JRC Scientific and Technical Reports

# Global Soil Organic Carbon Estimates and the Harmonized World Soil Database

**Roland Hiederer and Martin Köchy**

EUR 25225 EN - 2011

JRC
EUROPEAN COMMISSION

Institute for Environment and Sustainability

The mission of the JRC-IES is to provide scientific-technical support to the European Union's policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability

**Contact information**

R. Hiederer
European Commission Joint Research Centre
Institute for Environment and Sustainability
Via Enrico Fermi, 2749 - 21027 - Ispra (VA) – Italy
E-mail: roland.hiederer@jrc.ec.europa.eu

http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

**Legal Notice**

JRC 68528

EUR 25225 EN

ISBN 978-92-79-23108-7

ISSN 1831-9424

doi:10.2788/13267

Luxembourg: Publications Office of the European Union, 2011

*Printed in Italy*

This document may be cited as follows:

Hiederer, R. and M. Köchy[1] (2011) Global Soil Organic Carbon Estimates and the Harmonized World Soil Database. EUR 25225 EN. Publications Office of the European Union.79pp.

European Commission Joint Research Centre
Institute for Environment and Sustainability
TP 261
21027 Ispra (VA)
Italy

[1] Johann Heinrich von Thünen-Institut
Bundesforschungsinstitut für Ländliche Räume, Wald und Fischerei- Institut für Agrarrelevante Klimaforschung
Bundesallee 50
38116 Braunschweig
Germany

Front Page:

The graph shows soil organic carbon density ($t\ ha^{-1}$) for the amended HWSD data at 3 *arc second* grid spacing for the combined topsoil and subsoil layers. The SOC density is processed as height using the 3DEM terrain visualization software by Richard Horne (r.horne@verizon.net), which at the time of writing was no longer developed.

# Table of Contents

Page

# List of Figures

Page

# List of Tables

## List of Acronyms

| Acronym | Label |
|---|---|
| BIL | Band interleaved-by-line |
| DSMW | Digital Soil Map of the World (FAO) |
| EEA | European Environment Agency |
| ESDB | European Soil Database (JRC) |
| ETRS89 | European Terrestrial Reference System 1989 |
| FAO | Food and Agriculture Organization of the United Nations |
| GRS80 | Geodetic Reference System 1980 |
| HWSD | Harmonized World Soil Database |
| HWSDa | Harmonized World Soil Database with amendments |
| IGBP | International Geosphere-Biosphere Programme |
| IIASA | International Institute for Applied Systems Analysis |
| ISRIC | International Soil Reference and Information Centre |
| ISSCAS | Institute of Soil Science, Chinese Academy of Sciences |
| ITRS | International Terrestrial Reference System |
| JRC | European Commission Joint Research Centre |
| MU | Mapping Unit |
| NSRC | Natural Resources Conservation Service |
| ORNL DAAC | Oak Ridge National Laboratory Distributed Active Archive Center |
| RDBMS | Relational database management system |
| SGDBE | Soil Geographic Database of Euroasia (JRC) |
| SMU | Soil Mapping Unit (SGDBE) |
| SOC | Soil Organic Carbon |
| SOM | Soil Organic Matter |
| SOTER | Global and national soils and terrain database |
| SOTWIS | Harmonized continental SOTER-derived database |
| STU | Soil Typological Unit (SGDBE) |
| UNEP | United Nations Environment Programme |
| UNESCO | United Nations Educational, Scientific and Cultural Organization |
| USDA | United States Department of Agriculture |
| WISE | World Inventory of Soil Emission Potentials |
| WRB | World Reference Base for Soil Resources |
| WRG84 | World Geodetic System 1984 |

**Naming conventions:**

Numbers are given with a comma (,) separator for thousands and a point (.) is used as a decimal separator. When a number represents a code rather than a continuous numeric value the separator for thousands does not apply irrespective of the field format. Hence, there can by 16,107 records in a table with a maximum identifier code of 16107.

Table and field names are spelled in CAPITALS. A field name can be either linked to the table, in which case the table name precedes the filed name using a point (.) separator, or given without a table name. Field names, also those including the table name, are enclosed by square brackets ([]).

Codes and ordinal data, such as soil types, are formatted in *italics*.

# 1 INTRODUCTION

Global estimates of soil organic carbon stocks have been produced in the past to support the calculation of potential emissions of $CO_2$ from the soil under scenarios of change land use/cover and climatic conditions (IPCC, 2006). Very few global estimates are presented as spatial data.

For global spatial layers on soil parameters, the most recent and complete dataset is available as the *Harmonized World Soil Database* (HWSD)[1] (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009). The database was developed by the Land Use Change and Agriculture Program of the *International Institute for Applied Systems Analysis* (IIASA) and the *Food and Agriculture Organization of the United Nations* (FAO) in collaboration with the *International Soil Reference and Information Centre* (ISRIC) -World Soil Information, the European Commission Joint Research Centre (JRC) and the *Institute of Soil Science, Chinese Academy of Sciences* (ISSCAS). The data used in this evaluation are from V.1.1 as published on 29.03.2009.

The HWSD uses a raster format to present the spatial extent of the soil mapping units. In this respect the database deviates from previous pan-national soil databases, such as the *European Soil Database* (ESDB) of the *European Commission Joint Research Centre* (JRC) or the *Digital Soil Map of the World* (DSMW) of *Food and Agriculture Organization of the United Nations* (FAO), which store the spatial units in vector format. Using a raster format simplifies integrating spatial entities coming from diverse sources and facilitates processing numeric attributes in the spatial domain of environmental models. However, since the source data defines the boundaries of spatial units which are significantly larger than the grid size the raster format is more demanding in terms of file size and traditionally less closely associated with database management than the vector format of the source data.

In common with the ESDB and the DSMW is the storage of the soil properties in form of a table of typological units. One or more typological units are linked to the spatial units to define the soil characteristics. While there are strong similarities in the data model of the HWSD with the ESDB the process of harmonizing the parameters defining the typological units introduced significant changes to the source data. As a consequence estimates of soil carbon stocks calculated from the HWSD may differ from those calculated from the source data. This study investigates the HWSD with respect to the parameters used to calculate soil organic carbon density and compares the results with estimates derived from other global data sets.

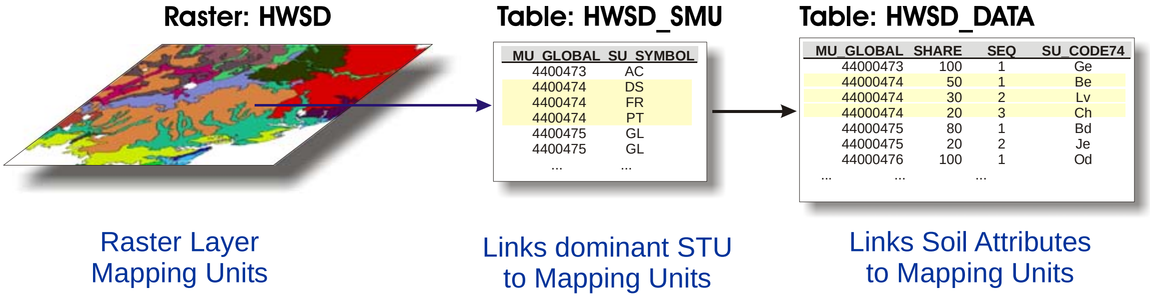
This study further investigates the organic carbon density for the topsoil (0 – 30cm) and the subsoil layer (30 – 100cm) from the amended HWSD with estimates derived from other global data sets for these depth layers.

---

[1] http://www.iiasa.ac.at/Research/LUC/luc07/External-World-soil-database/HTML/index.html?sb=1

# 2 HWSD DATA ORGANIZATION

The HWSD database comprises the mapping units (MUs) as a spatial raster layer (HWSD.RASTER) and the data characterizing to the MUs as attributes in form of tables (HWSD_DATA). An additional table (HWSD_SMU) specified the source of the data and the MU dominant soil type. The organization of the HWSD in schematic form is shown in Figure 1.



| MU_GLOBAL | SU_SYMBOL |
|---|---|
| 4400473 | AC |
| 4400474 | DS |
| 4400474 | FR |
| 4400474 | PT |
| 4400475 | GL |
| 4400475 | GL |
| ... | ... |

| MU_GLOBAL | SHARE | SEQ | SU_CODE74 |
|---|---|---|---|
| 44000473 | 100 | 1 | Ge |
| 44000474 | 50 | 1 | Be |
| 44000474 | 30 | 2 | Lv |
| 44000474 | 20 | 3 | Ch |
| 44000475 | 80 | 1 | Bd |
| 44000475 | 20 | 2 | Je |
| 44000476 | 100 | 1 | Od |
| ... | ... | ... | |

**Figure 1: Schematic Organization of Spatial and Typological Units in Harmonized World Soil Database**

Meta-data on the meaning of attributes given in categorical values are provided by a number of dictionary tables.

## 2.1 Spatial Layer

The spatial layer HWSD_RASTER is provided in band interleaved-by-line (BIL) format. With only a single band in the file the data can be processed as a binary layer. The dimensions of the data are given in the HWSD.HDR file (rows: 21,600, columns: 43,200, 16 bit). The geographic properties of the data are stored in the HWSD.BLW file (global extent with 0.008333 degree grid spacing). The un-projected raster layer has thus a resolution of 30 arc seconds, which corresponds to a grid size of approx. 1x1 km at the Equator.

In using a raster format for the MUs the HWSD differs from other spatial soil databases, such as the *Soil Geographic Database of Eurasia* (SGDBE) of the *European Soil Database* (ESDB) or the *Digital Soil Map of the World* (DSMW) of the FAO, where the spatial units are stored in vector format. With respect to organizing the attribute data in tabular format and linking the soil attributes to the spatial layer there is no major difference between using a raster or vector format for the spatial units. However, restrictions to processing the information and the range of available options for the analysis can apply depending on the software used.

## 2.2 Attribute Tables

The soil property information is arranged in two main tables, one related to the mapping units (HSWU_SMU) and one to the information of the properties of the typological units (HWSD_DATA). Codes and class values are stored in dictionary tables with explanatory comments.

### 2.2.1 Dominant Mapping Unit: HWSD_SMU

The table HWSU_SMU contains information on the source of the data in the [COVERAGE] field and the dominant soil group of the MU in the [SU_SYMBOL] and [SU_CODE] fields. The information on the soil group is thus duplicated and either field would have served the purpose. The primary key is given by [MU_GLOBAL] field.

### 2.2.2 Typological Units: HWSD_DATA

The table HWSD_DATA contains the characteristics of the soil units, which are found within an MU. The primary key of the table is a combination of the MU_GLOBAL and the [SEQ] fields. The [SEQ] field contains the order of the share of a soil unit within an MU. In a deviation from the data model used for the SGDBE the HWSD does not contain a separate table for the soil typological units and the link to the MUs. The corresponding data is recorded in the HWSD_DATA table for each MU without a further link table, such as the STUORG table of the SGDBE.

The fields [HWSD_SMU.SU_SYMBOL] and [HWSD_SMU.SU_CODE] contain the soil type of the dominant mapping unit to the WRB classification scheme. For sub-dominant soil units the soil type is also recorded following according to the FAO74 or FAO90[2] classification in the HWSD_DATA table.

## 2.3 Linking Spatial Layer and Attribute Tables

The data model of the HWSD, excluding the spatial layer, is presented in Figure 2.

---

[2] FAO85 is used at times, but complemented by FAO90 codes.

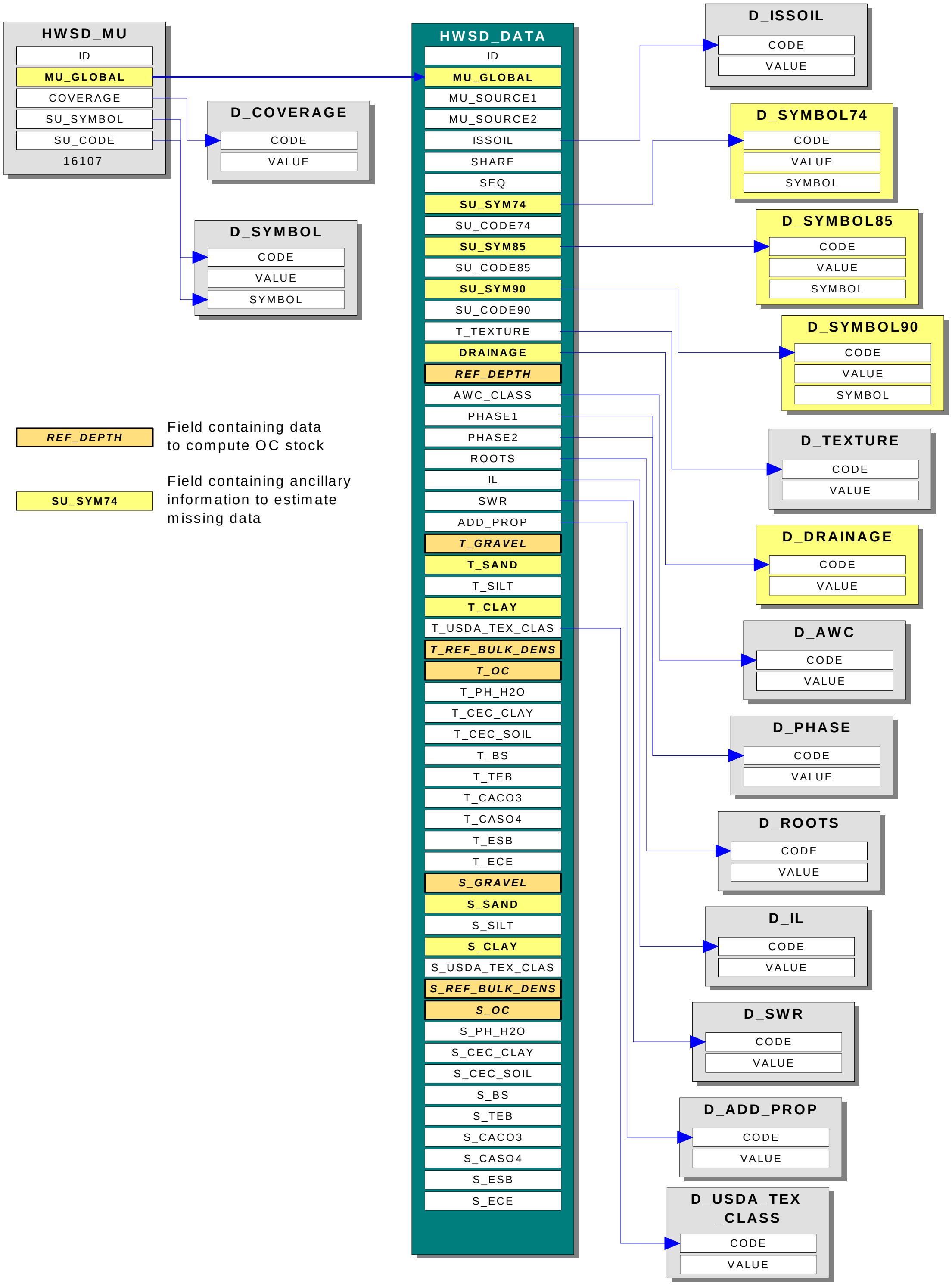


**Figure 2: HWSD Data Organization (only general and topsoil links shown)**

The attribute data are arranged in two main tables, one related to the mapping units (HSWU_MU) and one to the information of the properties of the typological units (HWSD_DATA). Soil attributes recorded as discrete values, such as codes and classes, are stored in dictionary tables with explanatory comments. In Figure 2 the fields used to generate the organic carbon (OC) maps are indicated by their colour: an orange background marks fields used to compute OC stocks while a yellow background marks ancillary information used to estimate missing data of the parameters needed to compute OC stocks.

The values of attribute table can link directly to the spatial layer via the [MU_GLOBAL] field. The [MU_GLOBAL] field contains entries between 2 and 31,773, which allows the field to be defined as a 16-bit integer format. This largely simplifies storing the MU identifiers directly in the raster layer and linking to the attribute table. The attribute tables contain 16,107 unique [MU_GLOBAL] identifiers, while the spatial layer contains 5 identifiers less (missing MUs: 57, 194, 969, 3084, 3113, 3326 and 4269). Since in terms of linking data in a relational database the field [MU_GLOBAL] forms the key and the raster data acts as the “parent” with the attribute tables the “child” tables the conditions of referential integrity are not met. In practical terms this may only be of consequence when the raster layer is represented as a table or attributes are derived from the raster layer, such as the area of the MUs, and then processed in a database. Where it is necessary to enforce referential integrity the MUs in the attribute table with missing correspondents in the raster data should be removed.

A 1:n relationship exists between the spatial layer / HWSD_SMU table and the attributes of the HWSD_DATA table. This condition makes mapping the complete range of attributes characterizing a mapping unit a non-trivial task. One approach to the situation is to link only the attributes of the dominant typological unit, as identified in the HWSD_SMU table, to the spatial layer. Mapping all data pertaining to mapping unit can be achieved for continuous numeric data by computing a weighted average for the area. For categorical data a translation of the table data into a spatial database requires generating 10 spatial layers for each of the categorical attributes. This can considerably increase storage requirements when using spatial layers directly to store soil attribute data.

# 3 SOIL ORGANIC CARBON COMPUTATION

SOC stocks were computed separately for both layers from SOC content, gravel content, soil depth and bulk density data.

SOC stock estimates were computed as:

$$SOC_S = SOC_C \times BD \times \left(1 - \frac{VS}{100}\right) \times LD \times 10^2 \ \ (t\ ha^{-1})$$

where

$SOC_S$: total amount of soil organic carbon to given depth ($t\ ha^{-1}$)
$SOC_C$: soil organic carbon content for given depth ($\%$)
$BD$: dry bulk density ($g\ cm^{-3}$)
$VS$: volume of stones ($\%$)
$LD$: Depth of soil layer ($m$)

According to the data recorded in the HWSD the total SOC stock were computed separately for the topsoil layer (0 - 30cm) and the subsoil layers (30 - 100cm). Where the soil depth was less than 100cm stocks were computed to that depth. The SOC stocks thus computed for the two layers were then combined to provide an estimate of SOC stock in $t\ ha^{-1}$ to a nominal depth of 1m.

## 3.1 Separation of Soil from Other Areas

A mapping unit may cover partially or completely areas which are not soil, such as water, rock or ice. For these areas no data on SOC parameters are available and yet they need specific consideration when computing SOC content and stock. The separation of soils from other areas is coded in the database in more than one table.

### 3.1.1 Non-Soil Surface Types Specified by the ISSOIL Field

Most explicitly areas of soils are given for the typological units through the field [HWSD_DATA.ISSOIL]. The table contains 624 records where the field [ISSOIL] is set to 0, of which 5 are organic soils (*fibric Histolos, Hfs*). The remaining 619 records mainly cover 10% (331) of the area linked to mapping units. Mapping units with a 100% share of non-soil surface cover types are 84. The sum of all shares comes to 100% in all but one case ([MU_GLOBAL]: 31538; sum of shares = 53%).

Soils could also be separated from non-soils by using the dictionary table SU_SYMBOL linked to the HWSD_SMU table. For reasons which could not be established some typological units were declared non-soils, in particular as sand dunes and rocky outcrops.

An overview of the entries in the field [HWSD_SMU.SU_SYMBOL] with the field [HWSD_DATA.ISSOIL] gave the combinations given in Table 1.

**Table 1:** **Classification of Soil and Surface Types of [HWSD_SMU.SU_SYMBOL] to Field [HWSD_DATA.ISSOIL]**

| [SU_SYMBOL] | Value | [ISSOIL] Entry | |
|---|---|---|---|
| | | *Non-Soil* | *Soil* |
| AC | Acrisols | 1 | 2334 |
| AL | Alisols | | 112 |
| AN | Andosols | 2 | 471 |
| AR | Arenosols | 8 | 3095 |
| AT | Anthrosols | 2 | 151 |
| CH | Chernozems | 6 | 711 |
| CL | Calcisols | 27 | 1882 |
| CM | Cambisols | 38 | 5794 |
| DS | Sand Dunes | 48 | 104 |
| FL | Fluvisols | | 1560 |
| FR | Ferralsols | | 1847 |
| GG | Glaciers | 9 | 2 |
| GL | Gleysols | 45 | 2325 |
| GR | Greyzems | 1 | 165 |
| GY | Gypsisols | 2 | 194 |
| HS | Histosols | 20 | 589 |
| IS | Island | 1 | |
| KS | Kastanozems | 6 | 826 |
| LP | Leptosols | 84 | 6985 |
| LV | Luvisols | 5 | 4567 |
| LX | Lixisols | | 1884 |
| NI | No data | 14 | 1 |
| NT | Nitisols | | 967 |
| PD | Podzoluvisols | 11 | 905 |
| PH | Phaeozems | | 1276 |
| PL | Planosols | | 713 |
| PT | Plinthosols | | 401 |
| PZ | Podzols | 91 | 1765 |
| RG | Regosols | 72 | 2868 |
| RK | Rock Outcrop | 68 | 103 |
| SC | Solonchaks | 11 | 594 |
| SN | Solonetz | 3 | 751 |
| ST | Salt Flats | 3 | |
| UR | Urban, mining, etc. | 6 | |
| VR | Vertisols | | 1163 |
| WR | Water Bodies | 40 | |

When classifying the records in the 3 soil classification dictionary tables into soil and non-soil and linking the information with the corresponding fields in the data table HWSD_DATA the combinations shown in Table 1 were found.

**Table 2: Soil and Non-Soil Combinations for FAO Soil Classes**

| **Field** | | | | **COUNT** |
|---|---|---|---|---|
| *[ISSOIL]* | *[FAO_74]* | *[FAO_85]* | *[FAO_90]* | *No.* |
| 0 | | | False | 153 |
| **0** | | | **True** | 4 |
| 0 | | False | False | 409 |
| **0** | | **True** | **True** | 1 |
| 0 | False | | | 57 |
| 1 | | | True | 26,325 |
| 1 | | **False** | **True** | 9 |
| 1 | | True | True | 11,389 |
| 1 | True | | | 9,384 |

The table shows a consistent result in the separation of soil from non-soil classes between the tables only for the FAO74 field. For FAO85 and FAO90 entries inconsistencies were found with the ISSOIL field in the data table and the classified soil entries in the dictionary tables in 5 cases. All cases concern *Histosols* and the reason for their classification as non-soil in the [ISSOIL] field is not evident.

Also found were 9 cases of inconsistent classifications with respect to soil / non-soil between the FAO85 and the FAO90 data. All cases relate to entries indicating “no data” in the FAO85 field, while the FAO90 field indicates a soil type (8 for “*ATa*” and 1 for “*ATc*”).

## 3.1.2 Coherence of Non-Soil Surface Types in Soil Classification

Coherence between the entries in the dictionary tables and the data table was assessed by analyzing data integrity for the [SYMBOL] field in the dictionary table and the corresponding fields in the data table.

- **[D_SYMBOL74]**
  All entries in the field [HWSD_DATA.SU_SYM74] had corresponding entries in the filed [D_SYMBOL74.SYMBOL]. Conversely, 7 entries in the field

[D_SYMBOL74.SYMBOL] ("*??*", "*Cg*", "*D*", "*M*", "*Mg*", "*Pf*" and "*Wx*") did not occur in the corresponding field of the data table.

- **[D_SYMBOL85]**
  Correspondence between the field [HWSD_DATA.SU_SYM85] and the linked field in the dictionary table could not be established for the entry indicating "No data" ("*ND*", 9 occurrences). The code in the dictionary table [D_SYMBOL85] for "No data" is "*NI*", while the data table uses "*NI*" and "*ND*", of which the latter is not defined in the dictionary table. The inverse relationship was not complete, because for 9 entries in the field [D_SYMBOL85.SYMBOL] ("*Bf*", "*Bm*", "*Dgd*", "*Eu*", "*Gms*", "*H*", "*MA*", "*NS*" and "*Uk*") no correspondence in the data table was found.
- **[D_SYMBOL90]**
  The situation for the field containing the FAO 90 classification codes was more confusing. There were 21 codes in the dictionary table without correspondence in the data table. There were also two codes ("*Glu*", 1 case and "*NI*", 12 cases) in the data table without corresponding entries in the dictionary table. The entry "*Glu*" is most likely a typing error for an "*Umbric Gleysols*", for which the code is "*GLu*".

### 3.1.3 Separation of Non-Soil Areas in SOC Processing

Those inconsistencies between the field [HWSD_SMU.ISSOIL] and the dictionary tables diminish the use of the field to separate soil from non-soil records in the data table.

The consequence of these conditions found in the database for the separation of soil from non-soil areas is to

a) use only the [HWSD_DATA.ISSOIL] field after amending the 5 cases of organic soils with a [ISSOIL] value of 0;

or

b) build a dictionary table with the soil codes by classification scheme and an additional field for the soil / non-soil information.

In the analysis of the data and processing for SOC maps the second option was used, because it offers a more flexible approach to defining coherent combinations of soil classes between classification schemes and because it does not modify any data in the HWSD_DATA table.

