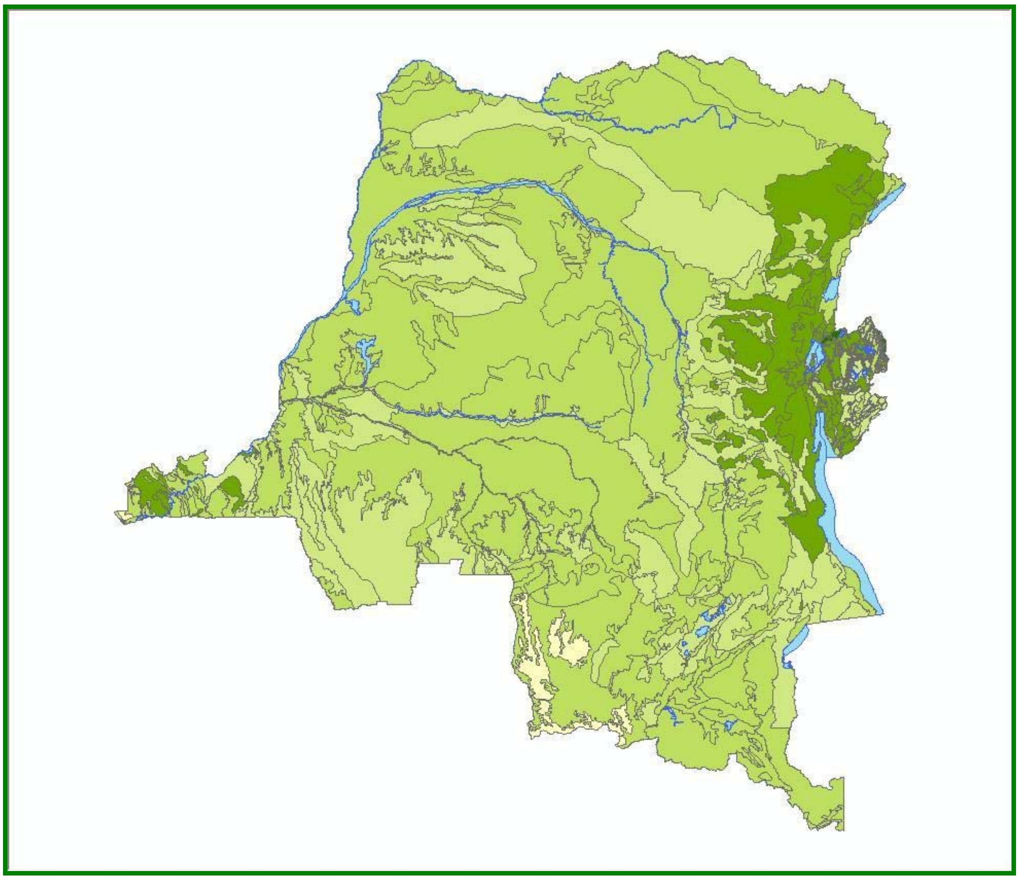
Report 2007/02

SOTER-based soil parameter estimates for Central Africa – DR of Congo, Burundi and Rwanda

(SOTWIScaf, version 1.0)

Niels H Batjes

(May 2007)



ISRIC World Soil Information

Disclaimer:

Correct citation:

Batjes NH 2007. SOTER-based soil parameter estimates for Central Africa — DR of Congo, Burundi and Rwanda (SOTWIScaf, ver. 1.0). Report 2007/02, ISRIC – World Soil Information, Wageningen (**http://www.isric.org/isric/Webdocs/Docs/ISRIC_Report_2007_02.pdf**)

Inquiries:

c/o Director, ISRIC – World Soil Information
PO B08 353
6700 AJ Wageningen
The Netherlands
Telefax: +31-(0)317-471700
E-mail: soil.isric@wur.nl
Web: www.isric.org

Front cover: Soil organic carbon content in first 20 cm of the dominant soil units of Central Africa, derived from SOTERCAF and WISE data.

Contents

List of Tables

List of Figures

SUMMARY

This report describes a harmonized set of soil parameter estimates for Central Africa, comprising Burundi and Rwanda (scale 1:1 million) and the Democratic Republic of the Congo (scale 1:2 million). The data set was derived from the Soil and Terrain Database for Central Africa (FAO *et al.* 2007) and profiles held in the ISRIC-WISE database, using standardized taxonomy-based pedotransfer (*taxotransfer*) procedures (Batjes *et al.* 2007).

The land surface of Central Africa, comprising some 2.4 million km^2, has been characterized using 244 unique SOTER units, corresponding with 504 polygons. Each SOTER unit may consist of up to 6 soil components; each of these has been characterized by a representative profile. The main soil units mapped for the region have been characterized using 167 real profiles, selected by soil experts as being regionally representative for these units; soil analytical data were derived from soil survey reports. These sources seldom hold all the physical and chemical attributes ideally required by SOTER. The primary database also includes 129 virtual profiles for which only the FAO classification is known; inherently, there are no measured data for these profiles. Gaps in the soil profile data have been filled using taxotransfer procedures; these were developed using soil profiles from WISE (*n*= 5672) having similar FAO soil unit names as those mapped for the SOTERCAF region.

Parameter estimates are presented by soil unit for fixed depth intervals of 0.2 m to 1 m depth for: organic carbon, total nitrogen, C/N ratio, pH(H_2O), CEC_{soil}, CEC_{clay}, base saturation, effective CEC, aluminium saturation, $CaCO_3$ content, gypsum content, exchangeable sodium percentage (ESP), electrical conductivity of saturated paste (ECe), bulk density, content of sand, silt and clay, content of coarse fragments (> 2 mm), and available water capacity (-33 kPa to -1.5 MPa). These attributes have been identified as being useful for agro-ecological zoning, land evaluation, crop growth simulation, modelling of soil carbon stocks and change, and analyses of global environmental change.

Soil parameter estimates are presented as summary files that can be linked to the SOTERCAF map using GIS, through the unique SOTER-unit code; applications should consider the full map unit composition and depth range.

Individual parameter estimates should be seen as best estimates based on the current selection of soil profile data; each taxotransfer rule has been flagged to provide an indication of confidence in the derived data. The secondary data presented here are considered appropriate for exploratory studies at the continental scale (≤ 1:1 million). Correlation of soil analytical data should be done more rigorously when more detailed scientific work is considered.

Keywords: secondary data set, soil parameter estimates, Central Africa, Rwanda, Burundi, Democratic Republic of the Congo, environmental modelling, WISE database, SOTER database

1 INTRODUCTION

ISRIC and FAO and UNEP, under the aegis of the International Union of Soil Sciences (IUSS), are updating the information on world soil resources in the World Soils and Terrain Digital Databases (SOTER) project. Once global coverage has been attained, SOTER is to supersede the 1:5 million scale Soil Map of the World (Nachtergaele and Oldeman 2002; Oldeman and van Engelen 1993).

