A China Dataset of Soil Properties for Land Surface Modeling

Article *in* Journal of Advances in Modeling Earth Systems · June 2013

DOI: 10.1002/jame.20026

CITATIONS
105

READS
650

23 authors, including:

Wei Shangguan
Sun Yat-Sen University
23 PUBLICATIONS **923** CITATIONS
SEE PROFILE

Yongjiu Dai
Sun Yat-Sen University
102 PUBLICATIONS **4,410** CITATIONS
SEE PROFILE

Baoyuan Liu
Beijing Normal University
67 PUBLICATIONS **2,272** CITATIONS
SEE PROFILE

A-Xing Zhu
University of Wisconsin–Madison
276 PUBLICATIONS **4,369** CITATIONS
SEE PROFILE

A China Dataset 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[1], 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], Yizhou Zhang[2]

[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

Submitted to [Journal of Advances in Modeling Earth Systems] – AGU

(revised)

4 February 2013

*Corresponding author:
Yongjiu Dai
College of Global Change and Earth System Science
Beijing Normal University
No. 19, Xinjiekouwai St.
Beijing 100875
China
E-mail: yongjiudai@bnu.edu.cn
Tel: +86-10-5880-5436
Fax: +86-10-5880-0156

Abstract

A comprehensive 30×30 arc-second resolution gridded soil characteristics dataset of China has been developed for use in the land surface modeling. It includes physical and chemical attributes of soils derived from 8,979 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 dataset includes 28 attributes for 8 vertical layers at the spatial resolution of 30×30 arc-seconds. Based on this dataset, 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 dataset can be incorporated into land models to better represent the role of soils in hydrological and biogeochemical cycles in China.

1. Introduction

As land surface models (LSMs) for use in numerical weather prediction models (NWPs) 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 datasets 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; *Batjes*, 2006; *Global Soil Data Task*, 2000; *Reynolds et al.*, 2000; *van Engelen et al.*, 2005; *Webb et al.*, 1991; *Webb et al.*, 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/NWPs that depend on these soil data are questionable.

The soil map data and soil attribute data of the sampled profiles are two source databases for developing the spatial soil property dataset. 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 effort has been devoted to developing procedures for referencing between these schemes and other internationally more widely used ones [*Shi et al.*, 2004; *Shi et al.*, 2006a; *Shi et al.*, 2006b; *Shi et al.*, 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.

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.

The Harmonized World Soil Database (HWSD) is a newly released global soil dataset [*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.

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.

This paper describes the development of the China soil characteristics dataset for use in land models in NWPs/ESMs. This effort is unique in that, for the first time, the China NWPs/ESMs community will have access to a dataset 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

The soils in 2,444 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 6 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 dataset for land surface modeling.

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 non soil 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.

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 Auxiliary materials). 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 8,979 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

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

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

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.

3) About three fourths of the soil profiles have a depth less than 1 meter, and 90 % of them have a depth less than 1.5 meters. 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.

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.

5) Many land surface models 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.

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.

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

The soil profile dataset 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.

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.

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:

$$\frac{A_i - A_{i-1}}{A_{i+1} - A_i} = \frac{d_{i-1}}{d_i} \quad (1)$$

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

$$d_i = \frac{b_{i+1} - b_{i-1}}{2} \quad (2)$$

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

The soil characteristics of soil profiles are divided into 8 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 was excluded. For brevity and comparison with other datasets, we also use a two-layer scheme (i.e. 0-0.3 m and 0.3- 1 m) to display the results.

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.

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 dataset. 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 5 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:

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

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.

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 is 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.

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].

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 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.

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

2.3. Quality control information

Quality control information (QC) was provided in numerical symbols. The symbol '11' indicates that the map unit is non-soil; otherwise, numerical symbols have 6 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 (15km). 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.

The quality control 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

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 dataset (as listed in Table 1) is given in the auxiliary materials.

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, non-linear feedbacks in the biogeochemical cycles makes such a 'spin-up' more time-consuming and less reliable for initiating soil nutrients. The dataset can be an important benchmark for initial or calibration variables.

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

Soil pH value is one of the most important chemical properties as it controls

many other soil physical, 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].

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^{o}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 dataset (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

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 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.

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 semi-arid 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 (10-20me/100 g) (Table 3; Figure 4a, b).

3.3 Soil organic carbon (SOC)

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.

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).

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)

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].

3.4.1 Soil nitrogen (N)

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 Figure 6(a, b), 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 class 2 to class 5 (Table 3).

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

The total P in the topsoil is in the range of 0.01 to 0.25 %, and the available P is in the range of 0.9 to 16.5 mg/kg. As shown in Figure 6(c), 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 6(d) 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 in classes 2 to 5, while the available P is mainly distributed in classes 2 and 3 (Table 3).

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

The range of total K in the topsoil is 0.19 to 3.72 %, and available K is 21.9 to 703.8 mg/kg. As shown in Figure 6 (e, f), 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

The quality of the data varies across China in the dataset. As an example, Figure 7 shows the data quality control (QC) information 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 dataset; 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.

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

The soil dataset 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.

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; *Shi et al.*, 2006b; *Shi et al.*, 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].

The accuracy of the data is, of course, limited by the errors and uncertainties of the raw data and those introduced during data processing. Quality control 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 3 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.

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 auxiliary materials) 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 dataset [*Batjes*, 2003]. For example, most of the soil pH values are measured using H_2O as the standard solution in our dataset, 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].

This dataset 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 dataset 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.

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 dataset. 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].

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 un-published reports). This is a valuable source for soil information, but most such information is in hard-copy, and requires digitization, geo-referencing, quality control and standardization. A China Digital Soil Map dataset 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

We used soil profiles and the 1:1 million soil map of China to develop a soil property dataset for regional land surface models. This dataset 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 land surface models in numerical

weather prediction models 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 quality control, and a reasonable methodology of linkage between profile data and map. Quality control 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 dataset remains more qualitative than quantitative. In many cases, it may only tell us an approximate magnitude and spatial distribution.

Efforts will be needed to improve the dataset 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 dataset 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 dataset 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.

The data is available at http://globalchange.bnu.edu.cn, and will be the only China Soil Dataset for land modeling purpose, that can be freely downloaded.

Acknowledgements: 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. Technical Paper 27, International Soil Reference and Information Centre (ISRIC), Wageningen.

Batjes, N. H., F. G., N. F.O., S. V.S., and v. V. H.T. (1997), Soil data derived from WISE for use in global and regional AEZ studies (ver. 1.0), Interim Report IR-97-025, FAO/ IIASA/ ISRIC, Laxenburg

Batjes, N. H. (2002), Soil parameter estimates for the soil types of the world for use in global and regional modelling (Version 2.1), ISRIC Report 2002/02c, International Food Policy Research Institute (IFPRI) and International Soil Reference and Information Centre (ISRIC), Wageningen.

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

Batjes, N. H. (2006), 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).

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, Elsevier 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, Elsevier Academic Press, New York.

Dai, Y., Z. X., D. R. E., 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. Yang (2003), The Common Land Model, *American Meteorological Society*, *8*, 1013~1023.

FAO/IIASA/ISRIC/ISS-CAS/JRC (2009), Harmonized World Soil Database (version1.1), FAO, Rome, Italy and IIASA, Laxenburg, Austria.

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

Global Soil Data Task (2000), 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.

Gong, Z., Z. Chen, X. Shi, G. Zhang, J. Zhang, W. Zhao, G. Luo, Y. Gao, S. Cao, Z. Cao, and W. Lei (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 Science Society of America Journal*, *75*(4), 1201-1213.

Guo, J. H., X. J. Liu, Y. Zhang, J. L. Shen, W. X. Han, W. F. Zhang, P. Christie, K. W. T. Goulding, P. M. Vitousek, and F. S. Zhang (2010), Significant acidification in

major Chinese croplands, *Science*, *327*, 1008-1010.

Hengl, T. (2006), Finding the right pixel size, *Computers & Geosciences*, *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, Elsevier Academic Press, New York.

Katschinski, N. A. (1956), Die mechanische bodenanalyse und die klassifikation der böden nach ihrer mechanischen zusammensetzung, Paris, 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. National. 570*, edited, Center for Atmospheric Research (NCAR), Boulder, Colorado.

Kuehni, R. G. (2002), The early development of the Munsell system, *Color Research and Application*, *27*(1), 20-27.

Lambers, H., F. S. C. 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, *Climate Dynamics*, *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 Agriculture Press, Beijing, China.

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., Y. Dai, and Co-authors (2004), Technical Description of the Community Land Model (CLM). NCAR Tech. Note TN-461+STR, Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder.

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, *Journal of Geophysical Research*, *110*, G01006.

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 Science Society of America Journa*l 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, third ed*, ITC Lecture Notes SOL.27. ITC, Enschede, The Netherlands.

Shangguan, W., and Y. Dai (2009), Investigation of different models to describe soil particle-size distribution for sparse experimental data, *Journal of Beijing Normal University (Natural Science)*, *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 Agricultural 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 Survey 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 Science Society of America Journal*, *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, Elsevier 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.

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.

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, *Journal of Geography Science*, *11*, 3-13.

Wang, T., Y. Yang, and W. Ma (2008), Storage, patterns and environmental controls of soil phosphorus in China, *Acta Scientiarum Naturalium Universitatis 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, *Biogeosciences Discussions*, *6*, 9891-9944.

Webb, R. S., C. E. Rosenzweig, and E. R. Levine (1991), A global data set of soil particle size properties, NASA.

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 Biogeo. Cyc.*, *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, *Journal of Climatology*, *5*, 119-143.

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

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

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

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

Changing World, Brisbane, Australia, 1-6 August, 2010.

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 Science Society of America Journal*, *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 Science Bulletin*, *56*(3748-3758).

Zhou, C., Q. Zhou, and S. Wang (2003), Estimating and Analyzing the Spatial Distribution of Soil Organic Carbon in China, *Ambio: A Journal of the Human Environment*, *32*, 6-12.

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

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 [^{o}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.

Measured Physical and Chemical Attributes in Horizon Layers:

No.	Attribute	Unit	Number of records [a]	Maps [b]
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/cm^3	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 33010 records in total.

[b] Maps of soil properties interpolated in land model standard layers [Figure S1- S26 in Auxiliary materials].

[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.

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 6 digits. From left to right are *d1- d6*.

Table 3. Comparison of the area percentage and storage of soil pH value, CEC, C/N/P/K from of our dataset with the HWSD

Attribute	Dataset	Percentage of soil area in each class in topsoil (0- 0.3 m) [a]					Total storage in soil (0- 1 m) (Pg) [b]
		Class 1	Class 2	Class 3	Class 4	Class 5	
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 captions:

Figure 1. Geographic distribution of soil profiles. The number of soil profiles from different provinces varies from about 160 to 400, except Tibet has 2583 profiles. Soil profiles are denser in the east and south of Tibet, and in east China, and are rarer in the northwest.

Figure 2. Three data gap filling treatments for the soil profiles. *A* is the soil property of a layer, and *a* is the observed value of the soil property.

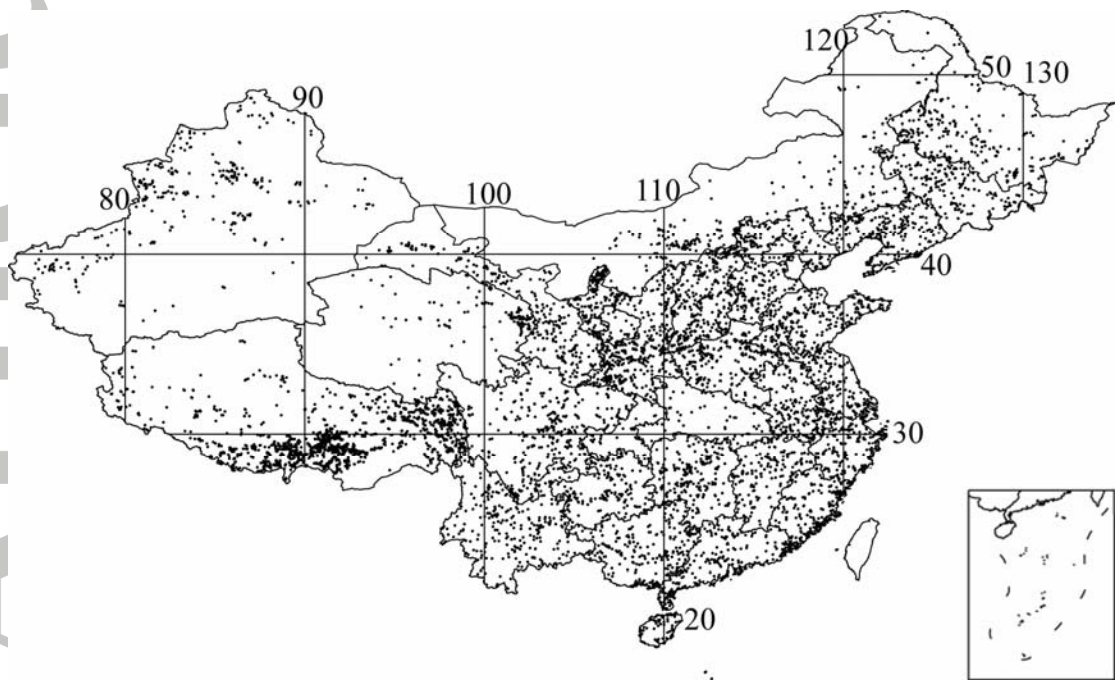
Figure 3. Spatial distribution of soil pH value (H_2O) in topsoil (0-0.3 m) in China. a, our study; b, Harmonized World Soil Database.

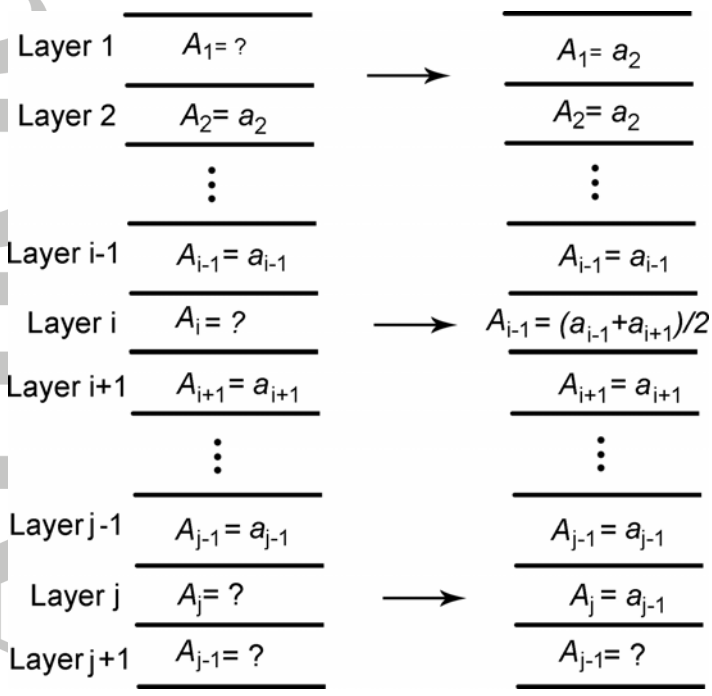
Figure 4. Spatial distribution of cation exchange capacity (me /100 g) in topsoil (0-0.3 m) in China. a, our study; b, Harmonized World Soil Database.

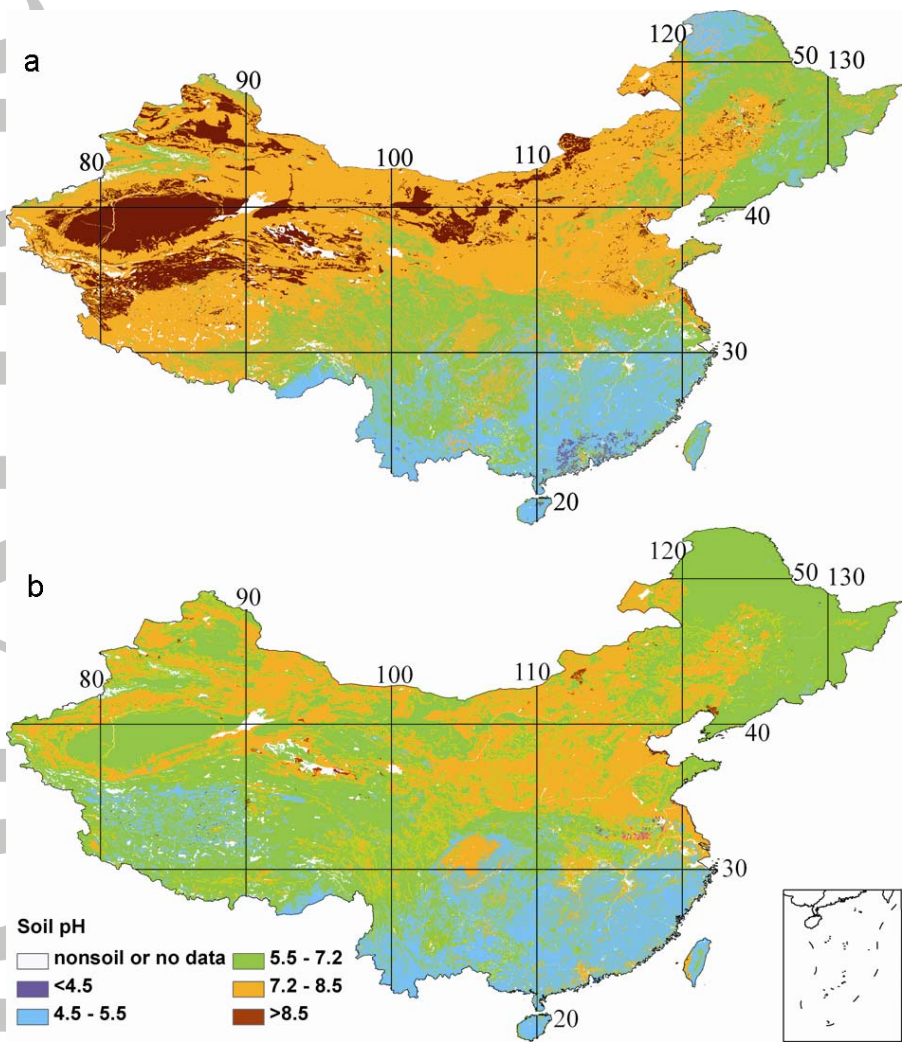
Figure 5. Spatial distribution of soil organic carbon (%) in topsoil (0-0.3 m) in China. a, our study; b, Harmonized World Soil Database.