The presence of non-soil areas in a mapping unit impacts on computing and mapping SOC content and stocks. Using all typological units of a mapping unit the area-weighted SOC content is then the mean SOC content for the share of the mapping unit which is soil. In case the SOC content assigned to a mapping unit denotes the mean SOC content of the total mapping unit the figure has to be adjusted by the share of the non-soil area within the mapping unit. As a consequence, whenever there is a share of non-soil areas in a mapping unit the mean SOC content of the soil share in the mapping unit is always greater than the mean SOC content of the area of the total mapping unit. These options of expressing SOC content for mapping units has to be considered when calculating SOC stocks for mapping units which are expressed in area units, such as OC in $t\ ha^{-1}$.

## 3.2 Completeness of Parameter Data

As a first step the completeness of the data as regards the parameters used was assessed. An overview of the completeness check performed on the entries in the HWSD_DATA table is given in Table 3.

**Table 3: Completeness of Parameters for Computing Soil Organic Carbon Stock**

| Parameter | Soil Layer |  | Comment on Missing Data |
|---|---|---|---|
|  | *Topsoil* | *Subsoil* |  |
| SOC Content | 100.0% | 99.7% | Top: 3 records where T_OC = 0.0<br>Sub: 119 records |
| Gravel | 99.2% | 98.9% | Top: 358 org. soils, 5 min soils (Andosols)<br>Sub: 343 org. soils, 108 mineral soils |
| Bulk Density | 100.0% | 99.4% | Top: 2 records<br>Sub: 229 records<br>Substitution for organic soils |
| Depth | 100.0% |  | Limits of 10, 30 and 100cm, complete |

The percentages of completeness are based on 47,106 records for the topsoil (0 - 30 cm) and 39,681 (>30 – 100 cm) records for the subsoil layer. During processing blank entries in a field were treated as missing data and not set to be treated as zero (0), while zero (0) entries were treated as data. The evaluation of the completeness of the data to compute SOC stocks provided some information on peculiarities in the database, which were treated depending on the parameter concerned.

### 3.2.1 SOC Content

For the topsoil SOC content was defined for all typological units. For 3 records a value of 0 was given in the field [T_OC] (soil type "*CMx*" and "*ARo*"). By comparing the defining parameters of the typological units with similar ones the records containing 0 entries were re-assigned to the lowest values for [T_OC] other than zero.

For the subsoil data 12 records with 4 soil types (*“ARg”*, *“ARo”*, *“CMx”* and *“LXf”*) had values of 0 assigned in the field [S_OC]. By evaluating typological units with similar parameters to those with zeros for [S_OC] the zero entries were modified to values between 0.12 and 0.34%.

For the subsoil layer 107 records had no OC content information. The data were missing for 29 different soil types and from the properties of the typological units concerned particular reasons for the absence of data were not evident. The missing data were estimated mainly by using topsoil values to guide the search for typological units with comparable combinations of parameters. As ancillary information characterizing soil conditions parameters on drainage condition, texture and pH were used. Information of the parent material would have been valuable, but such data were largely unavailable for the cases where no OC values were recorded. For the analysis of the subsoil OC content the topsoil values were also taken into consideration.

Some doubts about the OC content of organic soils recorded in the database were raised when investigating the relationship. In the topsoil layer 10 records with organic soils had an OC content of <20%, 4 of <12%. For the subsoil layer 32 records indicating organic soils recorded values of SOC content of <20% and 12 of <12%. In 3 cases the OC content was <12% for both layers and it could be argued that these soils could be classified differently. In the analysis the soil classes were used as found.

Another area of uncertainty is the distribution of OC contents for organic soils. The relative frequency of SOC values in the HWSD_DATA table is presented in Figure 1

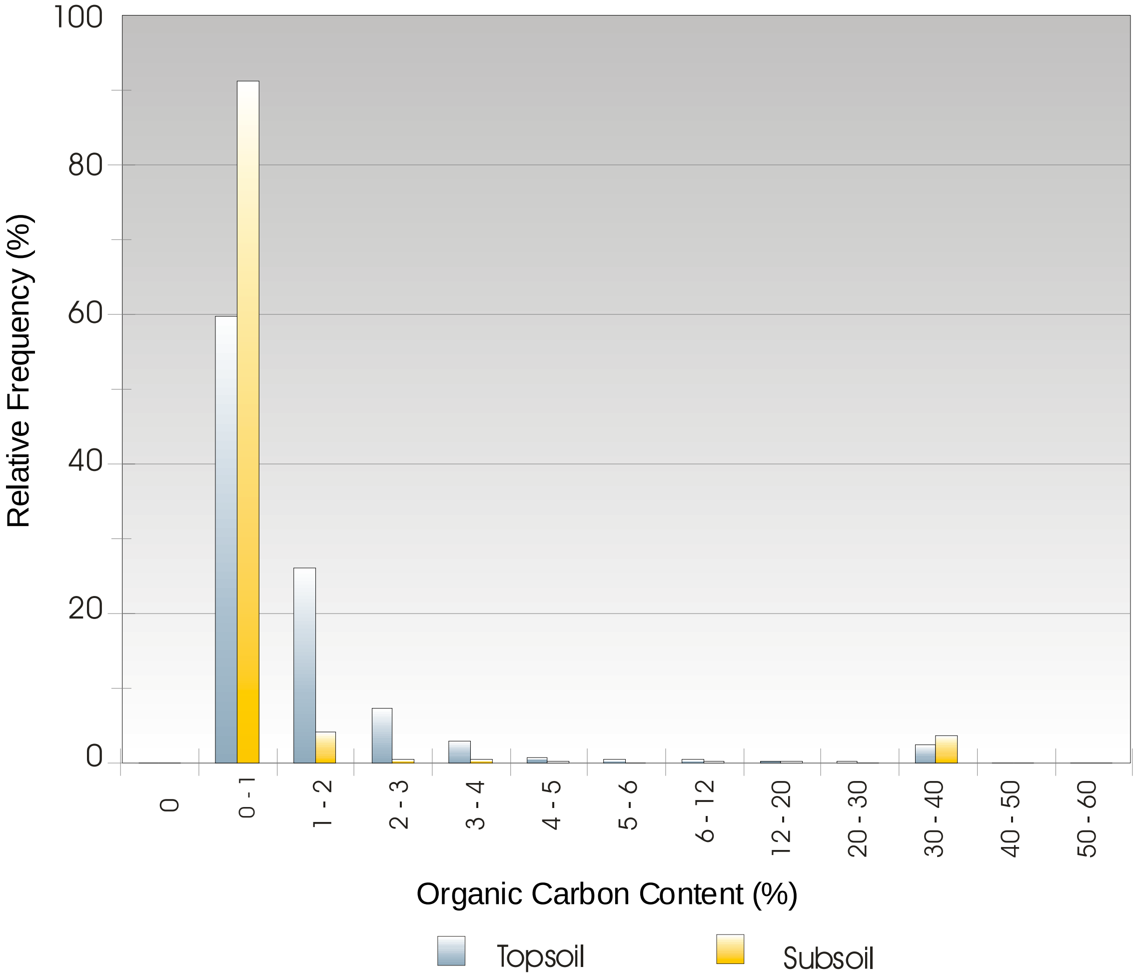


**Figure 3: Relative Frequency (%) of OC Content of Typological Units**

For the topsoil the majority of OC values (60%) fall into the range of 0 – 1%. For 85% of all typological units the OC content is < 2%. For the subsoil the OC value is less than 1% for 91% of the typological units and 95% have a value < 2%.

There are hardly any soils with an OC value of 6 – 30% (0.7% for topsoil, 0.2% for subsoil typological units). The OC values for organic soils are concentrated in the range of 30-40% OC content. The relative portion of typological units in this range is 2.6% for the topsoil and 3.6% for the subsoil. The relative portion does not reflect a higher occurrence of typological units in this range of OC values. The number of typological units is comparable (topsoil: 1,131 vs. subsoil: 1,134), but the difference in the relative portions is due to the lower number of typological units for the subsoil.

Within the range of 30-40% OC content the data concentrate on few values. For the topsoil 40.3% of the typological units have an OC content value of 33.63% and 33.3% of 35.27%. The subsoil OC content values are concentrated on 32.89% (35.0% of typological units) and on 39.16% (27.7% of typological units).

The spatial distribution of the two main OC content values for the topsoil is given in Figure 4.

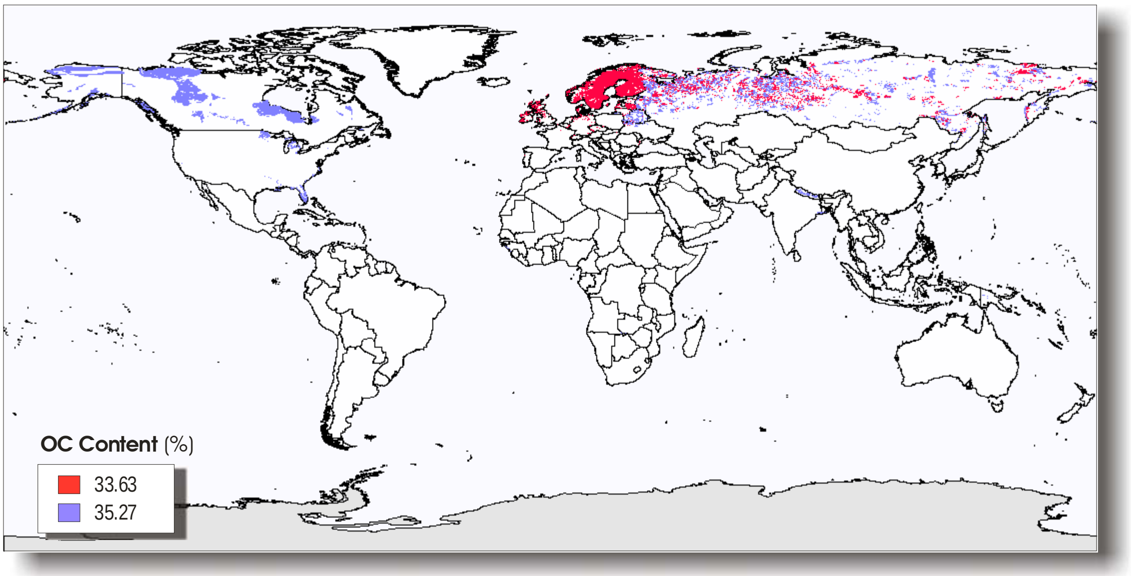


**Figure 4: Distribution of OC content 33.63% and 35.27% in Typological Units for Topsoil Displayed by Mapping Unit**

The map shows that a value of 33.63% OC content is only found in mapping units, which originate from the ESDB. The value of 35.27% OC content is found in the USA, Canada and Nepal. It is also found for typological units of the extension of the ESDB to the CIS countries, but not in the countries covered in V1.0 of the database.

The distribution of the soils with an OC content >12% by source is presented in Table 4.

Arranging the values of OC and clay content by source provides insight into a tendency for some combinations to occur in clusters and a dependency on the data source. For China only 1 organic soil class (*fibric Histosol*, *HSf*) is reported. For data originating from the DSMW 4 fixed combinations of OC and clay content are included. Data for *gelic Histosols* (*Ox*) soils are identical to the values given for the ESDB, while data for soils classified as *Histosol* (*O*) correspond to the *Eutric Histosols* (*Oe*) of the ESDB DATA. The typological data derived from the ESDB contains mainly 3 combinations of clay and OC content (21/35.27, 32/39.40 and 40/33.63). Notable for the data derived from SOTWIS is that *Histosols* (*H***) rarely contain data on clay content, which is not in line with the data coming form other sources.

**Table 4: Distribution of Soils with Entries in Field T_OC >12% and with more than 1 Occurrence**

| **Field in HWSD_DATA Table** | | | | | **REGION LABEL** | **Count** |
|---|---|---|---|---|---|---|
| *[SU_ SYM74]* | *[SU_ SYM85]* | *[SU_ SYM90]* | *[T_CLAY]* | *[T_OC]* | | *No* |
| **O** | | | **39** | **35.27** | **DSMW** | 35 |
| **Od** | | | **41** | **30.73** | **DSMW** | 123 |
| **Oe** | | | **40** | **38.37** | **DSMW** | 27 |
| **Ox** | | | **21** | **35.27** | **DSMW** | 31 |
| | | HS | 0 | 33.74 | SOTWIS | 12 |
| | | HSf | 0 | 29.22 | SOTWIS | 3 |
| | | HSf | 0 | 33.87 | SOTWIS | 47 |
| | | HSf | 0 | 47.24 | SOTWIS | 2 |
| | | HSf | 21 | 22.43 | SOTWIS | 15 |
| | | HSi | 0 | 19.97 | SOTWIS | 2 |
| | | HSt | 0 | 27.63 | SOTWIS | 4 |
| | | HSt | 0 | 30.73 | SOTWIS | 3 |
| | | HSt | 0 | 46.70 | SOTWIS | 2 |
| | | HSs | 0 | 39.94 | SOTWIS | 19 |
| | | HSs | 32 | 24.43 | SOTWIS | 2 |
| | | HSs | 67 | 27.15 | SOTWIS | 4 |
| | | **HSf** | **40** | **33.63** | **China** | 2 |
| | Od | **HSf** | **40** | **33.63** | **ESDB** | 393 |
| | Odp | **HSf** | **40** | **33.63** | **ESDB** | 2 |
| | Oe | HSi | 21 | **35.27** | **ESDB** | 67 |
| | Ox | HSi | 21 | **35.27** | **ESDB** | 217 |
| | Oe | HSl | 39 | **35.27** | **ESDB** | 28 |
| | **O** | **HSs** | **32** | **39.40** | **ESDB** | 3 |
| | **Oe** | **HSs** | **32** | **39.40** | **ESDB** | 117 |
| | | **LPq** | **24** | **14.00** | **SOTWIS** | 38 |
| | | **ANu** | **35** | **19.20** | **SOTWIS** | 5 |
| | | **CMu** | **20** | **18.23** | **SOTWIS** | 2 |
| | | **CMu** | **25** | **28.03** | **SOTWIS** | 4 |
| | | **GLu** | **31** | **19.37** | **SOTWIS** | 2 |

Globally, there are only 5 typological units with an OC content > 40% with a maximum at 47.24%. It would seem that the distribution of the OC content values and the range of the values does not represent the OC content in organic soil. For peat one could have expected an OC content closer to 58%. This value is found in soil profiles data for peat in the boreal and arctic region, but also the tropical peat in south-East Asia. The concentration on mid-range values for OC content for organic soils does not automatically result in an underestimation of the global SOC stocks. The global stocks could still be estimated acceptably when the distribution of OC content values were evenly distributed around the central values. However, the profile data for Europe indicates a prevalence of OC content values >40% for organic soils (Hiederer, 2009).

A comparison of the relative frequency of the OC content for organic soils between the typological units of the HWSD and the soil profile data of the ISRIC-WISE V3.1 data is presented in Figure 5.

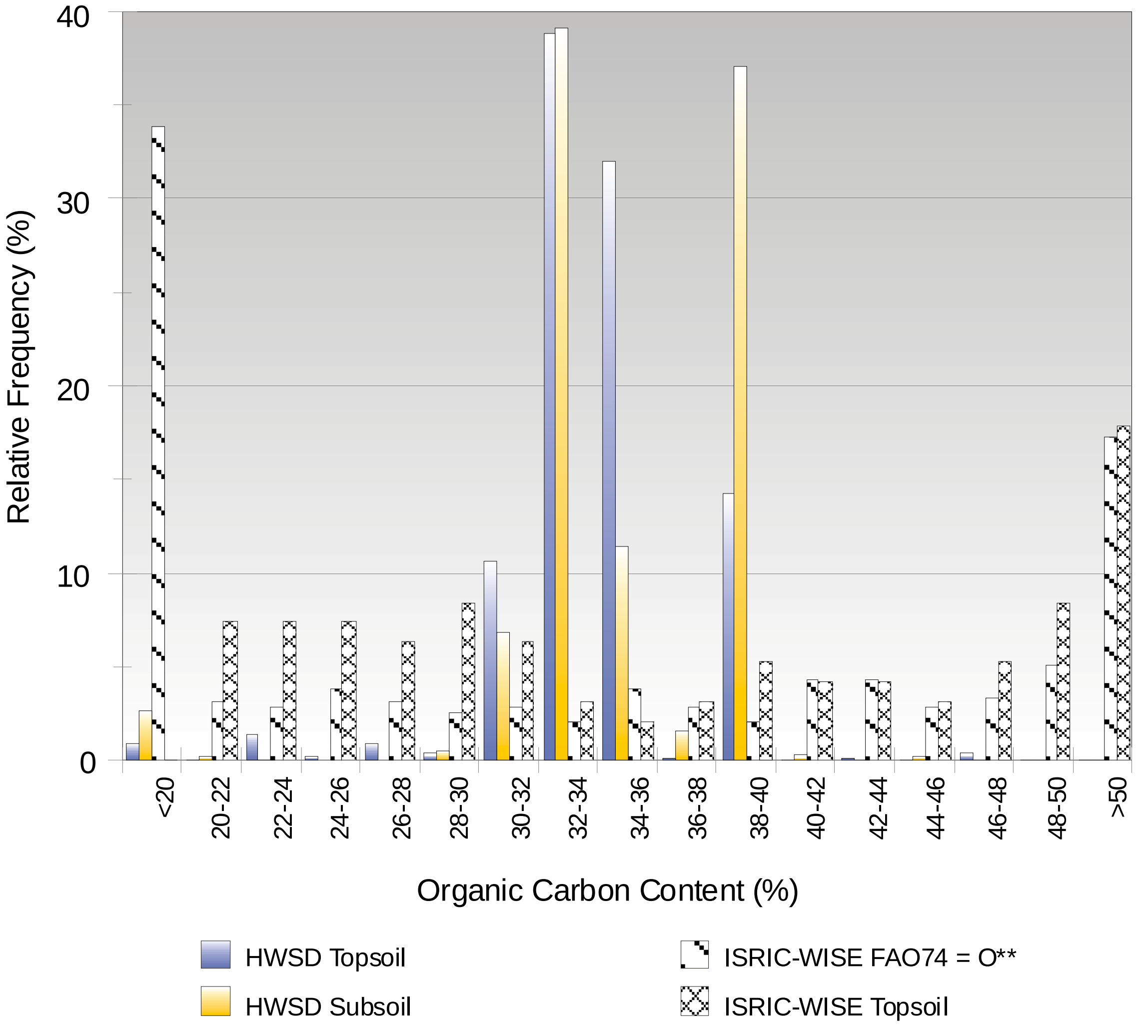


**Figure 5: Relative Frequency (%) of OC Content of HWSD Typological Units and ISRIC-WISE V3.1 Profile data for Organic Soils**

The graph shows the relative frequency of topsoil and subsoil OC content values occurring within bins with a range of 2% for soils where the field SU_SYM74 contains a value of “O**” (where ** means any character) or the field SU_SYM90 a value of “H*”. Also given are the relative frequencies of OC content given in the ISRIC-WISE V3.1 database in the field WISE3_HORIZON.ORG_C for entries in the field

WISE3_SITE.FAO_74 of “O**” and in the processed data for the topsoil layer for the same condition. A condition for the OC content to be > 20% was set for all data.

The concentration on few values in the HWSD is obvious from the graph. The ISRIC-WISE V3.1 profiles have OC content values across the whole range and 18% of the profiles have OC content values > 50%.

### 3.2.2 Gravel Content

Data for Gravel in the topsoil were recorded for 99.2% of the typological units. A distinction in treating the missing information was made between mineral and organic soils:

> ***Mineral Soil Type***
>
> For 3 cases the soil type was “*Th*” (*Humic Andosol*; FAO74). Usually (all other 36 cases), this soil type is assigned a gravel content of 8%. For the two cases of “*Tm*” (*Mollic Andosol*; FAO74) without data one would have expected a gravel content of 4%, which is given for all other 17 cases of the soil type. With the gravel content being assigned by a pedo-transfer rule the reason for the partial absence of a value for the gravel content for those two soil types is not immediately obvious.
>
> ***Organic Soil Type***
>
> Of the cases without data 358 records related to an organic soil type. It was assumed that the gravel content could be set to 0.

While the absence of other parameters used to compute SOC stocks result in missing data for the typological unit, the absence of data on the gravel content (volume of stones) should only result in the expression (1-*VS*/100) to become 1.

### 3.2.3 Bulk Density

A value for bulk density was not recorded in the database for 128 records in the topsoil and 229 records in the subsoil layer. No zero entries were recorded for either layer. Since the PTRs for bulk density and most functions usually only cover mineral soils the approach to estimating data for the missing fields records with mineral soil types differed from the one applied to estimate the parameter for an organic soil.

> ***Mineral Soil Type***
>
> For the topsoil layer the missing data were estimated mainly from other typological units with comparable values of key parameters. The main

> parameters were texture and OC content. Data missing in the subsoil data could be estimated by comparing typological units using the topsoil bulk density.
>
> ***Organic Soil Type***
>
> In cases where a figure of bulk density was missing for organic soils the values were estimated by a function based on SOC content. This function was also applied to existing values of bulk density for organic soils, thus substituting non-zero entries in the database. This approach was taken because the database contained 1,086 records of organic soils in the topsoil and 1,056 in the subsoil layer with a bulk density ranging between 1.15 and 1.41 $g\ cm^{-3}$ (query based on OC content >18%) This range of values is obviously incorrect for organic soils and using the figures would lead to a considerable over-estimation of the SOC stock in the affected soil mapping units by a factor of 10.

From the analysis of the records of organic soils with bulk density it would appear that a PTF was used to estimate the values. This was confirmed in the documentation for the HWSD V.1.1, where as a reference of the values an on-line bulk density calculator was given (http://www.pedosphere.com/resources/bulkdensity/index.html). The basis for the bulk density calculation is the estimation of pore space from soil texture, where bulk density is defined as BD = (1-pore space)*2.65 ($g\ cm^{-3}$) (Saxton, *et al*., 1986). According to Saxton, *et al*. (1986) the PTF was developed for estimating soil water characteristics for agricultural soils with a range of SOM of 0-3% (equivalent to SOC 0 - 1.8%) and with a limited range of textural values.

The PTF seems to have been applied whenever sufficient data were available, i.e. when the texture components recorded. Since those parameters were also recorded for some typological units with organic soils the rule intended for mineral soils was applied to organic soil, too. Substituting the existing values with those derived from a function should rectify the situation.

Values for bulk density also exist for textures outside the valid range of the calculator. An analysis of the frequency of values occurring suggests that a PTR was also used in some cases to estimate bulk density. In the topsoil layer a frequency of >1,000 (69% of STUs) was found for 14 values and a value of 1.39 is recorded for 3,935 typological units (8.4%).

### 3.2.4 Depth of Soil Layer

MU attributes are provided for a topsoil and, where appropriate, a subsoil layer. The limits of the depth is set in the field [REF_DEPTH]. It is either 10cm (4,488 records), 30cm (2,938 records) or 100cm (39,681 records). For 619 records no depth value is available. In all these cases the [ISSOIL] field is set to 0. There are 5 cases where a value of 100 is given in the field [REF_DEPTH], although the [ISSOIL] field is set to 0.

All cases concern organic soils (FAO90: *Hsf*). No cases were found where a value is given for the subsoil and the [REF_DEPTH] entry is not set to 100.

In general, when a value for a soil attribute is given for the subsoil a related value should also exist for the topsoil. This is not always the case. For the soil attributes used to compute OC stocks a situation of "*no data for the topsoil where a value exists for the subsoil*" was found as:

**Gravel**: 18 cases (3 for FAO74: *Th*; 15 for FAO90: *Hsf*)

**Bulk Density**: 8 cases (1 for FAO90 *ARa*; 1 for FAO90 *ARb*; 3 for FAO90 *ARh*; 1 for FAO90 *ARo*, 1 for FAO90 *PZh*; 1 for FAO90 *RGd*)

**OC**: 0 cases (for all cases with [S_OC] > 0 a value [T_OC] > 0 exists)

Cases of a missing related topsoil value for an attribute can exist, for example the absence of gravel in the topsoil for organic soils. The absence of data in the topsoil can be of consequence when combining data for the two layers. In a processing environment where blank fields are treated as missing data an out join has to be set to combine the layer data. With the number of cases with missing topsoil data found in the table HWSD_DATA the practical consequences are deemed to be of no significant impact on the computations.

# 4 HWSD SOC PARAMETER MAPS

The maps of the parameters used to compute SOC stocks are generated using the complete set of up to 10 typological data pertaining to a mapping unit. The parameters are all in continuous numeric format and it is therefore possible to integrate all data to a single value and thus a single spatial layer. The integration of all typological data was preformed by weighing the parameter values according to the share of the typological unit in the associated mapping unit, first separately for the topsoil and subsoil layer and then for the combination of both layers to a nominal depth of 1m.

Since typological units are geographically positioned within a mapping unit only the mean values from all typological units of a mapping unit can be assigned to a grid cell. Two approaches to integrating the data can be distinguished:

- integrating only data for soil typological units;
- integrating data from all typological units.

The first option generates comparable data for soils, but for computing carbon stocks the areas of the mapping units have to be adjusted from the surface area. The second option simplifies the computation of carbon stocks by the surface area of a grid cell or mapping unit, but the integrated parameter values are not characteristics of the soil units. With the focus on calculating SOC stocks the second processing option was chosen. However, care should be taken when overlaying the gridded SOC content data with thematic layers containing non-soil areas form other sources, such as urban zones from land cover datasets. A simple masking procedure for non-soil areas does not adequately address the situation and would lead to artefacts and reduced estimates of SOC contents.

