RESEARCH ARTICLE

Key Points:
- A global soil data set was developed for earth system modeling
- Various data sources were harmonized using consistent processes
- Examples of the data set were given show the vertical and horizontal variations

Supporting Information:
- Readme
- Parsed Citations
- OAA Complete

Correspondence to:
Y. Dai,
yongjiudai@bnu.edu.cn

A global soil data set for earth system modeling

Wei Shangguan[1], Yongjiu Dai[1], Qingyun Duan[1], Baoyuan Liu[2], and Hua Yuan[1]

[1]College of Global Change and Earth System Science, Beijing Normal University, Beijing, China, [2]School of Geography, Beijing Normal University, Beijing, China

Abstract We developed a comprehensive, gridded Global Soil Dataset for use in Earth System Models (GSDE) and other applications. The GSDE provides soil information, such as soil particle-size distribution, organic carbon, and nutrients, and quality control information in terms of confidence level at $30'' \times 30''$ horizontal resolution and for eight vertical layers to a depth of 2.3 m. The GSDE is based on the Soil Map of the World and various regional and national soil databases, including soil attribute data and soil maps. We used a standardized data structure and data processing procedures to harmonize the data collected from various sources. We then used a soil type linkage method (i.e., taxotransfer rules) and a polygon linkage method to derive the spatial distribution of the soil properties. To aggregate the attributes of different compositions of a mapping unit, we used three mapping approaches: the area-weighting method, the dominant soil type method, and the dominant binned soil attribute method. The data set can also be aggregated to a lower resolution. In this paper, we only show the vertical and horizontal variations of sand, silt and clay contents, bulk density, and soil organic carbon as examples of the GSDE. The GSDE estimates of global soil organic carbon stock to the depths of 2.3, 1, and 0.3 m are 1922.7, 1455.4, and 720.1 Gt, respectively. This newly developed data set provides more accurate soil information and represents a step forward to advance earth system modeling.

1. Introduction

Rapid changes are occurring in our global ecosystem, and stresses on human well being, such as climate regulation and food production, are increasing [*Millennium Ecosystem Assessment*, 2005]. Soil is a major factor of the ecosystem that determines the modeling capabilities on the global and regional scale, and there is a soaring demand for up-to-date and relevant soil information [*Sanchez et al.*, 2009]. Earth System Models (ESMs) require detailed information of the physical and chemical properties of soil, such as soil color, which affects ground surface albedo, soil texture (percentages of sand, silt, and clay), organic carbon, rock fragments, bulk density, and soil nutrients (e.g., carbon (C), nitrogen (N), phosphorus (P), potassium (K), and sulfur (S)). These soil properties function as model parameters [*Dai et al.*, 2003; *Oleson et al.*, 2010], initial variables [*Thornton and Rosenbloom*, 2005] or benchmark data sets for calibration, validation, and comparison [*Todd-Brown et al.*, 2013]. However, the popular soil data sets used in modeling, including data sets by *Wilson and Henderson-Sellers* [1985], *Zöbler* [1986], *Webb et al.* [1993], *Miller and White* [1998], *Reynolds et al.* [2000], *Global Soil Data Task* [2000], and *Batjes* [2006], are often based on limited soil profiles and coarse resolution soil maps. They are outdated and should no longer be adequate for use. Currently, the most detailed world soil data set is the Harmonized World Soil Database (HWSD) [*FAO/IIASA/ISRIC/ISS-CAS/JRC*, 2009, 2012]. This database is produced by harmonizing regional soil databases and the Soil Map of the World [*FAO*, 1995, 2003] using a standardized structure. However, it contains limited soil properties and lacks coverage by local soil databases in some regions such as North America and Australia. Thus, it remains crucial to update and expand soil databases that are specifically developed to meet the needs of various types of ecosystem modeling, such as land surface modeling [*Dai et al.*, 2003; *Oleson et al.*, 2004], hydrology modeling [*Gassman et al.*, 2007], and biogeochemical modeling[*Wang et al.*, 2009]. Another effort is the production of a new fine resolution, three-dimensional grid of the functional properties of soils using digital soil mapping methods. The GlobalSoilMap.net project, comprising eight nodes worldwide, intends to lay a foundation for this effort [*Sanchez et al.*, 2009]. Specification of this project has been made [*Global Soil Map*, 2013] but only several initial products have been produced, such as the Africa soil profiles database [*Leenaars*, 2012] and the soil organic carbon map of the US [*Odgers et al.*, 2012]. It is estimated that 300 million dollars are needed to map the entire world over 5 years, while most project nodes still lack abundant financial support. So a practical solution is to utilize the regional legacy soil data available in digital formats and harmonize them into consistent global soil property maps.
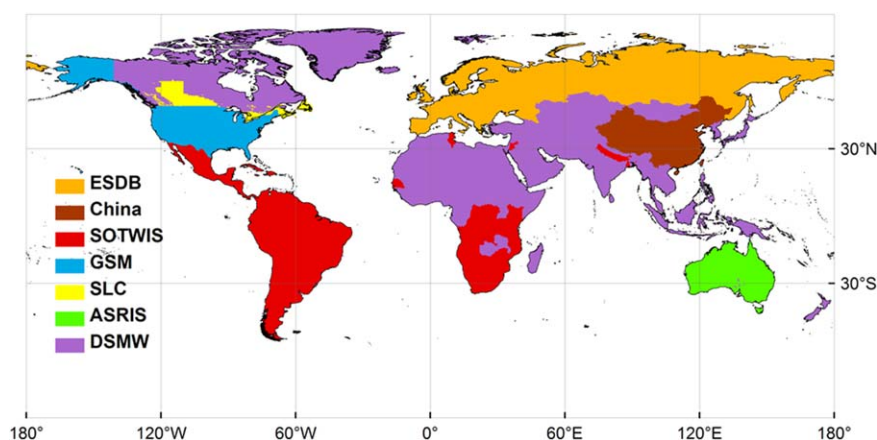
Figure 1. Data sources for the Global Soil Dataset for ESMs (GSDE). They are: ESDB (European Soil Database), GSM (General Soil Map of United States), SLC (Soil landscapes of Canada), China (The soil database of China for land surface modeling), ASRIS (the ASRIS (Australian Soil Resource Information System) polygon attributed surfaces), and SOTWISE (soil property estimates derived from the WISE and SOTER (Soil Terrain Database)).

This paper focuses on the development of a comprehensive Global Soil Dataset for ESMs (GSDE) by using a framework similar to that of the HWSD. In the HWSD, regional Soil and Terrain databases (SOTER)[*Batjes*, 2007], the European Soil Database (ESDB)[*ESB*, 2004], and the 1:1 million scale Soil Map of China[*Shi et al.*, 2004] are incorporated with the Digital Soil Map of the World (DSMW) [*FAO*, 1995, 2003] to generate a merged soil map. The soil parameters are estimated based on the World Inventory of Soil Emission Potential (WISE) profile database and the SOTER using uniform taxonomy-based pedotransfer (taxotransfer) rules and expert rules [*Batjes*, 2002, 2007]. The soil map is linked to the estimated soil parameters through either the FAO-74 or the FAO-90 soil unit symbol by soil texture classes. Where information from the FAO-74 or the FAO-90 are unavailable, soil correlation (or taxonomy reference) with FAO classifications is performed based on soil characteristics and other available classifications [*FAO/IIASA/ISRIC/ISS-CAS/JRC*, 2012]. In absence of more detailed textual data, all representative profiles are assumed to have the modal textural class of the corresponding soil unit. The HWSD contains three blocks of data: the merged soil map, general information on the soil mapping unit composition and information related to phases, and physical and chemical characteristics of topsoil (0–30 cm) and subsoil (30–100 cm).

We improve the protocol of the HWSD in the following ways. First, we focus on deriving soil property maps but do not require a FAO classification or a complete set of all soil properties in the source data sets. Therefore, soil data sets without a FAO classification or even without any classification can be utilized. Second, we use the NCSS (National Cooperative Soil Survey of United States) profile database [*NCSS*, 2012] to make more soil property maps available and to make soil types represented by more profiles. Third, we expand the spatial coverage of local soil maps, and incorporate the attribute tables linked to these maps, if there are any. Fourth, we do not require the soil correlation with an international classification where both soil map and abundant soil profiles are available in the same local classification, such as the United States [*USDA-NCSS*, 2006] and China [*Shangguan et al.*, 2012]. Thus, the linkage between the soil map and the soil profiles based on the original local soil classification can avoid the error brought by taxonomy reference, because the referencing ability is usually far below 100% [*Shi et al.*, 2010]. In addition, most local soil classifications have not been correlated with a universal classification such as the FAO legends and the World Reference Base (WRB) for Soil Resources, which are designed as tools for better correlation between national systems [*FAO*, 1998], or there is only a tentative correlation. The work of taxonomy reference also requires large efforts by experienced soil experts and is time consuming and expensive. Fifth, we used eight layers to better represent the vertical variation of soil properties. Finally, we also provide quality control information in terms of confidence level as a reference for users.

