Chapter 10
Overview of the U.S. Rapid Carbon Assessment Project: Sampling Design, Initial Summary and Uncertainty Estimates

Skye Wills, Terrance Loecke, Cleiton Sequeira, George Teachman, Sabine Grunwald, and Larry T. West

Abstract The Rapid Carbon Assessment (RaCA) project was undertaken by the Soil Science Division of the Natural Resource Conservation Service (NRCS) to capture baseline soil carbon stocks across the conterminous US (CONUS). A multi-level hierarchical design was used to ensure that samples were distributed across regions, soils and land use/land cover classes (LULC). Within those strata, sites were selected at random locations where five pedons were described and sampled at 0–5 cm and by genetic horizon from 5 to 100 cm. A total of 6,148 sites, 32,084 pedons and 144,833 samples were described. Bulk density was calculated for samples from the upper 50 cm and predicted for deeper samples using pedon and horizon information in a regression tree developed with random forests. Soil organic carbon (SOC) concentration was predicted for each sample using processed Visible-Near Infrared spectra and a random forest model. Pedon SOC stocks were calculated by fixed depth to 100 cm. Expected variance was introduced into the stock calculations using analytical and modeling prediction errors (e.g., SOC concentration and bulk density measurements) and the stratified sampling design was partitioned using a hierarchical Bayesian modeling approach. Pedons were averaged by site. The mean of all RaCA site SOC stocks to 100 cm was 321.1, with a median

S. Wills (✉) • G. Teachman • L.T. West
National Soil Survey Center, Natural Resource Conservation Service – USDA,
100 Centennial Mall North, Federal Building, Room 152, MS 35, Lincoln, NE 68508, USA
e-mail: skye.wills@lin.usda.gov

T. Loecke
University of Nebraska-Lincoln, 616 Hardin Hall, 3310 Holdrege Street,
Lincoln, NE 68583-0996, USA
e-mail: tloecke2@unl.edu

C. Sequeira
Koch Industries, Wichita, KS, USA

S. Grunwald
University of Florida, Gainesville, FL 32611, USA

A.E. Hartemink and K. McSweeney (eds.), *Soil Carbon*. Progress in Soil Science,
DOI 10.1007/978-3-319-04084-4_10,

of 173.3 and range of 2 to over 5,000 Mg ha^{-1}. Geometric means of soil groups and LULC classes were used to extrapolate results to all assessed areas. Further work is needed to properly weight averages by areal extent and assess the cause of higher than expected site SOC stock values.

Keywords Soil organic carbon • Carbon stocks • Uncertainty

Introduction

Understanding soil C stocks and the uncertainty around these stocks is necessary for national C inventories (IPCC 2007), inputs to earth system models (Todd-Brown et al. 2013), and to determine soil organic carbon (SOC) temporal changes (Kravchenko and Robertson 2011). Current assessments of national scale carbon stocks rely on a variety of methods. Most efforts use published maps of ecosystem, climate and soils (e.g. Eswaran et al. 1993; Post et al. 1982; Guo et al. 2006), sometimes legacy pedons are used (e.g. Kern 1994; Wu et al. 2003) and some supplement those with additional sampling (e.g. Bellamy et al. 2005; Bradley et al. 2005). Recent estimates of SOC for the conterminous US (CONUS) (Guo et al. 2006) are based on the STATSGO database which provides an upper and lower range for SOC for each map unit. No probability distributions are assigned to the STATSGO ranges, thus it is unknown what these uncertainty estimates represent or if they are narrow or broad relative to reality. Soil C stocks in temperate regions are dynamic (Bellamy et al. 2005); however, the underlying soil data for the STATSGO database were collected over several decades (1950s–2000s, L. T. West personal communication) and thus the STATSGO SOC estimates may not represent a static inventory. The Rapid Carbon Assessment (RaCA) project was undertaken by the Soil Science Division of the Natural Resource Conservation Service (NRCS) to capture baseline soil carbon stocks across the CONUS at a single point in time using a robust sampling design.

Project Sampling Design

The initial RaCA project emphasis was on soil organic carbon (SOC) stocks, or the amount of SOC in a certain volume (area and depth of soil). The effort was conceptualized by staff at the National Soil Survey Center (NSSC) and carried out by Natural Resource Conservation Service (NRCS) soil scientists. A multi-level stratified random sampling scheme was created to both maximize geographical/spatial sample coverage with a maximum of conditions represented while giving a framework for aggregating information into regional areas. The first level of strata were RaCA regions based on major land resource area (MLRA) regional offices (USDA 2010). Within each region, samples were further stratified by a combination of soil groups and land use/land cover classes.

