



# A soil particle-size distribution dataset for regional land and climate modelling in China

Wei Shangguan<sup>a,\*</sup>, Yongjiu Dai<sup>a</sup>, Baoyuan Liu<sup>b</sup>, Aizhong Ye<sup>a</sup>, Hua Yuan<sup>a</sup>

<sup>a</sup> State Key Laboratory of Land Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, 100875, Beijing, China

<sup>b</sup> School of Geography, Beijing Normal University, 100875, Beijing, China

## ARTICLE INFO

### Article history:

Received 28 April 2010

Received in revised form 13 January 2011

Accepted 19 January 2011

Available online 22 February 2011

### Keywords:

Soil particle-size distribution map

Map polygon linkage

Soil type linkage

Taxotransfer scheme

## ABSTRACT

We developed a multi-layer soil particle-size distribution dataset (sand, silt and clay content), based on USDA (United States Department of Agriculture) standard for regional land and climate modelling in China. The 1:1,000,000 scale soil map of China and 8595 soil profiles from the Second National Soil Survey served as the starting point for this work. We reclassified the inconsistent soil profiles into the proper soil type of the map as much as possible because the soil classification names of the map units and profiles were not quite the same. The sand, silt and clay maps were derived using the polygon linkage method, which linked soil profiles and map polygons considering the distance between them, the sample sizes of the profiles, and soil classification information. For comparison, a soil type linkage was also generated by linking the map units and soil profiles with the same soil type. The quality of the derived soil fractions was reliable. Overall, the map polygon linkage offered better results than the soil type linkage or the Harmonized World Soil Database. The dataset, with a 1-km resolution, can be applied to land and climate modelling at a regional scale.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Particle-size distribution (PSD) is a basic physical property of soils that affects many important soil attributes. The PSDs of soils have been widely used for estimating various soil hydraulic properties (Arya and Paris, 1981; Haverkamp and Parlange, 1986; Minasny and McBratney, 2007). The percentage of sand, silt and clay within a soil profile is frequently required to describe the physical processes in soil by land and climate models at regional and global scales (Dickinson et al., 1993; Dai, 2003; Sitch et al., 2003; Gassman et al., 2007). Despite the importance of having proper soil properties for use in these models, there is a dearth of spatial information on the physical and hydraulic properties of soil, especially for China. Webb et al. (1993) produced a global dataset for the top and bottom soil depths, with a 1° by 1° spatial resolution, that included the percentages of sand, silt and clay of individual soil horizons for 106 soil types by combining the Soil Map of the World of FAO-UNESCO (Food and Agriculture Organization of the United Nations/United Nations Educational, Scientific and Cultural Organization) with the World Soil Data File of Zabler (1986). Miller and White (1998) developed the CONUS-SOIL dataset, which includes the sand and clay fractions for 11 standard layers. Reynolds et al. (2000) produced the FAO-UNESCO global 5-minute distributions of the sand and clay fractions for two layers (0–30 and 30–100 cm).

Batjes (2006) derived soil properties for the 106 soil units shown on the Soil Map of the World for fixed depth intervals of 20 cm up to a depth of 100 cm. Dijkshoorn et al. (2008) developed a soil and terrain database at a scale of 1:1,000,000 for China with 1430 profiles (Zhang and Zhao, 2008). While FAO et al. (2009) used the 1:100,000 scale soil map of China and soil profiles from the World Inventory of Soil Emission Potential (WISE), which included only 61 profiles from China, to develop the Harmonized World Soil Database (HWSD), they indicated a need for more soil profiles from China. The existing soil datasets are based on limited profile data and a coarse resolution of spatial data; therefore, they cannot satisfy the requirements of regional modelling for China. Thus, it remains crucial to update and expand soil PSD databases that are specifically designed for modelling applications.

The goal of this study is to develop a practical 1-km resolution dataset of particle-size distribution of soil for China that is suitable for regional land and climate modelling.

## 2. Data and methods

### 2.1. Data source and preparation

The 1:1,000,000 soil map of China was compiled by the Institute of Soil Science, Chinese Academy of Sciences (Shi et al., 2004) based on the results of the Second National Soil Survey of China. This map is the most detailed soil map in China at the national scale. It is classified using the Genetic Soil Classification of China (GSCC), which includes 12 orders, 61 great groups, 235 sub-great groups, and 909 families.

\* Corresponding author at: College of Global Change and Earth System Science, Beijing Normal University, No. 19, Xijiekouwai St., Beijing 100875, China. Fax: +86 10 6220 5274.

E-mail address: [shanggv@hotmail.com](mailto:shanggv@hotmail.com) (W. Shangguan).

There are 94,303 map polygons in the map, including 85,257 soil map polygons and 9046 non-soil map polygons. More than half of the soil map polygons are at the sub-great group level, and the others are at the great group or family level (Table 1). The latitude and longitude of the centres of the map polygons were extracted from the coverage file using GIS tools.