SOTER databases are composed of two main elements: a geographic and an attribute data component. The *geographical database* holds information on the location, extent and topology of each SOTER unit. The *attribute database* describes the characteristics of the spatial unit and includes both area data and point data. A geographical information system (GIS) is used to manage the geographic data, while the attribute data are handled in a relational database management system. Methodological details may be found in the SOTER Procedures Manual (van Engelen and Wen 1995). Issues of data acquisition, quality control and sharing in the context of SOTER projects have been discussed by Batjes (1999).

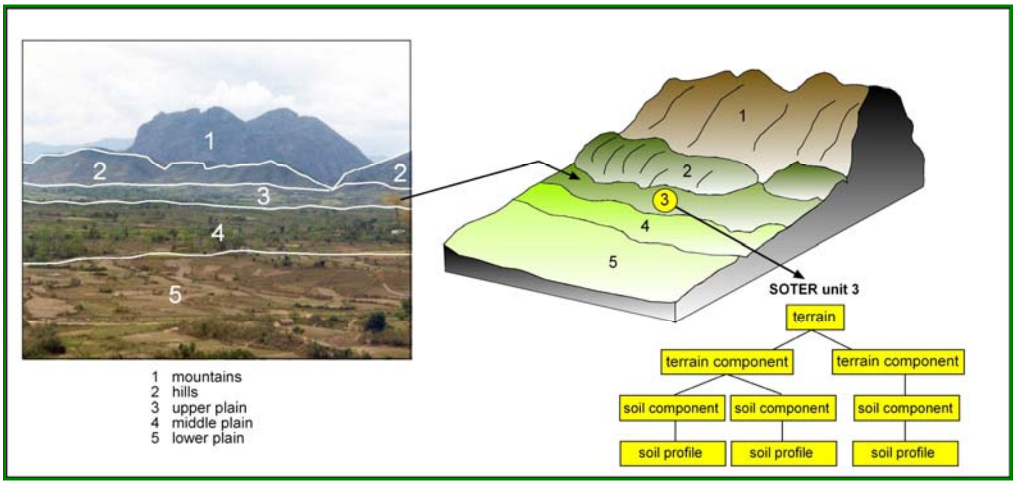


Figure 1. Representation of SOTER units and their conceptual structure

Soil components of individual SOTER units are characterized by a representative soil profile (Figure 1). These are selected from available soil survey reports as the SOTER program does not involve new ground surveys. As a result, there are often gaps in the measured (i.e. *primary*) analytical data, in particular the soil physical data. This precludes the direct use of primary SOTER data in models, so far requiring varying approaches to gap-plugging. ISRIC has therefore developed a uniform methodology for filling gaps in primary SOTER databases, for general purpose applications. This taxotransfer rule-based procedure draws heavily on soil analytical held in the ISRIC-WISE soil profile database, as detailed in Batjes (2003).

This report discusses the application of the taxotransfer procedure to the *primary* SOTER data for Central Africa (SOTERCAF, FAO *et al.* 2007; van Engelen *et al.* 2006). The resulting *secondary* data comprise soil parameter estimates for 19 soil attributes for up to five standardized depth ranges of 20 cm each. Secondary SOTER data having the same format already exist for Latin America and the Caribbean, Central and Eastern Europe, and Southern Africa and other areas with SOTER-like databases; for details see Van Engelen *et al.* (2005).

Chapter 2 describes the materials and methods with special focus on the procedure for preparing the secondary SOTER data. Results are discussed in Chapter 3, while concluding remarks are drawn in Chapter 4. The structure of the various output tables and installation procedure are documented in the Appendices.

2 MATERIALS AND METHODS

2.1 Source of primary SOTER data

Release 1.0 of the SOTER database for Central Africa provided the basis for this study (FAO *et al.* 2007). The study area covers Burundi, the Democratic Republic of the Congo, and Rwanda or some 2.4 million km^2 in total. The available soil geographical and attribute data were harmonized into SOTER format by various organizations, using standardized procedures (van Engelen *et al.* 2006). Although the map has a generalized scale of 1:1 million (Rwanda and Burundi) and 1:2 million (Democratic Republic of the Congo), the detail and quality of the primary information varies widely between and within the three countries. Whereas most soil data for the Congo originated from surveys carried out before 1960 more recent materials were available for Rwanda and Burundi. By implication, the soil profiles represented in the data set were collected between 1940 to 2001.

Profiles in SOTERCAF were characterised according to the Revised Legend of FAO (1988) and World Reference Base for Soil Resources (FAO 2006). The Revised Legend, however, was used to display/map the soil units to ensure consistency with existing SOTER databases.

2.2 Preparation of secondary SOTER data

2.2.1 Checking of primary data

Primary SOTER data are routinely submitted to automated integrity checks (Tempel 1997). Nonetheless, when glancing through the primary SOTERCAF data it soon appeared that were several errors or inconsistencies in the PROFILE and HORIZON tables, thus hampering the application of the taxotransfer procedure. Therefore, all data were re-checked for possible errors using data-integrity checks developed for the WISE database (Batjes *et al.* 1995).

Corrections were made where: the sum of (sand + silt + clay) differed greatly from 100 %; NULL fields were erroneously filled with 0's (i.e. pH=0 instead of NULL); values for exchangeable-Al^{+++} where reported to be greater than those reported for exchangeable-acidity (i.e. exchangeable H^+ + Al^{+++}); C/N ratios were unrealistically low or high. Typographical errors in FAO classification codes (e.g. Glu instead of GLu) or non-existing codes (e.g. 'PTp' was changed to PTd) were also corrected, as these would adversely affect the outcome of the taxotransfer procedure. In some instances, it has been necessary to consult the compilers of the SOTERCAF data as these had privileged access to the original, legacy data.

2.2.2 Procedure for filling gaps in the measured soil data

Being based on available soil survey reports, there are always gaps in the measured soil analytical data as required by SOTER. The standard procedure for filling such gaps, for selected soil attributes, includes three stages (Batjes 2003; Batjes *et al.* 2007):

a) Collate (additional) measured soil data where these exist, in uniform SOTER format;

b) Use expert estimates and common sense to fill selected gaps in a secondary, working data set;

c) Use taxotransfer rule-derived soil parameter estimates for similar FAO soil units (as mapped for the study area), based on a selection of profiles held in a working copy of the WISE profile database.

Being based on a recently published SOTER database, this study only involved stage c). There are 167 real profiles in SOTERCAF; the profile density is lowest for the Democratic Republic of Congo (96). Virtual profiles (129) have been introduced when the FAO classification for a given soil unit was known from soil maps, but there were no measured data to characterize these units. All parameter estimates presented here for virtual profiles thus had to be derived through taxotransfer procedures. Further, the soil drainage class and maximum depth for virtual profiles had to be inferred using expert judgement.