Table 10.1 National Land Cover Dataset (NLCD) codes and corresponding RaCA land use – land cover (LULC) classes

NLCD code	NLCD class	RaCA LULC class
11	Open water	
12	Perennial ice/snow	
21	Developed – open space	
22	Developed – low intensity	
23	Developed – medium intensity	
24	Developed – high intensity	
31	Barren land	
41	Deciduous forest	Forestland
42	Evergreen forest	Forestland
43	Mixed forest	Forestland
51	Dwarf scrub	Rangeland
52	Shrub/scrub	Rangeland
71	Grassland/herbaceous	Rangeland[a]
81	Pasture hay	Pastureland
82	Cultivated crop	Cropland
90	Woody wetlands	Wetland
95	Emergent herbaceous wetlands	Wetland

RaCA classes reflect NRI classes and definitions
[a]In the Eastern CONUS (regions 11, 12, 14, and 18), the Grassland/herbaceous NLCD class was assigned to pasture land instead of rangeland to better match NRI assessments

Soil groups were created using a statistical algorithm as described by Wills et al. (2013). Soil information from official series descriptions (Soil Survey Staff 2010a) and the soil data access (SDA) portal (Soil Survey Staff 2010b) was translated into scores that related to the amount of SOC the soil was expected to contain. The scores were then used in a statistical clustering algorithm to create 8–20 groups for each RaCA region.

Land use-land cover (LULC) information was assigned in order to align with NRCS National Resource Inventory (NRI) classes and definitions (USDA 2007). In order to obtain complete spatial coverage of LULC, the national land cover dataset (NLCD) was used (Fry et al. 2011). NLCD classes were relabeled as RaCA land use/cover (LULC) classes to correspond with the classes used in NRI (Table 10.1). There was some geographic variation in the correspondence of NRI and NLCD classes. In the Eastern CONUS (regions 11, 12, 14, and 18), the Grassland/herbaceous NLCD class was assigned to pasture land instead of rangeland to better match NRI assessments.

The pool of potential RaCA sites was created using the NRI sampling framework (Nusser et al. 1998). The primary sampling units of NRI are arranged randomly within geographic strata in a way that provides complete coverage of CONUS. One point was randomly generated within each primary sampling unit. The soil group was assigned by performing a spatial join with SSURGO (Soil Survey Staff 2010b). The land use/cover of the nearest NRI point was used to assign LULC. In cases where that was not possible (within areas of federal lands that are not assessed for NRI), a spatial join with the NLCD coverage was used to assign a LULC class.

The number of points in each soil group and LULC strata was determined using their relative extent within each region. Soil group and LULC strata were attached to a raster that combined SSURGO (Soil Survey Staff 2010b) as of October 2010 – current versions available as gSSURGO (Soil Survey Staff 2013a) – and NLCD on a 10 m grid.

A target number of sites to sample per soil group and LULC along with a randomized list of potential sites were supplied to each region. An excess of sites was supplied so that as sites were rejected (due to lack of access, safety or unexpected land use/cover) they could be replaced by the next random site on the list.

RaCA Site Data Collection and Pedon Sampling

Five pedons were sampled at each site: one at the plot center and one 30 m away in each cardinal direction (unless the arrangement was altered due to obstructions such as fences or roads). Each pedon was described according to the Field Book for Describing and Sampling Soils v.2 (Schoeneberger et al. 2002) and assigned to the most likely soil series given the information available. Minimum required information for each horizon included: horizon nomenclature, depths, color, texture, rock fragment modifier (% volumetric coarse fragments), redox features, and structure (where possible). Small pits were excavated to 50 cm or a root limiting layer such as bedrock or cemented soil. Samples were collected from 0 to 5 cm and from 5 to 50 cm by genetic horizon. Probes or augers were used to sample genetic horizons from 50 to 100 cm. Volumetric samples were collected for samples from 0 to 50 cm in the most appropriate manner . Samples were labeled, sealed in air-tight bags and transported to the regional office for processing. Complete sampling instructions can be found at the RaCA website (Soil Survey Staff 2013b).

