JOURNAL OF ADVANCES IN MODELING EARTH SYSTEMS, VOL. 5, 212–224, doi:10.1002/jame.20026, 2013

A China data set of soil properties for land surface modeling

Wei Shangguan,[1] Yongjiu Dai,[1] Baoyuan Liu,[2] Axing Zhu,[3] Qingyun Duan,[1]
Lizong Wu,[4] Duoying Ji,[1] Aizhong Ye,[1] Hua Yuan,[2] Qian Zhang,[2] Dongdong Chen,[2]
Ming Chen,[2] Jianting Chu,[2] Youjun Dou,[2] Jianxia Guo,[2] Haiqin Li,[2] Junjia Li,[2]
Lu Liang,[2] Xiao Liang,[1] Heping Liu,[2] Shuyan Liu,[2] Chiyuan Miao,[1] and Yizhou Zhang[2]

Received 20 October 2012; revised 18 February 2013; accepted 7 March 2013; published 6 May 2013.

[2] A comprehensive 30×30 arc-second resolution gridded soil characteristics data set of China has been developed for use in the land surface modeling. It includes physical and chemical attributes of soils derived from 8979 soil profiles and the Soil Map of China (1:1,000,000). We used the polygon linkage method to derive the spatial distribution of soil properties. The profile attribute database and soil map are linked under the framework of the Genetic Soil Classification of China which avoids uncertainty in taxon referencing. Quality control information (i.e., sample size, soil classification level, linkage level, search radius and texture) is included to provide "confidence" information for the derived soil parameters. The data set includes 28 attributes for 8 vertical layers at the spatial resolution of 30×30 arc-seconds. Based on this data set, the estimated storage of soil organic carbon in the upper 1 m of soil is 72.5 Pg, total N is 6.6 Pg, total P is 4.5 Pg, total K is 169.9 Pg, alkali-hydrolysable N is 0.55 Pg, available P is 0.03 Pg, and available K is 0.61 Pg. These estimates are reasonable compared with previous studies. The distributions of soil properties are consistent with common knowledge of Chinese soil scientists and the spatial variations over large areas are well represented. The data set can be incorporated into land models to better represent the role of soils in hydrological and biogeochemical cycles in China.

Citation: Shangguan, W. et al. (2013), A China data set of soil properties for land surface modeling, *J. Adv. Model. Earth Syst.*, *5*, 212–224, doi:10.1002/jame.20026.

1. Introduction

[3] As land surface models (LSMs) for use in numerical weather prediction models (NWPMs) and Earth system models (ESMs) become more sophisticated, they need more detailed information on physical and chemical properties of soil, including information on how soil properties change with depth in the soil and across geographic areas. The soil data sets now used in global land modeling were derived from the FAO/UNESCO (Food and Agriculture Organization, United Nations Educational, Scientific and Cultural Organization) world soil map (1:5 million) and limited soil profile data [*Batjes*, 2002, 2006; *Global Soil Data Task*, 2000; *Reynolds et al.*, 2000; *van Engelen et al.*, 2005; *Webb et al.*, 1991, 1993; *Wilson and Henderson-Sellers*, 1985; *Zobler*, 1986]. The development of these data sets used little data from the China national soil surveys in 1950s and later. A soil scientist, Pedro Sanchez, said "We know more about soils of Mars than about soils of Africa" (http://www.globalsoilmap.net/). The information about Chinese soils used by LSMs may be not better than that used for Africa. Thus, any results from ESMs/NWPMs that depend on these soil data are questionable.

[4] The soil map data and soil attribute data of the sampled profiles are two source databases for developing the spatial soil property data set. The soil map is composed of mapping units, and each mapping unit is composed of soil units. The linkage between soil unit and soil attribute is determined according to the classification of soils. Unfortunately, a variety of soil classification schemes have been developed by different organizations and sometimes with different purposes. Chinese scholars have developed two schemes, the Genetic Soil Classification of China (GSCC) and the Chinese Soil Taxonomy (CST) classification. However, most soil profile information in China has been referenced to the GSCC, which was used in the Second National Soil Survey of China. Although considerable

---

Additional Supporting Information may be found in the online version of this article.

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

[2] School of Geography, Beijing Normal University, Beijing, China.

[3] Department of Geography, University of Wisconsin, Madison, Wisconsin, USA.

[4] Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China.

1942-2466/13/10.1002/jame.20026

effort has been devoted to developing procedures for referencing between these schemes and other internationally more widely used ones [*Shi et al*., 2004, 2006a, 2006b, 2010], such "taxonomy referencing" is not needed here because our objective is not to develop maps for soil classification but for the physical and chemical properties of soil needed by a land model. Thus use of the GSCC rather than some other classification scheme introduces the least error in mapping these properties.

[5] Soil profiles have commonly been used to assign measured properties to each classification element (soil map unit), i.e., using "taxotransfer rules" that ignore the spatial variability within map units [*Batjes et al*., 1997; *Batjes*, 2002]. However, with enough profile information, it should be possible to include spatial pattern variability among soil polygons of the same map unit. *Shangguan et al*. [2012] have developed such a "polygon linkage" method, which is also used here.

[6] The Harmonized World Soil Database (HWSD) is a newly released global soil data set [*FAO/IIASA/ISRIC/ISS-CAS/JRC*, 2009, 2012] available for land modeling use. It was established by combining existing regional and national updates of soil information, in particular, the Soil Map of China at 1:1 million [*National Soil Survey Office*, 1995; *Shi et al*., 2004]. The Soil Map of China was polygon-mapped using the GSCC, most detailed at the soil family level. This paper uses the HWSD for reference in evaluating our efforts.

[7] The HWSD only incorporates the Soil Map of China, but not the abundant soil profile information. In the HWSD, version 2.0 of WISE (World Inventory of Soil Emission Potentials) database (comprising 9607 profiles in the world and about 60 profiles in the China domain), has been used to derive topsoil and subsoil parameters using uniform taxonomy-based pedotransfer (taxotransfer) rules. WISE provides very limited soil attributes, and lacks attributes such as consistence, structure, total P, total K, and exchangeable cations.

[8] This paper describes the development of the China soil characteristics data set for use in land models in NWPMs/ESMs. This effort is unique in that, for the first time, the China NWPMs/ESMs community will have access to a data set of physical and chemical properties of soil specifically designed for modeling applications. This work should provide an infrastructure for further development of soil data for land modeling use, including easy and reliable future incorporation into it of more soil profile data and higher resolution soil map of China.

2. Materials and Methodology

2.1. China Soil Profiles and Soil Map

[9] The soils in 2444 counties, 312 national farms, and 44 forest farms within China were surveyed during the Second National Soil Survey (1979–1985). Using this survey, a soil map at a scale of 1:1 million for China was published. It is the most detailed soil map at the national level. Soil profile data are in six monographs in hard copy [*National Soil Survey Office*, 1996] at the national level, dozens at provincial level and thousands at prefectural and county level. With these data, it has become possible to build a comprehensive grid-based China soil data set for land surface modeling.

[10] The soil map of China was digitized by the Institute of Soil Science of the Chinese Academy of Sciences. Soil map units are delineated using the GSCC classification. GSCC exists in a four-level hierarchical structure from low level to high level (family, subgroup, great group and order). There are 12 orders, 61 great groups, 235 subgroups, and 909 families in the database [*Shi et al*., 2004]. However, there are only 925 soil map units, which are at family, subgroup and great group levels, and 11 nonsoil map units (i.e., glacier, river, lake and man-made reservoir, rock debris/detritus, coral reef and islet, salt desert and crust, coastal salt marsh, in-river sand bar and islet, urban and built-up lands, coastal aquatic farm, and coastal ocean) in the soil map. Each map unit has only one component in the soil map. There are 94,303 polygons in the soil map with 85,257 soil polygons.