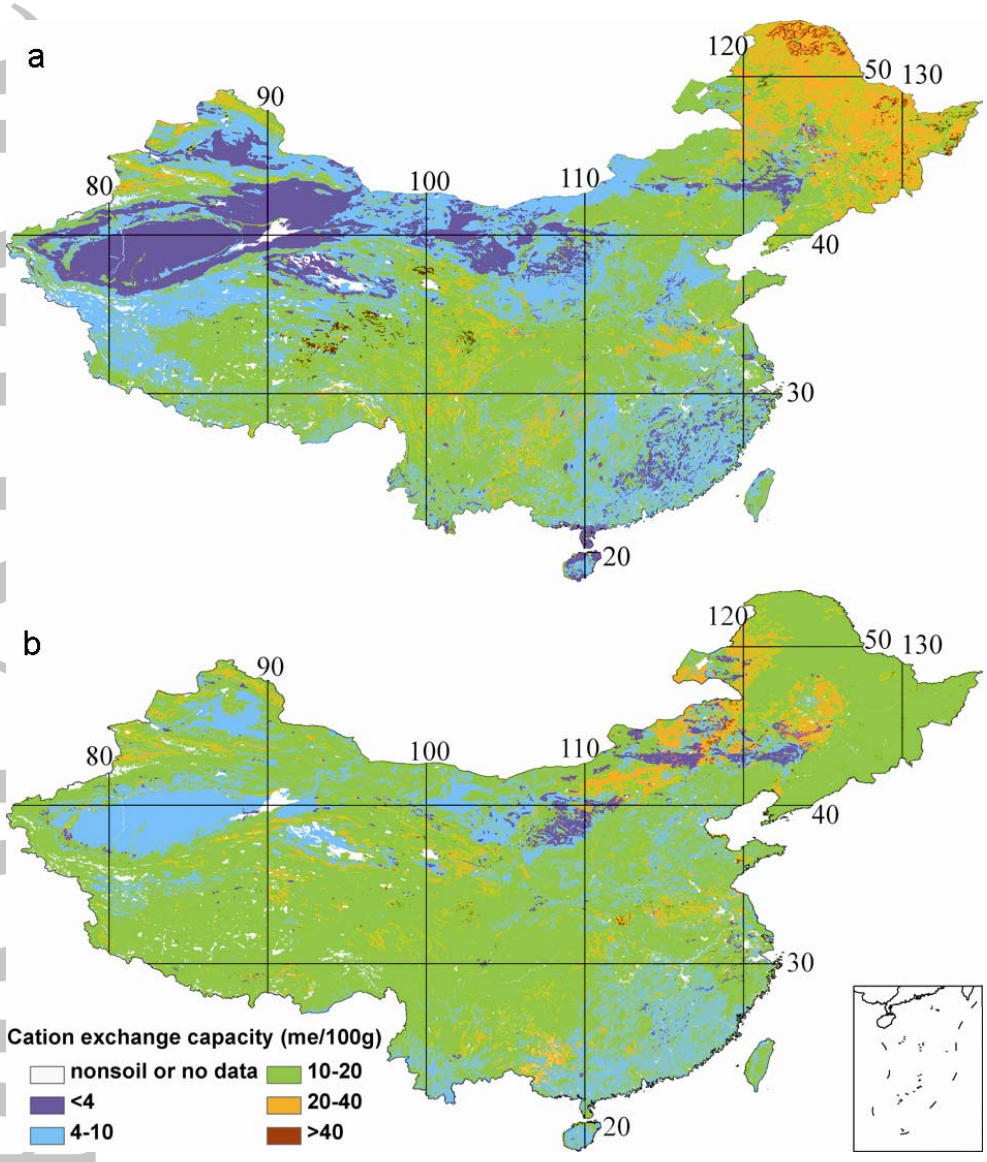
Figure 6. Spatial distribution of soil total nitrogen (a, %), Alkali-hydrolysable nitrogen (b, mg/kg), total phosphorus (c, %), available phosphorus (d, mg/kg) total potassium (e, %) and available potassium (f, mg/kg) in topsoil (0-0.3 m) in China.

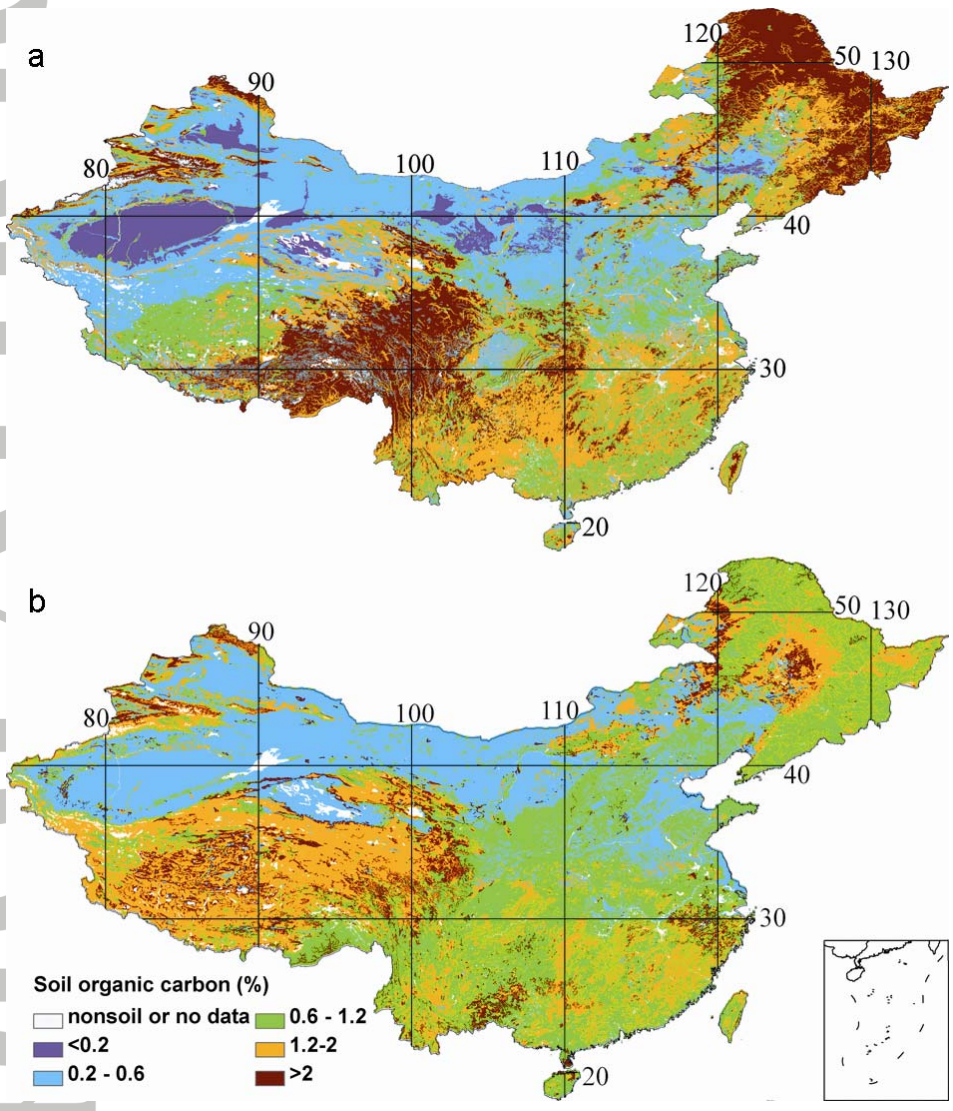
Figure 7. Quality control information (QC) of soil pH value (H_2O) in topsoil (0-0.3 m) by soil polygon linkage method. The meaning of the QC index is given in Table 2; only the first four digits of the QC index are given in the figure; and values 0 and 1 of the last digit are combined for brevity.

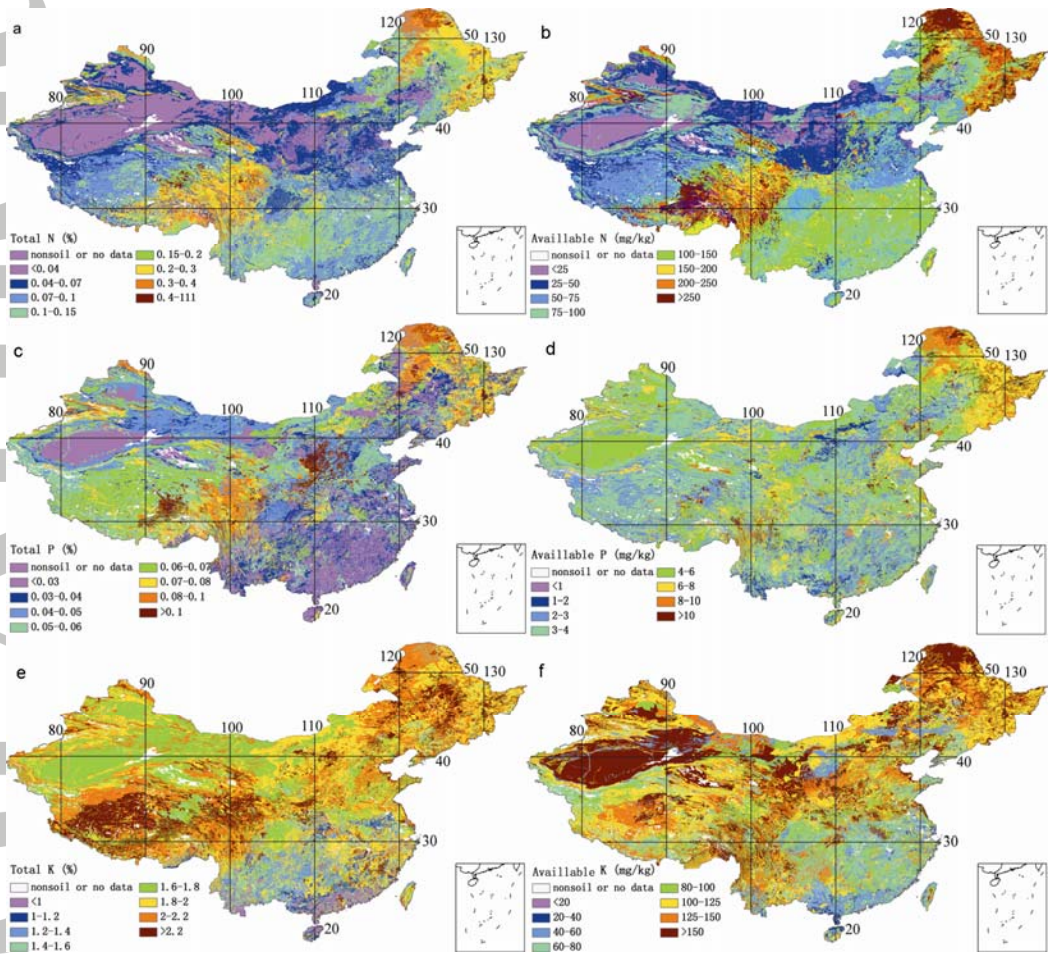


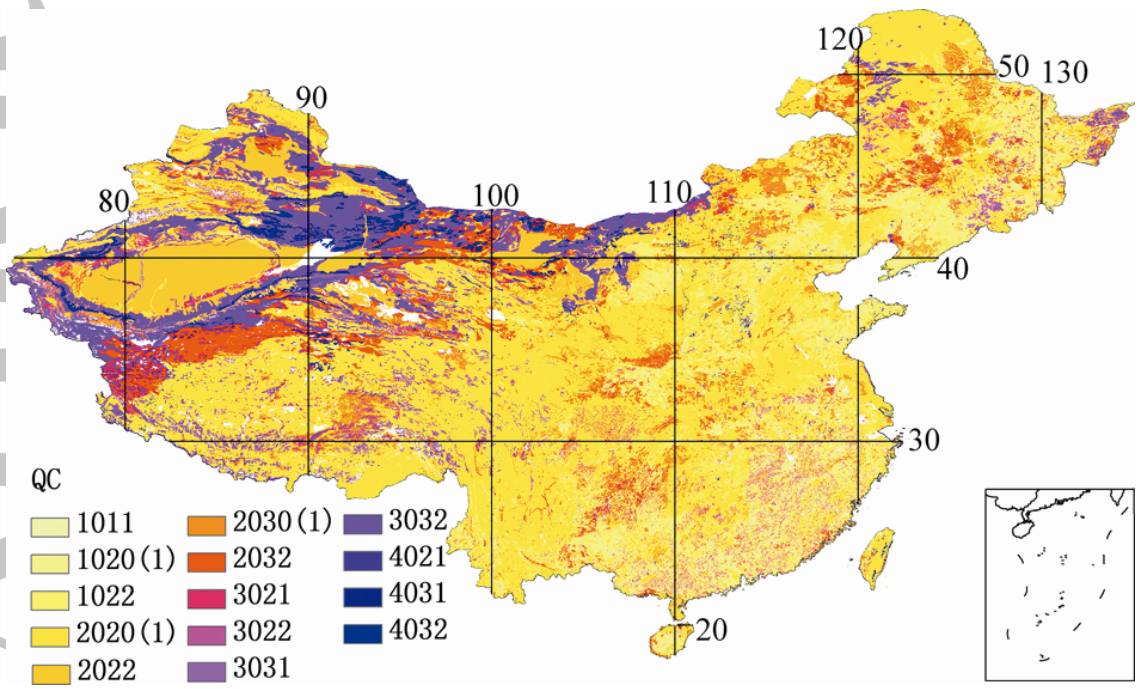












Section I. Source Data Citations

The following are books and unpublished documentations used in constructing the China soil profile attributes dataset:

Published books:

1. Soil Map of People's Republic of China. 1995. The Office for the Second National Soil Survey of China, Mapping Press, Xi'an (60pp.)
2. Soil Species of China, I-VI. 1993-1996. The Office for the Second National Soil Survey of China, Chinese Agriculture Press, Beijing (924pp. + 739pp. + 744pp. + 806pp. + 886pp. + 880pp.)
3. Soil Species of Shanxi, 1990. Soil Survey Office of Shanxi and Soil Workstation of Shanxi, Shanxi Science and Technology Press.
4. Soil Species of Zhejiang, 1990. Soil Survey Office of Zhejiang, Zhejiang Science and Technology Press.
5. Soil Species of Shandong, 1991. Soil and Fertilizer Workstation of Shandong, Chinese Agriculture Press.
6. Soil Species of Shaanxi, 1993. Soil Survey Office of Shaanxi, Shaanxi Science and Technology Press.
7. Soil Species of Qinghai, 1995. Agriculture Resource Planning Office of Qinghai, Chinese Agriculture Press.
8. Soil Species of Inner Mongolia, 1994. Soil Survey Office of Inner Mongoniai, Chinese Agriculture Press.
9. Soil Resource of Tibet, 1994. Land Admission of Tibet, Science and Technology Press.
10. Soil of Tibet, 1985. Qinghai-Tibet Plateau General Science Expedition of Chinese Academy of Science, Science and Technology Press.
11. Soil Species of Tibet, 1994. Land Admission of Tibet, Science and Technology Press.
12. Soil Species of Gansu, 1992. Soil Survey Office of Gansu, Science and Technology Press of Gansu.
13. Soil Species of Guangdong, 1991. Soil Survey Office of Gansu, Science and Technology Press.
14. Soil Species of Guangxi, 1993. Soil and Fertilizer Workstation of Gansu, Science and Technology of Guangxi.
15. Soil Species of Jiangsu, 1996. Soil Survey Office of Jiangsu, Science and Technology Press of Jiangsu.
16. Soil Species of Jiangxi, 1991. Land Use Administration of Jiangxi and Soil Survey Office of Jiangxi, Chinese Agriculture Science and Technology Press.
17. Soil Species of Henan, 1995. Soil Survey Office of Henan, Chinese Agriculture Press.
18. Soil Species of Hebei, 1992. Science and Technology Press of Hebei.
19. Soil Species of Guizhou, 1993. Soil Survey Office of Guizhou, Science and Technology Press of Guizhou.
20. Soil Species of Sichuan, 1993. Department of Agriculture of Sichuan and Soil Survey Office of Sichuan, Science and Technology Press of Sichuan.

21. Soil Species of Liaoning, 1991. Soil and Fertilizer Center of Liaoning, Liaoning University Press.
22. Soil Species of Jilin, 1992. Soil and Fertilizer Center of Jilin, Science and Technology Press of Jilin.
23. Soil Species of Xinjiang, 1993. Department of Agriculture of Xinjiang, Science, Technology and Health Press of Xinjiang.
24. Soil Species of Anhui, 1996. Soil Survey Office of Anhui, Science Press.
25. Soil Species of Yunnan, 1994. Soil Survey Office of Yunnan, Science and Technology Press of Yunnan.
26. Soil Species of Shannan of Tibet, 1990. Chinese Agriculture Science and Technology Press.