Soil parameter estimates used in this study were derived from statistical analyses of some 5762 profiles taken from WISE — these only included those profiles that have similar FAO classification as mapped for the SOTERCAF area (Table 1). Thereby, the current approach is a refinement of earlier work that considered the full complement of profiles in WISE (see Batjes 2003). The measured values in WISE that underlie the taxotransfer scheme — like those held in SOTERCAF — will reflect both variations inherent to the soil unit and those that can be ascribed to the methods of measurement.

For reasons outlined earlier (Batjes 2002, p. 6-11), a pragmatic approach to the comparability of soil analytical data has been adopted for use with small scale SOTER databases. The analytical data for each combination of soil unit, texture class and depth layer were screened using a robust outlier scheme, by attribute (Batjes 2003). This type of approach is considered appropriate for soil data applications at a small scale — correlation of soil analytical data, however, should be done more accurately when more precise scientific research is considered.

Table 1. FAO soil units mapped in SOTERCAF and number of similar soil profiles in WISE used for taxotransfer rule development

FAO soil units	SOTERCAF [a]	WISE [b]
ACf	1 (1/0)	247
ACh	78 (25/53)	359
ACu	12 (5/7)	61
ALp	1 (1/0)	6
ANm	4 (4/0)	44
ANu	2 (1/1)	119
ARa	2 (2/0)	32
ARg	1 (1/0)	74
ARh	2 (2/0)	213
ARl	2 (1/1)	164
ARo	9 (7/2)	136
CMd	7 (5/2)	216
CMe	1 (1/0)	346
CMg	1 (1/0)	132
CMo	17 (3/14)	154
CMu	10 (10/0)	105
CMv	3 (1/2)	76
FLd	2 (2/0)	84
FLm	1 (1/0)	39
FRg	2 (2/0)	38
FRh	48 (14/34)	206
FRp	1 (1/0)	13
FRr	7 (4/3)	105
FRu	18 (17/1)	47
FRx	9 (8/1)	133
GLd	3 (1/2)	129
GLe	2 (1/1)	284
GLm	3 (3/0)	110
GLu	2 (1/1)	55
HSf	1 (1/0)	67
HSs	2 (2/0)	93
HSt	1 (1/0)	9
KSh	1 (1/0)	26
LPd	1 (1/0)	92
LVv	1 (1/0)	43
LXf	1 (1/0)	167

FAO soil units	SOTERCAF [a]	WISE [b]
LXh	2 (2/0)	249
LXp	1 (1/0)	18
NTh	6 (3/3)	56
NTr	1 (1/0)	45
NTu	2 (2/0)	39
PHc	1 (1/0)	35
PHh	4 (4/0)	122
PHl	1 (1/0)	187
PLe	1 (1/0)	100
PTd	2 (2/0)	46
PZc	1 (1/0)	36
RGd	6 (5/1)	64
SCg	2 (2/0)	35
VRe	6 (6/0)	336
VRk	1 (1/0)	170

[a] First number is for total number of soil profiles in SOTERCAF; the first number in brackets is for measured profiles, the second for virtual profiles (i.e. profiles for which there are no measured data; these have codes like RW*syn*1)

[b] Number of profiles from WISE considered in the taxotransfer scheme (*n* = 5672)

2.2.3 List of soil variables

Special attention has been paid to those key attributes (Table 2) that are commonly required in studies of agro-ecological zoning, food productivity, soil gaseous emissions/sinks and environmental change (see Batjes *et al.* 1997; Bouwman *et al.* 2002; Cramer and Fischer 1997; Easter *et al.* 2007; Fischer *et al.* 2002; Scholes *et al.* 1995).

Table 2. List of soil variables considered in secondary SOTER data sets

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}) [a] [b]
Base saturation (as % of CEC_{soil}) [b]
Effective cation exchange capacity (ECEC) [b c]
Aluminium saturation (as % of ECEC) [b]
$CaCO_3$ content
Gypsum content
Exchangeable sodium percentage (ESP) [b]
Electrical conductivity of saturated paste (ECe)
Bulk density

Coarse fragments (> 2mm, volume %)
Sand (mass %)
Silt (mass %)
Clay (mass %)
Available water capacity (cm^3 cm^{-3} 10^2; -33 kPa to -1.5 MPa) [b d]

[a] CEC_{clay} was calculated from CEC_{soil} by assuming a mean contribution of 240 $cmol_c$ kg^{-1} OC, the common range being from 150 to over 750 $cmol_c$ kg^{-1} (Klamt and Sombroek 1988; Schachtschabel *et al.* 1998).

[b] Calculated from other measured soil properties.

[c] ECEC is defined as exchangeable (Ca^{++}+ Mg^{++}+ K^++ Na^+) + exchangeable (H^++ Al^{+++})(van Reeuwijk 2002).

[d] Limits for soil water potential for Available Water Capacity (AWC) conform to USDA standards (Soil Survey Staff 1983).

Table 2 does not include soil hydraulic properties because measured data for the latter were generally lacking in the systematic soil survey reports that underlie SOTER and WISE.

3 RESULTS AND DISCUSSION

3.1 Map unit composition

Central Africa has been described using 244 unique map units or SOTER units. These comprise 478 terrain components and 601 soil components. There are 504 polygons in total. At the small scale under consideration, most mapping units will be compound. Some 20% of the map units have been mapped as single units, while over 43% consist of more than three different soil units (i.e. soil components). This map unit complexity must be considered when using the data; typically, this will have to be done using specifically designed applications (e.g. Batjes *et al.* 2007; Easter *et al.* 2007).

The full composition of each SOTER unit has been summarized in table *SOTERunitComposition* (Appendix 1). This table lists the name and relative area of the main major FAO soil group for each map unit, as well as the type and relative area of all the component soil units (Figure 2).