The soil profiles were from the Chinese soil profile database, which was also established using the results of the Second National Soil Survey of China conducted in the 1980s. It contains data for 33,039 soil layers representing 8979 profiles. The data were published by the National Soil Survey Office (1993a, b, 1994, 1995a, b, 1996), provincial soil survey offices and the soil survey offices of some Tibetan counties. However, PSD data are not always available for each layer. Thus, the number of samples varies with soil type and depth. Fine size fractions were determined using the hydrometre or pipette method, whereas coarse size fractions were obtained through sieving (National Soil Survey Office, 1992). Particle-size fraction data were classified by several schemes including the ISSS (International Society of Soil Science) and Katschinski's schemes. For modelling purposes, these particle-size data were converted to the FAO-USDA (United States Department of Agriculture) System (Shangguan and Dai, 2009; Shangguan and Dai, 2010).

The latitude and longitude of the soil profiles were derived at different levels of spatial precision from their geographic location descriptions. The spatial precision of profile locations was broken down into three classifications: A, B and C, which had errors below 15 km, between 15 km and 60 km, and above 60 km, respectively.

The soil classification system used for the aforementioned soil profiles and soil map was the GSCC. However, there are some inconsistencies: there were profiles with a classification at a specific soil type level (e.g., soil family) that had no corresponding map unit of the same type level, and vice versa; different names were used for the same soil type. The inconsistent soil profiles were reclassified into the proper soil type from the soil map at different soil type level (i.e., great group, sub-great group or family). Basically, soil type names were modified in light of the principle of approximation of naming and use of bynames for soils. For example, 'ploughed diluvium sandy thin meadow soil' was modified as 'sandy thin meadow soil'. For soil map units with parent material information, the parent material of the corresponding soil profile was also used to modify the soil type names.

The PSD data were interpolated to 2 and 11 standard layers by a depth-weighted method for their convenience of use in land and climate models (Reynolds et al., 2000). Many grid-based models are designed as equal-compartment layers (Dickinson et al., 1993; Dai, 2003). However, the great range and diversity of soil profile layer thicknesses make them inconvenient to use in these models without additional analyses. The 2 layers were the topsoil (0–30 cm) and subsoil (30–100 cm), and the 11 layers were the same as the CONUS-SOIL dataset standard, which retains a better vertical variation (Miller and White, 1998). For brevity, only the 2-layer dataset is shown in this paper because it is easier to compare with other datasets.

Soil profiles without PSD data or consistent soil classification information were excluded, leaving 8595 soil profiles from which to derive sand, silt and clay maps by a linkage method.

**Table 1**  
Numbers of different soil type level of linkage at different soil type level of map polygons.

| Soil type level of map polygons | Soil type level of linkage |                 |             |          |
|---------------------------------|----------------------------|-----------------|-------------|----------|
|                                 | Family                     | Sub-great group | Great group | Subtotal |
| Family                          | 20,161                     | 4575            | 5           | 24,741   |
| Sub-great group                 | –                          | 42,815          | 489         | 43,304   |
| Great group                     | –                          | –               | 13,812      | 13,812   |
| Subtotal                        | 20,161                     | 47,390          | 14,306      | 81,857   |

## 2.2. Methods

### 2.2.1. Existing linkage method and problems

In previous studies, the linkage method has usually been accomplished by linking soil map units and profiles following the so-called taxotransfer rules (Reynolds et al., 2000; Batjes, 2003; FAO et al., 2009). We called this method the soil type linkage. Soil type linkage gave soil parameter estimates by soil units for each soil layer, usually with reference to a soil's textural class. The variation in soil properties across different map polygons of the same soil type that actually existed was not considered.

### 2.2.2. Polygon linkage method

**2.2.2.1. Basic Idea.** In this study, a method of linking soil profiles to individual polygons instead of map units was developed. In order to preserve the spatial variation in soil properties as much as possible, two aspects other than the soil classification information were taken into account (i.e., the sample sizes of profiles and the distances between soil polygons and profiles). The distance was used to determine the order of priority of soil profiles to link to a map polygon. The likelihood of linkage decreased as distance increased, and the variation of soil properties among polygons of the same soil type was retained. The possible effects of regional variation in environmental factors (e.g., climate and vegetation) were also implicitly considered.

To represent a map polygon, a minimum sample size of soil was needed. Scholes et al. (1995) insisted that at least 30 pedons per soil unit were necessary to provide adequate representation for the 106 soil units of the Soil Map of the World at a 1:5,000,000 scale. For the Soils and Terrain Database (SOTER), each soil component of a map unit was characterised by a typical soil profile (Batjes et al., 2007). In order to fill in the gaps in the primary SOTER database, Batjes (2003) performed a taxotransfer scheme using the median of more than 5 profiles for the considered combination of FAO soil unit (or soil grouping), attribute, depth zone and soil texture class. Map units with higher soil type levels need more samples to cover the variation in soil properties. In this study, we aimed for at least 40 profiles for a great group, 10 profiles for a sub-great group and 3 profiles for a soil family.

**2.2.2.2. Linkage process.** First, as a reference of polygon linkage, the Euler distance between a map polygon and a soil profile with the same soil type was calculated. Soil profiles without sufficient precision for their location were not involved in the selection of linkages.