A similar choice of processing options is open for the treatment of the layer depth. The parameters can be expressed for the depth of the soil layer or a given fixed volume. In the processing performed the parameters are integrated over to the available depth of the soil layer and not scaled to a fixed depth.

## 4.1 SOC Content

The SOC content is recorded directly in the HWSD_DATA table as a parameter of the typological units and given separately for the topsoil and the subsoil. In addition to the topsoil and subsoil data the mean SOC content for both depth layers was computed. A single figure of SOC content for the three depth layers was computed by integrating data from all typological units associated with a mapping unit, using weights by depth and area. The resulting layer of SOC content to a maximum soil depth of 1m is depicted in Figure 6.

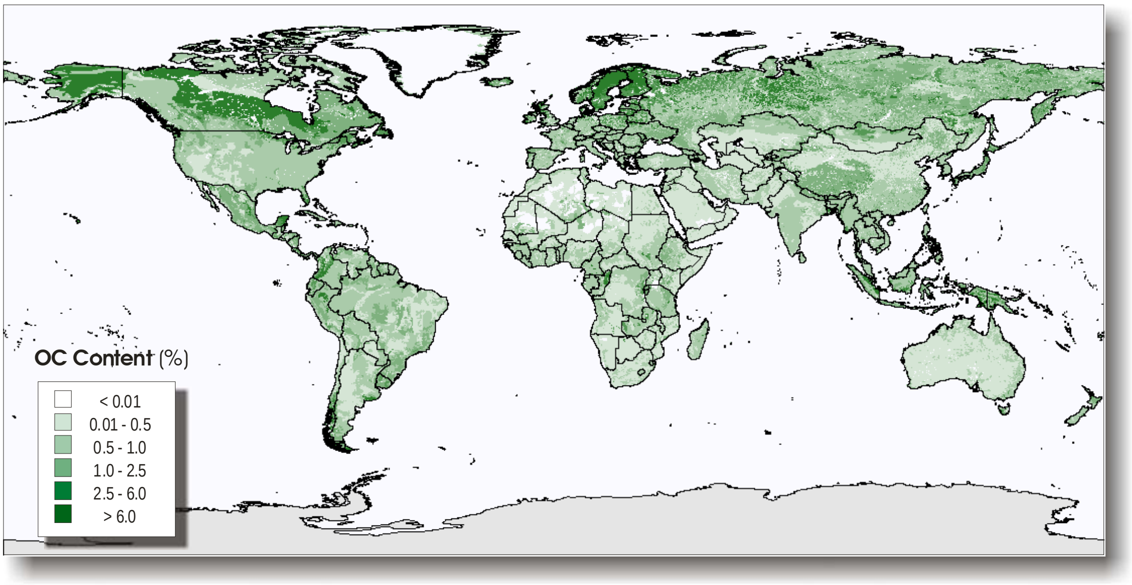


**Figure 6: Global Soil Organic Carbon Content Estimates (%) for Combined Topsoil and Subsoil Layers**

The map shows the mean SOC content for the mapping units computed from typological units pertaining to soil and non-soil areas and the depth of the soil up to 1m. The limit to the soil depth rather than scaling the SOC content to a fixed depth leads to some areas with shallow soils to have similar values to those with deep soils. Thus, the soils of the border region between Algeria and Niger show almost the same SOC content as the soils in the central Amazonian basin in Brazil (1.5%), but the former are only 10 cm deep.

## 4.2 Gravel Content

A value for the gravel content is given for 46,749 typological units for the topsoil. For 3,664 typological units (7.8%) a value of 0 is recorded. A value of 1% for the topsoil is given for 30% of the typological units and is the most frequently used value. Other frequently used values are 10% (6.0% of typological units), 20% (8.2% of typological units) and 30% (2.8% of typological units). This concentration on single values indicates that the gravel content is estimated to the nearest 10% for those typological units and not the nearest 1%. There is no discernable spatial pattern to the use of the dominant values for gravel content for the topsoil. For the subsoil gravel content the concentration on single values is far less notable with a wider spread of values across the range.

The gravel content of the topsoil and subsoil layers was combined for the available soil depth up to 100 cm. The spatial distribution of the gravel content for combined layers is presented in Figure 7.

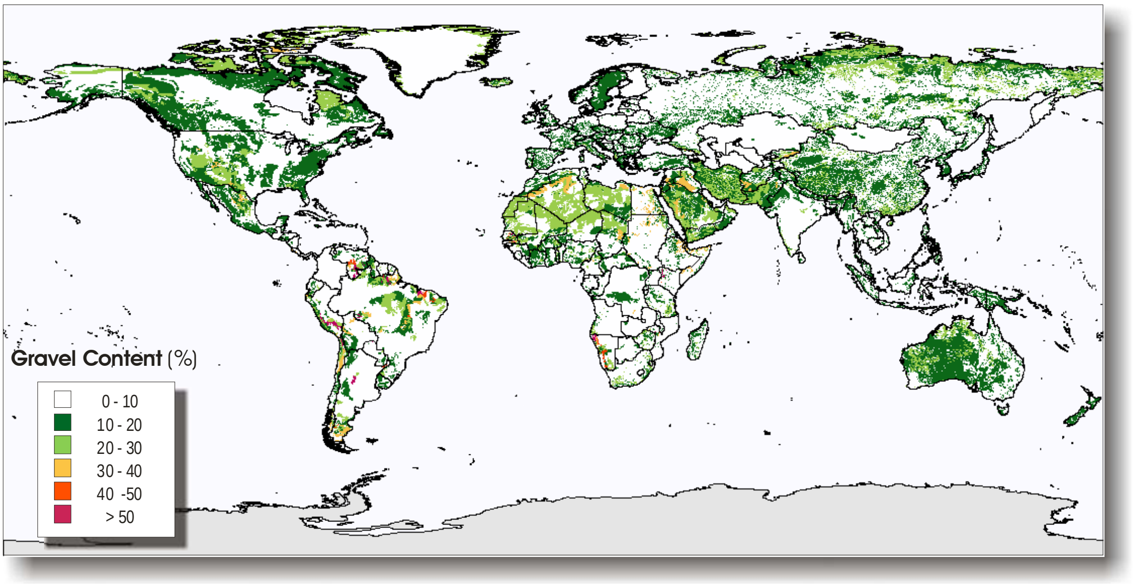


**Figure 7: Distribution of Gravel Content (%) in Combined Topsoil and Subsoil Layer**

Gravel contents of >30% are mainly found in some regions of South America, Africa and the Arabian peninsula. No particular spatial patter of distribution of gravel content is evident in the graph.

## 4.3 Bulk Density

The values of bulk density for organic soils were found to be unrealistically high in the HWSD for both, the topsoil and the subsoil typological units, mainly because where values were estimated they follow the PTF defined for mineral soils (Saxton, *et al.*, 1986). An alternative is to replace the bulk density values for soils with a higher OC content in the HWSD with data from a PTF based on profile data, which provide reasonable estimates of bulk density from OC content for these soils. As is the case with the PTF given by Saxton *et al.* (1986) the PTFs from other sources concentrate on estimating bulk density for mineral soils. Frequently, information on soil texture forms a parameter in the PTF, such as the PTF of Rawls (1983), which uses mineral matter and a value of 0.244 $g\ cm^{-3}$ for the bulk density of organic matter, or Kaur, *et al.* (2002) or Tranter, *et al.* (2007). A PTF using a fixed values for soil texture and a variable value for organic matter was developed by Adams (1973) and then modified by de Vos *et al.* (2005). The modified PTF uses a fixed bulk density of 1.661 $g\ cm^{-3}$ for the mineral material and a value of 0.312 *g cm*$^{-3}$ for organic material. De Vos, *et al*. (2005) also provides a list of PTFs from various sources. For the purpose of estimating bulk density mainly for soils high in organic carbon content the functions can be divided into those using only organic carbon as a parameter (or organic matter (OM) from loss on ignition) and those including texture information as a defining parameter. In case of a bivariate relationship of OC or OM and bulk density, i.e. where information on texture is not included, three types of models can be distinguished:

a) Transformation of OC or OM (logarithmic)

$$\rho_b = a * \ln(OC) + b$$

b) Transformation of bulk density (logarithmic or power variable)

$$\rho_b = e^{(a * OC + b)}$$

c) Reciprocal

$$\rho_b = (a * OC + b)^{-1}$$

The model of using a linear regression between the log-transformation of bulk density ($\rho_b$) and OC content is conceptually comparable to the relationship used by Ruehlmann & Körschens (2009). The reciprocal functions are largely derivates of Adams (1973) with a fixed value for bulk density of the mineral material. Adams (1973) uses a value of 0.311 *g cm*$^{-3}$ as a default for the bulk density of organic matter, whereas Rawls (1983) used a default of 0.244 *g cm*$^{-3}$ and Rawls & Brakensiek (1985) a value of 0.224 *g cm*$^{-3}$.

The bulk density estimated by the PTFs for organic matter varies between 0.10 *g cm*$^{-3}$ for the log-transformed 2nd order polynomials to 0.19 *g cm*$^{-3}$ for the subsoil layer estimated by Harrison & Bocock (1981). PTFs using a power transformation (square root) yield negative bulk densities for the organic matter.

These functions concentrate on estimating bulk density for mineral soils, not organic soils. The transformation of the organic carbon stretches the range of organic carbon for such soils, but results in a compressed range for organic soils. Therefore, for the purpose of estimating bulk density in the absence of information on the mineral material the functions should be based on a transformation of the bulk density parameter. Furthermore, some of the PTFs were developed with data from restricted areas or specific land uses, such as European forests, and may thus not appropriate to estimating bulk density for organic soils at global scale. Rather than adapting one of the PTFs a function tailored to the purpose was defined from soil profile data.

In the course of this evaluation the logarithmic transformations of the OC and bulk density with subsequent linear regression and the simple reciprocal model were used. The Bleasdale yield-density model was used only to evaluate whether a more complex model would improve the estimates.

### 4.3.1 Soil Profile Data from SPADE/M

The *Soil Profile Analytical Database for Europe for Measured Data* (SPADE/M) of the ESDB contains typical soil profile data for a wide range of soil types. The profiles were used to support the development of PTRs of the ESDB to extend the range of parameters. OC content and bulk density of the profiles were largely measured, although there are also cases where the bulk density values were derived from PTFs. To provide an estimate of the distribution of OC content and bulk density suitable for use with the HWSD layers the horizon data were processed to conform with the depth ranges of the topsoil and subsoil. The relationships between OC content and bulk density of 308 profiles with data for the layers are presented in Figure 8.

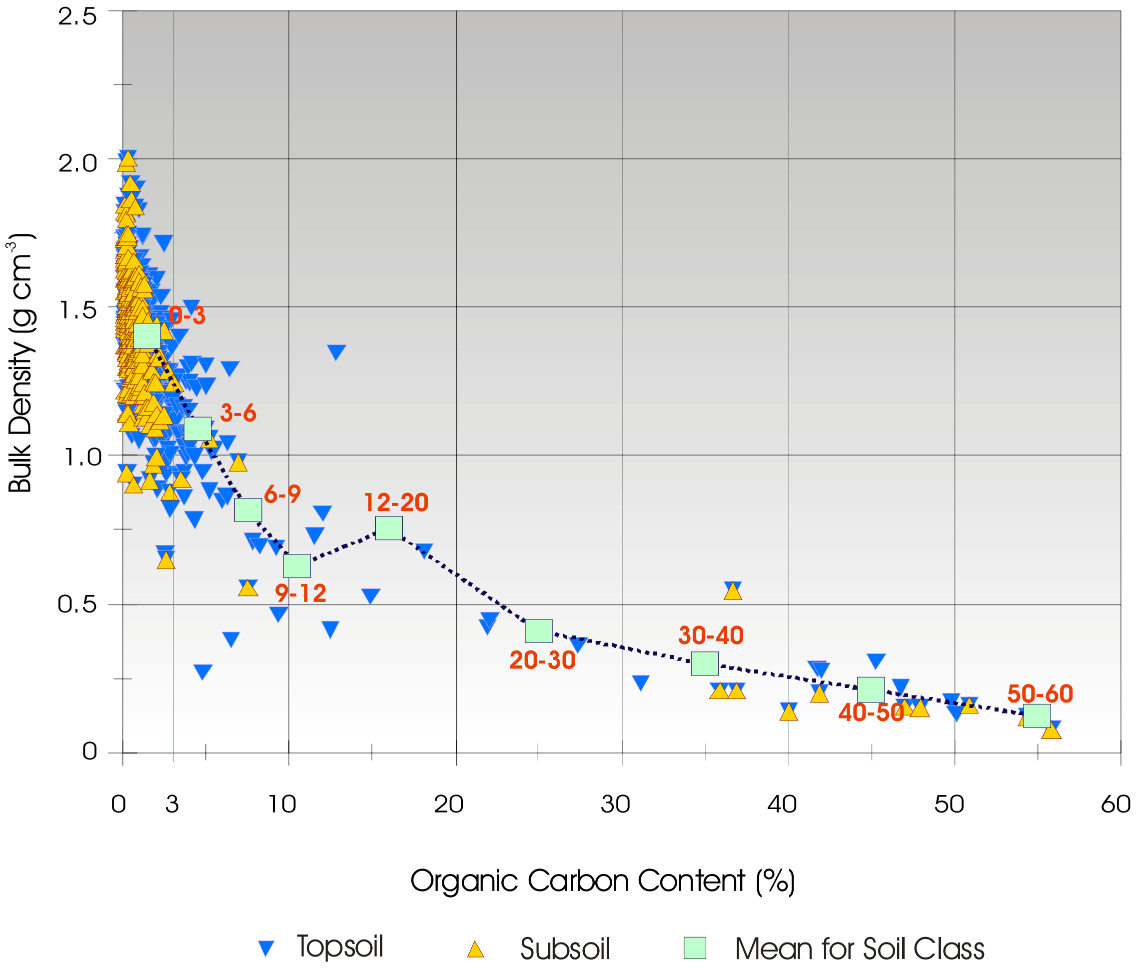


**Figure 8: Relationship between Organic Carbon Content (%) and Bulk Density for SPADE/M Profiles for Calculated Topsoil (0-30cm) and Subsoil (30-100cm) Layers**

The graph suggests a non-linear relationship between OC content and bulk density when the whole range of OC content values is considered. The class means follow close to linear trends for profiles with an OC content of 0 – 12% and 20-60%, albeit with a different slope. The class mean of the bulk density of OC contents between 12-20% is very much defined by a profile with an unusually high bulk density for the amount of OC content.

The application of the parameter treatment and model used to estimate the relationship between OC content and bulk density in the SPADE/M data is given in Table 5.

**Table 5: Relationship between Organic Carbon Content and Bulk Density for Log-transformed Organic Carbon, Log-transformed Bulk Density and Simple Reciprocal Model for Combined topsoil and Subsoil Layers derived from SPADE/M Profiles**

| Model to Estimate Bulk Density | Minimum OC | Parameter | | Coeff. of Determ. | OM Bulk Dens. |
|---|---|---|---|---|---|
| | | Coefficient | Constant | | |
| | % | *a* | *b* | $r^2$ | $g\ cm^{-3}$ |
| $\rho_b = a * ln(OC) + b$ | All Data | -0.208 | 1.342 | 0.62 | **0.50** |
| | > 3% | -0.360 | 1.574 | 0.78 | 0.11 |
| $\rho_b = e^{(a * OC + b)}$ | All Data | -0.047 | 0.364 | 0.83 | 0.09 |
| | > 3% | -0.042 | 0.189 | 0.85 | 0.11 |
| $\rho_b = (a * OC + b)^{-1}$ | All Data | 0.074 | 0.639 | 0.72 | 0.20 |
| | > 3% | 0.079 | 0.611 | 0.78 | 0.19 |

The relationships are defined by data from 589 topsoil and subsoil layers belonging to 308 profiles. Presented are the relationships when all data are used and when the fit is based only on data with an OC content > 3%. The 3% threshold was selected since at this value it continues from the range of the PTF used in the HWSD to estimate bulk density.

For the log-transformed bulk density model the parameters are broadly comparable to the values derived from the means for OC content and bulk density published by Ruehlmann & Körschens (2009) in Table 1 for the global parameterization data set ($BD = 1.411\ x\ e^{-0.064xOC}$).

For all relationships the bulk density of OM (58% OC content) is computed. Notable is the high value of 0.50 $g\ cm^{-3}$ for the log-transformed OC when using all data. It is caused by the variability of in the OC vs. BD relationship in profiles with low OC content. The fit improves significantly when using the 3% threshold for data included in the analysis and the bulk density for OM becomes 0.11 $g\ cm^{-3}$. The highest and consistent values for the coefficient of determination is found for the linear relationship of the log-transformed bulk density. Bulk density for peat varies by region, also in Europe. For peat in Scotland bulk densities of 0.10 $g\ cm^{-3}$ are reported, while for peat in England and Wales values of up to 0.40 $g\ cm^{-3}$ is found (Bradley, *et al.*, 2005). All models, with the exception of the log-transformed OC without restriction, could therefore be rejected based on the fit of the data or the estimated bulk density for organic matter.

The relationships were further evaluated by their residuals. A graphical presentation of the mean of the residual by OC class (MEAN[*measured* – *estimated*]) is given in Figure 9.

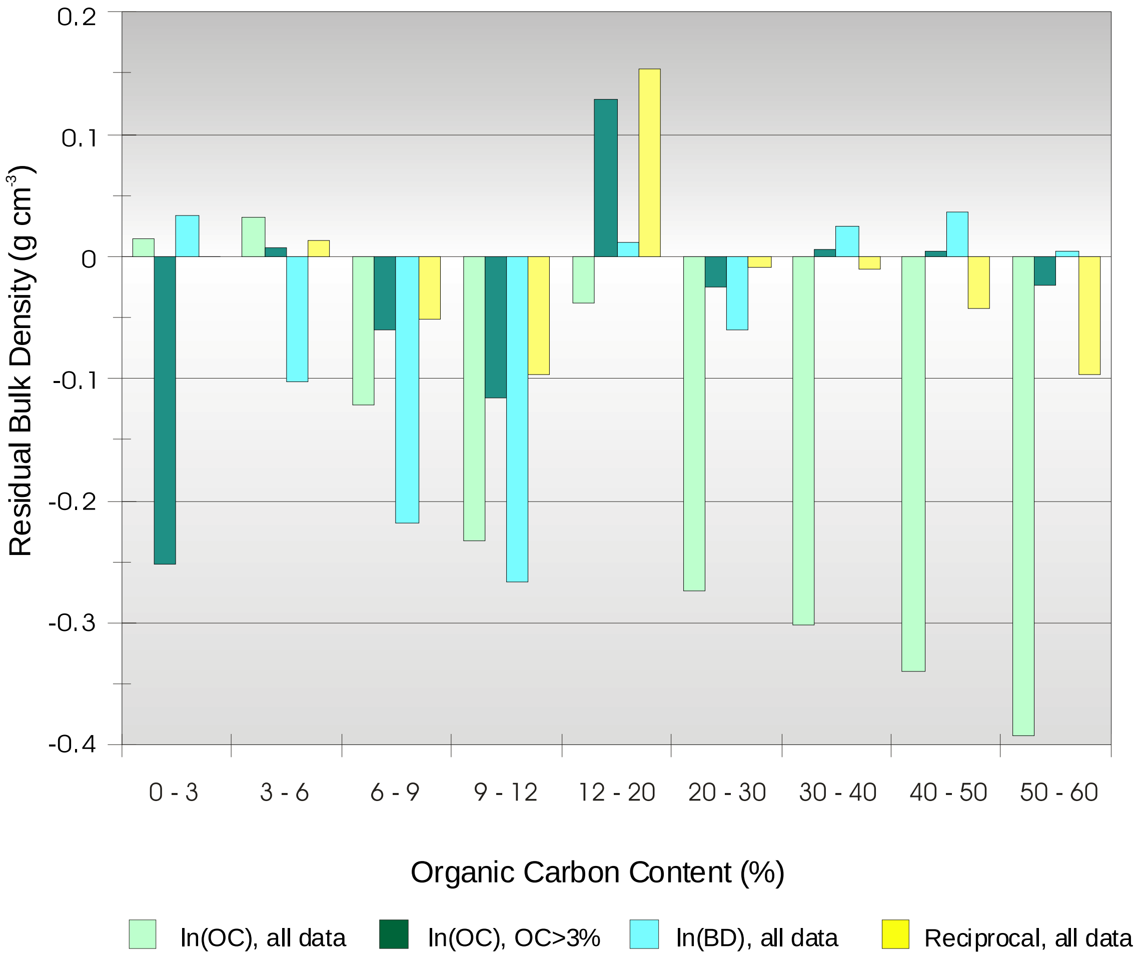


**Figure 9: Mean of Residuals of Bulk Density Estimates by Organic Carbon Content for SPADE/M Combined Topsoil and Subsoil Layers**

The graph illustrates a marked over-estimation of the bulk density when using all data and the log-transformed OC content to characterize the relationship. Yet, when limiting the data to layers with an OC content >3% the model describes well the bulk density for organic soils, at the expense of estimating the bulk density for mineral soils with an OC content <3%. This tendency of improving the fit is also found for the log-transformed bulk density and reciprocal model, but to a much lesser extent and with only marginally lower residuals for organic soils. A common tendency of all models used is the increase in residuals in mineral soils with increasing OC content to 12%. This common tendency may indicate that the relationship between OC content and bulk density for mineral soils differs to some degree from the relationship for organic soils.

No significant difference was found between the regression coefficients when analyzing the topsoil and subsoil layer data separately (95% confidence level). However, for all models

- the subsoil layer fits closer the model;
- the estimated bulk density of OM is lower in the subsoil than the topsoil layer;
- the differences are more pronounced for the relationship using a 3% OC threshold.

Using only a single data set with restricted geographic coverage would not allow substantiating the general applicability of the characteristics found.

### 4.3.2 Soil Profile Data from ISRIC/WISE V3.1

Because the profiles of the SPADE/M data are restricted to soils found in Western Europe the relationship between OC content and bulk density was also assessed using the ISRIC-WISE V3.1 (Batjes, 2008) data set which provides global soil profile data. Data from the pedological horizons of the profiles were re-arranged into topsoil and subsoil layers as for the SPADE/M profiles was performed.

The relationship between OC and bulk density for the global ISRIC-WISE V3.1 data by soil layer is presented in Figure 10.

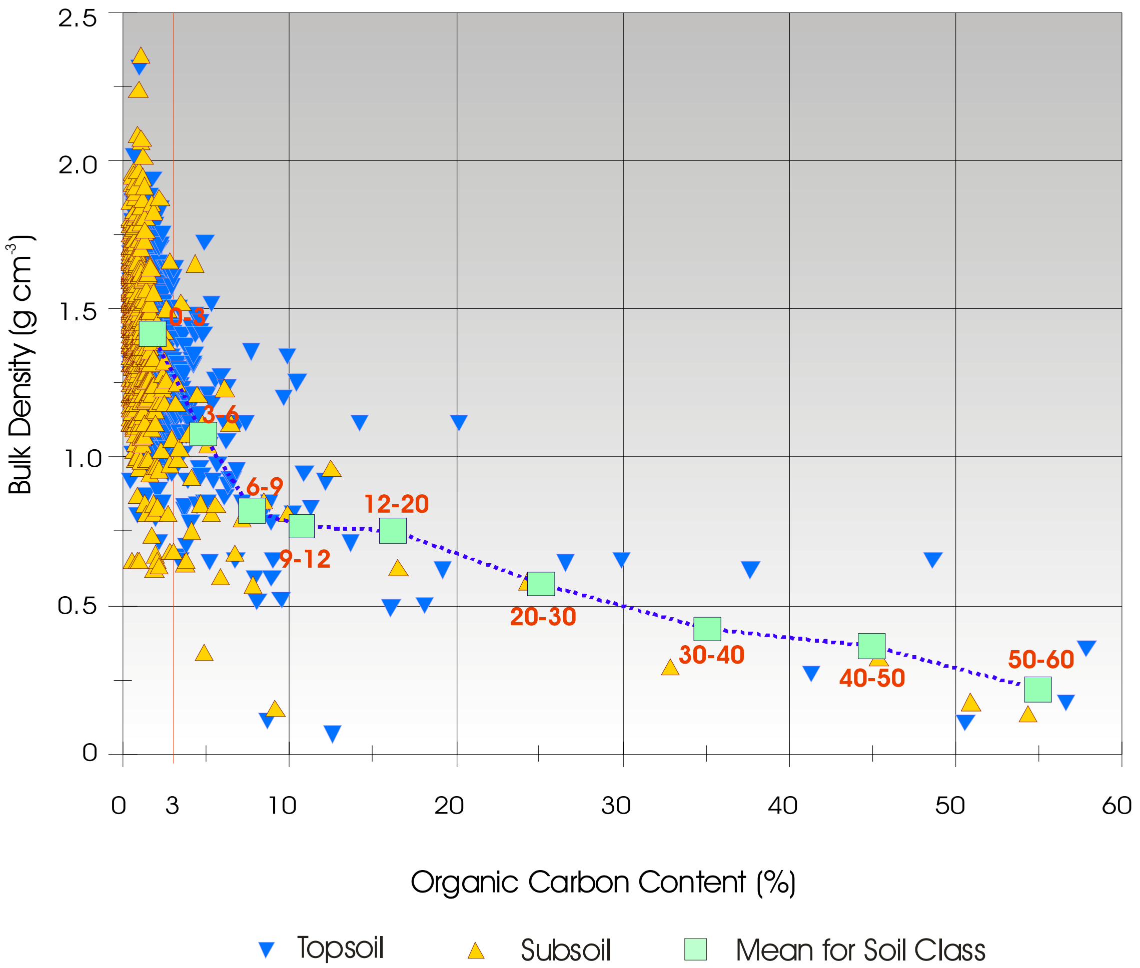


**Figure 10:** **Relationship between Organic Carbon Content (%) and Bulk Density for ISRIC-WISE V3.1 Profiles for Estimated Topsoil (0-30cm) and Subsoil (30-100cm) Layers (Global Data Subset)**

For the sake of clarity only every 5th profile of the database are shown in the graph. Compared to the topsoil and subsoil data from the SPADE/M profiles the ISRIC-WISE profiles show a larger spread of values and in particular the occurrence of bulk densities <0.75 $g\ cm^{-3}$ for OC contents <5%. For the ISRIC-WISE profile data the method used to establish the bulk density values is not always known. It may be assumed from the meta-data of other soil profile databases that the bulk densities given are based in part on PTFs rather than actual measurements. The data set also covers a wider range of soils and conditions, which may lead to the larger variability of the data.