[11] A tedious and labor intensive effort was needed to produce the soil profile database, which has taken almost 8 years and over 50 people to collect, digitize, standardize and geo-reference the soil profiles. All the soil books at the national and provincial levels were collected, and the soil books at prefectural and county levels of Tibet were collected (S-I in Supporting Information). These soil books, most covered by dust, were obtained from various libraries and old bookstores. Dozens of people spent countless hours to search and digitize their related data using a uniform procedure. The digitized data were thoroughly checked and quality controlled to reduce mistakes and redundancy. Repeated soil profiles in books from different administrative levels were combined to yield 8979 distinct soil profiles with 33,010 horizons. There are about 3.7 horizons per profile on average. The coordinates of the soil profiles were retrieved manually from a description of the site location point by point using the 1:250,000 topographic maps of China and administrative maps (Figure 1). Table 1 shows the site information of soil profiles and the soil characteristics for each horizon of a profile, including physical properties, chemical properties, and fertility.

2.2. Data Processing

[12] The soil database of China could not be directly used for regional land surface modeling without first dealing with some issues including:

[13] (1) Soil particle size distribution is given under the International Society of Soil Science (ISSS) and the Katschinski's schemes [*Katschinski*, 1956]. However, most LSMs require soil texture data in the FAO-USDA (United States Department of Agriculture) System.

[14] (2) Soil profiles differ from each other in terms of number, sequence, thickness and depth to the top and bottom soil layers. In addition, not all the layers have data for all soil characteristics. There are more data available for layers near the surface than for deeper layers.

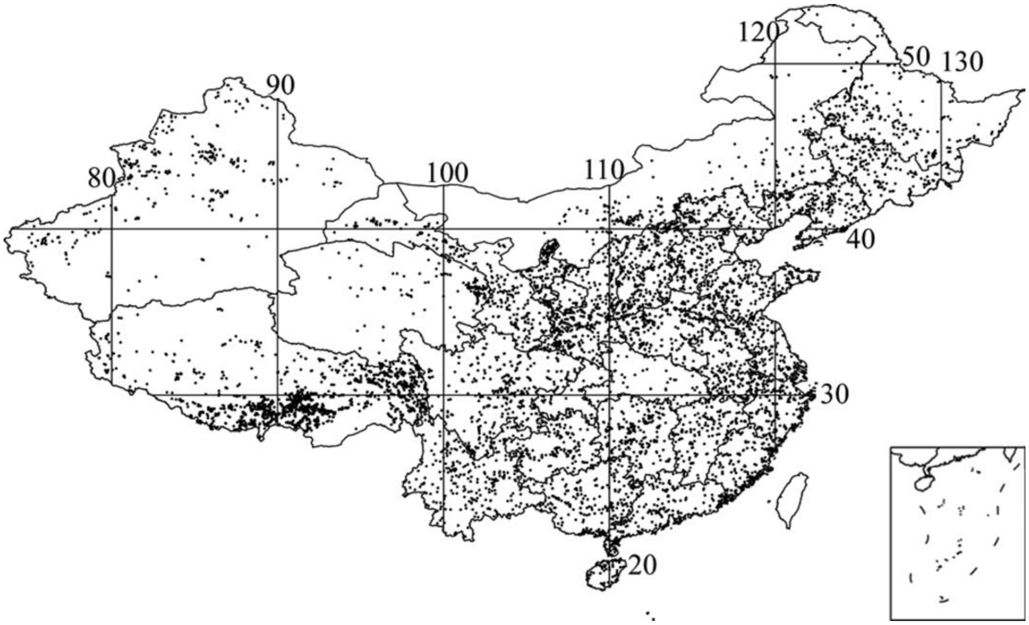

Figure 1. Geographic distribution of soil profiles. The number of soil profiles from different provinces varies

[15] (3) About three fourths of the soil profiles have a depth less than 1 m, and 90% of them have a depth less than 1.5 m. These depths of soil profiles are the depths to which the soils were examined, but in most cases, they are not the depth-to-bedrock.

[16] (4) The soil map provides a horizontal pattern of soil type information but soil profiles provide a vertical variation of soil characteristics at point locations. There is no spatially continuous information of soil properties.

[17] (5) Many LSMs require uniform grid cells or raster format, while map units of the 1:1 million soil map of China are defined as polygons in a vector format.

[18] The aim of this work is to derive a coverage map for soil characteristics based on the legacy soil data of China that can be conveniently used by regional modelers. In this section, we describe methods for preparing the soil data to ensure it will be suitable for land modeling purposes.

[19] The original ISSS and Katschinski particle-size distribution data could not be used by most LSMs so they were converted to the FAO-USDA System using several particle-size distribution models [*Shangguan and Dai*, 2009].

[20] The soil profile data set lacked soil characteristics information for some layers. To achieve vertical completeness of soil properties, we filled these data gaps in soil profiles. The gap filling was based on the assumption that neighboring layers have similar soil properties that change gradually with depth. Abrupt change of soil properties may happen in nature, but our assumption is more realistic than assigning values arbitrarily.

[21] For quantitative soil properties that do not change monotonically with depth, including particle-size distribution, rock fragment, pH value, bulk density, porosity, total K, exchangeable Al^{3+}, Ca^{2+}, Mg^{2+}, Na^{+}, and CEC, we filled the gaps with values from their neighboring layers. Figure 2 shows three data gap filling treatments for the soil profiles. If the first layer lacked data but the second layer had data, values of the soil properties of the second layer were assigned to the first layer. If some middle layers lacked data but had two neighboring layers with data, the average of these neighbors was assigned to the layer that lacked data. If only one neighboring layer had data, values from the neighbor were assigned to the layer lacking data. The above process was done first from the top to the bottom layer and then from the bottom to the top layer to provide every layer with data.

[22] For quantitative soil properties that change monotonically with depth, including soil organic matter, total N, total P, alkali-hydrolysable N, available P, available K exchangeable H^{+} and K^{+}, we filled in data-lacking layers using a linear depth weighting method. We assumed that the soil property of a layer was represented by the value of the center of the layer. The soil properties of a natural soil layer (A) were derived through the following relationship:

$$\frac{A_{\mathrm{i}}-A_{i-1}}{A_{\mathrm{i}+1}-A_{i}}=\frac{d_{i-1}}{d_{i}}, \tag{1}$$

where i is the ith layer, and d is the distance between the centers of soil layers, which was calculated by the following equation:

$$d_{i}=\frac{b_{\mathrm{i}+1}-b_{i-1}}{2}, \tag{2}$$

where b is the depth to the bottom of a layer. The soil properties of a data-lacking layer were calculated by

Table 1. List of Information of Soil Profile Data

Site Information

1. Sampled profile ID.
2. Source of data (national, provincial, prefectural or county level books).
3. Soil group (classification of Genetic Soil Classification of China (GSCC)).
4. Soil subgroup (GSCC).
5. Soil family (GSCC).
6. Soil species (GSCC).
7. Sampling date (year/month/day).
8. Site location description (provincial, prefectural, county and landform).
9. Longitude and latitude (derived from site location description).
10. Precision of geographic coordinates (classified as three classes according to estimated errors: < 15 km, 15–60 km, and >60 km).
11. Mean annual temperature (°C)
12. Annual precipitation (mm)
13. Accumulated temperature (≥ 10°C) (degree-days).
14. Frost-free days (days).
15. Soil parent material.
16. Water table depth (m).
17. Elevation (m).
18. Topography description.
19. Slope gradient (degree)
20. Slop orientation (degree, clockwise from North).
21. Vegetation coverage.

Number	Attribute	Unit	Number of records[a]	Maps[b]
Measured Physical and Chemical Attributes in Horizon Layers				
1	Horizon thickness	cm	32,208	Figure S1
2	pH value (H_2O)	pH units	29,668	Figure S2
3	Soil organic matter	g/100g	30,018	Figure S3
4	Total N	g/100g	29,237	Figure S4
5	Total P	g/100g	28,226	Figure S5
6	Total K	g/100g	22,910	Figure S6
7	Alkali-hydrolysable N	mg/kg	12,533	Figure S7
8	Available P	mg/kg	17,920	Figure S8
9	Available K	mg/kg	17,976	Figure S9
10	Cation exchange capacity (CEC)	me/100 g	22,327	Figure S10
11	Exchangeable H^+	me/100 g	2,060	Figure S11
12	Exchangeable Al^{3+}	me/100 g	2,021	Figure S12
13	Exchangeable Ca^{2+}	me/100 g	3,470	Figure S13
14	Exchangeable Mg^{2+}	me/100 g	3,417	Figure S14
15	Exchangeable K^+	me/100 g	3,380	Figure S15
16	Exchangeable Na^+	me/100 g	3,327	Figure S16
17	Soil texture		28,580	
18	Particle-size distribution[c]	g/100g	28,903	Figure S17–19
19	Rock fragment	g/100g	6,374	Figure S20
20	Bulk density	g/cm^3	4,296	Figure S21
21	Porosity	cm^3/cm3	2,247	Figure S22
22	Color (water condition unclear) [c]	hue, value, chroma	8,070	
23	Dry color[d]	hue, value, chroma	7,334	Figure S23
24	Wet color[d]	hue, value, chroma	11,140	Figure S23
25	Dominant structure[e]		29,343	Figure S24
26	Second structure[e]		866	Figure S24
27	Consistency		26,219	Figure S25
28	Root abundance description		23,998	Figure S26

[a] There are 33,010 records in total.