Samples were air dried and sieved to a size of <2 mm. A sub-sample was oven dried to calculate bulk density as given in the Soil Survey Lab. Manual (Burt 2004). A LabSpec® 2500 spectrometer (Analytical Spectral Devices, Boulder, CO), which we refer to as VNIR, was used to scan all mineral samples. Each region had an identical VNIR and measured both reference and high/low QC sample checks in order to maintain consistency and comparability across regions. Organic horizon samples and 3 % of mineral samples were sent the Kellogg Soil Survey Laboratory. Those samples were scanned with another VNIR and carbon was measured according to the Soil Survey Laboratory Methods Manual (Burt 2004). Total carbon was measured with dry combustion and inorganic carbon content was measured by manometer after HCl treatment.

For samples that were not collected volumetrically (including all samples collected below 50 cm), a suite of pedotransfer function (PTFs) were developed to predict bulk density (Sequeira et al. 2014a). Twelve PTFs were developed using 2,680 pedons (20,045 horizons) from the National Cooperative Soil Survey characterization database (NCSS 2011). All bulk density measurements in the NCSS database were determined at −33 kPa matric potential using the clod method (Burt 2004). The PTFs used information known to be available for samples collected using the RaCA protocols. The PTFs had 9 input variables (regressors): 4 from the horizon for which bulk density would be predicted [designation, textural class, depth (at the middle of the horizon), and thickness] and 5 from a neighbor horizon (bulk density, horizon designation, textural class, depth, and thickness). The Random Forest algorithm (Breiman 2001) was used to develop the PTFs. The accuracy of these models was good, with predictions errors of 0.10–0.15 g cm^{-3} (Sequeira et al. 2014a). Huang et al. (2003) reported that the prediction error should be below 20 % to be declared acceptable for bulk density prediction. The measured bulk density data were used when available with outliers removed if the measured bulk density was outside of the 1st and 99th percentiles of samples from the NCSS database by simplified horizon nomenclature. The bulk density of all other samples was estimated with the PTFs described above.

Similarly, the NCSS characterization data were used to create carbon prediction models using the spectra of each sample's VNIR scan. The VNIR scans were taken on air-dry, <2 mm samples using the LabSpec® 2500 spectrometer (Analytical Spectral Devices, Boulder, CO) with spectral range of 350–2,500 nm, acquired at 1 nm increments. Preprocessing methods were applied to the spectra to reduce non-constituent-related interferences (e.g., light scattering, light path-length, spectrum baseline drift) that decrease the signal-to-noise ratio. Additionally, the artifact bands generated around the edge of the three detectors built-in the spectrometer were also removed to improve model performance (Duckworth 2004). Two multi-variate techniques, partial least squares regression and random forest regression trees were used to model SOC from VNIR scans modified with a variety of pre-processing techniques.18 VNIR-based SOC prediction models were tested. The best performing models used random forest with VNIR spectra that had been preprocessed by clipping out artifact bands and smoothing with using a Savitzky-Golay first derivative (Sequeira et al. 2014b).

No satisfactory model was developed for predicting SOC with VNIR spectra in organic horizons. Therefore, all O horizons were sent to the Kellogg Soil Survey lab for SOC determination by combustion (Burt 2004). When samples were collected from multiple segments of a horizon (e.g. 0–5 cm and 5–25 cm samples for a Horizon described from 0 to 25 cm) but no other volume or description information was provided to differentiate them, the average SOC content was recorded.

SOC Stock Calculations

Soil organic carbon stocks were calculated by multiplying horizon bulk density and SOC concentration, adjusting for coarse fragments and then summing the horizons stocks to a fixed depth of 100 cm for each pedon (Ellert et al. 2008). Where horizon depth did not match the fixed depth increment, the within horizon bulk density and C concentration was assumed to be constant. Variance introduced into the stock calculations by analytical and modeling prediction errors (e.g., SOC concentration and bulk density measurements) and the stratified sampling design was partitioned using a hierarchical Bayesian modeling approach implemented in WinBUGS (Clark 2005; Gelman and Hill 2007; Cressie et al. 2009) and R (Sturtz et al. 2005; R Development Core Team 2011). Specifically, measurement and model prediction errors were used as informative priors and hyperpriors whereas the hierarchical levels of the sampling design (e.g., region, soil group, and LULC) were given uninformative priors. Stock model parameters were estimated with three Markov Chain Monte Carlo (MCMC) chains each with 5,000 updates, a 100 update burn in, and a 50 % thinning rate to reduce auto-correlation within the chains. Convergence of the three MCMCs was determined using the Rhat parameter (Gelman and Hill 2007). This approach allows for complete error propagation and is a flexible method for estimating uncertainty in complex environmental data and models (Clark 2005).