To model the relationship of OC content with bulk density for the global data the same models used for the SPADE/M profiles were used (log-transformed OC content, log-transformed bulk density and simple reciprocal model). Results from the Bleasdale yield density model were found to be very close to those of the simple reciprocal model and are not further specified in this evaluation. The parameters for models used are summarized in Table 6.

**Table 6:** **Relationship between Organic Carbon Content and Bulk Density for Log-transformed Organic Carbon, Log-transformed Bulk Density and Simple Reciprocal Model for Combined topsoil and Subsoil Layers derived from ISRIC-WISE V3.1 Profiles**

| Model to Estimate Bulk Density | Minimum OC | Parameter | | Coeff. of Determ. | OM Bulk Dens. |
|---|---|---|---|---|---|
| | | Coefficient | Constant | | |
| | % | *a* | *b* | $r^2$ | $g\ cm^{-3}$ |
| $\rho_b = a * ln(OC) + b$ | > 0.1% | -0.151 | 1.314 | 0.32 | **0.70** |
| | > 3% | -0.308 | 1.482 | 0.41 | 0.23 |
| $\rho_b = e^{(a * OC + b)}$ | > 0.1% | -0.044 | 0.343 | 0.46 | 0.11 |
| | > 3% | -0.034 | 0.100 | 0.46 | 0.15 |
| $\rho_b = (a * OC + b)^{-1}$ | > 0.1% | 0.066 | 0.661 | 0.39 | 0.22 |
| | > 3% | 0.060 | 0.702 | 0.40 | 0.24 |

In the analysis of the parameters the data were limited to those layers with an OC content > 0.1%. This threshold reduces the large amount of layer data with an ambiguous relationship between OC content and bulk density.

For the treatments the fit between the profile and the modelled data was much less well defined than for the SPADE/M profiles. The coefficient of determination does not exceed 0.46, i.e. lass than half the variation in the profile data could be explained by the modelled data. For the bulk density of organic matter (58% OC assumed) the log-transformed OC content resulted in an unrealistically high value of 0.70 $g\ cm^{-3}$ when the layers included in the analysis have an OC content of > 0.1%. The range of the OM bulk density of the other 5 treatments ranges from 0.11 to 0.24 $g\ cm^{-3}$. These values are close those reported by Rawls (1983) and Ruehlmann & Körschens (2010) for global data sets.

The two profile databases provide higher values of the coefficient of determination for the log-transformed bulk density than for either the log-transformed OC content and the reciprocal. To better evaluate the fit of the data the mean of the residuals were plotted for the log-transformed linear and the reciprocal model. The resulting graph is presented in Figure 11.

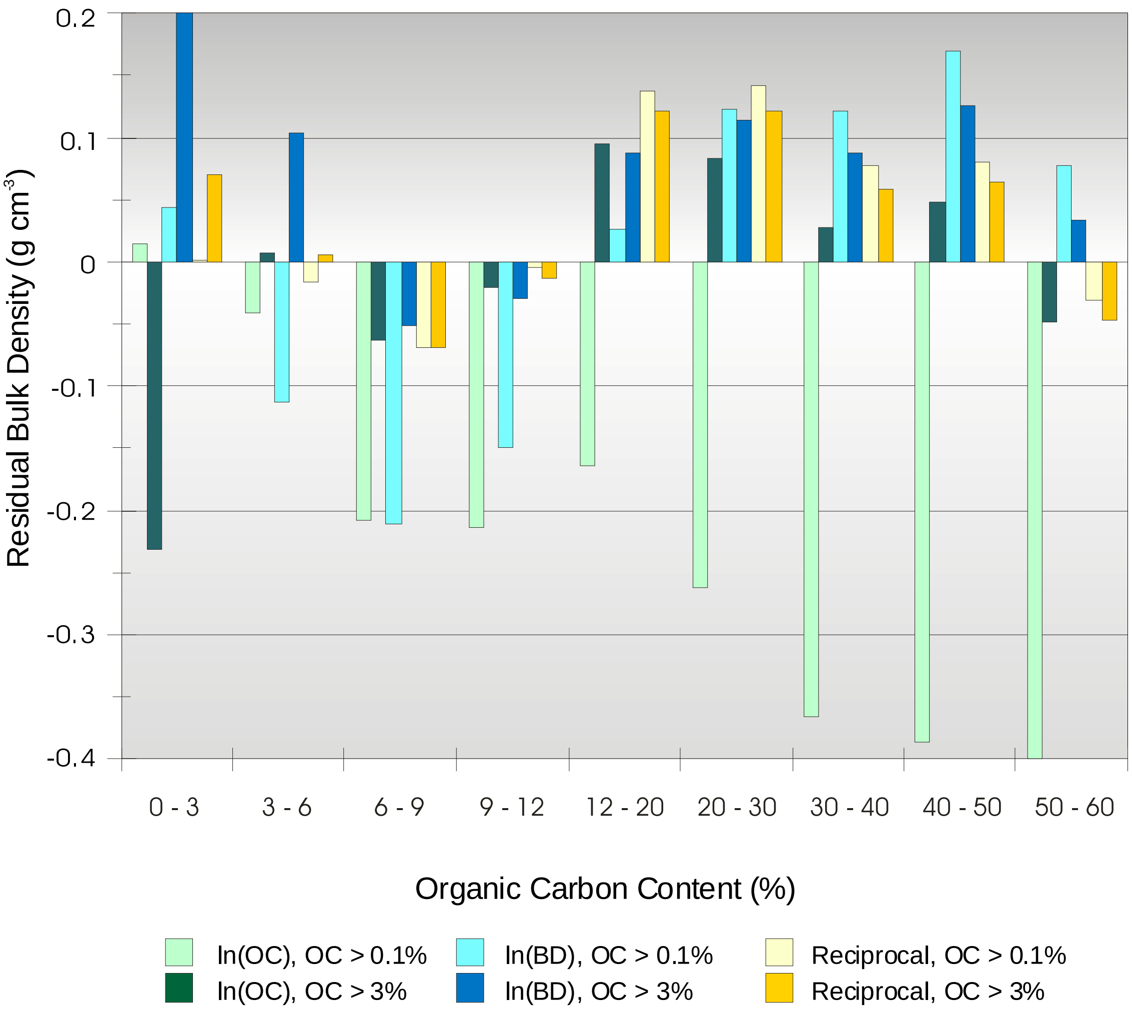


**Figure 11:** **Residuals from Log-transformed OC Content, Bulk Density and Simple Reciprocal Model Applied to Topsoil and Subsoil Data Derived from ISRIC-WISE V3.1 Combined Topsoil and Subsoil Layers for 0.1 and 3% OC Content Thresholds**

The graph shows a common tendency for the models to over-estimate the bulk density for mineral soil with an OC contents from 3 to 12% and to largely under-estimate the bulk density for OC contents for organic soils. Across the range of OC values the log-transformed OC content shows the worst performance when using all layer data with an OC content > 0.1%. However, when limiting the layers included in the analysis to an OC content > 3% the model estimates the bulk density for organic soils with the least deviation. For the other models there is a tendency to improve the fit when limiting the layers included in the analysis from 0.1% to 3%.

Overall, the models do not seem to emulate the relationship between OC content and bulk density very well for organic soils. For estimating the whole range of values the trend in residuals suggests that the profile data should be separated into at least two groups with a threshold value of approx. 12%. To better asses the use of a minimum OC content for the layers used in the definition of the regression model parameters on the fit of the estimated bulk density for organic soils the threshold was increased in steps of 3% for the 3 models. The parameters were determined for the topsoil and subsoil layers separately as well as for the combined layer data.

The results of the variations in the OC threshold used in the parameterization of the regression models and the effect on global SOC stocks are presented in Table 7 to Table 9.

**Table 7: Effect of Organic Carbon Threshold on Bulk Density Estimation and Global SOC Stocks for Regression Parameters of Log-Transformed Organic Carbon Content**

| **Parameter** | **Unit** | **OC Threshold** | | | | |
|---|---|---|---|---|---|---|
| | | *0.1%** | *3.0%* | *6.0%* | *9.0%* | *12.0%* |
| | | *Combined Layers* | | | | |
| Coefficient *a* | - | -0.151 | -0.308 | -0.271 | -0.312 | -0.390 |
| Constant *b* | $g\ cm^{-3}$ | 1.314 | 1.482 | 1.377 | 1.519 | 1.801 |
| Coeff. of Determ. $r^2$ | - | 0.32 | 0.41 | 0.30 | 0.33 | 0.36 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.70 | 0.23 | 0.28 | 0.25 | 0.22 |
| Topsoil SOC Stock | *Pg* | 786 | 653 | 679 | 688 | 691 |
| Subsoil SOC Stock | *Pg* | 1,060 | 740 | 764 | 763 | 766 |
| Global SOC Stock | *Pg* | 1,846 | 1,394 | 1,444 | 1,451 | 1,457 |
| | | *Topsoil* | | | | |
| Coefficient *a* | - | -0.160 | -0.314 | -0.255 | -0.284 | -0.337 |
| Constant *b* | $g\ cm^{-3}$ | 1.339 | 1.517 | 1.349 | 1.449 | 1.634 |
| Coeff. of Determ. $r^2$ | - | 0.34 | 0.41 | 0.25 | 0.26 | 0.28 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.69 | 0.24 | 0.31 | 0.30 | 0.27 |
| Topsoil SOC Stock | *Pg* | 784 | 660 | 689 | 697 | 698 |
| | | *Subsoil* | | | | |
| Coefficient *a* | - | -0.161 | -0.272 | -0.287 | -0.348 | -0.486 |
| Constant *b* | $g\ cm^{-3}$ | 1.280 | 1.335 | 1.386 | 1.611 | 2.126 |
| Coeff. of Determ. $r^2$ | - | 0.30 | 0.37 | 0.39 | 0.44 | 0.49 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.63 | 0.23 | 0.22 | 0.20 | 0.15 |
| Subsoil SOC Stock | *Pg* | 1,003 | 722 | 726 | 732 | 750 |
| Global SOC Stock | *Pg* | 1,787 | 1,382 | 1,415 | 1,429 | 1,448 |

* Parameters from OC > 0.1% threshold applied to HWSD with OC > 3%.

**Table 8: Effect of Organic Carbon Threshold on Bulk Density Estimation and Global SOC Stocks for Regression Parameters of Log-Transformed Bulk Density**

| Parameter | Unit | OC Threshold | | | | |
|---|---|---|---|---|---|---|
| | | *0.1%* | *3.0%* | *6.0%* | *9.0%* | *12.0%* |
| *Combined Layers* | | | | | | |
| Coefficient *a* | - | -0.044 | -0.034 | -0.030 | -0.031 | -0.033 |
| Constant *b* | $g\ cm^{-3}$ | 0.343 | 0.100 | -0.073 | -0.023 | 0.062 |
| Coeff. of Determ. $r^2$ | - | 0.46 | 0.46 | 0.39 | 0.42 | 0.39 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.11 | 0.15 | 0.17 | 0.17 | 0.16 |
| Topsoil SOC Stock | *Pg* | 634 | 630 | 652 | 663 | 666 |
| Subsoil SOC Stock | *Pg* | 666 | 691 | 697 | 701 | 706 |
| Global SOC Stock | *Pg* | 1,301 | 1,321 | 1,349 | 1,364 | 1,372 |
| *Topsoil* | | | | | | |
| Coefficient *a* | - | -0.043 | -0.033 | -0.027 | -0.027 | -0.027 |
| Constant *b* | $g\ cm^{-3}$ | 0.337 | 0.129 | -0.082 | -0.077 | -0.042 |
| Coeff. of Determ. $r^2$ | - | 0.46 | 0.42 | 0.32 | 0.33 | 0.31 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.12 | 0.17 | 0.20 | 0.20 | 0.20 |
| Topsoil SOC Stock | *Pg* | 640 | 641 | 663 | 673 | 674 |
| *Subsoil* | | | | | | |
| Coefficient *a* | - | -0.047 | -0.034 | -0.032 | -0.035 | -0.039 |
| Constant *b* | $g\ cm^{-3}$ | 0.349 | -0.028 | -0.113 | 0.026 | 0.199 |
| Coeff. of Determ. $r^2$ | - | 0.46 | 0.49 | 0.45 | 0.51 | 0.48 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.09 | 0.13 | 0.14 | 0.13 | 0.13 |
| Subsoil SOC Stock | *Pg* | 649 | 660 | 664 | 673 | 684 |
| Global SOC Stock | *Pg* | 1,290 | 1,301 | 1,327 | 1,346 | 1,359 |

* Parameters from OC > 0.1% threshold applied to HWSD with OC > 3%.

**Table 9:** **Effect of Organic Carbon Threshold on Bulk Density Estimation and Global SOC Stocks for Regression Parameters of Reciprocal Model**

| Parameter | Unit | OC Threshold | | | | |
|---|---|---|---|---|---|---|
| | | *0.1%* | *3.0%* | *6.0%* | *9.0%* | *12.0%* |
| | | *Combined Layers* | | | | |
| Coefficient *a* | - | 0.066 | 0.060 | 0.042 | 0.045 | 0.056 |
| Constant *b* | $g\ cm^{-3}$ | 0.661 | 0.702 | 0.900 | 0.817 | 0.517 |
| Coeff. of Determ. $r^2$ | - | 0.39 | 0.40 | 0.30 | 0.33 | 0.35 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.22 | 0.24 | 0.30 | 0.29 | 0.27 |
| Topsoil SOC Stock | *Pg* | 637 | 642 | 682 | 691 | 687 |
| Subsoil SOC Stock | *Pg* | 701 | 715 | 773 | 772 | 759 |
| Global SOC Stock | *Pg* | 1,338 | 1,358 | 1,455 | 1,463 | 1,447 |
| | | *Topsoil* | | | | |
| Coefficient *a* | - | 0.061 | 0.060 | 0.037 | 0.038 | 0.044 |
| Constant *b* | $g\ cm^{-3}$ | 0.659 | 0.680 | 0.928 | 0.890 | 0.727 |
| Coeff. of Determ. $r^2$ | - | 0.42 | 0.40 | 0.25 | 0.26 | 0.28 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.24 | 0.24 | 0.32 | 0.32 | 0.30 |
| Topsoil SOC Stock | *Pg* | 647 | 646 | 692 | 700 | 698 |
| | | *Subsoil* | | | | |
| Coefficient *a* | - | 0.089 | 0.056 | 0.052 | 0.057 | 0.080 |
| Constant *b* | $g\ cm^{-3}$ | 0.654 | 0.828 | 0.860 | 0.688 | 0.041 |
| Coeff. of Determ. $r^2$ | - | 0.35 | 0.37 | 0.38 | 0.42 | 0.49 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.17 | 0.25 | 0.26 | 0.25 | 0.21 |
| Subsoil SOC Stock | *Pg* | 642 | 717 | 734 | 735 | 718 |
| Global SOC Stock | *Pg* | 1,289 | 1,363 | 1,425 | 1,435 | 1,415 |

* Parameters from OC > 0.1% threshold applied to HWSD with OC > 3%.

The SOC stocks in Table 7 to Table 9 were estimated for the 30 *arc second* grid size and integrating all typological data associated with a mapping unit. The method of defining the parameters for the OC vs. bulk density relationship from layer data with an OC > 0.1% and applying the relationship to layers with an OC > 3% leads to the highest global SOC estimates (1,846 *Pg*). This is caused by improbably high values of bulk density for organic soils. Apart from this configuration there is a general trend for SOC stocks to increase with higher thresholds for the OC content used in the parameterization of the models from 1,321 to 1,457 *Pg* (for parameterization of combined topsoil and subsoil layers).

For a given model the shift in the OC threshold from 3 to 6, 9 or 12% for applying the estimates of bulk density lead to differences between 51 to 105 *Pg*. The results obtained from the log-transformed OC content and the reciprocal model are generally close compared to the results obtained from the log-transformed bulk density. Estimating the global SOC stock from parameterizing the models separately for the topsoil and the subsoil layer results in estimates which are between 5 *Pg* higher and 32 *Pg* lower than the estimates obtained when using the combined layer data.

The changes in the residual means for thresholds of 6 and 12% OC content are shown in Figure 12.

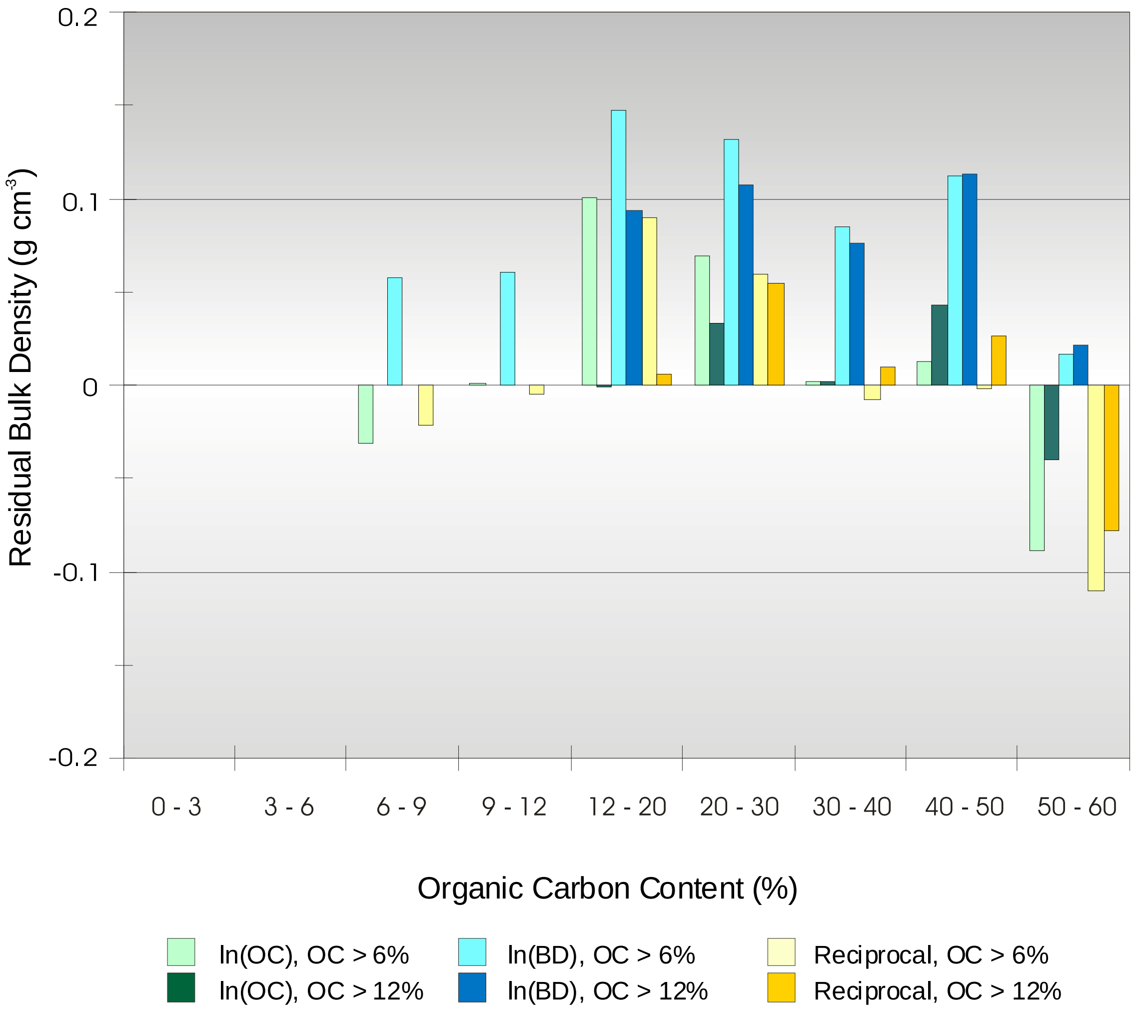


**Figure 12: Residuals from Log-transformed OC Content, Bulk Density and Simple Reciprocal Model Applied to Topsoil and Subsoil Data Derived from ISRIC-WISE V3.1 Combined Topsoil and Subsoil Layers for 6 and 12% OC Content Thresholds**

The residual means indicate a tendency for lower residuals of the log-transformed OC content and the reciprocal model than the log-transformed bulk density for most of the range of OC content values, in particular for the important range of 30-40% OC content. However, the latter provides better estimates for OM (OC = 58%), for which the other two models overestimate by 0.05 to 0.1 $g\ cm^{-3}$.

With bulk densities from 0.15 to 0.17 $g\ cm^{-3}$ the estimates correspond better to values found in the literature. Chambers, *et al*. (2010) found ash-free bulk densities from 4,697 samples to range from <0.01 to 0.23 $g\ cm^{-3}$ with a median of 0.1 $g\ cm^{-3}$. Their literature review concludes that the bulk density for peat is typically 0.05 to 0.2 $g\ cm^{-3}$ not only for peat in high latitudes, but also for peat of tropical areas. Chambers, *et al*. (2010) also conclude that for 110 samples of the West Siberian Lowlands the OC content of the peat OM varies between 0.50 and 0.58 $g\ OC\ g\ OM^{-1}$, depending on the botanical composition of the material, with peat formed by *Sphagum* having decidedly and consistently lower values. It may thus be argued that the reciprocal model derived from the profile data leads to over-estimating SOC stocks in peat of higher latitudes when OM determined by from loss on ignition is converted to OC content using a commonly applied factor of 0.58 $g\ OC\ g^{-1}\ OM$. For the ISRIC-WISE profile data the method used to assess bulk density and the conversion factor between OC and OM content is not evident since the horizon data contains values for OC content up to 96.1%.

With respect to identifying the most suitable method for estimating bulk density for the HWSD for mineral soils high in OC and organic soils it should be considered that the global ISRIC-WISE profile data set contains considerable variations in the relationship of OC vs. bulk density, which results in correspondingly elevated uncertainties in the estimates. There would be no significant difference in the estimates if a simple piece-wise linear regression would be used with the range separated at 12% OC content.

In the subsequent analysis of this study the bulk density was estimated only for layers with an OC content >12% instead of the 3% indicated as the valid range of the model used in the HWSD. The higher threshold was applied to minimize any changes to the data of the HWSD and still provide a reasonable estimate of bulk density for soils high in OC. With the data for organic soil mainly within the range of 30-40% OC content the log-transformed OC content shows the least deviations from the ISRIC-WISE profile data for this range. The consistent difference in parameters when treating topsoil and subsoil layers separately also suggests to use discrete regressions for the layers.

An alternative view of the data and a possible method to reduce the influence of the uneven distribution of profiles in the data set for OC content is to apply a weight to the data in the regression analysis. One procedure is to compute the mean OC content and bulk density for ranges of OC content. The distribution of the means and the residuals from the three regression models applied to the combined topsoil and subsoil layers are presented in Figure 13.

