



# The Unified North American Soil Map and its implication on the soil organic carbon stock in North America

S. Liu<sup>1,\*</sup>, Y. Wei<sup>1</sup>, W. M. Post<sup>1</sup>, R. B. Cook<sup>1</sup>, K. Schaefer<sup>2</sup>, and M. M. Thornton<sup>1</sup>

<sup>1</sup>Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

<sup>2</sup>National Snow and Ice Data Center, University of Colorado, Boulder, CO 80309, USA

\* now at: College of Resources and Environment, Huazhong Agricultural University, Wuhan, Hubei 430070, China

Correspondence to: S. Liu (carol.shishi@gmail.com)

Received: 26 September 2012 – Published in Biogeosciences Discuss.: 30 October 2012

Revised: 25 March 2013 – Accepted: 4 April 2013 – Published: 2 May 2013

**Abstract.** The Unified North American Soil Map (UNASM) was developed to provide more accurate regional soil information for terrestrial biosphere modeling. The UNASM combines information from state-of-the-art US STATSGO2 and Soil Landscape of Canada (SLCs) databases. The area not covered by these datasets is filled by using the Harmonized World Soil Database version 1.21 (HWSD1.21). The UNASM contains maximum soil depth derived from the data source as well as seven soil attributes (including sand, silt, and clay content, gravel content, organic carbon content, pH, and bulk density) for the topsoil layer (0–30 cm) and the subsoil layer (30–100 cm), respectively, of the spatial resolution of 0.25 degrees in latitude and longitude. There are pronounced differences in the spatial distributions of soil properties and soil organic carbon between UNASM and HWSD, but the UNASM overall provides more detailed and higher-quality information particularly in Alaska and central Canada. To provide more accurate and up-to-date estimate of soil organic carbon stock in North America, we incorporated Northern Circumpolar Soil Carbon Database (NCSCD) into the UNASM. The estimate of total soil organic carbon mass in the upper 100 cm soil profile based on the improved UNASM is 365.96 Pg, of which 23.1 % is under trees, 14.1 % is in shrubland, and 4.6 % is in grassland and cropland. This UNASM data will provide a resource for use in terrestrial ecosystem modeling both for input of soil characteristics and for benchmarking model output.

## 1 Introduction

Analyses of the global carbon cycle suggest a significant role of North America as a biospheric sink of atmospheric carbon dioxide (CO<sub>2</sub>) in the overall carbon budget in the world (Prentice, 2001; Gurney et al., 2002; CCSP, 2007). Given the crucial role of North America in global carbon dynamics, North America has become a focus of a US interagency research initiative aimed at quantifying sources and sinks of carbon and the mechanisms underlying continental-scale carbon balance (Wofsy and Harris, 2002).

Soil characteristics, an important input for terrestrial ecosystem models, are mainly used to simulate soil water balance. Generally, the soil water balance is modeled in two approaches in terrestrial ecosystem models. Models, such as CENTURY (Parton et al., 1987), the Carnegie-Ames-Stanford (CASA) (Potter et al., 1996), and BIOME-BGC (Kimball et al., 1997), consider the soil mainly as a storage reservoir, or “bucket”. The bucket fills with water when precipitation is greater than evaporation, and no movement of soil water between adjacent soil layers occurs. The soil storage is defined by a prescribed soil depth and field capacity. Precipitation is passed to the next soil layer until the soil is at field capacity. Although the bucket model is simple to use, it is not sophisticated enough to describe complex soil water process. Therefore, some terrestrial ecosystem models adopt more complicated approaches to simulate soil water balance. For example, the Integrated Biosphere Simulator (IBIS) and Simple Biosphere Model (SiB) employ Darcy’s law to simulate the vertical movement of soil water (Foley et al., 1996; Sellers et al., 1986). The Lund-Potsdam-Jena (LPG) (Sitch et

al., 2003), the Terrestrial Ecosystem Model (TEM) (McGuire et al., 1993), and BIOME2 (Prentice et al., 1992; Haxeltine et al., 1996) simulate soil water balance based on the empirical model developed by Haxeltine et al. (1996), which is originated from Neilson et al. (1995) for the Mapped Atmosphere-Plant-Soil System (MYPASS). However, whichever method is used in terrestrial ecosystem models to simulate soil water balance, they all require soil texture and depth information to determine hydrologic parameters in the water balance model and to further estimate soil moisture. While most models only require soil texture and depth (e.g. LPJ, IBIS), some models also require gravel fraction (e.g. BIOME-BGC) and organic matter content. These soil properties indirectly affect leaf phenology, carbon dynamics, and soil nutrients pools through effects of soil water balance model outputs. In addition, terrestrial ecosystem models require existing or potential stores of soil nutrients (e.g. carbon and nitrogen content) when simulating photosynthesis, respiration, or other biosphere processes (Cramer and Fischer, 1996).

Unfortunately, a lack of comprehensive gridded information about North American soil properties based on US and Canadian soil datasets has impeded the understanding and improvement of modeling carbon dynamics in North America. Currently, North American carbon modeling relies on the spatial subset of different world soil maps, such as digitized Food and Agriculture Organization – United Nations Educational, Science and Cultural Organization (FAO-UNESCO) soil map (Bouwman et al., 1993; McGuire et al., 1993), a world dataset of derived soil properties by FAO-UNESCO soil unit for global modeling (Batjes, 1997), the World Inventory of Soil Emission Potentials (WISE) (Gijssman et al., 2007), and the Harmonized World Soil Database (HWSD) (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2013). In addition, there are some efforts to develop global and regional soil gridded datasets for ecology, climate and hydrology applications. Webb et al. (1993) produced a global dataset, at  $1^\circ \times 1^\circ$  spatial resolution, of soil physical properties by combining FAO Soil Map of the World with the World Soil Data File of Zobler (1986). This data provides the top and bottom depths and percentages of sand, silt, and clay of individual soil horizons for 106 soil types. Kern (1995) estimated geographic patterns of soil-water-holding capacity for the conterminous US, using the US Department of Agriculture–Natural Resources Conservation Service (USDA–NRCS) National Soil Geographic Database Major Land Resource Areas as a geographic base and the 1992 National Resources Inventory and the associated Soil Interpretations Record (SIR). Batjes (1996) developed a world dataset of soil water retention properties using pedotransfer rules based on WISE. Miller and White (1998) developed a multilayer soil dataset for the conterminous US, with a 1 km spatial resolution and based on the STATSGO. The soil properties provided by this dataset include soil texture, rock fragment classes, depth-to-bedrock, bulk density, porosity, rock fragment volume, sand, silt, and clay fractions, available water capacity, and

hydrologic soil group. As a part of the Land Data Assimilation (LDAS) project (Mitchell et al., 2003), the soil characteristics database were created at  $1/8$  degree spatial resolution for hydrologic modeling, in which the conterminous US are based on Miller and White (1998) soil dataset and the other areas are based on FAO soil map. The Vegetation/Ecosystem Modelling and Analysis Project (VEMAP) (Kittel et al., 1995) developed a soil dataset that included soil texture, depth, and other properties for the dominant soil types based on Kern's (1994, 1995) 10 km gridded Soil Conservation Service national-level (NATSGO) database.

The increasing amount of soil observations makes the FAO-UNESCO Soil Map of the World obsolete, but none of the available world soil maps incorporate the more detailed and up-to-date US and Canada soil datasets. This paper describes the development of a two-layer gridded soil characteristic dataset in North America for use in terrestrial ecosystem and the related modeling. This Unified North American Soil Map (UNASM) provides the commonly used soil properties that can be found in all the data sources. Hydrologic properties are not provided by some of our data sources and most ecosystem models have their own methods to estimate soil hydrologic properties. Thus, soil hydrologic properties, although can be useful in hydrologic, agricultural, and some ecosystem models, are not included in this dataset. In this paper, the UNASM is compared with the subset of HWSD 1.21 and the differences between these two datasets are evaluated.

A soil map is usually used to initialize or validate models that study hydrology, photosynthesis, vegetation dynamics, and any applications involving soil moisture. A soil map is also essential to determine soil organic carbon stock, which is the largest pool of terrestrial organic carbon and is affected by land use/land cover change and climate change. Many global and regional soil organic carbon estimates are available (Post et al., 1982; Sombroek et al., 1993; Jobbágy et al., 2000; Tarnocai, 2009). However, few 0.5 degree or finer resolution map of the size and spatial distribution of soil organic carbon pools in North America exist. In this study, we estimate North American soil organic carbon stock based on the Northern Circumpolar Soil Carbon Database (NCSCD)-modified UNASM and the subset of HWSD 1.21 for major vegetation types, and analyze the spatial distribution of soil organic carbon in North America.

## 2 Data

The UNASM encompasses the US (including Alaska), Canada, Mexico, and a part of Guatemala. It spans from 84 degrees to 10 degrees latitude, and from  $-170$  degrees to  $-50$  degrees longitude. Below are the descriptions of the source datasets.

## 2.1 US General Soil Map (STATSGO2)

The US General Soil Map (STATSGO2) was developed by the National Cooperative Soil Survey and supersedes the State Soil Geographic dataset published in 1994. It consists of a broad based inventory of soils and non-soil areas that occur in a repeatable pattern on the landscape and that can be cartographically shown at the scale map (Soil Survey Staff, Natural Resources Conservation Service, <http://soildatamart.nrcs.usda.gov>). The dataset was created by generalizing more detailed soil survey maps. Where more detailed soil survey maps were not available, data on geology, topography, vegetation, and climate were assembled, together with Land Remote Sensing Satellite (LANDSAT) images. Map unit composition was determined by transecting or sampling areas on the more detailed maps and expanding the data statistically to characterize the whole map unit. The spatial scale of this dataset is 1 : 250 000.

