SOIL Discuss., https://doi.org/10.5194/soil-2018-32
Manuscript under review for journal SOIL
Discussion started: 15 October 2018

A review on the global soil datasets for earth system modeling

Yongjiu Dai[1*], Wei Shangguan[1*], Dagang Wang[2], Nan Wei[1], Qinchuan Xin[2], Hua Yuan[1], Shupeng Zhang[1], Shaofeng Liu[1], Fapeng Yan[3]

[1] Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, China.
[2] School of Geography and Planning, Sun Yat-sen University, Guangzhou, China.
[3] College of Global Change and Earth System Science, Beijing Normal University, Beijing, China

Correspondence to: Wei Shangguan (shgwei@mail.sysu.edu.cn) and Yongjiu Dai(daiyj6@mail.sysu.edu.cn)

Abstract. Global soil dataset is a pillar to the challenge of earth system modeling. But it is one of the most important uncertainty sources for Earth System Models (ESMs). Soil datasets function as model parameters, initial variables and benchmark datasets for model calibration, validation and comparison. For modeling use, the dataset should be geographically continuous, scalable and with uncertainty estimates. The popular soil datasets used in ESMs are often based on limited soil profiles and coarse resolution soil maps. Updated and comprehensive soil information needs to be incorporated in ESMs. New generation soil datasets derived by digital soil mapping with abundant soil observations and environmental covariates are preferred to those by the linkage method for ESMs. Because there is no universal pedotransfer function, an ensemble of them may be more suitable to provide secondary soil parameters to ESMs. Aggregation and upscaling of soil data are needed for model use but can be avoid by taking a subgrid method in ESMs at the cost of increases in model complexity. Uncertainty of soil data needs to be incorporated in ESMs.

1 Introduction

Soil or pedosphere is a key component of Earth system, and plays an important role in the water, energy and carbon balances and biogeochemical processes. An accurate description of soil properties is essential in advancing the modeling capabilities of Earth System Models (ESMs) to predict land surface processes at the global and regional scales. Soil information is required by the land surface models (LSMs), which is a component of ESMs. With the help of computer-based geographic systems, many researchers have produced geographical databases to organize and harmonize large amount of soil information generated from soil surveys during the last decades (Batjes, 2017; Hengl et al., 2017). However, soil dataset used in ESMs is not well updated nor well utilized yet. The popular soil datasets used in ESMs are outdated and with limited accuracy. Some available soil properties such as gravel (or coarse fragment) and depth to bedrock are not utilized in most ESMs. Meanwhile, it is needed to change ESMs' schemes and structure to better represent the soil process in utilizing new soil information (Brunke et al., 2016; Oleson et al., 2010). Better soil information with high resolution and better representation of soil in models have improved and will improve the performance in simulating the Earth system (Kearney and Maino, 2018).

ESMs require detailed information on the soil physical and chemical properties. Site observations (called soil profiles) from soil surveys include soil properties such as soil depth, soil texture (sand, silt and clay fractions), organic matter, coarse fragments, bulk density, soil colour, soil nutrients (carbon (C), nitrogen (N), phosphorus (P), potassium (K) and sulfur (S)), etc. However, soil hydraulic and thermal parameters as well as biogeochemical parameters are usually not observed in soil surveys, which need to be estimated by pedotransfer functions (PTFs) (Looy et al., 2017). This review focus on the soil data (usually time-invariant) from soil surveys, while variables such as soil temperature and soil moisture are beyond this paper's scope.

Soil properties are functioned in three aspects in ESMs:

1) Model parameters. The soil thermal (soil heat capacity and the thermal conductivity) and hydraulic characteristics (empirical parameters of soil water retention curve and hydraulic conductivity) are usually obtained by fitting equations (PTFs) to easily measured and widely available soil properties, such as sand, silt and clay fractions, organic matter content, rock fragments and bulk density (Clapp and Hornberger, 1978; Farouki, 1981; Vereecken et al., 2010; Dai et al., 2013). Soil albedos are significantly correlated with Munsell soil color value (Post et al., 2000).

2) Initial variables. The nutrient (C, N, P, K, S, etc.) amounts and the nutrients associated parameters (pH, cation-exchange capacity, etc.) in soils can be used to initialize the simulations. Generally, their initial values are assumed to be at steady state by running model over thousands of model years (i.e., spin-up) until no trend of change in pool sizes (McGuire et al., 1997; Thornton and Rosenbloom, 2005; Doney et al., 2006; Luo et al., 2016). To initialize nutrient amounts using the reliable soil data as background field could largely reduce the times of model spin-up, and also could avoid the possibility of the non-linear singularity evolution of the modeling, and

then provide better estimate of the true terrestrial nutrient state.

3) Benchmark data. Soil data, as observations, could serve as a reference for modeling calibration, validation and comparison. Soil carbon stock is one of the most frequently used soil properties as benchmark data (Todd-Brown et al., 2013).

Soil properties are of great spatial heterogeneity both horizontally and vertically. As a result, ESMs usually incorporate soil property maps for multiply layers rather than a global constant. ESMs, especially LSMs, are evolving towards hyper-resolutions of 1km or finer with more detailed parameterization schemes to accommodate the land surface heterogeneity (Singh et al., 2015; Ji et al., 2017). So spatially explicit soil data at high resolutions are necessary to improve land surface representation and simulation. Because soil properties are observed at individual locations, soil mapping or spatial prediction model is needed to derive the 3D representation of soil distribution. The traditional way (i.e., the linkage method) is to link soil profiles and soil mapping units on soil type maps, sometimes with ancillary maps such as topography and land use (Batjes, 2003). In the past decades, various digital soil mapping technologies were proposed by finding the relationships between soil and environmental covariates (usually remote sensing data) such as climate, topography, land use, geology and so on (McBratney et al., 2003).

Soil datasets are usually not ready for the use of ESMs and some upscaling issues need to be addressed. The soil datasets produced by the linkage methods are polygon-based and need to be converted to fit the grid-based ESMs. This conversion can be done by either subgrid method or spatial aggregation. The up-to-date soil data are provided at a resolution of 1km or finer, while the LSMs are mostly ran at a coarser resolution. So upscaling of soil data is necessary before it can be used by ESMs. Proper upscaling methods need to be chosen carefully to minimize uncertainty in the modeling results introduced by them (Hoffmann and Christian Biernath, 2016; Kuhnert et al., 2017).

This paper is organized in the following sections. In section 2, we first introduce soil datasets at global and national scales produced by the linkage method and digital soil mapping technology and then the soil datasets that have already been incorporated in ESMs. Section 3 presents PTFs that are used in ESMs to estimate soil hydraulic and thermal parameters. Section 4 describes how to deal with soil data derived by the linkage methods. Section 5 introduces the upscaling of high resolution soil data to the coarse resolution of ESMs. Section 6 gives the summary and an outlook of further improvements.

2 Soil datasets used in ESMs

2.1 Global and national soil datasets

Two kinds of soil data are generated from soil surveys: soil polygon maps representing distribution of soil types and soil profiles with observations of soil properties. ESMs usually require the spatial distribution of soil properties, or soil property maps rather than soil classification information. Two kinds of methods, i.e. the linkage method and the digital soil mapping method, are used to derive soil property maps.

SOIL Discuss., https://doi.org/10.5194/soil-2018-32
Manuscript under review for journal SOIL
Discussion started: 15 October 2018

Soil maps show the geographical distribution of soil types, which are compiled under a certain soil classification system. At the global level, there is only one generally accepted global soil map, i.e., the FAO-UNESCO Soil Map of the World (SMW) (FAO, 1971-1981). It was made based on soil surveys conducted between the 1930s and the 1970s, and technology available in 1960s. Several versions exist in the digital format (FAO, 2003b, 1995; Zöbler, 1986). At the regional and national level, there are many soil maps based on either national or international soil classifications. Here are some examples of major soil maps available in digital formats: the Soil and Terrain Database (SOTER) databases (Van Engelen and Dijkshoorn, 2012) for different regions, the European Soil Database (ESB, 2004), the 1: 1 million Soil Map of China (National Soil Survey Office, 1995), the U.S. General Soil Map (GSM), the 1:1 million Soil Map of Canada (Soil Landscapes of Canada Working Group, 2010) and the Australian Soil Resource Information System (ASRIS) (Johnston et al., 2003).

Soil profiles are composed of multiply layers called soil horizons. For each horizon, soil properties are observed. At the global level, several soil profile databases exist. Here we only discuss the two most comprehensive ones. The World Inventory of Soil Emission Potentials (WISE) database was developed as a homogenized set of soil profiles (Batjes, 2008). The newest version (WISE 3.1) contains 10,253 soil profiles and 26 physical and chemical properties. The soil profiles database of World Soil Information Service (WoSIS) contains the most abundant profiles (about 118,400) from national and global databases including most of the databases mentioned below, though only a selection of important soil properties are included (Batjes, 2017). At the regional and national level, there are many soil profile databases, usually with soil classifications corresponding to the local soil maps. Here are some examples: the USA National Cooperative Soil Survey Soil Characterization database (http://ncsslabdatamart.sc.egov.usda.gov/), profiles from the USA National Soil Information System (http://soils.usda.gov/technical/nasis/), Africa Soil Profiles database (Leenaars, 2012), the Australian Soil Resource Information System (Karssies, 2011), the Chinese National Soil Profile database (Shangguan et al., 2013), soil profile archive from the Canadian Soil Information System (MacDonald and Valentine, 1992), soil profiles from SOTER (Van Engelen and Dijkshoorn, 2012), the soil profile analytical database for Europe (Hannam et al., 2009), the Mexico soil profile database (Instituto Nacional de Estadística y Geografía, 2016), and the Brazilian national soil profile database (Cooper et al., 2005).