Any products should be tailored to the specific needs of end users. The products should be easily converted to different resolutions and different layers. In this paper, the data set is prepared at 30″ × 30″ resolution with eight layers to a depth of 2.3 m.
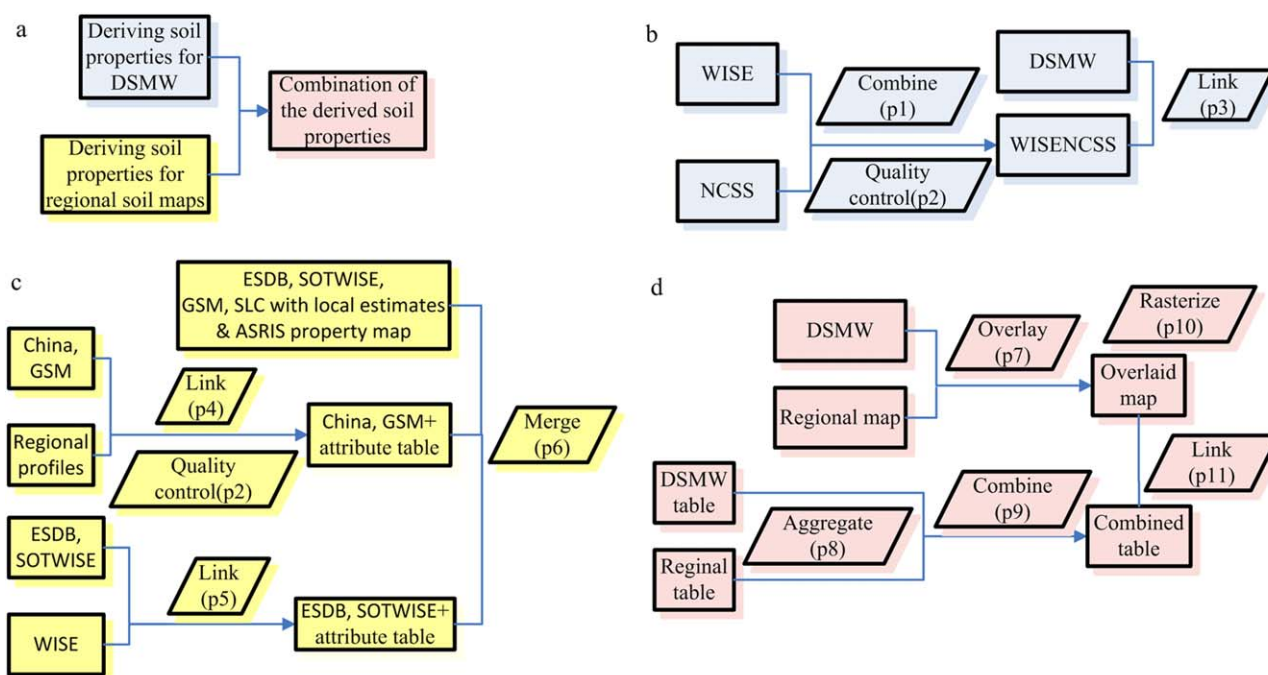
Figure 2. Data processing processes. There are (a) three major parts, (b) deriving soil properties for DSMW (blue background), (c) deriving soil properties for regional soil maps (yellow background) and (d) combination of the derived soil properties (red background). The steps p1–p11 and other details are given in the text.

2. Materials and Methodology

2.1. Data Sources

The 1:5 million scale Digital Soil Map of the World (DSMW) is used as a basic soil map, and various regional and national soil databases or soil maps are also used to compile the GSDE (Figure 1). Details about the data sources are given in the supporting information. The soil mapping units in the soil maps are composed of one or more components (Figure 3). Each component occupies a certain percentage of the mapping unit but their location is unclear. The components usually have the same soil type or the same combination of soil type and other taxonomy information such as land use and texture class. The soil data sets, which are used in the HWSD, are included, i.e., the European Soil Database (ESDB), the 1:1 million Soil Map of China, and some SOTER-derived databases (referred as SOTWIS) [*Batjes*, 2007, 2008b; *FAO/ISRIC*, 2003; *FAO/UNEP/ISRIC/CIP*, 1998]. The newly included soil data sets are the U.S. General Soil Map (GSM) [*USDA-NCSS*, 2006], the Soil Landscapes of Canada (SLC, version 3.2) [*Soil Landscapes of Canada Working Group*, 2010], the Australian Soil Resource Information System (ASRIS) polygon attributed surface [*CSIRO*, 2001a, 2001b, 2001c], the soil database of China for land surface modeling [*Shangguan et al.*, 2013], and the SOTWIS of the Indo-Gangetic Plains [*Batjes et al.*, 2004], Jordan [*Batjes et al.*, 2003], and Kenya [*Batjes and Gicheru*, 2004]. The DSMW is produced using the FAO-74 legend. The 1:1 million ESDB covers Europe and northern Eurasia with soil classification information of FAO-90. The soil database of China for land surface modeling was developed with the soil polygon linkage method based on the Genetic Soil Classification of China (GSCC) [*Shangguan et al.*, 2013]. The soil properties of the SOTWIS are based on the FAO-90 classification using WISE-derived estimates to fill the gaps in the SOTER attribute data at scale from 1:250,000 to 1:5 million. The GSM of the US at 1:250,000 scale was developed to supersede the State Soil Geographic (STATSGO) dataset using the Soil Taxonomy (ST) [*Soil Survey Staff*, 1999]. The SLC at 1:1 million scale covers the major agricultural areas of Canada (about 2,000,000 km^2) using the Canadian System of Soil Classification. The ASRIS polygon attributed surfaces, including soil thickness, bulk density, sand, silt, and clay fractions of the topsoil and subsoil, were constructed with the best available soil survey information from various state and federal agencies. There are corresponding soil property tables linked to the soil maps for the ESDB, GSM, and SLC, though the available properties are quite different and do not cover the entire soil maps.
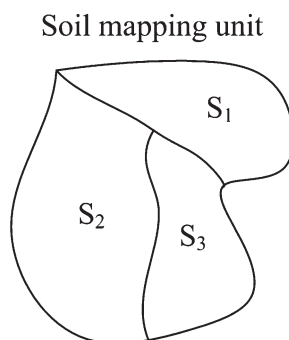
Soil mapping unit

S_1, S_2, S_3: soil types in the mapping unit.
A_1, A_2, A_3: the area percentage of the soil types.
P_1, P_2, P_3: the soil property value of the soil types.
C_1, C_2: pre-defined classes of the soil property.
V: the value assigned to the soil mapping unit.

Method A:
  $V = A_1 \times P_1 + A_2 \times P_2 + A_3 \times P_3$
Method D:
  $\because A_2 > A_3 > A_1$
  $\therefore V = P_2$
Method B:
  $\because A_1 \in C_1, A_2 \in C_2, A_3 \in C_1, A_1 + A_3 > A_2$
  $\therefore V = C_1$

Figure 3. An example of a soil mapping unit and the aggregating methods.

We used two soil profile databases, version 3.1 of the WISE [*Batjes*, 2008a] and version 2011 of the NCSS [*NCSS*, 2012], to derive the soil properties for the soil maps in this study. We combined these two profile databases in a uniform data structure (p1 of Figure 2). The WISE 3.1 holds 10,253 profiles with FAO-74 and FAO-90 legends collected worldwide from 1925 to 2005. The NCSS holds 41,218 profiles, approximately 1600 of which are collected outside the United States. The soil classification of the NCSS is the ST. After excluding soil profiles without soil classification or soil property measurement, 31,339 profiles remain. Of the 41,592 profiles, 36,638 in the WISE and NCSS have geographic coordinates. Soils in areas that have denser soil profiles tend to be better represented. Soil properties in WISE and NCSS were measured with different methods depending on the laboratories and time. Data quality of the NCSS is better because soil analyses in the NCSS were carried out in accordance with predefined procedures while soil analyses in the WISE took place in at least 150 laboratories worldwide, using a range of different methods. [*Batjes*, 2008a]. The attributes of a soil profile are not always available for each horizon, especially for deep soils. The representation of different soil classes also varies with soil attributes.