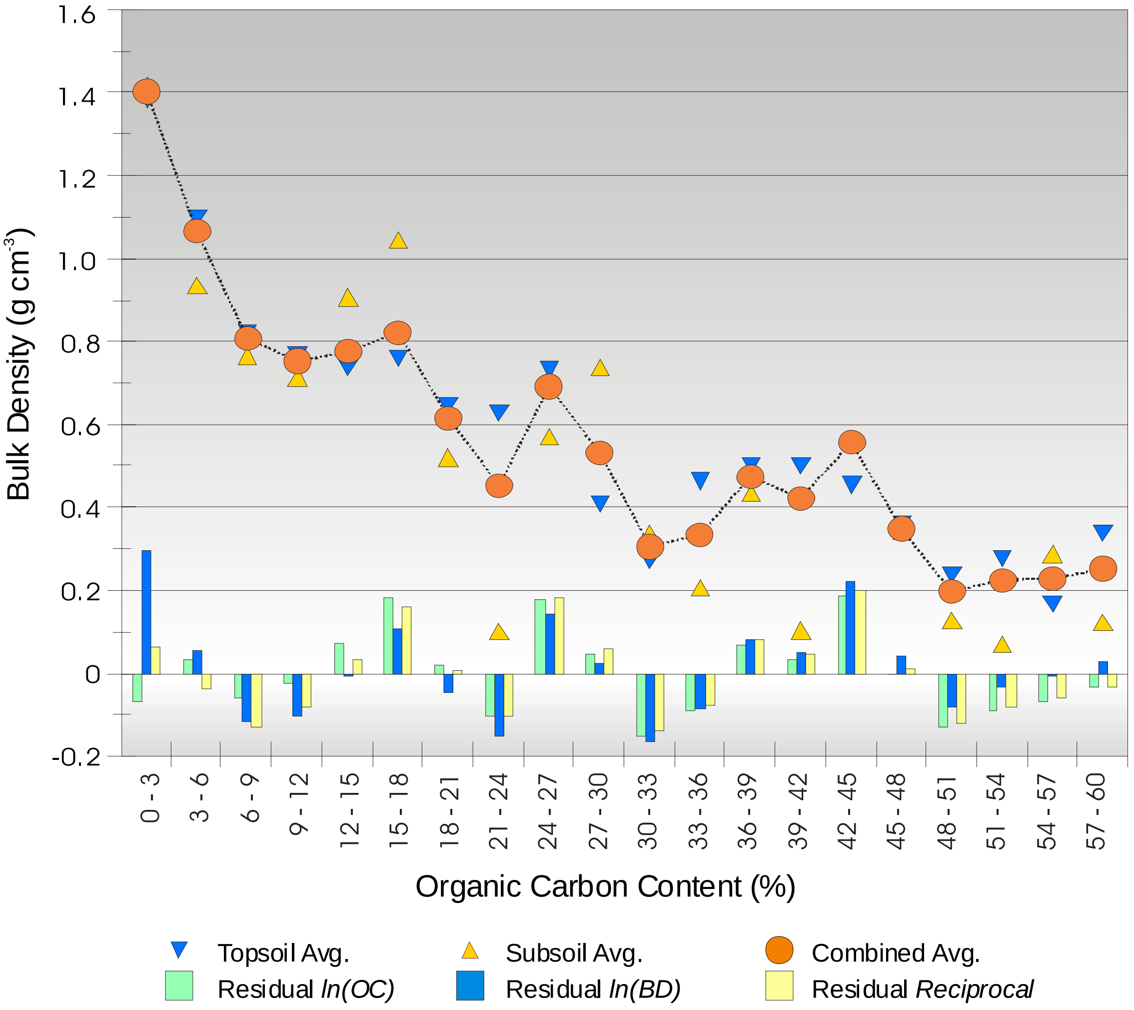


**Figure 13: Relationship between Mean Organic Carbon Content (%) and Mean Bulk Density for ISRIC-WISE V3.1 Profiles for Topsoil (0-30cm) and Subsoil (30-100cm) Layers (3% OC intervals) and Regression Model Residuals**

The first class of averaged OC content and related bulk densities (OC 0 – 3%) contains 90% of all topsoil and subsoil layers used in the analysis. This contrasts with 1.6% of the layers with OC >12%. The variation in the OC vs. bulk density relationship for organic soils is not least due to the low number of profiles with data in each class. For the majority of classes the number of the combined topsoil and subsoil layers is 5 or less. The variation of the residuals graph does not recommend one regression model to be more suitable to estimate bulk density from OC content than any other. Only for layers with an OC content < 3% are the estimates from the log-transformed bulk density notably lower than the estimates from the other two models.

The regression parameters derived from the mean class values of OC content and corresponding bulk densities from all layers with OC content > 0.1% for the regression models and the effect of substituting bulk densities for layers with OC content > 12% on the global SOC stock estimates are given in Table 10.

**Table 10: Estimation of Global SOC Stocks for Regression Models Parameterized from Mean Class Values with 3% OC Intervals and Applied to Layer with OC>12%**

| Parameter | Unit | Regression Model | | |
|---|---|---|---|---|
| | | *Log(OC)* | *Log(BD)* | *Reciprocal* |
| | | *Combined Layers* | | |
| Coefficient *a* | - | -0.281 | -0.028 | 0.049 |
| Constant *b* | $g\ cm^{-3}$ | 1.424 | 0.125 | 0.707 |
| Coeff. of Determ. $r^2$ | - | 0.89 | 0.83 | 0.89 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.29 | 0.22 | 0.28 |
| Topsoil SOC Stock | *Pg* | 693 | 694 | 689 |
| Subsoil SOC Stock | *Pg* | 777 | 770 | 765 |
| Global SOC Stock | *Pg* | 1,471 | 1,464 | 1,454 |
| | | *Topsoil* | | |
| Coefficient *a* | - | -0.285 | -0.027 | 0.046 |
| Constant *b* | $g\ cm^{-3}$ | 1.457 | 0.120 | 0.711 |
| Coeff. of Determ. $r^2$ | - | 0.90 | 0.80 | 0.89 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.30 | 0.24 | 0.30 |
| Topsoil SOC Stock | *Pg* | 699 | 701 | 695 |
| | | *Subsoil* | | |
| Coefficient *a* | - | -0.291 | -0.039 | 0.062 |
| Constant *b* | $g\ cm^{-3}$ | 1.389 | 0.127 | 0.687 |
| Coeff. of Determ. $r^2$ | - | 0.73 | 0.60 | 0.73 |
| $\rho_b$(OM) | $g\ cm^{-3}$ | 0.21 | 0.12 | 0.23 |
| Subsoil SOC Stock | *Pg* | 718 | 664 | 717 |
| Global SOC Stock | *Pg* | 1,417 | 1,365 | 1,412 |

There are hardly any differences between the regression models for global SOC stock estimates when applying a single regression to the topsoil and subsoil layers. More variability is found in the subsoil layer when defining the parameters separately for the soil layers. Using the log-transformed bulk density model the estimates of global SOC stocks are almost 100 *Pg* lower for the separate layer regressions than for the combined regression. This development was not found when analyzing the individual layer data, although the global SOC stock estimated from the separate regression of the means by soil layer are close to the estimates obtained from the regression of individual layers for both, the regression parameters of the combined layers and the parameters derived from separate layers.

For the specific conditions of the distribution of OC contents in the HWSD and the results from the analysis of individual profile layers (for parameters see Table 7, Table 8 and Table 9) the regression model selected to substitute unrealistic bulk densities in the database was the log-transformed mean OC content as defined by:

$$\rho_s(t) = -0.285 \times \ln(OC_{>12}) + 1.457 \;\; g\ cm^{-3} \text{ and}$$

$$\rho_s(s) = -0.291 \times \ln(OC_{>12}) + 1.389 \;\; g\ cm^{-3}$$

where

| | |
|---|---|
| $\rho_s(t)$ | topsoil bulk density ($g\ cm^{-3}$) |
| $\rho_s(s)$ | subsoil bulk density ($g\ cm^{-3}$) |
| $OC_{>12}$ | organic carbon content (%) of layer with OC>12% |

The effect of the grid size and the method used to aggregate the SOC stock parameters were assessed for the three model options when substituting bulk densities for layers with an OC content > 12%.

For reasons of spatial compatibility with other global data sets, such as the FAO data on land degradation assessment in drylands (LADA) data, the 30 *arc second.* HWSD raster was reduced by a factor of 10 to a raster size of 5 *arc minute*. This size corresponds to approximately 81 km$^2$ at the Equator. The grid size was reduced by thinning the original raster layer by a factor of 10 in x- and y-direction, i.e. taking every 10th pixel in both directions. This not only reduces the size of the layer by a factor of 100, but also allows using 16-bit integer values for the layer dimension. The aggregation of data from the multi-link between mapping and typological units of the HWSD distinguishes between integrating the information from all related units and the use of only the dominant typological unit. The results of the various processing options on global soil carbon stocks to 1m are summarized in Table 11.

**Table 11: Selection of Bulk Density Model and Data Processing Options on Global SOC Stocks using Modelled Bulk Density for Layers with OC > 12%**

| Model | Soil Layer | Grid Size | | | |
|---|---|---|---|---|---|
| | | 30 arc second | | 5 arc minute | |
| | | Dominant | Integration | Dominant | Integration |
| | | Pg | Pg | Pg | Pg |
| None | Topsoil | 902 | 967 | 900 | 965 |
| | Subsoil | 1,313 | 1,502 | 1,309 | 1,498 |
| | **Combined** | **2,215** | **2,469** | **2,209** | **2,463** |
| Log(OC) | Topsoil | 676 | 699 | 675 | 698 |
| | Subsoil | 664 | 718 | 662 | 716 |
| | **Combined** | **1,340** | **1,417** | **1,337** | **1,414** |
| Log(BD) | Topsoil | 677 | 701 | 675 | 699 |
| | Subsoil | 620 | 664 | 618 | 663 |
| | **Combined** | **1,296** | **1,365** | **1,297** | **1,362** |
| Reciprocal | Topsoil | 672 | 695 | 671 | 694 |
| | Subsoil | 663 | 717 | 661 | 715 |
| | **Combined** | **1,335** | **1,412** | **1,332** | **1,409** |

For the original data the reduction in grid spacing by a factor of 10 to 5 *arc minutes* on the global SOC stocks resulted in a topsoil SOC stock of 965 *Pg* C and a subsoil stock of 1,498 *Pg* C, giving a total of 2,463 *Pg* C. This amounts to a difference of 6.3 *Pg* C or 0.26% of the global SOC stock computed for the 30 *arc second* raster. For the amended data reducing the grid spacing results in lower global SOC stocks by 1 - 4 *Pg*. For the change in spatial resolution no discernible consistent difference between the topsoil and the subsoil layer response to the change in spatial resolution was found. The difference is considered inconsequential when assessing global SOC stocks.

When using only the information from the dominant typological unit the global SOC stocks are lower by 254 *Pg* C for the original data and on average 77 *Pg* for the amended data as compared to the integration of the information from all typological units. This difference is comparable to the variations in global OC stock following the use of alternative OC thresholds and models. 70% of this difference is due to lower subsoil OC stock. This tendency is found irrespective of the spatial resolution and the regression model used to estimate bulk density for higher OC contents. The cause of the lower stock values when using only the dominant typological unit is attributed to the tendency of having organic soils as sub-dominant components in mapping units with more than one typological unit. As a consequence, the selection of the dominant typological unit introduces a bias in the distribution of soil types against organic soils and lower SCO stock estimates.

The spatial distribution of the differences in the bulk density values between the original HWSD and the amended data for the topsoil and subsoil layers are presented in Figure 14.

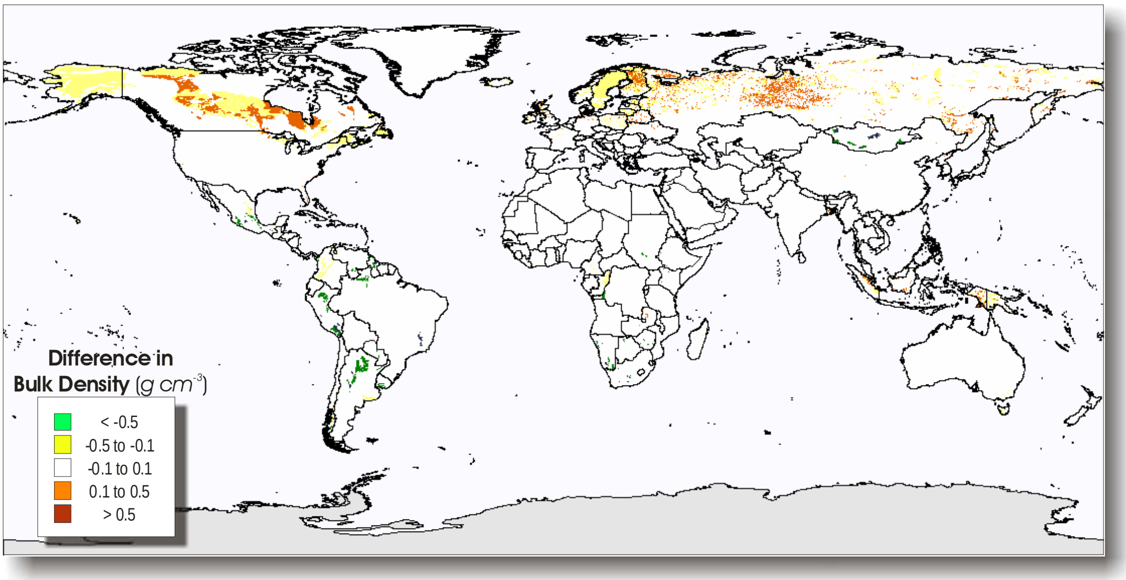


**Figure 14: Difference in Bulk Density between HWSD and Amended Soil Profile Data for Soil Layer 0 – 100cm**

The map shows the bulk density of the combined topsoil and subsoil layer where the value for the mapping unit is derived from integrating the values of all associated typological units. A negative difference in the map indicates areas where the bulk density estimated from the amended soil profile data is higher than the value given in the HWSD. The difference mainly arises from providing estimates in the amended data for values missing in the horizon data of the HWSD and treating blank data as missing, i.e. not as zero. Such cases are found in some areas in South America and Mongolia. Positive differences are areas where the HWSD bulk density is higher than the value estimated by the function. These areas are largely in the northern hemisphere, but concern also tropical peat in South-East Asia.

## 4.4 Depth

The depth of a soil is given as either 10cm, 30cm or 100m. Where more than one typological unit is associated with a MU and when integrating the information the average depth of the soil in the mapping unit can take on intermediate values. This average soil depth for the mapping units is given in Figure 15.

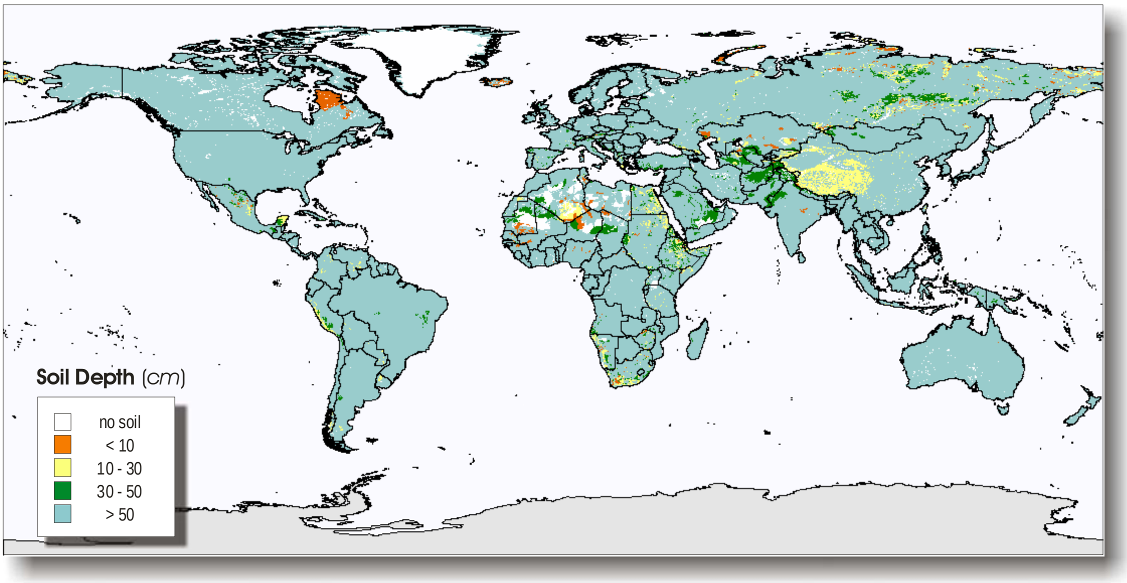


**Figure 15:** **Average Depth of Soil in Mapping Unit (cm) Adjusted to Total Mapping Unit Area**

The average depth is computed from all typological units with data and refers to the total area of the mapping unit, i.e. including non-soil areas. The layer can therefore be used directly for overlay computations in a GIS with the other spatial layers, without further adjustments for the grid cell area.

## 4.5 SOC Density and Stock

The global SOC stocks for the HWSD using the 30 *arc second* grid without any amendments is calculated at 2,469.5 *Gt*. The topsoil layer (0-10 or 0-30 cm) SOC stock comes to 967.3 *Gt* C and the subsoil layer to 1,502.2 *Gt* C. The subsoil layer thus contains about 1.5 times the amount of OC of the topsoil. This ratio is higher than the shares given by other studies ((Batjes, 1996; FAO, 2001; Jobbagy & Jackson, 2000). It is not sufficiently explained by a reduced number of areas without subsoil OC, since the proportion of MUs without subsoil OC in the database is 3.8% of all MUs. However, the area represented by these MUs without subsoil is 13.2% of the total area.

A map of the SOC stocks for the combined topsoil and subsoil layers as computed from the HWSD data is presented in Figure 16.

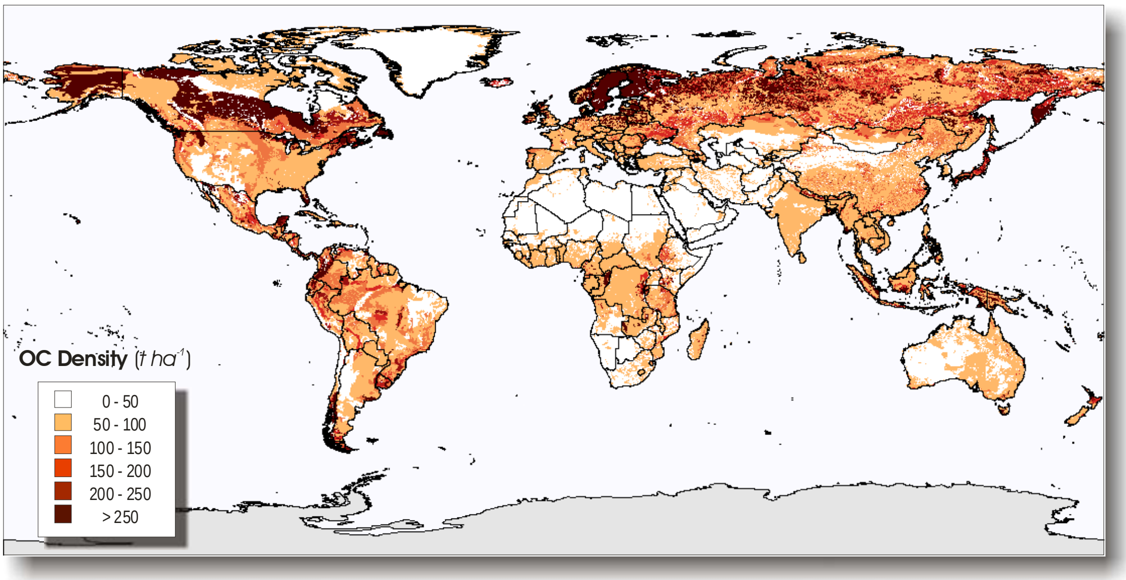


**Figure 16: Soil Organic Carbon Density (*t ha*$^{-1}$) for Combined Topsoil and Subsoil Layer (0 – 100cm) from HWSD V1.1**

The map shows the combined area-weighted OC density for the soil and the non-soil areas within a mapping unit. This allows computing OC stocks by simply multiplying the density with the area, but means that the density values relate to the whole mapping area, not only the soil component. Therefore, the SOC density in the actual soil portion of a mapping unit with both, soil and non-soil components, is higher than the figure computed for the total mapping unit although the amount of OC for the area is identical.

Using the amended typological data and a bulk density estimated by the regression model derived from the global ISRIC-WISE profile datasets the global SOC stocks are calculated as 1,417 *Pg* C for a spatial layer of 30 *arc second* and 1,414 *Pg* C for the 5 *arc minute* layer. These figures are lower than the widely quoted 1,500 *Pg* C (Batjes, 1996), although they remain within the range of 1,115 – 2,200 *Pg* C given for the uncertainty of 1 standard deviation from the mean by Batjes (1992).

The distribution of the OC density in the combined topsoil and subsoil layers from the completed and modified data is presented in Figure 17.

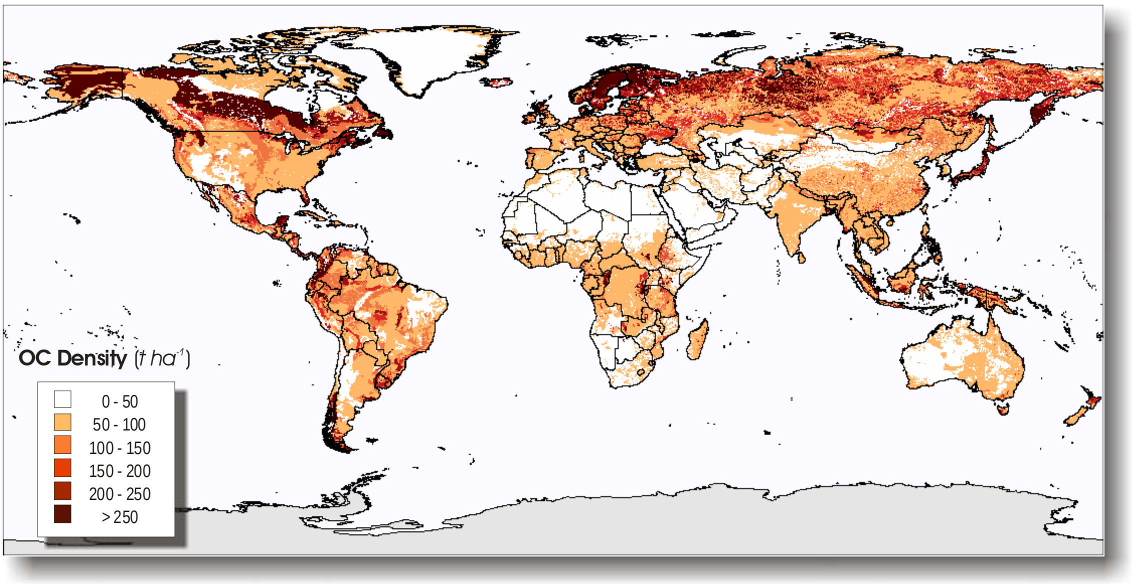


**Figure 17: Soil Organic Carbon Density (*t ha*$^{-1}$) for Combined Topsoil and Subsoil Layer from Amended HWSD**

To appreciate the result of complementing and substituting the data the difference in SOC stock between the processed data and the original data with missing parameters is given in Figure 18.

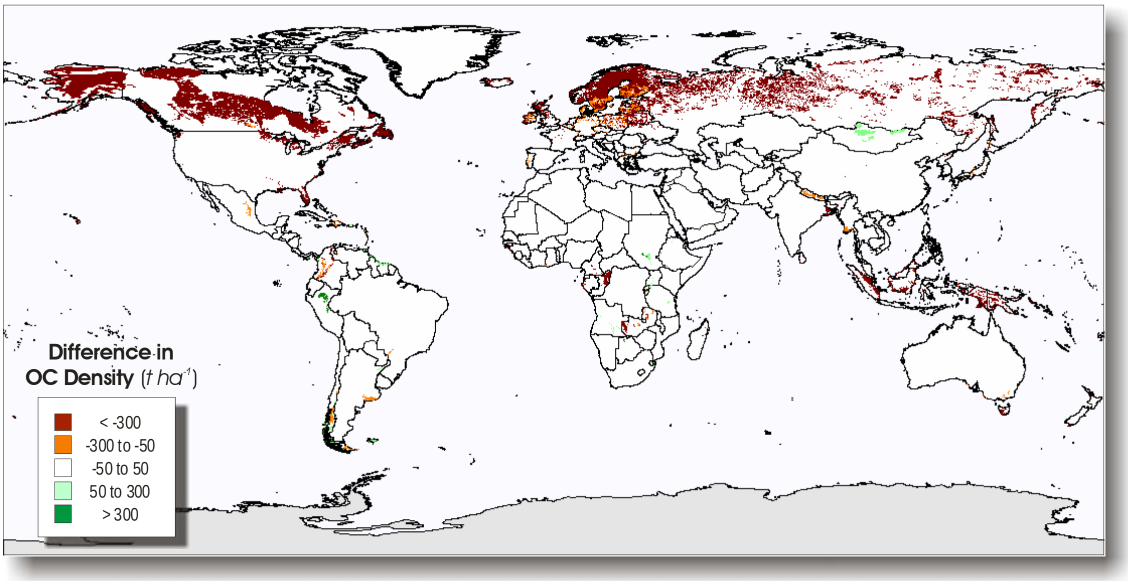


**Figure 18:** **Difference in SOC Density between HWSD and Amended Parameters with Modification to Bulk Density (*t ha*$^{-1}$)**

The map shows that the effect of complementing missing data and adjusting the SOC content for organic soils on the distribution of SOC density mainly affects areas in the northern hemisphere. An increase in SOC density of the modified data is largely restricted to areas in Peru, Sudan, Angola and Mongolia. Changes in these areas are the consequence of completing the typological units with values, for which then a SOC density can be computed. Compared to the SOC stock changes introduced by the adjustment of bulk density for organic soils completing the typological unit data has a very small effect on the global estimates.

## 4.6 SOC Stocks and Bulk Density Model

The evaluation of the relationship between OC content and bulk density for soils high in organic carbon based on global data did not result in the identification of one model as being significantly different from a liner relationship. When estimating SOC density and subsequently stocks from modelled bulk densities the influence of the model used to describe the relationship between OC content and bulk density should be taken into account. Because of the definition of SOC density an increase in OC content does not necessarily result in an increase in the amount of OC in a given volume of soil. A linear relationship between OC content and bulk density for organic soils leads to a quadratic relationship between OC content and density.