[b] Maps of soil properties interpolated in land model standard layers (Figures S1–S26 in Supporting Information).

[c] Soil particle size distribution was given under the International Society of Soil Science (ISSS) and the Katschinski's schemes with different separating limits.

[d] These three color are the soil colors described with "unclear", dry and wet water condition.

[e] For soil with multiple structure classes, we retain the first two.

transformation of equation (1). For brevity, we do not list the transformation for all the situations that were encountered. Two layers above and two layers below a layer were used at the most. If negative values appeared, they were set to zero.

[23] The soil characteristics of soil profiles are divided into eight standard layers (i.e., 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) for convenience of use in the Common Land Model (CoLM) [*Dai et al.*, 2003] and the Community Land Model (CLM) [*Oleson et al.*, 2004]. Because the first two layers of CoLM/CLM are too thin, they were combined. Since the last layer of CoLM/CLM has no data for almost all soil profiles, it

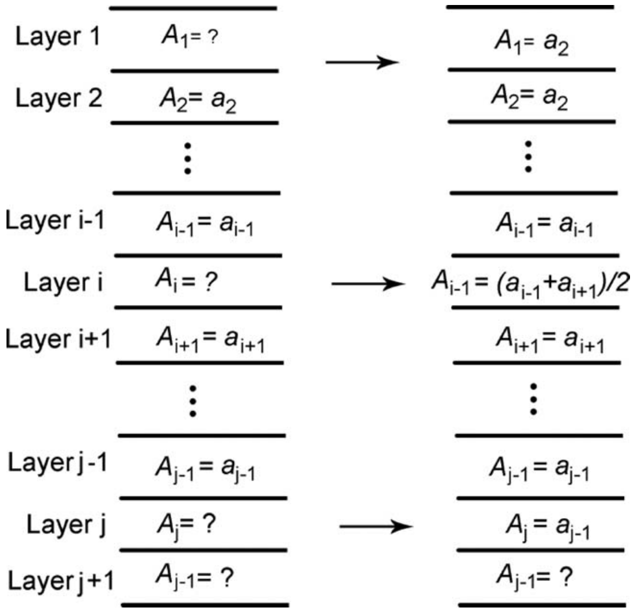

Figure 2. Three data gap filling treatments for the soil

was excluded. For brevity and comparison with other data sets, we also use a two-layer scheme (i.e., 0–0.3 m and 0.3–1 m) to display the results.

[24] For quantitative soil properties, the data were interpolated from natural soil horizons to the standard layers using the equal-area quadratic smoothing spline functions, which proved to be advantageous in predicting the depth function of soil properties including soil pH, electrical conductivity, clay content, organic carbon content, and gravimetric water content [*Bishop et al.*, 1999; *Malone et al.*, 2009; *Odgers et al.*, 2012]. This method guarantees mass conservation for a soil property of a layer under the assumption of continuous vertical variation of soil properties. The smoothing parameter of the spline was set as 0.1 [*Bishop et al.*, 1999]. The spline was then used to estimate the values of soil properties in the standard layers. Negative values were set to zero.

[25] For categorical soil properties that cannot be converted into quantitative values (including consistency and structure), the percentage of each class was calculated; some of the categorical soil properties, i.e., color and root abundance, were converted into quantitative values before they were interpolated into standard layers. There are very few data available for root size, so this property was not retained in the final data set. Soil color is represented by the Munsell notation with three dimensions: hue, value, and chroma [*Kuehni*, 2002]. Value and chroma are quantitative but not hue. The dimension of hue is a horizontal circle, which is divided into five principal hues: red, yellow, green, blue, and purple, along with five intermediate hues halfway between adjacent principal hues. These hues were represented by numbers between 0 and 10 when they were converted into quantitative values. However, as the dimension of hue is a circle, these numbers were converted into vectors before the equal-area spline as follows:

$$H(x,y)=\left(\cos\frac{2\pi h}{10},\sin\frac{2\pi h}{10}\right), \tag{3}$$

where h is hue represented as numbers, which is reversible. After the calculation of equation (3), the vector was converted back into numbers, and the numbers were converted back into hues.

[26] As there is no information about depth-to-bedrock, we only tabulated the soil profile depth for each soil type. This depth only represents a possible minimum depth-to-rock. In addition, the thicknesses of horizons were also derived for each soil type. The Munsell color can be converted into quantitative soil color systems such as RGB first (i.e., red, green and blue) [*Viscarra Rossel et al.*, 2006] and then the color averaging can be performed. However, we did not use this approach for several reasons. Our approach based on Munsell color has several advantages: (1) offering data in Munsell color are more direct while the conversion will introduce errors; (2) previous accepted methods for estimating albedo from Munsell color [*Post et al.*, 2000] can be utilized to derive the soil albedo needed by LSMs; (3) the main disadvantages of RGB are the high degree of correlation and the high influence of illumination intensity on each of the dimensions. If users need data in RGB or another color system, the conversion can be still done after the averaging based on Munsell color. Ultimately, which color system is best suited for an application depends on the purpose.

[27] The soil-type linkage method and the soil polygon linkage method were used to derive the spatial distribution of soil characteristics [*Shangguan et al.*, 2012]. The soil-type linkage method was accomplished by linking soil map units (soil types) and soil profiles according to taxonomy-based pedotransfer rules [*Batjes*, 2003]. The soil-type linkage method gave soil property estimates by soil type, textural class and depth zone. The topsoil (0–30 cm) texture class, as required by the linkage, was provided based on the specification of the HWSD and soil profile data of China [*FAO/IIASA/ISRIC/ISS-CAS/JRC*, 2009, 2012].

[28] The soil polygon linkage method works by linking a soil polygon with several closest soil profiles that have the same soil type and texture classes as the soil polygon. This method can account for spatial variation of the soil profiles corresponding to a specific soil type. The effects of climate, topography, land use, and parent material on soil properties are implicitly considered by this method [*Shangguan et al.*, 2012]. The linkage procedure, as described by *Shangguan et al.* [2012], incrementally enlarges the search radius for profiles until the target sample size is reached. The key difference between the polygon linkage method and the type linkage method is that the polygon linkage method can represent the spatial variation in soil properties across different map polygons of the same soil type, while the type linkage assigns an identical value to all soil polygons of the same soil type. The initial search radius is set as 15 km in order to represent the spatial variation

Table 2. Quality Control Information of the Derived Soil Properties

Digit[a]	Name	Code
d1	Linkage level	1: family; 2: subgroup; 3: great group; 4: order; 5: (non-)acid; 6: Andosols; 7: Histosols. Andosols and Histosols are separated for their rather specific behavior.
d2	Texture consideration	0: texture was considered in the linkage; 1: texture was not considered in the linkage.
d3	Sample size level 1	1: 3N or more; 2: N-(3N-1); 3: 0-(N-1); 4: no data. *N* is the target sample size and has different value due to the linkage level. At (non-)acid level, *N* = 400; at order level, *N*= 200; at great group level, *N* = 40; at subgroup level, *N* = 10; and at family level, *N* = 5.
d4	Search radius flag	0: search radius is in the initial radius (15 km); 1: search radius is larger than the initial radius but in the soil map extent (i.e., at least one soil profile in the same soil type of a map unit are not included in the linkage); 2: the linkage takes place in the whole map.
d5	Map unit level	1–7: The meanings of the numbers are the same as those in *d1*.
d6	Sample size level 2	1: 30 or more; 2: 15–29; 3: 5–14; 4: 1–4; 5: no data.