## 2.2 Soil Landscapes of Canada (SLC) version 3.2 and 2.2

The SLC V3.2 and V2.2 are standardized datasets consisting of the major characteristics of soil and land for Canada. SLCs were compiled at a scale of 1 : 1 000 000, and information is organized according to a uniform national set of soil and landscape criteria based on permanent natural attributes. Each polygon on the map is described by a standard set of attributes and the associated landscape, such as surface form, slope, permafrost and so on. Updates and corrections to boundaries, attributes, and file structures have taken place over the years. SLC V3.2 is the latest revision of the Soil Landscapes of Canada, which was developed by Agriculture and Agri-Food Canada to provide information about the country's agricultural soils and the provincial and national levels (Soil Landscapes of Canada Working Group, 2010). SLC V2.2 is the latest complete coverage of Canada, including areas outside the agricultural regions of the country (Centre for Land and Biological Resources Research, 1996). Both versions of soil landscapes data are linked to Canada soil name and soil layer table V2.0, which contain comprehensive soil attributes along vertical direction for all soils in Canadian National Soil DataBase (NSDB).

## 2.3 Harmonized World Soil Database (HWSD) Version 1.21

The HWSD is a 30 arc-second (ca. 1 km) raster database with over 16 000 different soil mapping units that combines existing regional and national updates of the soil information worldwide (Soil and Terrain Database (SOTER), European Soil Database (ESD), Soil Map of China, World Inventory of Soil Emission Potential database (WISE)) with the information contained within the 1 : 5 000 000 scale FAO-UNESCO soil map of the world (FAO/IIASA/ISRIC/ISS-CAS/JRC,

2013). The soil-mapping-units raster layer can be linked to harmonized attribute data, which contains 16 physical and chemical soil properties. The HWSD contains two standard depths – the topsoil layer ranges from 0 to 30 cm and the subsoil layer ranges from 30 to 100 cm.

The HWSD 1.21 is the latest version of HWSD. One of the improvements of HWSD 1.21 is that it provides both reference bulk density values and the secondary Soil and Terrain Database (SOTER) bulk density. The reference bulk density values are calculated from equations developed by Saxton et al. (1986) that relate to the texture of the soil only (FAO/IIASA/ISRIC/ISSCAS/JRC, 2013). These estimates, although generally reliable, overestimate the bulk density in soils that have a high porosity (Andosols) or that are high in organic matter content (Histosols) (FAO/IIASA/ISRIC/ISSCAS/JRC, 2013). The SOTER bulk density has been estimated by soil type and depth, based on available analyzed soil data in the SOTER database of soil texture, organic matter, and porosity (FAO/IIASA/ISRIC/ISSCAS/JRC, 2013). Thus, we use SOTER bulk density in this dataset.

## 2.4 The Northern Circumpolar Soil Carbon Database (NCSCD)

One of the important objectives of this study is to quantify the soil organic carbon storage in North America. However, none of the datasets described above provide sufficient and accurate soil organic carbon information for the high-latitude permafrost region. To provide a more accurate estimate of the soil organic carbon storage in North America, we combined the NCSCD with soil organic carbon content derived from the UNASM.

The NCSCD was developed to quantify the soil organic carbon stocks in the circumpolar permafrost region ( $18.7 \times 10^6 \text{ km}^2$ ). The NCSCD links organic carbon measurements from 1647 pedons in the northern permafrost regions to several digitized regional/national soil maps to produce a combined circumpolar coverage (Hugelius et al., 2013). Together these datasets have been used to quantify soil organic carbon stock in the topsoil (0–30 cm depth) and down to a depth of 1 m. The NCSCD provides both GIS-polygon files and gridded datasets. As the spatial resolution of the UNASM is 0.25 degrees, we used the 0.25 degree gridded NCSCD (in NetCDF format) in this study for estimating North American soil organic carbon stock.

## 3 Methods and procedures

STATSGO2 and SLCs provide more detailed and accurate soil information than HWSD. However, STATSGO2 and SLCs are not easy for terrestrial ecosystem modelers to use directly. First, many terrestrial ecosystem and related models require a uniform grid cell or raster format different from the format of STATSGO and SLCs, which are defined as

polygons in a vector geographic information system environment. Second, the number, thickness, and depth of soil layers vary widely from one soil component to another, but most models need soil data with harmonized layers that have uniform depths. Third, the location information is not provided for each component, and only the relative amount of each component within a map unit is specified. Consequently, our objective is to produce a gridded soil map that would meet the needs of modelers, combining information of STATSGO2 and SLCs and filling the rest of the area with HWSD1.21.

The compilation of the UNASM required a marked degree of data integration and generalization of the geographical distribution of soil types to a regionally representative pattern. Developing the UNASM involved four stages:

1. Convert STATSGO2, SLCs, and HWSD 1.21 into 0.25 degree gridded format by selecting the dominant soil type in each cell.
2. Merged STATSGO2, SLCs and HWSD1.21 into a seamless map that can best represent soil in North America.
3. Harmonize the North American soil map developed in stage 2 into two standard layers, 0–30 cm and 30–100 cm.
4. Quality control.

### 3.1 Stage I: convert STATSGO2, SLCs, and HWSD 1.21 into 0.25 degree gridded format

For both STATSGO2 and SLCs, the required soil properties were linked to the soil maps, which contain polygon features of soil and nonsoil map units on the landscape. Each 0.25 degree grid cell may overlap with one or more soil unit polygon feature and each of these soil unit polygon features may contain one or more soil components, such as sand/silt/clay fractions and bulk density. We first evaluated all the unique soil types and their combined area fractions in a 0.25 degree cell. The soil type with the highest area fraction was selected as the dominant soil type in the cell. We then selected the soil component that has the largest area of the dominant soil type. Last, the selected soil component's vertical soil layers information along with their detailed soil properties were assigned to the target 0.25 degree grid cell. The data structure of HWSD 1.21 is similar to those of STATSGO2 and SLCs. The difference is that the HWSD 1.21 is represented in a 30 arc-second resolution raster map instead of vector polygons. We applied a similar approach to that used for the two polygonal datasets to select the dominant soil type and soil component for each target 0.25 degree cell and then linked to the detailed soil attributes.

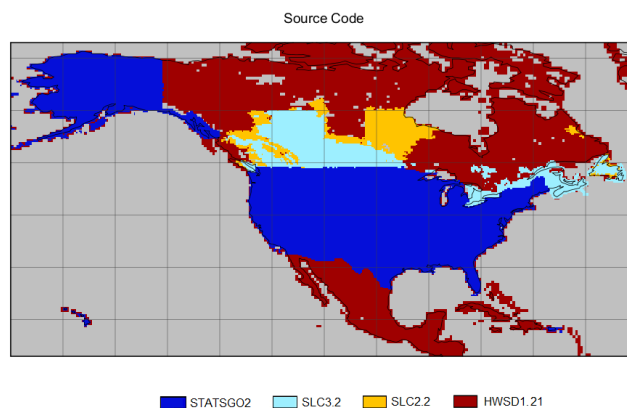


Fig. 1. The spatial distribution of data sources for the UNASM.

### 3.2 Stage II: integrating different soil datasets into a seamless product

The gridded STATSGO2, SLCs and HWSD 1.21 were integrated into a 0.25 degree North American grid with a total of 142 080 grid cells. The HWSD for North America has not been updated since the FAO Soil Map of the World was developed in 1970s. Thus, we give the lowest priority to HWSD. STATSGO was first published in 1994, and in 2006 the STATSGO spatial and tabular data (which contains soil attribute values) were revised and updated, and has been renamed to STATSGO2. SLC3.2, published in 2011, provides soil data for the significant agricultural regions of Canada. SLC3.2 is based on SLC3.0 (released in 2005), but SLC3.2 added some additional agricultural areas with new soil attribute data. SLC2.0, released in 1996, not only covers the agricultural region but also some other areas in Canada. However, SLC2.0 is less reliable than SLC3.2 based on the personal contact with the officers in Agriculture and Agri-Food Canada. Therefore, based on whether the data is reliable and up-to-date, we merged different soil data sources with the following priority: STATSGO2 > SLC 3.2 > SLC 2.2 > HWSD 1.21. For the cells that have both STATSGO2 and SLC3.2, the fraction of the cell within Canada and U.S. was calculated. If most of the area of a cell is within the US, we will assign the STATSGO2 data to the cell. Otherwise, SLC3.2 will be assigned to the cell.

The goal of this selection was to take advantage of more precise soil information from STATSGO2 and SLCs, and fill the rest of the area with the HWSD 1.21. The source dataset used by each cell can be traced back using the source code variable. Figure 1 illustrates the spatial distribution of data sources for the UNASM. Among all the cells that have valid values, STATSGO2 accounts for 40 % of the total land area, SLC3.2 and SLC2.2 accounts for 8.5 % and 4.3 %, respectively, and HWSD 1.21 accounts for 47.2 % of total land area.

### 3.3 Stage III: harmonization with depth

After integrating the different soil datasets together, the seamless dataset from Stage II still has inconsistent vertical soil layers. Grid cells assigned with soil properties from STATSGO2 may contain up to 11 soil layers, those from SLCs may contain up to 9 soil layers and cells from HWSD have two soil layers. The thickness of soil layers also varies across cells. To harmonize the vertical structure within these three datasets in the UNASM, we converted the profile data into two standard layers that are consistent with the HWSD 1.21. The top layer is from 0 to 30 cm and the sub layer is from 30 to 100 cm.