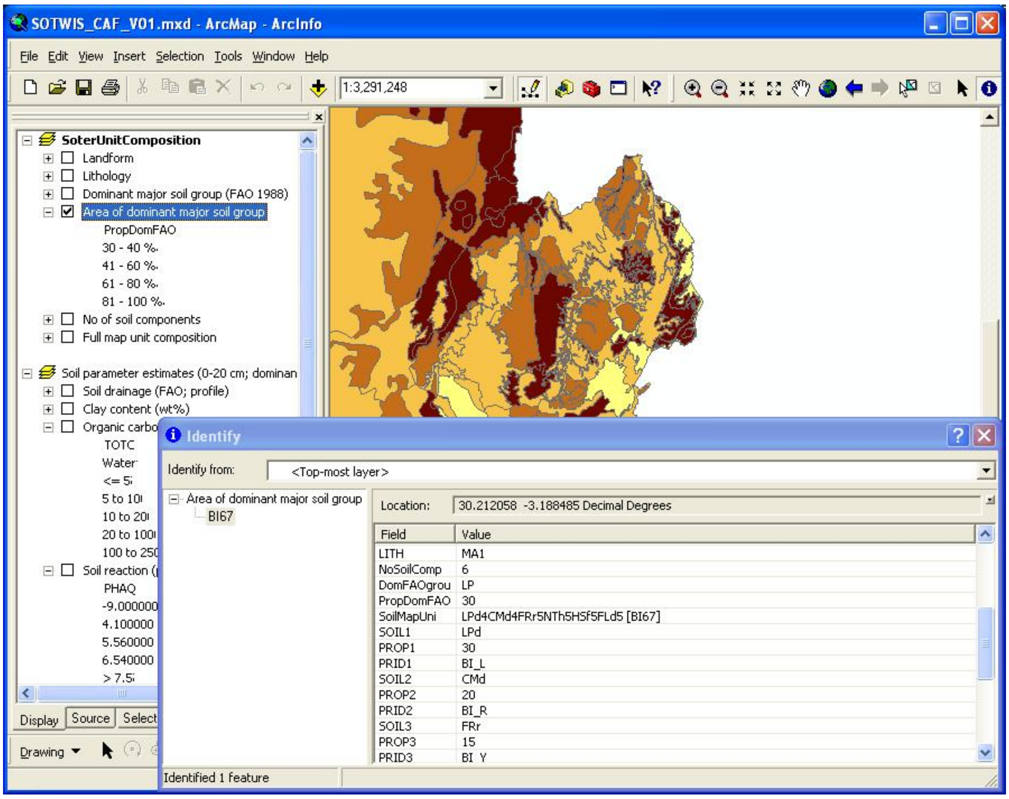


Figure 2. Complexity of SOTER mapping units

3.2 Soil Parameter estimates

Depth-weighed, derived data for 19 soil variables (Table 1) are presented in table *SOTERparameterEstimates* (Appendix 2), for up to 5 fixed depth intervals — D1 (0-20 cm), D2 (20-40 cm), D3 (40-60 cm), D4 (60-80 cm) and D5 (80-100 cm).

In case of missing measured values, the cut-off point for applying any taxotransfer rule is $n_{WISE} < 5$; that is there should be at least 5 cases in the WISE subset for the corresponding combination of soil unit, soil variable, soil layer and soil textural class in order to apply the substitution procedure. Soil textural classes were defined in accordance with current SOTER standards (Appendix 5).

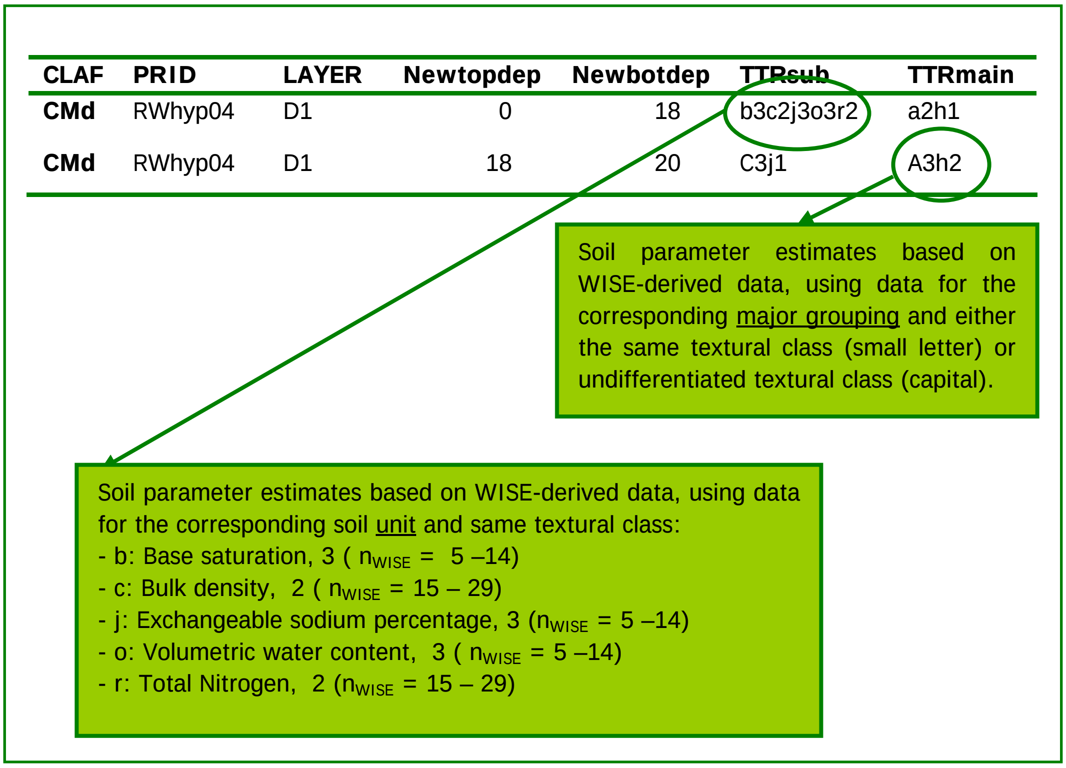


Figure 3. Schematic representation of taxotransfer procedure for filling gaps in SOTER

Where applicable, the type of taxotransfer rules applied has been flagged by profile and depth layer in table *SOTERflagTTRrules* (Appendix 3). The text-strings listed under field TTRsub indicate that the data substitution for a given attribute, in the secondary SOTER set, was based on WISE-derived parameter estimates for similar soil units. Otherwise, if the corresponding population in WISE was considered too small ($n_{WISE} < 5$) for a meaningful substitution, the rules were flagged under TTRmain as described in Batjes (2003). In such cases, the taxotransfer procedure was based on parameter estimates for the corresponding major soil group, as best available default values for the substitution. The overall procedure is depicted in Figure 3 for the upper 0 to 20 cm (layer D1) of a hypothetical profile from Rwanda (RWhyp04).

Each flag listed under *TTRsub* and *TTRmain* consists of a sequence of letters followed by a numeral, for example *A3h2*. The letters indicate soil attributes for which a taxotransfer rule has been applied; coding conventions are explained in Appendix 3. The number code reflects the size of the sample population in WISE, after outlier rejection, on which the statistical analyses that underlie taxotransfer scheme were based (Table 3).

Table 3. Criteria for defining confidence in the derived data

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

[a] n_{WISE} is the sample size after the screening or outlier rejection procedure
[b] The cut-off point in the TTR-approach is $n_{WISE} < 5$

When a *small letter* is used, the substitution was based on median data for the corresponding soil unit, depth layer and textural class (for example, Rhodic Ferralsols (FRr), 0-20cm (D1), Fine and $n_{WISE} > 5$). Otherwise, when a *capital* is used, this indicates that the substitution for the given soil attribute was based on the whole set for the corresponding soil unit and depth layer, irrespective of soil texture (i.e. undifferentiated or *u*). The same coding conventions apply for TTRmain, but substitutions then consider derived soil data for the corresponding major FAO soil group.