The linkage method (called the taxotransfer rule-based method) is to link soil mapping units or soil polygons and soil profiles according to taxonomy-based pedotransfer (taxotransfer) rules (Batjes, 2003). The criteria used in the linkage could be one or many factors as following: soil class, soil texture class, depth zone, topographic class, distance between soil polygons and soil profiles and so on (Shangguan et al., 2012). Each soil type is represented by one or a group of soil profiles that meet the criteria, and usually the median or mean value of a soil property is assigned to the soil type. There are many sources of uncertainty in the linkage method (Shangguan et al., 2012). The major source is spatial errors of soil maps, i.e. the location of soil types, as the estimation relies heavily on the soil map and the purity of soil map units is likely to be around 50 to 65%. Because the linkage method assigned only one value or a statistical distribution to a soil type in soil polygons, the intra-polygonal variation is not taken into account. At the global level, many databases were derived by the linkage method: the FAO Soil Map of the World with derived soil properties (FAO, 2003a), the Soil and Terrain Database (Van Engelen and Dijkshoorn, 2012) for multiply regions and countries, the ISRIC-WISE derived soil property maps (Batjes, 2006), the Harmonized World Soil Database (HWSD) (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012), and the Global Soil Dataset for Earth System Model (GSDE) (Shangguan et al., 2014). Two most recent ones are HWSD and GSDE. HWSD was built via combining the existing regional and national updates of soil information. GSDE as an improvement of HWSD incorporated more soil maps and more soil profiles related to the soil maps, with more soil properties. GSDE accomplished the linkage based on the local soil classification, which required no correlation between classification systems and avoided the error brought by taxonomy reference. In addition, GSDE provided estimation of eight layers to the depth of 2.3 m, while HWSD provided estimation of two layers to the depth of 1 m. Many national and regional agencies around the world have organized their soil surveys by linking soil maps and soil profiles, including the USA State Soil Geographic Database (STATSGO2) (Soil Survey Staff, 2017), Soil Landscapes of Canada (Soil Landscapes of Canada Working Group, 2010), the ASRIS (Johnston et al., 2003), the Soil-Geographic Database of Russia (Shoba et al., 2008) the European Soil Database (ESB, 2004), the China dataset of soil properties (Shangguan et al., 2013) and so on.

Digital soil mapping (McBratney et al., 2003) is the creation and the population of a geographically referenced soil database, generated at a given resolution by using field and laboratory observation methods coupled with environmental data through quantitative relationships (http://digitalsoilmapping.org/). Usually, the soil datasets derived by digital soil mapping provide grid-based spatial continuous estimation while the soil datasets derived by the linkage method provide estimations with abrupt changes at the boundary of soil polygons. The uncertainty could be estimated quantitatively by methods such as geostatistical methods and quantile regression forest (Vaysse and Lagacherie, 2017). The GlobalSoilMap is a global consortium that aims to create global digital maps for key soil properties (Sanchez et al., 2009). This global effort takes a bottom-up framework and will produce the best available map of soil at a resolution of 3 arc sec (about 100 m) along with the 90% confidence of predictions. Soil properties will be provided for six soil layers (i.e. 0–5, 5–15, 15–30, 30– 60, 60–100, and 100–200 cm). Many countries have produced soil maps following the GlobalSoilMap specifications (Odgers et al., 2012; Viscarra Rossel et al., 2015; Mulder et al., 2016; Ballabio et al., 2016; Ramcharan et al., 2018; Arrouays, 2018). The Soilgrids system (https://www.soilgrids.org) is another global soil mapping project (Hengl et al., 2014; Hengl et al., 2015; Hengl et al., 2017). The newest version (Hengl et al., 2017) at a resolution of 250 m was produced by fitting an ensemble of machine learning methods based on about 150,000 soil profiles and 158 soil covariates, which is currently the most detailed estimation of global soil distribution.

The new generation soil dataset produced by digital soil mapping method gave a quite different distribution of soil properties from those produced by the linkage method. Figure 1 shows soil sand and clay fraction at the surface 0-30 cm layer from Soilgrids, IGBP-DIS (Data and Information System of International Geosphere-Biosphere Programme) and GSDE. Figure 2 shows soil organic carbon and bulk density at the surface 0-30 cm layer from Soilgrids and GSDE. Significant differences are visible in these datasets. This will lead to different modelling results in ESMs. (Marwa et al., 2018) found that the global soil organic carbon stocks at 1m depth is 3,400 Pg estimated by Soilgrids while it is 2500 Pg by HWSD, and the estimates by Soilgrids are closer to the observations.

2.2 Soil dataset incorporated in ESMs

Table 1 shows several most popular ESMs (specifically, their land surface models) and the input soil datasets. Land surface models (LSMs) are key tools to predict the dynamic of land surface under climate change and land use. Five datasets are widely used, e.g., the datasets by Wilson and Henderson-Sellers (1985), Zöbler (1986), Webb et al. (1993), Reynolds et al. (2000), Global Soil Data Task (2000), and Miller and White (1998). Except HWSD and STATSGO (Miller and White, 1998) for USA, these datasets were derived from the Soil Map of the World (FAO, 1971-1981) and limited soil profile data (no more than 5,800 profiles), which gained popularity because its simplicity and ease of use. But they are outdated and should no longer be used because much better soil information as introduced in Section 2.1 can be incorporated.

In recent years, efforts were taken to improve the soil data condition in ESMs. This was started by the Land-Atmosphere Interaction Research Group at Beijing Normal University (BNU, now at Sun Yat-sen University). Shangguan et al. (2012, 2013) developed a China dataset of soil properties for land surface modeling based on 8,979 soil profiles and the Soil Map of China using the linkage method. Dai et al. (2013) derived soil hydraulic parameters using pedotransfer functions based on the soil properties by Shangguan et al. (2013). Shangguan et al. (2014) further developed a comprehensive global dataset for ESMs. The above soil datasets were widely used in the ESMs. Soil properties from these soil datasets, including soil texture fraction, organic carbon, bulk density and derived soil hydraulic parameters, were implemented in the Common Land Model Version 2014 (CoLM2014, http://land.sysu.edu.cn/). Li et al. (2017) shows that CoLM2014 was more stable than the previous version and had comparable performance to that of CLM4.5 which may be attributed in part to the new soil parameters as input. Wu et al. (2014) shows that soil moisture values are closer to the observations when simulated by CLM3.5 with the China dataset than those simulated with FAO. Zheng and Yang (2016) estimated effects of soil texture datasets from FAO and BNU on regional terrestrial water cycle simulations with the Noah-MP land surface model. Tian et al. (2012) used the China soil texture data in a land surface model (GWSiB) coupled with a groundwater model. Lei et al. (2014) used the China soil texture data in CLM to estimate the impacts of climate change and vegetation dynamics on runoff in the mountainous region of the Haihe River basin.

SOIL Discuss., https://doi.org/10.5194/soil-2018-32
Manuscript under review for journal SOIL
Discussion started: 15 October 2018

Zhou et al. (2015) estimated age-dependent forest carbon sink with a terrestrial ecosystem model utilizing the soil carbon data of China. Dy and Fung (2016) updated the soil data for the Weather Research and Forecasting model (WRF).

Researchers have also put efforts to update ESMs with other soil data. Lawrence and Chase (2007) used MODIS data to derive soil reflectance, which was used as a soil colour parameter in Community Land Model 3.0 (CLM). De Lannoy et al. (2014) updated the Catchment land surface model of the NASA with soil texture and organic matter data from HWSD and STATSGO2. Livneh et al. (2015) evaluated the influence of soil textural properties on hydrologic fluxes by comparing the FAO data and STATSGO2. Folberth et al. (2016) evaluated the impact of soil input data on yield estimates in a global gridded crop model. Slevin et al. (2017) utilized the HWSD to simulate global gross primary productivity in the JULES land surface model. Trinh et al. (2018) proposed an approach that can assimilate coarse global soil data by finer land use and coverage dataset which improved the performance of hydrologic modeling at watershed scale. Kearney and Maino (2018) incorporated the new generation of soil data produced by digital soil mapping method into a climate model and found that, compared to the old soil information, this improved the simulation of soil moisture at fine spatial and temporal resolution over Australia. A global gridded hydrologic soil groups (HYSOGs250m) was developed based on soil texture and depth to bedrock of Soilgrids (Hengl et al., 2017) and groundwater table depth (Fan et al., 2013) for curve-number based runoff modeling of U.S. Department of Agriculture (Ross et al., 2018).