The changes in OC density with OC content for the three models evaluated and a linear relationship are presented in Figure 19.

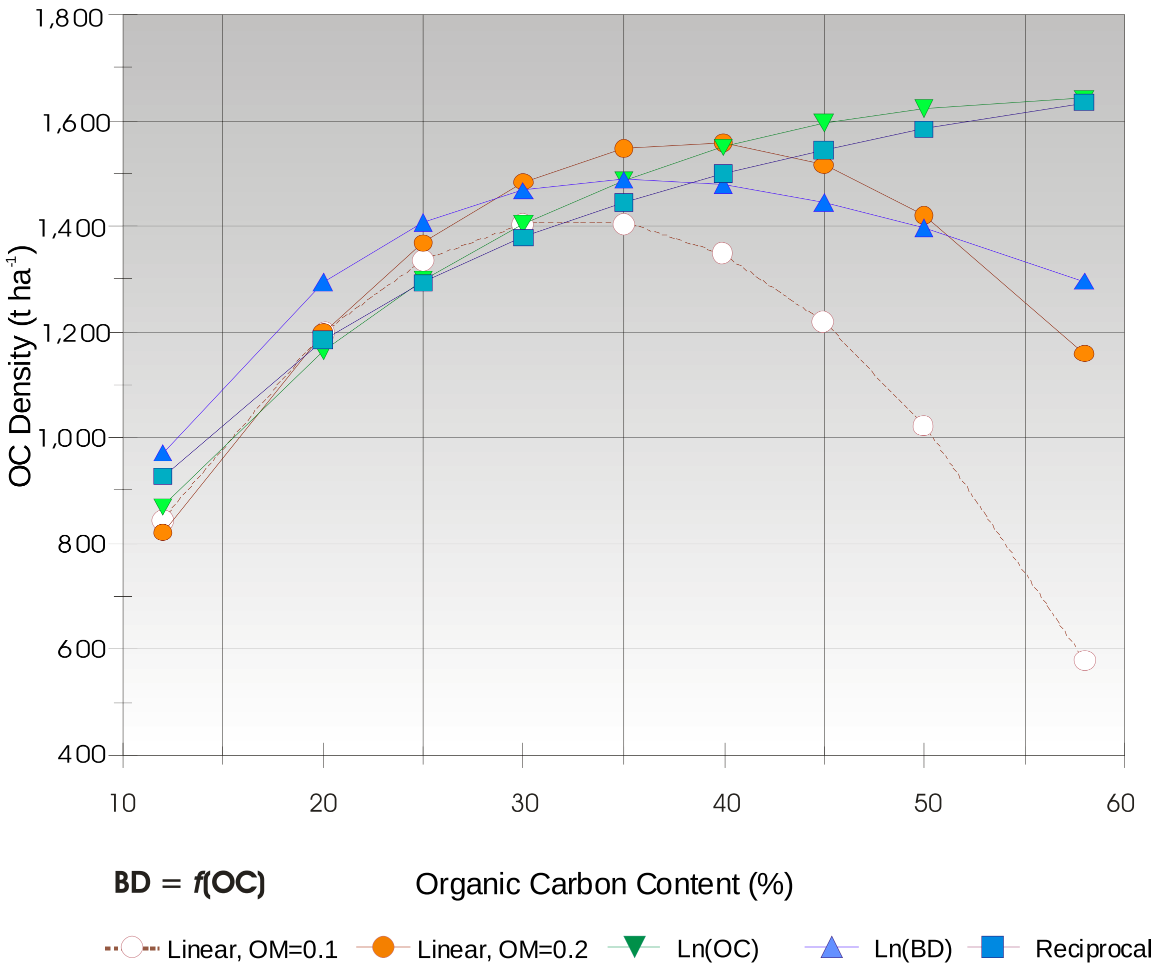


**Figure 19:** **Changes in SOC Density with Content for Different Bulk Density Models for Organic Soils (fixed depth of 100 cm)**

The parameters for the bulk density model are taken from Figure 10 for the combined topsoil and subsoil layer. Also included is a linear relationship between OC content and bulk density with OC contents for OM of 0.1 and 0.2 $g\ cm^{-3}$. The lower value is quoted in the literature and corresponds to the value estimated from SPADE/M while the higher value is closer to the estimate derived from the global soil profile data. SOC densities are given for a soil free of coarse fragments and a depth of 100cm.

The graph shows that an increase in OC content may not necessarily lead to an increase in the amount of OC stock. For a linear relationship the SOC density peaks at about 35% OC content. For a bulk density of 0.1 $g\ cm^{-3}$ the SOC density for OM (58% OC content) is half of the SOC density at 20% OC content and a bulk density of 0.6 $g\ cm^{-3}$. Comparable relationships are also found the log-transformed bulk density, but within the range of values not for the log-transformed OC content and reciprocal models.

This counterintuitive relationship of an increase in OC content leading to a decrease in OC stock is a result of defining OC density for a fixed depth from the surface and is prevalent for organic soils. When bulk density decreases to total volume (depth) of soil increases and, consequently, the amount of SOC in a fixed value decreases. Simply put, for a fixed soil depth layer one is looking at less soil material when bulk density decreases, which, depending on the relationship between OC content and bulk density, can lead to lower estimates for SOC stocks. A linear distribution of the clusters of OC content in soil data may result in a decrease in SOC stocks within the soil layers of fixed depth. It is therefore rather complex when trying to define an unbiased mean for the OC content of organic soils. To account for the changes in soil material with varying bulk density when using a fixed thickness a method using equivalent soil mass could be applied (Ellert & Bettany, 1995)..

## *4.7 Comparison of HWSD SOC with other Global Data*

There are several spatial data sets with global estimates of SOC content or stock available from various sources. For global spatial layers on soil organic carbon two principal sources of information were evaluated:

- Global Soil Organic Carbon map[3] from *Natural Resources Conservation Service* (NRCS) of the Unites States Department of Agriculture;
- FAO Organic Carbon Pool;
- WISE5by5MIN;
- Digital Soil Map of the World (Ver. 3.6)

Most of the data sets use a much coarser grid size, mainly 5 *arc minute* resolution. To compare the estimates derived from the amended HWSD with spatial layers of SOC content or stock from other sources all spatial data were processed to a standard spatial frame with grid spacing of 5 *arc minute* to allow the identification of areas where differences occur.

### 4.7.1 Standard Spatial Layers Properties

The spatial data layers with global coverage use a common raster format with standardized characteristics. All layers use a regular grid size of 5 *arc minute*. This grid spacing corresponds to approx. 10 km at the equator. The data are arranged in geographic co-ordinates using the *European Terrestrial Reference System 1989* (ETRS89). With a difference of less than 1m to the *World Geodetic System 1984* (WGS84) (shift of approx. 2.5 $cm\ year^{-1}$ due to continental drift with parity at epoch

---

[3] http://soils.usda.gov/use/worldsoils/mapindex/soc.html
Data received from Paul Reich, World Soil Resources, USDA in May, 2009.

1989) the difference to WGS84 can be neglected for the purposes of this study and assumed to be sufficiently similar to the *International Terrestrial Reference System* (ITRS), which is stipulated to be used for areas outside the geographical scope of ETRS89[4].

The specifications of the spatial frame for the data layers are given in Table 12.

**Table 12: Specifications of Spatial Data Layers**

| Feature | Value |
|---|---|
| Data type | 16-bit integer or real* |
| File type | binary |
| No. of columns | 4320 |
| No of rows | 2160 |
| Reference system | ETRS89 |
| Reference units | degree |
| Min. X co-ordinate | -180.0000 |
| Max. X co-ordinate | 180.0000 |
| Min. Y co-ordinate | -90.0000 |
| Max. Y co-ordinate | 90.0000 |

* Depending on the information to be stored data formats are either 16-bit integer or floating-point (real). The range limits of integer values is -32768 and +32767, while it is 1.0E-38 to 1.0E+38 with 7 significant digital for data in floating-point format (IEEE 754 single; Goldberg, 1991).

To avoid arbitrary results in coastal areas all spatial layers were adjusted to a standard land/sea mask. The mask was generated from the global *Geographical Information System at the Commission* (GISCO) country coverage at scale 1:1mio. (GISCO.CNTR_RG_01M_2006) Coastal areas in the thematic data layers were revised by using a distance function to allocate layer attributes to the common mask.

### 4.7.2 Re-Scaling of Layer Geometry

The global map product is available in form of a single file. Each grid location contains the value corresponding to one of the legend classes. When reducing the spatial resolution of the data several options of aggregating data are available:

a) Class of central pixel -> single data layer

---

[4] OJ L 323, 8.12.2010, p. 11. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0011:0102:EN:PDF

b) Class of majority -> single data layer
c) Proportional distribution -> multiple data layers

The first two methods lead to a single data layer to represent the original data. Those methods are relatively rapid to generate and provide considerable savings in data storage. The disadvantage of the methods is that the resulting data is biased against classes with a low portion of occurrence. The third method maintains the information on the proportional distribution of the classes and thus also includes those with a low occurrence, but has higher requirements for processing and data storage.

### 4.7.3 Natural Resources Conservation Service Global Soil Organic Carbon

The *Natural Resources Conservation Service* world soil resources map on soil organic carbon (NRSC SOC) map was produced based on the FAO-UNESCO Soil Map of the World and a climate map. The original map was produced in 1997, based on the FAO-UNESCO Soil Map of the World. The data layer was revised in 2000, which is the version used in this project. SOC density to a depth of 1 m is presented in 9 classes of “soil organic carbon volume” using $kg\ m^{-2}$ as the reporting unit. The original data were reformatted from the Miller projection and 2 *arc minute* resolution to the common layer specifications of 5 *arc minute* by retaining the central pixel.

The distribution of the SOC density values across the global map reprocessed to the standard spatial specification is presented in Figure 20.

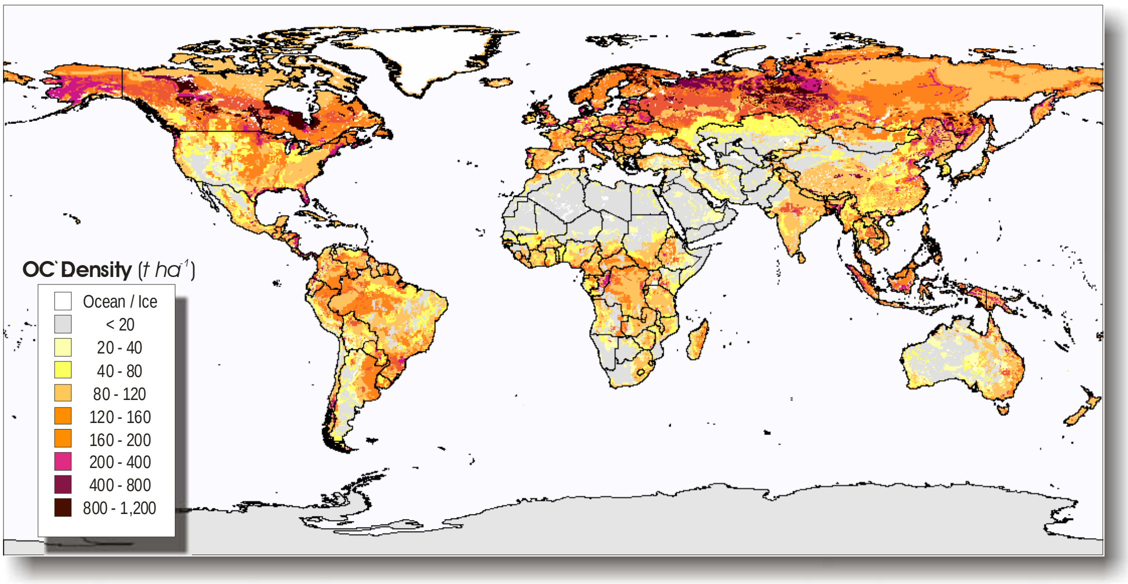


**Figure 20:** **Map of SOC Density from Natural Resources Conservation Service (NRCS) of the Unites States Department of Agriculture (revised version from year 2000; resampled to 5 *arc minute*)**

In the preparation of the map the original SOC data were modified by a soil climate map. The influence of the ancillary data on SOC densities is clearly visible in the low values prevalent in dry and warm areas and the concentration of SOC in areas of high precipitation rates and/or low temperatures. A detailed description of the method applied for producing the NRCS data was not available and queries should be addressed to the guardian of the data (paul.reich@wdc.usda.gov).

Estimating global SOC stocks from class values is a matter of how the class range values are converted into specific quantities. A simple and readily applicable procedure is to use the mean of the class range. This approach resulted in a global SOC stock of 1,376 *Pg*.

### 4.7.4 Comparing HWSD with NRCS Global Soil Carbon Layers

Since NRCS data were only available as classes the global SOC stocks were estimated from the mean value of the class ranges. This procedure can introduce bias in the estimates and for reasons of comparability the amended HWSD layers were re-classified following the ranges and mean values used for the NRCS data HWSDa$^{classified}$).

The results obtained from global SOC stocks by the two standardized data sets are:

| | |
|---|---|
| NRSC | 1,399 *Pg* C 0-100 cm |

$\text{HWSDa}^{\text{classified}}$ 1,392 *Pg* C, topsoil and subsoil

The global SOC stock estimated from the classified amended HWSD includes 3 *Pg* C found outside the land/sea mask of the standard data. For the NRSC the SOC stock outside the land/sea mask was 0.4 *Pg* C. The difference introduced by using the central class value instead of the continuous range of values for the SOC density to the global SOC stocks from the HWSD is thus 25 *Pg* C.

For the NRCS SOC map the source data from the FAO-UNESCO Soil Map of the World was modified by soil climate conditions. The effect of including soil climate in the computations is a deviation from the default values of the input data, which can lead to an increase or decrease in local SOC density. As a consequence, there are spatially variable differences between the NRCS and the HWSD, although the overall estimates are similar. A comparison of the relative differences to the HWSD data ($[\text{HWSD}_{\text{Class\_Mean}}\text{-NRCS}_{\text{Class\_Mean}}]/\text{HWSD}_{\text{Class\_Mean}}*100$) shows the influence of the soil climate data on the SOC stocks, as given in Figure 21.

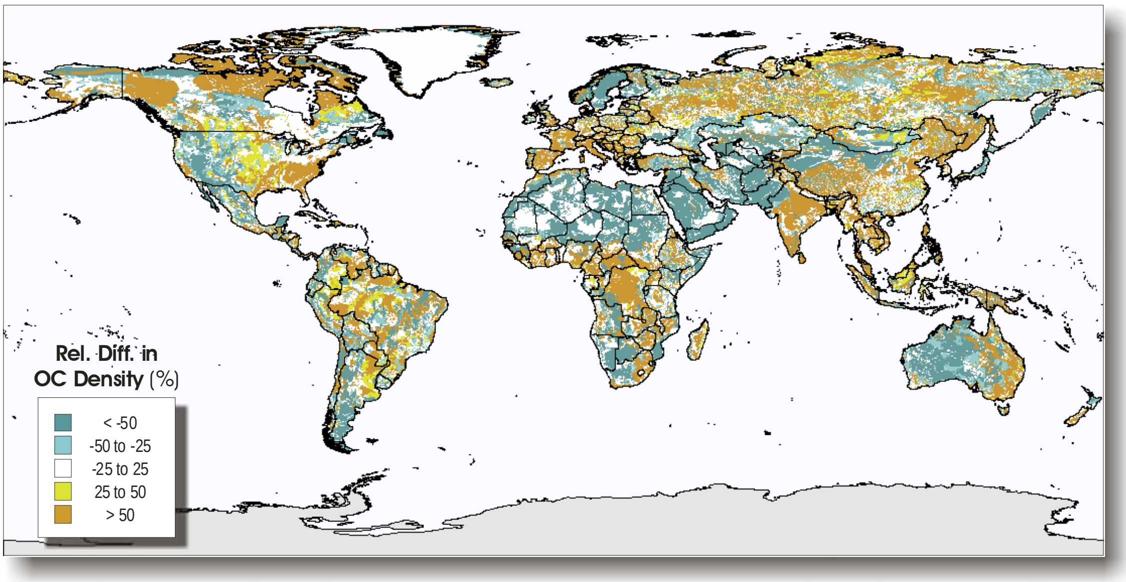


**Figure 21: Relative Difference in SOC Stock in between NRCS and Amended HWSD Data**

The map shows that despite comparable overall global SOC stocks there are significant local differences between the NRCS and the HWSD data (Std. Dev.: 143 $t\ ha^{-1}$). The differences are closely related to the soil moisture regime presented by NRCS[5]. The moisture and temperature regimes are also found in maps of climate classification, such as the one used by Köppen&Geiger (Kottek, *et al.*, 2006; Peel *et al.*, 2007). Areas classified as arid (Main climate type B) show lower SOC stocks in the NRCS map while other climate types have generally higher quantities.

---

[5] http://soils.usda.gov/use/worldsoils/mapindex/smr.html

### 4.7.5 FAO Organic Carbon Pool

From the FAO-UNESCO Soil Map of the World maps of soil carbon pools for the topsoil and subsoil layer were generated using the "Derived Soil Properties" attached to the raster data. Edition 3.6 from 2007 is available for download from the FAO Geonetwork[6] site. The data layers use a 5 *arc minute* grid resolution and are in geographic coordinates with WGS84 reference system.

Then number and ranges of the classes differ between the topsoil and the subsoil layer data of the FAO data and quite notably from the 9 classes of the NRSC SOC map, as presented in Table 13.

**Table 13: Classes and Ranges for SOC Density of FAO Organic Carbon Pool Maps**

| Class Label | Topsoil SOC Density Min $t\ ha^{-1}$ | Topsoil SOC Density Max $t\ ha^{-1}$ | Topsoil Mean Value $t\ ha^{-1}$ | Subsoil SOC Density Min $t\ ha^{-1}$ | Subsoil SOC Density Max $t\ ha^{-1}$ | Subsoil Mean Value $t\ ha^{-1}$ |
|---|---|---|---|---|---|---|
| 1 | 0 | 18 | 9.0 | 0 | 36 | 18.0 |
| 2 | 18 | 36 | 27.0 | 36 | 75 | 55.5 |
| 3 | 36 | 75 | 55.5 | 75 | 150 | 112.5 |
| 4 | 75 | 150 | 112.5 | 150 | 350 | 250.0 |
| 5 | 150* | 350 | 350.0 | 350 | | 600.0 |

* Range not defined in map legend file, but in file “carbon_pool.txt”.

Despite the specifications given in the metadata of the download site and the file “carbon_pool.txt” accompanying the data files the topsoil layer data does not contain a class for SOC densities of 150 - 350 $t\ ha^{-1}$. Therefore, when re-classifying the HWSD any values >150 $t\ ha^{-1}$ was set to 350 $t\ ha^{-1}$. For the subsoil layer any value >350 $t\ ha^{-1}$ was set to 600 $t\ ha^{-1}$, which approximates the value of the final two classes in the NRCS SOC map.

Classifying the HWSD according to the ranges of the FAO maps resulted in similarities between the general pattern of the distribution of SOC densities, but the classes were largely different, as shown in Figure 22.

The classified FAO OCP data were then assigned to SOC densities using the mean class values given in Table 13. From the topsoil and subsoil layers the global SOC stocks were then estimated. The settings following the classification specified in the meta-data gave approx. 5 times as much SOC in the subsoil as compared to the topsoil. Together with the inconsistency in the classes of the individual soil layers the result was taken as an indicator for some anomaly in the data available for download from the Web-site.

---

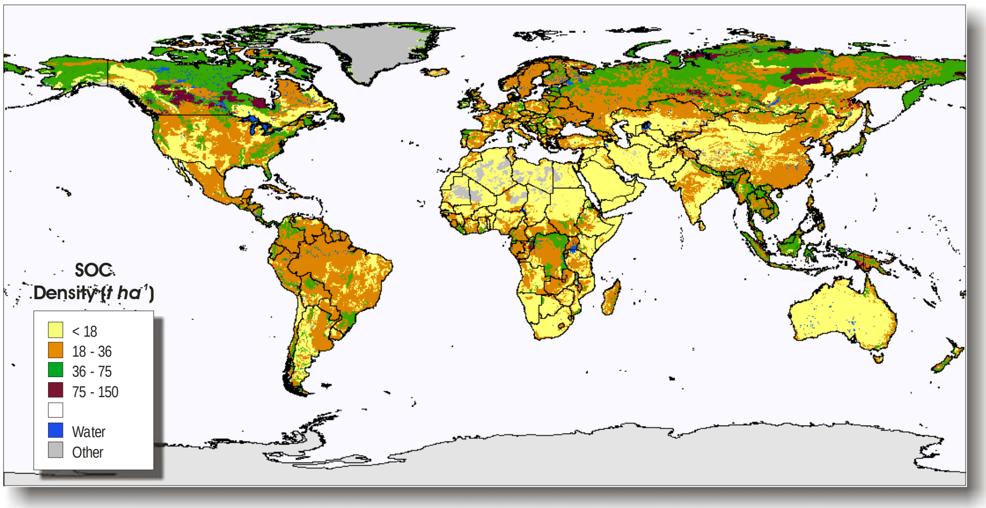
[6] http://www.fao.org/geonetwork/srv/en/main.home

As a trial the classifications between topsoil and subsoil were reversed and the quantification was repeated using the class means of inverted legends. The results on global SOC stocks by soil layer obtained from the assignments are presented in Table 14.

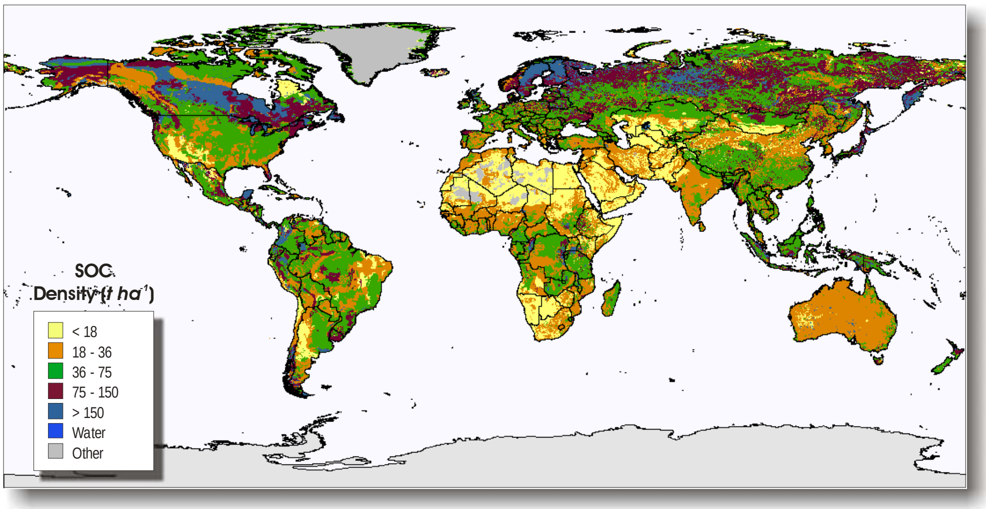
**Table 14: Global SOC Stocks from FAO Organic Carbon Pool Maps from Defined and Inverted Classification Schemes**

| Soil Layer | Classification | SOC Stock Layer *Pg C* | SOC Stock Total *Pg C* |
|---|---|---|---|
| Topsoil | Topsoil | 334 | |
| Subsoil | Subsoil | 1,515 | 1,849 |
| Topsoil | Subsoil | 710 | |
| Subsoil | Topsoil | 746 | 1,459 |

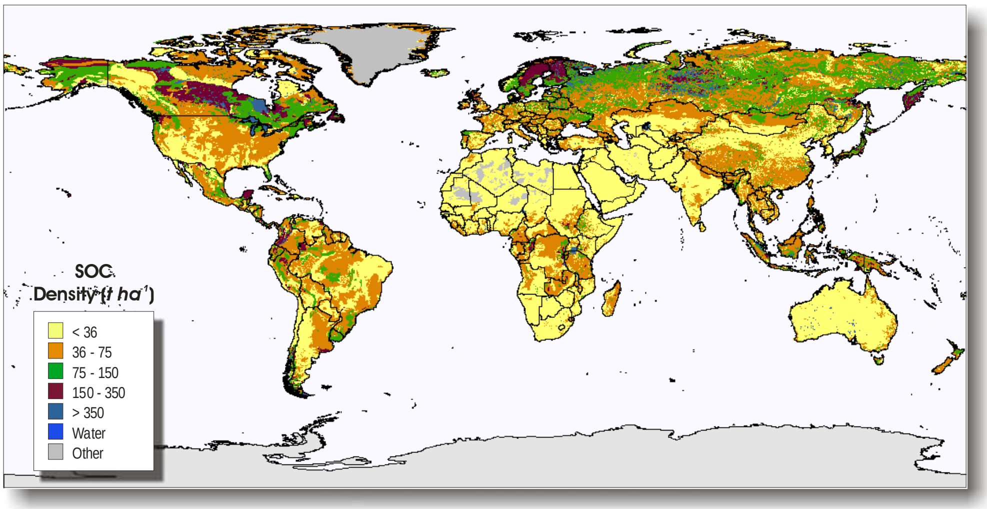
The distribution in global SOC stocks between the topsoil and subsoil layers suggested to assess the correspondence of the FAO data with the *amended HWDS* (HWSDa). The soil layers of the HWSDa were classified according to the defined classification scheme and to the interchange schemes. A comparison of the FAO OCP topsoil map following the specified legend and the HWSDa data classified according to the topsoil and subsoil classes is presented in Figure 22. The graphs show a visibly closer resemblance of the HWSDa data with the FAO OCP Topsoil map using the subsoil classification for the HWSDa topsoil data. This is also the case for the subsoil data, which are not shown here.



a) FAO Orcanig Carbon Pool, Topsoil



b) HWSD Topsoil, FAO Topsoil Classification



c) HWSD Topsoil, FAO Subsoil Classification

**Figure 22: FAO Organic Carbon Pool. Topsoil and HWSD Topsoil Classified according to FAO Topsoil and Subsoil Legend Values**

The agreement between the FAO OCP and the classified HWSDa layers was then assessed by using cross-classifications for which the overall Kappa was determined. The results are presented in Table 15.