Unpublished documentations:

1. Soil Species of Fujian, 1989. Soil Survey Office of Fujian.
2. Soil Species of Ningxia, 1991. Agriculture Survey and Design Institute of Ningxia.
3. Soil Species of Hunan, 1990. Department of Agriculture of Hunan.
4. Soil Species of Heilongjiang, 1990. Soil Survey Office of Heilongjiang and Land Survey and Design Technology Center of Heilongjiang.
5. Soil Species of Lhasa, 1991. Agriculture and Animal Husbandry Bureau of Lhasa.
6. Soil Species of Nakqu Prefecture of Tibet, 1991. Zoning Office of Nakqu Prefecture of Tibet.
7. Soil Species of Ali Prefecture of Tibet, 1990. Agriculture and Animal Husbandry Bureau of Ali Prefecture of Tibet.
8. Soil Species of Shigatse Prefecture of Tibet, 1991. Agriculture and Animal Husbandry Bureau of Shigatse Prefecture of Tibet.
9. Soil Species of Linzhi Prefecture of Tibet, 1990. Zoning Office of Linzhi Prefecture of Tibet.
10. Soil Species of Changdu Prefecture of Tibet, 1990. Agriculture and Animal Husbandry Bureau of Changdu Prefecture of Tibet.
11. Soil of Changdu, 1987. Zoning Office of Changdu Prefecture.
12. Soil, Land Use and Land Resource Assessment of Yanjing County, 1990. Zoning Office of Changdu Prefecture and Agriculture and Animal Husbandry Bureau of Changdu Prefecture.
13. Soil of Mangkang, 1990. Zoning Office of Changdu Prefecture and Agriculture and Animal Husbandry Bureau of Changdu Prefecture.
14. Agriculture land Resource of Gongjue, 1989. Government of Gongjue and Resource Survey Team of Yibin.
15. Soil of Bitu, 1990. Zoning Office of Changdu Prefecture.
16. Soil of Basu, 1989. Agriculture Zoning Office of Changdu Prefecture.
17. Soil of Zuogong, 1990. Agriculture and Animal Husbandry Bureau of Changdu Prefecture.
18. Soil of Shengdao County, 1989. Agriculture Zoning Office of Changdu Prefecture.

19. Soil of Biandao, 1989. Agriculture Zoning Office of Changdu Prefecture.
20. Soil of Luolong, 1989. Agriculture Zoning Office of Changdu Prefecture.
21. Soil of Jiangdao, 1989. Agriculture Zoning Office of Changdu Prefecture.
22. Soil of Leiwuqi, 1987. Agriculture Zoning Office of Changdu Prefecture.
23. Soil and Land Assessment of Dingqing County, 1989. Agriculture Zoning Office of Changdu Prefecture.
24. Soil of Luolong, 1989. Agriculture Zoning Office of Changdu Prefecture.
25. Soil of Chaya, 1989. People's Government of Chaya, Helping Tibet Team of Agriculture University of Sichuan.
26. Soil of Tuoba, 1989. Agriculture Zoning Office of Changdu Prefecture and Helping Tibet Team of Sichuan.
27. Soil of Chengguan District of Lhasa, 1990. Agriculture Zoning Office of Lhasa.
28. Soil of Linzhou County, 1989. Agriculture Zoning Office of Lhasa.
29. Soil of Daozi County, 1990. Agriculture and Animal Husbandry Bureau of Lhasa.
30. Soil of Qushui, 1990.Qinghai-Tibet Plateau General Science Expedition of Chinese Academy of Science.
31. Soil of Mozhugongka County, 1990. Agriculture and Animal Husbandry Bureau of Lhasa.
32. Soil of Duilongdeqing County, 1990. Agriculture and Animal Husbandry Bureau of Lhasa.
33. Land Resource Survey Integrated Report of Lang County, 1990. Agriculture and Animal Husbandry Bureau of Linzhi Prefecture.
34. Land Resource Survey of Chayu County, 1990. Agriculture and Animal Husbandry Bureau of Linzhi Prefecture.
35. Soil of Naidong, 1990. Helping Tibet Team of Sichuan of Southwest Agriculture University.
36. Soil of Qiongjie, 1989. Land Resource Survey Team of Shannan Prefecture of Tibet.
37. Soil of Cuomei, 1988. Land Resource Survey Team of Shannan Prefecture of Tibet.
38. Soil of Jiacha County, 1988. Zoning Office of Shannan Prefecture of Tibet.
39. Soil of Gongga County, 1990. Agriculture and Animal Husbandry Bureau of Shannan Prefecture of Tibet.
40. Soil of Luozha County, 1989. Agriculture and Animal Husbandry Bureau of Shannan Prefecture.
41. Soil of Qusong County, 1991. Agriculture and Animal Husbandry Bureau of Shannan Prefecture.
42. Soil of Sangri County, 1990. Agriculture and Animal Husbandry Bureau of Shannan Prefecture.
43. Soil of Zhanong County, 1991. Land Resource Survey Team of Shannan Prefecture.
44. Soil of Cuona County, 1989. Land Resource Survey Helping Tibet Team of Yichang District of Hubei Province.
45. Soil of Nongzi County, 1989. Helping Tibet Team of Huanggang.

46. Soil of Langkazi County, 1991. Agriculture and Animal Husbandry Bureau of Shannan Prefecture.
47. Soil of Rikaze, 1991. Agriculture and Animal Husbandry Bureau of Rikaze Prefecture.
48. Soil of Dingjie County, 1989. Agriculture Natural Resource Survey Team of Rikaze Prefecture.
49. Soil of Sajia County, 1990. Agriculture Natural Resource Survey Team of Rikaze Prefecture.
50. Soil of Jiangzi County, 1990. Agriculture Natural Resource Survey Team of Rikaze Prefecture and Helping Tibet Team of Southwest Agriculture University.
51. Soil of Kangma County, 1988. Agriculture Natural Resource Survey Team of Rikaze Prefecture and Helping Tibet Team of Heze Prefecture of Shandong.
52. Soil of Nielamu County, 1991. Agriculture and Animal Husbandry Bureau of Rikaze Prefecture.
53. Soil of Jinong County, 1990. Agriculture and Animal Husbandry Bureau of Rikaze Prefecture and Helping Tibet Team of Southwest Agriculture University.
54. Soil Survey and Land Resource of Xiemengtong County, 1990. Agriculture Natural Resource Survey Team of Rikaze Prefecture.
55. Soil Survey and Assessment Report of Anren County, 1990. Agriculture Natural Resource Survey Team of Rikaze Prefecture.
56. Soil of Gangba County, 1988. Agriculture Land Resource Survey Team of Rikaze Prefecture and Helping Tibet Team of Heze Prefecture District of Shandong.
57. Soil of Zhongba County, 1989. Agriculture Natural Resource Survey Team of Rikaze Prefecture.
58. Soil Survey and Land Resource Assessment Report of Saga County, 1989. Agriculture Natural Resource Survey Team of Rikaze Prefecture.
59. Soil Survey and Assessment Report of Bailang County, 1990. Agriculture Natural Resource Survey Team of Rikaze Prefecture.
60. Soil of Nanmulin County, 1990. Agriculture and Husbandry Bureau of Rikaze Prefecture.

Section II. Description of Measurement

The description of the measurement method of soil properties is mainly extracted from handbooks of soil survey in China [*Li*, 1983; *National Soil Survey Office*, 1992; *Shi and Bao*, 1980].

1. pH Value

Two major methods, i.e. the colorimetric method and the electrometric method, are used to measure pH value.

The colorimetric method is a fast cheap method making a visual comparison with a colored standard for each 0.2 pH value interval. The precision is relatively low.

The electrometric method is the most frequently used method for soil analysis as it is more precise than the colorimetric method. An electrode is plunged into the soil solution to measure the potential difference E (volts) as a function of the concentration in H^+ ions. The measurement can be carried out in different solution including water, KCl, NaCl and CaCl. In addition, the ratio of soil to water can be different. As a result, different measurements may give different pH values. In our dataset, pH value is measured in suspension of soil in water.

2. Organic Matter Fraction

The organic matter in soil consists of: living plant roots, dead but little-altered plant remains, partly decomposed plant remains, colloidal organic matter, living micro-organisms and macro-organisms, inactive or inert organic matter. The litter lying on the top of the soil surface is ignored in the measurement.

Two major methods, i.e. the heating method (oxidation with potassium dichromate, $K_2Cr_2O_7^{2-}$), and the hydrothermal synthesis method (oxidation with potassium dichromate), are used to measure organic matter fraction.

In the heating method, carbon in the soil organic matter is oxidized and correspondingly, $Cr_2O_7^{2-}$ is reduced in the H_2SO_4-$K_2Cr_2O_7^{2-}$ solution by heating, and the remaining $Cr_2O_7^{2-}$ is titrate by adding standard Fe^{2+} standard solution. Organic matter is calculated by:

$$SOM = (a-b)N_{Fe} \times 0.003 \times 1.724 \times 1.08 / W \qquad \text{(S1)}$$

where SOM is soil organic matter, *a* is the volume of Fe^{2+} standard solution in the blank titration, *b* is the volume of Fe^{2+} standard solution in the soil sample titration, N_{Fe} is the equivalent concentration of the Fe^{2+} standard solution, 0.003 is the equivalent gram of carbon, 1.724 is the factor between soil organic carbon and soil organic matter, 1.08 is the oxidation correcting coefficient, and W is the weight of dry soil sample. The allowed difference of two measures of a soil is 0.05 % when SOM is less than 3 %, and 0.1- 0.3 % when SOM is large than 3 %. It should be noted that the direct measurement is soil organic carbon instead of soil organic matter.

The hydrothermal synthesis method has the same principle as the heating method. However, it utilizes the heat from the chemistry reaction instead of heating outside. The oxidation of organic carbon is more incomplete (about 77 %), and the oxidation

correcting coefficient is 1.33.

3. Total N

Two major methods, i.e. the Kjeldahl method: the diffusion absorption method and the distillation method, are used to measure total N.

The aim of Kjeldahl method is to transform all N forms into ammonium-N form using a wet method in a concentrated sulphuric acid medium in the presence of catalysts. In the distillation method, the ammonium-N form is transferred to a basic medium by steam distillation, and then titrated volumetrically in the presence of an indicator. Total N is calculated as:

$$TN = (V - V_0)N \times 0.014 \times T / W \quad \text{(S2)}$$

where TN is total N, N is equivalent concentration of standard acid, V is the volume of standard acid consumed by soil sample, V_0 is the volume of standard acid in blank titration, 0.014 is equivalent weight of N, W is the weight of dry soil sample, and T is the ration of total reaction medium to the consumed medium. The allowed difference of two measures of a soil is 0.005 %.

In the diffusion absorption method, the ammonium N is absorbed by boric acid, and then titrated. The diffusion absorption method has similar calculation and precision to the distillation method.

4. Total P

Two major methods, i.e. the $HClO_3$-H_2SO_4 digestive and Mo-Sb colorimetry method, and the NaOH alkali fusion and Mo-Sb colorimetry method, are used to measure total P.

In the $HClO_3$-H_2SO_4 digestive and Mo-Sb colorimetry method, soil is mixed with $HClO_3$-H_2SO_4 at a high temperature, and phosphorus forms are transformed into orthophosphate form, and then orthophosphate is measured by Mo-Sb colorimetry. In most cases, 97-98 % of P in soil is transformed. The allowed difference of two measures of a soil is 0.005 %.

The process of the NaOH alkali fusion and Mo-Sb colorimetry method is similar to the $HClO_3$-H_2SO_4 digestive and Mo-Sb colorimetry method. However, this method mixed soil with NaOH.

5. Total K

Two major methods, i.e. the NaOH alkali fusion and flame photometer method, and the HF-$HClO_4$ sodium tetraphenylborate method, are used to measure total K.

In the NaOH Alkali fusion and flame photometer method, poorly soluble potassium transformed into soluble forms after alkali fusion, and potassium is measured by flame photometer method. The allowed difference of two measures of a soil is 0.05% when the content of potassium is less than 1 %.

In the HF-$HClO_4$ sodium tetraphenylborate method, potassium is transformed into soluble forms, and sodium tetraphenylborate is used to transformed soluble potassium into deposit.

6. Alkali-hydrolysable N

Two major methods, i.e. the Conway diffusion method and the distillation method, are used to measure alkali-hydrolysable N.

In the Conway diffusion method, alkali-hydrolysable N is transformed into ammonium N in alkaline solution, absorbed by boric acid, and then titrated. The calculation of alkali-hydrolysable N is similar to total N. The allowed difference of two measures of a soil is 5 ppm.

In the distillation method, alkali-hydrolysable N is extracted by H_2SO_4, transformed into ammonium salt, distillated, absorbed by boric acid, and then titrated.

7. Available P

Two major methods, i.e. the NH_4F - HCL method (acid soil) and the $NaHCO_3$ method (basic and carbonate soil), are used to measure available P.

In both methods, some active forms of P are extracted by reagents. $NaHCO_3$ and NH_4F - HCL can extract some active Ca-P, Fe-P and Al-P. It should be noted that the measurements of available P are just indicators. It is meaningful to compare the results only when the measuring method is the same.

8. Available K

Two major methods, i.e. the NH_4OAc extracting and flame photometer method, and the Na_2SO_4 extracting and sodium tetraphenylborate turbidimetry method, are used to measure available K.

In the NH_4OAc extracting and flame photometer method, potassium is extracted by NH_4OAc solution, and available potassium is measured using flame photometer. Potassium measured by this method correlates well with K fertilizer. The allowed difference of two measures of a soil is 2 ppm.

In the Na_2SO_4 extracting and sodium tetraphenylborate turbidimetry method, potassium is extracted by Na_2SO_4 solution, and soluble potassium and sodium tetraphenylborate react producing deposit. Available K measured by this method is less than the NH_4OAc extracting and flame photometer method, and the precision is lower.

9. Cation Exchange Capacity

Three major methods, i.e. the ammonium acetate method (pH7.0), the natrium aceticum method (pH8.2) and the summation method are used to measure cation exchange capacity. Methods for measuring CEC thus depend on measuring conditions including pH, buffered medium, nature and concentration of cations, pretreatments and so on.

In the Ammonium acetate method and the natrium aceticum method, the soil is saturated by the ammonium or natrium counter-ion at pH7.0 or pH8.2, respectively. Ammonium or natrium is adsorbed and an equivalent quantity of cation is moved. The exchangeable cations are titrated in the percolation solution.

In the summation method, exchangeable cations (Ca^{2+}, Mg^{2+}, K^+, Na^+ ,H^+ , Al^{3+} , Fe^{2+}, Fe^{3+} , Mn^{2+}) are summed up to get the effective CEC.

10. Exchangeable H^+ and Al^{3+}

One major method, i.e. the KCl titration method, is used to measure exchangeable H^+.

The sample is percolated with a KCl solution which enables extraction of exchangeable acidity (H^+ and Al^{3+}). The solution is titrated by NaOH to get the sum of exchangeable H^+ and Al^{3+}. Another percolated solution is added with NaF to transform Al^{3+} into $[AlF_6]^{3-}$, so that exchangeable H^+ can be titrated by NaOH.

11. Exchangeable Ca^{2+} and Mg^{2+}

Two major methods, i.e. the EDTA volume method and the atomic absorption spectroscopy method, are used to measure exchangeable $Ca^{2+}$and Mg^{2+}.

In both methods, the soil is saturated by the NH_4^+ cation to move the exchangeable cations into the solution. In the EDTA volume method, the solution is titrated by EDTA at different pH value to measure exchangeable $Ca^{2+}$and Mg^{2+}. In the atomic absorption spectroscopy method, exchangeable $Ca^{2+}$and Mg^{2+} are measured by atomic absorption spectroscopy.

12. Exchangeable Na^+ and K^+

Two major methods, i.e. the flame photometer method and the atomic absorption spectroscopy method, are used to measure exchangeable Na^+ and K^+.

In both methods, the soil is saturated by the NH_4^+ cation to move the exchangeable cations into the solution. In the flame photometer method, the exchangeable $Na^+$and K^+ are measured by flame photometer. In the atomic absorption spectroscopy method, exchangeable $Na^+$and K^+ are measured by atomic absorption spectroscopy.

13. Particle-Size Distribution and Gravel

Three major methods, i.e. the sieving method, the pipette method and the hydrometer method, are used to measure particle-size distribution.

The sieving method is used to measure the particles with diameter larger than 0.25 mm (including gravel). This method separates soils on sieves with different mesh sizes.

The pipette method and the hydrometer method are used to measure small particles with diameter less than 0.25 mm in soil based on Stokes' Law. This law states that denser (larger, usually) particles sink farther than less dense (smaller) particles when suspended in a liquid. The pipette method measures the actual percent by weight of each particle size class in soil sample. The hydrometer method uses the density of the soil/water mixture. For the pipette method, the allowed difference of two measures is 1 % for clay and 2 % for sand and silt. For the hydrometer method, allowed difference of two measures is 3 % for clay and 4 % for sand and silt.

Particle-size Distribution and soil texture class is measured under 17 schemes of the ISSS (International Society of Soil Science) or Katschinski's soil texture classification system. The dominance are 2- 0.2- 0.02- 0.002 mm, 2- 0.02- 0.002 mm and 1- 0.25- 0.05- 0.01- 0.005- 0.001 mm. For model use, all schemes are converted into sand, silt and clay of USDA standard (i.e. 2- 0.5- 0.002 mm). Gravel is defined soil

fraction larger than 2 mm and 1mm in ISSS and Katschinski's schemes, respectively.

14. Bulk Density

Two major methods, i.e. the ring method (core method) and the wax immersion method, are used to measure bulk density.