2.2. Data Processing

The GSDE was developed with a uniform data structure through a consistent data processing procedure on data sources with various formats. The linkage methods, including the soil type linkage method (i.e., taxo-transfer rules) and the polygon linkage method, were used to derive the soil property estimates of the soil type or soil polygon in the soil map. The soil type linkage method works by linking soil map units (soil types) and soil profiles according to soil type and textual class. The soil polygon linkage method works by linking soil polygons and soil profiles considering the distance between them in addition to the soil classification information. Details about the methods were described by *Shangguan et al.* [2012].

2.2.1. General Information of Soil Mapping Unit Composition and Soil Properties

The attribute tables of the soil mapping units and soil properties were prepared in a structure similar to the HWSD. General information for each soil mapping unit composition included the records identifiers, source of the record, flag for soil or nonsoil, sequence within the mapping unit, the percentage of the soil unit (or soil type), soil unit symbol according to FAO74, FAO85, and FAO90, topsoil texture class, soil drainage class, reference depth, available water storage capacity class, soil phase, obstacles to roots, impermeable layer, soil water regime, additional property relevant for agriculture use, and classification information about the Canadian System of Soil Classification (only for SLC) and the ST (only for GSM). The abundance of the above information varied with data sources. A number of the 34 soil properties considered in this study are given in Table 1. An eight standard layer scheme [0–0.045, 0.045–0.091, 0.091–0.166, 0.166–0.289, 0.289–0.493, 0.493–0.829, 0.829–1.383, and 1.383–2.296 m] was used to retain the vertical variation of the soil properties [*Shangguan et al.*, 2013]. This layer scheme is modified from the Common Land Model (CoLM) [*Dai et al.*,

Table 1. List of Soil Properties Presented in the Data Sources

Soil Property	Unit	Soil Data Set[a]							
		WISE	NCSS	ESDB	GSM	SLC	China	ASRIS	SOTWISE
Total carbon	%		√						
Organic carbon	%	√	√	√	√		√		√
Total nitrogen	%	√	√				√		√
Total sulfur	%		√						
Calcium carbonate content	%	√	√		√	√			√
Gypsum content	%	√	√		√				√
pH, measured in water	none	√	√	√	√	√	√		√
pH, measured in KCL solution	none	√	√	√					
pH, measured in CaCl$_2$ solution	none	√	√	√	√	√			
Electrical conductivity	ds/m or mmho/cm	√	√		√	√			√
Exchangeable calcium	cmol/kg	√	√				√		
Exchangeable magnesium	cmol/kg	√	√				√		
Exchangeable sodium	cmol/kg	√	√				√		
Exchangeable potassium	cmol/kg	√	√				√		
Exchangeable aluminum	cmol/kg	√	√				√		
Exchangeable acidity	cmol/kg	√					√		
Cation exchange capacity	cmol/kg	√	√		√	√	√		√
Base saturation, expressed as % of CEC	%	√	√			√			√
Sand content	%	√	√	√	√	√	√	√	√
Silt content	%	√	√	√	√	√	√	√	√
Clay content	%	√	√	√	√	√	√	√	√
Gravel content	% in volume	√	√	√	√	√	√		√
Bulk density	g/cm^3	√	√	√	√	√	√	√	
Volumetric water content at −10 kPa	%	√	√		√	√			√
Volumetric water content at −33 kPa	%	√	√		√	√			√
Volumetric water content at −1500 kPa	%	√	√		√	√			√
The amount of phosphorous using the Bray1 method	ppm		√						
The amount of phosphorous by Olsen method	ppm		√						
Phosphorous retention by New Zealand method	%		√						
The amount of water soluble phosphorous	ppm		√						
The amount of phosphorous by Mehlich method	ppm		√						
Exchangeable sodium percentage	%		√						
Total phosphorus	%		√				√		
Total potassium	%		√				√		

[a]The soil data sets are: WISE (the World Inventory of Soil Emission Potential profile database), NCSS (the National Cooperative Soil Survey of United States profile database), ESDB (European Soil Database), GSM (General Soil Map of United States), SLC (Soil landscapes of Canada), China (The soil database of China for land surface modeling), ASRIS (the ASRIS (Australian Soil Resource Information System) polygon attributed surfaces), and SOTWISE (soil property estimates derived from the WISE and SOTER (Soil Terrain Database)).

2003] and the Community Land Model (CLM) [*Oleson et al.*, 2004]. The standardization of the soil layers were accomplished using the equal-area quadratic smoothing spline functions [*Bishop et al.*, 1999].

2.2.2. Overall Process

The data processing procedures are shown in Figure 2. We first describe the process and then give details of the steps in the subsequent sections. Step 1 to step 11 correspond to p1 to p11 in Figure 2.

Step 1: The WISE and NCSS soil profile databases were combined as one uniform database.

Step 2: The soil profiles included in the WISE, NCSS, and regional soil databases were quality controlled before the linkage between the soil maps and soil profiles was implemented.

Step 3: We derived soil parameter estimates for the DSMW using the WISE and NCSS by the linkage method based on the FAO74 legends. As a result, the spatial distribution of the soil properties with quality control information (QC) was derived for the DSMW.

Step 4: For China and the US, soil parameter estimates were derived using regional soil profile databases and regional soil maps by the linkage method based on the regional soil classification systems (i.e., GSCC and ST).

Step 5: For the ESDB and SOTWISE areas, soil parameter estimates were derived using the WISE profiles and regional soil maps by the linkage method based on the FAO90 legends.

Step 6: The three types of soil data sets were merged according to their priority described in section 2.2.5. As a result, the spatial distribution of the soil properties with quality control information (QC) was derived for the merged regional map.

Step 7: The DSMW and the merged regional soil map were overlaid.

Step 8: The derived soil property tables of the DSMW and the regional soil maps were aggregated by soil map units separately. As a result, each soil map unit has unique values of soil properties.

Step 9: The aggregated tables of the DSMW and the regional soils were combined.

Step 10: The overlaid soil map was rasterized.

Step 11: The overlaid soil map and the combined table were linked according to the linkage implemented in step 3, step 4, and step 5.

2.2.3. Quality Control

All profile data were quality controlled manually or automatically for possible inconsistencies or errors (p2 of Figure 2). These procedures were adopted from the WISE [*Batjes*, 2008a]. The following were included in the quality control. Values out of the reasonable range of a property were excluded and extreme values within the range were checked. Soil properties were converted to the uniform units in Table 1. Soil organic matter in the source data sets were converted to soil organic carbon by a factor of 0.58 [*Pribyl*, 2010]. The depth, boundary, and sequence of the soil horizons were checked and corrected. Soil horizons with a known top depth and an unknown bottom depth were assumed to have a 0.3 m horizon depth. If the sum of sand, silt, and clay fractions was <98% or >102%, these values were excluded; otherwise, they were normalized to 100%. When there are two or more measurements of the same attribute using different analytic methods, measurements by the standard preparations were preferred [*NRCS*, 2004]. In addition, the unrealistic bulk densities for soils with an organic carbon content >12% were corrected using regression models [*Hiederer and Köchy*, 2012].

2.2.4. Deriving Soil Properties for DSMW

The NCSS and WISE were used to derive the soil parameter estimates for the DSMW (p3 of Figure 2). The NCSS profiles with a ST classification were correlated to the 26 major soil groups of the FAO74 according to a tentative approximation [*Batjes*, 2011; *Spaargaren and Batjes*, 1995]. The use of the NCSS made more soil properties available, which were not included in the WISE, and improved the representation of the major soil groups of the FAO74. The topsoil texture (coarse, medium, or fine) of a profile was calculated according to the sand, silt, and clay fractions. If there was no data on the soil fractions, the texture was assumed to be medium. The soil properties by soil unit, topsoil texture class, and depth zone were estimated following taxotransfer rules (i.e., the soil type linkage method) based on the FAO74 legend. Assumptions were used to fulfill the soil properties for two nonsoil map units of the FAO74 [*Batjes*, 2006], and no assumptions were made to fulfill the other nonsoil map units due to lack of credible evidence. For dune sand, soil parameters of Cambic Arenosols with coarse texture were used as the default, except that organic carbon content was set to 0.2% for layers 1–4 and 0.1% for layers 5–8; the content of sand, silt, and clay were set to 98%, 1%, and 1%, respectively. For salt flat, the soil parameters of Orthic Solonchaks with medium texture were used as the default.