3.3 Type and number or taxotransfer rules used

Table 4 lists how often each taxotransfer rule has been applied as a percentage of the total number of "horizons/layers/depth" combinations in the secondary SOTERCAF set; details may be found in table *SOTERflagTTRrules* (Appendix 3).

Table 4. Type and frequency of taxotransfer rules (TTR) and expert rules (XR*i*) applied

SOTNAM	Frequency of occurrence (%)			
	TTRsub	TTRmain	TTRtotal	Expert rules
XR0-Text	-	-	-	36
XR1-Alsa	-	-	-	36
XR2-Bsat	-	-	-	0
XR3-Elco	-	-	-	91
XR4-Gyps	-	-	-	91
XR5-CaCo	-	-	-	91
XR6-CECc	-	-	-	4
XR7-Hist	-	-	-	1
XR8-LAC	-	-	-	0
XR9-ECEC	-	-	-	0
TTR-ALSA	56	9	65	-
TTR-BSAT	61	0	61	-
TTR-BULK	97	2	99	-
TTR-CECC	42	1	43	-
TTR-CECS	36	0	36	-
TTR-GRAV	37	2	39	-
TTR-CLAY	35	0	35	-
TTR-ECEC	66	1	67	-
TTR-ELCO	5	27	32	-
TTR-ESP	60	1	61	-
TTR-GYPS	4	17	21	-
TTR-PHAQ	35	0	35	-
TTR-SAND	35	0	35	-
TTR-SILT	35	0	35	-
TTR-TAWC	93	5	98	-
TTR-TCEQ	6	16	22	-
TTR-TOTC	39	0	39	-
TTR-TOTN	69	0	69	-
TTR-CN	66	0	66	-

Note: For definitions of abbreviations see text and Table 4; see also see the footnote in Appendix 3. A '-' stands for not applicable.

For example, the aluminium saturation percentage (ALSA) has been estimated in 65% of the cases, using data for similar soil units (56% of cases, see under TTRsub) resp. similar major soil groups (9%). Further, expert rules for ALSA (XR1-Alsa) have been applied in 36% of the cases. Conversely, for bulk density (BULK) taxotransfer rules have been applied in 97% of the cases, which reflects that there are almost no measured bulk density in SOTERCAF. Conversely, TTRs were applied in some 35% of the cases for sand, silt and clay content (SDTO, STPC, CLPC); these cases mainly correspond with the 129 virtual profiles.

3.4 Main assumptions and limitations

The overall assumption has been that the confidence in a TTR-based parameter estimate should increase with the size of the corresponding sample populations present in WISE, after outlier rejection. In addition, the confidence in soil parameter estimates listed under TTRsub should be higher than for those listed under TTRmain.

A high confidence rating for a given parameter estimate, however, does not necessarily imply that this estimate will be representative for the soil unit under consideration. Profile selection for SOTER and WISE, as for many other small-scale soil databases, is not probabilistic but based on available data and expert knowledge. Several of the soil attributes under consideration in Table 2 are not diagnostic in the Revised Legend (FAO 1988). In addition, some soil properties are readily modified by changes in land use or management, for example soil pH, aluminium saturation and organic matter content. Information on land use/management history, however, is not available in SOTERCAF and, as such, this aspect could not be considered in the taxotransfer procedure.

It should be noted that the adoption of different criteria for clustering the data would inherently lead to varying parameter estimates. For example, selecting a different soil classification system (e.g., FAO 1988 *versus* WRB 2006), limits for depth layers (e.g., 0-20 cm intervals up to 100 cm *versus* 0-30 cm and 30-100 cm), criteria for defining soil textural classes (e.g., 5 classes in SOTER *versus* 3 classes for the FAO Soil Map of the World), choice of critical limits for applying taxotransfer rules (i.e. reject when $n_{WISE} < 5$ or $n_{WISE} < 15$), as well as the type of outlier-rejection and statistical procedures used, or the number of WISE profiles under consideration. Most importantly, however, the outcome will primarily be determined by the number and quality of profiles stored in the underpinning, primary SOTER database; in particular the degree to which the various fields have been filled and the comparability of the analytical methods used. There are no data, however, to indicate that the available profile descriptions are statistically representative of the regional distribution of the soil units in the region, nor is the land use history known for most profiles; these problems are commonly encountered and recognized in small scale, soil databases (Batjes and Dijkshoorn 1999; Lal *et al.* 1995; Sombroek *et al.* 1993).

3.5 Linkage to GIS

Most SOTER units mapped for Central Africa are compound, comprising up to six soil components. The full map unit composition has been summarized in one single table (*SOTERunitComposition*, see Appendix 1). Results of the taxotransfer procedure in table *SOTERparameterEstimates* (Appendix 2) can be linked to the digital map using GIS through the unique *NEWSUID* (see *SOTERsummaryFile*, Appendix 4). The overall procedure is visualized in Figure 4 for a hypothetical database.