Except soil properties, the estimation of underground boundaries including the groundwater table depth, the depth to bedrock (DTB) and depth to regolith and its implementation in ESMs is also a new focus. Fan et al. (2013) compiled global observations of water table depth and inferred the global patterns using a groundwater model. Pelletier et al. (2016) developed a global DTB dataset by using process based models for upland and an empirical model for lowland. This dataset was implemented in the CLM4.5 and found that there were significant influences on water and energy simulations compared to the default constant depth (Brunke et al., 2015). Shangguan et al. (2017) developed a global DTB by digital soil mapping based on about 1.7 million observations from soil profiles and water wells, which has a much higher accuracy than the dataset by Pelletier et al. (2016). Vrettas and Fung (2016) shows that the weathered bedrock stores a significant fraction (more than 30%) of the total water despite its low porosity. Jordan et al. (2018) estimated global permeability of the unconsolidated and consolidated earth for groundwater modelling. However, due to the lack of data, an accurate global estimation of depth to regolith is not feasible. Caution should be paid to use of the products of so-called soil depth in ESMs. Soil depth maps are usually estimated based on observations from soil survey, and soil depth (or depth to the R horizon) is assumed to be equal to DTB. However, these observations are usually less than 2 meters and usually do not meet the depth to bedrock (Shangguan et al., 2017). Thus, soil depth maps based on soil profiles only are significantly underestimated (one order of magnitude lower) compared to the actual depth to bedrock and should not be taken as the lower boundary of ESMs.

For the convenience of ESMs' application, we present basic descriptions about the new soil datasets in Table 2 and 3. As described in section 2.1, three available global soil datasets, i.e. HWSD, GSDE and Soilgrids, have been developed in the last several years (Table 2). Table 3 shows the available soil properties of these soil datasets. Though all three databases do not contain uncertainty estimation, Soilgrids is considered to be the most accurate one. The explained variance of soil properties in Soilgrids is between 56% and 83%, while HWSD and GSDE do not offer quantitative accuracy assessment. GSDE has the largest number of soil properties, while Soilgrids contains only ten primary soil properties.

3 Estimating secondary parameters using pedotransfer functions

Earth system modellers have employed different pedotransfer functions (PTFs) to estimate soil hydraulic parameters (SHP), soil thermal parameters (STP), and biogeochemical parameters (Looy et al., 2017;Dai et al., 2013). Almost all ESMs incorporated SHPs and STPs estimated by PTFs but not biogeochemical parameters. PTFs are the empirical functions that account for the relationships between these secondary parameters and more easily obtainable soil property data. Direct measurement of these parameters is difficult, expensive and in most cases impractical to take sufficient samples to reflect the spatial variation. Thus, most soil databases do not contain these secondary parameters. PTFs provide the alternative to estimate them. In ESMs, SHPs and STPs are usually derived using simple PTFs taking only soil texture data as the input. As more soil properties become available globally, including gravel, soil organic matter and bulk density, more sophisticated PTFs using additional soil properties can be utilized in ESMs.

PTFs can be expressed as either numerical equations or by machine learning methodology which is more flexible to simulate the highly nonlinear relationship in analysed data. PTFs can also be developed based on soil processes. Most researches did not indicate where the PTFs can potentially be used, and the accuracy of a PTF outside of its development dataset is essentially unknown McBratney et al. (2011). Therefore, they should never be considered as an ultimate source of parameters in soil modelling. Looy et al. (2017) reviewed PTFs extensively in earth system science and emphasized that PTF development has to go hand in hand with suitable extrapolation and upscaling techniques such that the PTFs correctly represent the spatial heterogeneity of soils in ESMs. Though the PTFs were evaluated, it is not clear which are the best set of PTFs for global applications. Due to these limitations, a better way to estimate the secondary parameters may be to use an ensemble of PTFs, which can give the variability of parameters. Dai et al. (2013) derived a global soil hydraulic parameters using the ensemble method. Selection of PTFs was carried out based on the following rules, including the consistent physic definition, large enough training sample and positive evaluations in comparison with other PTFs. The PTFs selected included not only those in equations but also PTFs of machine learning. As a result, the modellers could use these parameters as inputs instead of calculating them in ESMs every time running the model.

The new generation soil information has already been utilized to derive SHPs and STPs in some researches. Montzka et al. (2017) produced a global map of SHPs at a resolution of 0.25° based on the SoilGrids 1km dataset. Tóth et al. (2017) calculated SHPs for Europe with the EU-HYDI PTFs (Tóth et al., 2015) based on SoilGrids 250 m. Wu et al. (2018) used an integrated approach that ensembles PTFs to map field capacity of China based on multi-source soil datasets.

The performance of PTF in ESMs is evaluated in many researches, though PTFs has not been fully exploited and integrated into ESMs (Looy et al., 2017). Here are some examples. Chen et al. (2012) incorporated soil organic matter to estimate soil porosity and thermal parameters for the use of land surface models. Zhao et al. (2018a) evaluated PTFs performance to estimate SHPs and STPs for land surface modelling over the Tibetan Plateau. Zheng et al. (2018) developed PTFs to estimate the soil optical parameters to derive soil albedo for the Tibetan Plateau, and the PTFs incorporated into an eco-hydrological model which improved the model simulation of surface energy budget. Looy et al. (2017) envisaged two possible approaches to improve parameterization of Earth system models by PTFs. One is to replace constant coefficients in the current ESMs with spatially distributed values by PTFs. The other is to develop spatially exploitable PTFs to parameterize specific processes using knowledge of environmental controls and variation of soil properties.

4 Model use of soil data derived by the linkage method

In the soil map, a soil mapping unit (SMU) is composed of more than one component in most cases, and thus a one-to-many relationship exists between the SMU and the profile attribute. This condition makes representing attributes characterizing a SMU a non-trivial task. To keep the whole variation of soil in a SMU, the best way is using the subgrid method in ESMs (Oleson et al., 2010), i.e. aggregate values of soil properties and provide the area percentage of each value. However, this will increase the computing time and the complexity of ESMs' structure, which needs to implement the soil processes over each subgrid soil column within a grid instead of the entire model grid.

Usually, the ESMs uses grid data as input and each grid cell has one unique value of a soil property. Three spatial aggregation methods were proposed to aggregate compositional attributes in a SMU to a representative value (Batjes, 2006; Shangguan et al., 2014). The area-weighting method (method A) takes area-weighting of soil attributes. The dominant type method (method D) takes the soil attribute of the dominant type. The dominant binned method (method B) classifies the soil attribute into several preselected classes and takes the dominant class. All three methods can be applied to quantitative data, while the method D and the method B can be applied to categorical data. The advantages and disadvantages of these methods were discussed (Batjes, 2006; Shangguan et al., 2014). The choice should be made according to the specific applications (Hoffmann and Christian Biernath, 2016). The method B provides binned classes, which are not convenient for modelling, though method B is considered more appropriate to represent a grid cell. The method A keeps mass conservation, which can meet most demands of model applications. However, the method A may be misleading in cases when extreme values appeared in a SMU. For the linkage method, the uncertainty is usually estimated by giving the 5 and 95 percentile soil properties of the soil profiles that linked to a SMU. Because the frequency distribution of soil properties within a SMU is usually not a normal distribution or any other typical statistic distribution, the application of statistic such as standard deviation in model use is not proper. This means that the uncertainty of soil dataset derived by the linkage method can not be incorporated into ESMs in a straight forward way.

The basic soil properties are often used to derive secondary parameters including SHPs and STPs by PTFs and soil carbon stock or other nutrient stocks by certain equations (Shangguan et al., 2014). This procedure could be done either before or after the aggregation (here referred to ‘‘aggregating after” and ‘‘aggregating first’’). Because the relationship between the soil basic properties and the derived soil parameters is usually nonlinear, the ‘‘aggregating first’’ method should be taken. This was also proved by case studies (Romanowicz et al., 2005; Shangguan et al., 2014). However, some researchers (Hiederer and Köchy, 2012) were not aware of this and used the ‘‘aggregating after” method producing misleading results.

The aggregation smooths the variation of soil properties between soil components within a given SMU (Odgers et al., 2012). To avoid the aggregation, the spatial disaggregation of soil type maps can be used to determine the location of the SMU components, though the location error may be high in some cases (Thompson et al., 2010; Stoorvogel et al., 2017). This method depends on high density of soil profiles to establish soil and landscape relationships. Folberth et al. (2016) shows that the correct spatial allocation of the soil type to present cropland was very important in global crop yield simulations. Currently, aggregation is still the pragmatic way at the global scale due to lack of data.

5 Upscaling detailed soil data for model use

The updated soil datasets derived by both the linkage method and digital soil mapping are usually at a high resolution from 1 km to 100 m, and upscaling or aggregation is required to derive lower resolution datasets for model use. The aggregation methods mentioned above can be used. Moreover, there are plenty of upscaling methods such as the window median, variability-weighted methods (Wang et al., 2004), variogram method (Oz et al., 2002), fractal theory (Quattrochi et al., 2001) and Miller–Miller scaling approach (Montzka et al., 2017). However, few studies have been devoted to test out which upscaling methods are suitable for soil data. A preliminary effort was done by (Shangguan, 2014). Five upscaling methods compared were the window average, widow median, widow modal, arithmetic average variability-weighted method and bilinear interpolation method. Differences between aggregation methods varied from 10% to 100% for different parameters. The upscaling methods affected the data derived by the linkage method more than the data by digital soil mapping. The window average, window median and arithmetic average variability-weighted method performed similar in upscaling. The root mean square error increased rapidly when the window size was less than 40 pixels. Similar to the aggregation of SMUs, the “aggregating first” method is recommended when secondary soil parameters are derived. Again, alternative to avoid the aggregation into one single value for a grid cell is to use the subgrid methods in ESMs.