The ring method uses steel cylinders of known volume that are driven in the soil vertically or horizontally by percussion. The following equation is used to calculate bulk density

$$BD = (M - G) \times 100 / V / (100 + W) \quad \text{(S3)}$$

where BD is bulk density, M is the weight of cylinder and soil, G is the weight of cylinder, V is the volume of cylinder, and W is percent of soil water content. The allowed difference of two measures of a soil is 0.02 g/cm^3, and the average is used as the final measurement. The method cannot be used when the soil is too hard or fragile.

When the soil is too hard or fragile, the wax immersion method is used. This method used wax to wrap up the soil sample and measured its volume in water based on Archimedes' principle. The allowed difference of two measures of a soil is 0.03 g/cm^3, and the average is used as the final measurement.

15. Porosity

Porosity is calculated by the following equation:

$$P = (1 - BD / PD) \times 100 \quad \text{(S4)}$$

where PD is particle density.

16. Color

Soil color is recorded according by comparing with a Munsell system color chart. Usually, the corresponding moisture condition is also recorded as wet or dry, as the moisture condition can change the soil color. The Munsell notation with three characteristics of the color: hue, value, and chroma are recorded.

17. Structure

In soils that have structure, the shape and size of the structure units are described (grade is not included in our dataset). Soils lacking structure are referred as structureless. The structure is recommended to observe at moderate moisture content and to use fingers or a tool to touch the soil to disperse.

We reclassified structure into 20 classes: Angular, angular blocky, blocky, big blocky, crumb granular, columnar, crumb, granular, granular blocky, massive, nutty, nutty blocky, nutty grain, no structure, platy, prismatic, small blocky, single grain, sheet, and squamose.

18. Consistency

Consistency is highly dependent on the soil-water state and the description has little meaning unless the water state class is specified or is implied by the test.

However, the soil water condition is not given in our dataset, as the default is the natural water condition.

We reclassified consistency into 15 classes: compacted, extremely firm, extremely loose, firm, hard, loose, loose and soft, slightly firm, slightly hard, slightly loose, slightly loose and soft, slightly soft, soft, sticky and firm, and very firm.

19. Root quantity

Root quantity and size is described qualitatively in the data sources. We reclassified root quantity into 7 classes: none, very few, moderately few, a few, moderate, many, a great many, and dense.

REFERENCES

Li, Y. (1983), Methods of soil agrochemical analysis, Science press, Beijing.

National Soil Survey Office (1992), Soil Survey Technology of China, Agriculture Press, Beijing.

Shi, R., and S. Bao (1980), Soil Agrochemical Analysis, Agriculture Press, Beijing.

Section III. Maps of Soil Properties of China

1. Soil profile depth and horizon thickness

Soil depth determines the lower boundary depth in land surface modeling. However, measurements of soil depth are rare and it is difficult to assess the exact soil depth. As a result, LSMs usually set the lower boundary as a certain default value [*Dai et al.*, 2003]. A soil body is composed of several horizons divided by the physical, chemical or biological differences. The inclusion of horizon thickness in LSMs may improve the simulation of water movement [*Francés and Lubczynski*, 2011].

In our study, the distribution of soil profile depth was derived. We also derived the structure of natural horizon for all soil types. First, the natural horizons that have horizon names are reclassified into seven basic horizons, denoted as K, O, H, A, E, B, C, R, which represent mineral crust horizon, litter or grass mat horizon, peat horizon, humus horizon, eluvial horizon, illuvial horizon, parent material horizon and parent rock horizon, respectively [*National Soil Survey Office*, 1992; *Li et al.*, 2004], and three transition horizons, i.e. AB, AC and BC, which have characteristics of two basic horizons. Then, the standard horizon sequence is defined as K, O, H, A, E, AB, AC, B, BC, C, and R. And the average thickness for each horizon with the same soil type is calculated. As there is a large variety of horizons among soil profiles with the same soil type, the derived structure of soil profiles can be only used as a reference.

Figure S1 shows the distribution soil profile depth by soil type. Soils on Qinghai-Tibet Plateau are quite shallow. The soil seems to be deeper in the north than in the south. It should be noted that this depth cannot indicate the depth to rock, and we do not know how deep the soil is below the observed soil profile depth.

2. Soil pH value (H2O)

Figure S2 shows the soil pH distribution in LSM's 3 standard layers in China. Soils in the south and northeast are acid while soils in the north and northwest are basic or alkaline. The pH value does not change a lot with soil depth.

3. Soil organic matter (SOM)

SOM dissolved organic chemicals act to glue soil particles together, enhancing aggregation and increasing overall soil aeration, and water infiltration and retention. The dark consistency of humus causes soils high in SOM to be dark brown or black in color, increasing the amount of solar radiation absorbed by the soil and thus, soil temperature. SOM persists mainly because of physicochemical and biological influences from the surrounding environment.

Figure S3 shows the SOM distribution in LSM's 3 standard layers in China. SOM is high in the southeastern part of Qinghai-Tibet Plateau and in northeast China, and low in the north and northwest, especially in the deserts. It decreases with soil depth.

4. Soil Nitrogen (N), phosphorus (P) and potassium (K)

Figure S4-6 shows the total N, total P, and total K distribution in LSM's 3 standard layers in China. Total N and total P decrease with depth, but total K does not. Total N is high in the eastern part of Qinghai-Tibet Plateau and northeast, and it is low in the north, northwest and deserts. Total P is high in the northeast and Loess Plateau, and it is low in the south and deserts. Total K is high in the northeast and Qinghai-Tibet Plateau and it is low in the south.

Figure S7-9 shows alkali-hydrolysable N, available P and available K. All of them decrease rapidly with the depth. High alkali-hydrolysable N appears in the northeast and the east part of Qinghai – Tibet Plateau, and the lowest appears in the deserts. High available P appears in the northeast, and low available P appears in the deserts, Qinghai–Tibet Plateau and the south. High available K appears in the north, and low available K appears in the deserts.

5. Cation exchange capacity and soil exchangeable cations

Figure S10 shows CEC distribution of China. High CEC is found in central and northeast and low CEC is found in the north and south, especially in the deserts.

Figure S11-16 shows exchangeable H^+, Al^{3+}, Ca^{2+}, Mg^{2+}, K^+ and Na^+. Exchangeable H+ and K+ decrease with depth for most areas. Exchangeable Al^{3+}, Ca^{2+}, Mg^{2+}, and Na^+ do not decrease with depth apparently for most areas and in some areas they even increase with depth. High exchangeable H^+ appears in the desert, south east of Qinghai- Tibet Plateau and the south, and low exchangeable H^+ appears in the north. High exchangeable Al^{3+} appears in the south, and low exchangeable appears in the north. The opposite happens to exchangeable Ca^{2+} and Mg^{2+}. High exchangeable K^+ appears in the north, and low exchangeable K^+ appears in the deserts and south. High exchangeable Na^+ appears in the northeast, and low exchangeable Na^+ appears in the deserts and north and south.

6. Soil sand, silt, clay and gravel fraction

Particle-size distribution or soil texture is a basic physical property of soil required by most land surface models to simulate the hydraulic and thermal processes. We derived sand, silt and clay fractions under FAO-USDA System, which is most commonly used in land surface models. In previous study, we used the particle-size distribution as an example of the linkage method, which is also employed in this study. The validation showed that our estimates were more reliable compared to Harmonized World Soil Database [*Shangguan et al.*, 2012].

Figure S17-19 show sand, silt and clay distribution of China. High sand content is found in the west and the north of China, while high clay content is found in the east and the south of China. High silt content is found in the middle latitude area, and low silt content is found in the desert areas. Soil fractions do not change a lot for most of the area in China.

Gravel is composed of materials in a soil that have a particle size larger than 2 mm. Gravel changes soil physical properties such as bulk density, and water movement in soil.

Figure S20 shows gravel content of China. Gravel content is low in the agriculture area of east China, and it is high on the Qinghai-Tibet Plateau. In most areas, gravel content increases with depth.

7. Soil bulk density and porosity

Bulk density is often used to predict soil water retention properties in LSMs. However, measurements of bulk density are limited and global datasets such as HWSD have only reference bulk density calculated from particle size distribution. Because the lack of data, LSMs also use pedotransfer function to estimate bulk density, which brings uncertainty into the estimates [*Dai et al.*, 2003]. We derived bulk density based on measured data from more than 1,000 profiles in China. Porosity is a primary determining factor in hydraulic conductivity. Porosity is inversely related to bulk density and is usually calculated from bulk density (BD) and particle density (PD). However, there are no records of particle density in our dataset, either, and there are records of porosity for only a small portion of profiles.

Figure S21 shows bulk density of China derived directly and by pedotransfer function. The difference is quite large in the south and northeast. It is recommended to use direct estimates to avoid uncertainty brought by equation 6. Bulk density is low in the northeast and south, and high in the desert areas and deep soil of Qinghai-Tibet Plateau. In most areas, bulk density increases with depth.

8. Soil color

Though soil color is not seen as an agent affecting soil behavior, soil color of the surface soil layer, together with soil moisture, determines soil albedo and thus affect the radiation balance in LSMs. Soil color is associated with soil water condition. Soil color is recorded under wet, dry and unclear soil water condition, and the derived results were given in these three cases. As soil water condition changes with time, it was not derived. Soil color was offered in Munsell color system and can be converted to computer-friendly RGB triplets (i.e., red, green and blue) [*Viscarra Rossel et al.*, 2006]. The soil albedo can be estimated from the Munsell value color component. [*Post et al.*, 2000]

Figure S23 shows the surface wet and dry color of soil in China. The three dimension of Munsell color system were given separately. The dominant hues are between 5YR (Yellow Red) and 10YR. Dry color is more near 10YR than wet color. Value or lightness ranges from black (value 0) to white (value 10). It is mostly around 4 in China. High value appears in the deserts, and low value appears in the northeast and east part of Qinghai-Tibet Plateau. Value of dry soil is usually higher than wet soil. Chroma represents the purity of color, with high chroma being high pure. High chroma appears in the south, and low chroma appears in the northeast, central and east part of Qinghai-Tibet Plateau.

9. Soil structure and consistency

The first two components with the highest percentage were given for structure and

consistency in the derived dataset. Structure, in conjunction with texture, controls porosity of the soil and thus affects water relations, aeration, root penetration, and the metabolic activities of soil flora and fauna. Consistency has important effect on water holding capacity and water movement through the soil. Structure and consistency are described as categories. However, there is no universal standard to describe and measure structure and consistency, and thus it is hard for LSMs to utilize them.

Figure S24 shows the structure of the second layer (4.5-9.1 cm). The distribution of structure varies across the country, and the proportion of each component is rather low, even for the dominant class.

Figure S25 shows the consistency of the second layer. The distribution of consistency seems to be homogeneous, and the main classes of consistency are loose and slightly firm. Most of the dominant classes are above 50 %.

10. Root quantity

Root quantity (abundance) is needed in LSMs to calculate the transpiration of root from each soil layer [*Jackson et al.*, 1999; *Schenk and Jackson*, 2002; *Zeng*, 2001]. In our dataset, root abundance is described as none, few, moderate, moderately few, moderate, many, great many and dense in the dataset. These descriptions were set as numbers between 0 and 7, and standardized as quantitative values.

Figure S26 shows root abundance of China. Most of the near surface soils have many roots and most of the layer 6 soils have few roots.

REFERENCES

Dai, Y., Z. X., D. R. E., 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. Yang (2003), The Common Land Model, American Meteorological Society, 8, 1013~1023.

Francés, A. P., and M. W. Lubczynski (2011), Topsoil thickness prediction at the catchment scale by integration of invasive sampling, surface geophysics, remote sensing and statistical modeling, Journal of Hydrology, 405(1-2), 31-47.

Jackson, R. B., L. A. Moore, W. A. Hoffmann, W. T. Pockman, and C. R. Linder (1999), Ecosystem rooting depth determined with caves and DNA, Proceedings of the National Academy of Sciences, 96, 11387-11392.

Li, T., Y. Zhao, K. Zhang, Y. Zheng, and Y. Wang (2004), Soil Geography (3ed.), Higher Education Press, Beijing.

National Soil Survey Office (1992), Soil Survey Technology of China, Agriculture Press, Beijing.

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 Science Society of America Journal 64(3), 1027-1034.

Schenk, H. J., and R. B. Jackson (2002), Rooting depths, lateral root spreads and below-ground/above-ground allometries of plants in water-limited ecosystems, Journal of Ecology, 90, 480- 494.

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.

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.

Zeng, X. (2001), Global vegetation root distribution for land modeling, Journal of Hydrometeorology, 2, 525-530.

Figure S1. Soil profile depth by soil type in China

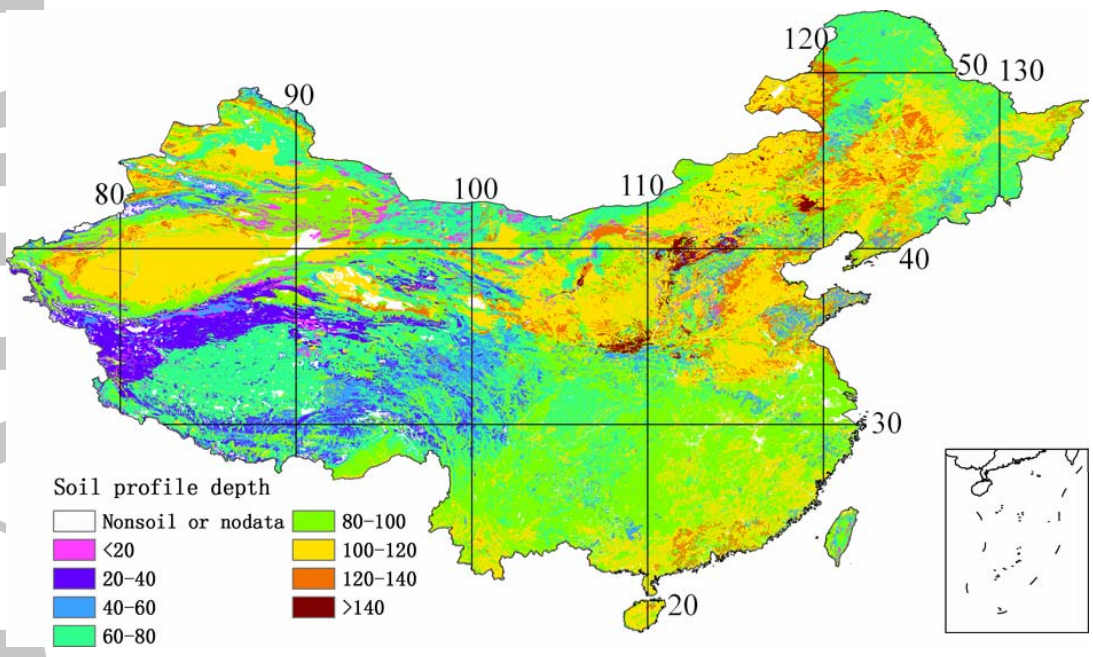


Figure S2. Spatial distribution of soil pH value (H2O) in China. a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

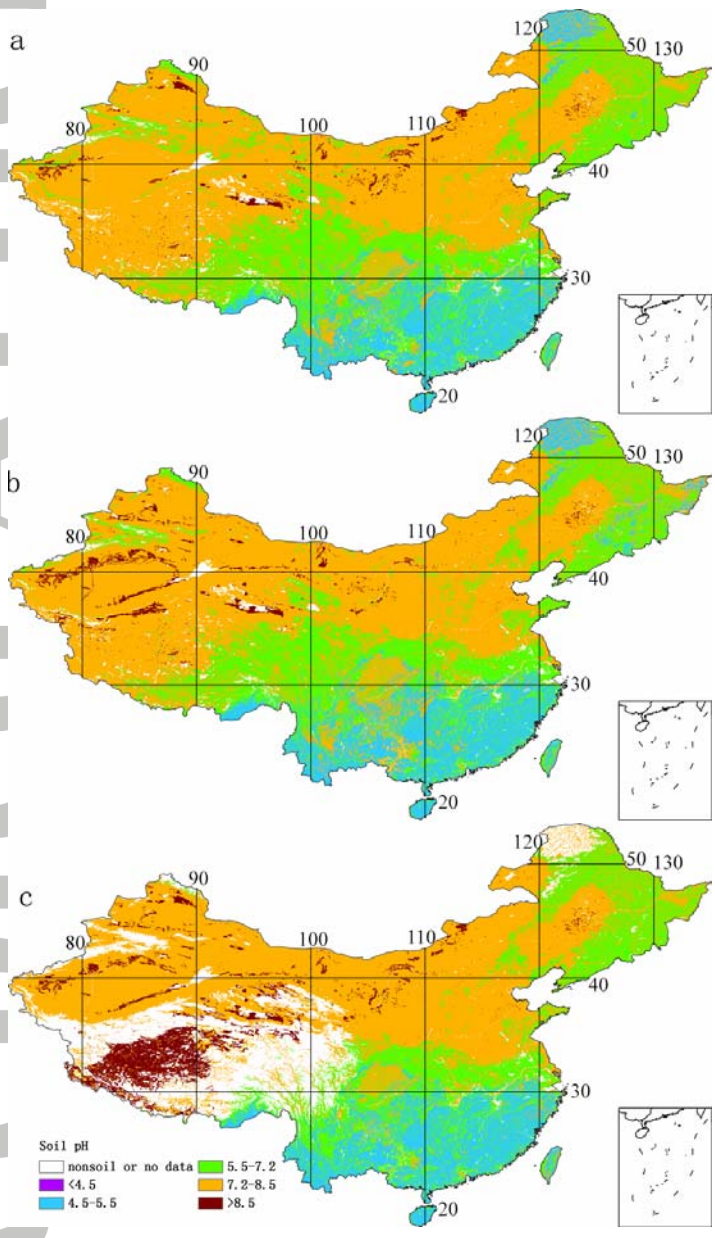


Figure S3. Spatial distribution of soil organic matter (SOM) in China (%). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