We used the depth-weighted averaging method to interpolate the layers from the Stage II dataset to the two standard layers for the volume-related properties gravel fraction and pH, and used the mass-weighted averaging method for the other soil properties that are related to soil weight. The standard UNASM layers were compared to each layer in the Stage II soil dataset. If the UNASM layer was entirely contained within a single unified soil data layer from Stage II, we used the Stage II layer value for the standard layer. Otherwise, all layers from Stage II that were fully or partially included within the standard UNASM layers were identified, and a portion to each of the standard layers was used as the weighting to determine the standard layer properties. If the soil thickness is less than 30 cm or 100 cm, the maximum soil depth would be used for harmonization rather than extending the attributes to 100 cm. Therefore, if the soil thickness was less than 30 cm, the weighted average of soil properties in different layers would be assigned to the topsoil layer, but the subsoil layer (30–100 cm) would be filled with missing value, which is –999.0 in this study. The maximum soil depth derived from the data sources is kept as a separate variable in the UNASM.

### 3.4 Stage IV: quality control

All fields in the UNASM were checked for the minimum and the maximum, which were then compared to the value range for each soil property in the source datasets. The cell values in the UNASM falling out of soil property value ranges were treated as errors. Only few errors were found, and these errors have been corrected.

Missing values resulting from empty entries in the data source were filled with –999.0, except for soil texture fields in the surface organic layer. Zeros in sand, silt, clay content in the surface organic layer from STATSGO2 and SLCs are valid values for soil since soil texture is not applicable in this case.

The sum of sand, silt and clay fractions in top and subsoil layers was corrected to 100 % in the cases where necessary due to rounding errors. Similar to what was done in the development of the HWSD 1.21, when the sum was less than 100 %, the largest percentage was increased to obtain 100 %.

**Table 1.** Soil depth, source code, and attributes of topsoil layer (0–30 cm) and subsoil layer (30–100 cm).

Soil Attribute	Abbreviation	Units
Maximum Soil Depth	Soil Depth	cm
Source Code	Source	na
Topsoil Sand Fraction	$t_{\text{sand}}$	% weight
Topsoil Silt Fraction	$t_{\text{silt}}$	% weight
Topsoil Clay Fraction	$t_{\text{clay}}$	% weight
Topsoil Gravel Fraction	$t_{\text{gravel}}$	% volume
Topsoil Organic Carbon	$t_{\text{oc}}$	% weight
Topsoil pH (H <sub>2</sub> O)	$t_{\text{ph}}$	–log (H <sup>+</sup> )
Topsoil Bulk Density	$t_{\text{bd}}$	g cm <sup>–3</sup>
Subsoil Sand Fraction	$s_{\text{sand}}$	% weight
Subsoil Silt Fraction	$s_{\text{silt}}$	% weight
Subsoil Clay Fraction	$s_{\text{clay}}$	% weight
Subsoil Gravel Fraction	$s_{\text{gravel}}$	% volume
Subsoil Organic Carbon	$s_{\text{oc}}$	% weight
Subsoil pH (H <sub>2</sub> O)	$s_{\text{ph}}$	–log (H <sup>+</sup> )
Subsoil Bulk Density	$s_{\text{bd}}$	g cm <sup>–3</sup>

When the sum exceeded 100 %, the highest value was reduced to obtain a sum of 100 %.

### 3.5 Soil properties

This study considers seven soil attributes that are commonly required in terrestrial biosphere modeling studies (Table 1). In addition, we provide the maximum soil depth from the sources and the source code that specifies the origin of the attributes values. The detailed explanations of each soil property can be found in Appendix A.

### 3.6 Computation of soil carbon content and soil carbon mass

The soil organic carbon content (SOCC, g cm<sup>–2</sup>) was calculated for each cell of the UNASM using the formula:

$$\text{SOCC} = \text{OC} \times \text{BD} \times T \times (1 - \text{Gravel}), \quad (1)$$

where OC is the soil organic carbon concentration (% weight), BD is the bulk density (g cm<sup>–3</sup>),  $T$  is the soil layer thickness (either 30 or 70 cm), and Gravel is the gravel fraction (% volume). Using this information, the SOCC was calculated for the 0–30 and 30–100 cm layers for all cells. The total SOCC of the upper 100 cm soil profile is the sum of SOCC in the topsoil and subsoil layers. To provide a more detailed estimate of soil organic carbon, the SOCC values in the high-latitude cells in the UNASM were then replaced by the SOCC values in the corresponding cells in the 0.25 degree NCSCD. The soil organic carbon mass (SOCM, Pg) was determined by multiplying the SOCC by the area of each cell.

## 4 Results

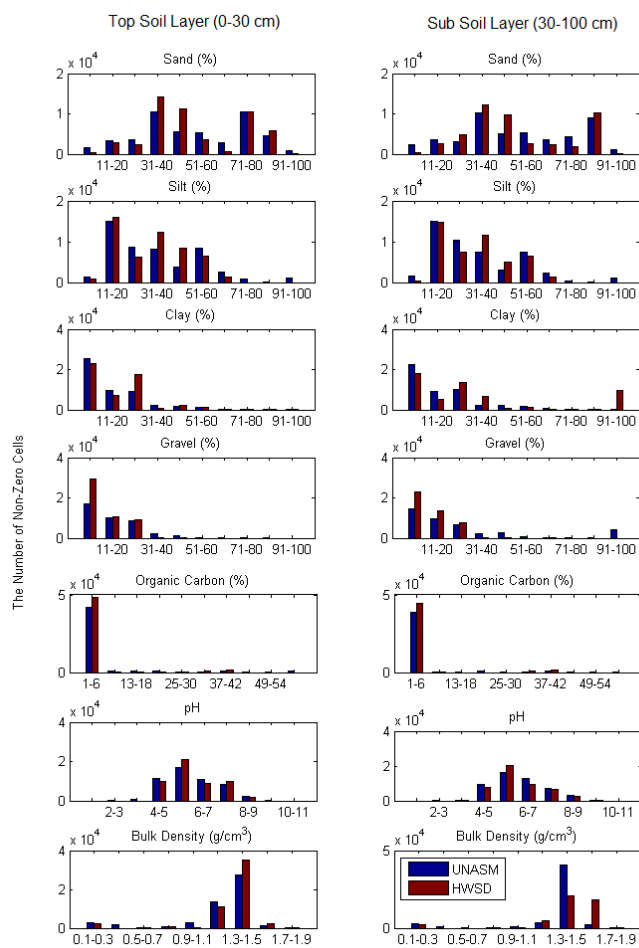
### 4.1 The comparisons of soil properties between UNASM and HWSD

Figure 2 shows the histogram of soil property values in the topsoil layer (0–30 cm) and the subsoil layer (30–100 cm) in the UNASM and HWSD. Overall, the distribution of soil property values in the UNASM agrees with those in the HWSD, but the distributions of gravel fraction, organic carbon concentration, and bulk density have greater discrepancies between the UNASM and the HWSD. The UNASM has higher gravel fraction, with 56 % non-zero cells within 10–100 % value range in the topsoil layer and 68 % of non-zero cells in the subsoil layer. In contrast, the HWSD only has 45 % of non-zero cells lie within 10–100 % value range in the topsoil layer and 56 % in the subsoil layer. The UNASM also has higher organic carbon concentration. The maximum soil organic carbon concentration in the UNASM is 58 % in the topsoil layer and 60 % in the subsoil layer, but in the HWSD the maximum is 38 % in the topsoil layer and 39 % in the subsoil layer. In the topsoil layer, 13 % of non-zero cells in the UNASM lie within 5–60 % value range while only 6 % of non-zero cells in the HWSD lie within the same high value range. Both STATSGO2 and SLCs contain the organic layer. When harmonizing the unified data into two standard layers, we combined organic layer with the mineral layers below. Usually the organic layer did not extend past the top 30 cm, so only the topsoil layer was affected, resulting in the higher soil organic carbon concentration in the UNASM. However, occasionally, the organic layer extends below 30 cm depth, affecting the density of both the topsoil and subsoil layers in the UNASM. Unfortunately, there is no information available about whether the HWSD takes into account the organic layer during harmonization. Because of the higher soil organic carbon concentration, the UNASM has lower bulk density than the HWSD, with a mean of  $1.2 \text{ g cm}^{-3}$  in the topsoil layer and  $1.3 \text{ g cm}^{-3}$  in the subsoil layer.

The difference map, which is the result of UNASM minus HWSD 1.21, demonstrates the spatial distributions of differences between the UNASM and the HWSD 1.21 for all soil properties (Fig. 3).

#### 4.1.1 Soil texture

The UNASM has lower sand content in the central and eastern US, Alaska, and some areas in Canada, but higher sand content in the coastal area of southeastern US and the areas near Great Slave Lake and Lake Winnipeg in Canada. Clay content in general is slightly lower in the UNASM. The locations of high clay content are in the north of Lake Winnipeg (Canada), along the Mississippi River (US), and in central Montana (US). The difference in spatial distribution of soil texture would affect the estimates of hydrological properties of soil, such as porosity and hydraulic conductivity, and