The upscaling effect of soil data on model simulation has been investigated in previous studies with controversial conclusions. For example, Melton et al. (2017) used two linked algorithms to provide tiles of representative soil textures for subgrids in a terrestrial ecosystem model and found that the model is relatively insensitive to subgrid soil textures compared to a simple grid-mean soil texture at a global scale. However, the treatment without soil subgrid structure in JULES resulted in soil-moisture dependent anomalies in simulated carbon flux (Park et al., 2018). Further researches are necessary to investigate the upscaling effect on models

6 Summary and outlook

This paper reviews the status of soil datasets and their usage in ESMs. Soil physical and chemical properties served as model parameters, initial variables or benchmark datasets in ESMs. Soil profiles, soil maps and soil datasets derived by the linkage method and digital soil mapping are reviewed at national, regional and global levels. The soil datasets derived by digital soil mapping are considered to provide more realistic estimation of soils than those derived by the linkage method.

The popular soil datasets used in ESMs are outdated and there are soil datasets available for the updates. In the recent several years, efforts were taken to update the soil data in ESMs. The effects of updated soil properties which are used to estimate soil hydraulic and thermal parameters were evaluated. Other major updates include soil reflectance, ground water tables and depth to bedrock.

Pedotransfer functions (PTFs) are employed to estimate secondary soil parameters, including soil hydraulic and thermal parameters, and biogeochemical parameters. PTFs can take more soil properties (i.e., soil organic carbon, bulk density etc.) as input in addition to soil texture data. An ensemble of PTFs may be more robust to provide secondary soil parameters as direct input to ESMs.

Soil data derived by the linkage methods and high resolution data can be aggregated by different methods to fit the use in ESMs. The aggregation should be done after the secondary parameters are estimated. However, the aggregation will omit the variation of soil properties. To avoid the aggregation, the subgrid method in ESMs is an alternative which increases the model complexity. The effect of different upscaling methods on the performance of ESMs needs to be investigated further.

Because digital soil mapping has many advantages compared to the traditional linkage method, especially in representing spatial heterogeneity, the new generation soil datasets derived by digital soil mapping needs to gain popularity in ESMs. As a data-driven method, digital soil mapping requires soil profiles observations and environmental covariates (in which the importance of soil maps is low), and including more these data in mapping will improve the global predictions (Hengl et al., 2017). The temporal variation of global soil is quite challenging due to lack of data. Soil image fusion is also needed to merge the local and global soil maps. Mapping the soil depth and depth to bedrock separately at the global level is also still challenging due to lack of data and the understanding of relevant processes. Uncertainty estimation should be included in the soil datasets developed in the future.

The gap between soil data existence and data availability is huge. The soil profiles included by global soil databases such as WoSIS make up a very small fraction of the soil pits dug by human beings. For example, there are more than 100,000 soil profiles from the second national soil survey of China (Zhang et al., 2010) and no more than 9,000 were used to produce the national soil property maps freely available (Shangguan et al., 2013). In the last century, national soil survey was accomplished widely, majorly for agriculture purpose. However, most of these legacy data are not digitalized and they are usually not made available to the science community even if digitalized. How to flush out these hidden soil data requires some mechanism such as government mandatory regulations and investing money on making them available. In addition, investments on new soil samplings should be made, especially in the under-represented areas. A good example is the US, which has the most abundant soil data freely available like many other data. Data compatibility of different analysis methods and different description protocols including soil classifications is also an important issue and data harmonization is necessary when the data are made available to public.

The gap between the soil data availability and data usage in ESMs is still large. Most popular ESMs have not utilized the recent available global soil datasets yet. Another challenge may be to incorporate the spatial uncertainty of soil properties and the statistical distribution of soil properties in a grid cell in ESMs.

Acknowledgements. This work was supported by the Natural Science Foundation of China (under grants 41730962 and 41575072) and the National Key Research and Development Program of China under grants 2017YFA0604303 and 2016YFB0200801.

SOIL Discuss., https://doi.org/10.5194/soil-2018-32
Manuscript under review for journal SOIL
Discussion started: 15 October 2018

References

Arrouays, D., Savin, I., Leenaars, J. , McBratney, A.: GlobalSoilMap - Digital Soil Mapping from Country to Globe, CRC Press, London, 2018.

Ballabio, C., Panagos, P., and Monatanarella, L.: Mapping topsoil physical properties at European scale using the LUCAS database, Geoderma, 261, 110-123, 2016.

Batjes, N. H.: A taxotransfer rule-based approach for filling gaps in measured soil data in primary SOTER databases, International Soil Reference and Information Centre, Wageningen, 2003.

Batjes, N. H.: ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid. Report 2006/02, ISRIC- World Soil Information, Wageningen (with data set), 2006.

Batjes, N. H.: ISRIC-WISE harmonized global soil profile dataset (ver. 3.1). Report 2008/02, ISRIC - World Soil Information, Wageningen, 2008.

Batjes, N. H., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., Mendes de Jesus, J.: WoSIS: Serving standardised soil profile data for the world, Earth Syst. Sci. Data, 9, 1-14, 2017.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description– Part 1: Energy and water fluxes, Geosci. Model Dev., 4, 677-699, 10.5194/gmd-4-677-2011, 2011.

Blyth, E. M. a. C.: JULES: A new community land surface mode. Global Change Newsletter, NO. 66, IGBP, Stockholm, Sweden, 9-11, 2006.

Brunke, M. A., Tucson, A., Broxton, P. D., Pelletier, J., Gochis, D. J., Hazenberg, P., Lawrence, D. M., Niu, G. Y., Troch, P. A., and Zeng, X.: Implementation and testing of variable soil depth in the global land surface model CLM4.5, 27th Conference on Climate Variability and Change, Phoenix, 2015,

Brunke, M. A., Broxton, P., Pelletier, J., Gochis, D., Hazenberg, P., Lawrence, D. M., Leung, L. R., Niu, G.-Y., Troch, P. A., and Zeng, X.: Implementing and evaluating variable soil thickness in the Community Land Model version 4.5 (CLM4.5), Journal of Climate, 29, 3441–3461, doi:10.1175/JCLI-D-15-0307.1, 2016.

Chen, F., and Dudhia, J.: Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity, Monthly Weather Review, 129, 569-585, 2001.

Chen, Y., Yang, K., Tang, W., Qin, J., and Zhao, L.: Parameterizing soil organic carbon's impacts on soil porosity and thermal parameters for Eastern Tibet grasslands, Science China Earth Sciences, 55, 1001-1011, 10.1007/s11430-012-4433-0, 2012.

Clapp, R. W., and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water Resources Res., 14, 601-604, 1978.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geosci. Model Dev., 4, 701-722, 10.5194/gmd-4-701-2011, 2011.

Cooper, M., Mendes, L. M. S., Silva, W. L. C., and Sparovek, G.: A national soil profile database for brazil available to international scientists, Soil Science Society of America Journal, 69, 649–652, 2005.

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J.: The impact of new land surface physics on the GCM sensitivity of climate and climate sensitivity, Climate Dynamics, 15, 183-203, 1999.

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G., Oleson, K. W., Schlosser, C. A., and Yang, Z.: The Common Land Model, Bull. Amer. Meteor. Soc., 84, 1013-1023, 2003.

Dai, Y., Shangguan, W., Duan, Q., Liu, B., Fu, S., and Niu, G.: Development of a China Dataset of Soil Hydraulic Parameters Using Pedotransfer Functions for Land Surface Modeling, Journal of Hydrometeorology, 14, 869–887, 2013.

De Lannoy, G. J. M., Koster, R. D., Reichle, R. H., Mahanama, S. P. P., and Liu, Q.: An updated treatment of soil texture and associated hydraulic properties in a global land modeling system, Journal of Advances in Modeling Earth Systems, 6, 957-979, 10.1002/2014ms000330, 2014.

Dickinson, R. E., Henderson-Sellers, A., and Kennedy, P. J.: Biosphere-Atmosphere Transfer Scheme (BATS) Version 1e as Coupled to the NCAR Community Climate Model. NCAR-TN-387+STR, National Center for Atmospheric Research, Boulder, Colorado, 88, 1993.

Doney, S. C., Lindsay, K., Fung, I., and John, J.: Natural variability in a stable, 1000-yr global coupled climate-carbon cycle simulation, Journal of Climate, 19, 3033-3054, 2006.

Dy, C. Y., and Fung, J. C. H. C. J.: Updated global soil map for the Weather Research and Forecasting model and soil moisture initialization for the Noah land surface model, Journal of Geophysical Research: Atmospheres, 121, 8777-8800, 10.1002/2015jd024558, 2016.