The map polygon linkage process was performed as follows:

1. Soil profiles of the same soil type for each map polygon were searched at a 15-km radius. We assumed that all profiles in this range (about the county size) should be used to represent a map polygon. In addition, the profiles in this radius were likely to be within or near the linked map polygon, as the average size of map polygons was about 10 km by 10 km. If there were enough soil profiles of the same soil type in this radius, these profiles were linked to the map polygon. Otherwise, we continued to step 2. Soil profiles with spatial precisions B and C were not involved in this step.
2. The search radius was enlarged until it was greater than the whole soil map of China or the target number of soil profiles was reached. If the target number of soil profiles was reached, these soil profiles were linked to the map polygon and the final radius was recorded. If the search resulted in insufficient soil profiles, these soil profiles were also linked to the map polygon and marked as insufficiently represented. If there was no soil profile in the whole soil map, we proceeded to step 3.
3. The grouping level of soil type was expanded and the search was restarted at step 1. For example, if there was no profile for a map polygon at the family level, the search was restarted at the sub-great group level. The linkage started at the lowest soil type level of a map polygon and continued upwards to the great group level, with different linkages stopping at different levels.

**2.2.2.3. Obtaining a representative value.** The medians, means, ranges, variances and sample sizes for the sand, silt and clay contents of the linked soil profiles were calculated both for the topsoil (0–30 cm) and subsoil (30–100 cm). The maps of sand, silt and clay content were derived by using the median value for each soil polygon (Batjes, 2006) because the influence of extreme values is partially ignored compared to a mean value. The PSD computed based on medians rarely summed up to 100%. To guarantee that the sum of three fractions totalled 100%, the following process was adopted. First, the median of each fraction for linked soil profiles was calculated. Then, the sum of squared deviation (*SSD*) of the medians was calculated for each linked profile based on the following formula:

$$SSD_i = (s_{a_i} - s_{a_m})^2 + (s_{i_i} - s_{i_m})^2 + (s_{c_i} - s_{c_m})^2 \quad (1)$$

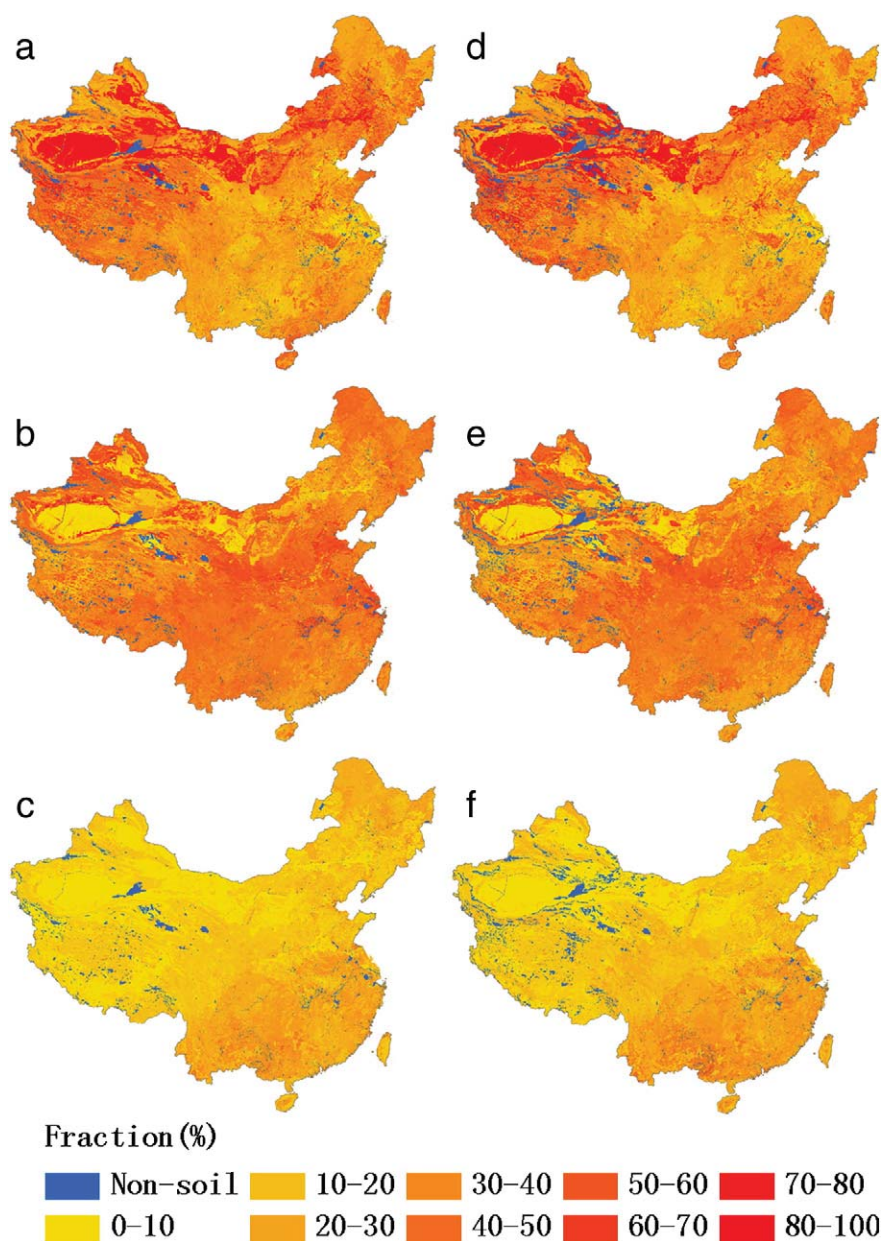
where  $s_{a_i}$ ,  $s_{i_i}$  and  $s_{c_i}$  are the sand, silt and clay fraction of the  $i$ th linked profile, respectively, and  $s_{a_m}$ ,  $s_{i_m}$  and  $s_{c_m}$  are the medians of the sand, silt

and clay fractions for the linked soil profiles, respectively. Finally, the profile with the minimum *SSD* was used to represent the soil polygon.