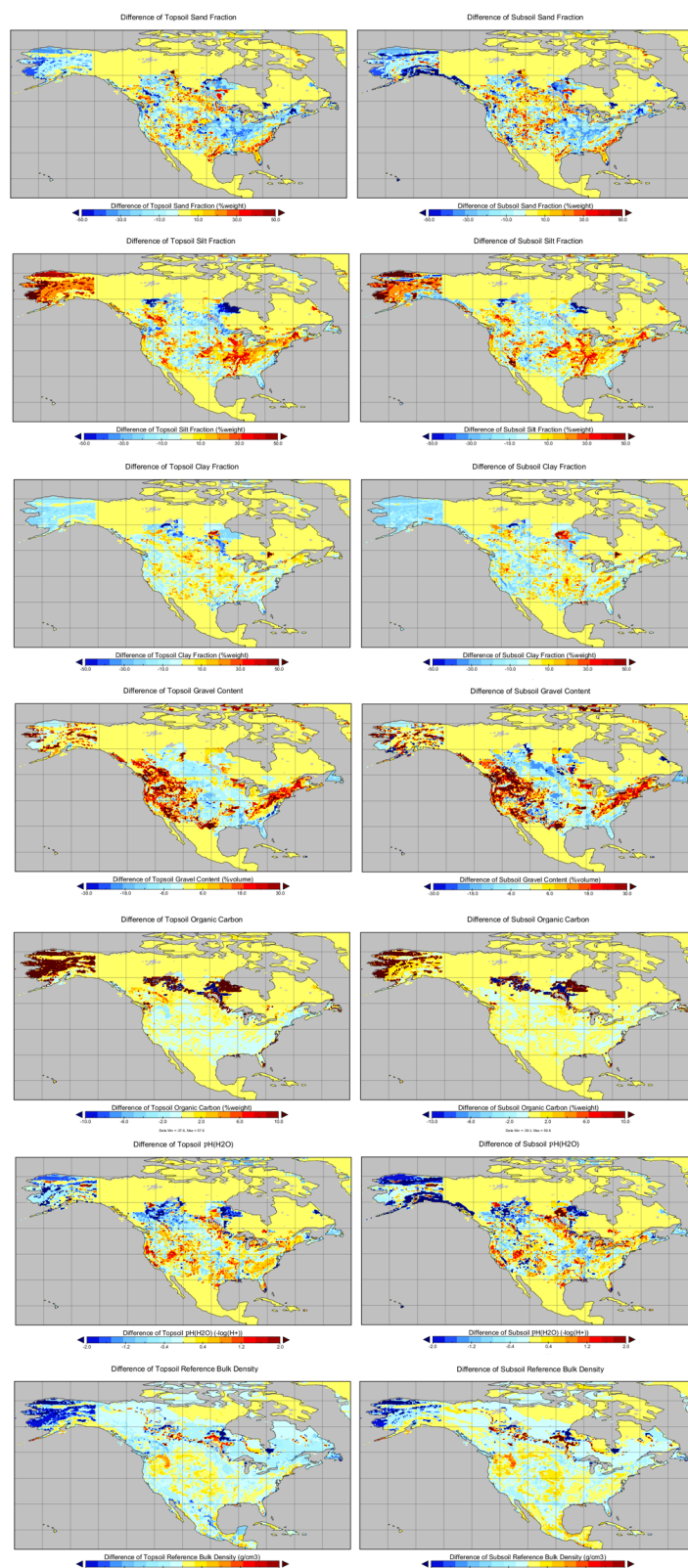
**Fig. 2.** The histogram of soil properties of the topsoil layer (0–30 cm) and subsoil layer (30–100 cm) in the UNASM and the subset of the HWSD 1.21.

hence influence the modeling of water availability and energy partitioning.

#### 4.1.2 Gravel fraction

Gravel fraction is much higher in the western and northeastern US, some areas of Alaska, and southwestern Canada in the UNASM. We compared the gravel fractions derived from these two datasets with the data reported in a few studies. In the northwestern Sonoran Desert, California, Young and Nobel (1986) reported that the average gravel content from 0–50 cm is about 35 % in the study sites. In the UNASM, the gravel fraction in Sonoran Desert ranges from 22 % to 51 %, but the average gravel fraction in HWSD in the same area is about 4 %. Simanton et al. (1994) reported up to 60 % gravel in the upper 10 cm of the soil profiles and less than 40 % gravel in the underlying parts of the profiles in the study sites in southeastern Arizona. In the same area, gravel fraction in the UNASM ranges from 6.8 % to 51 %, but in the HWSD it ranges from 4 % to 20 %. Vasek (1980) reported 24 %–57 %





**Fig. 3.** The difference map between the UNASM and the subset of the HWSD 1.21 for each soil property. The topsoil ranges from 0 to 30 cm and the subsoil ranges from 30 to 100 cm.

gravel in the study sites in the San Bernardino County of Mojave Desert, California. In the UNASM, the gravel fraction is between 9 % and 51 %, but in the HWSD, the gravel fraction is between 4 % and 20 % in the same area. The gravel fraction data provided in a few studies suggest that the HWSD may underestimate the gravel fraction in western US.

#### 4.1.3 Soil organic carbon concentration

The UNASM has much higher soil organic carbon concentration in most area of Alaska, and the lake area in the central Canada, but slightly lower values in the conterminous US for the topsoil layer. A few studies reported the soil organic carbon concentration and bulk density of the soil samples taken at sites in Alaska (Table 2). With these data, we calculated the soil organic carbon concentration and bulk density for the surface layer (0–30 cm) and the subsoil layer (30–100 cm) using the same method as we used for UNASM. In the UNASM, soil organic carbon concentration values range from 0.87 to 51 % and the median is 17 % in Alaska, which agrees well with the field studies data listed in Table 2. This qualitative comparison suggests that the HWSD may underestimate the soil organic carbon concentration in Alaska.

#### 4.1.4 pH

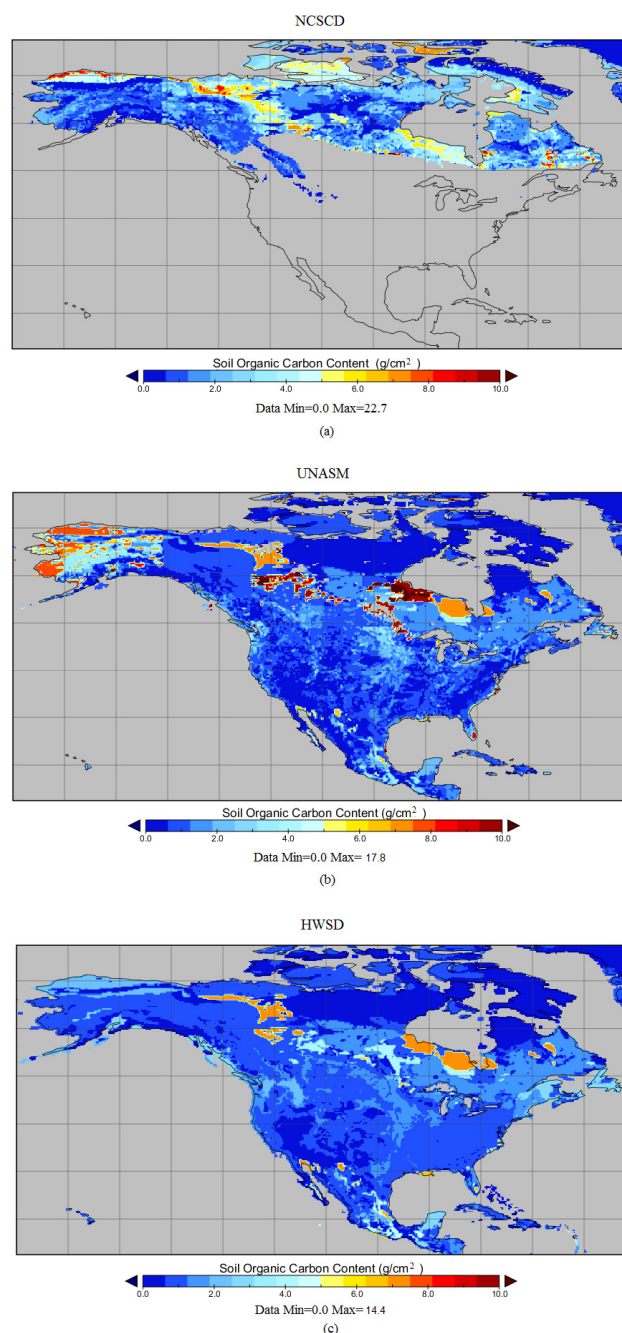
We compared the pH values in the UNASM and the HWSD with the median pH values provided in Soil Test Levels in North America (Fixen, 2005). We found that both datasets agree well with values in Fixen's report, but pH in the UNASM is slightly lower than pH in the HWSD in most areas of southwestern Canada, Alaska, and some areas of US, while slightly higher in southwestern and southeastern US.

#### 4.1.5 Bulk density

Bulk density diverges from HWSD to a greater degree than for the other soil properties. The UNASM bulk density is the mass-weighted average and includes the soil organic layer as part of the profile. In the area filled with the values from HWSD, bulk density is also recalculated using our method. The values in the area filled with the HWSD are slightly lower than the values in the original HWSD. The most significant differences of bulk density lie in Alaska, US and northern Alberta and southern Northwest Territories, Canada. Where the soil organic carbon concentration is higher in the UNASM, the bulk density is lower than in the HWSD.

### 4.2 Spatial distribution of SOCC

Regional patterns of soil carbon content from 0–100 cm based on the NCSCD, UNASM, and HWSD are shown in Fig. 4. The UNASM provides more details in conterminous US than HWSD does, particularly showing higher values in central US, the coastal area of Washington state, Minnesota,



**Fig. 4.** SOCC in the top 100 cm soil profile derived from (a) NCSCD, (b) UNASM, and (c) HWSD 1.21.

coastal area of North Carolina, and southern Florida. In northern circumpolar region, all three maps demonstrate the high SOCC in coastal area of Ontario, Manitoba and around Great Bear Lake, Great Slave Lake and Lake Winnipeg in Canada. However, the NCSCD provides pronounced details of SOCC distribution in Canada especially in northern Canada where both UNASM and HWSD lack data.



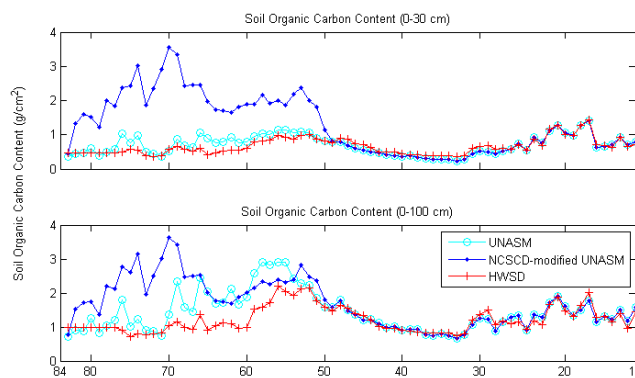
**Table 2.** Soil organic carbon concentration (OC) and bulk density (BD) of field studies in Alaska.