Elguindi, N., Bi, X., Giorgi, F., Nagarajan, B., Pal, J., Solmon, F., Rauscher, S., Zakey, A., O'Brien, T., Nogherotto, R., and Giuliani, G.: Regional climatic model RegCM Reference Manual version 4.6, ITCP, Trieste, 33, 2014.

Fan, Y., Li, H., and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth, Science, 339, 940-943, 10.1126/science.1229881, 2013.

FAO: Soil Map of the World, UNESCO, Paris. Vol. 110, 1971-1981.

FAO: Digitized Soil Map of the World and Derived Soil Properties, FAO, Rome, 1995.

FAO: Digital soil map of the world and derived soil properties, Food and Agriculture Organization of the United Nations, Land and Water Digital Media Series, CD-ROM, 2003a.

FAO: The Digitized Soil Map of the World Including Derived Soil Properties (version 3.6), FAO, Rome, 2003b.

FAO/IIASA/ISRIC/ISS-CAS/JRC: Harmonized World Soil Database (version1.2), FAO, Rome, Italy and IIASA, Laxenburg, Austria, 2012.

Farouki, O. T.: Thermal Properties of Soils. Monograph, No. 81-1, U.S. Army Cold Regions Research and Engineering Laboratory, 1981.

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M., and van der Velde, M.: Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations, Nature Communications, 7, 11872, 10.1038/ncomms11872, 2016.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572-597, doi:10.1002/jame.20038, 2013.

Global Soil DataTask: Global Soil Data Products CD-ROM (IGBP-DIS). International Geosphere-Biosphere Programme - Data and Information Services, Available online at from the ORNL Distributed Active Archive Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A., 2000.

Gurney, K. R., Baker, D., Rayner, P., and Denning, S.: Interannual variations in continental-scale net carbon exchange and sensitivity to observing networks estimated from atmospheric CO2 inversions for the period 1980 to 2005, Global Biogeochemical Cycles, 22, doi:10.1029/2007GB003082, 2008.

Hagemann, S., Botzet, M., Dümenil, L., and Machenhauer, B.: Derivation of global GCM boundary conditions from 1 km land use satellite data. MPI Report No. 289, 34, 1999.

Hagemann, S.: An Improved Land Surface Parameter Dataset for Global and Regional Climate Models. MPI Report No. 336, 28, 2002.

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

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., and Gonzalez, M. R.: SoilGrids1km — Global Soil Information Based on Automated Mapping, PLoS ONE, 9, e105992, 10.1371/journal.pone.0105992, 2014.

Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., Jesus, J. M. d., Tamene, L., and Tondoh, J. E.: Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions, PLOS ONE, 10, e0125814, 2015.

Hengl, T., J., M. d. J., Heuvelink, G. B. M., Gonzalez, R., M., K., M. , Blagotic, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: global gridded soil information based on Machine Learning, PLOS One, 12, 2017.

Hiederer, R., and Köchy, M.: Global Soil Organic Carbon Estimates and the Harmonized World Soil Database, Publications Office of the European Union, Luxembourg, 79, 2012.

Hoffmann, H., G. Zhao, S. Asseng, M. Bindi, and Christian Biernath, J. C., Elsa

SOIL Discuss., https://doi.org/10.5194/soil-2018-32
Manuscript under review for journal SOIL
Discussion started: 15 October 2018

Coucheney, Rene Dechow, Luca Doro, Henrik Eckersten, Thomas Gaiser, Balázs Grosz, Florian Heinlein, Belay T. Kassie, Kurt-Christian Kersebaum, Christian Klein, Matthias Kuhnert, Elisabet Lewan, Marco Moriondo, Claas Nendel, Eckart Priesack, Helene Raynal, Pier P. Roggero, Reimund P. Rötter, Stefan Siebert, Xenia Specka, Fulu Tao, Edmar Teixeira, Giacomo Trombi, Daniel Wallach, Lutz Weihermüller, Jagadeesh Yeluripati, Frank Ewert: Impact of Spatial Soil and Climate Input Data Aggregation on Regional Yield Simulations, Plos One, 11, e0151782, 2016.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., and Swanson, D. K.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth Syst. Sci. Data, 5, 3-13, 10.5194/essd-5-3-2013, 2013.

Ji, P., Yuan, X., and Liang, X.-Z.: Do Lateral Flows Matter for the Hyperresolution Land Surface Modeling?, Journal of Geophysical Research: Atmospheres, 122, 12,077-012,092, doi:10.1002/2017JD027366, 2017.

Johnston, R. M., Barry, S. J., Bleys, E., Bui, E. N., Moran, C. J., Simon, D. A. P., Carlile, P., McKenzie, N. J., Henderson, B. L., Chapman, G., Imhoff, M., Maschmedt, D., Howe, D., Grose, C., and Schoknecht, N.: ASRIS: the database, Australian Journal of Soil Research, 416, 1021-1036, 2003.

Instituto Nacional de Estadística y Geografía: Conjunto de Datos de Perfiles de Suelos Escala 1: 250 000 Serie II (Continuo Nacional), INEGI, Aguascalientes, Ags. Mexico, 2016.

Jordan, H., Tom, G., Jens, H., and Janine, B.: Compiling and Mapping Global Permeability of the Unconsolidated and Consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0), Geophysical Research Letters, 45, 1897-1904, doi:10.1002/2017GL075860, 2018.

Karssies, L.: CSIRO National Soil Archive and the National Soil Database (NatSoil). No. v1 in Data Collection, CSIRO, Canberra, 2011.

Kearney, M. R., and Maino, J. L.: Can next-generation soil data products improve soil moisture modelling at the continental scale? An assessment using a new microclimate package for the R programming environment, Journal of Hydrology, 561, 662-673, https://doi.org/10.1016/j.jhydrol.2018.04.040, 2018.

Koster, R. D., and Suarez, M. J.: Modeling the land surface boundary in climate models as a composite of independent vegetation stands, Journal of Geophysical Research: Atmospheres, 97, 2697-2715, doi:10.1029/91JD01696, 1992.

Kuhnert, M., Yeluripati, J., Smith, P., Hoffmann, H., van Oijen, M., Constantin, J., Coucheney, E., Dechow, R., Eckersten, H., Gaiser, T., Grosz, B., Haas, E., Kersebaum, K.-C., Kiese, R., Klatt, S., Lewan, E., Nendel, C., Raynal, H., Sosa, C., Specka, X., Teixeira, E., Wang, E., Weihermüller, L., Zhao, G., Zhao, Z., Ogle, S., and Ewert, F.: Impact analysis of climate data aggregation at different spatial scales on simulated net primary productivity for croplands, European Journal of Agronomy, 88, 41-52, https://doi.org/10.1016/j.eja.2016.06.005, 2017.

Lawrence, P. J., and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), Journal of Geophysical Research, 112, 10.1029/2006JG000168, 2007.

Leenaars, J. G. B.: 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 report 2012/03, Africa Soil Information Service (AfSIS) project and ISRIC - World Soil Information, Wageningen, the Netherlands, 2012.
Lei, H., Yang, D., and Huang, M.: Impacts of climate change and vegetation dynamics on runoff in the mountainous region of the Haihe River basin in the past five decades, Journal of Hydrology, 511, 786-799, http://dx.doi.org/10.1016/j.jhydrol.2014.02.029, 2014.
Li, C., Lu, H., Yang, K., Wright, J. S., Yu, L., Chen, Y., Huang, X., and Xu, S.: Evaluation of the Common Land Model (CoLM) from the Perspective of Water and Energy Budget Simulation: Towards Inclusion in CMIP6, Atmosphere, 8, 141, 2017.
Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, Journal of Geophysical Research: Atmospheres, 99, 14415-14428, doi:10.1029/94JD00483, 1994.
Livneh, B., Kumar, R., and Samaniego, L.: Influence of soil textural properties on hydrologic fluxes in the Mississippi river basin, Hydrological Processes, 29, 4638-4655, dx.doi.org/10.1002/hyp.10601, 2015.
Looy, K. V., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y. A., Padarian, J., Schaap, M. G., Tóth, B., Verhoef, A., Vanderborght, J., Ploeg, M. J., Weihermüller, L., Zacharias, S., Zhang, Y., and Vereecken, H.: Pedotransfer Functions in Earth System Science: Challenges and Perspectives, Reviews of Geophysics, 55, 1199-1256, doi:10.1002/2017RG000581, 2017.
Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E. A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J. W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R. B., Jones, C. D., Lara, M. J., Liang, J., McGuire, A. D., Parton, W., Peng, C., Randerson, J. T., Salazar, A., Sierra, C. A., Smith, M. J., Tian, H., Todd-Brown, K. E. O., Torn, M., van Groenigen, K. J., Wang, Y. P., West, T. O., Wei, Y., Wieder, W. R., Xia, J., Xu, X., Xu, X., and Zhou, T. C. G. B.: Toward more realistic projections of soil carbon dynamics by Earth system models, Global Biogeochemical Cycles, 30, 40-56, 10.1002/2015gb005239, 2016.
MacDonald, K. B., and Valentine, K. W. G.: CanSIS/NSDB. A general description (Centre for Land and Biological Resources Research), Research Branch, Agriculture Canada, Ottawa, 1992.
Marwa, T., Bertrand, G., and Christine, H.: Large Differences in Global and Regional Total Soil Carbon Stock Estimates Based on SoilGrids, HWSD, and NCSCD: Intercomparison and Evaluation Based on Field Data From USA, England, Wales, and France, Global Biogeochemical Cycles, 32, 42-56, doi:10.1002/2017GB005678, 2018.
McBratney, A. B., Santos, M. L. M., and Minasny, B.: On digital soil mapping, Geoderma, 117, 3-52, 10.1016/s0016-7061(03)00223-4, 2003.
McBratney, A. B., Minasny, B., and Tranter, G.: Necessary meta-data for pedotransfer functions, Geoderma, 160, 627-629, 2011.
McGuire, A. D., Melillo, J. M., Kicklighter, D. W., Pan, Y. D., Xiao, X. M., Helfrich, J.,

SOIL Discuss., https://doi.org/10.5194/soil-2018-32
Manuscript under review for journal SOIL
Discussion started: 15 October 2018

Moore, B., Vorosmarty, C. J., and Schloss, A. L.: Equilibrium responses of global net primary production and carbon storage to doubled atmospheric carbon dioxide: sensitivity to changes in vegetation nitrogen concentration, Global Biogeochem. Cycles, 11, 173-189, 1997.