**Table 15: Correspondence between FAO Organic Carbon Pool Topsoil and Classified HWSDa Data by Defined and Interchanged Classification Schemes**

| HWSD | | FAO Organic Carbon Pool | Overall Kappa |
|---|---|---|---|
| *Layer* | *Classification* | *Layer* | |
| Topsoil | Topsoil | Topsoil | 0.1050 |
| Topsoil | Subsoil | Topsoil | 0.3065 |
| | Subsoil* | | 0.3311 |
| Subsoil | Subsoil | Subsoil | 0.0258 |
| Subsoil | Topsoil** | Subsoil | 0.2231 |

* Class 5 (>350 $t\ ha^{-1}$) merged with Class 4 (>150-350 $t\ ha^{-1}$)

** Class 5 as >150 $t\ ha^{-1}$

For the computation of the Kappa any non-soil classes in the layers were excluded. There is a notable increase in the overall Kappa index of agreement when interchanging the classification schemes. However, the agreement of the interchanged classification of the HWSDa with the FAO OCP topsoil layer is still rated as only poor to fair.

In the light of these findings it would seem useful to re-process the FAO-UNESCO Soil Map of the World data with the carbon pool values given in the file “*CP.xls*” of the data download package. This task, however, was outside the scope of this evaluation.

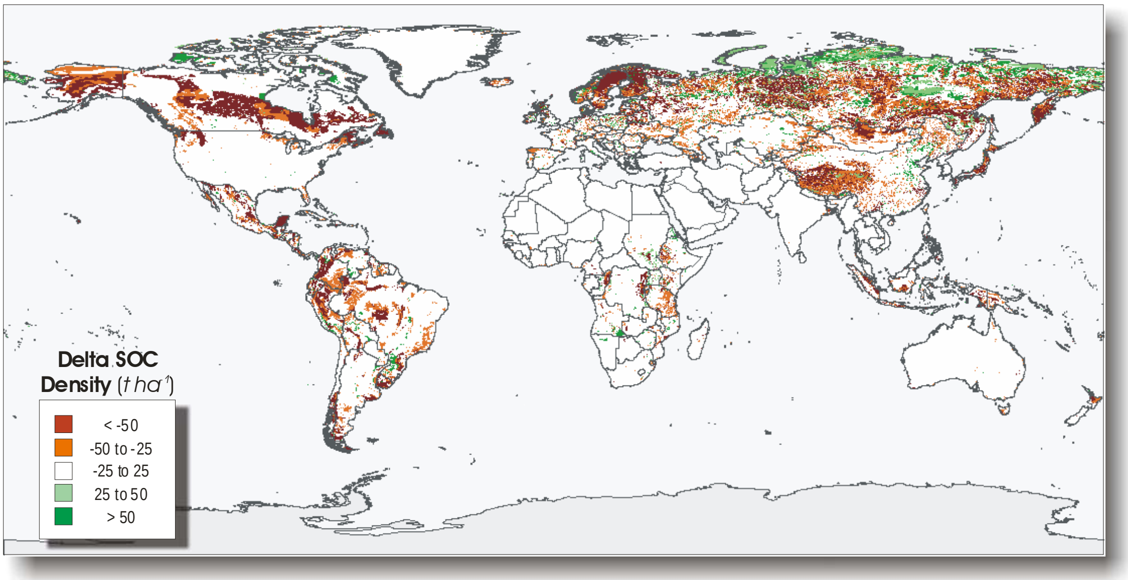
### 4.7.6 WISE5BY5MIN

The ISRIC-WISE data set of derived soil properties on a 5x5 *arc minute* global grid (WISE5by5MIN, ver. 1.1) was compiled from combining the FAO-UNESCO Soil Map of the World and parameters estimated from the WISE database (Batjes, 2006).

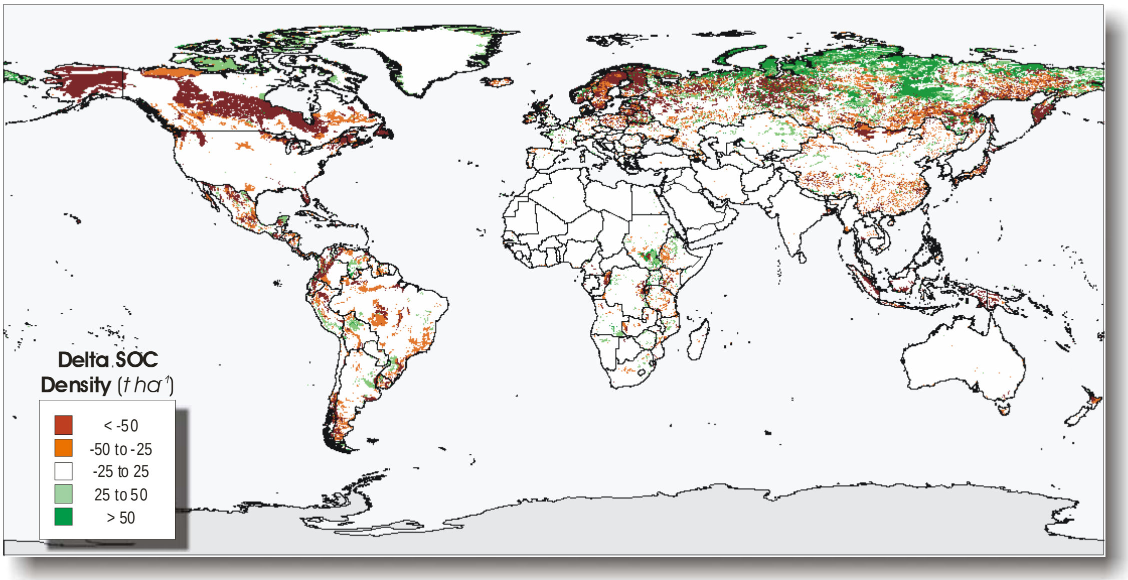
Soil properties are estimated for 5 layers of fixed depth in 20 cm intervals. To account for shallow soils the soil depth of the first layer can be either 10 or 20cm. For deeper soils the depth is given depending on the presence of data for a layer. For each layer more than one set of soil properties may be defined, where the distribution of the soil property set is given as a proportion within the depth layer. This arrangement of a one-to-many relationship is comparable to the use of typological units in other soil data sets.

For the comparison of the WISE5by5MIN data with the HWSDa the SOC density of the layers was arranged into topsoil and subsoil layers. To obtain an estimate for two layers the density of the depth layer D2 (20 – 40cm) was equally distributed between the topsoil and subsoil layer. The process resulted in a global SOC stock of 991 $t\ ha^{-1}$, with 504 $t\ ha^{-1}$ in the topsoil and 487 $t\ ha^{-1}$ in the subsoil layer.

These figures are considerably lower than global SOC stock estimates obtained from other data sources. The geographic variations in the differences between the WISE5by5MIN and the HWSDa data for the topsoil and subsoil OC density are presented in Figure 23.



a) Delta of WISE5by5MIN - HWSDa SOC Density, Topsoil



b) Delta of WISE5by5MIN - HWSDa SOC Density, Subsoil

**Figure 23: Difference in SOC Stock in between WISE5by5MIN and Amended HWSD Data for Topsoil and Subsoil Layers**

In many areas of the world the WISE5by5MIN data presents lower values for SOC density, except in the high northern latitudes, where the values are higher than those of the HWSDa. For the areas of common source data, such as North America, Western Africa, Southern and Eastern Asia and Australia, the difference is due to lower bulk densities for organic soils. In areas where the source data differ, such as South America and Europe, the soil types and their distribution very much deviate. The relationship between OC content and bulk density is presented in Figure 24.

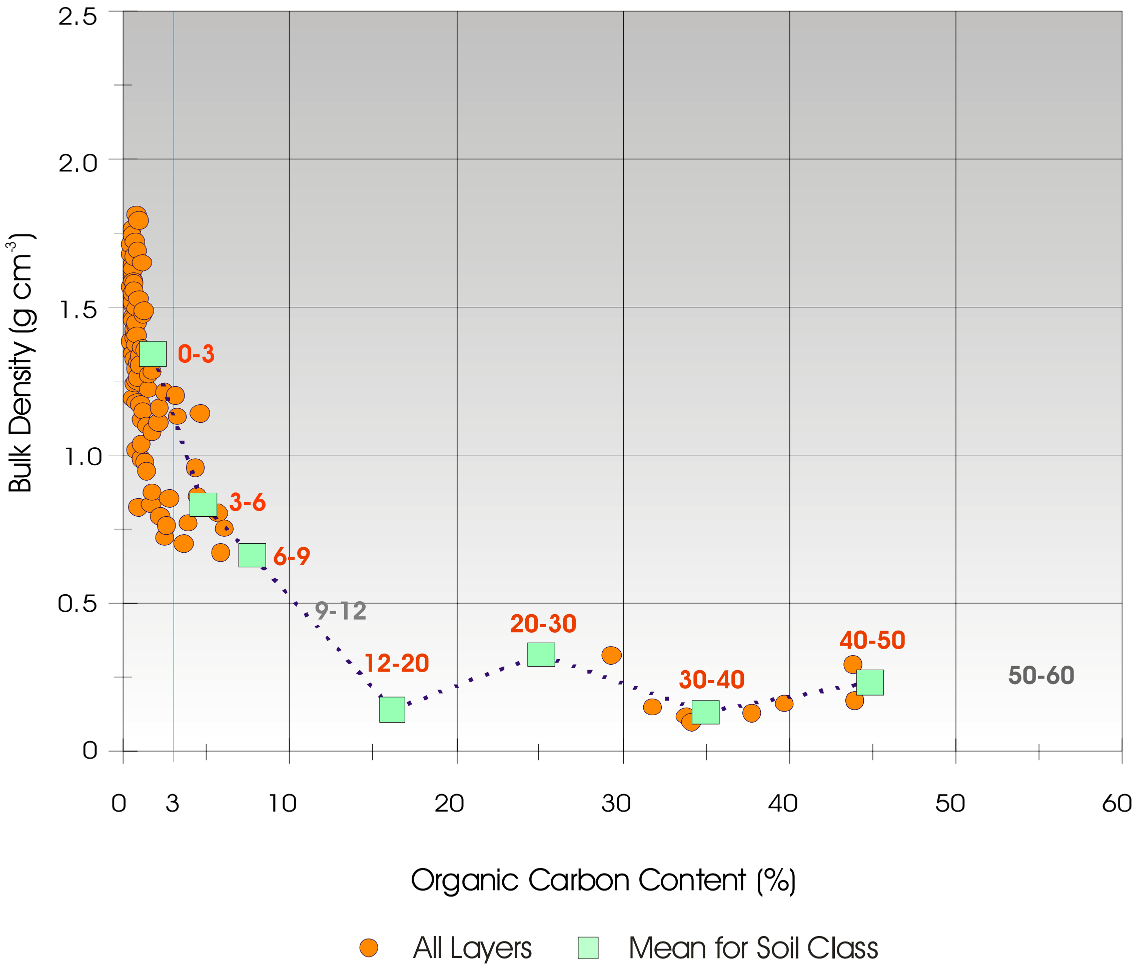


**Figure 24: Relationship between Organic Carbon Content (%) and Bulk Density for ISRIC-WISE5by5MIN Gridded Data for all Layers and by Interval Mean**

The graph shows a clear separation between mineral and organic soils. There are no values in the range 9-12% OC content just one value from 12-20% (16.6%) and from 20-30% (28.28%), although there are several occurrences of these values. Notable in the distribution of OC content and bulk density presented in Figure 24 is also the absence of layers with an OC content > 44.0%. For the 1,948 soil layers with an OC content >20% 95% have OC contents <40%.

For layers with > 20% OC content 8 different values for bulk density are present in the data set, ranging from 0.12 to 0.34 $g\ cm^{-3}$. Of these more than 60% have a bulk density of 0.15 $g\ cm^{-3}$ or less. This contrasts quite strongly with the distribution of the bulk density used for the HWSDa and the ISRIC-WISE V3.1 profile database. The ISRIC-WISE profile data contains several instances with profiles of a bulk density <0.2 $g\ cm^{-3}$. The re-processed data estimates topsoil bulk densities <0.2 $g\ cm^{-3}$ for 15 profiles and for 20 profiles in the subsoil. However, of these 5 have OC contents < 12% in the topsoil or the subsoil and 4 in both layers.

The number of bulk density values in the data set is further restricted by depth layer. For soil units with an OC content >12% the number of bulk density values by depth layer is either 2 (D1, D2 and D4) or 3 (D3 and D5) values. With layer depth these values generally decrease with a limited number of changes from all possible combinations. The values for OC contents of these layers show an increase from D1 to D2 followed by a decrease with layer depth. For the combination of bulk density and OC content the soil units of each layer contain 4 pairs of values. These pairs are presented in Table 16.

**Table 16: Combinations of Bulk Density and Organic Carbon Content in WISE5byMIN Data for Soil Units with OC Content > 12%**

| Layer | Bulk Density | Organic Carbon Content | Soil Units | Organic Carbon Density | |
|---|---|---|---|---|---|
| | | | | Layer | Layer Mean |
| | $g\ cm^{-3}$ | $g\ kg^{-1}$ | Count | $t\ ha^{-1}$ | $t\ ha^{-1}$ |
| D1 | 0.31 | 439 | 43 | 1361 | |
| | 0.31 | 327 | 79 | 1012 | |
| | 0.31 | 384 | 82 | 1190 | |
| | 0.34 | 293 | 202 | 995 | 1,077 |
| D2 | 0.19 | 440 | 43 | 837 | |
| | 0.19 | 359 | 79 | 682 | |
| | 0.19 | 373 | 82 | 708 | |
| | 0.12 | 341 | 202 | 409 | 568 |
| D3 | 0.15 | 378 | 43 | 567 | |
| | 0.15 | 356 | 79 | 533 | |
| | 0.18 | 398 | 82 | 716 | |
| | 0.12 | 340 | 202 | 408 | 511 |
| D4 | 0.14 | 383 | 43 | 536 | |
| | 0.14 | 340 | 79 | 476 | |
| | 0.17 | 318 | 82 | 541 | |
| | 0.14 | 338 | 202 | 473 | 494 |
| D5 | 0.14 | 378 | 43 | 529 | |
| | 0.14 | 318 | 79 | 445 | |
| | 0.16 | 162 | 82 | 259 | |
| | 0.12 | 320 | 202 | 383 | 386 |

The OC densities was computed from the pairs of OC content and bulk density for a depth layer without considering a value for coarse fragments and a standard depth of 100cm. This depth was used in preference to the layer thickness (10 or 20 cm) for reason s of comparability with other data. The final column contains the mean value for OC density for each layer taking only the soil unit data into account, but not the spatial distribution of the units in the map layer. This mean OC density is highest in the uppermost D1 layer (1,077 $t\ ha^{-1}$) and falls to half the value (568 $t\ ha^{-1}$) for the D2 layer data. It decreases further with increasing depth of the layer. While a decrease in bulk density with depth is not uncommon for organic soils (Hiederer, 2009) the generally occurring decrease from D1 to D2 is steep. This is combined with a decrease in OC content from D2 to lower layers, which subsequently results in a decrease in OC density with layer depth.

Characterizing the relationship between OC content and bulk density by a regression model is determined by the absence of a defined trend for soil with an OC content >9%. The regression models estimate bulk densities between 0.06 – 0.13 $g\ cm^{-3}$ for organic matter, or less than half of what is estimated from the ISRIC-WISE V3.1 profile database. Such values are not outside the range found in the literature, although they contribute to the lower global SOC stocks.

This combination of relatively low bulk density and limit in the OC content lowers the estimates for global SOC stocks compared to the HWSDa where the source data is the FAO-UNESCO Soil Map of the World. In areas where the HWSD uses different data sources organic soils are less widely distributed. In Central and South America data from the *Harmonized continental SOTER-derived database* (SOTWIS) data set contains areas with soil >6% OC content, which are few in the WISE5by5MIN data.

Higher SOC densities in the WISE5by5MIN than in the HWSDa data are found in Arctic regions. With 16 to 60 $t\ ha^{-1}$ the difference to the HWSDa is relatively small, considering that the SOC densities are already high and hence the effect on global estimates does not compensate for the lower values in other areas of Eurasia.

### 4.7.7 DSMW

The FAO-UNESCO *Digital Soil Map of the World* (DSMW)[7] is based on the FAO-UNESCO Soil Map of the World. The soil map was compiled from a collection of soil profile data and published in 1974 and 1978. The DSMW further includes information from the. *Global and national soils and terrain database* (SOTER; UNEP/ISSS/ISRIC/FAO, 1993). The first version became officially available in 1991 at 1:5 mil. scale, initially in vector format and later also in raster format with a grid size of 5 *arc minute*. In this evaluation Ver.3.6 is used, which was completed in January, 2003. The files can be downloaded in various GIS formats from the FAO Geonetwork site[8].

---

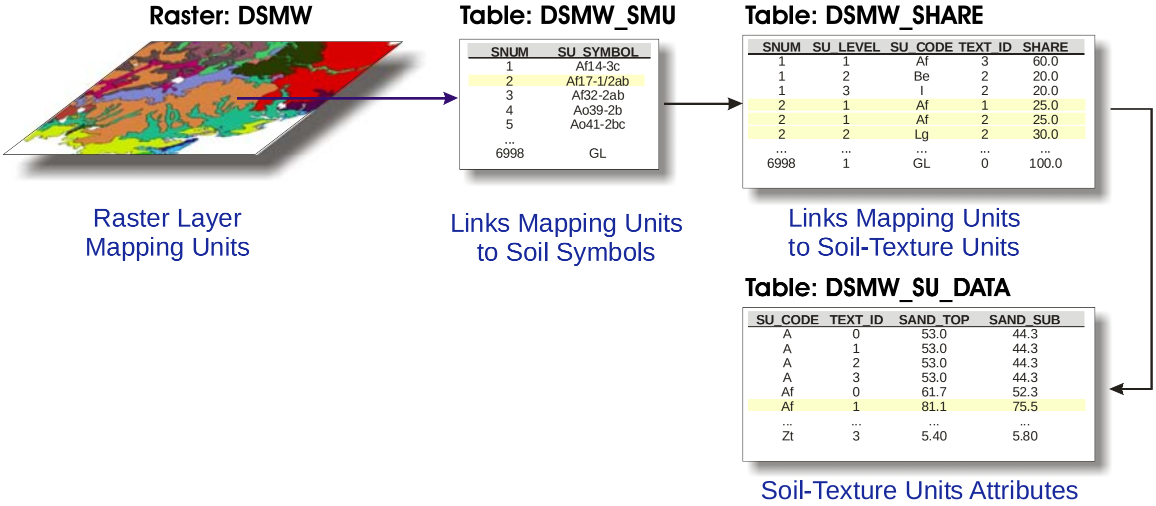
[7] Source: Land and Water Development Division, FAO, Rome

[8] http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116&currTab=distribution

Data from the DSMW for the basis of derived products of global SOC maps, such as the NRCS Global Soil Carbon Layers, the FAO Global Carbon Pools or the WISE5by5MIN SOC data. Also the HWSD uses the DSMW in several regions. This common source of data for various products of SOC stocks implies a level of correlation in the values. Therefore, global SOC stocks from the products are not independent from each other and closeness of values should not be treated as confirmation of any value or range of values.

The attributes of the 4,931 soil mapping units of the spatial layer are defined by characteristics of the soil units. The DSMW defines 135 soil units, which include areas of non-soil, such as rock and water. Each soil unit related to soil has three expressions, which are varied by the texture classes, and one default expression. This organization of data results in the definition of characteristics for 512 soil unit–texture combinations. Information on slope is not used. Up to 8 soil unit–texture combinations, one dominant and 7 associated soil units, can be linked to a mapping unit. The share of the soil unit–texture combinations in a mapping unit is assembled by combining the proportions from the soil unit share that belongs to a texture-slope class by only the texture class. The resulting table defines 16,445 records, which link the soil attributes with the mapping units of the spatial layer.

For further processing the spreadsheet format of the downloaded files were transferred to tables of a *Relational Database Management System* (RDBMS). The attribute data to the spatial layer were normalized using the soil unit code and texture class as table keys. The schematized structure of the processing database for the DSMW data is presented in Figure 25.



| SNUM | SU_SYMBOL |
|---|---|
| 1 | Af14-3c |
| 2 | Af17-1/2ab |
| 3 | Af32-2ab |
| 4 | Ao39-2b |
| 5 | Ao41-2bc |
| ... | |
| 6998 | GL |

| SNUM | SU_LEVEL | SU_CODE | TEXT_ID | SHARE |
|---|---|---|---|---|
| 1 | 1 | Af | 3 | 60.0 |
| 1 | 2 | Be | 2 | 20.0 |
| 1 | 3 | I | 2 | 20.0 |
| 2 | 1 | Af | 1 | 25.0 |
| 2 | 1 | Af | 2 | 25.0 |
| 2 | 2 | Lg | 2 | 30.0 |
| ... | ... | ... | ... | ... |
| 6998 | 1 | GL | 0 | 100.0 |

| SU_CODE | TEXT_ID | SAND_TOP | SAND_SUB |
|---|---|---|---|
| A | 0 | 53.0 | 44.3 |
| A | 1 | 53.0 | 44.3 |
| A | 2 | 53.0 | 44.3 |
| A | 3 | 53.0 | 44.3 |
| Af | 0 | 61.7 | 52.3 |
| Af | 1 | 81.1 | 75.5 |
| ... | ... | ... | ... |
| Zt | 3 | 5.40 | 5.80 |

**Figure 25: Schematic Organization of Spatial and Typological Units in the Re-Designed Digital Soil Map of the World**

The structure resembles the organization of the HWSD and the ESDB. A difference is that more than one attribute of the DSMW_SU_DATA table can be linked to a level of a mapping unit in the DSMW_SHARE table.

The DSMW does not include data on soil depth for the mapping or typological units. To assess SOC stocks by topsoil and subsoil layer information on effective soil depth from the *Derived Soil Properties*[9] data set was used. For an appreciation of the amount of coarse fragments the PTF used by Reynolds *et al*. (1999) was re-constructed from the attribute file for *Coarse Fragments*[10]. The coarse fragments are defined by the soil symbols used to identify the Soil-Texture unit attributes. The values are given in Table 17.

**Table 17: PTF for Coarse Fragments (based on Reynolds, *et al*., 1999)**

| Soil Unit | | Coarse Fragments | Soil Unit | | Coarse Fragments |
|---|---|---|---|---|---|
| *Code* | *Name* | *%* | *Code* | *Name* | *%* |
| A | Acrisols | 5 | N | Nitosols | 5 |
| B | Cambisols | 5 | O | Histosols | 5 |
| C | Chernozems | 5 | P | Podzols | 5 |
| D | Podzoluviosols | 5 | Q | Arenosols | 5 |
| E | Rendzinas | 50 | R | Regosols | 40 |
| F | Ferralols | 5 | S | Solonetz | 5 |
| G | Gleysols | 5 | T | Andosols | 5 |
| H | Phaeozems | 10 | U | Rankers | 40 |
| I | Lithosols | 40 | V | Vertisols | 5 |
| J | Fluvisols | 5 | W | Planosols | 5 |
| K | Kastanozems | 5 | X | Xerosols | 5 |
| L | Luvisols | 5 | Y | Yermosols | 10 |
| M | Greyzems | 5 | Z | Solonchaks | 5 |

The default value for any soil unit is 5% coarse fragments. This contrasts with data from the HWSD where no coarse fragments are recorded in the database. It is not generally recognizable whether the absence of data on coarse fragments is due to the absence of the parameter or lack of data. The difference in data on coarse fragments directly impacts on the local estimates of SOC density and potentially on global SOC stocks.

The global SOC stock estimates calculated from the DSMW with the additions from the PTF are:

---

[9] http://www.fao.org/geonetwork/srv/en/resources.get?id=30581&fname=depth.zip&access=private

[10] http://www.ngdc.noaa.gov/ecosys/cdroms/reynolds/reynolds/coarse_fragments/coarsfrg.csv
Note: inconsistent spelling of file from Web-site (coarsfrag.csv vs. coarsefrg.csv).

Topsoil: 574 *Pg* C

Subsoil: 632 *Pg* C

The estimate for the global SOC stocks to a depth of 1 m is thus 1,206 *Pg* C. This stock is 211 *Pg* C lower than the estimate obtained from amended HWSDa data.