[a]The quality control information is composed of six digits. From left to right are *d1*–*d6*.

of the same soil type as much as possible. If this radius was set to a very large value, the ability to characterize the spatial variability would be diminished. For soil polygons of a specific soil type, the soil property derived by the polygon linkage method would be similar to that derived by the type linkage method if the total number of considered soil profiles is small. The results of the polygon linkage method become the same as the type linkage method in an extreme case when the search radius covers the entire domain.

[29] Because LSMs are usually grid based, we used rasterized soil maps with spatial grids at a resolution of 30 ×30 arc-seconds (about 1 km × 1 km at equator), which is the same as that of the HWSD.

2.3. Quality Control Information

[30] Quality control (QC) information was provided in numerical symbols. The symbol "11" indicates that the map unit is nonsoil; otherwise, numerical symbols have six digits. Table 2 shows the codes of the digits. The linkage level (*d1*) represents the soil classification level at which the linkage is performed. The texture consideration (*d2*) represents whether the soil texture is considered in the linkage. The sample size level (*d3* and *d6*) represents how many soil profiles are used to represent a soil map unit or soil polygon. We provide two kinds of sample size levels: *d6* is taken from *Batjes* [2002], and *d3* is set according to the linkage level (*d1*) because there is more variation of soil properties at higher soil type levels, which needs more samples to be representative. The search radius flag (*d4*) represents whether the search radius is in the initial radius (15 km). The map unit level (*d5*) represents the soil classification level of soil map unit. The digit *d5* is related to the detail level of the soil categorical map and the other digits are related to the linkage method.

[31] QC information serves as an indicator of "confidence" level in the derived soil parameters. The underlying assumptions are as follows. The confidence in the derived results should be higher when the linkage happened at a lower soil classification level (*d1*). If the linkage level (*d1*) does not reach the soil map unit level (*d5*), for example, a soil polygon at the soil family level of soil map unit was linked by profiles from corresponding soil subgroup, it has a potential to increase the confidence level by collecting samples at the soil map unit level (at the soil family level in the example). The confidence level should be higher when soil texture is considered in the linkage process (*d2*). The confidence should increase with the sample size of soil profiles (*d3* or *d6*). The spatial variation of soil type should be more reliable when the search radius is smaller (*d4*). The importance of the above factors is assumed to decrease in the following order: the linkage level (*d1*), soil texture consideration (*d2*), sample size (*d3* or *d6*), and search radius (*d4*). For each factor, the code is better when the corresponding numerical number is smaller, except for *d1* of volcanic ash soils and peat soils (corresponding to Andosols and Histosols in the World Reference Base for soil resources (WRB), respectively [*Shi et al*., 2010]).

3. Results

[32] This section presents a few results of soil chemical or fertility properties, i.e., soil organic carbon (SOC), nitrogen (N), phosphorus (P), potassium (K), cation exchange capacity (CEC) and the pH value in the topsoil (0–0.3 m), and comparisons with previous estimates. The full data set (as listed in Table 1) is given in the Supporting Information.

[33] The soil pH value is used as a nutrient solubility parameter, and CEC is as an indication of fertility and nutrient retention capacity in modeling. Soil C, N, and P are the key parameters and prognostic variables in biogeochemical process modeling with LSMs (currently, soil K is still not considered in LSMs). These soil nutrients can be calculated by running the models for thousands of years until an equilibrium state is reached (also called as model "spin-up") [*Kluzek*, 2010; *Parton et al*., 1988; *Wang et al*., 2009; *Xu and Prentice*, 2008]. However, nonlinear feedbacks in the biogeochemical cycles makes such a "spin-up" more time-consuming and less reliable for initiating soil nutrients. The data set can be an important benchmark for initial or calibration variables.

[34] The spatial distribution of soil properties is consistent with that given by Chinese soil scientists [*Shen*, 1998; *Xiong and Li*, 1987], which incorporates common knowledge of Chinese soil scientists from field surveys over many years.

3.1. Soil pH Value

[35] Soil pH value is one of the most important chemical properties as it controls many other soil physical,

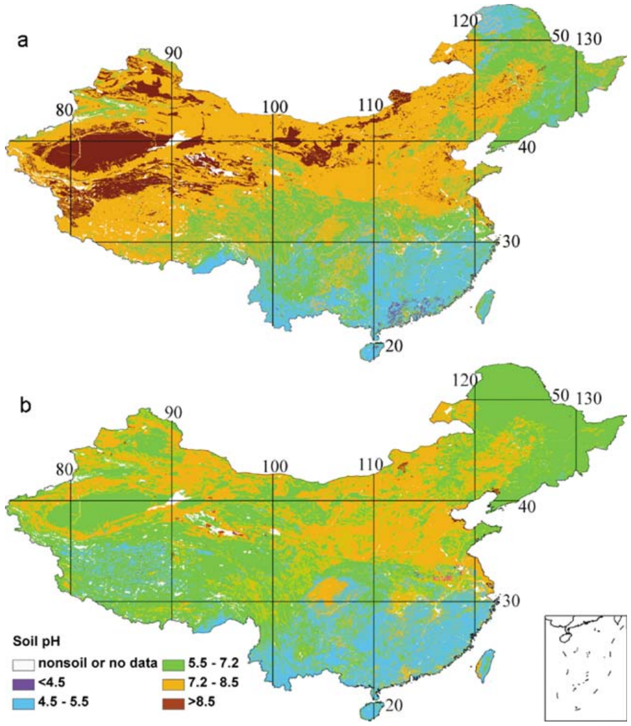

Figure 3. Spatial distribution of soil pH value (H_2O)

chemical, and biological properties. The pH value of a soil depends on the CO_2 concentration in the soil air and the salt concentrations in the soil solution, and these are constantly changing [*Bolan and Kandaswamy*, 2004]. Soil pH value is also significantly influenced by the overuse of nitrogen fertilizers [*Guo et al.*, 2010].

[36] The range of the pH values (H_2O) in the topsoil ranges from 4.2 to 9.8. Figure 3a shows that soils to the south of 30°N are acid to strongly acid, while soils in the north and northwest are typically basic or alkaline. In some southern mountainous and northeastern forested areas, soil appears to be acid (pH<7.2). In some northern areas, especially, in the deserts areas, soil appears alkaline (pH>7.2). The distribution is in accord with the common knowledge that alkali conditions occur in regions with low amounts of precipitation, acid conditions occur in regions with high amounts of precipitation. The HWSD showed acid conditions in the north of Qinghai-Tibet Plateau and desert areas (Figure 3b) that are contrary to our data and common knowledge. In China, soils are acid (pH< 7.2) in the HWSD data over 66% of the area, but only 42% of the soil areas are acid in our data set (Table 3). Our data show that 11% of the area has strong alkaline (pH> 8.5) soils, but the HWSD does not. Based on the common knowledge of Chinese soil scientists [*Shen*, 1998; *Xiong and Li*, 1987], our pH data are reasonable, but not that in the HWSD.

3.2. Soil Cation Exchange Capacity

[37] Cation exchange capacity is the sum of exchangeable cations that a soil can absorb. It is seen as a measure of fertility nutrient retention capacity and buffer capacity, and thus affects the growth of plants. A low CEC indicates that the soil can store a small amount of

Table 3. Comparison of the Area Percentage and Storage of Soil pH Value, CEC, C/N/P/K from of Our Data Set With the HWSD