Location	Land Cover	Depths (cm)	BD (g cm <sup>-3</sup> )	OC (%)	Reference
Barrow	Moist Acidic tundra	0–30	0.96	12.4	Michaelson et al. (1996)
		30–100	0.97	8.9	
Barrow	Wet acidic tundra		0.58	11.72	
			1.11	9.08	
Nello Pingo	Moist acidic tundra		0.41	31.82	
			0.40	7.46	
Betty Pingo	Wet non acidic tundra	0–30	0.60	21.3	
		30–100	0.38	11.7	
Betty Pingo	Moist non acidic tundra		0.78	11.27	
			0.38	12.40	
AK Pipeline Mile 24	Wet non acidic tundra		0.39	13.60	
			0.89	3.17	
Sagwon Hills	Moist non acidic tundra		0.59	7.95	
			1.00	4.38	
Toolik Lake	Moist acidic tundra		0.47	11.65	
			0.52	5.63	
Toolik Lake, Alaska	Wet acidic fen	0–30	0.47	10.8	
		30–100	0.30	8.9	
Imnaviat Creek	Wet acidic tundra		0.09	41.81	
			0.26	34.08	
Imnaviat Creek, Alaska	Moist acidic tundra	0–30	0.62	11.6	Ping et al., 1997
		30–100	0.83	8.8	
Sag River	Riparian shrubland		1.15	1.00	
			0.98	0.80	
Sag River	Wet acidic tundra		1.03	4.84	
			0.92	2.42	
Happy Valley	Moist acidic tundra		1.00	4.75	
			0.71	8.84	
Atigun River Gorge, Alaska	Alpine	0–30	1.5	2.5	
		30–100	2.03	0.9	
Betty Pingo, Alaska	Marsh	0–30	0.62	21.6	
		30–100	1.53	3.68	
Happy Valley			0.62	21.64	
			1.53	3.68	
Toolik Lake			0.86	9.07	
			1.47	2.70	
Coldfoot, Alaska	Boreal Forest	0–30	0.25	48.59	
		30–100	1.39	5.80	
Smith Lake, Alaska	Bog	0–30	1.08	16.48	
		30–100	1.13	10.3	
Delta Junction			0.66	3.82	
			1.23	1.13	
Nancy, Alaska	Boreal Forest	0–30	0.77	7.64	
		30–100	0.95	0.99	
Pt. MacKenzi, Alaska	Bog	0–30	0.15	52.92	
		30–100	0.37	39.00	
Sukoi			0.56	14.51	
			0.93	3.04	
Tanana Valley, Alaska	Black Spruce	Organic Horizon	0.11	34.5	O'Neil et al. (2003)

Interestingly, the estimate based on the original UNASM shows higher SOCC in Alaska than that in NCSCD does. In Alaska, SOCC derived from the UNASM ranges from 1.5–8.1 g cm<sup>-2</sup> in the northern tundra zone, to 3.7–7.1 g cm<sup>-2</sup> in the interior of Alaska, and 0.67–9.6 g cm<sup>-2</sup> in the southern coastal area, but in the NCSCD, SOCC value ranges from 1.0–8.3 g cm<sup>-2</sup> in the northern tundra zone, to 0.6–2.6 g cm<sup>-2</sup> in the interior of Alaska, and 0.60–9.6 g cm<sup>-2</sup> in the southern coastal area. The estimate derived from the UNASM roughly agrees with the field sampling result of Ping et al. (1997), in which they reported soil carbon content of 6.92, 5.99, and 3.14 g cm<sup>-2</sup>, respectively in three tundra vegetation zones in northern Alaska, 7.87, 1.69, and 12.92 g cm<sup>-2</sup> in the interior forest/taiga zone, and 2.4, 12.6, and 4.37 g cm<sup>-2</sup> in the southern coastal zone. However, Ping et al. (1997) also mentioned in the paper that their result was slightly higher than other studies (Oechel and Billings, 1992; Chapin and Matthews, 1993). The HWSD has much lower SOCC values in Alaska mainly due to insufficient data. To provide a more detailed estimate of soil organic carbon stock in North America, we incorporated the Alaska and Canada soil organic carbon data of NCSCD into the SOCC map derived from the UNASM. The NCSCD-modified UNASM soil organic carbon map is only different from the original UNASM in Alaska and Canada where the UNASM was developed without sufficient data. The rest of the analyses use the NCSCD-modified UNASM soil organic carbon map instead.

Figure 5 illustrates the mean SOCC within each degree latitudinal band for the upper 30 cm and the upper 100 cm soil profile. The NCSCD-modified UNASM soil organic carbon map shows significantly high values within 65°–80° latitudinal bands where the permafrost soil in Alaska and Canada is high in soil organic carbon. The relatively high potential for organic matter storage in cool and humid high latitudinal soils is mainly due to climatic conditions that cause the slow decomposition rate in the balance between carbon inputs and carbon losses (Post et al., 1982; Carter et al., 1997). However, the HWSD illustrates low values in this region due to insufficient data, but shows a peak in 48°–58° latitudinal bands.

Within 48°–25° latitudinal band, the NCSCD-modified UNASM soil organic carbon map is the same as the original UNASM, which is mainly derived from STATSGO2. In this region, the UNASM has lower average SOCC in the top-soil layer due to the high gravel fraction, but shows little discrepancy from the pattern of the HWSD in the upper 100 cm soil profile. Although soil organic carbon concentration, bulk density, and gravel fraction show pronounced regional differences between UNASM and HWSD within the 48°–25° latitudinal bands, the combined effect of these three properties diminishes the differences in the SOCC in the upper 100 cm soil profile. The increase in SOCC within 21°–19° latitudinal bands is the result of relatively higher SOCC in the central and southeastern Mexico.



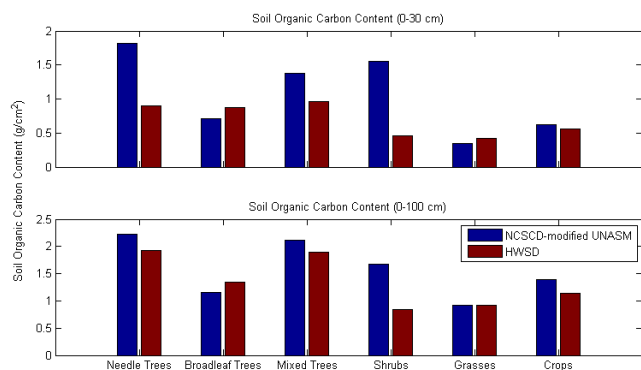
**Fig. 5.** The latitudinal mean SOCC in (a) the 0–30 cm and (b) the 0–100 cm soil profile.

#### 4.3 SOCC summarized by vegetation types

Although the spatial distribution of soil organic carbon is primarily controlled by precipitation, temperature and clay content (Oades, 1988; Sala et al., 1988; Amundson, 1989; Paul, 1984), the type of vegetation or crop and the type of land use or agricultural management system can also influence soil organic matter content (Carter et al., 1997). We calculated the average SOCC for major vegetation types (Fig. 6), including needle leaf trees, broad leaf trees, mixed trees, shrubs, grasses, and crops, based on the International Satellite Land Surface Climatology Project (ISLSCP II) MODIS (Collection 4) IGBP Land Cover, 2000–2001 (Friedl et al., 2002, 2010). Needle leaf trees have the highest average SOCC as the result of slow decomposition under cool temperatures at high latitudes. Broad leaf trees, on the other hand, have lower average SOCC. Shrubs have higher SOCC in the NCSCD-modified UNASM soil carbon map than those in the HWSD, mainly because the shrubs not only live in the mid-latitude semi-arid and arid areas but also exist in high-latitude areas, in which stores a large amount of soil organic carbon. Grasses have the lowest average SOCC values in the NCSCD-modified UNASM. Crops have lower average SOCC values than trees and shrubs but have higher values than grasses.

#### 4.4 SOCM

In the NCSCD-modified UNASM soil carbon map, the total SOCM in the upper 100 cm soil profile is 365.96 Pg, and in the HWSD, the total SOCM is 296.70 Pg. Approximately 74.4 % of the carbon pool in the NCSCD-modified UNASM soil carbon map and 52.2 % of the carbon pool in the HWSD are held in the top 30 cm, the layer which is most prone to changes upon land use/land cover conversion. Table 3 summarizes the SOCM for six major vegetation types, including needle trees, broad leaf trees, mixed trees, shrubs, grasses, and crops. The upper 100 cm soil under needle trees stores



**Fig. 6.** The mean SOCC for major vegetation types in (a) the 0–30 cm and (b) the 0–100 cm soil profile.

about 68.27 Pg carbon mass based on the NCSCD-modified UNASM, and the soil under shrubs is the second largest carbon pool that stores about 51.63 Pg carbon. Based on the estimates from the NCSCD-modified UNASM soil organic carbon map, about 23.1 % of soil organic carbon in the upper 100 cm is stored under trees, about 14.1 % of soil organic carbon is stored under shrubs, and only 1.3 % is stored under grasses and 3.6 % under crops.

## 5 Discussion

### 5.1 The difference between UNASM and HWSD 1.21

We present the 0.25 degree Unified North American Soil Map that combines US STATSGO2, Canada SLC3.2 and SLC2.2, and HWSD1.21. The HWSD 1.21 is the most recent soil dataset, but to our knowledge it is based on measurements from North America that were made in the 1970s. Therefore, the UNASM provides more up-to-date and detailed information for the US and Canada. The pronounced difference between UNASM and HWSD occurs in Alaska and central Canada around the major lakes. The difference between UNASM and HWSD in conterminous US is less obvious than that in Alaska, US and some areas in Canada, but the difference, especially in soil texture and soil organic carbon concentration, may result in the large differences in estimating water and energy budgets and SOCM (Fernandez-Illescas et al., 2001; Abu-Hamdeh, 2003).