All GIS manipulations were done using ArcGIS (ESRI Inc. Redlands, CA). All prediction and error assessment were conducted using R (R Development Core Team 2011).

Initial SOC Stocks and Uncertainties

A total of 6,148 sites, 32,084 pedons and 144,833 samples were described for the RaCA project. In order to summarize SOC stocks, pedons were averaged by site. Site averages are displayed by GPS coordinates taken at the time of sampling (Fig. 10.1). Summary statistics for site SOC stocks are given in Table 10.2. The mean of all RaCA site SOC stocks to 100 cm was 345.4 Mg ha^{-1}, with a median of 183.2 Mg ha^{-1}. These values are much higher than previous estimates based on SSURGO data, but only slightly higher than reported in the NCSS database by Wills et al. (2013). They report that NCSS SOC stocks to 100 cm ranged from 0 to more than 8,000 Mg C ha^{-1}; with an average of 136.5 and median of 152.3 Mg C ha^{-1}. The NCSS database and SSURGO do not have a full range of LULC classes present and likely include 'typical' or 'representative' conditions. The RaCA project was designed to capture a range of LULC classes, soil groups and expected SOC stocks. Specifically, we found that within and nested-among site variation account for 34 and 60 % of the total variation in SOC stocks to 1 m, respectively, while the remaining variation was attributable to LULC within soil groups and regions (5 %), soil groups within regions (0.5 %), and regional differences (0.04 %).

To extrapolate to all sampled strata, pedon stocks were first natural log transformed, to better approach normality and avoid the skew of extremely large values. Transformed values were averaged by site and site averages were further averaged by soil group and LULC. The average SOC stock value for each soil group and

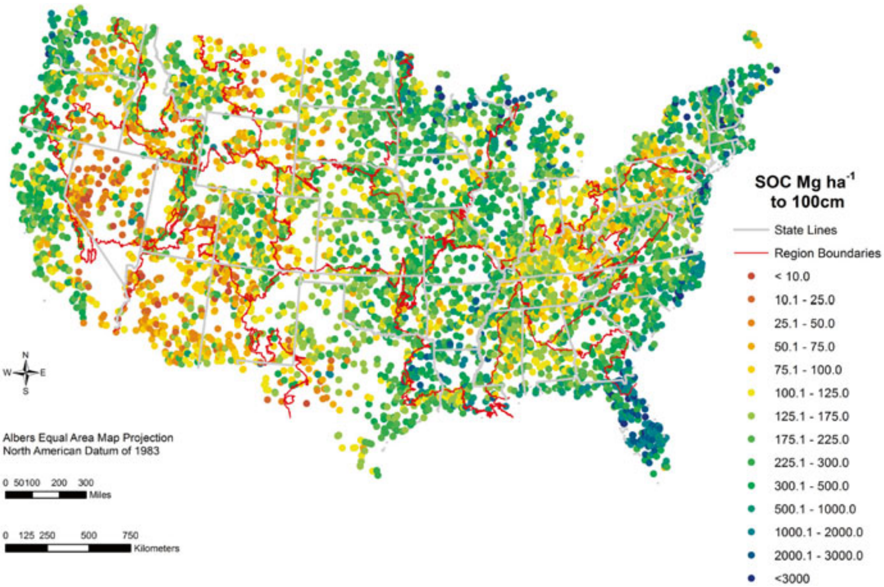

Fig. 10.1 Soil organic carbon stocks to 100 cm for each RaCA site

Table 10.2 Soil organic carbon stocks (Mg C ha^{-1} to 100 cm) summary for original values (stock), log normal transformed values (ln(stock)), and back-transformed values exp$^{(\ln(\text{stock}))}$

	Stock	ln(stock)	Exp$^{(\ln(\text{stock}))}$
Minimum	2.2	0.7	2.1
25th quartile	111.9	4.7	106.5
Median	183.2	5.2	174.3
Mean	345.4	5.2	321.8
75th quartile	317.2	5.7	300.1
Maximum	5,567.0	8.6	5,560.0

LULC combination is displayed on the Jan 2012 SSURGO-NLCD raster grid (Fig. 10.2). Only assessed areas were considered (pixels with both an assigned soil group and a relevant NLCD class). Values were back transformed from natural log to whole stock values before being reported and mapped. Thus, reported means represent geometric means as described by Crawley (2013). Soil group and LULC class geometric means ranged from 14.9 to 3366.7 Mg C ha^{-1}, with an overall mean of 284.6 Mg C ha^{-1}.