		Percentage of Soil Area in Each Class in Topsoil (0–0.3 m)[a]					
Attribute	DATA SET	Class 1	Class 2	Class 3	Class 4	Class 5	Total Storage in Soil (0–1 m) (Pg)[b]
pH	Range	<4.5	4.5–5.5	5.5–7.2	7.2–8.5	>8.5	
	HWSD	0.07	17.28	49.12	33.18	0.35	
	This data	0.34	12.55	28.86	47.44	10.81	
CEC (me/100 g)	Range	<4	4–10	10–20	20–40	>40	
	HWSD	2.02	19.98	70.33	7.49	0.17	
	This data	6.13	37.37	45.13	10.75	0.62	
SOC (%)	Range	<0.2	0.2–0.6	0.6–1.2	1.2–2	>2	
	HWSD	0.00	25.48	34.78	31.21	8.53	67.06
	This data	6.05	24.48	26.70	23.62	19.15	72.50
TN (%)	Range	<0.05	0.05–0.1	0.1–0.2	0.2–0.4	>0.4	
	This data	16.55	30.81	35.83	14.97	1.84	6.61
TP (%)	Range	<0.02	0.02–0.04	0.04–0.06	0.06–0.08	>0.08	
	This data	0.14	17.97	29.48	33.91	18.50	4.45
TK (%)	Range	<1.2	1.2–1.6	1.6–2	2–2.4	>2.4	
	This data	3.63	9.35	52.67	28.71	5.64	169.90
AN (mg/kg)	Range	<20	20–50	50–80	80–120	>120	
	This data	6.24	20.37	20.83	27.10	25.46	0.553
AP (mg/kg)	Range	<2	2–4	4–6	6–8	>8	
	This data	3.11	48.57	31.13	11.61	5.57	0.030
AK (mg/kg)	Range	<50	50–100	100–150	150–200	>200	
	This data	3.66	34.79	39.44	16.29	5.82	0.611

[a] Null values were excluded in calculating the percentage of each class. The unit of range of each attribute is given in the first column. The selection of class limits partly followed the suggestions in HWSD. However, the classes are still somewhat arbitrary. CEC: cation exchange capacity. SOC: soil organic carbon. TN: total nitrogen. TP: total phosphorus. TK: total potassium. AN: Alkali-hydrolysable nitrogen. AP: available phosphorous. AK: available potassium.

[b] 1 Pg=10^{15} g.

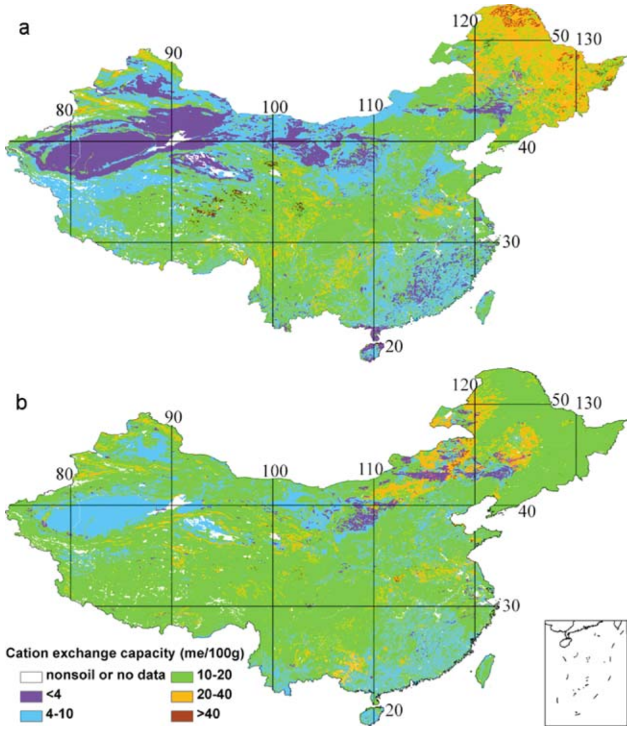

Figure 4. Spatial distribution of cation exchange

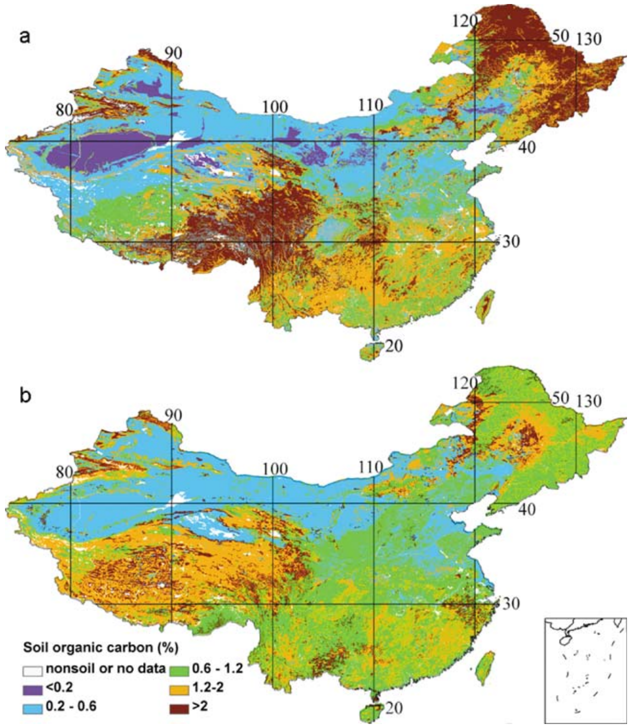

Figure 5. Spatial distribution of soil organic carbon

nutrients. In general, soils with higher amounts of clay and/or organic matter will typically have a higher CEC than more silty or sandy soils.

[38] The range of CEC in the topsoil (0–0.3 m) ranges from 1.5 to 50.5 me/100 g. Figure 4a shows that a high-CEC value in the topsoil is found in peat and forested areas in Qinghai-Tibet Plateau, central and northeast China, i.e., high-biomass [*Piao et al.*, 2005] or low-leaching areas, a low-CEC value is found in Ferrasols in the south, Semi-Aqueous soils in the north, and the north arid and semiarid area, and very low CEC is found in the deserts. CEC is highly dependent upon soil clay and SOC (Figure S19 and Figure 5a). The CEC of our data are predominantly distributed (83%) in class 2 and 3 (4–10, 10–20 me /100 g), whereas in the HWSD, 70% of the area is in classes 3 (10–20 me/100 g) (Table 3; Figures 4a and 4b).

3.3. Soil Organic Carbon (SOC)

[39] SOC (or soil organic matter) is important for the function of ecosystems and has a major influence on the soil thermal and hydraulic properties, thereby affecting the ground thermal and moisture regimes [*Lawrence and Slater*, 2008]. In addition, loss of SOC leads to a reduction in soil fertility, land degradation and even desertification, and an increase in CO_2 emissions into the atmosphere.

[40] The range of SOC in the topsoil ranges from 0.09% to 17.05%. As is shown in Figure 5a, the highest SOC in the topsoil appeared in the peat and forested areas in southeastern Tibet mountains and forested areas in northeast China, where there is less disturbed by human activities, and lower values are in the north and northwest, especially in the deserts. The SOC is quite evenly distributed from class 2 to class 5 (Table 3). The spatial pattern agrees well with that of biomass C density in China [*Piao et al.*, 2005]. In contrast, the HWSD showed more area represented by class 3(0.6%–1.2%) (Figure 5b, Table 3).

[41] The total storage of SOC in the upper 1 m of soil is 72.5 Pg (Table 3), which is in the range of 50–185 Pg given by previous studies [*Wang et al.*, 2001; *Wu et al.*, 2003; *Yu et al.*, 2005; *Zhou et al.*, 2003], which are largely different due to the different data and methods used to upscale soil C observations [*Zhao et al.*, 2006]. A widely accepted estimate is 90 Pg [*Zheng et al.*, 2011]. The estimate from the HWSD (67.06 Pg) is close to our studies.

3.4. Soil Nitrogen (N), Phosphorus (P), and Potassium (K)

[42] Soil N, P and K are three major nutrients for plant growth and health. The total N, P, and K (or available N, P and K for uptake by plants) are related to amounts of soil minerals and soil organic matter, atmospheric wet/dry deposition, soil leaching, soil biotic processes and other factors [*Coyne and Frye*, 2004; *Huang et al.*, 2004; *Post et al.*, 1985; *Sims and Vadas*, 2004]. Human activities also have important influences. In the short term, recycling of nutrients from soil organic matter is the major direct source of soluble nutrients to the soils [*Lambers et al.*, 2006].