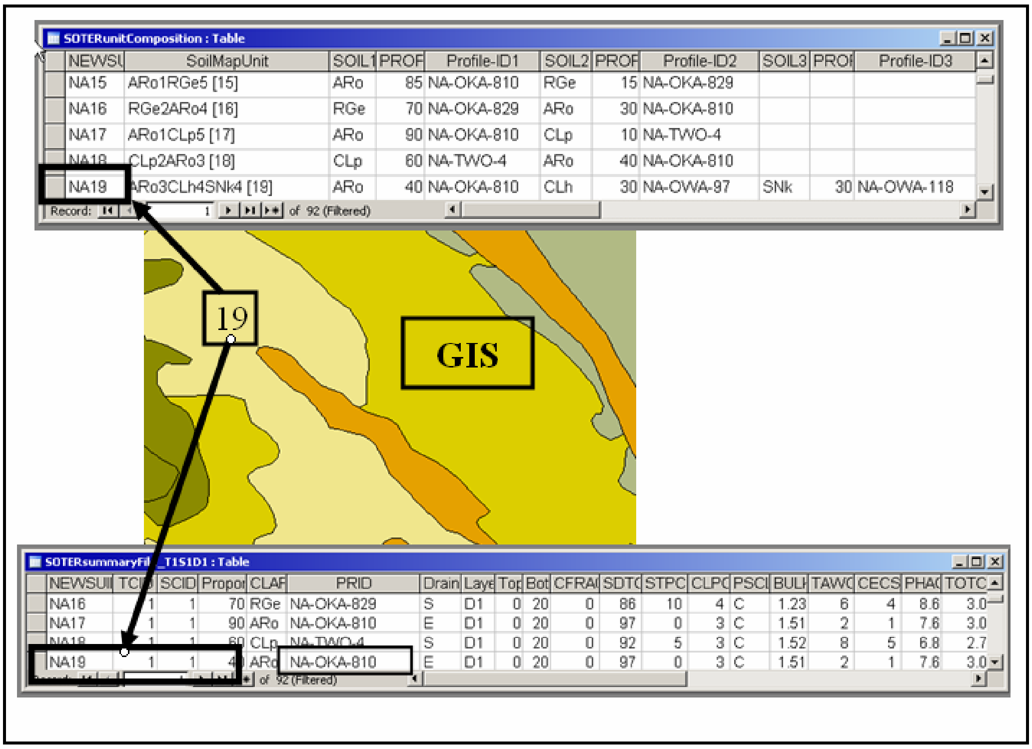


Figure 4. Schematized procedure for linking soil parameter estimates for the upper layer (D1) of the main soil unit (TCID=1; SCID=1) of a SOTER map unit with the geographical data

The dataset includes, as an example for visualization using GIS, a table listing derived soil data for the 0-20 cm layer of the dominant soil unit of each SOTER unit. In this simple case, the selection criteria will be: Terrain

Component (TCID) =1), Soil Component (SCID) = 1, and Layer-ID (D_i) = D1; see table *SOTERsummaryFile_T1S1D1* in Appendix 4. Typically, however, specific data selections that consider the full soil composition of individual SOTER units will have to be made before 'aggregated' model output can be coupled back to the GIS. Details of such an approach, as developed for the GEFSOC Soil Carbon Modelling System, may be found in Easter et al. (2007).

4 CONCLUSIONS

- The detail and quality of primary information available within the various countries of Central Africa resulted in a variable resolution of the secondary products presented.
- Linkage between the soil profile data and the spatial component of the SOTERCAF map required generalisation of measured soil (profile) data by soil unit and depth zone. This involved the transformation of variables that show a marked spatial and temporal variation and that have been determined in a range of laboratories, according to various analytical methods.
- A pragmatic approach to the comparability of soil analytical data has been adopted when developing the taxotransfer scheme. This was considered appropriate at the present scale, but such a comparison must be done more rigorously when more detailed scientific work is considered.
- End users should familiarize themselves with the assumptions and procedures used to derive soil parameter estimates prior to using them in models. It is important that such applications consider the full map unit composition, not merely the dominant soil unit of a SOTER unit.
- The soil parameter estimates for the SOTERCAF region are considered appropriate for applications at a scale of 1:1 million and smaller.

REFERENCES

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 2002. *Soil parameter estimates for the soil types of the world for use in global and regional modelling (Version 2.1)*. ISRIC Report 2002/02c (available through: http://www.isric.org, verified 02 March 2007), International Food Policy Research Institute (IFPRI) and International Soil Reference and Information Centre (ISRIC), Wageningen

Batjes NH 2003. *A taxotransfer rule-based approach for filling gaps in measured soil data in primary SOTER databases (GEFSOC Project)*. Report 2003/03 (available through: http://www.isric.org), ISRIC - World Soil Information, Wageningen

Batjes NH and Dijkshoorn JA 1999. Carbon and nitrogen stocks in the soils of the Amazon Region. *Geoderma* 89, 273-286

Batjes NH, Bridges EM and Nachtergaele FO 1995. World Inventory of Soil Emission Potentials: development of a global soil data base of process-controlling factors. In: Peng S, KT Ingram, HU Neue and LH Ziska (editors), *Climate Change and Rice*. Springer-Verlag, Heidelberg, pp 102-115

Batjes NH, Fischer G, Nachtergaele FO, Stolbovoy VS and van Velthuizen HT 1997. *Soil data derived from WISE for use in global and regional AEZ studies (ver. 1.0)*. Interim Report IR-97-025, FAO/ IIASA/ ISRIC, Laxenburg

Batjes NH, R. Al-Adamat, Bhattacharyya T, M. Bernoux, C.E.P. Cerri, P. Gicheru, P. Kamoni, 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* 112, 26-34

Bouwman AF, Boumans LJM and Batjes NH 2002. Modeling global annual N_2O and NO emissions from fertilized fields. *Global Biogeochemical Cycles* 16, 1080, doi:10.1029/2001GB001812

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

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

Easter M, Paustian K, Killian K, Williams S, Feng T, Al-Adamat R, Batjes NH, Bernoux M, Bhattacharyya T, Cerri CC, Cerri CEP, Coleman K, Falloon P, Feller C, Gicheru P, Kamoni P, Milne E, Pal DK, Powlson D, Rawajfih Z, Sessay M and Wokabi S 2007. The GEFSOC soil carbon modeling system: a tool for conducting regional-scale soil carbon inventories and assessing the impacts of land use change on soil carbon. *Agriculture, Ecosystems & Environment* 112, 13-25

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

FAO 2006. *World Reference Base for soil resources - A framework for international classification, correlation and communication*. World Soil Resources Reports 103, International Union of Soil Sciences, ISRIC - World Soil Information and Food and Agriculture Organization of the United Nations, Rome

FAO, ISRIC and UG 2007. *Soil and terrain database for central Africa (Burundi and Rwanda 1:1 million scale; Democratic Republic of the Congo 1:2 million scale)*. Land and Water Digital Media Series 33, Food and Agricultural Organization of the United Nations, ISRIC - World Soil Information and Universiteit Gent, Rome

Fischer G, van Velthuizen HT, Shah M and Nachtergaele FO 2002. *Global Agro-ecological Assessment for Agriculture in the 21st Century: Methodology and Results*. RR-02-02, International Institute for Applied Systems Analysis (IIASA) and Food and Agriculture Organization of the United Nations (FAO), Laxenburg

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

Lal R, Kimble J, Levine E and Whitman C 1995. Towards improving the global database on soil carbon. In: Lal R, J Kimble, J Levine and BA Stewart (editors), *Soils and Global Change*. Lewis Publishers, Boca Raton, pp 343-346

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

Marsman BA and de Gruijter JJ 1986. *Quality of soil maps: A comparison of soil survey methods in sandy areas*. Soil Survey Papers 15, Netherlands Soil Survey Institute, Wageningen

Nachtergaele FO and Oldeman LR 2002. World soil and terrain database (SOTER): past, present and future. *Transactions 17th World Congress of Soil Science*. International Union of Soil Sciences (IUSS), Bangkok, pp 653/1-653/10

Oldeman LR and van Engelen VWP 1993. A World Soils and Terrain Digital Database (SOTER) - An improved assessment of land resources. *Geoderma* 60, 309-335