Given that this study is the first comprehensive one-time sampling of the CONUS, no studies are available to directly compare with our uncertainty estimates. Whole country comparable uncertainty assessments of SOC stocks have been conducted elsewhere (e.g., Bellamy et al. 2005; Heikkinen et al. 2013) and serve as a basis for comparison upon further analysis of the RaCA data.

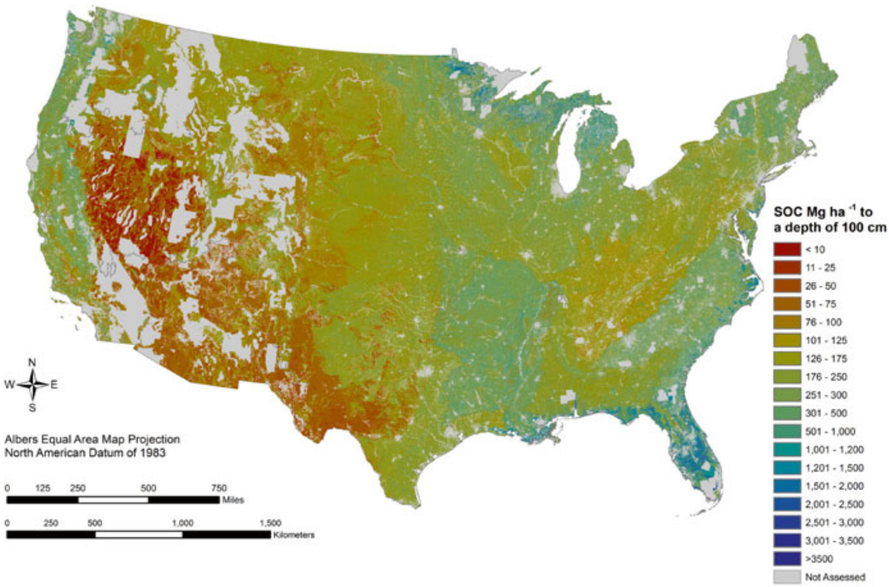

Fig. 10.2 Geometric mean of SOC stocks by soil groups and LULC

References

Bellamy PH, Loveland PJ, Bradley RI, Lark RM, Kirk GJD (2005) Carbon losses from all soils across England and Wales 1978–2003. Nature 437:245–248

Bradley RI, Milne R, Bell J, Lilly A, Jordan C, Higgins A (2005) A soil carbon and land use database for the United Kingdom. Soil Use Manage 21:363–369

Breiman L (2001) Random forests. Mach Learn 45:5–32

Burt R (2004) Soil survey laboratory methods manual. Soil survey investigations report no 42 version 4. U.S. Department of Agriculture, Natural Resources Conservation Service. http://soils.usda.gov/technical/lmm/

Clark JS (2005) Why environmental scientists are becoming Bayesians. Ecol Lett 8(1):2–14. Available at http://doi.wiley.com/10.1111/j.1461-0248.2004.00702.x. Verified 8 Mar 2012

Crawley MJ (2013) The R book, 2nd edn. Wiley, West Sussex, pp 116–118

Cressie N, Calder CA, Clark JS, Ver Hoef JM, Wikle CK (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecol Appl Pub Ecol Soc Am 19(3):553–570. Available at. http://www.ncbi.nlm.nih.gov/pubmed/19425416

Duckworth J (2004) Mathematical data preprocessing. In: Roberts CA, Workman J Jr, Reeves JB III (eds) Near-infrared spectroscopy in agriculture. ASA-CSSA-SSSA, Madison, pp 115–132

Ellert BH, Janzen HH, VandenBgaart AJ, Bremer E (2008) Measuring change in soil organic carbon storage. In: Carter MR, Gregorich EG (eds) Soil sampling and methods of analysis, 2nd edn. CRC Press, Boca Raton, pp 25–38

Eswaran H, Van Den Berg E, Reich P (1993) Organic carbon in soils of the world. Soil Sci Soc Am J 57:192–194