For grid-based model applications, the vector-format data were subsequently rasterised to spaced grids at a resolution of about 1-km (30 arc seconds by 30 arc seconds) for sand, silt and clay. There were non-soil map polygons (organic materials, water, rocks or other) and layers containing bedrock. As a result, the sum of the computed sand, silt and clay fractions was often less than 100% when rasterisation was done. The sand, silt and clay fractions were normalised to 100% (before rounding) if the sum of the fractions was less than 100% and greater than 50%. Otherwise, the fractions were set to zero (This normalisation may cause some false information to be included; Miller and White, 1998).

### 2.3. Validation and comparison

To evaluate and validate the results of the linkage method, an independent dataset was used. The data were collected from three areas in 2008 and 2009. There were 168, 163 and 58 samples from the



**Fig. 1.** Sand, silt, and clay fractions of China. (a) Sand fraction of the topsoil (0–30 cm). (b) Silt fraction of the topsoil (0–30 cm). (c) Clay fraction of the topsoil (0–30 cm). (d) Sand fraction of the subsoil (30–100 cm). (e) Silt fraction of the subsoil (30–100 cm). (f) Clay fraction of the subsoil (30–100 cm).



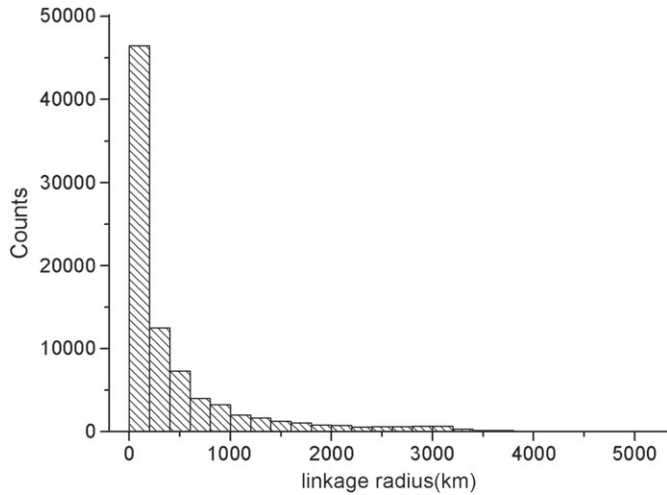


Fig. 2. The distribution of linkage radius between map polygons and soil profiles.

Bingxian county of Heilongjian province (3834 km<sup>2</sup>), Ansai county of Shaanxi province (3607 km<sup>2</sup>, including parts of the neighbouring counties) and Zitong county of Sichuan province (1435 km<sup>2</sup>), respectively. The samples were taken as a 5-km grid for the top soil layer (0–20 cm). The fine size fractions were determined using the hydrometre method, whereas the coarse size fractions were obtained through sieving. Bingxian is dominated by black soil (which is a black-coloured soil containing a high percentage of humus and high percentages of phosphoric acids, phosphorus and ammonia, corresponding to Phaeozems in World Reference Base for soil resources (WRB)), meadow soil (which contains a high percentage of humus with a high groundwater level and meadow vegetation, corresponding to Cambisols in WRB) and dark brown soil (which is a dark brown-coloured soil containing a high percentage of humus with vegetation of coniferous and broad-leaved mixed forest, corresponding to Cambisols in WRB). Ansai is dominated by loessial soil (which has apparent characteristics of parent material of loess, corresponding to Cambisols in WRB), and Zitong is dominated by purplish soil (which is developed from purplish shale and sandstone, and at the early stage of eluviations, corresponding to Cambisols in WRB). Though a soil great group in GSCC could be interpreted into several WRB soil groups (Shi et al., 2010), only the dominant one were given here. The cross-reference was also developed to relate GSCC with Soil Taxonomy of US and Chinese Soil Taxonomy (Shi et al., 2006a, b).

The quality of the linkage was evaluated based on the search radius, soil type level of the linkage and sample size. If the search stopped at a small radius, it is implied that the linked profiles are close

to the map polygons and offer good estimates. If the soil type level of linkage is low (such as soil family), the variation in soil properties is lower than at higher levels of soil type. If the target for sample size is reached, the polygon can be considered well represented.

For comparison, the soil type linkage method was also performed, following the methods of previous studies in which map units and profiles with the same soil classification information were linked (Reynolds et al., 2000; Batjes, 2003; FAO et al., 2009).

The results derived through the soil type and polygon linkage methods were compared with the independent samples using mean error (ME) and root mean square error (RMSE). The Harmonized World Soil Database (HWSD), which was derived by linking soil map units and profiles from WISE, was also compared with our results.