### 5.2 The implication on North American soil organic carbon stock

Both UNASM and HWSD lack sufficient data in high latitudes, particularly in northern Canada, where permafrost soils contain a significantly large amount of soil organic carbon (Tarnocai et al., 2009). To provide a more accurate and up-to-date estimate of soil organic carbon stock in North America, we incorporated an Alaska and Canada NCSCD

into a UNASM soil organic carbon map. The NCSCD-modified UNASM soil organic carbon map demonstrates more details in the spatial distribution of SOCC and the large potential of soil organic carbon stock in high latitudinal regions. However, the NCSCD-modified UNASM soil organic carbon map has lower values in Alaska than the SOCC derived from the original UNASM. The average SOCC in the upper 100 cm in Alaska is  $4.0 \text{ g cm}^{-2}$  based on the UNASM and  $1.8 \text{ g cm}^{-2}$  based on the NCSCD-modified UNASM. The average SOCC estimated from NCSCD-modified UNASM is also lower than the other published estimates. For example, Post et al. (1982) reported an average SOCC value of  $2.18 \text{ g cm}^{-2}$  for the Arctic tundra region. Ping et al. (2008) reported average SOCC to be  $3.48 \text{ g cm}^{-2}$  for 100 cm depth in Alaska. Johnson et al. (2011) stratified the state of Alaska into different ecoregions and reported average SOCC (for 100 cm depth) of  $5.33 \text{ g cm}^{-2}$ ,  $0.86 \text{ g cm}^{-2}$ ,  $2.1 \text{ g cm}^{-2}$ , and  $2.4 \text{ g cm}^{-2}$  for arctic tundra, intermontane boreal, Alaska range transition, and coastal rainforests, respectively. Mishra and Riley (2012) predicted SOCC in Alaska using spatially referenced environmental variables and pedon observations, resulting in the average SOCC of  $3.54 \text{ g cm}^{-2}$  for the active layer and  $5.36 \text{ g cm}^{-2}$  for the whole profile. The average SOCC estimated from the original UNASM is close to the case 3 ( $4.49 \text{ g cm}^{-2}$ ) result reported by Bliss and Maursestter (2010), probably because both the UNASM and the case 3 in Bliss and Maursestter (2010) used the soil order to link STATSGO soil map with soil property information. The differences with the estimates of Alaska average SOCC from these studies can be attributed to the number and quality of the pedon observations used to develop the dataset, the way to develop the soil map and estimate SOCC, and the spatial resolution of the soil map. The total SOCM across Alaska is 48.07 Pg based on the UNASM and 24.52 Pg based on the NCSCD-modified UNASM soil organic carbon map. In the conterminous US, the SOCC derived from the UNASM agrees with the one based on HWSD in most areas, except in the eastern US Minnesota, and Iowa. The UNASM shows lower SOCC in eastern US, but higher SOCC in Minnesota and Iowa.

The average SOCC for the major biome types in the NCSCD-modified UNASM roughly agrees with the result of Post et al.'s (1982) study, except shrubs in NCSCD-modified UNASM have much higher SOCC values than the cool temperate bush in Post et al.'s (1982) study. Needle trees have higher mean SOCC than the other vegetation types mainly due to the slow decomposition resulted from the cool weather. However, there are very few studies of the impact of vegetation on the spatial distribution of soil organic carbon, because vegetation and climate typically covary in a complex fashion to impact the accumulation of soil organic carbon. Post et al. (1982) estimated soil carbon density for vegetation life zones and studied its relationship with climatic factors. Quideau et al. (2001a, b) suggested that the mosaic of vegetation has significant impact on the accumulation and

**Table 3.** The total SOCM and SOCM for major vegetation types in the top 0–30 cm and the 0–100 cm soil profile in North America derived from the NCSCD-modified UNASM soil carbon map and HWSD 1.21.

Soil Layer	Needle Trees (Pg)	Broad Leaf Trees (Pg)	Mixed Trees (Pg)	Shrubs (Pg)	Grasses (Pg)	Crops (Pg)	Total (Pg)
NCSCD-modified UNASM Soil Organic Carbon Map							
0–30 cm	53.08	8.02	2.37	44.43	1.73	5.72	272.50
0–100 cm	68.27	12.93	3.36	51.63	4.64	13.04	365.96
HWSD 1.21							
0–30 cm	28.45	9.14	1.58	15.22	1.93	5.05	154.86
0–100 cm	59.37	14.55	3.06	29.17	4.21	10.28	296.70

turnover of soil organic matter directly by determining the quality and the pathway of biomass incorporation into soil. Given the limited studies of the impact of vegetation on soil organic carbon spatial distribution, our results can only be explained by the climate factors, but more studies on finer scales are needed to evaluate the effect of vegetation on soil organic carbon spatial distribution under the same climatic condition.

The total SOCM in the upper 100 cm soil profile is 365.96 Pg based on the NCSCD-modified UNASM, higher than the estimates of 296.70 Pg derived from the HWSD 1.21. Since the NCSCD does not provide other soil properties information required for terrestrial biosphere modeling except for SOCC, the NCSCD-modified UNASM soil organic carbon map is a separated dataset from the UNASM. Although there are estimates for various subregions of the North America domain (Kern, 1994; Bliss and Maurestter, 2010; Johnson et al., 2011; Mishra and Riley, 2012), the estimate based on UNASM is the first and the most up-to-date estimate of soil organic carbon in North America based on the observed soil type distribution, combining the latest national soil surveys of the US and Canada. It provides more details of spatial distribution of soil organic carbon in North America and can be used as a benchmark for the terrestrial biosphere modeling research.

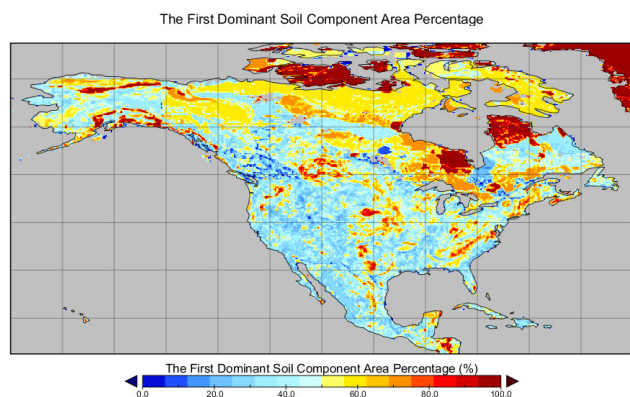
### 5.3 The limitation and the uncertainty of UNASM

The UNASM is created for use in terrestrial biosphere modeling. Given the limited resources, the UNASM is developed at 0.25 degree in latitude and longitude, which limits the flexibility for users to downscale to any spatial resolution. Moreover, the UNASM is derived from different sources that use different soil taxonomies, and thus the current version of the UNASM does not provide uniform soil taxonomy.

Quantifying the uncertainty of the UNASM is quite complicated, as it is developed from four data sources without uncertainty for any of them. In this paper, we only use a simple index as the approximation of the certainty of the UNASM. We provide the percentage of the area of the

dominant component selected for each cell to the total land area of the cell. This percentage can be served to quantify the extent to which the selected dominant soil component represents the cell. The higher percentage implies that the selected soil component can be more representative to the soil properties within the cell and the soil information assigned for the cell is of more confidence. We found that 51.3 % of all land cells have the 1st dominant soil component that occupies over 50 % of the land area of the cell. The 1st dominant soil component selected for most of land cells can represent 40–70 % of the cell area (Fig. 7). The percentage that the dominant soil component occupies over 50 % of the land area of the cell increases to 85.2 % if we select the 1st and the 2nd dominant soil components for each cell, and it increases to 95.2 % if we select the three most dominant soil components for each cell. In other words, the 1st dominant soil component can represent about half of the land cells in UNASM, while three most dominant soil components can represent almost all land cells in UNASM. However, the current terrestrial biosphere models can only take one value of each property for each cell. In addition, soil information cannot be averaged across different soil types because a specific soil type is associated with specific soil properties. Though we provide soil information of the 1st, 2nd, and 3rd dominant soil components for each cell, respectively, it is hard to use them in modeling. Therefore, we will not provide further analysis about the variability of soil information determined by selecting different soil components for each cell in this paper. In terms of the source data quality, we have relatively higher confidence in the area based on STATSGO2 and SLC3.2, and lower confidence in the area based on SLC2.2 and the area filled with HWSD1.21. Especially the high-latitude area filled with HWSD 1.21 is of less confidence due to the lack of observations when developing HWSD.

For the future work, one useful activity would be to incorporate more soil properties, so that this soil dataset can be used not only in terrestrial ecosystem modeling but also in hydrologic and agricultural applications. Another useful activity would be a comprehensive sensitivity analysis of



**Fig. 7.** The percentage of the area of the first dominant soil component to the area of the cell.

terrestrial ecosystem model performance in relation to different soil inputs.

## 6 Conclusions

The North American soil map presented here fills a need of the modeling community for a dataset of soil physical properties specifically created for North America by combining state-of-the-art soil information from STATSGO2, two versions of SLCs, and HWSD1.21. The comparison between the UNASM and the HWSD illustrates the pronounced difference in the spatial distributions of soil properties and soil organic carbon, but the UNASM overall provides higher-quality and more detailed information particularly in Alaska and central Canada. The total SOCM in the upper 100 cm soil profile is 365.96 Pg based on the NCSCD-modified UNASM, among which 23.1 % is under trees, 14.1 % is under shrubs, and 1.3 % is under grasses and 3.6 % under crops. These new estimates will help provide a more reliable prediction for the effect of global climate change and land use management on the carbon budget.