2.2.5. Deriving Soil Properties for Regional Soil Maps

The standard procedure for deriving soil properties of the regional soil maps includes three stages:

1. Collate the regional soil property maps or local attribute estimates that have already been linked to the soil map in the uniform GSDE format;

2. Use the regional soil profile database to derive the soil properties of the regional soil map by the linkage method based on the regional soil classification system (p4 of Figure 2);

3. Use the WISE profile database to derive the soil properties of the regional soil map by the linkage method based on the FAO legend (p5 of Figure 2).

The data derived by these stages were merged, and only one of them was assigned to represent the soil property at a location. The priority was as follows: stage 1, stage 2, and stage 3. The gaps in the data derived by stage 1 were filled by the data derived by stage 2 or stage 3. In stage 2, it requires abundant soil profiles in the local soil classification to well represent the soil types. Table 1 shows the soil properties provided by each stage for each regional soil map. For the ESDB area, the SPADE-2 (Soil Profile Analytical Database for Europe Version 2.0) [*Hannam et al.*, 2009] of the ESDB covering 19 West and Middle Europe countries were used as the stage 1 data set, and the gaps were filled with the stage 3 data derived from the 1:1 million Soil

Table 2. Codes for the Soil Classification Levels of Different Data Sources

Data Source	Soil Classification	Code
DSMW, ESDB, SOTWISE	FAO74, FAO90	1: soil unit; 2: major soil group
China	Genetic Soil Classification of China	1: family; 2: subgroup; 3: great group; 4: order; 5: (non)acid
US	Soil Taxonomy	1: subgroup; 2: great group; 3: suborder; 4: order

Geographic Database of Europe (SGDBE) and the WISE using the type linkage method based on the FAO-90 classification. Because there are different soil properties for different land uses with the same soil type in a soil map unit component, the average of them was assigned to the component. For the SOWISE regions, the SOWISE itself provided the stage 1 data set, and the gaps were filled with the stage 3 data derived from the WISE using the type linkage method based on the FAO-90 classification. For the United States, the tabular tables of the GSM produced from the National Soil Information System (NASIS) were used as the stage 1 data set, and the gaps were filled with the stage 2 data derived from the NCSS using the type linkage method based on the ST classification. For the convenience of the linkage in stage 2, the classification in the NCSS and GSM were corrected and supplemented, and the texture classes were combined into 11 classes. For Canada, the SLC3.2 was used as the stage 1 data, but no attempts were made to link the soil map with the WISE and NCSS because the Canadian System of Soil Classification has not been well-correlated to the FAO legend. For Australia, the ASRIS polygon attributed surfaces were used as the stage 1 data, but soil property mapping based on the regional soil maps and soil profile database was not performed because the soil profile database was not downloadable and the percentages of the map unit compositions were not available. For China, the stage 2 data set was derived from the 1:1 million Soil Map of China and 8,979 profiles in China using the polygon linkage method based on the GSCC classification [*Shangguan et al.*, 2013].

2.2.6. Merging Regional Soil Maps
All original soil maps were available in vector format except the ASRIS polygon attributed surfaces, which were presented as raster format, though the original data were in vector format. The merging of the regional soil maps was performed in an ESRI ArcGIS environment (p6 of Figure 2) and included the following steps:

1. If necessary, the original soil maps were converted to geographic coordinates (i.e., longitude and latitude).

2. The ASRIS soil properties maps were converted to the vector format and overlaid. For each polygon, an identifier was assigned.

3. One of the source regional maps was used to represent the corresponding area. The priority was as follows: US, ESDB, China, Canada, ASRIS, and SOTWIS (Figure 1).

2.2.7. Combination of the Derived Soil Properties
The soil maps and derived soil attribute tables were combined separately. The merged regional soil map and the DSMW were overlaid (p7 of Figure 2). Each mapping unit in the overlaid map was identified by a unique ID. The overlaid map can be rasterized into 30″ using the unique ID (p10 of Figure 2). The soil attribute tables for the DSMW and the merged regional soil map were processed by the three aggregation methods (p8 of Figure 2, see section 2.4) and were combined (p9 of Figure 2). The soil property derived for the DSMW was used as a basic data set, and it was replaced by the regional data where available. Because the availability of soil properties is different, the coverage of the basic data and the regional data vary. The combined soil attribute tables and the overlaid soil map (vector or raster map) can be linked through the soil mapping unit identifier (p11 of Figure 2).

2.3. Quality Control Information
Quality control information was offered to provide "confidence" information on the derived soil parameters through sample size, soil classification level of mapping unit, soil classification level of the linkage, search radius, and texture consideration [*Shangguan et al.*, 2013, Table 2]. In most cases, the QC value has six digits representing the above indexes (only for derived data by the linkage method). The codes of the digits are given in Table 2 of *Shangguan et al.* [2013]. However, the codes of the soil classification level are different
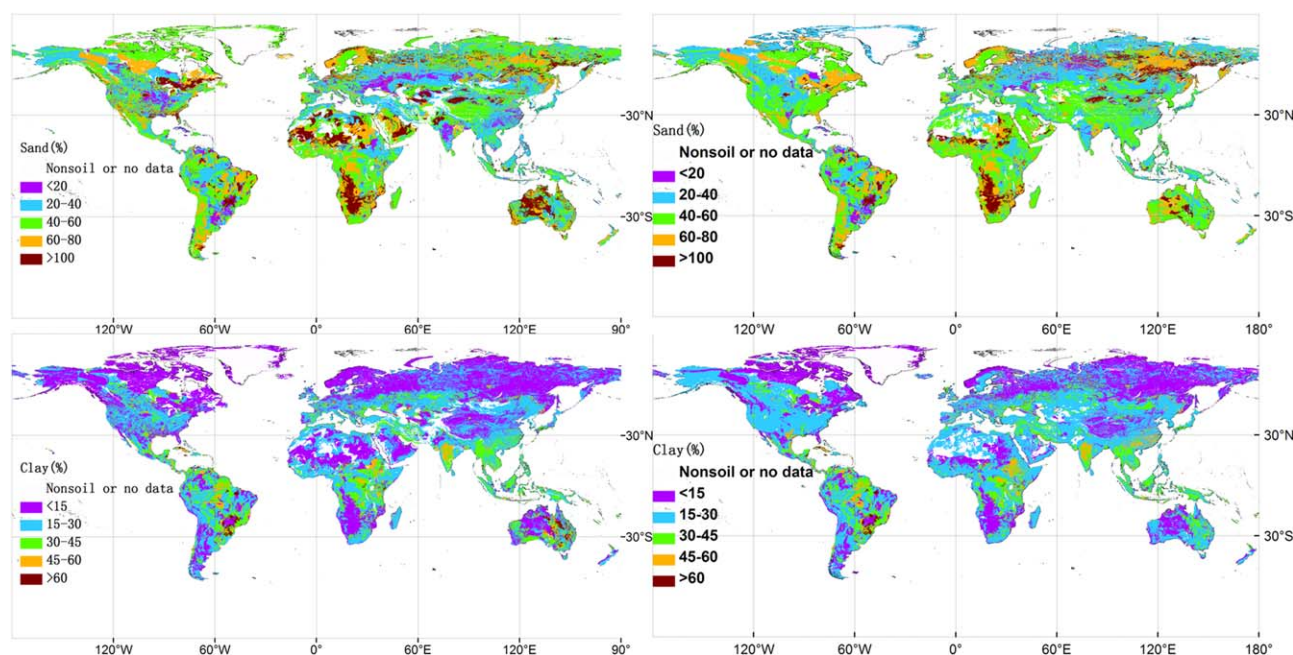
Figure 4. The geographic distribution of sand and clay content (%) of topsoil. (left) The GSDE and (right) HWSD.

according to the soil classification used in the data sources (Table 2). If the QC value is 1, the corresponding attribute is from the local soil database. If the QC value is 2, the corresponding attribute is from the local soil property map (only for Australia). If the QC value is 3, the corresponding attribute (only for bulk density) is corrected according to the soil organic carbon. If the QC value is 0 or 11, the corresponding attribute is unavailable as it is a nonsoil or no data are available. When the data derived for the DSMW is used, an additional digit "1" is added to the beginning of the QC number to indicate the use of the basic map.