Schachtschabel P, Blume H-P, Brummer G, Hartge KH and Schwertmann U 1998. *Lehrbuch der Bodenkunde (Scheffer/Schachtschabel, 14th Ed.)*. Ferdinand Enke Verlag, Stuttgart, 494 p

Scholes RJ, Skole D and Ingram JS 1995. *A global database of soil properties: proposal for implementation*. IGBP-DIS Working Paper 10, International Geosphere Biosphere Program, Data & Information System, Paris

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

Tempel P 1997. *Global and national Soil and Terrain Digital Databases (SOTER) - Scale 1:2,500,000 (Attribute database use manual adapted for SOVEUR Project)*. Report 97/09, ISRIC, Wageningen

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

van Engelen VWP, Batjes NH, Dijkshoorn K and Huting J 2005. *Harmonized Global Soil Resources Database (Final Report)*. Report 2005/06, FAO and ISRIC - World Soil Information, Wageningen

van Engelen VWP, Verdoodt A, Dijkshoorn K and van Ranst E 2006. *SOTER database for Central Africa -- DR Congo, Burundi and Rwanda (SOTERCAF; ver. 1.0)*. ISRIC REport 2006/07 (available through http://www.isric.org), Laboratory of Soil Science (University of Ghent), FAO and ISRIC - World Soil Information, Wageningen [

van Reeuwijk LP 2002. *Procedures for soil analysis (6th ed.)*. Technical Paper 9, ISRIC, Wageningen

APPENDICES

Appendix 1. Structure of table SOTERunitComposition

This table, in MS Access® format, gives the full composition of each SOTER unit in terms of its: landform, lithology (parent material), dominant major FAO soil group and its relative extent, then component in soil units with their relative extent, and the identifier for the corresponding representative profile. The relevant information was distilled from three primary SOTER tables, *viz. Terrain*, *SoilComponent* and *Profile*. The content of this table can be linked to the geographical data in a GIS through the unique SOTER unit code or NEWSUID, a combination of the fields for ISO and SUID.

Structure of table SOTERunitComposition[a]

Name	Type	Description
ISOC	Text	ISO-3166 country code (1994) or WD for World
SUID	Integer	The identification code of a SOTER unit on the map and in the database (corresponds with SNUM on the SOTER)
NEWSUID	Text	Globally unique code, comprising fields ISOC plus SUID (e.g. RU1234)
NoOfSoilComp	Text	Number of soil components in given SOTER unit
DomFAOgroup	Text	Dominant FAO major soil group in SOTER (Note: This need not always be SOIL1)
PropDomFAOGroup		Relative area of dominant major soil group in SOTER unit
SoilMapunit	Text	Aggregated code for map unit summarizing the overall composition [a]
SOIL1	Text	Characterization of the first (main) soil unit according to the Revised FAO-Unesco Legend
PROP1	Integer	Proportion, as a percentage, that the main soil unit occupies within the SOTER unit
PRID1	Text	Unique code for the corresponding measured resp. virtual soil profile (e.g. RWsyn1)
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

Name	Type	Description
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
SOIL9	Text	As above but for the next soil component
PROP9	Integer	As above
PRID9	Text	As above
SOIL10	Text	As above but for the next soil component
PROP10	Integer	As above
PRID10	Text	As above

[a] Note: Generally, not all 10 available fields for $SOIL_i$ will be filled in SOTER.
[b] These codes have the following format: VRe2GLe4. The relative extent of each soil unit (e.g., VRe) has been expressed in 5 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 percent; 4 – from 20 to 40 per cent, and 5 – less than 20 percent.

Appendix 2. Structure of table SOTERparameterEstimates

Table SOTER*parameterEstimates* lists parameter estimates for all soil units (represented by their PRID) that have been mapped for the study region. This information can be linked to the soil geographical data – in a GIS – through the unique profile code (PRID).

Structure of table SOTER*parameterEstimates*

Name	Type	Description
CLAF	Text	FAO-Unesco (1988) Revised Legend code
PRID	Text	profile ID (as listed in SOTERmapunitComposition)
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 (> 2mm)
SDTO	Integer	sand (mass %)
STPC	Integer	silt (mass %)
CLPC	Integer	clay (mass %)
PSCL	Text	SOTER texture class (see Appendix 5)
BULK	Single	bulk density (kg dm^{-3})
TAWC	Integer	available water capacity (cm^3 cm^{-3} 10^2, -33 kPa to -1.5 MPa conform to USDA standards)
CECs	Single	cation exchange capacity ($cmol_c$ kg^{-1}) for fine earth fraction
BSAT	Integer	base saturation as percentage of CECsoil
CECc	Single	CECclay, 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 N kg^{-1})
CNrt	Single	C/N ratio
ECEC	Single	effective CEC ($cmol_c$ kg^{-1})

Note: These are depth-weighted values. In view of the TTR-rules applied and depth weighting, parameters listed for TOTC and TOTN should not be used to compute median C/N ratios.

Contents of the above table should be consulted in conjunction with table *SOTERflagTTRrules* which lists the taxotransfer rules that have been applied for each profile, depth layer and soil attribute. Details are given in Appendix 3.

Appendix 3. Structure of table SOTERflagRules

Table SOTER*flagTTRrules* documents the type of taxotransfer that have been used to create table *SOTERparameterEstimates* (Appendix 2). Coding conventions are detailed in Tables 3 and 5.

Structure of table *SOTERflagTTRrules*

Name	Type	Description
CLAF	Text	FAO Legend code
PRID	Text	Unique identifier for representative profile
Layer	Text	code for depth layer (from D1 to D5; e.g. D1 is from 0 to 20 cm)
Newtopdep	Integer	Depth of top of layer (cm)
Newbotdep	Integer	Depth of bottom of layer (cm)
TTRsub	Text	Code showing the type of taxotransfer rule used (based on data for soil *units*; see text)
TTRmain	Text	Code showing the type of taxotransfer rule used (based on data for *major units*; see text)
TTRfinal	Text	Additional flags (based on 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. Finally, the CEC of the clay fraction (CEC_{clay}) has been re-calculated from the depth-weighted measured and TTR-derived data for CEC_{soil} and content of organic carbon, assuming a mean contribution of 240 $cmol_c$ kg^{-1} OC. When applicable, expert rules have been flagged in the field TTRfinal.