Fry J, Xian G, Jin S, Dewitz J, Homer C, Yang L, Barnes C, Herold N, Wickham J (2011) Completion of the 2006 national land cover database for the conterminous United States. Photogramm Eng Remote Sens 77(9):858–864

Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York

Guo Y, Amundson R, Gong P, Yu Q (2006) Quantity and spatial variability of soil carbon in the conterminous United States. Soil Sci Soc Am J 70:590–600

Heikkinen J, Ketoja E, Nuutinen V, Regina K (2013) Declining trend of carbon in Finnish cropland soils in 1974–2009. Global Change Biol: 1456–1469. Available at http://www.ncbi.nlm.nih.gov/pubmed/23505137. Verified 1 June 2013

Huang S, Yang Y, Wang Y (2003) A critical look at procedures for validating growth and yield models. In: Amaro A, Reed D, Soare P, Soares P (eds) Modeling forest systems. CAB International, Wallingford, pp 271–293

IPCC (2007) Climate change 2007: synthesis report. Contribution of the working group I, II, and III to the fourth assessment report of the intergovernmental panel on climate change. In: Pachauri RK, Reisinger A (eds) Core writing team. IPCC, Geneva, Switzerland, 10 pp

Kern JS (1994) Spatial patterns of soil organic carbon in the contiguous United States. Soil Sci Soc Am J 58:439–455

Kravchenko AN, Robertson GP (2011) Whole-profile soil carbon stocks: the danger of assuming too much from analyses of too little. Soil Sci Soc Am J 75(1):235. Available at https://www.soils.org/publications/sssaj/abstracts/75/1/235. Verified 1 June 2013

National Cooperative Soil Survey (NCSS) (2011) National cooperative soil characterization database. Available online at http://ncsslabdatamart.sc.egov.usda.gov

Nusser SM, Breidt FJ, Fuller WA (1998) Design and estimation for investigating the dynamics of natural resources. Ecol Appl 8(2):234–245

Post WM, Emanuel WR, Zinke PJ, Stangenberger AG (1982) Soil carbon pools and life zones. Nature 298:156–159

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Schoeneberger PJ, Wysocki DA, Benham EC, Broderson WD (eds) (2002) Field book for describing and sampling soils, Version 2.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE

Sequeira CH et al (2014a) Predicting soil bulk density for incomplete databases. Geoderma 213:64–73

Sequeira et al (2014b) Development and update process of VNIR-based models built to predict soil organic carbon. Soil Sci Soc Am J (in press)

Soil Survey Staff (2010a) Natural Resources Conservation Service, United States Department of Agriculture. Official soil series descriptions. USDA-NRCS, Lincoln, NE. http://soils.usda.gov/technical/classification/osd/index.html

Soil Survey Staff (2010b) Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) database. Available online at http://soildatamart.nrcs.usda.gov

Soil Survey Staff (2013a) Gridded Soil Survey Geographic (gSSURGO) database. United States Department of Agriculture, Natural Resources Conservation Service. Available online at: http://datagateway.nrcs.usda.gov/ (FY2013 official release)

Soil Survey Staff (2013b) Rapid Assessment of U.S. Soil Carbon (RaCA) project. United States Department of Agriculture, Natural Resources Conservation Service. Available online at. http://soils.usda.gov/survey/raca/. June 1 2013 (FY2013 official release)

Sturtz S, Ligges U, Gelman A (2005) R2WinBUGS: a package for running WinBUGS. J Stat Softw 12(3):1–17

Todd-Brown KEO, Randerson JT, Post WM, Hoffman FM, Tarnocai C, Schuur EAG, Allison SD (2013) Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. Biogeosciences 10(3):1717–1736. Available at http://www.biogeosciences.net/10/1717/2013/. Verified 22 May 2013

USDA (2007) Natural Resources Conservation Service, National Resources Assessment. National Resources Inventory. http://www.nrcs.usda.gov/wps/portal/nrcs/main/national/technical/nra/nri/

USDA (2010) Natural Resources Conservation Service. National soil survey handbook, title 430-VI. http://soils.usda.gov/technical/handbook/

Wills SA, Seybold C, Chiaretti J, Sequira C, West LT (2013) Quantifying tacit knowledge about soil organic carbon stocks using soil taxa and official soil series descriptions. Soil Sci Soc Am J. doi:10.2136/sssaj2012.0168

Wu H, Guo Z, Ping C (2003) Distribution and storage of soil organic carbon in China. Glob Biogeochem Cycles 17:1048–1060