### 3. Results and discussion

Fig. 1 shows the sand, silt and clay fractions linked by map polygons. The fraction maps display soil PSD distribution for China in great detail. Generally, north and west China have high sand fractions and low clay fractions, especially in the desert area, while the opposite was observed in south China. This is expected due to the physical and chemical weathering processes in different parts of China. The non-soil map units were assigned zero values for all three fractions.

The quality of the derived PSD dataset was assessed based on the level of soil type linkage, sample size and search radius. The linkage levels of map polygons are shown in Table 1. Most of the soil map polygons were linked at the same soil type level they belong to, which indicates that most of soil types had corresponding soil profiles. The linkage level was recorded for each soil map polygon for future reference. Lower level linkages had better estimates for PSDs. The target sample size was achieved, except for in 9.2%, 7.8% and 1.2% of the linkages at the levels of soil family, soil sub-group and soil group, respectively. The map polygons that were not linked at the same soil type level that they belong to or did not reach the target sample size need more profile samples examined in the future. On the other hand, the sample size was maintained at the target number to describe the variety of different map polygons of the same soil type, though Scholes et al. (1995) chose not to exclude additional soil profiles on the basis that a particular soil type was already well represented. Fig. 2 shows the counts of different linkage radius between map polygons and soil profiles. The median value was about 146 km and the 75th percentile was about 520 km, which indicates that most linkage happens at the climate zone scale. Although natural similarity and variety was mainly considered within the context of the soil map itself in previous studies (Webb et al., 1993; Reynolds et al., 2000; Batjes, 2006), the distance-

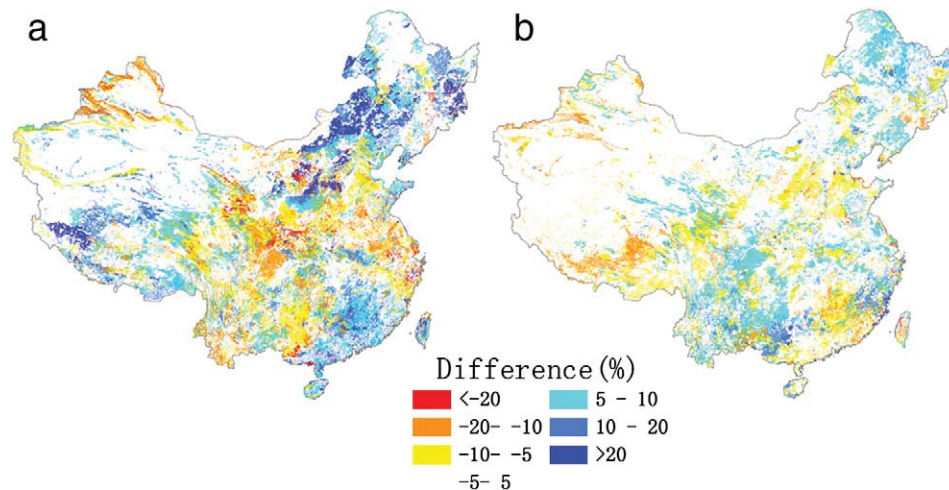


Fig. 3. Differences in soil fraction contents linked by map polygons and linked by soil types. (a) Sand fraction of the topsoil (0–30 cm). (b) Clay fraction of the topsoil (0–30 cm).

based linking method, which considers the distance between soil polygons and profiles, better presents these properties.

Fig. 3 was obtained by subtracting the sand and clay content derived by soil types from those derived by map polygons. The contents of these soil fractions were different in most of the map polygons. For the soil type linkage, the sand content in the north and southeast and the clay content in the southwest were underestimated compared to those of the map polygon linkage. The differences confirm that it is better to derive soil properties through a linkage method that considers the distance between profiles and map polygons to present the variation in soil properties of different polygons with the same soil type. When detailed datasets are available, the map polygon linkage method offers more spatial information for soil fractions. If there are not enough soil profiles, the map polygon linkage and soil type linkage will not appear to be significantly different, as the polygon linkage will not stop until it is outside of the whole soil map area in step 2. It is also reasonable that the soil texture for the same soil type varies within a certain range in different locations, which usually means there is a difference in soil formation factors, i.e., climate, organisms (including humans), relief, parent material and time (Jenny, 1941), particularly for soils grouped together at a high classification level.

Table 2 shows the ME and RMSE of soil fraction contents for the three datasets: HWSD, the polygon linkage dataset and the type linkage dataset. Overall, the polygon linkage method gave the most accurate estimation and HWSD gave the least accurate estimation, with the exception of the polygon linkage performing slightly worse than the other two datasets for estimating clay contents. Within the limits of our data, there were some differences in the PSDs of the top soil depth of samples collected for validation and the PSD maps. In the Ansai and Bingxian areas, all of the datasets overestimated sand and clay contents and underestimated silt contents. However, in the Zitong area, the polygon linkage and type linkage methods overestimated silt contents and underestimated sand contents, while the opposite happened with HWSD. This indicates that the performance of these datasets varies with soil type, as these areas have different soil types. In all areas, the polygon linkage estimates had the lowest RMSEs for sand and silt contents, but in the Ansai and Bingxian areas, the polygon linkage method did not perform the best of the three methods.