2.4. Soil Property Mapping Approaches

Through the data processing, the soil attribute databases for the DSMW and the merged regional soil maps were produced. Because a soil mapping unit is typically composed of more than one component for most databases except China and Australia, there is a one to n relation between a soil mapping unit (or a soil map polygon or a grid cell) and the attributes. This makes mapping the complete range of attributes characterizing a mapping unit a nontrivial task. The users can use the derived attribute table directly and relate a soil mapping unit with its compositional soil property information like the subgrid hierarchy (referred as the subgrid method) in land surface models [*Oleson et al.*, 2010]. However, a "one to one" relation between the mapping units or a grid cell and soil properties is needed in ESMs and many other applications. Three mapping approaches were adopted to aggregate the attributes of different compositions of a mapping unit (p8 of Figure 2). The three methods were the area-weighting method (Method A), the dominant soil type method (Method D), and the dominant binned soil attribute method (Method B). Method A uses area-weighting of a soil attribute considering all the mapping unit compositions to represent the mapping unit. Method D uses the soil attribute of the dominant soil type only. Method B takes the full mapping unit compositions into account and classified the soil attribute into several preselected classes, and the spatially dominant class is assigned to the mapping unit. The preselected classes were specified taking the standard by *FAO/IIASA/ISRIC/ISS-CAS/JRC* [2012] and *Batjes* [2006] as a reference with some modifications. The results of Method A and Method D can be classified into the same predefined classes in Method B, making them comparable. Users can define their own attribute classes if necessary. Figure 3 shows an example of a soil mapping unit and how the aggregating methods assign value to the mapping unit. For quantitative data, we tested all three methods but offered the results of Method A as the product. For categorical data, we adopted Method B for simplicity. However, it is possible to generate multiple spatial layers (each layer

Table 3. Comparison of the Area Percentage of Soil Properties of Our Data Set (GSDE) With the HWSD

Attribute	Data Set	Percentage of Soil Area in Each Class in Topsoil (0–0.3 m)					
		Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Sand (%)	Range	<20	20–40	40–60	60–80	>80	
	HWSD	3.0	30.8	41.9	16.9	7.4	
	GSDE	5.6	32.0	33.5	17.6	11.3	
Silt (%)	Range	<15	15–30	30–45	45–60	>60	
	HWSD	15.2	37.1	37.3	9.1	1.3	
	GSDE	20.4	33.0	32.6	13.1	0.9	
Clay (%)	Range	<15	15–30	30–45	45–60	>60	
	HWSD	27.2	54.2	13.8	4.3	0.5	
	GSDE	36.2	43.8	15.1	3.9	1.0	
SOC (%)	Range	<0.2	0.2–0.6	0.6–1.2	1.2–2	>2	
	HWSD	0.3	16.2	41.3	24.8	17.3	
	GSDE	3.6	23.3	31.0	19.9	22.1	
BD (g/cm^3)	Range	<0.4	0.4–0.9	0.9–1.2	1.2–1.4	1.4–1.6	>1.6
	HWSD	1.2	1.4	7.9	53.4	35.8	0.3
	GSDE	0.6	3.2	12.2	47.5	35.3	1.3

corresponding to one class) for each of the categorical attributes. Comparison between the aggregating methods is given in the supporting information.

2.5. Conversion in Resolutions

The current data set was provided at 30″ resolution. If a lower resolution is needed, the aggregation methods A and B can be used. Method D can be used in rasterizing the vector map. There are some up-scaling methods such as the window median, variability-weighted methods [*Wang et al.*, 2004] and complicated methods such as variograms [*Oz et al.*, 2002] and fractal theory [*Quattrochi et al.*, 2001]. The data set can also be prepared in different vertical layer schemes according to the ESM application requirements.

3. Results and Discussion

3.1. Examples of GSDE Data Set

Here we use some soil properties as examples (here referred to soil basic properties), including sand, silt, and clay content, bulk density and soil organic carbon, which are usually needed to estimate soil hydraulic properties in ESMs [*Dai et al.*, 2013]. For the convenience of comparison with the HWSD, we prepared the results for the topsoil (0–0.3 m) using depth weighting. It should be noted that there were some areas with no data in western Asia and Mongolia because soil properties are not available below layer 2 and were taken as nonsoil in the depth weighting. The GSDE and the HWSD were similar in some aspects because part of the source data was the same. The full dataset is given in the supporting information.

3.1.1. Sand, Silt, and Clay Content

Sand, silt, and clay content or particle-size distribution (PSD) is used as the only input of soil in most LSMs, except soil color or albedo. The PSD is used to the derive soil hydraulic parameters and other soil basic properties such as bulk density and porosity, in LSMs. The PSD impacts the redistribution of both soil water and energy.

Figure 4 shows the sand and clay content of the topsoil in the GSDE and the HWSD. In the GSDE, the areas with high sand content were found in the Africa, Canada, Australia, Central Asia, and the Arabian Peninsula, most of which are deserts. Usually, areas with high sand content had low clay content, and vice versa. High clay content appeared in the South America, Central Africa, India, and East Australia. Compared to the HWSD, the GSDE gave a higher estimation of sand content in Canada, Australia and the ESDB area while it gave a lower estimation in the middle of the US and East Asia. The GSDE had a higher value of clay content in the east of Australia, and a lower value in North America. Table 3 shows that the GSDE had a more even distribution of sand, silt, and clay content than the HWSD, i.e., the GSDE gave more extreme values. For sand content, the GSDE had much less area of soils with a medium value (i.e., class 3, 40%–60%) and had more soils in other classes. For silt content, the GSDE had less area of soils belonging to classes 2 and 3. For clay content, the GSDE had much less area of soils belonging to class 2. As sand, silt, and clay content are
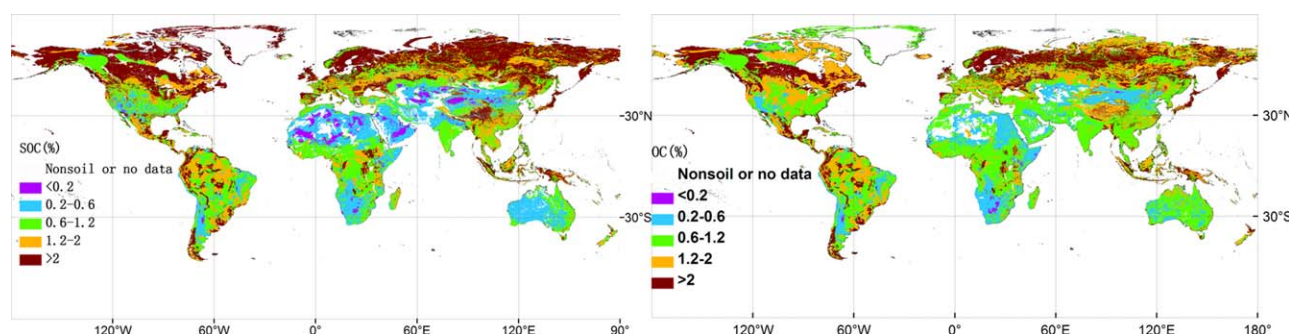
Figure 5. The geographic distribution of soil organic carbon (SOC, %) of topsoil. (left) The GSDE and (right) HWSD.

typically used to estimate soil hydraulic properties in LSMs, the LSMs are expected to give more extreme cases of soil water conditions if the GSDE is used as input instead of the HWSD.

3.1.2. Soil Organic Carbon

The soil organic carbon (SOC) increases the water-holding capacity of soil. The SOC is also important to plant growth and is a major factor in the overall health of soil. Usually, the SOC is not measured directly in a soil survey but is estimated from the soil organic matter [*Pribyl*, 2010].