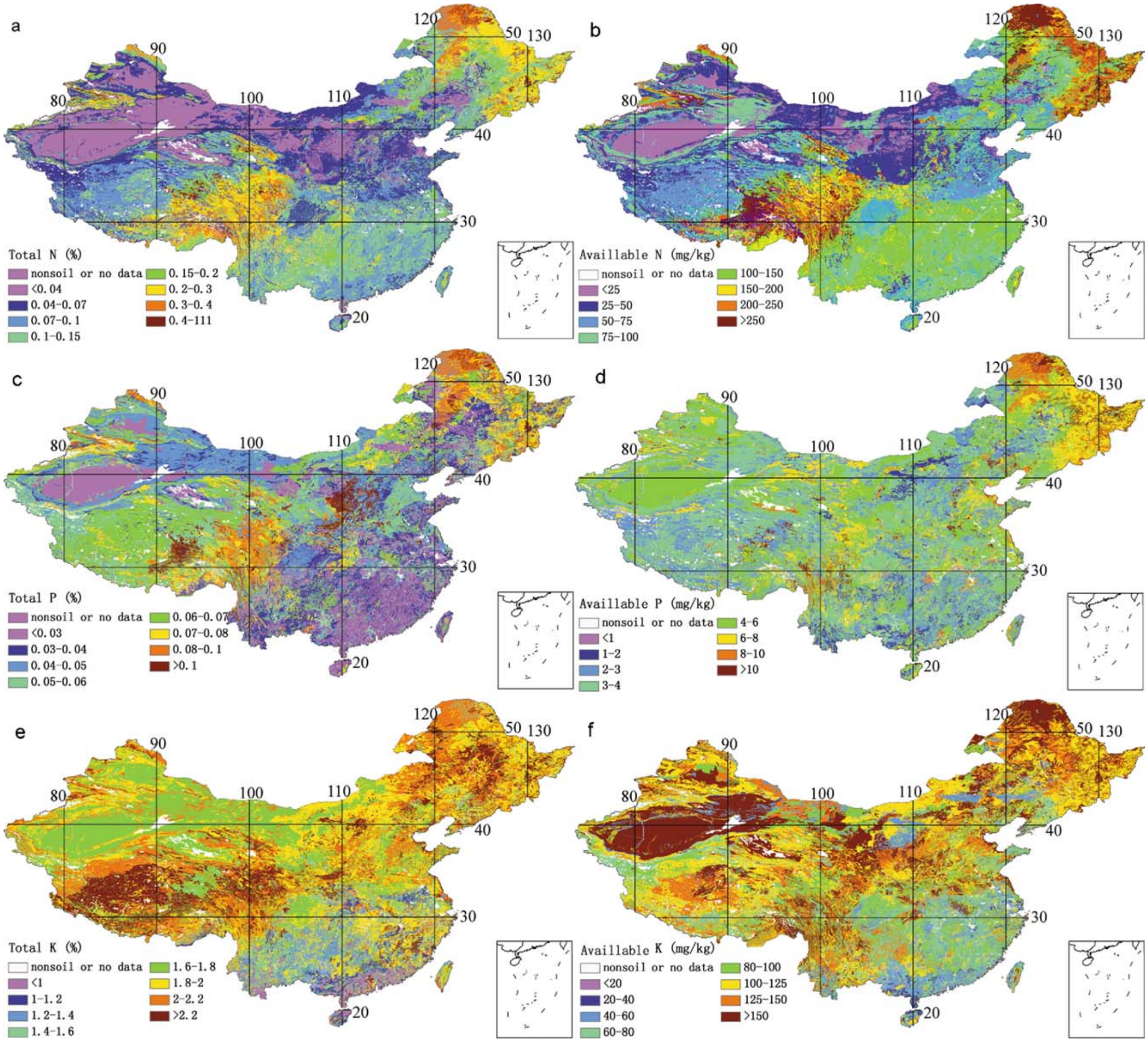

Figure 6. Spatial distribution of (a) soil total nitrogen (%), (b) Alkali-hydrolysable nitrogen (mg/kg), (c) total

3.4.1. Soil *Nitrogen* (N)

[43] The soil total N in the topsoil (0–0.3 m) ranges from 0.01% to 1.04%, alkali-hydrolysable N ranges from 9.8 to 648 mg/kg. As is shown in Figures 6a and 6b, distributions of total N and alkali-hydrolysable N are similar to the distribution of SOC. The total N is mainly distributed in classes 1, 2, and 3, while the alkali-hydrolysable N is quite evenly distributed from classes 2 to 5 (Table 3).

[44] The storage of total N in the upper 1 m of soil is 6.61 Pg, which is in the range of 4.5–52.5 Pg given by previous studies [*Tian et al.*, 2006], and is close to the estimates of 8.29 Pg by *Tian et al*. [2006] and 7.4 Pg by *Yang et al*. [2007]. The storage of alkali-hydrolysable N is 0.553 Pg, which is about 8.3% of total N.

3.4.2. Soil *Phosphorus*

[45] The total P in the topsoil is in the range of 0.01%–0.25%, and the available P is in the range of 0.9–16.5 mg/kg. As shown in Figure 6c, a high-total P value appears in the Qinghai-Tibet plateau, while a low total P appears in the south, the north and the desert areas. The total P decreases with increases in temperature and precipitation [*Wang et al.*, 2008]. However, Aeolian soils in the desert areas have a low total P. The definition of available P is not strict yet. It cannot be taken as an absolute value, but a relative index under specific conditions [*Xiong and Li*, 1987]. Figure 6d shows that available P is higher in the northeast than the other regions. There is no evident correlation between the total P and available P. The total P is mainly distributed

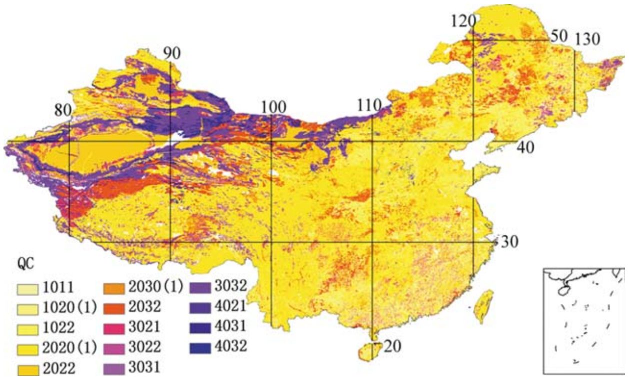

Figure 7. Quality control (QC) information of soil pH value (H_2O) in topsoil (0-0.3 m) by soil polygon linkage method. The meaning of the QC index is given in Table 2; only the first four digits of the QC index are given in the figure; and values 0 and 1 of the last digit are combined for brevity.

in classes 2–5, while the available P is mainly distributed in classes 2 and 3 (Table 3).

[46] The storage of total P in the upper 1 m of soil is 4.45 Pg, which is in the range of 3.5 to 5.3 Pg of previous studies [*Wang et al.*, 2008; *Zhang et al.*, 2005]. The available P is 0.03 Pg, which is about 0.7% of the total P.

3.4.3. Soil *Potassium*

[47] The range of total K in the topsoil is 0.19%–3.72%, and available K is 21.9–703.8 mg/kg. As shown in Figures 6e and 6f, the total and available K both decrease from the North to the South, although their distributions are rather different. Low-total K values appear in tropical areas, while high-total K values appear in the Qinghai-Tibet Plateau and the northeast of China. High available K values scatter across the north of China. The total K is mainly in classes 3 and 4, while available K is more scattered (Table 3). The storage of total K in the upper 1 m of soil is 169.9 Pg, and the available K is 0.611 Pg (Table 3), which is about 0.4% of the total K.

3.5. Data Quality

[48] The quality of the data varies across China in the data set. As an example, Figure 7 shows the data QC nformation of the soil pH value (H_2O) in the topsoil (0–0.3 m) derived by the polygon linkage. The numerical symbols of QC (i.e., *d1*, *d2*, *d3*, and *d4*) form an index, so that smaller number indicates higher data quality. Poor data quality appears in the northwest. Across China, about three fifths of soil polygons are linked at the subgroup level. About one fifth of them are linked at both great group level and family level. And only 2% of them are linked at the soil order level. All soil pH values are derived taking soil texture class into account. About four fifths of soil polygons are moderately represented by abundant soil samples exceeding the target sample size at a specific soil type level. About one fifth of them are poorly represented by small numbers of soil samples. And only a few of them are represented by a very large number of soil samples. About three fifths of the soil polygons are linked with a search radius smaller than the map extent, which indicates that the spatial variation of soil properties among the soil polygons of the same map unit is represented to some extent by the data set; about two fifths of them are linked by all soil samples at a specific soil type level, giving the same result as the soil type linkage method; and only scores of them are linked within the initial search radius, indicating that the local spatial variation of soil properties are mostly not represented due to the limited sampling density.