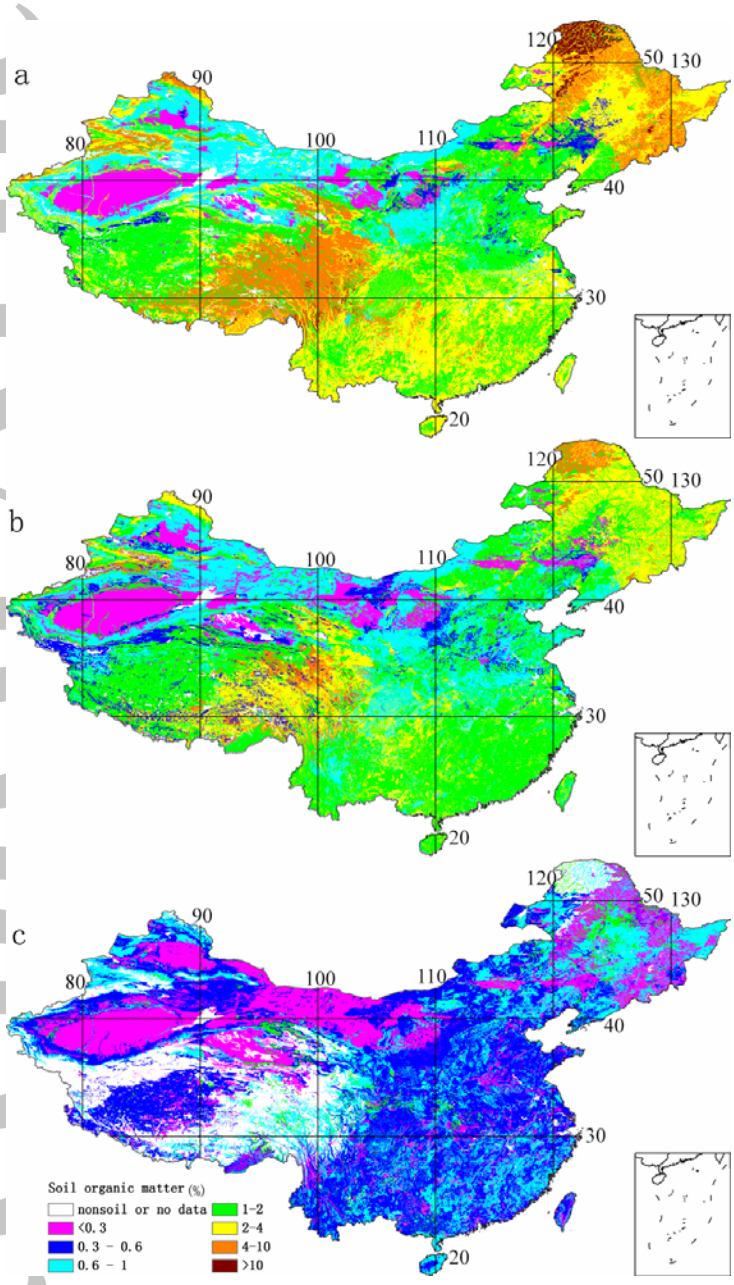


Figure S4. Spatial distribution of soil total nitrogen (N) in China (%). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

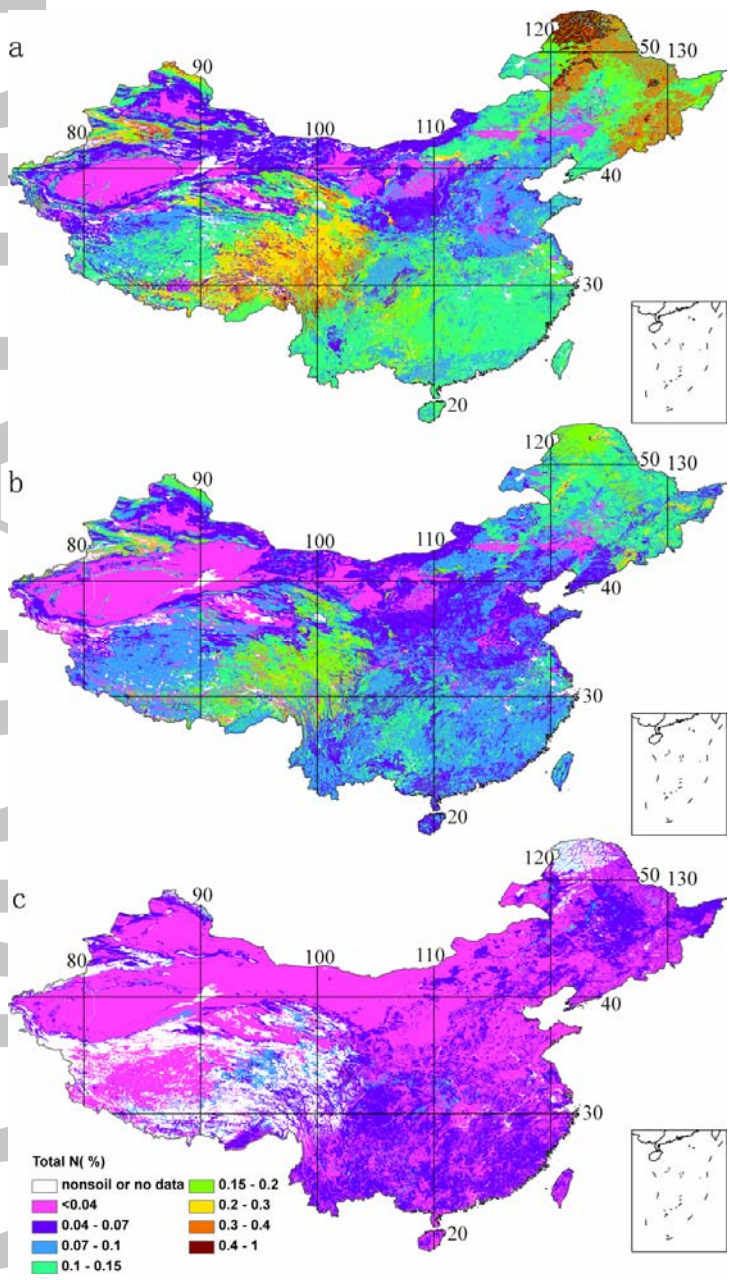


Figure S5. Spatial distribution of soil total phosphorus (P) in China (%). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

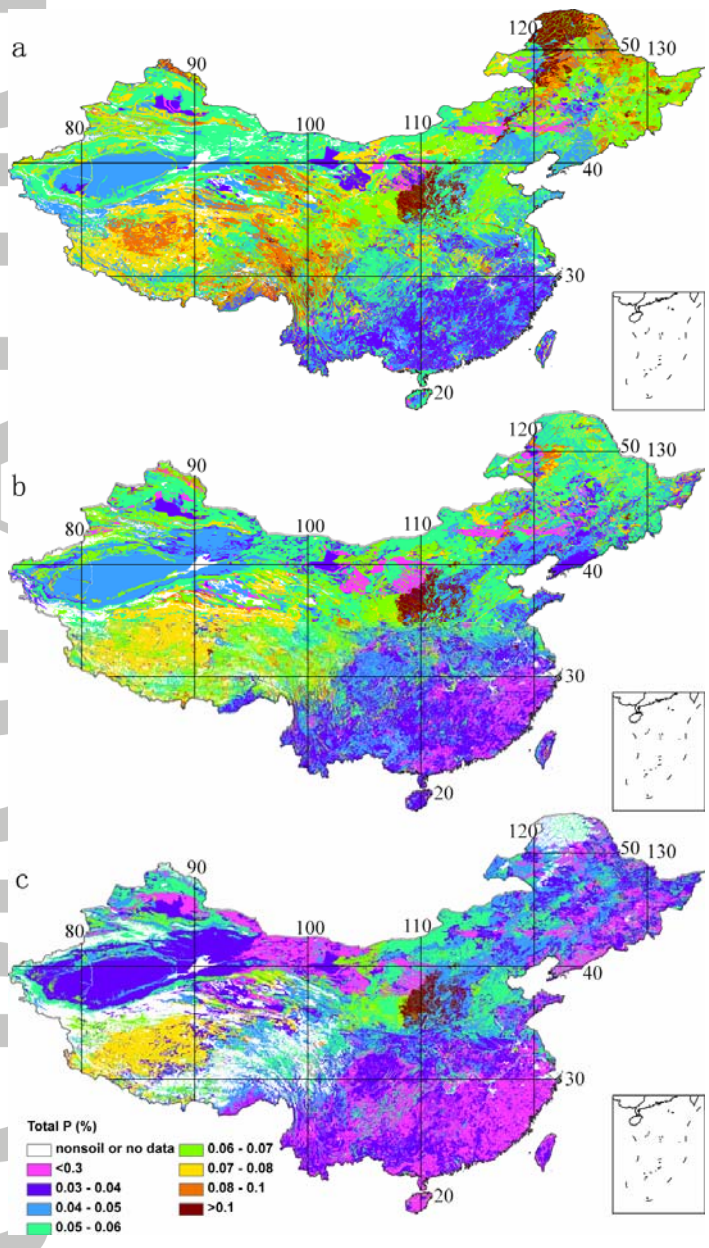


Figure S6. Spatial distribution of soil total potassium (K) in China (%). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

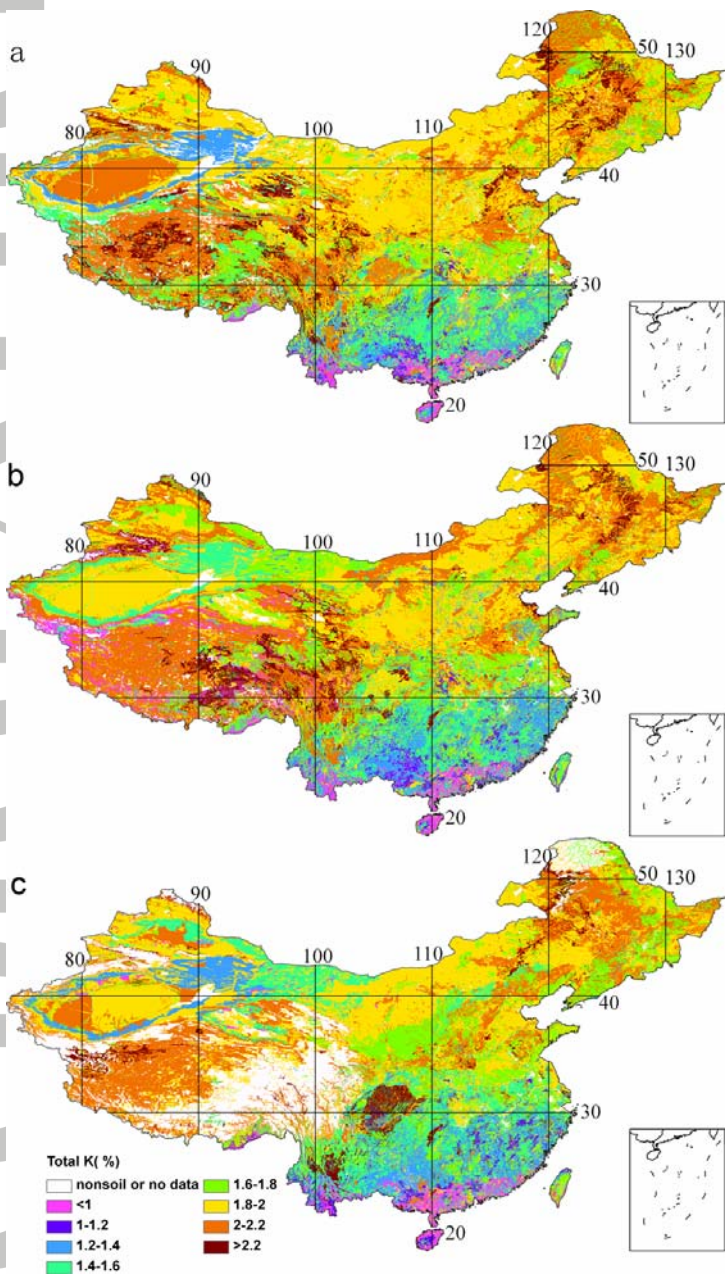


Figure S7. Spatial distribution of alkali-hydrolysable nitrogen (N) in China (mg/kg). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

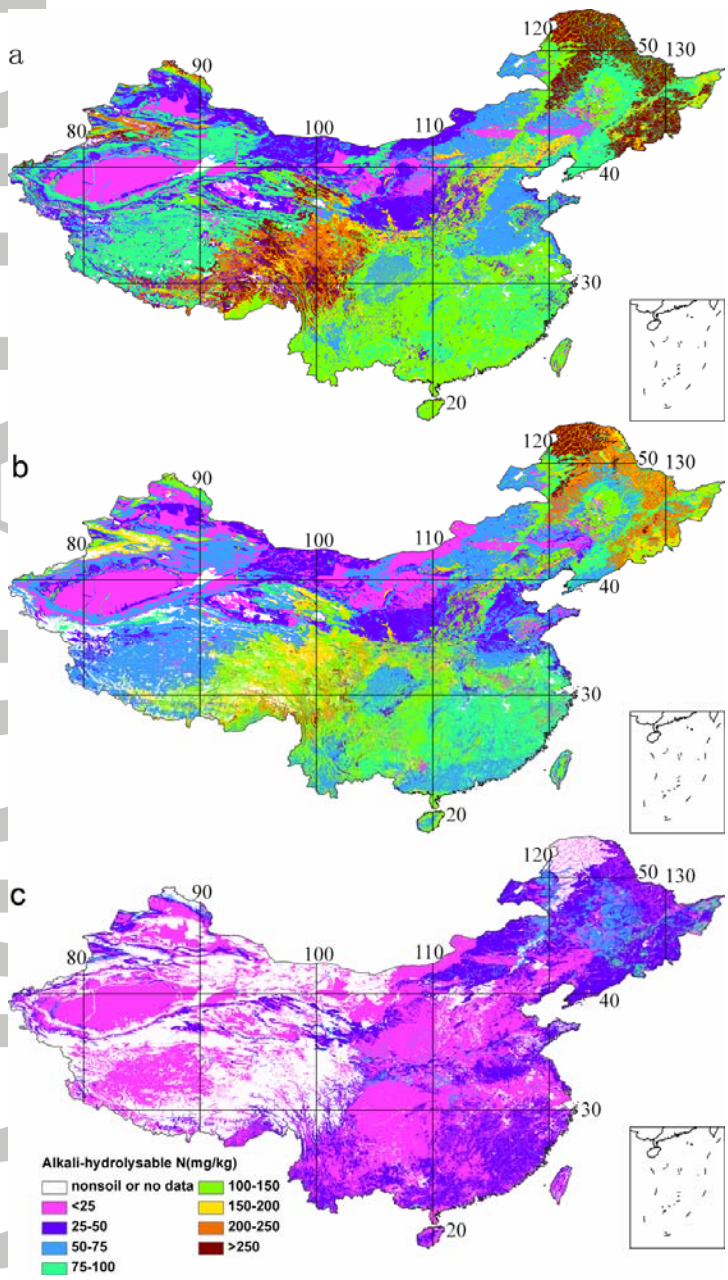


Figure S8. Spatial distribution of available phosphorus (P) in China (mg/kg). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

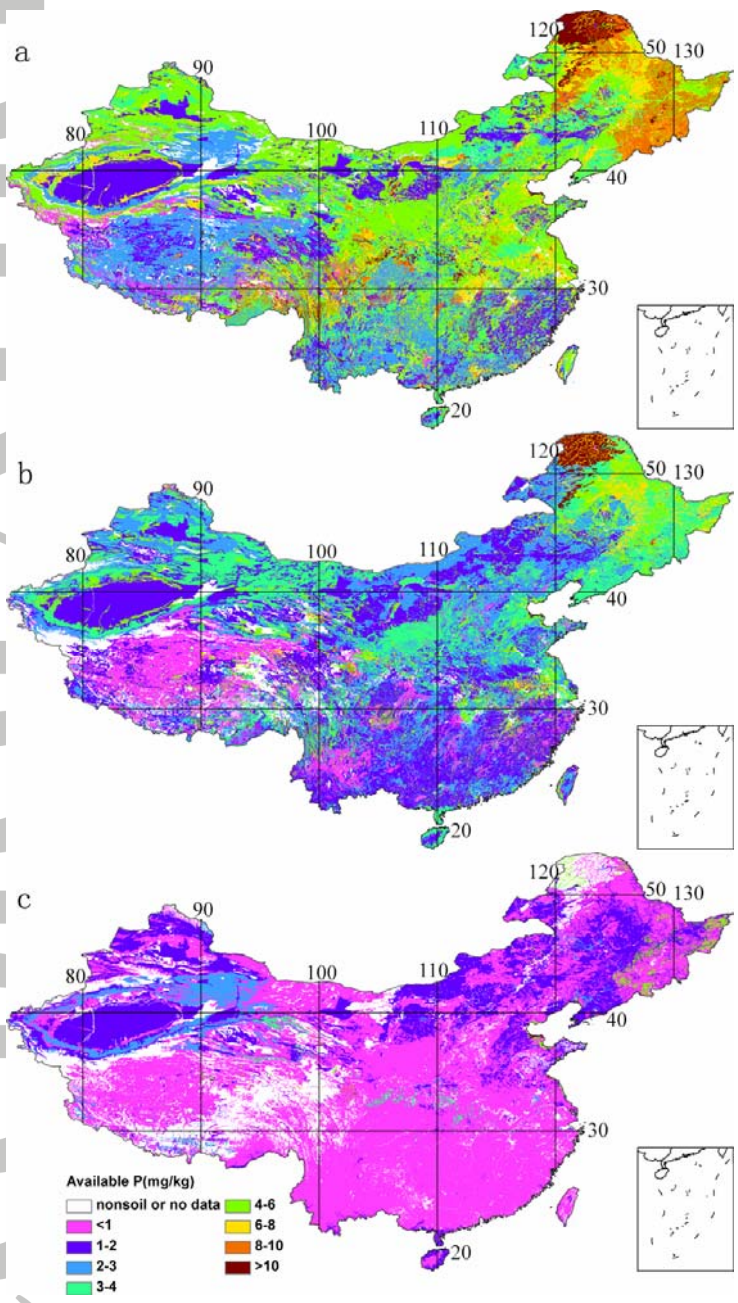


Figure S9. Spatial distribution of available potassium (K) in China (mg/kg). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