Figure 5 shows the SOC of the topsoil in the GSDE and HWSD. In the GSDE, high SOC was primarily found in the high latitudes of the Northern Hemisphere. Low SOC was found in the desert areas. Compared to the HWSD, the GSDE had higher values in the high latitudes of the Northern Hemisphere and the Qinghai-Tibet plateau, but lower values in the Middle East, North Africa, and Australia. Table 3 shows that the GSDE had less values in classes 3 and 4, and higher values in other classes. The GSDE gave a lower estimation of the SOC for the topsoil on the whole, compared to the HWSD. As soils with SOC <0.6% are considered to be poor, the estimation of the GSDE indicates that approximately 26.9% soils were unhealthy, while this figure in the HWSD was 16.5%. However, the GSDE had more soils with a SOC >2%.
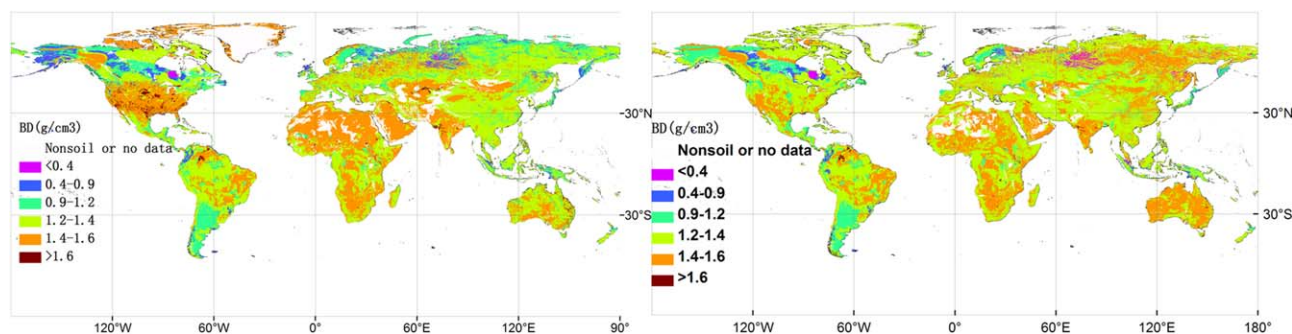
3.1.3. Bulk Density

The soil bulk density (BD) is the mass of soil material divided by the total volume of solids and pores. Because of lack of data, the BD is usually estimated from the sand, silt, and clay contents in LSMs. The BD is inversely related to the porosity of a soil. The BD impacts the soil water retention properties and soil hydraulic conductivity.

Figure 6 shows the BD of the topsoil in the GSDE and the HWSD. In the GSDE, high BD appeared in the US, and low BD appeared in the high latitudes of the Northern Hemisphere, North Africa, and the Middle East. Compared to the HWSD, the GSDE had more areas with a high BD and had a lower BD in the high latitudes of the Northern Hemisphere. Table 3 shows that the GSDE had more soils in classes 2, 3, and 6. 3.8% area of soils had a value lower than 0.9 g/cm^3 in the GSDE, which are usually organic soils, compared to 2.6% in the HWSD.

3.1.4. Quality Control Information

The data quality may vary with the depth and soil properties. For the same soil taxonomic class, the deeper the soil is, the poorer the data quality. This is because there are fewer samples in the deeper soils. Soil properties with more available samples tend to have better quality, such as sand, clay and SOC contents. Figure 7 shows the quality control information of the SOC in the topsoil. The first three digits of the QC (the first four if the data source is the DSMW) are shown, which are the most important QC indexes. The meaning of the digits is given by *Shangguan et al.* [2013]. The first digit (the linkage level) represents the soil classification level at which the linkage is performed (Table 3). The second digit (texture consideration) represents whether the soil texture is considered in the linkage. The third digit (the sample size level) represents the number of soil profiles used to represent a soil map unit or a soil polygon. For some areas of Canada and the ESDB region, and most areas of the US and SOTWIS region, the QC value was 1, and the data sources were linked with the local soil database (stage 1). For most areas of the ESDB region, some areas of the US and the SOTWIS region, and China, the QC was between 100 and 1000, and the data were derived using

Figure 6. The geographic distribution of bulk density (BD, g/cm^3) of topsoil. (left) The GSDE and (right) HWSD.

the regional soil maps (stage 2 or 3). For the rest of the world, the QC was above 1000, and the data were derived using the DSMW and the WISE and NCSS profile databases. For the stage 3 data of the ESDB and SOTWIS region, the linkage was primarily accomplished by soil unit. For the stage 2 data of the US, the linkage was primarily accomplished by soil subgroup. Most of the stage 2 and stage 3 data were derived by taking soil texture into account. For China, data with poor quality appeared in the northwest [*Shangguan et al.*, 2013]. Most of the stage 2 and stage 3 data were derived using abundant soil profiles at the corresponding linkage level. The quality of the DSMW data was rather poor, and almost half of these areas were derived by FAO major soil group, which need to be updated in the future.

3.2. Estimation of SOC Stocks

The SOC density (t ha^{-1}) for a given depth is estimated by the following equation:

$$SOCD = SOC \times BD \times \left(1 - \frac{GRAV}{100}\right) \times Dep \times 10^2 \tag{1}$$

where *SOC* is the soil organic carbon content for the given depth (%), *BD* is the bulk density (g cm^{-3}), *GRAV* is the gravel content (%), and *Dep* is the depth of a given soil layer (m).

Because there were multiple soil components in a map unit, only the mean values from all soil components can be assigned to a grid cell. Two approaches to aggregate the data can be distinguished: aggregating the input soil properties (i.e., SOC, BD, and GRAV) into a grid cell first and then computing the SOC density or computing the SOC density and then aggregating (here referred as "aggregating first" and "aggregating after"). We used the area-weighting method because it conserves the total carbon mass. Figure 8 shows the resulting SOC density to the depth of 1 m by the two approaches. The SOC density is higher in the high
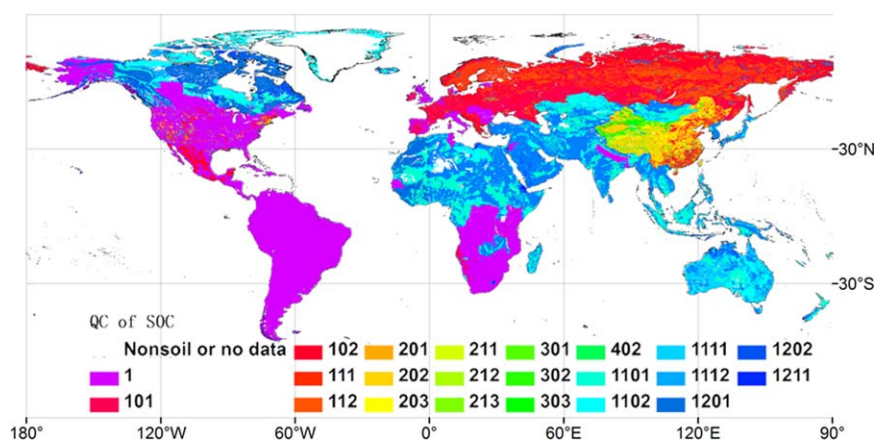
Figure 7. Quality control information (QC) of soil organic carbon (SOC) of topsoil (0–0.3 m). Smaller the QC, better the data quality.
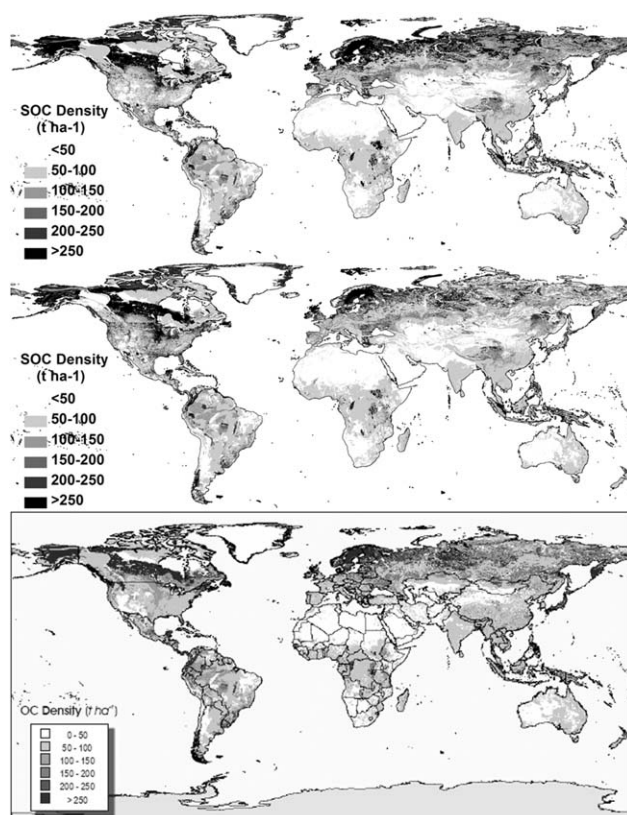
Figure 8. The SOC density to the depth of 1 m (top) by the "aggregating first" approach, (middle) by the "aggregating after" approach, and (bottom) by [*Hiederer and Köchy*, 2012] using the HWSD.

latitudes of the Northern Hemisphere and lower in the arid and semiarid areas in the midlatitudes. The "aggregating first" approach results in higher estimates than the "aggregating after" approach in the high latitudes of the Northern Hemisphere. However, opposite results are obtained in the US. The SOC stock to the depth of 2.3, 1, and 0.3 m are 2525.7, 1876.9, and 841.3 Gt, respectively, from the "aggregating first" approach, and they are 1922.7, 1455.4, and 720.1 Gt, respectively, from the "aggregating after" approach. The "aggregating after" approach, which first calculates the SOC density for each component of a map unit, ensures that there are no biases in the process of integration. In contrast, the "aggregating first" approach results in a significant overestimation, and so it is not regarded as a proper approach. The SOC stock to the depth of 1m from previous studies ranged from 991 to 1,849 Gt

[*Hiederer and Köchy*, 2012]. The estimation by the "aggregating after" approach is close to the results of *Hiederer and Köchy* [2012] (1,417 Gt) using the amended HWSD. However, *Hiederer and Köchy* [2012] used the "aggregating first" approach and produced a different spatial distribution (Figure 8). A defect in the SOC stock estimations is that they were based on the observed soil profile depth instead of the actual soil depth or the depth to bedrock. This led to an underestimation of the soil SOC stocks, especially for the deep soil layers. However, the deep soils contain very low SOCs.