Melton, J. R., Sospedra-Alfonso, R., and McCusker, K. E.: Tiling soil textures for terrestrial ecosystem modelling via clustering analysis: a case study with CLASS-CTEM (version 2.1), Geosci. Model Dev., 10, 2761-2783, 10.5194/gmd-10-2761-2017, 2017.

Miller, D. A., and White, R. A.: A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling, Earth Interactions, 2, 1-26, 10.1175/1087-3562(1998)002<0001:ACUSMS>2.3.CO;2, 1998.

Montzka, C., Herbst, M., Weihermüller, L., Verhoef, A., and Vereecken, H.: A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves, Earth Syst. Sci. Data, 9, 529-543, 10.5194/essd-9-529-2017, 2017.

Mulder, V. L., Lacoste, M., Richer-de-Forges, A. C., and Arrouays, D.: GlobalSoilMap France: High-resolution spatial modelling the soils of France up to two meter depth, Science of The Total Environment, 573, 1352-1369, http://dx.doi.org/10.1016/j.scitotenv.2016.07.066, 2016.

NationalSoilSurveyOffice: Soil Map of China (in Chinese), China Map Press, Beijing, 1995.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research: Atmospheres, 116, doi:10.1029/2010JD015139, 2011.

Odgers, N. P., Libohova, Z., and Thompson, J. A.: 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, 2012.

Oleson, K. W., Lawrence, D. M., B, G., Flanner, M. G., Kluzek, E., J., P., Levis, S., Swenson, S. C., Thornton, E., Feddema, J., Heald, C. L., Lamarque, J.-f., Niu, G.-y., Qian, T., Running, S., Sakaguchi, K., Yang, L., Zeng, X., and Zeng, X.: Technical Description of version 4.0 of the Community Land Model (CLM). NCAR Technical Note NCAR/TN-478+STR, National Center for Atmospheric Research, Boulder, 257, 2010.

Oleson, K. W., D.M. Lawrence, G.B. Bonan, B. Drewniak, M. Huang, C.D. Koven, S. Levis, F. Li, W.J. Riley, Z.M. Subin, S.C. Swenson, P.E. Thornton, A. Bozbiyik, R. Fisher, E. Kluzek, J.-F. Lamarque, P.J. Lawrence, L.R. Leung, W. Lipscomb, S. Muszala, D.M. Ricciuto, W. Sacks, Y. Sun, J. Tang, Z.-L. Yang: Technical Description of version 4.5 of the Community Land Model (CLM). Ncar Technical Note NCAR/TN-503+STR, National Center for Atmospheric Research, Boulder, CO, 422, 2013.

Oz, B., V. Deutsch, C., and Frykman, P.: A visualbasic program for histogram and variogram scaling, Computers & Geosciences, 28, 21-31, http://dx.doi.org/10.1016/S0098-3004(01)00011-5, 2002.

Park, J., Kim, H.-S., Lee, S.-J., and Ha, T.: Numerical Evaluation of JULES Surface Tiling Scheme with High-Resolution Atmospheric Forcing and Land Cover Data, SOLA, 14, 19-24, 10.2151/sola.2018-004, 2018.

Patterson, K. A.: Global distributions of total and total-avaiable soil water-holding capacities, Master, University of Delawar, Newark, DE, 1990.

Pelletier, J. D., P. D. Broxton, P. Hazenberg, X. Zeng, P. A. Troch, G.-Y. Niu, Z. Williams, M. A. Brunke, and D. Gochis: A gridded global data set of soil, immobile regolith, and sedimentary deposit thicknesses for regional and global land surface modeling, Journal of Advances in Modeling Earth Systems, 8, 10.1002/2015MS000526, 2016.

Post, D. F., Fimbres, A., Matthias, A. D., Sano, E. E., Accioly, L., Batchily, A. K., and Ferreira, L. G.: Predicting Soil Albedo from Soil Color and Spectral Reflectance Data, Soil Science Society of America Journal 64, 1027-1034, 2000.

Quattrochi, D. A., Emerson, C. W., Lam, N. S.-N., and Qiu, H.-l.: Fractal Characterization of Multitemporal Remote Sensing Data, in: Modelling Scale in Geographical Information System, edited by: Tate, N., and Atkinson, P., John Wiley & Sons, Lodon, 13-34, 2001.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., and Thompson, J.: Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution, Soil Science Society of America Journal, 82, 186-201, 10.2136/sssaj2017.04.0122, 2018.

Reynolds, C. A., Jackson, T. J., and Rawls, W. J.: 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, 2000.

Romanowicz, A. A., Vanclooster, M., Rounsevell, M., and Junesse, I. L.: Sensitivity of the SWAT model to the soil and land use data parametrisation: a case study in the Thyle catchment, Belgium, Ecological Modelling, 187, 27-39, 2005.

Rosenzweig, C., and Abramopoulos, F.: Land surface model development for the GISS GCM, J. Climate, 10, 2040-2054, 1997.

Ross, C. W., Prihodko, L., Anchang, J., Kumar, S., Ji, W., and Hanan, N. P.: HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling, Scientific Data, 5, 180091, 10.1038/sdata.2018.91, 2018.

Sanchez, P. A., Ahamed, S., Carré, F., Hartemink, A. E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A. B., McKenzie, N. J., Mendonça-Santos, M. d. L., Budiman Minasny, L. M., Okoth, P., Palm, C. A., Sachs, J. D., Shepherd, K. D., Vågen, T.-G., Vanlauwe, B., Walsh, M. G., Winowiecki, L. A., and Zhang, G.-L.: Digital soil map of the world, Science, 325, 680-681, 2009.

Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., Zhang, C., Collelo, G. D., and Bounoua, L.: A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: model formulation, Journal of Climate, 9, 676-705, 1996.

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

Shangguan, W., Dai, Y., Liu, B., Zhu, A., Duan, Q., Wu, L., Ji, D., Ye, A., Yuan, H., Zhang, Q., Chen, D., Chen, M., Chu, J., Dou, Y., Guo, J., Li, H., Li, J., Liang, L., Liang, X., Liu, H., Liu, S., Miao, C., and Zhang, Y.: A China dataset of soil properties for land surface

modeling, Journal of Advances in Modeling Earth Systems, 5, 212-224, 10.1002/jame.20026, 2013.
Shangguan, W.: Comparison of aggregation ways on soil property maps, 20th World Congress of Soil Science, Jeju, Korea, 2014,
Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H.: A global soil data set for earth system modeling, Journal of Advances in Modeling Earth Systems, 6, 249-263, 2014.
Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., and Dai, Y.: Mapping the global depth to bedrock for land surface modeling, Journal of Advances in Modeling Earth Systems, 9, 65-88, 10.1002/2016ms000686, 2017.
Shoba, S. A., Stolbovoi, V. S., Alyabina, I. O., and Molchanov, E. N.: Soil-geographic database of Russia, Eurasian Soil Science, 41, 907-913, 10.1134/s1064229308090019, 2008.
Singh, R. S., Reager, J. T., Miller, N. L., and Famiglietti, J. S.: Toward hyper-resolution land-surface modeling: The effects of fine-scale topography and soil texture on CLM4.0 simulations over the Southwestern U.S, Water Resources Research, 51, 2648-2667, doi:10.1002/2014WR015686, 2015.
Slevin, D., Tett, S. F. B., Exbrayat, J. F., Bloom, A. A., and Williams, M.: Global evaluation of gross primary productivity in the JULES land surface model v3.4.1, Geosci. Model Dev., 10, 2651-2670, 10.5194/gmd-10-2651-2017, 2017.
Soil Survey Staff, N. R. C. S., United States Department of Agriculture: Web Soil Survey. Available online at http://websoilsurvey.nrcs.usda.gov/. Accessed 1/1/2017, 2017.
Soil Landscapes of Canada Working Group: Soil Landscapes of Canada version 3.2., Agriculture and Agri-Food Canada, Ottawa, Ontario, 2010.
Stoorvogel, J. J., Bakkenes, M., Temme, A. J. A. M., Batjes, N. H., and Brink, B. J. E.: S-World: A Global Soil Map for Environmental Modelling, Land Degradation & Development, 28, 22-33, doi:10.1002/ldr.2656, 2017.
Thompson, J. A., Prescott, T., Moore, A. C., Bell, J., Kautz, D. R., Hempel, J. W., Waltman, S. W., and Perry, C. H.: Regional approach to soil property mapping using legacy data and spatial disaggregation techniques, 19th World Congress of Soil Science, Brisbane, Queensland, 2010,
Thornton, P. E., and Rosenbloom, N. A.: Ecosystem model spin-up: estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model, Ecological Modelling, 189, 25-48, 2005.
Tian, W., Li, X., Wang, X. S., and Hu, B. X.: Coupling a groundwater model with a land surface model to improve water and energy cycle simulation, Hydrol. Earth Syst. Sci. Discuss., 2012, 1163-1205, 10.5194/hessd-9-1163-2012, 2012.
Todd-Brown, K. E. O., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A. G., and Allison, S. D.: Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, Biogeosciences, 10, 1717-1736, 10.5194/bg-10-1717-2013, 2013.
Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., and Tóth, G.: New generation of hydraulic pedotransfer functions for Europe, European Journal of Soil Science, 66, 226-238, doi:10.1111/ejss.12192, 2015.

