in Appendix 4.
3) Pragmatically, the same legend has been used for a given soil variable, irrespective of the depth layer under consideration (D1 to D5). Not all these classes, however, need occur for all depth ranges.

Class limits for mapping soil pH are listed below, as an example. In addition to listing the pH class (*PHAQ*), table *WISE5bin_Di* also lists the number of soil units whose properties fall within the given class (*PHAQnum*) in the grid cell under consideration, and the relative area (*PHAQprop*) thereof. It also lists the total number of soil units that occur in the given grid cell (*COMPtot*; see Appendix 6, under Method B) and the proportion thereof that falls in the considered class (e.g. *PHAQnum*).

Example of classes used for mapping soil pH_{H2O}:

Class 1:	0 <= PHAQ < 4.5
Class 2:	4.5 <= PHAQ < 5.5
Class 3:	5.5 <= PHAQ < 6.5
Class 4:	6.5 <= PHAQ < 7.3
Class 5:	7.3 <= PHAQ < 8.5
Class 6:	8.5 <= PHAQ < 14.0 (symbolised by 10000 in above table)
Class -1:	Oceans and inland waters, predominant in grid cell
Class -2:	Glaciers and snowcaps, predominant in grid cell
Class -7:	Rock outcrops, respectively shallow rocky soils, predominant in grid cell for specified depth layer. For example, in case of Lithosols, soil property estimates for layer D2 to D5 have been set at -7, while for Rendzinas and Rankers this has been done for layers D3 to D5 (Class -3 stand for No Data).

The dominant, binned pH class for 0-20 cm is shown Figure 8 as an example. Figure 9 shows the proportion of the mapping unit for which the above classes apply, ranging from 25 to 100%. Similar maps can be created for the other layers and attributes using the information in tables *WISE5binDi*; see Appendix 6 for the table's structure.

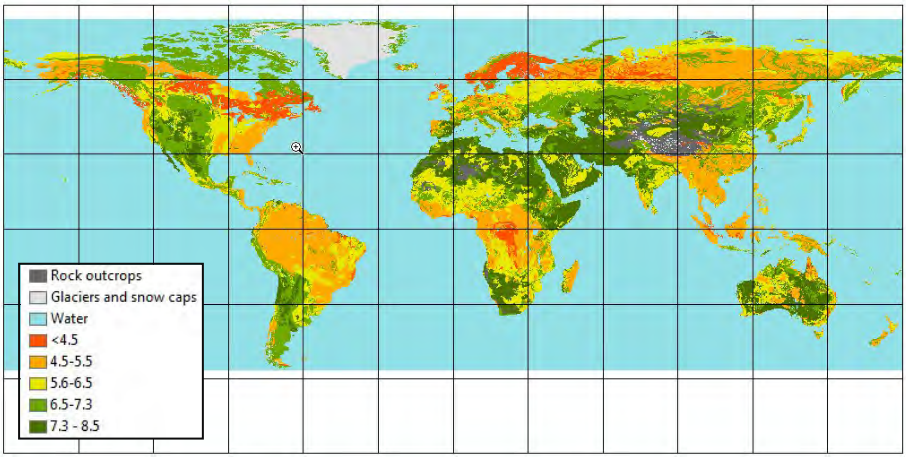


Figure 8
Binned soil pH classes for the world (0-20 cm).