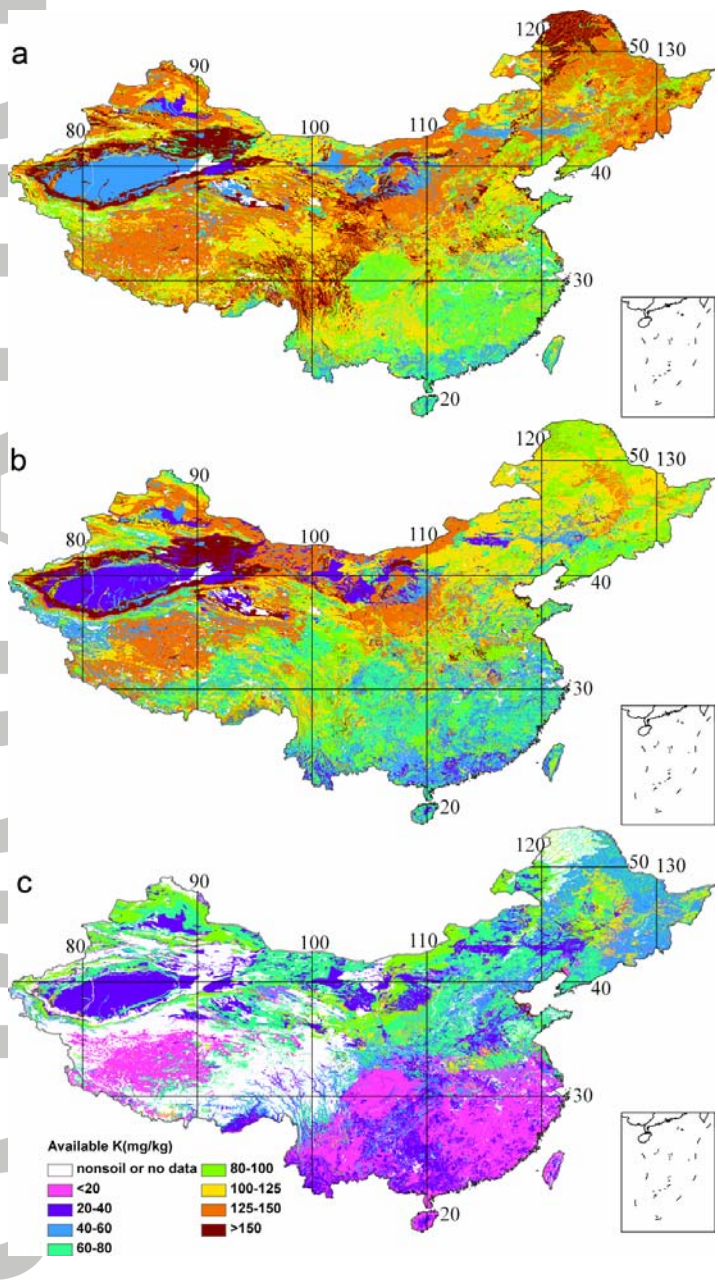


Figure S10. Spatial distribution of cation exchange capacity (CEC) in China (me/100 g). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

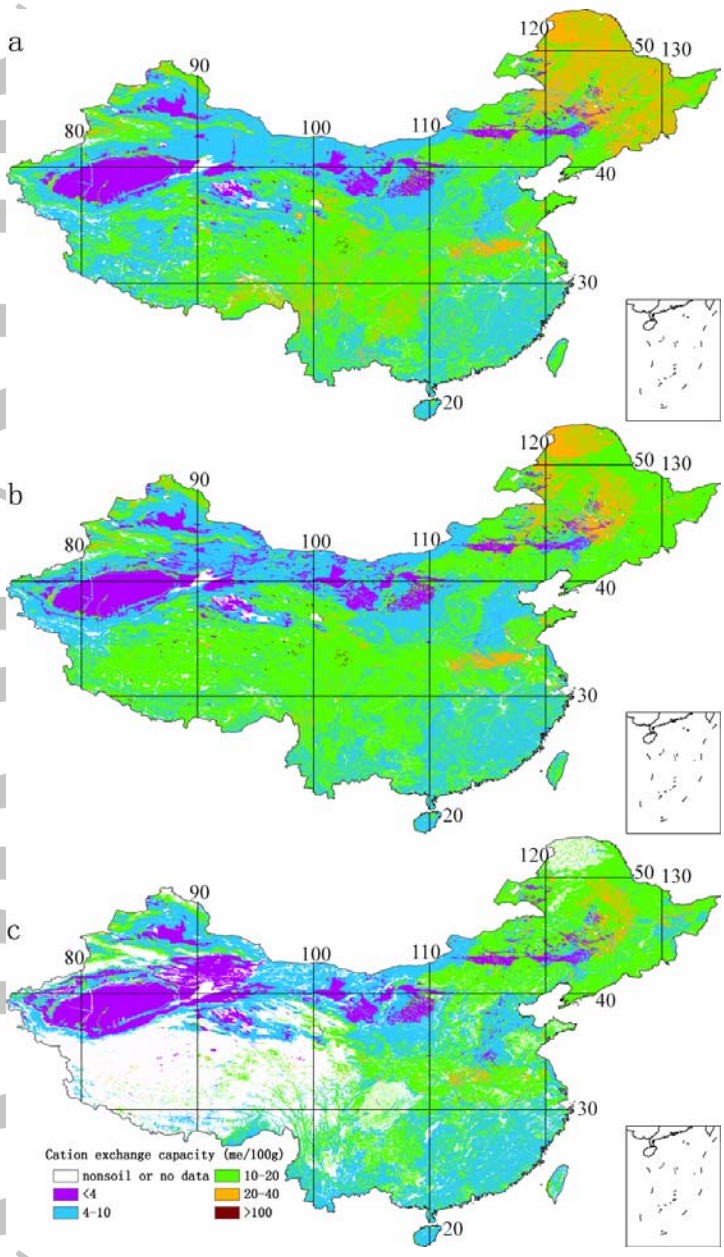


Figure S11. Spatial distribution of exchangeable H+ in China (me/100 g). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

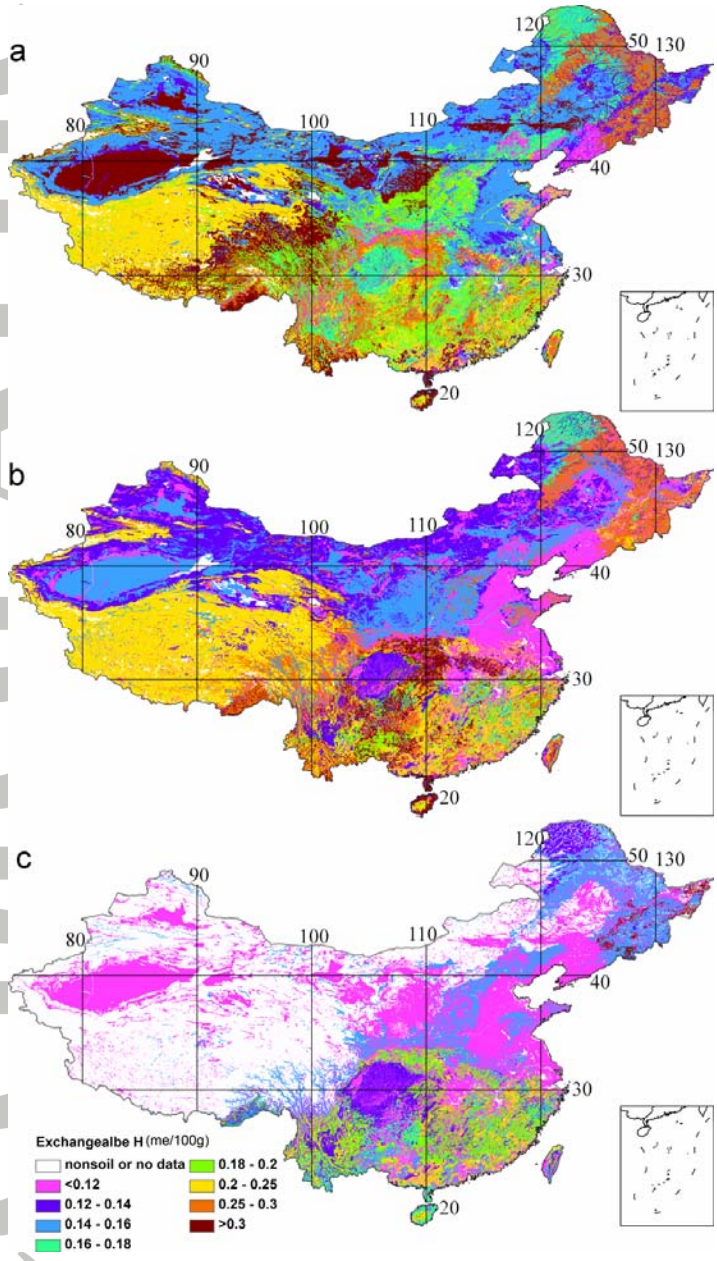


Figure S12. Spatial distribution of exchangeable Al3+ in China (me/100 g). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

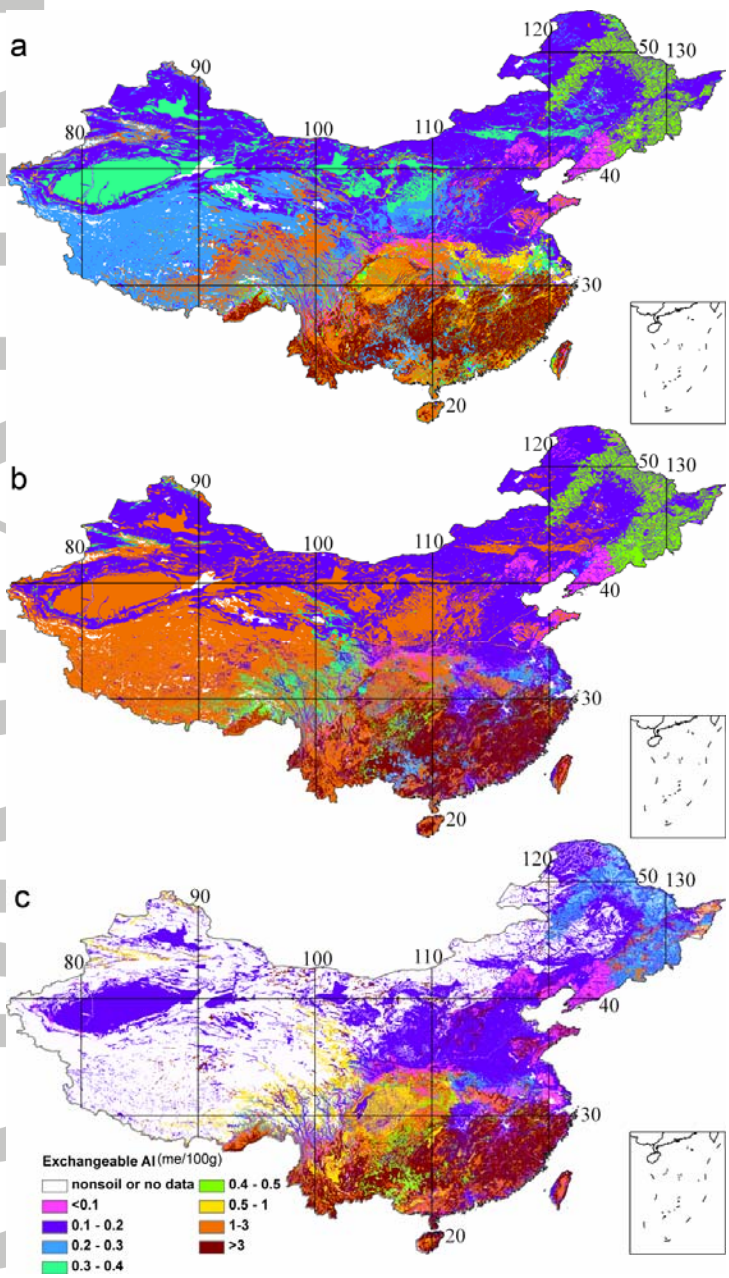


Figure S13. Spatial distribution of exchangeable Ca2+ in China (me/100 g). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

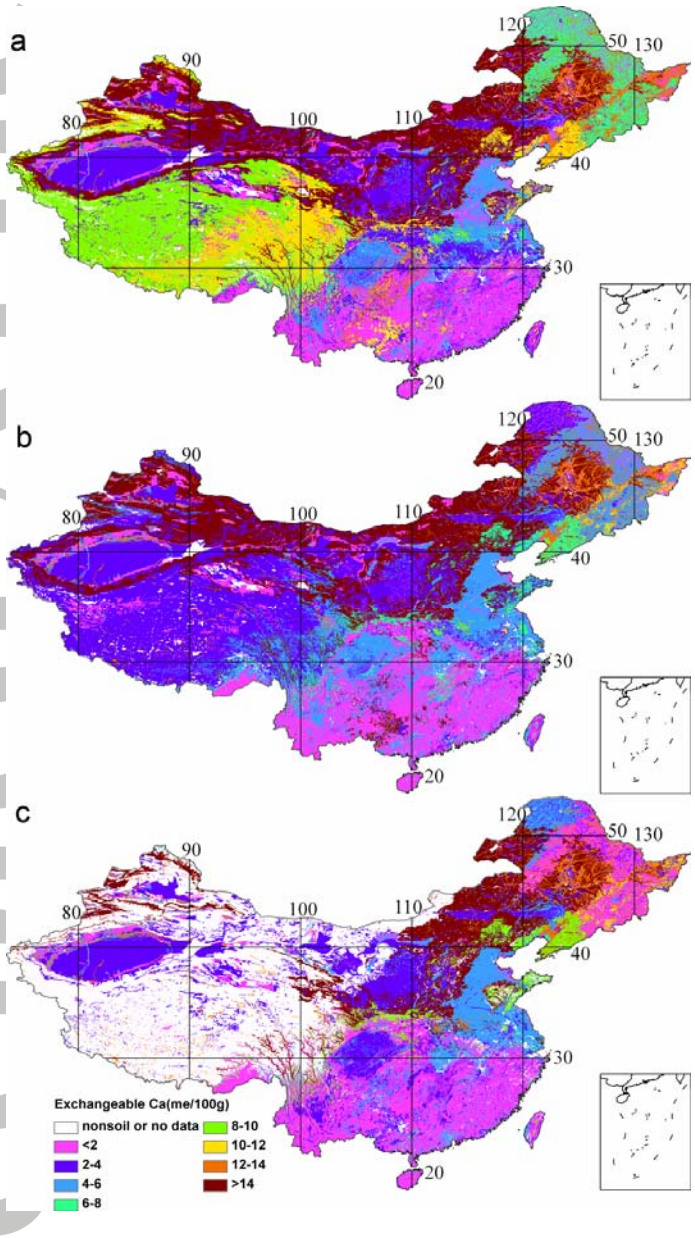


Figure S14. Spatial distribution of exchangeable Mg2+ in China (me/100 g). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

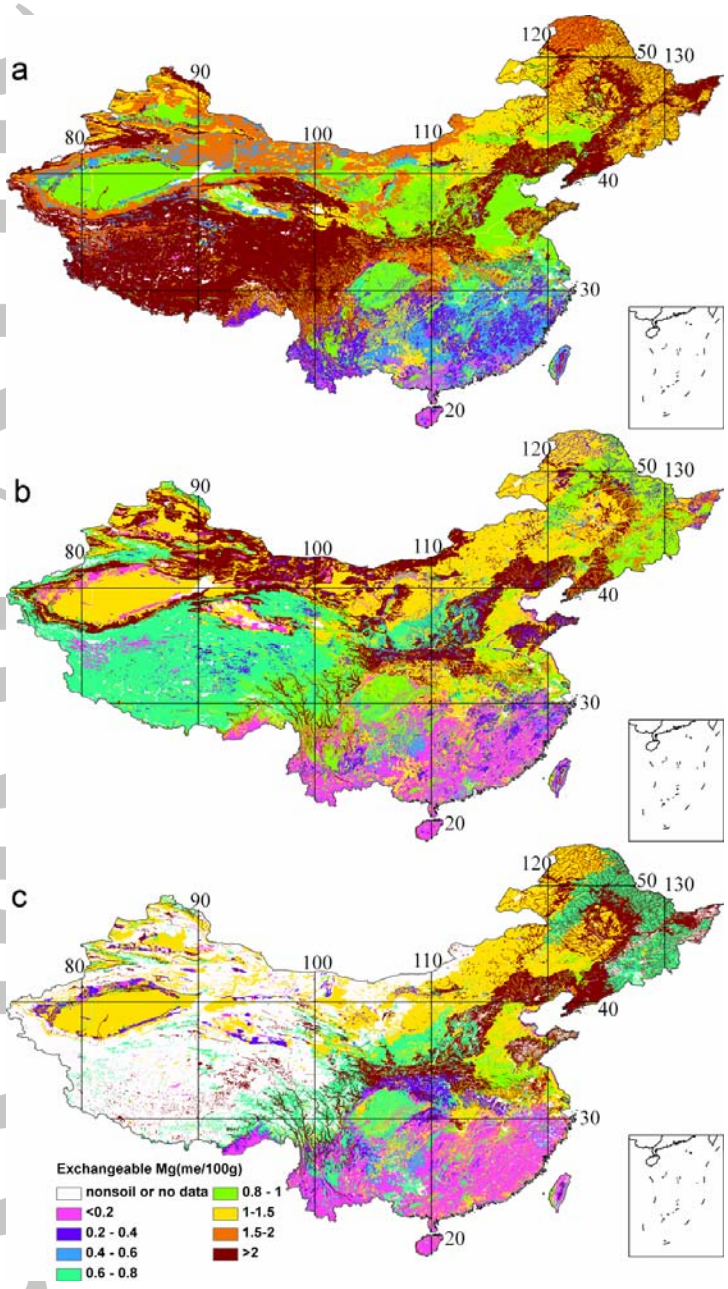


Figure S15. Spatial distribution of exchangeable K+ in China (me/100 g). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