SOIL Discuss., https://doi.org/10.5194/soil-2018-32
Manuscript under review for journal SOIL
Discussion started: 15 October 2018

Tóth, B., Weynants, M., Pásztor, L., and Hengl, T.: 3D soil hydraulic database of Europe at 250 m resolution, Hydrological Processes, 31, 2662-2666, doi:10.1002/hyp.11203, 2017.

Trinh, T., Kavvas, M. L., Ishida, K., Ercan, A., Chen, Z. Q., Anderson, M. L., Ho, C., and Nguyen, T.: Integrating global land-cover and soil datasets to update saturated hydraulic conductivity parameterization in hydrologic modeling, Science of The Total Environment, 631-632, 279-288, https://doi.org/10.1016/j.scitotenv.2018.02.267, 2018.

Van Engelen, V., and Dijkshoorn, J.: Global and National Soils and Terrain Digital Databases (SOTER), Procedures Manual, version 2.0. ISRIC Report 2012/04, ISRIC - World Soil Information, Wageningen, the Netherlands, 2012.

Vaysse, K., and Lagacherie, P.: Using quantile regression forest to estimate uncertainty of digital soil mapping products, Geoderma, 291, 55-64, https://doi.org/10.1016/j.geoderma.2016.12.017, 2017.

Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M. G., and Genuchten, M. T. v.: Using pedotransfer functions to estimate the van Genuchten-Mualem soil hydraulic properties: a review, Vadose Zone Journal, 9, 795-820, 2010.

Viscarra Rossel, R., Chen, C., Grundy, M., Searle, R., Clifford, D., and Campbell, P.: The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project, Soil Research, 53, 845-864, 2015.

Vrettas, M. D., and Fung, I. Y.: Toward a new parameterization of hydraulic conductivity in climate models: Simulation of rapid groundwater fluctuations in Northern California, Journal of Advances in Modeling Earth Systems, 7, 2105-2135, 10.1002/2015ms000516, 2016.

Wang, G., Gertner, G., and Anderson, A. B.: Up-scaling methods based on variability-weighting and simulation for inferring spatial information across scales, International Journal of Remote Sensing, 25, 4961- 4979, 2004.

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

Wilson, M. F., and Henderson-Sellers, A.: A global archive of land cover and soils data for use in general circulation climate models, Journal of Climatology, 5, 119-143, 1985.

Wu, L., Wang, A., and Sheng, Y.: Impact of Soil Texture on the Simulation of Land Surface Processes in China, Climatic and Environmental Research (in Chinese), 19, 559-571, doi:10.3878/j.issn.1006-9585.2013.13055, 2014.

Wu, X., Lu, G., Wu, Z., He, H., Zhou, J., and Liu, Z.: An Integration Approach for Mapping Field Capacity of China Based on Multi-Source Soil Datasets, Water, 10, 728, 2018.

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

Zhao, H., Zeng, Y., Lv, S., and Su, Z.: Analysis of soil hydraulic and thermal properties for land surface modeling over the Tibetan Plateau, Earth Syst. Sci. Data, 10, 1031-1061, 10.5194/essd-10-1031-2018, 2018a.

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., Chen, J.-H., Chen, X., Donner, L. J., Dunne, J. P., Dunne, K., Durachta, J., Fan, S.-M., Freidenreich, S. M., Garner, S. T., Ginoux, P., Harris, L. M., Horowitz, L. W., Krasting, J. P., Langenhorst, A. R., Liang, Z., Lin, P., Lin, S.-J., Malyshev, S. L., Mason, E., Milly, P. C. D., Ming, Y., Naik, V., Paulot, F., Paynter, D., Phillipps, P., Radhakrishnan, A., Ramaswamy, V., Robinson, T., Schwarzkopf, D., Seman, C. J., Shevliakova, E., Shen, Z., Shin, H., Silvers, L. G., Wilson, J. R., Winton, M., Wittenberg, A. T., Wyman, B., and Xiang, B.: The GFDL Global Atmosphere and Land Model AM4.0/LM4.0: 2. Model Description, Sensitivity Studies, and Tuning Strategies, Journal of Advances in Modeling Earth Systems, 10, 735-769, doi:10.1002/2017MS001209, 2018b.

Zheng, G., Yang, H., Lei, H., Yang, D., Wang, T., and Qin, Y.: Development of a Physically Based Soil Albedo Parameterization for the Tibetan Plateau, Vadose Zone Journal, 17, 10.2136/vzj2017.05.0102, 2018.

Zheng, H., and Yang, Z. L.: Effects of soil type datasets on regional terrestrial water cycle simulations under different climatic regimes, Journal of Geophysical Research: Atmospheres, Accepted, 10.1002/2016jd025187, 2016.

Zhou, T., Shi, P. J., Jia, G. S., Dai, Y. J., Zhao, X., Shangguan, W., Du, L., Wu, H., and Luo, Y. Q.: Age-dependent forest carbon sink: Estimation via inverse modeling, Journal of Geophysical Research-Biogeosciences, 120, 2473-2492, 10.1002/2015jg002943, 2015.

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

Table 1. Lists of the soil dataset used by land surface models (LSM) of Earth System Models (ESM) or climate models (CM)

ESM or CM	LSM	Input soil data	Dataset Used	Resolution
CESM **CCSM**	**CLM** (Oleson, 2013)	Soil texture (sand, clay)	IGBP-DIS (Global Soil DataTask, 2000)	5′
		Soil organic matter	ISRIC-WISE (Batjes, 2006) combined with NCSD (Hugelius et al., 2013)	5′; 0.25°
		Soil color class	dataset derived from MODIS (Lawrence and Chase, 2007)	0.05°
MPI-ESM **ICON-ESM**	**JSBACH** (Giorgetta et al., 2013)	Volumetric heat capacity and thermal diffusivity prescribed for 5 soil types OF FAO soil map	Dataset by Hagemann et al. (1999)	0.5°;
		Field capacity, Plant-available soil water holding capacity and wilting point prescribed for ecosystem type	Dataset by Hagemann (2002) based on Patterson (1990) and Hagemann et al. (1999)	0.5°
		Soil albedo	Dataset by Hagemann (2002)	0.5° (8km over Africa)
HadGEM2-ES **QUEST**	**JULES/MOSES** (Best et al., 2011;Clark et al., 2011)	Soil texture (Cox et al., 1999) (Blyth, 2006)	Dataset by Wilson and Henderson-Sellers (1985)	1°
			HWSD (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012)	30”
RegCM	**BATS** (Dickinson et al., 1993) or **CLM**3.5	Soil texture classes Soil color classes	Prescribed for BATS vegetation/land cover type (Elguindi et al., 2014).	
GFDL ESM	**GFDL LM4** (Zhao et al., 2018b)	soil texture type	HWSD (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012)	30”

SOIL Discuss., https://doi.org/10.5194/soil-2018-32
Manuscript under review for journal SOIL
Discussion started: 15 October 2018

GISS GCM ModelE	**GISS-LSM** (Rosenzweig and Abramopoulos, 1997)	soil textures	Dataset by Webb et al. (1993) and Zöbler (1986)	1°
WRF **CWRF**	**Noah** (Chen and Dudhia, 2001) **Noah-MP** (Niu et al., 2011) **CLM4** Other LSMs	Soil texture	FAO (outside Conterminous US) STATSGO (Miller and White, 1998)	5’ 30”
			GSDE (Shangguan et al., 2014)	30”
CAS_ESM **BNU_ESM** **GRAPES**	**CoLM** (Dai et al., 2003)	Soil texture, gravel, soil organic carbon, bulk density	GSDE (Shangguan et al., 2014)	30”
	SiB (Sellers et al., 1996; Gurney et al., 2008)	Soil texture classes	Dataset by Zöbler (1986)	1°
GLDAS	**Mosaic** (Koster and Suarez, 1992) **CLM2** **Noah** **VIC** (Liang et al., 1994)	Soil texture classes	Dataset by Reynolds et al. (2000)	5′
ACCESS-ESM	**CABLE**	Soil texture classes	Dataset by Zöbler (1986)	1°