3.3. Deriving Soil Hydraulic Data

The basic soil properties, such as sand, silt and clay content, bulk density, and soil organic carbon, are typically used to derive the soil hydraulic parameters required by land surface models with Pedo-Transfer Functions (PTFs) [*Shangguan et al.*, 2013]. A question arises when there are multiple components in a mapping unit. Is it better to aggregate the PTF results calculated for each mapping unit component or to aggregate the basic soil properties before deriving the soil hydraulic parameters? *Reybold and Tesselle* [1989] first calculated the soil available water content for each soil type and then aggregated the results for a mapping unit. *Webb et al.* [1993] first combined the soil basic properties of 106 soil units with the soil type map of *Zöbler* [1986] first and then calculated the soil hydraulic parameters. *Romanowicz et al.* [2005] suggested that the derived soil properties should be calculated before aggregation based on a case study. Because the relationship between the soil basic properties and the derived soil hydraulic parameters is nonlinear, the better way is to first calculate the soil hydraulic parameters and then aggregate them. However, the aggregation, which is needed in both ways, will filter out the differences of soil properties between soil components within a given mapping unit [*Odgers et al.*, 2012]. A possible solution is to avoid aggregation by using a subgrid structure (scheme) for the soil in models [*Dai et al.*, 2003], i.e., the soil hydrological processes are implemented over each subgrid soil column within a grid instead of the entire model grid. However, this will make the model structure more complicated because the subgrid of the soil column will be

overlaid with the land unit subgrid. Another possible solution is to incorporate the spatial disaggregation of the soil map to determine the location of the mapping unit components [*Thompson et al.*, 2010]. Soil-landscape patterns and relationships are needed to develop the spatial disaggregation rules. Because there are not abundant soil profile data available, the three aggregation methods are currently the most pragmatic.

4. Conclusions

The GSDE represents a step forward in providing a more realistic data set of the physical and chemical properties of global soils. This data set can also be used with empirical approaches (such as PTFs) to determine the secondary soil hydraulic and biogeochemical parameters that can be directly used by ESMs. The data set is available online at http://globalchange.bnu.edu.cn.

The characteristics of the GSDE are as follows:

1. The GSDE does not require a reference soil classification (such as the FAO legend). Instead, the GSDE uses the soil data with a local soil classification in the cases of China, the US, Canada, and Australia.

2. The GSDE incorporated an enriched soil profile database including 41,592 profiles of the WISE and NCSS, and more soil properties.

3. The GSDE uses more regional soil maps and attribute data including those from China, the US, Canada, Australia, Europe, and the SOTWIS.

4. The GSDE builds the linkage between soil maps and soil profiles based on the local soil classification for China and the US, avoiding errors that would be induced by taxonomy reference.

5. The GSDE has eight layers to better represent vertical variations.

6. The GSDE provides quality control information in terms of confidence level.

However, the GSDE has several limitations: (1) we did not attempt to convert the measurements through different analytical methods into a uniform standard; (2) we only provide a representative value (mean or median) of the soil properties for a specific soil type, not their ranges or standard deviations; (3) the quality of the data set is spatially uneven (Figure 7), depending on the availability of the source data; and (4) there exist uncertainties associated with the linkage method in deriving soil property maps [*Shangguan et al.*, 2012].

There are ongoing projects to develop high-resolution soil data sets, e.g., "A New Digital Soil Map of the World" by the global soil map project (http://www.globalsoilmap.net/). This project aims to provide a new digitized soil map of the world using the state of the art at a finer resolution (90 m). However, the time-consuming and expensive field surveys are not affordable in the near future. There are also some regional and national soil databases to be incorporated, such as the ASRIS and the Digital Soil Map data set of China at 1:50,000 scale [*Zhang et al.*, 2010]. The data set needs to be incorporated into the ESMs or LSMs to see how it influences the modeling results and can be used as a benchmark data set for modeling.

References

Batjes, N. H. (2002), Soil parameter estimates for the soil types of the world for use in global and regional modelling (version 2.1), *ISRIC Rep. 2002/02c*, Int. Food Policy Res. Inst. (IFPRI) and Int. Soil Ref. and Inf. Cent. (ISRIC), Wageningen, Netherlands.
Batjes, N. H. (2006), ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid, *Rep. 2006/02*, ISRIC World Soil Inf., Wageningen, Netherlands (with data set).
Batjes, N. H. (2007), SOTER-based soil parameter estimates for Central Africa: DR of Congo, Burundi and Rwanda (SOTWIScaf, version 1.0), Rep. 2007/02, ISRIC, World Soil Inf., Wageningen, Netherlands.
Batjes, N. H. (2008a), ISRIC-WISE harmonized global soil profile dataset (version 3.1), Rep. 2008/02, ISRIC, World Soil Inf., Wageningen, Netherlands.
Batjes, N. H. (2008b), SOTER parameter estimates for Senegal and The Gambia derived from SOTER and WISE (SOTWIS-Senegal, version 1.0), Rep. 2008/05, ISRIC, World Soil Inf., Wageningen, Netherlands.
Batjes, N. H. (2011), Overview of soil phosphorus data from a large international soil database, Rep. 2011/01, Plant Res. Int. (PRI), Wageningen UR, and ISRIC, World Soil Inf., Wageningen, Netherlands.
Batjes, N. H., and P. Gicheru (2004), Soil data derived from SOTER for studies of carbon stocks and change in Kenya (version 1.0), Rep. 2004/01, ISRIC, World Soil Inf., Wageningen, Netherlands.
Batjes, N. H., Z. Rawajfih, and R. Al-Adamat (2003), Soil data derived from SOTER for studies of carbon stocks and change in Jordan (version 1.0), Rep. 2003/04, ISRIC, World Soil Inf., Wageningen, Netherlands.

Batjes, N. H., T. Bhattacharyya, C. Mandal, K. Dijkshoorn, D. K. Pal, E. Milne, and K. S. Gajbhiye (2004), Soil data derived from SOTER for studies of carbon stocks and change in the Indo-Gangetic Plains (India) (version 1.0; GEFSOC Project), *Rep. 2004/06*, ISRIC, World Soil Inf., Wageningen, Netherlands.

Bishop, T. F. A., A. B. McBratney, and G. M. Laslett (1999), Modelling soil attribute depth functions with equal-area quadratic smoothing splines, *Geoderma*, *91*, 27–45.

CSIRO (2001a), *Soil Bulk Density for Australian Areas of Intensive Agriculture of Layer 1 and 2 (Derived from Soil Mapping)*, CSIRO, Land and Water, Canberra, Australia.

CSIRO (2001b), *Soil Thickness for Australian Areas of Intensive Agriculture of Layer 1 and 2 (Derived from Soil Mapping)*, CSIRO, Land and Water, Canberra, Australia.

CSIRO (2001c), *Soil Sand, Silt and Clay Content for Australian Areas of Intensive Agriculture of Layer 1 and 2 (Derived from Soil Mapping)*, CSIRO, Land and Water, Canberra, Australia.

Dai, Y., et al. (2003), The common land model, *Bull. Am. Meteorol. Soc.*, *84*, 1013–1023.

Dai, Y., W. Shangguan, Q. Duan, B. Liu, S. Fu, and G. Niu (2013), Development of a China dataset of soil hydraulic parameters using pedotransfer functions for land surface modeling, *J. Hydrometeorol.*, *14*, 869–887.