In order to provide end users the most functional dataset, the UNASM is in uniform 0.25 degree grid. The UNASM is in NetCDF format to meet the input requirement of most terrestrial biosphere models. This data will be archived in Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics (ORNL DAAC) (Liu et al., 2013).

## Appendix A

### The description of soil properties in the UNASM

#### A1 Sand, silt, and clay content

Sand, silt, and clay content, defined as the percentage of each size class based on weight, are the most important soil

attributes to quantify soil texture. Sand comprises particles or granules ranging in diameter from 0.05 mm to 2.0 mm. Silt size is between 0.002 and 0.0625 (In STATSGO2, the silt size is between 0.002 and 0.05 mm). Clay is composed primarily of fine-grained with the diameter less than 0.002 mm.

When harmonizing the unified soil dataset into two standard layers (the topsoil layer is from 0–30 cm, and the subsoil layer is from 30–100 cm), sand, silt and clay fractions were calculated as

$$\text{SandSiltClay \%} = \frac{M_{\text{sand, silt, clay}}}{M_{\text{min}}} \times 100 \%, \quad (\text{A1})$$

where  $M_{\text{sand, silt, clay}}$  represents the mass of sand, or silt, or clay in the standard layer, and  $M_{\text{min}}$  represents the total mass of mineral soil in the standard layer.  $M_{\text{min}}$  is determined as the product of the bulk density of mineral soil, the fraction of mineral soil (which is 1 minus the fraction of the organic matter), and the volume of the standard soil layer.

#### A2 Gravel content

Gravel content is the volume percentage gravel (diameter > 2 mm) in the soil. The depth-weighted average method was used to interpolate gravel content of different layers into two standard layers in the UNASM.

#### A3 Bulk density

Bulk density is the ratio of the mass of soil material to the total volume of solids plus pores. It is often used to estimate soil hydrological properties and to calculate the total amount of soil carbon. The reference bulk density measured at 0.33 bar water confront from STATSGO2 is used for the UNASM. Since bulk density is measured and estimated using different methods, potential bias among the bulk density measurements from different sources might exist.

When harmonizing the unified soil dataset into two standard layers, bulk density is calculated as:

$$\text{BD} = \frac{M_{\text{org}} + M_{\text{min}}}{V}, \quad (\text{A2})$$

where  $M_{\text{org}}$  and  $M_{\text{min}}$  are the mass of organic matter and mineral matter in the standard soil layer, and  $V$  is the volume of the standard layer.  $M_{\text{org}}$  is determined as the product of bulk density of organic matter, the fraction of organic matter, and the volume.  $M_{\text{min}}$  is determined as the product of the bulk density of mineral soil, the fraction of mineral soil (which is 1 minus the fraction of the organic matter), and the volume.

#### A4 Soil organic carbon

Soil organic carbon concentration is defined as the percentage of soil weight. However, estimates of soil organic carbon stock generally refer to a given depth of soil. Therefore,



the amount of soil in a given depth depends not only on soil organic carbon content as provided by this dataset, but also depends on bulk density, area, and depth.

The mass-weighted average of soil organic carbon concentration (OC) was calculated when the unified soil dataset was harmonized into two standard layers:

$$\text{OC} = 0.58 \times \frac{M_{\text{org}}}{M_{\text{org}} + M_{\text{min}}} \times 100, \quad (\text{A3})$$

where  $M_{\text{org}}$  and  $M_{\text{min}}$  are the mass of organic matter and mineral matter, respectively, the same as the ones used to calculate bulk density. The constant 0.58 in the above equation assumes that soil organic carbon accounts for 58% of soil organic matter (Mann, 1986).

## A5 pH

pH is a measure for the acidity and alkalinity of the soil. In the HWSD, pH is measured in a soil-water solution. However, in SLCs, pH is measured in calcium chloride ( $\text{CaCl}_2$ ). In STATSGO2, pH is measured in both ways. To keep the consistency with HWSD, we use pH measured in  $\text{H}_2\text{O}$  in the UNASM. Miller and Kissel's (2010) study demonstrated that pH measured in  $\text{H}_2\text{O}$  is significantly linearly related with pH measured in  $\text{CaCl}_2$ , only slightly diverging from the 1 : 1 line. The depth-weighted average method was used to interpolate pH values of different layers into two standard layers in the UNASM.

## A6 Soil depth

Here soil depth is the maximum soil depth of each cell before harmonization to standard layers. Maximum soil depth can be used as the approximate measure of the depth-to-bedrock. However, in the HWSD, the maximum soil depth is either 30 cm or 100 cm, which cannot represent the soil thickness in most cells.

*Acknowledgements.* The authors would like to acknowledge the financial support from the NASA's Terrestrial Ecology Program (TEP) for the Modeling and Synthesis Thematic Data Center (MAST-DC: project 09-TE09-26). We would like to thank the core group from the NASA TEP-funded Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP) for providing helpful suggestions to this study. This study was part of the North American Carbon Program.

Edited by: N. Ohte

## References

- Abu-Hamdeh, N. H.: Thermal properties of soils as affected by density and water content, *Biosystems Eng.*, 86, 97–102, 2003.
- Amundson, R. G., Chadwick, O. A., and Sowers, J. M.: A comparison of soil climate and biological activity along an elevational gradient in the eastern Mojave Desert, *Oecologia*, 80, 395–400, 1989.
- Batjes, N. H.: Development of a world data set of soil water retention properties using pedotransfer rules, *Geoderma* 71, 31–52, 1996.
- Batjes, N. H.: A world dataset of derived soil properties by FAO-UNESCO soil unit for global modeling, *Soil Use Manage.*, 13, 9–16, 1997.
- Bliss, N. and Maursetter, J.: Soil organic carbon stocks in Alaska estimated with spatial and pedon data., *Soil Sci. Soc. Am. J.*, 74, 565–579, 2010.
- Bouswman, A. F., Fung, I. E., and Matthews, J. J.: Global analysis of the potential for  $\text{N}_2\text{O}$  production in natural soils, *Global Biogeochem. Cy.*, 7, 557–597, 1993.
- Carbon Cycle Science Program (CCSP) The First State of the Carbon Cycle Report (SOCCR): The Northma American Carbon Budget and Implications for the Global Carbon Cycle, A report by the US Climate Change Science Program and the Subcommittee on Global Change Research, edited by: King, A. W., Dilling, L., Zimmerman, G. P., Fairman, D. M., Houghton, R. A., Marland, G., Rose, A. Z., and Wilbanks, T. J., National Oceanic and Atmospheric Administration, National Climate Data Center, Asheville, NC, USA, 242, 2007.
- Carter, M. R., Angers, D. A., Gregorich, E. G., and Bolinder, M. A.: Organic carbon and nitrogen stocks and storage profiles in cool, humid soils of eastern Canada, *Can. J. Soil Sci.*, 77, 205–210, 1997.
- Centre for Land and Biological Resources Research: Soil Landscapes of Canada, v.2.2, Research Branch, Agriculture and Agri-Food Canada, Ottawa, 1996.
- Chapin III, F. S. and Matthews, E.: Boreal carbon pools: Approaches and constraints in global extrapolations, in: *Proceedings of the International Workshop on Carbon Cycling in Boreal Forest and Sub-arctic Ecosystems: Biospheric responses and feedbacks to global climate change*, Corvallis, Oregon, US, 9–20, 1993.
- Cramer, W. and Fischer, A.: Data requirements for global terrestrial ecosystem modeling, in: *Global change and terrestrial ecosystem*, Cambridge University, Cambridge, UK, 1996.
- FAO/IIASA/ISRIC/ISS-CAS/JRC: Harmonized World Soil Database (version 1.21), FAO, Rome, Italy and IIASA, Laxenburg, Austria, 2013.
- Fernandez-Illescas, C. P., Porporato, A., Laio, F., and Rodriguez-Iturbe, I.: The ecohydrological role of soil texture in a water-limited ecosystem, *Water Resour. Res.*, 37, 2863–2872, 2001.
- Fixen, P. E.: Soil test levels in North America, *Better Crops*, 86, 12–15, 2002.
- Foley, J. A., Prentice, I. C., Ramankutty, N., Levis, S., Pollard, D., Sitch, S., and Haxeltine, A.: An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics, *Global Biogeochem. Cy.*, 40, 603–628, 1996.
- Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C.: Global