[49] For other soil properties and soil depths, the spatial distribution of QC information is similar, but the data quality varies. Soil properties with more records have a higher data quality. The data quality decreases with soil depth because there are fewer observations in the deep soil.

4. Discussion

[50] The soil data set developed in this study can be used as input parameters or initial variables, and calibration or evaluation data for LSMs. It is the first time that the spatial distribution of various soil characteristics in China at 30 arc-second resolution has become available to land surface modelers and biogeochemical researchers.

[51] We use the GSCC instead of the FAO symbols in the HWSD to link the soil map and soil attributes. Though different soil classifications can be correlated, their referencing ability is usually far below 100% [*Shi et al.*, 2006a, 2006b, 2010], so use of any correlated classification in the linkage method would lead to some errors. For this reason, it is better to use the local classification when soil profiles and a soil map using the same classification are available. It should be noted that China has developed a more advanced quantitative classification system, the Chinese Soil Taxonomy (CST), to replace the qualitative GSCC, although it is still difficult to apply this system nationally, mainly because of insufficient data [*Shi et al.*, 2006a].

[52] The accuracy of the data is, of course, limited by the errors and uncertainties of the raw data and those introduced during data processing. QC information is given to represent confidence levels in the derived attributes. As the number of soil profiles of different soil types is uneven, some soil types or map units have a lower linkage soil type level and more linked profiles and thus have a higher confidence level than the others. The confidence level also varies among soil attributes. Attributes with more records have a higher confidence. The confidence of attributes can be divided into three groups: attributes with a high confidence are particle size distribution, pH value, SOC, total N, total P, total K, CEC, structure, consistency and root abundance. Attributes with a medium confidence are alkali-hydrolysable N, available P, available K, soil color. And attributes with a low confidence are rock fragment, bulk density, porosity, and exchangeable cations. As

there are more records for the near-surface layers, the corresponding derived results are more reliable.

[53] There are many sources of uncertainty in deriving the attributes, including soil map, soil attribute measurement, distance between soil profile and map polygon, the classification system (GSCC) and the linkage method [*Shangguan et al*., 2012]. These uncertainties are hard or even impossible to quantify. The uncertainty of the spatial data (i.e., soil categorical map) is thought to be the largest. A third soil survey of China at the national scale to gain more reliable data seems to be far away from now because of its high cost. It should be also noted that some attributes are measured by different analytical methods (S-II of the Supporting Information) and the method used for a specific sample is usually not recorded. This reduces the comparability of soil analytical data, which brings inherent inconsistency in the attribute data set [*Batjes*, 2003]. For example, most of the soil pH values are measured using H_2O as the standard solution in our data set, but some records have no indication of the standard solution. In addition, soil pH values are measured using different soil/water ratios of suspensions, which deviate from each other [*Batjes*, 1995].

[54] This data set has been produced mainly for use in LSMs, and users should be careful when using the data for other applications, especially at local or detailed applications. The soil data set is appropriate for understanding how soil properties vary over large areas, but is not appropriate for small areas. The linkage method provides average estimates of attributes for a map unit of a map polygon. Therefore, the estimated values may be quite different from site-specific measurements or the actual properties at a given grid cell location.

[55] The resolution of the raster map (30 × 30 arc-seconds) is quite high for land surface modeling. It was determined according to the original soil map scale to preserve the most detailed information [*Hengl*, 2006]. As a result, there are many grid cells with identical attributes belonging to the same soil polygon on the original vector map, because the spatial variation within the soil polygon is not included in the data set. The average polygon area of the original soil map is about 101 km^2, and the lowest quartile, the median and the high quartile are 8, 20 and 55 km^2, respectively, which are better indexes to show the level of detail in the derived raster map [*Rossiter*, 2003].

[56] It is desirable to make the best use of all legacy data in China. There are more than 200,000 profiles collected in the Second National Soil Survey of China, in the study of Chinese Soil Taxonomy in 1990–1996 [*Gong et al*., 1999] and for other research-purposes, but they are scattered in various publications (books, papers, and unpublished reports). This is a valuable source for soil information, but most such information is in hard-copy, and requires digitization, geo-referencing, QC, and standardization. A China Digital Soil Map data set at 1:50,000 scale has been developed since 1999, and soil categorical and nutrient paper maps at 1:50,000 scale and soil profile records with various properties were collected from 2,300 counties of China [*Zhang et al*., 2010]. On the other hand, there has not been an effective platform to collect the new observations of soil in recent years, though some of them have been published in the literature.

5. Conclusion

[57] We used soil profiles and the 1:1 million soil map of China to develop a soil property data set for regional LSMs. This data set is intended to be as complete as possible and replace the outdated soil information that has been used widely by the land surface modeling community. It includes most of physical and chemical properties of soil that are required by LSMs in NWPMs and earth system models. It has much higher spatial precision than the HWSD, and more reasonable spatial patterns and magnitudes. These improvements are due to the use of more soil profile data, a higher resolution map, strict QC, and a reasonable methodology of linkage between profile data and map. QC information was provided to represent confidence levels in the derived attributes, though uncertainties inherent in the raw data cannot be fully quantified. A thorough evaluation of the overall data quality of all attributes is beyond the scope of this study. The data set remains more qualitative than quantitative. In many cases, it may only tell us an approximate magnitude and spatial distribution.

[58] Efforts will be needed to improve the data set by using more detailed soil information in China [*Zhang et al*., 2010] and related environmental factors such as climate and topography through digital soil mapping and modeling techniques [*Grunwald et al*., 2011]. These efforts are needed to produce soil property maps with a finer resolution of 3 × 3 arc-seconds required by the GlobalSoilMap.net specification. We expect that the use of the new data set instead of the old ones will result in a better performance of LSMs, which should be demonstrated through case studies. Future work will also focus on the development of a global soil data set integrating regional soil databases, such as State Soil Geographic Database (STATSGO) of the USA, the National Soil Database of Canada, the Australian Soil Resource Information System and others.

[59] The data are available at http://globalchange.bnu.edu.cn, and will be the only China soil data set for land modeling purpose, that can be freely downloaded.

[60] **Acknowledgments.** This work was supported by the Natural Science Foundation of China (under grants 41205037, 40875062, and 40225013), the R&D Special Fund for Nonprofit Industry (Meteorology, GYHY201206013, and GYHY200706025), and the Key International S and T Cooperation Project (2008DFA22180). We are grateful to Robert E. Dickinson and Guoyue Niu for their helpful discussions and English revision. We would like to extend our thanks to Dominique Arrouays for his helpful suggestion and insights. We also would like to thank the reviewers for their time and effort to thoroughly review the manuscript. Their suggestions have greatly improved the paper. The 1:1 million geo-referenced digital maps were purchased from the Institute of Soil Science of Chinese Academy of Sciences and the Department of Soil Environment of the Ministry of Agriculture of China.

References

Batjes, N. H. (1995), *A global data set of soil pH properties, Tech. Pap. 27*, Int. Soil Ref. and Inf. Cent. (ISRIC), Wageningen, Netherlands.

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. Inf. Cent. (ISRIC), Wageningen, Netherlands.

Batjes, N. H. (2003), *A Taxotransfer Rule-Based Approach for Filling Gaps in Measured Soil Data in Primary SOTER Databases*, Int. Soil Ref. Inf. Cent., 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., G. Fischer, F. O. Nachtergaele, V. S. Stolbovol, and H. T. van Velthuizen (1997), *Soil data derived from WISE for use in global and regional AEZ studies (ver. 1.0), Interim Rep. IR-97-025*, FAO/IIASA/ISRIC, Laxenburg.

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.

Bolan, N., and K. Kandaswamy (2004), *Encyclopedia of Soils in the Environment*, edited by D. Hillel, pp. 196–202, Academic Press, New York.

Coyne, M. S., and W. W. Frye (2004), Nitrogen in soils, in *Encyclopedia of Soils in the Environment*, edited by D. Hillel, pp. 13–21, Academic Press, New York.

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

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 (version 1.2)*, FAO, Rome, Italy.