ESB (2004), European Soil Database (version 2.0), European Comm., JRC Inst. for Environ. and Sustain., Ispra, Italy.

FAO (1995, 2003), *The Digitized Soil Map of the World Including Derived Soil Properties* (version 3.6), FAO, Rome, Italy.

FAO (1998), *World Reference Base for Soil Resources*, FAO, Rome, Italy.

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

FAO/IIASA/ISRIC/ISS-CAS/JRC (2012), *Harmonized World Soil Database (ver. 1.2)*, FAO, Rome, Italy.

FAO/ISRIC (2003), *Soil and Terrain Database for Southern Africa*, FAO, Rome, Italy.

FAO/UNEP/ISRIC/CIP (1998), *Soil and Terrain Digital Database for Latin America and the Caribbean at 1:5 Million Scale*, FAO, Rome, Italy.

Gassman, P. W., M. R. Reyes, C. H. Green, and J. G. Arnold (2007), The soil and water assessment tool: Historical development, applications, and future research directions, *Trans. ASABE*, *50*(4), 1211–2150.

Global Soil Data Task (2000), Global Soil Data Products CD-ROM (IGBP-DIS), International Geosphere-Biosphere Programme—Data and Information Services, Available online at the ORNL Distributed Active Archive Center, Oak Ridge Natl. Lab., Oak Ridge, Tenn.

Global Soil Map (2013), Specifications for GlobalSoilMap products, Release 2.3, USDA NRCS, Lincoln, Nebr.

Hannam, J. A., J. M. Hollis, R. J. A. Jones, P. H. Bellamy, S. E. Hayes, A. Holden, M. H. Van Liedekerke, and L. Montanarella (2009), *SPADE-2: The soil profile analytical database for Europe, version 2.0, Beta Version March 2009*, unpublished Report, 27 pp.

Hiederer, R., and M. Köchy (2012), *Global Soil Organic Carbon Estimates and the Harmonized World Soil Database*, 79 pp., Publ. Off. of the Eur. Union, Luxembourg.

Leenaars, J. G. B. (2012), Africa Soil Profiles Database, Version 1.0. A compilation of geo-referenced and standardized legacy soil profile data for Sub Saharan Africa (with dataset), *ISRIC Rep. 2012/03*, Afr. Soil Inf. Ser. (AfSIS) Proj. and ISRIC - World Soil Inf., Wageningen, Netherlands.

Millennium Ecosystem Assessment (2005), *Ecosystems and Human Well-Being: A Framework for Assessment*, Island Press, Washington, D. C.

Miller, D. A., and R. A. White (1998), A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling, *Earth Interact.*, *2*, 1–26.

NCSS (2012), *National Cooperative Soil Characterization Database*, Natl. Coop. Soil Surv, Lincoln, Nebr.

NRCS (2004), Soil Survey Laboratory Methods Manual, Nat. Resour. Conserv. Serv., USDA, Lincoln, Nebr.

Odgers, N. P., Z. Libohova, and J. A. Thompson (2012), Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale, *Geoderma*, *189–190*, 153–163.

Oleson, K. W., et al. (2004), Technical Description of the Community Land Model (CLM), *NCAR Tech. Note TN-461+STR*, Clim. and Global Dyn. Div., Natl. Cent. for Atmos. Res., Boulder, Colo.

Oleson, K. W., et al. (2010), Technical Description of version 4.0 of the Community Land Model (CLM), *NCAR Tech. Note NCAR/TN-478+STR*, 257 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.

Oz, B., C. V. Deutsch, and P. Frykman (2002), A visual basic program for histogram and variogram scaling, *Comput. and Geosci.*, *28*(1), 21–31.

Pribyl, D. W. (2010), A critical review of the conventional SOC to SOM conversion factor, *Geoderma*, *156*(3–4), 75–83.

Quattrochi, D. A., C. W. Emerson, N. S.-N. Lam, and H.-l. Qiu (2001), Fractal characterization of multitemporal remote sensing data, in *Modelling Scale in Geographical Information System*, edited by N. Tate and P. Atkinson, John Wiley, London, U.K., pp. 13–34.

Reybold, W. U., and G. W. Tesselle (1989), Soil geographic data bases, *J. Soil Water Conserv.*, *43*, 226–229.

Reynolds, C. A., T. J. Jackson, and W. J. Rawls (2000), Estimating soil water-holding capacities by linking the Food and Agriculture Organization Soil map of the world with global pedon databases and continuous pedotransfer functions, *Water Resour. Res.*, *36*, 3653–3662.

Romanowicz, A. A., M. Vanclooster, M. Rounsevell, and I. L. Junesse (2005), Sensitivity of the SWAT model to the soil and land use data parameterisation: A case study in the Thyle catchment, Belgium, *Ecol. Modell.*, *187*, 27–39.

Sanchez, P. A., et al. (2009), Digital soil map of the world, *Science*, *325*, 680–681.

Shangguan, W., Y. Dai, B. Liu, A. Ye, and H. Yuan (2012), A soil particle-size distribution dataset for regional land and climate modelling in China, *Geoderma*, *171–172*, 85–91.

Shangguan, W., et al. (2013), A China dataset of soil properties for land surface modeling, *J. Adv. Model. Earth Syst.*, *5*, 212–224.

Shi, X., D. Yu, E. D. Warner, X. Pan, G. W. Petersen, Z. G. Gong, and D. C. Weindorf (2004), Soil database of 1:1,000,000 digital soil survey and reference system of the Chinese genetic soil classification system, *Soil Surv. Horizons*, *45*, 129–136.

Shi, X. Z., D. S. Yu, S. X. Xu, E. D. Warner, H. J. Wang, W. X. Sun, Y. C. Zhao, and Z. T. Gong (2010), Cross-reference for relating Genetic Soil Classification of China with WRB at different scales, *Geoderma*, *155*, 344–350.

Soil Landscapes of Canada Working Group (2010), Soil Landscapes of Canada version 3.2., Agric. and Agri-Food Can, Ottawa, Ontario.

Soil Survey Staff (1999), *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, U.S. Dep. of Agric., Nat. Resour. Conserv. Serv., Fort Worth, Tex.

Spaargaren, O. C., and B. H. Batjes (1995), *Report on the Classification into FAO-UNESCO Soil Units of Profiles Selected from the NRCS pedon Database for IGBP-DIS, ISRIC*, Wageningen, Netherlands.

Thompson, J. A., T. Prescott, A. C. Moore, J. Bell, D. R. Kautz, J. W. Hempel, S. W. Waltman, and C. H. Perry (2010), Regional approach to soil property mapping using legacy data and spatial disaggregation techniques, paper presented at 19th World Congress of Soil Science, IUSS, Brisbane, Queensland, Australia.

Thornton, P. E., and N. A. Rosenbloom (2005), Ecosystem model spin-up: Estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model, *Ecol. Modell.*, *189*, 25–48.

Todd-Brown, K. E. O., J. T. Randerson, W. M. Post, F. M. Hoffman, C. Tarnocai, E. A. G. Schuur, and S. D. Allison (2013), Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, *Biogeosciences*, *10*, 1717–1736.

USDA-NCSS (2006), *Digital General Soil Map of U.S.*, U.S. Dep. of Agric., Nat. Resour. Conserv. Serv., Fort Worth, Tex.

Wang, G., G. Gertner, and A. B. Anderson (2004), Up-scaling methods based on variability-weighting and simulation for inferring spatial information across scales, *J. Remote Sens.*, *25*, 4961–4979.

Wang, Y. P., R. M. Law, and B. Pak (2009), A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere, *Biogeosciences*, *6*, 9891–9944.

Webb, R. S., C. E. Rosenzweig, and E. R. Levine (1993), Specifying land surface characteristics in general circulation models: Soil profile data set and derived water-holding capacities, *Global Biogeochem. Cycles*, *7*, 97–108.

Wilson, M. F., and A. Henderson-Sellers (1985), A global archive of land cover and soils data for use in general circulation climate models, *J. Climatol.*, *5*, 119–143.

Zhang, W. L., A. G. Xu, H. J. Ji, R. L. Zhang, Q. L. Lei, H. Z. Zhang, L. P. Zhao, and H. Y. Long (2010), Development of China digital soil map at 1:50,000 scale, paper presented at 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, 1–6 August.

Zöbler, L. (1986), A world soil file for global climate modeling, *NASA Tech. Memo. 87802*, 33 pp., NASA, New York.