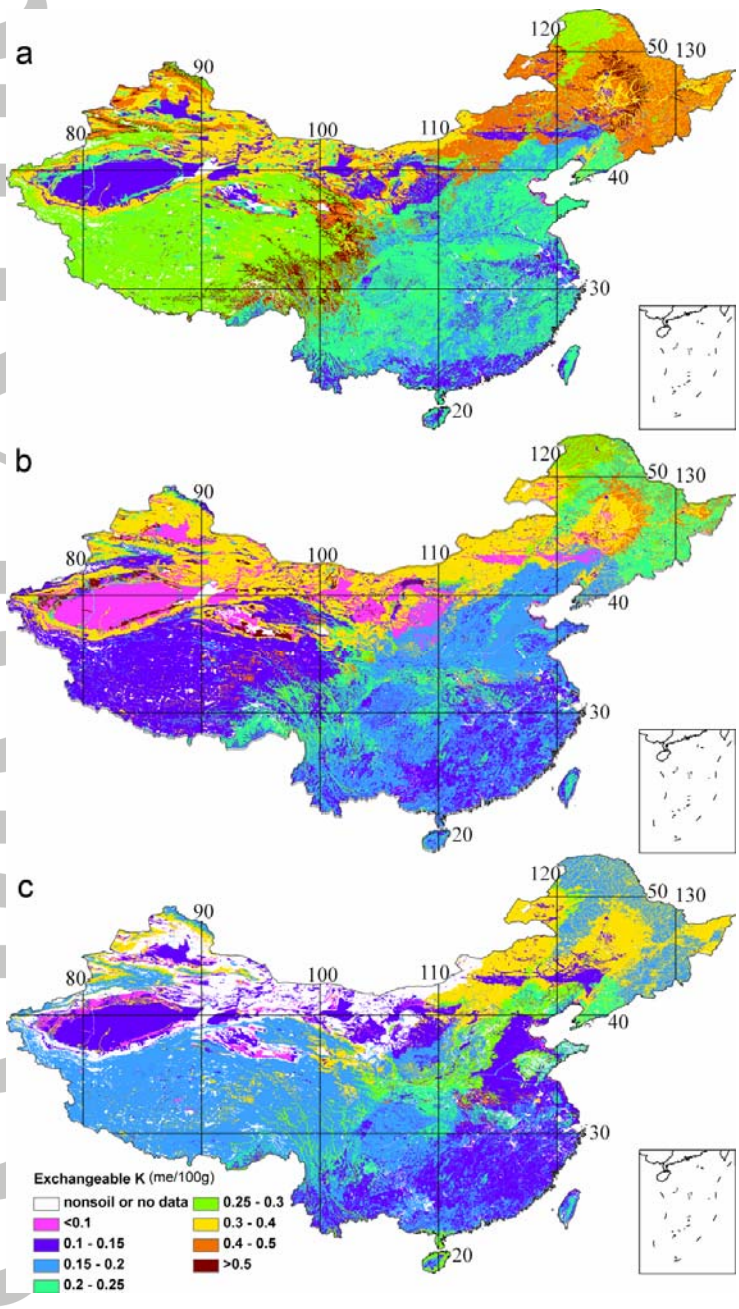


Figure S16. Spatial distribution of exchangeable Na+ in China (me/100 g). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

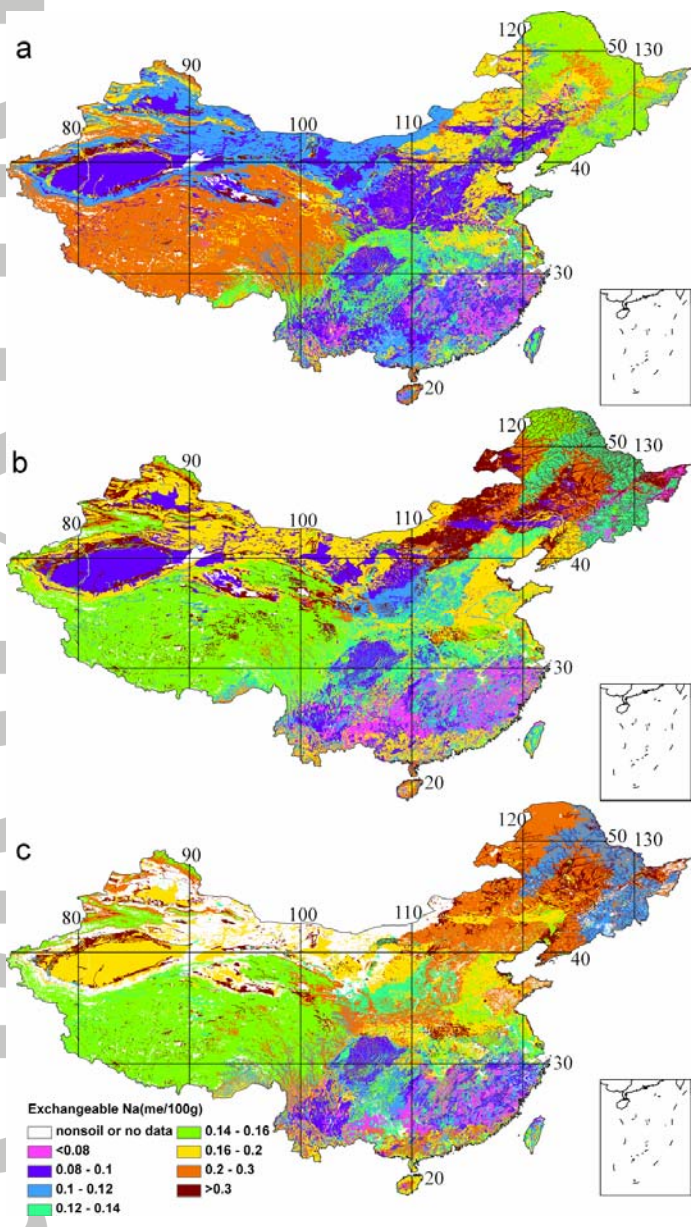


Figure S17. Spatial distribution of sand fraction in China (%). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

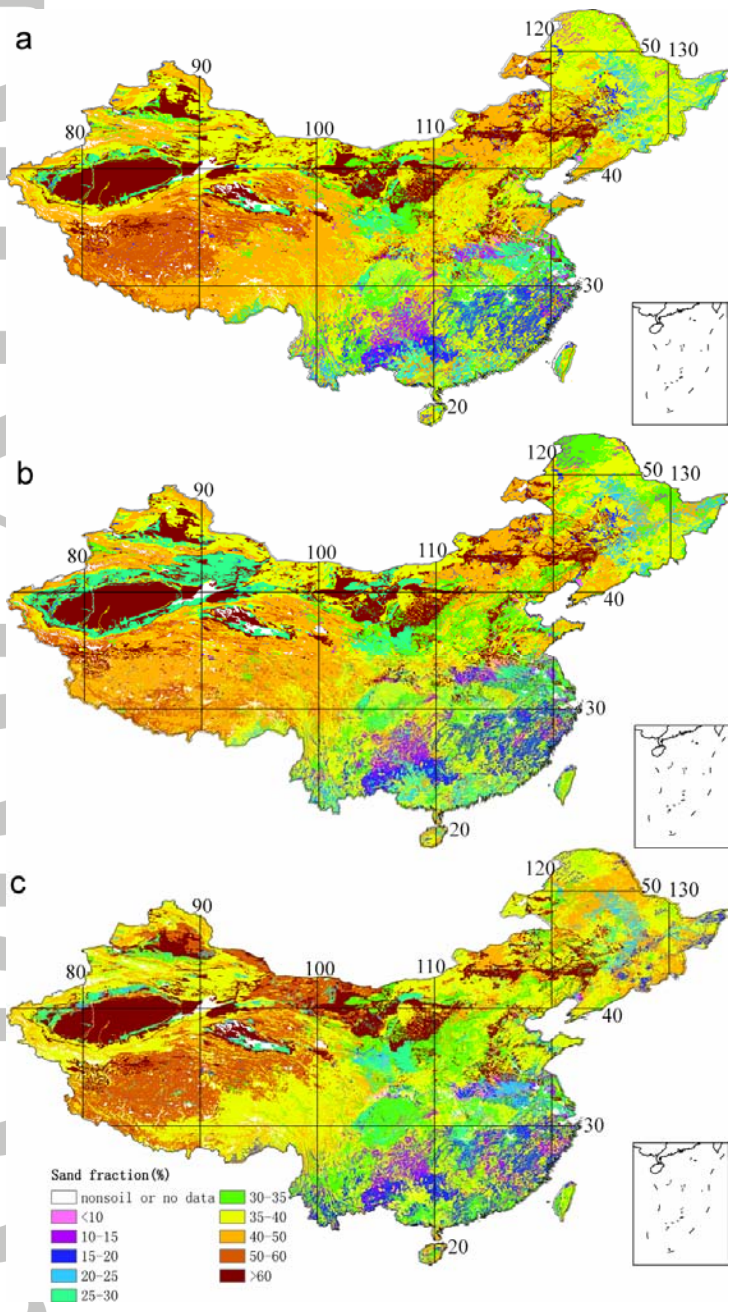


Figure S18. Spatial distribution of silt fraction in China (%). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

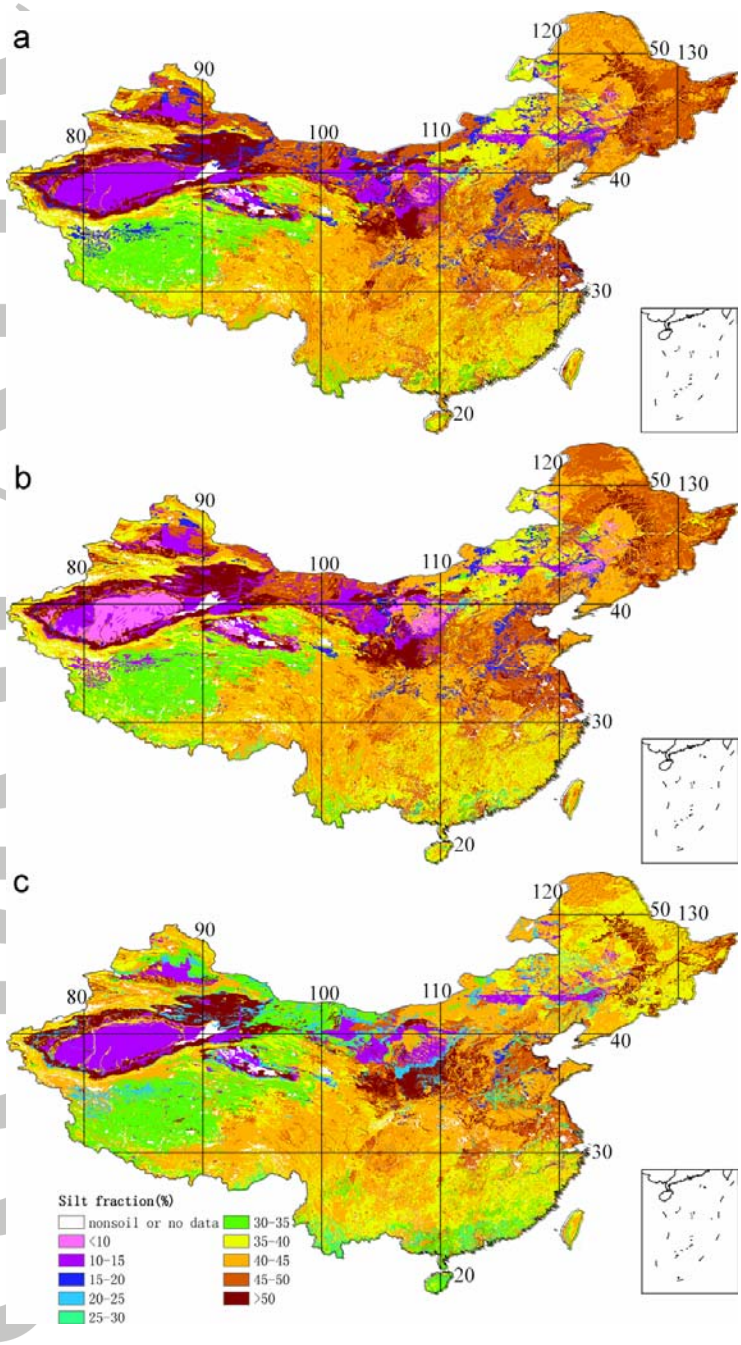


Figure S19. Spatial distribution of clay fraction in China (%). a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

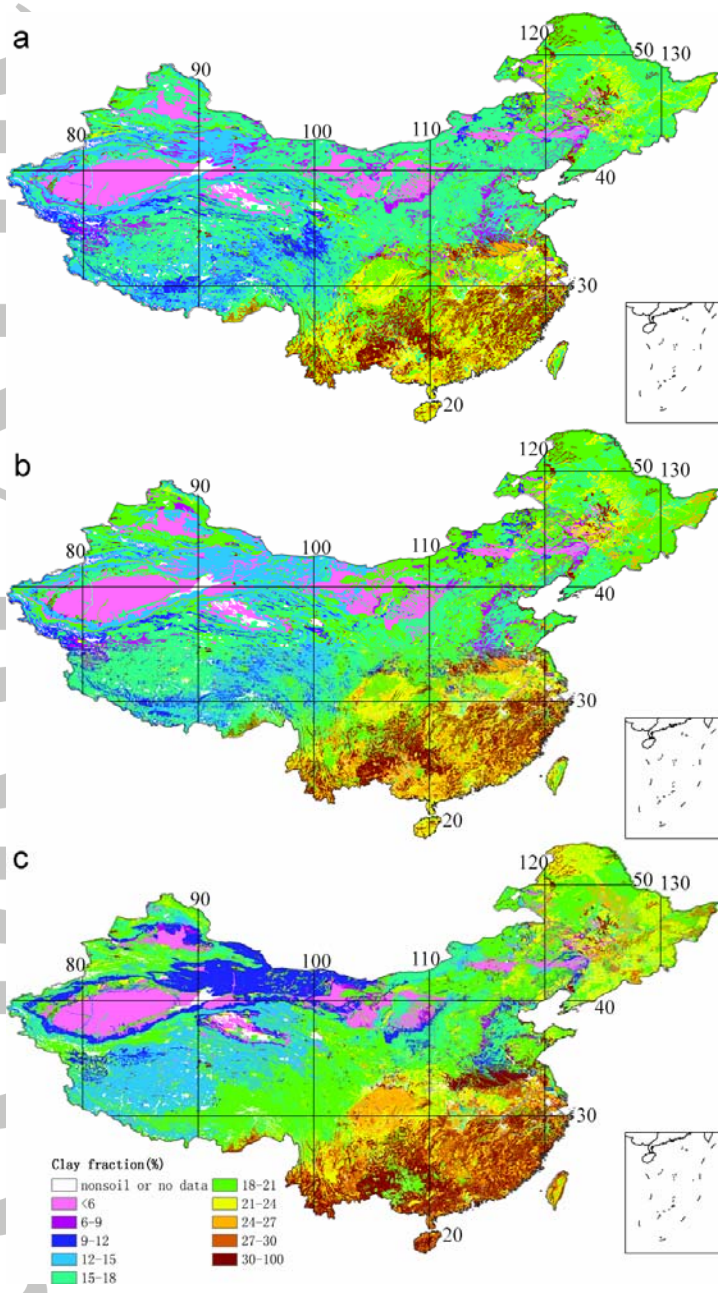


Figure S20. Spatial distribution of soil gravel content in China. a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

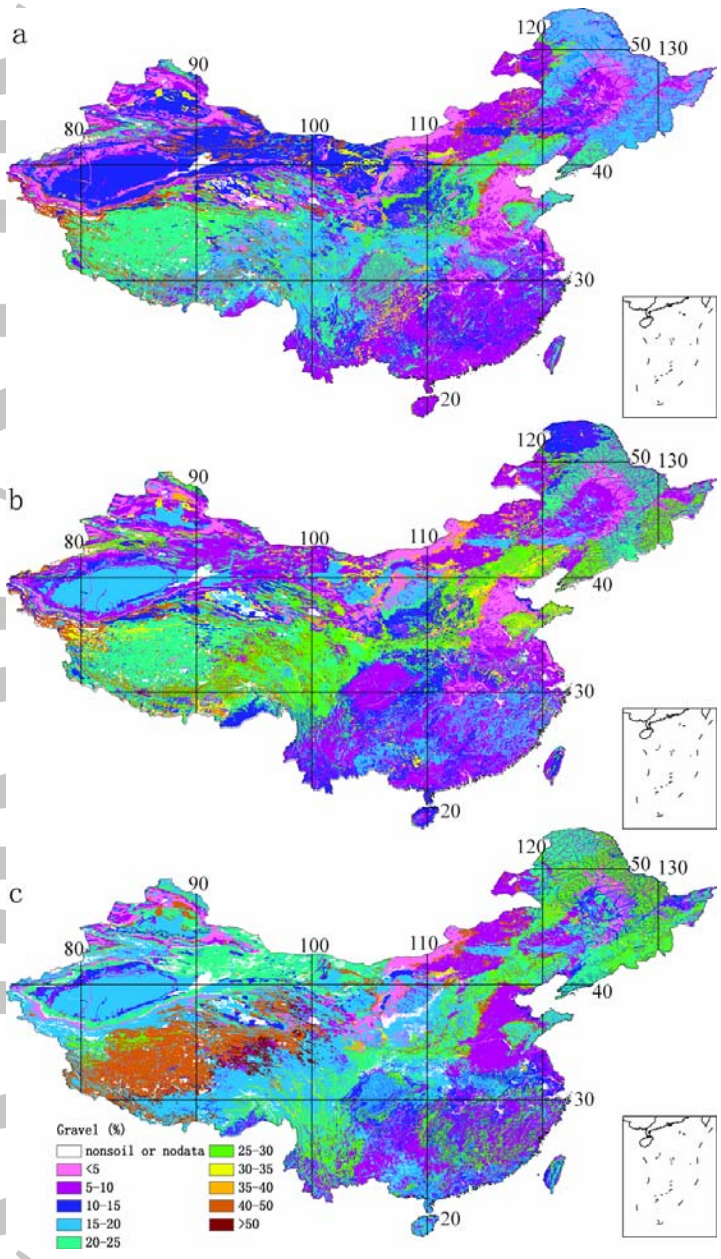


Figure S21. Spatial distribution of bulk density in China. a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm).