Table 5. Conventions used for coding soil attributes in the taxotransfer scheme

TTRflag	SOTnam	WISnam	SoilVariable	Comments
A	ALSA	ALSA	ALSAT	Exch. Aluminum percentage (% of ECEC)
B	BSAT	BSAT	BSAT	base saturation (% of CECs)
C	BULK	BULK	BULKDENS	Bulk density
D	CECC	CECC	CECCLAY	cation exchange capacity of clay fraction
E	CECS	CECS	CECSOIL	cation exchange capacity
F	CFRAG	GRAV	GRAVEL	coarse fragments
G	CLPC	CLAY	CLAY	clay % (see also y for texture (g, m & n)
H	ECEC	ECEC	ECEC	effective CEC
I	ELCO	ECE	ECE	electrical conductivity
J	ESP	ESP	ESP	exchangeable Na percentage (% of CECs)
K	GYPS	GYPS	GYPSUM	gypsum content (g C kg^{-1})
L	PHAQ	PHH2	PHH2O	pH in water

TTRflag	SOTnam	WISnam	SoilVariable	Comments
M	SDTO	SAND	SAND	sand %
N	STPC	SILT	SILT	silt %
O	TAWC	TAWC	TAWC	Vol. water content (-33 kPa to -1.5 MPa)
P	TCEQ	CACO	CACO3	carbonate content (g kg^{-1})
Q	TOTC	ORGC	ORGC	organic carbon content (g C kg^{-1})
R	TOTN	TOTN	TOTN	total nitrogen content (g N kg^{-1})
Y	---	---	---	PSCL estimated from TTR-derived sand, silt and clay content (where applicable)
Z	CNrt	CN	CN	C/N ratio

Abbreviations: TTRflag = code for TTR-rule; SOTnam = codes used in SOTER; WISnam= codes used in WISE; SoilVariable= soil variables as described in Table 2

Appendix 4. Structure of table SOTERsummaryFile

Table *SOTERsummaryFile* has been created to facilitate access to the derived data. For each SOTER unit (NEWSUID) on the map, it lists the un-binned soil parameter estimates by component soil unit and depth layer.

Structure of table SOTER*summaryFile*

Name	Type	Description
ISOC	Text	ISO-3166 country code (1994)
SUID	Integer	The identification code of a SOTER on the map and in the database
NEWSUID	Text	Globally unique map unit code, comprising fields ISOC plus SUID
TCID	Integer	Number of terrain component in given map unit
SCID	Integer	Number of soil unit within the given SOTER unit
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 SOTER unit
CLAF	Text	FAO-Unesco Revised Legend code
PRID	Text	Profile ID (see table *SOTERunitComposition*)
Drain	Text	FAO soil drainage class
TopDep	Integer	Upper depth of layer (cm)
BotDep	Integer	Lower dept of layer (cm)
CFRAG	Integer	Coarse fragments (% > 2mm)
SDTO	Integer	Sand (mass %)
STPC	Integer	Silt (mass %)
CLPC	Integer	Clay (mass %)
PSCL	Text	FAO texture class (see Appendix 5)
BULK	Single	Bulk density (kg dm^{-3})
TAWC	Integer	Available water capacity (cm^3 cm^{-3} 10^2, -33 kPa to -1.5 MPa)
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 ($cmol_c$ 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 kg^{-1})
TOTN	Single	Total nitrogen (g kg^{-1})
CNrt	Single	C/N ratio
ECEC	Single	Effective CEC ($cmol_c$ kg^{-1})

Notes:

1) The soil components that occur within a SOTER unit are numbered sequentially, starting with the spatially dominant one. The sum of the relative proportions of

all component soil units is always 100 per cent. This total will also include a number of unnamed 'impurities', commonly in excess of 15 to 30 percent of the map unit (Landon 1991 p. 16-17; Marsman and de Gruijter 1986).

2) Each map unit in the geographic database has a unique identifier (NEWSUID) consisting of the country ISO code (ISOC) and the SOTER unit-ID (SUID); this primary key provides a link to the attribute data for the constituent terrain, terrain component(s) (TCID) and soil components (SCID) (see Figure 1).
3) Tables with the same structure have been prepared for the DOMINANT soil unit only, by depth layer (i.e., for layer D1 to D5, see e.g. file *SOTERsummaryFile_T1S1D1*) to facilitate visualization using GIS, as example only. Comprehensive studies should consider the full map unit composition and depth range!
4) A limited of records may contain a negative value; this indicates that it has not yet been possible to plug the corresponding gaps using the taxotransfer scheme due to a lack of measured data in SOTERCAF and WISE. Whenever possible, the virtual profiles should be replaced with real, measured profiles in the primary SOTERCAF set after which new secondary data may be generated.
5) Parameter estimates are depth-weighted values, per 20 cm layer.

Appendix 5. Soil textural classes

Soil textural classes (PSCL) are in accordance with revised SOTER criteria (Figure 5). The following abbreviations are used: C–coarse, M–medium, Z–medium fine, F–fine and V–very fine. Further, the symbol *u* is used for undifferentiated (i.e., C + M + F + Z + V). In addition, all Histosols data have been flagged as consisting of organic materials (O) eventhough this may not always be the case, for the whole profile, in a strict taxonomic sense (see FAO 1988 , p. 39)

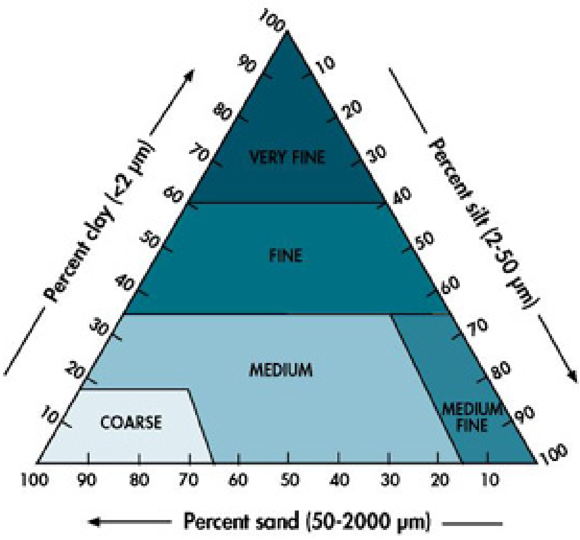


Figure 5. SOTER soil texture classes (Source: CEC 1985)

Appendix 6. Installation procedure

The derived data set and GIS-files are presented in one single zip file: *SOTWIS_CAF_v1.zip.* By default, this compressed file will be unzipped to folder *X:\SOTWIS_CAF*, where *X* is the actual location.

This new folder will contain 3 subfolders:

- *GISfiles* comprising the shape files with their metadata, layer files and the project (*SOTWIS_CAF_V01.mxd*) file
- *SOTWIS* with the derived soil data in MS Access format (*SOTWIS_CAF_1.mdb*)
- *REPORTS* with the documentation in PDF format.

The GIS project file (mxd) includes several derived data sets for the top layer (0-20 cm) of the dominant soil unit of each SOTER unit, as examples. Some of the themes are based on joins; please note that some fields in the attribute table(s) may not be visible. Actual data applications should always consider the full map unit composition and depth range; see text for details.

The following software packages have been used to create the dataset: MS Access® and ArcGIS9.2®; the shapefiles may also be accessed using ArcView3.3®.