Global Soil Data Task (2000), Global Soil Data Products CD-ROM (IGBP-DIS), International Geosphere-Biosphere Programme—Data and Information Services, ORNL Distributed Active Arch. Cent., Oak Ridge Nat. Lab., Oak Ridge, Tennessee. http://daac.ornl.gov/SOILS/guides/igbp.html. Last accessed March 39, 2013.

Gong, Z., et al. (1999), *China Sol Taxonomy: Theory, Methodology and Practice*, Science Press, Beijing.

Grunwald, S., J. A. Thompson, and J. L. Boettinger (2011), Digital soil mapping and modeling at continental scales: Finding solutions for global issues, *Soil Sci. Soc. Am. J.*, *75*(4), 1201–1213.

Guo, J. H., et al. (2010), Significant acidification in major Chinese croplands, *Science*, *327*, 1008–1010.

Hengl, T. (2006), Finding the right pixel size, *Comput. Geosci.*, *32*(9), 1283–1298.

Huang, P. M., J. M. Zhou, J. C. Xie, and M. K. Wang (2004), Potassium in soils, in *Encyclopedia of Soils in the Environment*, edited by D. Hillel, pp. 303–314, Academic Press, New York.

Katschinski, N. A. (1956), Die mechanische bodenanalyse und die klassifikation der böden nach ihrer mechanischen zusammensetzung, *Pari, B*, 321–327.

Kluzek, E. (2010), Spinning up the carbon-nitrogen dynamic global vegetation model (CNDV spinup), in *CCSM Research Tools: CLM4.0 User's Guide Documentation*, edited by E. Kluzek, pp. 40–41, Nat. Cent. for Atmos. Res. (NCAR), Boulder, Colo.

Kuehni, R. G. (2002), The early development of the Munsell system, *Color Res. Appl.*, *27*(1), 20–27.

Lambers, H., F. S. Chaplin, III, and T. L. Pons (2006), *Plant Physiological Ecology*, Springer.

Lawrence, D. M., and A. G. Slater (2008), Incorporating organic soil into a global climate model, *Clim. Dyn.*, *30*, 145–160.

Malone, B. P., A. B. McBratney, B. Minasny, and G. M. Laslett (2009), Mapping continuous depth functions of soil carbon storage and available water capacity, *Geoderma*, *154*, 138–152.

National Soil Survey Office (1995), *Soil Map of China (in Chinese)*, China Map Press, Beijing.

National Soil Survey Office (1996), *China Soil Species Description (in Chinese)*, vol. *1-6*, China Agric. Press, Beijing.

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., Nat. Cent. for Atmos. Res., Boulder, Colo.

Parton, W. J., J. W. B. Stewart, and C. V. Cole (1988), Dynamics of C, N, P and S in grassland soils: A model, *Biogeochemistry, 5*, 109–131.

Piao, S., J. Fang, B. Zhu, and K. Tan (2005), Forest biomass carbon stocks in China over the past 2 decades: Estimation based on integrated inventory and satellite data, *J. Geophys. Res.*, *110*, G01006, doi:10.1029/2005JG000014.

Post, D. F., A. Fimbres, A. D. Matthias, E. E. Sano, L. Accioly, A. K. Batchily, and L. G. Ferreira (2000), Predicting soil albedo from soil color and spectral reflectance data, *Soil Sci. Soc. Am. J.*, *64*(3), 1027–1034.

Post, W. M., J. Pastor, P. J. Zinke, and A. G. Stangenberger (1985), Global patterns of soil nitrogen storage, *Nature*, *317*(6038), 613–616.

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.

Rossiter, D. G. (2003), *Methodology for Soil Resource Inventories*, 3rd ed., ITC Lecture Notes SOL.27., ITC, Enschede, Netherlands.

Shangguan, W., and Y. Dai (2009), Investigation of different models to describe soil particle-size distribution for sparse experimental data, *J. Beijing Normal Univ. (Nat. Sci.)*, *45*, 279–283.

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.

Shen, S. M. (1998), *China Soils Fertility (in Chinese)*, 484 pp., China Agric. Press, Beijing.

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., D. Yu, G. Yang, H. Wang, W. Sun, G. Du, and Z. Gong (2006a), Cross-reference benchmarks for translating the genetic soil classification of china into the chinese soil taxonomy, *Pedosphere*, *16*(2), 147–153.

Shi, X. Z., D. S. Yu, E. D. Warner, X. Z. Pan, X. Sun, G. W. Petersen, Z. G. Gong, and H. Lin (2006b), Cross reference system for translating between genetic soil classification of China and soil taxonomy, *Soil Sci. Soc. Am. J.*, *70*, 78–83.

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.

Sims, J. T., and P. A. Vadas (2004), Phosphorus in soils, in *Encyclopedia of Soils in the Environment*, edited by D. Hillel, pp. 202–210, Academic Press, New York.

Tian, H., S. Wang, J. Liu, S. Pan, H. Chen, C. Zhang, and X. Shi (2006), Patterns of soil nitrogen storage in China, *Global Biogeochem. Cycles*, *20*(1), GB1001, doi:10.1029/2005GB002464.

van Engelen, V. W. P., N. H. Batjes, J. A. Dijkshoorn, and J. R. M. Huting (2005), *Harmonized Global Soil Resources Database*, FAO and ISRIC, Wageningen, Netherlands.

Viscarra Rossel, R. A., B. Minasny, P. Roudier, and A. B. McBratney (2006), Colour space models for soil science, *Geoderma*, *133*(3-4), 320–337.

Wang, S. Q., C. H. Zhou, K. R. Li, S. L. Zhou, and F. H. Huang (2001), Estimation of soil organic carbon reservoir in China, *J. Geogr. Sci.*, *11*, 3–13.

Wang, T., Y. Yang, and W. Ma (2008), Storage, patterns and environmental controls of soil phosphorus in China, *Acta Sci. Naturalium Univ. Pekinensis*, *44*, 945–952.

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

Webb, R. S., C. E. Rosenzweig, and E. R. Levine (1991), *A Global Data Set of Soil Particle Size Properties*, NASA, New York.

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.

Wu, H., Z. Guo, and C. Peng (2003), Distribution and storage of soil organic carbon in China, *Global Biogeochem. Cycles*, *17*(2), 1048–1058.

Xiong, Y., and Q. Li (1987), *Chinese Soils (in Chinese)*, Science Press, Beijing.

Xu, R., and I. C. Prentice (2008), Terrestrial nitrogen cycle simulation with a dynamic global vegetation model, *Global Change Biol.*, *14*, 1745–1764.

Yang, Y. H., W. H. Ma, A. Mohammat, and J. Y. Fang (2007), Storage, patterns and controls of soil nitrogen in China, *Pedosphere*, *17*, 776–785.

Yu, D., X. Shi, W. Sun, H. Wang, Q. Liu, and Y. Zhao (2005), Estimation of China soil organic carbon storage and density based on 1:1 000 000 soil database, *Yingyong Shengtai Xuebao*, *16*(12), 2279–2283.

Zhang, C., H. Tian, J. Liu, S. Wang, M. Liu, S. Pan, and X. Shi (2005), Pools and distributions of soil phosphorus in China, *Global Biogeochem. Cycles*, *19*(1), GB1020, doi:10.1029/2004GB002296.

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 Aug.

Zhao, Y. C., X. Z. Shi, D. C. Weindorf, D. S. Yu, W. X. Sun, and H. J. Wang (2006), Map scale effects on soil organic carbon stock estimation in North China, *Soil Sci. Soc. Am. J.*, *70*(4), 1377–1386.

Zheng, J. F., K. Cheng, G. X. Pan, P. Smith, L. Lianaing, Z. Xuhui, Z. Jinwei, H. Xiaojun, and D. Yanling (2011), Perspectives on studies on soil carbon stocks and the carbon sequestration potential of China, *Chinese Sci. Bull.*, *56*(3748-3758).

Zhou, C., Q. Zhou, and S. Wang (2003), Estimating and Analyzing the Spatial Distribution of Soil Organic Carbon in China, *Ambio J. Hum. Environ.*, *32*, 6–12.

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

Corresponding author: Y. Dai, College of Global Change and Earth System Science, Beijing Normal University, No. 19, Xinjiekouwai St., Beijing 100875, China. (yongjiudai@bnu.edu.cn)