The sources of uncertainty in the linkage methods have been discussed in previous studies (Batjes, 2002; Batjes, 2006). Errors in spatial data are much more important than those in soil analytical methods because of the purity of soil map units, which is likely to be around 50 to 65% (Landon, 1991). The 1:1,000,000 scale soil map of China was compiled through the cartographic generalisation of 1,500,000 scale maps. However, this results in the loss of soil type makeup information, leaving only a single soil type per map polygon, which degrades the quality of the spatial data. The impurity in soil map units, which is not taken into account in the linkage methods, can cause significant errors in estimating soil fractions, as other soil types within a

map unit or map polygon may have quite different textures than the one to which it is linked. The linkage method may be improved by adding all soil profiles within a map polygon to the linked profiles to determine the representative values of soil properties. However, because there is only less than 0.1 soil profile per map polygon in the database for China, any improvement would be very small.

In addition to the two sources of uncertainty mentioned above, the accuracy of the distances between map polygons and profiles, the soil classification system and the linkage method itself can also carry uncertainty. The coordinates of soil profiles were not very accurate because they were derived from the location description. The centre, rather than the boundary, of a map polygon was used to calculate the distances between map polygons and profiles. Therefore, the distances had some associated error. In the polygon linkage, distance was used to determine whether a profile should be linked to a map polygon. Since we abandoned profiles that did not have sufficient precision of location and did not weight any of the distances, the effect of distance errors on the linking results was rather small.

The GSCC system has some shortcomings (Gong, 1999). It is based on the soil genetic hypothesis, which may result in the same soil being classified as different soil types. For example, albic soils were classified as podzolic soils because the albic process and podzolic processes were not yet distinguished in the 1950s. In addition, the GSCC emphasises the importance of climate and vegetation while ignoring the time factor. Therefore, it may end up confusing soil-forming processes that have already happened with those which have not yet occurred. For example, under extreme conditions, it may even classify a purplish soil as a yellow soil (which has an intensive eluviation with high content of goethite, corresponding to Cambisols in WRB). The GSCC emphasises the Central Concept, which states that, while soil type can be very clear, the boundaries between types may be unclear, making some soils hard to classify as a specific soil. The GSCC also lacks quantitative indices, which causes its information system to be difficult to build. Because of the shortcomings mentioned above, it is hard to avoid errors in compiling soil maps and classifying soil profiles. In addition, the soil survey employed a bottom-up procedure starting from the county or town level, and as a result, inconsistencies are inevitable due to differences in the personal judgments of the data collectors. As previously mentioned, soil type names of soil profiles were modified to be consistent with the soil map at different levels of soil type (not always the lowest level of soil type), which can also cause some uncertainty.

While the linkage method used a single soil content value to represent a polygon or a map unit, the texture of soil can vary spatially (sometimes significantly) within a specific polygon or map unit. In this study, the polygon linkage method was used to take into account the inter-polygonal variation within a map unit. This was not considered by the soil type linkage method. The polygon linkage method cannot represent the spatial variation within a polygon, even though its statistical variation can be given by the variance and range of the linked samples. An alternative method that takes intra-polygonal variation into account is the Bayesian Maximum Entropy method (BME), which uses soil texture class maps as input data. (D'Or and Bogaert, 2003). In China, soil texture class map of high precision does not exist; however, the linkage method can create one. The underlying assumption of BME is the continuous change of soil properties inside the polygon, which often applies at a fine scale (as in the case in d'Or and Bogaert, 2003). At a coarse scale, such as the 1:1,000,000 scale in our study, this assumption is more likely to be wrong. Since the average polygon size is about 100 km<sup>2</sup>, the values of soil fractions are not likely to change gradually from the centre to the boundary, but rather through a series of ups and downs along the way.

The process of choosing the linkage between soil map polygons and profiles was subjective, as the target sample size for a soil type level and the search radius at step 1 of the linkage process was to some extent arbitrarily decided. In addition, the linkage method does not apply to the non-soil map units, such as city areas that do have soil, which were set to a value of zero for all soil fractions.

**Table 2**  
Accuracy of soil fraction contents from three datasets in three areas.

| Area     | Source <sup>a</sup> | Sand  |      | Silt  |      | Clay |      |
|----------|---------------------|-------|------|-------|------|------|------|
|          |                     | ME    | RMSE | ME    | RMSE | ME   | RMSE |
| Ansai    | Polygon             | 1.1   | 13.6 | -7.2  | 12.8 | 6.1  | 7.1  |
|          | Type                | 13.6  | 17.0 | -14.0 | 16.3 | 0.3  | 2.8  |
|          | HWSD                | 16.9  | 18.9 | -17.7 | 20.1 | 2.1  | 4.6  |
| Bingxian | Polygon             | 1.9   | 11.0 | -8.2  | 12.0 | 6.3  | 10.9 |
|          | Type                | 7.8   | 17.7 | -13.3 | 16.2 | 5.5  | 13.2 |
|          | HWSD                | 10.5  | 18.1 | -19.3 | 22.2 | 8.9  | 10.2 |
| Zitong   | Polygon             | -2.2  | 13.0 | 5.6   | 12.0 | -3.4 | 7.1  |
|          | Type                | -11.0 | 18.0 | 7.6   | 13.2 | 3.4  | 7.5  |
|          | HWSD                | 12.8  | 18.5 | -8.0  | 13.3 | -4.8 | 7.8  |
| Total    | Polygon             | 0.5   | 12.7 | -5.4  | 12.3 | 4.8  | 9.7  |
|          | Type                | 7.0   | 17.3 | -10.0 | 15.6 | 2.9  | 9.1  |
|          | HWSD                | 13.1  | 18.4 | -16.6 | 20.0 | 4.1  | 8.3  |

<sup>a</sup> "Polygon" was derived by linking soil map polygons and profiles. "Type" was derived by linking soil map units and profiles. "HWSD" is the Harmonized World Soil Database.