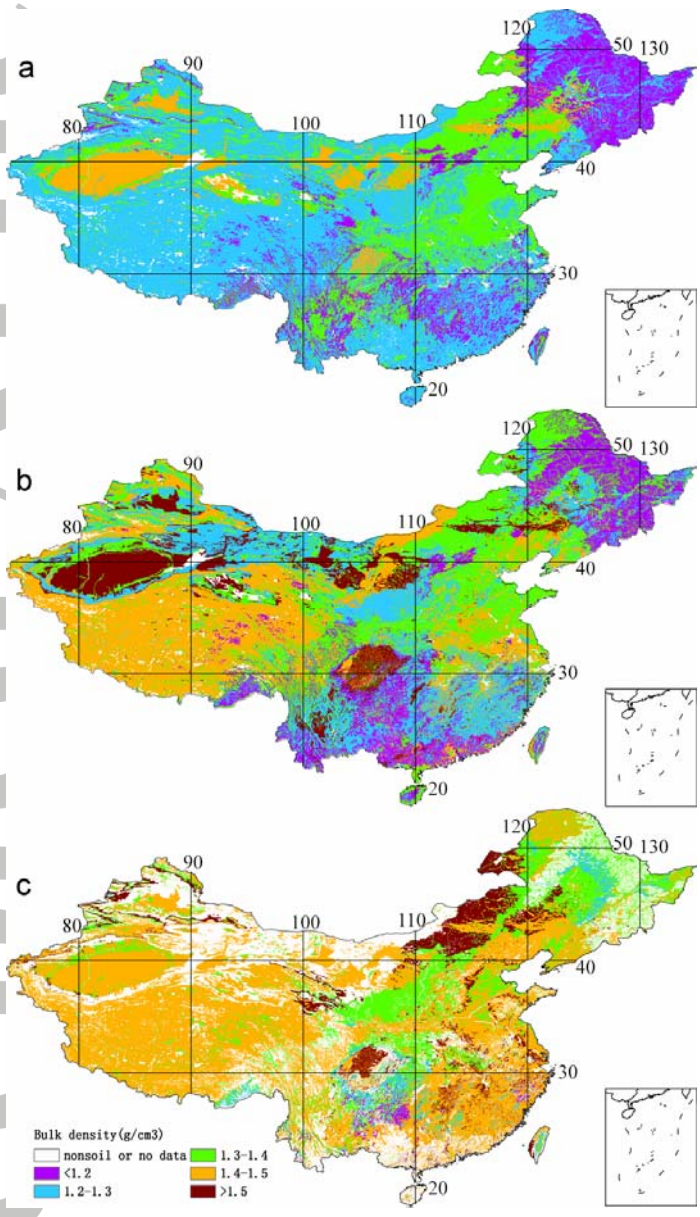


Figure S22. Spatial distribution of porosity in China (4.5-9.1 cm). a, porosity estimated directly; b, porosity calculated from bulk density and particle density; c, difference between the above two (b minus a).

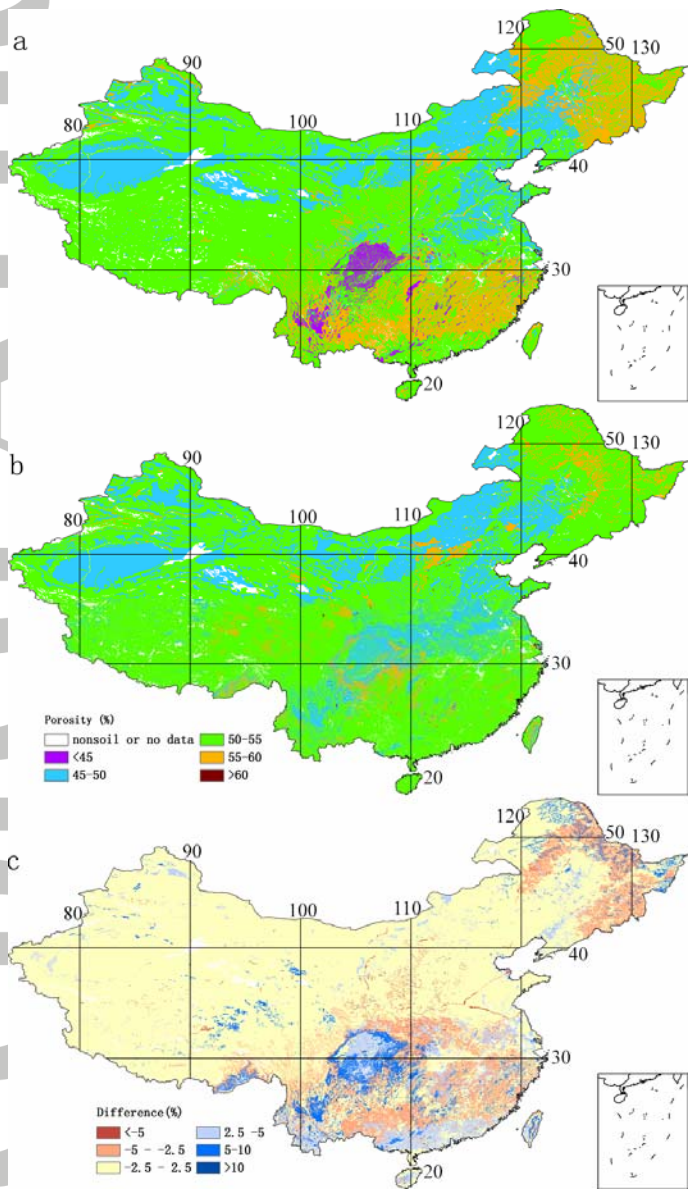


Figure S23. Spatial distribution of soil color of the surface layer in China. a, hue of wet soil; b, hue of dry soil; c, value of wet soil; d, value of dry soil; e, chrome of wet soil; f, chrome of dry soil. In hue, Y represents yellow, G represents green, and R represents red.

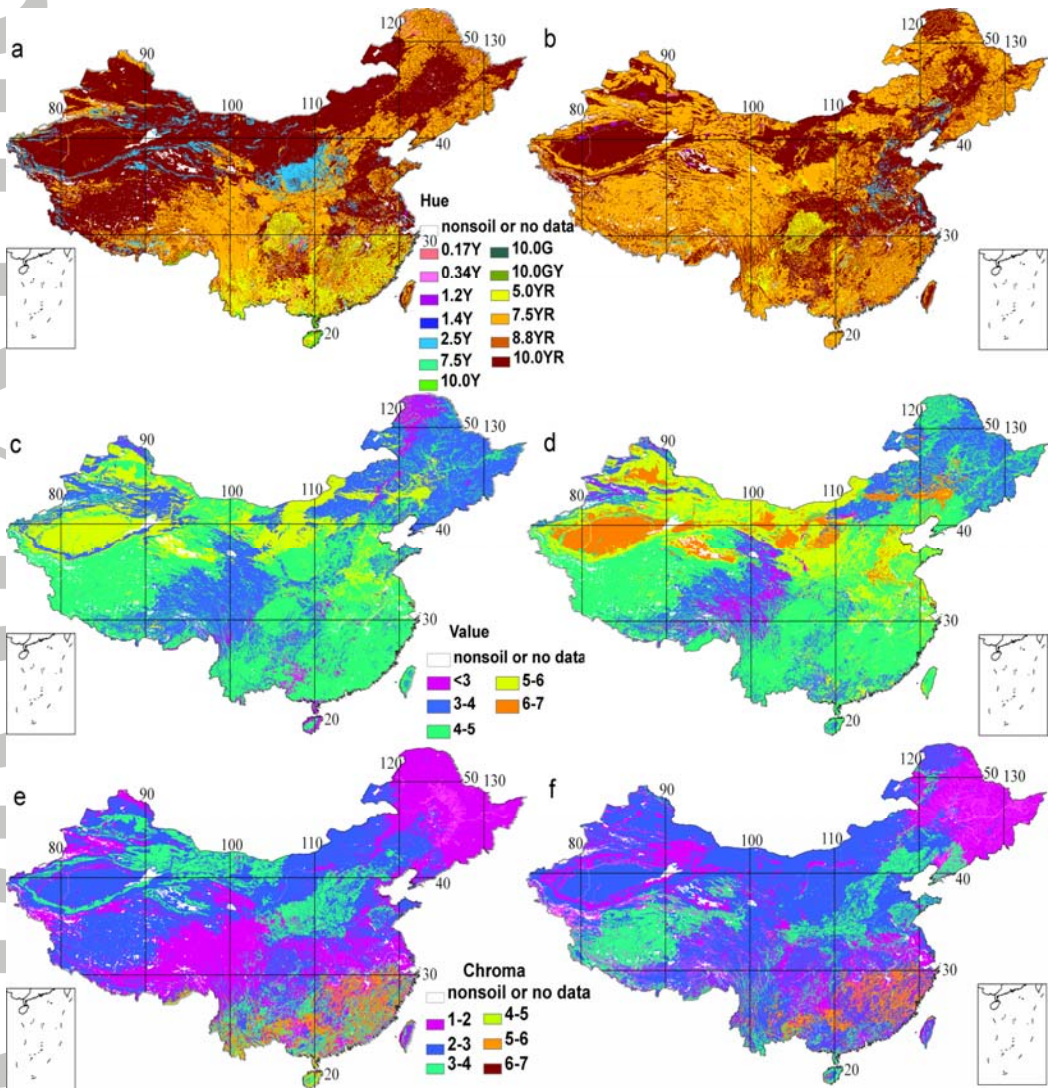


Figure S24. Spatial distribution of structure of soil (4.5-9.1 cm) in China. a, Dominant structure; b, Portion of Dominant structure; c, The second dominant structure; d, Portion of the second dominant structure.

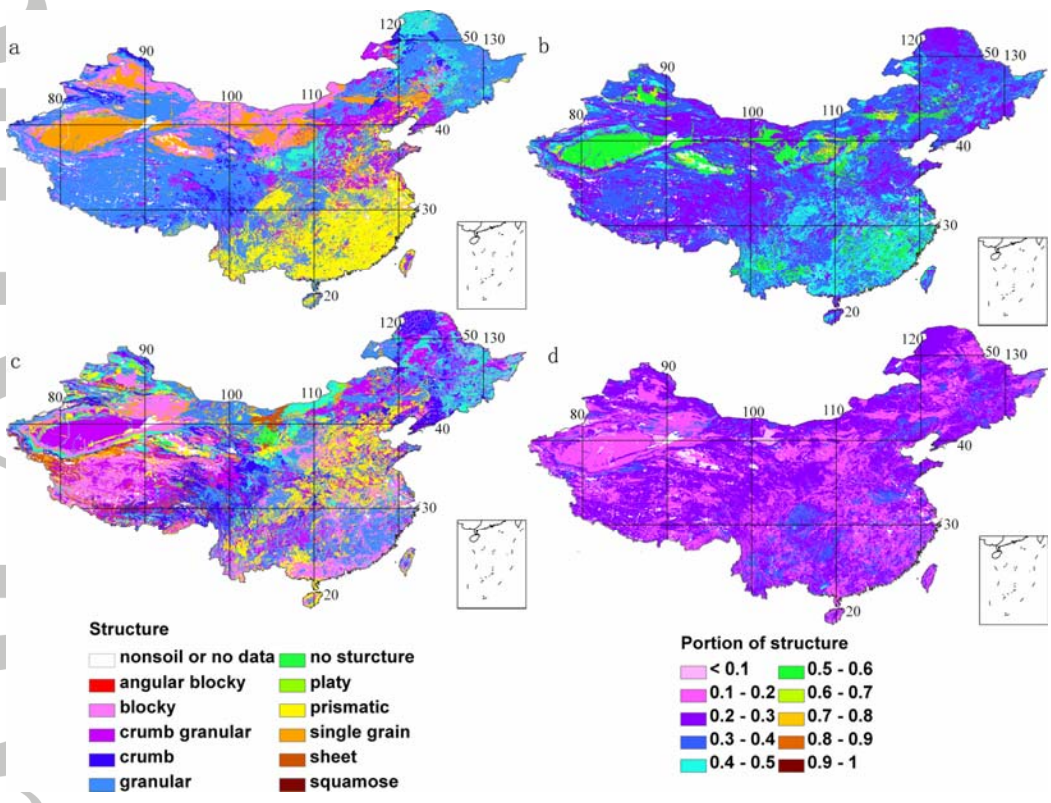


Figure S25. Spatial distribution of consistency of soil (4.5-9.1 cm) in China. a, Dominant consistency; b, Portion of Dominant consistency; c, The second dominant consistency; d, , Portion of the second dominant consistency.

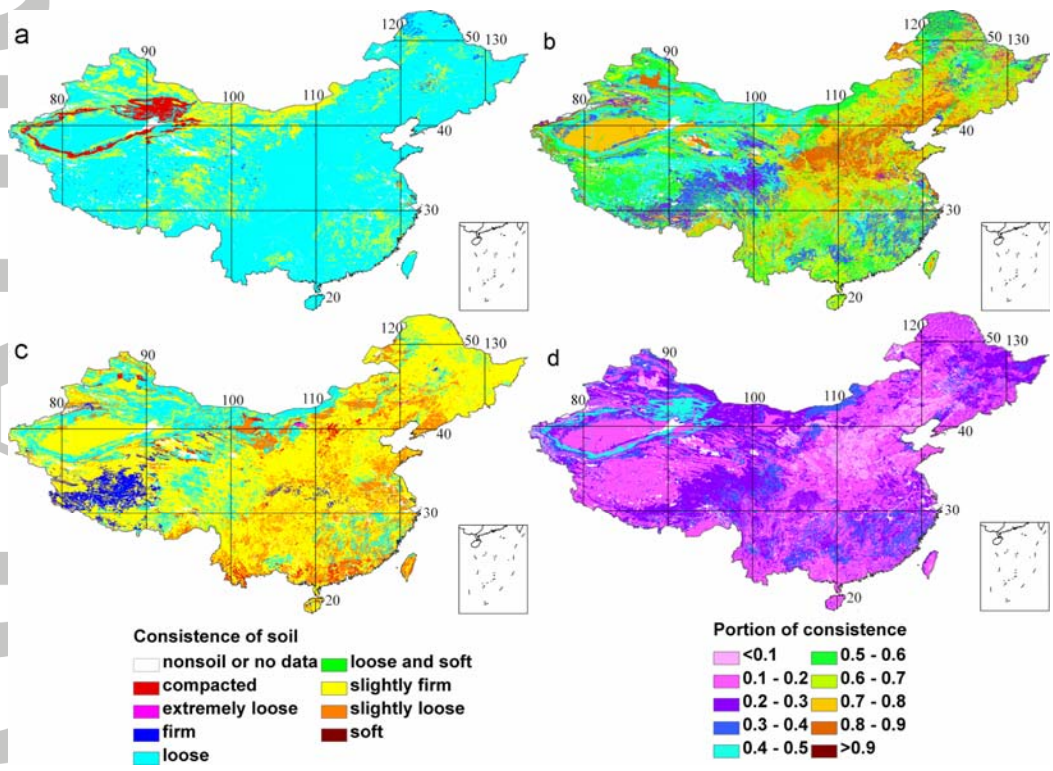


Figure S26. Spatial distribution of root quantity (abundance) in China. a, layer 2 (4.5-9.1 cm); b, layer 4 (16.6-28.9 cm); c, layer 6 (49.3-82.9 cm). The qualitative expression was converted into quantitative values, where 0 = none, 1 = very few, 2 = moderately few, 3 = few, 4 = moderate, 5 = many, 6 = great many and 7 = dense.