CLM = Community Land Model. The current released version is CLM4.5.
CCSM = Community Climate System Model.
CESM = Community Earth System Model.
IGBP-IDS = Data and Information System of International Geosphere-Biosphere Programme
ISRIC-WISE = World Inventory of Soil Emission Potentials of International Soil Reference and Information Centre
NCSD = Northern Circumpolar Soil Carbon Database
MPI-ESM = The Max Planck Institute for Meteorology Earth System Model

ICON-ESM = icosahedral non-hydrostatic Earth System Model
JSBACH = Jena Scheme of Atmosphere Biosphere Coupling in Hamburg
HWSD = Harmonized World Soil Database
FAO = the Food and Agriculture Organization (FAO-UNESCO) digital Soil Map of the World (SMW) at 1:5 million scale.
HadGEM2-ES = Hadley Global Environment Model 2 - Earth System
QUEST = Quantifying and Understanding the Earth System
JULES/MOSES= Joint UK Land Environment Simulator/Met Office Surface Exchange Scheme
RegCM = Regional Climate Model
BATS = Biosphere-Atmosphere Transfer Scheme
GFDL = Geophysical Fluid Dynamics Laboratory
GISS = Goddard Institute for Space Studies
STATSGO = the State Soil Geographic Database
WRF = Weather Research and Forecasting Model
Noah-MP = Noah-multiparameterization
GSDE = Global Soil Dataset for Earth System Model
CAS_ESM = Chinese Academy of Sciences Earth System Model
BNU_ESM = Beijing Normal University Earth System Model
GRAPES = Global/Regional Assimilation Prediction System
CoLM = Common Land Model. The current version is CoLM2014.
SiB = Simple Biopshere Model
GLDAS = Global Land Data Assimilation System
ACCESS = Australia Community Climate and Earth System Simulator
CABLE = Community Atmosphere Biosphere Land Exchange

Table 2 Three new global soil datasets for the updates of ESMs.

Dataset	Resolution	Number of layers	Number of properties	depth to the bottom of a layer (cm)	Mapping method
HWSD	1km	2	22	30, 100	Linkage method
GSDE	1km	8	39	0, 4.5, 9.1, 16.6, 28.9, 49.3, 82.9, 138.3, 229.6	Linkage method
Soilgrids	250m	6	7	5, 15, 30, 60, 100, 200	Digital soil mapping

SOIL Discuss., https://doi.org/10.5194/soil-2018-32
Manuscript under review for journal SOIL
Discussion started: 15 October 2018

Table 3 Available soil properties of three global soil datasets.

Soil property*	HWSD	GSDE	Soilgrids	Soil property*	HWSD	GSDE	Soilgrids
Drainage class	√	√		Total carbon		√	
AWC class	√	√		Total nitrogen		√	
Soil phase	√	√		Total sulfur		√	
Impermeable layer	√	√		pH(KCL)		√	√
Obstacle to roots	√	√		pH(Cacl$_2$)		√	
Additional property	√	√		Exchangeable Ca		√	
Soil water regime	√	√		Exchangeable Mg		√	
Reference soil depth	√	√		Exchangeable K		√	
Depth to bedrock			√	Exchangeable Na		√	
Gravel	√	√	√	Exchangeable Al		√	
Sand, Silt, Clay	√	√	√	Exchangeable H		√	
Texture class**	√			VWC at -10 kPa		√	
Bulk density	√	√	√	VWC at -33 kPa		√	
Organic Carbon	√	√	√	VWC at -1500 kPa		√	
pH(H_2O)	√	√	√	Phosphorous by Bray method		√	
CEC (clay)	√			Phosphorous by Olsen method		√	
CEC (soil)	√	√		Phosphorous by New Zealand method		√	
Base saturation	√	√		Water soluble phosphorous		√	
TEB	√			Phosphorous by Mechlich method		√	
Calcium Carbonate	√	√		Total phosphorous		√	
Gypsum	√	√		Total Potaasium		√	
Sodicity (ESP)	√			Salinity (ECE)	√	√	

*CEC is cation exchange capacity. The base saturation measures the sum of exchangeable cations (nutrients) Na, Ca, Mg and K as a percentage of the overall exchange capacity of the soil (including the same cations plus H and Al). TEB is total exchangeable base including Na, Ca, Mg and K. ESP is exchangeable sodium percentage, which is calculated as Na*100/CECsoil. ECE is electrical conductivity. AWC is the available water storage capacity. The first 9 soil properties on the left including drainage class, AWC class and so on are available for soil types, while the other properties are available for each layer.

**texture class can be calculated using sand, silt and clay content

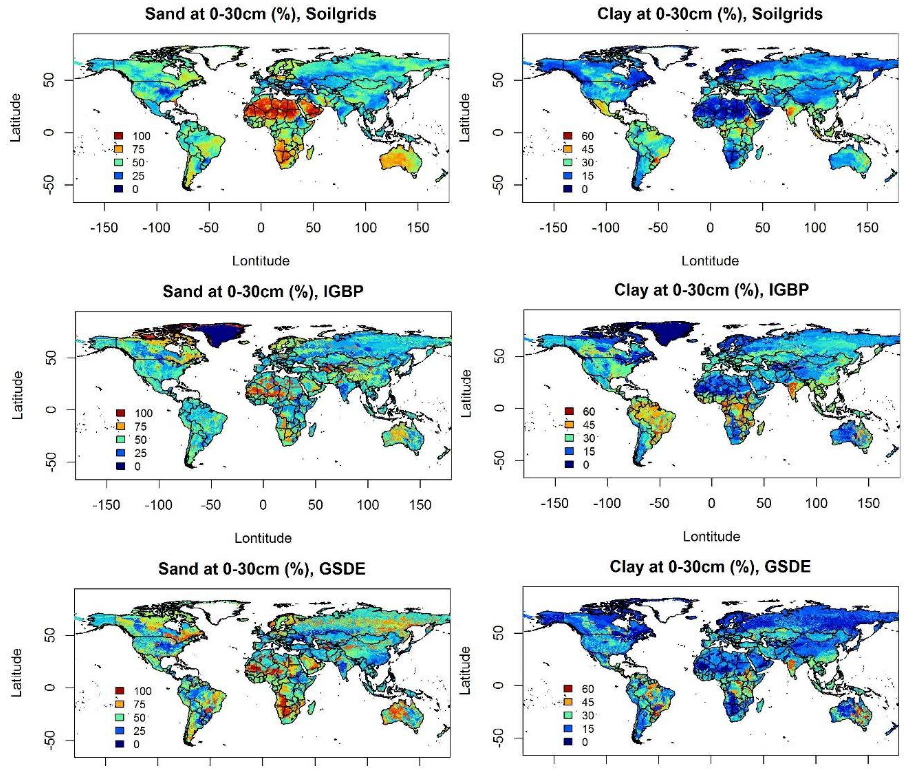

Figure 1 Soil sand and clay fraction at the surface 0-30 cm layer from Soilgrids, IGBP-DIS and GSDE. The difference among them will lead to different modeling results for ESMs. IGBP-DIS is Data and Information System of International Geosphere-Biosphere Programme, and GSDE is Global Soil Dataset for Earth System Model.

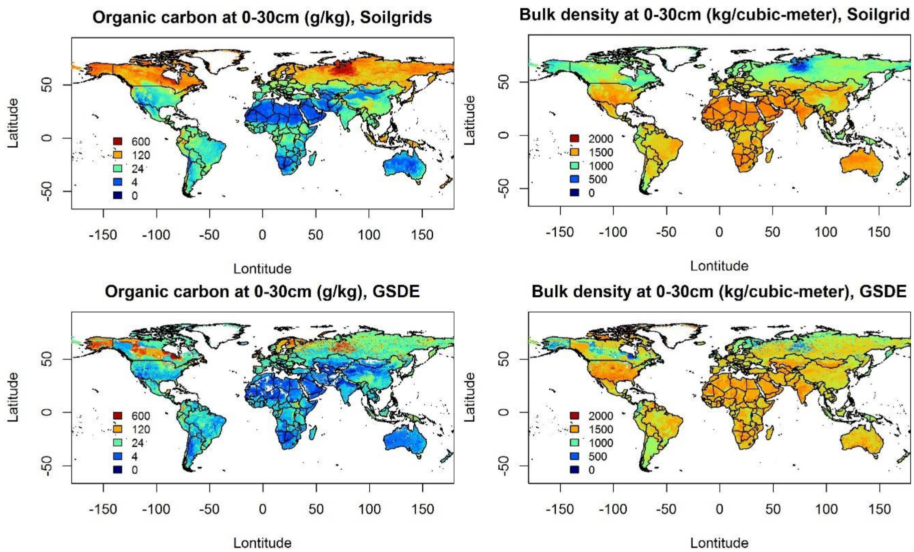

Figure 2 Soil organic carbon and bulk density at the surface 0-30 cm layer from Soilgrids and GSDE.