The polygon linkage considers the soil type and distance between soil polygons and profiles and indirectly takes into account environment-related factors. As the spatial variation of climate and vegetation is relatively small, our distance-based method may be sufficient because most of the linkage happens at the climate zone scale. In the Second National Soil Survey of China, topographic maps, or air photos, were used as the base maps of soil type maps, with geological maps as a reference. To some extent, the factors of topography, land use and parent material were implicitly considered. However, these factors vary greatly, so a distance-based linkage is not able to capture them well. In the future, it will be necessary to explicitly consider these environmentally related factors. In this context, the polygon linkage is a typical example of the scorpan paradigm of quantitative empirical digital soil mapping (McBratney et al., 2003).

Zhao et al. (2006) developed a pedological knowledge-based method, which considers soil classification information and the locations of profiles. The profiles within a county were all linked to map polygons in the same county in that study, since the Second National Soil Survey of China was implemented from the county level. Though Zhao et al. (2006) indirectly considered spatial location, the actual spatial pattern of soil properties was not confined by administrative division boundaries. It is better to take location into account through the distance between profiles and map polygons, like in our study.

#### 4. Conclusions

A soil PSD dataset with 1-km resolution was developed for its application in land and climate modelling by using the most detailed soil map of China at the national scale and a large soil profile database. The polygon linkage method provides more information about the distribution of soil PSDs than the soil type linkage method. The overall assumptions are that a soil map polygon can be represented by a minimum sample size of soil and that soil fractions vary due to the soil type and location. According to the soil type level of linkage, sample size and search radius, the quality of the data was reliable. Overall, the map polygon linkage offered better results than the soil type linkage or Harmonized World Soil Database. However, we need to put effort into improving the data quality and the product accuracy. The dataset is available for free download from <http://globalechange.bnu.edu.cn>.

Future efforts will be made to improve the quality of the dataset. This linkage method may be used in future work to derive other soil properties, such as rock fragment content, soil depth, and soil carbon and nitrogen contents. It is also necessary to consider environment-related factors directly under the scorpan framework to improve the prediction of soil properties (McBratney et al., 2003).

#### Acknowledgements

This work was supported by funding agencies including, the NSFC under Grant 40775041, the R&D Special Fund for Nonprofit Industry (Meteorology, GYHY200706005 and GYHY200706025) and MOST No. 2010CB951802. We would like to acknowledge A-Xing Zhu and Lizong Wu for their assistance and helpful discussions. We would also like to thank two anonymous reviewers for their thorough and constructive reviews. A group of students at Beijing Normal University helped with the data collection. The authors thank Peggy Chen for help correcting the English writing.

#### References

Arya, L.M., Paris, J.F., 1981. A physicoempirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data. *Soil Sci. Soc. Am. J.* 45, 1023–1030.