- Land Cover Mapping from MODIS: Algorithms and Early Results, *Remote Sens. Environ.*, 83, 287–302, 2002.
- Friedl, M. A., Strahler, A. H., and Hodges, J.: ISLSCP II MODIS (Collection 4) IGPB Land Cover, 2000–2001, in: ISLSCP Initiative II Collection, Data set, edited by: Hall, G., Forrest, G., Colatz, G., Meeson, B., Los, S., Brown de Colstoun, E., and Landis, D., available at: <http://daac.ornl.gov/> from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, doi:10.3334/ORNLDAAAC/968, 2010.
- Gijsman, A. J., Thornton, P. K., and Hoogenboom, G.: Using the WISE database to parameterize soil inputs for crop simulation models, *Comput. Electron. Agr.*, 56, 85–100, 2007.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.: Towards robust regional estimates of CO<sub>2</sub> sources and sinks using atmospheric transport models, *Nature*, 415, 626–630, 2002.
- Haxeltine, A., Prentice, I. C., and Creswell, I. D.: A coupled carbon and water flux model to predict vegetation structure, *J. Veg. Sci.*, 7, 651–666, 1996.
- Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., and Swanson, D. K.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, *Earth Syst. Sci. Data*, 5, 3–13, doi:10.5194/essd-5-3-2013, 2013.
- Jobbágy, E. G. and Jackson, R. B.: the vertical distribution of soil organic carbon and its relation to climate and vegetation, *Ecol. Appl.*, 10, 423–436, 2000.
- Johnson, K. D., Harden, J., McGuire, A. D., Bliss, N. B., Bockheim, J. G., Clark, M., Mettleton-Hollingsworth, T., Jorgenson, M. T., Kane, E. S., Mack, M., O'Donnell, J., Ping, C. E., Schuur, A. G., Turetsky, M. R., and Valentine, D. W.: Soil carbon distribution in Alaska in relation to soil-forming factors, *Geoderma*, 167, 71–84, 2011.
- Kern, J. S.: Spatial patterns of soil organic carbon in the contiguous United States, *Soil Sci. Soc. Am. J.*, 58, 439–455, 1994.
- Kern, J. S.: Geographic patterns of soil water-holding capacity in the contiguous United States, *Soil Sci. Soc. Am. J.*, 59, 1126–1133, 1995.
- Kimball, J. S., White, M. A., and Running, S. W.: BIOME-BGC simulations of stand hydrologic process for BOREAS, *J. Geophys. Res.*, 102, 29043–29051, 1997.
- Kittel, T. G. F., Rosenbloom, N. A., Painter, T. H., and Schimel, D. S.: The VEMAP integrated database for modeling United States ecosystem/vegetation sensitivity to climate change, *J. Biogeogr.*, 22, 857–862, 1995.
- Liu, S., Post, W. M., Wei, Y., and Cook, R. B.: NACP: MsTMIP Unified North American Soil Map, Data set, available at: <http://daac.ornl.gov/> from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, in press, 2013.
- Mann, L. K.: Changes in soil carbon storage after cultivation, *Soil Sci.*, 142, 279–288, 1986.
- McGuire, A. D., Joyce, L. A., Kichlighter, D. W., Melillo, J. M., Esser, G., and Vorosmarty, C. J.: Productivity response of climax temperate forests to elevated temperature and carbon dioxide: a North American comparison between two global models, *Clim. Change*, 24, 287–310, 1993.
- Michaelson, G. J., Ping, C. L., and Kimble, J. M.: Carbon storage and distribution in tundra soils of Arctic Alaska, USA, *Arct. Alp. Res.*, 28, 414–424, 1996.
- Mishra, U. and Riley, W. J.: Alaskan soil carbon stocks: spatial variability and dependence on environmental factors, *Biogeosciences*, 9, 3637–3645, doi:10.5194/bg-9-3637-2012, 2012.
- Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey, A. A.: The Multi-institution North American Land Data Assimilation System (NLDAS): Utilizing Multiple GCIP Products and Partners in a Continental Distributed Hydrological Modeling System, *J. Geophys. Res.*, 109, D07S90, doi:10.1029/2003JD003823, 2003.
- Miller, D. A. and White, R. A.: A conterminous United States multi-layer soil characteristics dataset for regional climate and hydrology modeling, *Earth Interactions*, 2, 1–26, doi:10.1175/1087-3562(1998)002<0001:ACUSMS>2.3.CO;2, 1998.
- Miller, R. O. and Kissel, D. E.: Comparison of soil pH methods on soils of North America, *Soil Sci. Soc. Am. J.*, 74, 310–316, 2010.
- Neilson, R. P.: A model for predicting continental scale vegetation distribution and water balance, *Ecol. Appl.*, 5, 362–386, 1995.
- Oades, J. M.: The retention of organic matter in soils, *Biogeochem.*, 5, 33–70, 1988.
- Oechel, W. C. and Billings, W. D.: Effects of global change on the Carbon balance of Arctic plants and ecosystems, in: *Arctic Ecosystems in a Changing Climate*, Academic Press, 139–168, 1992.
- O'Neil, K., Kasischke, P. E. S., and Richter, D. D.: Seasonal and decadal patterns of soil carbon uptake and emission along an age sequence of burned black spruce stands in interior Alaska, *J. Geophys. Res.*, 108, 8155, doi:10.1029/2001JD000443, 2003.
- Parton, W. J., Schimel, D. S., Cole, C. V., and Ojima, D. S.: Analysis of factors controlling soil organic levels of grasslands in the Great Plains, *J. Soil Sci. Soc. Am.*, 51, 1173–1179, 1987.
- Paul, E. A.: Dynamics of soil organic matter, *Plant Soil*, 76, 275–285, 1984.
- Ping, C. L., Michaelson, G. J., and Kimble, J. M.: Carbon storage along a latitudinal transect in Alaska, *Nutr. Cycl. Agroecosys.*, 49, 235–242, 1997.
- Ping, C. L., Michaelson, G. J., Jorgenson, M. T., Kimble, J. M., Epstein, H., Romanovsky, V. E., and Walker, D. V.: High stocks of soil organic carbon in the North American arctic region, *Nat. Geosci.*, 1, 615–619, 2008.
- Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., Solomon, A. M.: A global biome model based on plant physiology and dominance, soil properties and climate, *J. Biogeogr.*, 19, 117–134, 1992.
- Post, W. M., Emanuel, W. R., Zinke, P. J., and Stangenberger, A. G.: Soil carbon pools and world life zones, *Nature*, 298, 156–159, doi:10.1038/298156a0, 1982.
- Potter, C. S., Matson, P. A., Vitousek, P. M., and Davidson, E. A.: Process modeling of controls on nitrogen trace gas emissions from soils world wide, *J. Geophys. Res.*, 101, 1361–1377, 1996.
- Prentice, I. C., Farquhar, G. D., Fasham, M. J. R., Goulden, M. L., Heimann, M., Kheshi, H. S., Quere, Le, C., Scholes, R. J., Wal-

- lace, D. W. R., Archer, D., Ashmore, M. R., Aumont, O., Baker, D., Battle, M., Bender, M., Bopp, L. P., Bousquet, P., Caldeira, K., Ciais, P., Cramer, W., Dentener, F., Enting, I. G., Field, C. B., Holland, E. A., Houghton, R. A., House, J. I., Ishida, A., Jain, A. K., Janssens, I., Joos, F., Kaminski, T., Keeling, C. D., Kicklighter, D. W., Kohfeld, K. E., Knorr, W., Law, R., Lenton, T., Lindsay, K., Maier-Reimer, E., Manning, A., Matear, R. J., McGuire, A. D., Melillo, J. M., Meyer, R., Mund, M., Orr, J. C., Piper, S., Plattner, K., Rayner, P. J., Sitch, S., Slater, R., Taguchi, S., Tans, P. P., Tian, H. Q., Weirig, M. F., Whorf, T., and Yool, A.: Climate change 2001: the scientific basis: contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Houghton, J. T., Cambridge, Cambridge University Press, 183–237, 2001.
- Quideau, S. A., Chadwick, O. A., Trumbore, S. E., Johnson-Maynard, J. L., Graham, R. C., and Anderson, M. A.: vegetation control on soil organic matter dynamics, *Organic Geochem.*, 32, 247–252, 2001a.
- Quideau, S. A., Chadwick, O. A., Beines, A., Graham, R. C., and Anderson, M. A.: A direct link between forest vegetation type and soil organic matter composition, *Geoderma*, 104, 41–60, 2001b.
- Raich, J. W. and Schlesinger, W. H.: The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate, *Tellus*, 44, 81–99, 1992.
- Sala, O. E., Parton, W. J., Joyce, L. A., and Lauenroth, W. K.: Primary production of the central grassland region of the United States, *Ecology*, 69, 40–45, 1988.
- Saxton, K. E., Rawls, W. J., Romberger, J. S., and Papendick, R. I.: Estimating generalized soil-water characteristics from texture, *Soil Sci. Soc. Am. J.*, 50, 1031–1036, 1986.
- Sellers, P. J., Mintz, Y., Sud, Y. C., and Dalcher, A.: A Simple Biosphere Model (SiB) for use within general circulation models, *J. Atmos. Sci.*, 43, 505–531, 1986.
- Simanton, J. R., Renard, K. G., Christiansen, C. M., and Jane, L. J.: Spatial distribution of surface rock fragments along catenas in Semiarid Arizona and Nevada, USA, *Catena*, 23, 29–42, 1994.
- 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., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, *Global Change Biol.*, 9, 161–185, 2003.
- Soil Landscapes of Canada Working Group: Soil Landscapes of Canada version 3.2. Agriculture and Agri-Food Canada, (digital map and database at 1:1 million scale), 2010.
- Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture, US General Soil Map (STATSGO2), available at: <http://soildatamart.nrcs.usda.gov>, last access: August 2011.
- Sombroek, W. G., Nachtergaele, F. O., and Hebel A.: Amounts, dynamics and sequestrations of carbon in tropical and subtropical soils, *Ambio*, 22, 417–426, 1993.
- Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region, *Global Biogeochem. Cy.*, 23, GB2023, doi:10.1029/2008GB003327, 2009.
- Vasek, F. C.: Creosote Bush: long-lived clones in the Mojave Desert, *Am. J. Bot.*, 67, 246–255, 1980.
- Webb, R. S., Rosenzweig, C. E., and Levine, E. R.: Specifying land surface characteristics in general circulation models: Soil profile data set and derived water-holding capacities, *Global Biogeochem. Cy.*, 7, 97–108, 1993.
- Wofsy, S. C. and Harris, R. C.: The North American Carbon Program (NACP), Report of the NACP Committee of the US inter-agency Carbon Cycle Science Program, US Global Change Research Program, 2002.
- Young, D. R. and Nobel, P. S.: Predictions of soil-water potentials in the north-western Sonoran Desert, *J. Ecol.*, 74, 143–154, 1986.
- Zobler, L.: A world soil file for global climate modeling, NASA Tech. Memo, 87802, 32–14, 1986.