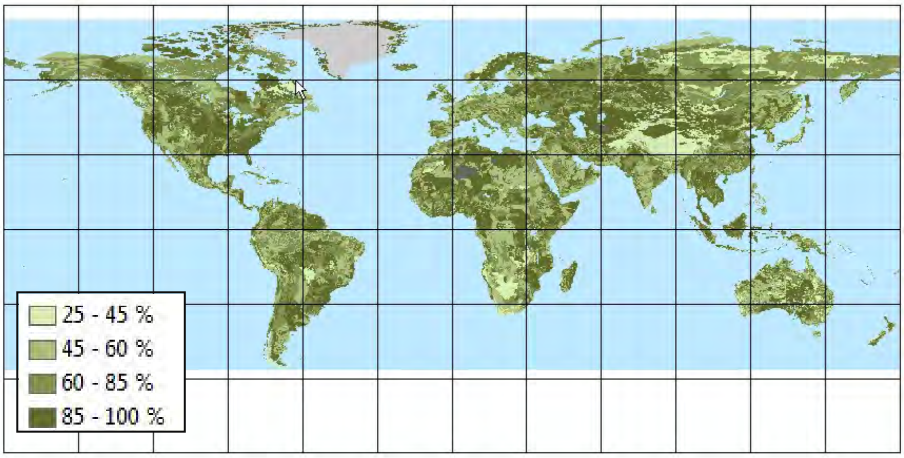


Figure 9
Proportion of mapping unit represented by pH classes shown in Figure 8.

Appendix 8. Revision notes

Latest version: Updates and corrections to the ISRIC-WISE 5 by 5 arc-minutes dataset (ver. 1.2, 28 Feb. 2012) include:

a) The taxotransfer procedure for application with the FAO74 Legend was screened and minor bugs corrected. The procedure was revised to better suit application to *virtual* profiles necessary to describe properties for the various soil units of the Digital Soil Map of the World (for which there are no measured data), as opposed to *real* profiles (that may have gaps in the measured data) as considered in SOTER databases.
b) Taxotransfer rules are based on substitution of mean values derived from analyses of (clustered) soil profile data held in version 3.1 of the ISRIC-WISE database.
c) The classification code (CLAF) for Dune Sands was set to "DS", with profile-ID (PRID) of "WD-DS".
d) A section on "Appropriate use of the data" has been added to the report. Besides describing the scales at which the data may be used (<1:5M; 5 by 5 arc-minutes and coarser), it specifically states why this type of generalised database may not be used to calculate soil organic carbon stocks and similar that involve the analysis of co-varying soil profile data.

Previous version:
Ver. 1.1, November 2006 (superseded by ver. 1.2).

Appendix 9. Installation procedure

The data set is provided in one single zip file, W*ISE5by5min_v1b.zip.* By default, it will be de-compressed (unzipped) to folder *X:\WISE5by5min,* where *X* is the actual location.

The data set includes the range of files described in Appendix 1 to 7 (see WISE5by5min.mdb for the corresponding data) as well as FAO-DSMW's 5 by 5 arc minute raster data (see: …\GRID\smw5by5min). Using ArcGIS® and similar, users may join the raster data to various derived soil properties as presented in the MSAccess® database.

Linkage is through the map unit code or grid cell identifier (*VALUE*) of the raster set and the *SUID* (or FAO's *SNUM*) of the various soil property files (see Appendix 2 and 6).

The various files present derived soil data for five 20 cm depth layers, coded D1 to D5, to 100 cm depth, either for the dominant soil unit only (see App. 6. Method A) or as spatially dominant class (see Appendix 6, method B) of a given mapping unit. Criteria used for making the binned data sets are presented in Appendix 7 (i.e., Legends for the respective properties).

Depending on the proposed applications, users may select the appropriate data set(s) with due consideration for the issues raised in section 3.5 on 'appropriate use of the derived data'.

Methodological and technical details are provided in the documentation:

Batjes NH 2012. ISRIC-WISE global data set of derived soil properties on a 5 by 5 arc-minutes grid (ver. 1.2). Report 2012/01, ISRIC – World Soil Information, Wageningen (with data set, available at http://www.isric.org)

ISRIC – World Soil Information has a mandate to serve the international community as custodian of global soil information and to increase awareness and understanding of soils in major global issues.

More information: www.isric.org

WAGENINGEN UR
For quality of life

ISRIC – World soil Information has a strategic association with Wageningen UR (University & Research centre)