- Batjes, N.H., 2002. Soil parameter estimates for soil types of the world for use in global and regional modelling (version 2.1). available through ISRIC Report2002/02c. International Food Policy Research Institute (IFPRI) and International Soil Reference and Information Centre (ISRIC, Wageningen. <http://www.isric.org>).
- Batjes, N.H., 2003. A taxotransfer rule-based approach for filling gaps in measured data in primary SOTER databases. ISRIC Report 2003/03. ISRIC-World Soil Information, Wageningen.
- Batjes, N.H., 2006. ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (version 1.0). available through ISRIC Report2006/02. ISRIC-World Soil Information, Wageningen. <http://www.isric.org> (with dataset).
- Batjes, N.H., Al-Adamat, R., Bhattacharyya, T., Bernoux, M., Cerri, C.E.P., Gicheru, P., Kamoni, P., Milne, E., Pal, D.K., Rawajfih, Z., 2007. Preparation of consistent soil data sets for SOC modelling purposes: secondary SOTER data sets for four case study areas. *Agric. Ecosyst. Environ.* 122, 26–34.
- Dai, Y., 2003. The common land model. *Bull. Amer. Meteor. Soc.* 84, 1013–1023.
- Dickinson, R.E., Henderson-Sellers, A., Kennedy, P.J., Wilson, M.F., 1993. Biosphere atmosphere transfer scheme (BATS) version 1e as coupled for Community Climate Model. NCAR Tech. Note NCAR/TN-378+STR, p. 72.
- Dijkshoorn, J.A., van Engelen, V.W.P., Huting, J.R.M., 2008. Soil and landform properties for LADA partner countries (Argentina, China, Cuba, Senegal, The Gambia, South Africa, Tunisia). ISRIC report 2008/06, GLADA report 2008/03. ISRIC – World Soil Information, FAO, Wageningen (23 pp with data set).
- D'Or, D., Bogaert, P., 2003. Continuous-valued map reconstruction with the Bayesian Maximum Entropy. *Geoderma* 112, 169–178.
- FAO, IASA, ISRIC, ISS-CAS, JRC, 2009. Harmonized World Soil Database (version1.1). FAO, Rome, Italy and IASA, Laxenburg, Austria.
- Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, applications and future research directions. *Trans. Am. Soc. Agric. Biol. Eng.* 50, 1211–1250.
- Gong, Z., 1999. Chinese soil taxonomy: theory, methodology and practice. Science Press of China, Beijing. 26–28.
- Haverkamp, R., Parlange, J.Y., 1986. Predicting the water-retention curve from a particle-size distribution: 1. Sandy soils without organic matter. *Soil Sci.* 142, 325–339.
- Jenny, H., 1941. Factors of Soil Formation, A System of Quantitative Pedology. McGraw-Hill, New York.
- Landon, J.R., 1991. Booker Tropical Soil Manual. Longman Scientific & Technical, New York.
- McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52.
- Miller, D.A., White, R.A., 1998. A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. Available online at Earth Interactions 2. <http://EarthInteractions.org>.
- Minasny, B., McBratney, A.B., 2007. Estimating the water retention shape parameter from sand, clay content. *Soil Sci. Soc. Am. J.* 71, 1105–1110.
- National Soil Survey Office, 1992. China soil survey technology. China Agriculture Press, Beijing, China. in Chinese.
- National Soil Survey Office, 1993a. China soil species description, Vol. 1. China Agriculture Press, Beijing, China. in Chinese.
- National Soil Survey Office, 1993b. China soil species description, Vol. 2. China Agriculture Press, Beijing, China. in Chinese.
- National Soil Survey Office, 1994. China soil species description, Vol. 3. China Agriculture Press, Beijing, China. in Chinese.
- National Soil Survey Office, 1995a. China soil species description, Vol. 4. China Agriculture Press, Beijing, China. in Chinese.
- National Soil Survey Office, 1995b. China soil species description, Vol. 5. China Agriculture Press, Beijing, China. in Chinese.
- National Soil Survey Office, 1996. China soil species description, Vol. 6. China Agriculture Press, Beijing, China. in Chinese.
- Reynolds, C.A., Jackson, T.J., Rawls, W.J., 2000. Estimating soil water-holding capacities by linking the Food Agriculture Organization soil map of the world with global pedon databases, continuous pedotransfer functions. *Water Resour. Res.* 36, 3653–3662.
- Scholes, R.J., Skole, D., Ingram, J.S., 1995. A Global Database of Soil Properties: Proposal for Implementation. Report of the Global Soils Task Group, IGBP-DIS. 35 pp.
- Shangguan, W., Dai, Y., 2009. Investigation of different models to describe soil Particle-size distribution for sparse experimental data. *J. Beijing Norm. Univ. Nat. Sci.* 45, 279–283.
- Shangguan, W., Dai, Y., 2010. An investigation of soil particle-size distribution models for the conversion of soil texture classification from ISSS, Katschinski's to USDA System. In preparation.
- Shi, X.Z., Yu, D.S., Warner, E.D., Pan, X., Petersen, G.W., Gong, Z.G., Weindorf, D.C., 2004. Soil database of 1:1,000,000 digital soil survey and reference system of the Chinese Genetic Soil Classification System. *Soil Surv. Horiz.* 45, 129–136.
- Shi, X.Z., Yu, D.S., Warner, E.D., Sun, W.X., Petersen, G.W., Gong, Z.T., 2006a. Cross-reference system for translating between Genetic Soil Classification of China and Soil Taxonomy. *Soil Sci. Soc. Am. J.* 70, 78–83.
- Shi, X.Z., Yu, D.S., Xu, S.X., Warner, E.D., Wang, H.J., Sun, W.X., Zhao, Y.C., Gong, Z.T., 2010. Cross-reference for relating Genetic Soil Classification of China with WRB at different scales. *Geoderma* 155, 344–350.
- Shi, X.Z., Yu, D.S., Yang, G.X., Wang, H.J., Sun, W.X., Du, G.H., Gong, Z.T., 2006b. Cross-reference benchmarks for translating the genetic soil classification of China into the Chinese Soil Taxonomy. *Pedosphere* 16, 147–153.
- Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J.O., Levis, S., Lucht, W., Sykes, M.T., Thonicke, K., Venevsky, S., 2003. Evaluation of ecosystem

- dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Glob. Change Biol.* 9, 161–185.
- Webb, R.S., Rosenzweig, C.E., Levine, E.R., 1993. Specifying land surface characteristics in general circulation models: soil profile data set and derived water-holding capacities. *Glob. Biogeo. Cyc.* 7, 97–108.
- Zhang, G., Zhao, Y., 2008. SOTER database for China, scale 1:1 million. Institute of Soil Science, Chinese Institute of Soil Science, Nanjing.
- Zhao, Y., Shi, X., Weindorf, D.C., Yu, D., Sun, W., Wang, H., 2006. Map scale effects on soil organic carbon stock estimation in North China. *Soil Sci. Soc. Am. J.* 70, 1377–1386.
- Zobler, L., 1986. A World Soil File for Global Climate Modelling. NASA Technical Memorandum # 87802. NASA Goddard Institute for Space Studies, New York, U.S.A.