The local differences in SOC density estimates from the DSMW and the HWSDa for the combined topsoil and subsoil layers are presented Figure 26.

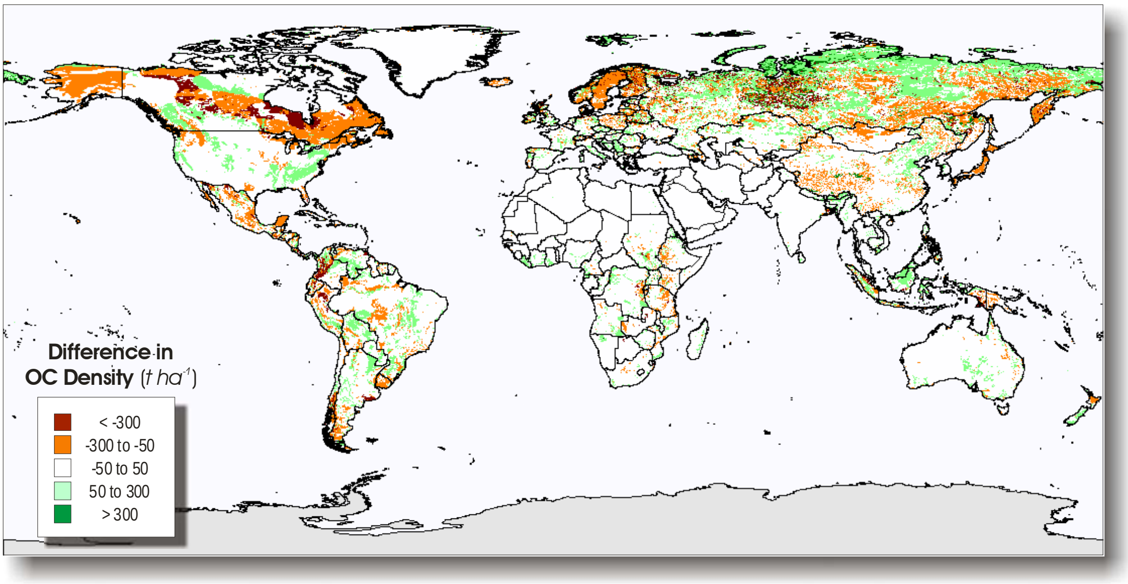
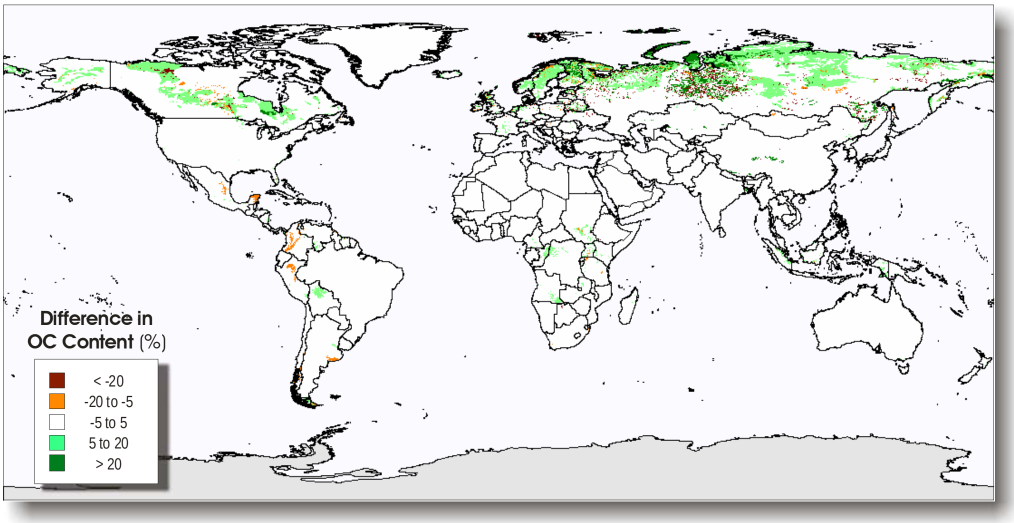
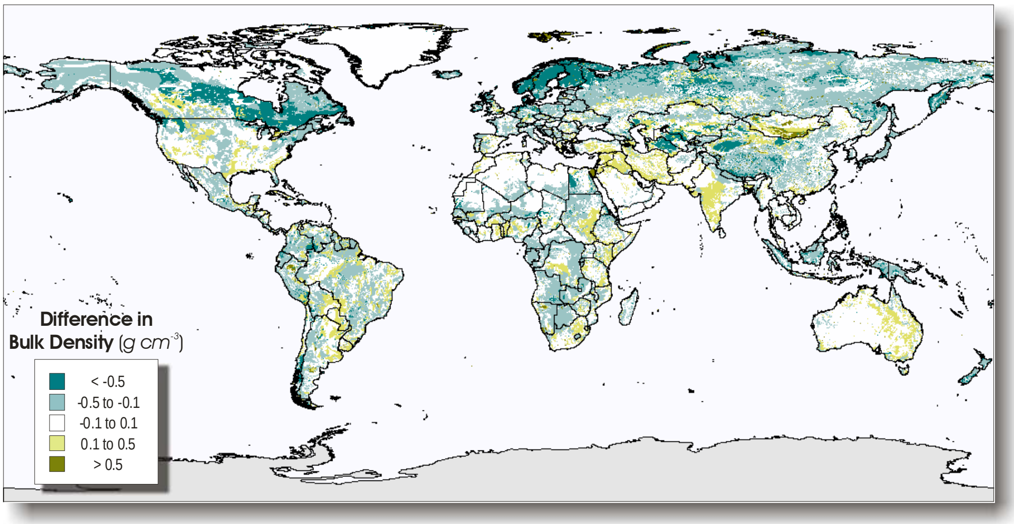


**Figure 26: Difference in SOC Density between DSMW and Amended HWSD Data for Combined Topsoil and Subsoil Layers**

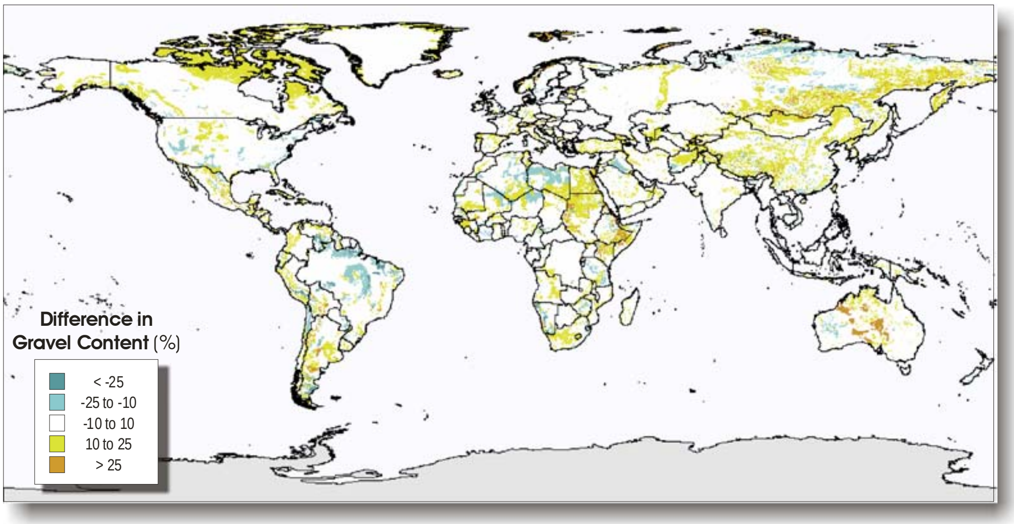
The map indicates the main differences between the data sets for areas with a high proportion of organic soils at high latitudes in the northern hemisphere. In the DSMW the values are higher than the HWSDa for large parts in Siberia and lower for Scandinavia, Canada and Alaska. However, there are also notable differences in South America and the areas of peat in South-East Asia. The differences in SOC density are mainly caused by variations in the bulk density parameter. These differences between the DSMW and the HWSDa data for the parameters OC content, bulk density and coarse fragments for the combined topsoil and subsoil layers are presented in Figure 27.



a) Delta of OC Content between DSMW and HWSDa



b) Delta of Bulk Density between DSMW and HWSDa



c) Delta of Coarse Fragments / Gravel between DSMW and HWSDa

Figure 27: Differences between DSMW and Amended HWSD Data for SOC Density Parameters

The graphs illustrate that the parameters determining SOC density vary by parameter and locally for each parameter. In the visual interpretation of differences based on the graphs it should be considered that the classification thresholds very much influence the appearance of differences in the graphs.

Differences in OC content of >5% are very much limited to areas in Canada, northern Europe and Siberia. Differences of this magnitude in other areas are limited to few mapping units and are found in South America, Africa and South-East Asia. For the bulk density parameter differences of 0.1 to 0.5 $g\,cm^{-3}$ between the DSMW and the HWSDa are quite widespread. Areas where variances in bulk density exceed 0.5 $g\,cm^{-3}$ are found in Canada, Scandinavia and western Siberia. For the volume of coarse fragments or gravel content the DSMW shows higher values for most areas. Exceptions are mainly found in Brazil and northern Africa.

The differences in SOC densities in the northern Hemisphere are due to a combination of higher OC contents in the DSMW than the HWSDa and generally lower values for bulk density. The differences in coarse fragments or gravel on SOC density are of minor significance in those areas. The differences in OC content in northern Europe and Siberia may be attributed to the use of alternative sources for the data. The observed deviations of OC content for mapping units in Canada are unexpected, since for this area the DSMW is the underlying data source.

The higher values for OC content and lower values for bulk density in the DSMW in the northern hemisphere offset to some degree the differences in SOC density. In other regions of the world the differences in SOC density are largely affected by bulk density alone. Variations in coarse fragments or gravel content are less consequential since the areas concerned are mainly arid with low OC.

# 5 SUMMARY AND CONCLUSIONS

The various sources of spatial data for estimating global SOC stocks used in this evaluation needed further processing or amendments of data to provide coherent and comparable results. In particular the use of ancillary data to complete the parameters for calculating SOC density (bulk density, coarse fragments and soil depth) very much influences the global estimates. The results presented here are therefore only valid for the conditions under which the data were evaluated, for example the processing to topsoil and subsoil layers.

Estimating bulk density for soils with an OC content > 3% from profile data by a regression model has proven to depend on the data used, the model assumed and the aggregation method applied. Using the European SPADE/M profile data the relationship between OC content and bulk density was best modelled by a log-transformed bulk density function. For the profiles a relationship between the parameters was also found for organic soils with a bulk density of organic matter of 0.1 $g\ cm^{-3}$. For the global profiles of the ISRIC/WISE data set a much more varied relationship was found. The three regressions models included in the evaluation provide indistinct results for capturing the variation in the data, but the log-transformed bulk density function provides a bulk density for organic matter which is about 50% less than the estimates of the other two functions.

A summary of the estimates of global SOC stocks by topsoil (0-30cm), subsoil (30 – 100cm) and the combined layers (0 - 100cm or when less to depth of soil layer) is presented in Table 18.

**Table 18: Summary of Estimates of Global SOC Stocks in Topsoil, Subsoil and Combined Layers**

| Layer | Data Source | | | | | |
|---|---|---|---|---|---|---|
| | HWSD | HWSDa | NRCS | FAO* | WISE 5by5MIN | DSMW** |
| | *Pg C* | *Pg C* | *Pg C* | *Pg C* | *Pg C* | *Pg C* |
| Topsoil | 967 | 699 | - | 710 | 504 | 574 |
| Subsoil | 1,502 | 718 | - | 746 | 487 | 632 |
| 0-100cm | 2,469 | 1,417 | 1,399 | 1,459 | 991 | 1,206 |

* Inverted classification for topsoil and subsoil. Using original classification: 1,849 *Pg* C

** Depth from the *Derived Soil Properties*[11], coarse fragments from PTF by Reynolds *et al*. (1999)

---

[11] http://www.fao.org/geonetwork/srv/en/resources.get?id=30581&fname=depth.zip&access=private

The estimates derived from the HWSDa, the NRCS and FAO are within 60 *Pg* C. Despite the similarity of the global figures for these data there are notable local variations in SOC density. Some can be linked to specific processing additions, such as using climatic conditions to modify OC content as in the case of the NRCS data, or to the use of ordinal scales to represent the data, as for the FAO maps.

The cause of the lower global estimates derived from the DSMW data is primarily attributed to lower values for bulk density in the data set when compared to the HWSD. The estimates derived from the WISE5by5MIN data are considerably lower as a consequence of generally low values for bulk density for organic soils and low values for OC content in the soil. Affected by the combination of unusual parameter values is in particular the subsoil layer, for which a SOC stock below the amount found in the topsoil is computed.

The HWSD represents a step forward towards spatially more detailed and thematically more refined set of global soil data. It successfully integrates into a common structure data from very diverse origins. The use of a raster format to map the spatial units instead of the vector format employed by the DSMW or the ESDB simplifies the implementation of mathematical functions in a GIS as spatial overlay operations when integrating soil properties with other thematic data. During the evaluation of the data for generating estimates of global SOC stocks Version 1.1 of the HWSD revealed also some rough edges. Without further commenting on the database model and for the set of parameters analyzed, these concern inconsistencies in the data nomenclature and the presence and treatment of missing data. However, the number of inconsistencies and omissions is relatively small and should not critically influence global estimates.

However, the values of bulk density attributed to soils high in organic carbon are erroneous and with serious consequences on estimates of local and global SOC stocks. The method used to estimate bulk density should be restricted to the soils for which it was developed, i.e. mineral soils, and an alternative method should be applied for soils high in organic carbon.

The findings from the investigation of estimating bulk density by OC content using the SPADE/M and the ISRIC-WISE profile data sets suggest that using only the OC content parameter to model bulk density is not applicable to soils with an OC content <3%. They also agree that the relationship between the two parameters for organic soils is not necessarily different from being linear. Separate relationships could be used to model the relationship for soils with a threshold of the OC content between 9 - 12%. When defining a single relationship the regression analysis should use a method appropriate to avoid an undue influence of the relationship for mineral soils on organic soils.

All spatial global data sets on the distribution of OC were found to have their particularities. One rather curious and consequential characteristic is that parameter values which determine SOC stocks seem to be defined by a rule-based method. This leads to few distinct values, although not integers, with a high frequency of occurrence. While it may be assumed that bulk density is modelled from other soil parameters or using a PTR based on soil taxonomy, finding a comparably limited number of values for OC content was unexpected. For organic soils the values for OC content in the typological data frequently fall short of the range found in the profile data. This would lead to an under-estimation of global SOC stocks unless the values present in the data sets represent an unbiased mean when calculating SOC densities. The regression model adopted to estimate bulk densities for organic soils is also of concern, because for the function parameters derived from the profile data the SOC densities peak for the linear or log-transformed bulk density function for soils with around 35 - 45% OC content, yet for the function parameters derived for the log-transformed OC content and the reciprocal functions SOC density continues to increase with OC content. For organic soils the variations in the amount of carbon is thus closely linked to the amount of material included in a fixed volume rather than to carbon content of a fixed amount of soil material. This incidentally makes calculating changes in carbon stocks an unsuitable method for estimating $CO_2$ emissions from organic soils.

The evaluation has demonstrated that bulk density is the most important factor for estimating SOC density and, subsequently, SOC stocks. In contrast to the effect of the variability of the volume of coarse fragments on SOC density the uncertainty of estimating bulk density is highest where the values are lowest. In the databases the volume of coarse fragments is high predominantly in dry areas, where the OC content is low and soils are frequently shallow. In these areas a difference in bulk density of 0.1 $g\ cm^{-3}$ would lead to a difference in SOC density of less than 10%. In areas with soils high organic carbon such a difference can lead to a doubling of SOC stocks. Therefore, the differences in SOC stocks between the various data sets are most prominent in areas with soils high in organic carbon. The presence of such spatially variable differences are also present in data for which similar estimates of global SOC stocks are computed. Apart from the spatial compensation of variations in SOC density it was also found that differences in parameters can offset the impact on the computed value, such as higher OC content combined with lower bulk density. There is therefore more variation between the data sets used than the global estimate of SOC stock indicates.

As a consequence, where changes in one of the parameters determining SOC stocks are spatially variable the changes are likely to generate different responses in SOC stock between the data sets used. This is the case when modelling the effect of changes in climatic conditions on soil organic carbon or changes in land use. Bulk density is closely linked to SOC content for soils where changes in OC content mostly affects soil carbon stocks. Thus, the evaluation suggests that while the focus in projecting the effect of climate change or land use is on the fate of soil organic carbon the main uncertainty for global SOC stocks stems from the vagueness in determining bulk density for organic soils.

## References

Adams, W.A. (1973) The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. Journal of Soil Science, 24, pp. 10-17.

Batjes, N.H. (1992) Organic matter and carbon dioxide. In: A Review of Soil Factors and Processes that Control Fluxes of Heat, Moisture and Greenhouse Gases. Eds. N.H. Batjes & E.M. Bridges. Technical Paper 23, International Soil Reference and Information Centre, Wageningen. pp. 97-148.

Batjes, N.H. (1996) Total carbon and nitrogen in the soils of the World. European Journal of Soil Science, June 1996, Vol. 47(2), pp. 151-163. doi:10.1111/j.1365-2389.1996.tb01386.x.

Batjes, N.H. (2008) ISRIC-WISE Harmonized Global Soil Profile Dataset (ver. 3.1). Report 2008/02. ISRIC – World Soil Information, Wageningen, The Netherlands. 52pp. http://www.isric.org/isric/Webdocs/Docs/ISRIC_Report_2008_02.pdf

Batjes, N.H. (2006) ISRIC-WISE derived soil properties on a 5 by 5 arcminutes global grid (ver. 1.1). Report 2006/02 ISRIC – World Soil Information, Wageningen.

Harmonized Global Soil Profile Dataset (ver. 3.1). Report 2008/02. ISRIC – World Soil Information, Wageningen, The Netherlands. 52pp.

Bradley, R.I., R. Milne, J. Bell, A. Lilly, C. Jordan and A. Higgins (2005) A soil carbon and land use database for the United Kingdom. Soil Use and Management (2005) 21, pp. 363–369.

Chambers, F.M, D.W. Beilman and Z. Yu (2010) Methods for determining peat humification and for quantifying peat bulk density, organic matter and carbon content for palaeostudies of climate and peatland carbon dynamics. Mires and Peat, Volume 7 (2010/11), Article 07, pp. 1–10. ISSN 1819-754X.

Driessen, P. J. Deckers and F. Nachtergaele (eds) (2001) Lecture notes on the major soil types of the World. FAO, Publishing and Multimedia Service, Information Division, , Viale delle Terme di Caracalla, 00100 Rome, Italy. ISBN 925-104637-9. http://www.fao.org/docrep/003/y1899e/y1899e00.htm#toc

Ellert, B.H. and J.R. Bettany (1995) Calculation of organic matter and nutrients stored in soils under contrasting management regimes. Can. J. Soil Sci. (75) p.529-538.

FAO/IIASA/ISRIC/ISS-CAS/JRC (2009) Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

FAO (2001) Soil carbon sequestration for improved land management. Food and Agriculture Organization of the United Nations, Rome.

Fischer, G., F. Nachtergaele, S. Prieler, H.T. van Velthuizen, L. Verelst and D. Wiberg (2008) Global Agro-ecological Zones Assessment for Agriculture. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Global Ecosystems Database Project (2000) Global Ecosystems Database Version II: Database, User's Guide, and Dataset Documentation. US Department of Commerce, National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, Colorado. KGRD #35. Two CDROMs and publication on the World Wide Web.

Global Soil Data Task (2000) Global Soil Data Products CD-ROM (IGBP-DIS). CD-ROM. International Geosphere-Biosphere Programme, Data and Information System, Potsdam, Germany. Available from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. [http://www.daac.ornl.gov].

Goldberg, D. (1991) What every computer scientist should know about floating-point arithmetic. Computing Surveys, Vol. 23, March, 1991. Association for Computing Machinery, Inc., New York, USA. pp. 5 - 48. http://portal.acm.org/citation.cfm?id=103162.103163&coll=portal&dl=ACM&idx=J204&part=journal&WantType=Journals&title=ACM%20Computing%20Surveys%20%28CSUR%29&CFID=://csur.acm.org/&CFTOKEN=csur.acm.org/

Hiederer, R. (2009) Distribution of Organic Carbon in Soil Profile Data. EUR 23980 EN. Luxembourg: Office for Official Publications of the European Communities. 126pp.

Holdridge, L.R. (1947) Determination of world plant formations from simple climatic data. Science, 105, pp.367-368.

Intergovernmental Panel on Climate Change (IPCC) (2006) 2006 Guidelines for National Greenhouse Gas Inventories. Eggelstone, S., L. Buemdia, K. Miwa, T. Ngara and K. Tanabe (Eds.). IPCC/OECD/IEA/IGES, Hayama, Japan.

Intergovernmental Panel on Climate Change (IPCC) (2003) Good Practice Guidance for Land Use, Land-Use Change and Forestry. Penman J., M. Gytarsky, T. Hiraishi, T. Krug, D. Kruger, R. Pipatti, L. Buendia, K. Miwa, T. Ngara, K. Tanabe and F. Wagner (Eds.). IPCC/OECD/IEA/IGES, Hayama, Japan.

Jobbágy E.G and R.B. Jackson (2000) The Vertical Distribution of Soil Organic Carbon and Its Relation to Climate and Vegetation. Ecological Applications, Vol. 10, No. 2, pp. 423-436.

Kaur, R., S. Kumar and H.P. Gurung (2002) A pedo-transfer function (PTF) for estimating soil bulk density from basic soil data and its comparison with existing PTFs. Austr. J. Soil Res. 40, pp. 847-857.

Kottek, M., J. Grieser, C. Beck, B. Rudolf and F. Rubel (2006) World Map of the Köppen-Geiger climate classification updated. Meteorol. Z., 15, 259-263. DOI: 10.1127/0941-2948/2006/0130.

Lal, R. (2005) Encyclopedia of Soil Science – 2nd Edition. CRC Press. 2,060pp. ISBN: 9780849338304.

Peel, M. C., B.L. Finlayson and T.A. McMahon (2007) Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, pp. 1633–1644. ISSN 1027-5606.

Rawls, W.J. (1983) Estimating soil bulk-density from particle-size analysis and organic-matter content. Soil Science, 135(2): 123-125.

Reynolds, C.A., T. J. Jackson, and W.J. Rawls (1999) Estimating Available Water Content by Linking the FAO Soil Map of the World with Global Soil Profile Databases and Pedo-transfer Functions. Proceedings of the AGU 1999 Spring Conference, Boston, MA. May 31-June 4, 1999. available from: http://www.ngdc.noaa.gov/ecosys/cdroms/reynolds/reynolds/reynolds.htm#tr

Ruehlmann, J. and M. Körschens (2009) Calculating the effect of soil organic matter concentration on soil bulk density. Soil Sci. Soc. Am. J. (73); pp. 876-885.

Saxton, K.E., W.J. Rawls, J.S. Romberger, and R.I. Papendick. (1986) Estimating generalized soil-water characteristics from texture. *Soil Sci. Soc. Am. J.* 50(4), pp. 1031-1036.

Tranter G, B. Minasny, A.B. McBratney, B. Murphy, N.J. McKenzie, M. Grundy and D. Brough (2007) Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use and Management, 23(4), pp. 437-443.

UNEP/ISSS/ISRIC/FAO (1993) Global and national soils and terrain database (SOTER). Procedures manual. World Soil Resources Report 74. FAO, Rome. 122pp.

European Commission

**EUR 25225 EN – Joint Research Centre – Institute for Environment and Sustainability**

Title: Global Soil Organic Carbon Estimates and the Harmonized World Soil Database

Author(s): R. Hiederer, M. Köchy

Luxembourg: Publications Office of the European Union

2011 – 79 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online), ISSN 1018-5593 (print)

ISBN 978-92-79-23108-7

doi:10.2788/13267

## Abstract

Global estimates of soil organic carbon stocks have been produced in the past to support the calculation of potential emissions of $CO_2$ from the soil under scenarios of change land use/cover and climatic conditions (IPCC, 2006), but very few global estimates are presented as spatial data. For global spatial layers on soil parameters, the most recent and complete dataset is available as the *Harmonized World Soil Database* (HWSD). The HWSD represents a step forward towards a spatially more detailed and thematically more refined set of global soil data.

During the evaluation of the data for generating estimates of global SOC stocks Version 1.1 of the HWSD revealed also some rough edges. Consequential are the values of bulk density attributed to soils high in organic carbon. In the evaluation the database was completed and suitable substitutions for bulk density for soil high in organic carbon were investigated. For the amended data the global SOC stock to 100cm soil depth is estimated at 1,417 Pg C, although this estimate very much dependent on the ancillary data used.

The amended HWSD was compared to data from 4 other global data sets on SOC stocks. The comparative evaluation has demonstrated that bulk density is the most important factor for estimating SOC stocks and mainly responsible for the differences between estimates. Most affected from the variability in bulk density are SOC stocks in areas with soils which are high in organic carbon.

## How to obtain EU publications

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

LB-NA-25225-EN-N

EUROPEAN COMMISSION

Publications Office

ISBN 978-92-79-23108-7

9 789279 231087