USDA
United States Department of Agriculture

Rapid Carbon Assessment (RaCA):

Methodology, Sampling, and Summary

National
Cooperative
Soil Survey

Natural
Resources
Conservation
Service

nrcs.usda.gov/

Skye Wills, Soil Scientist (editor), United States Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center

Terrance Loecke, Assistant Professor, University of Kansas

Stephen Roecker, GIS Specialist and Soil Scientist, United States Department of Agriculture, Natural Resources Conservation Service, Indianapolis, IN

Dylan Beaudette, Soil Scientist, United States Department of Agriculture, Natural Resources Conservation Service, Davis, CA

Cleitein Seqeira, Soil Scientist, United States Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center

Ellis Benham, Soil Scientist (retired), United States Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center

Richard Ferguson, Chemist, United States Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center

Ken Scheffe, Soil Scientist, United States Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center

Larry West, National Leader for Research and Laboratory (retired), United States Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center

Cover photos: Courtesy of Amy Langner, Soil Scientist, United States Department of Agriculture, Natural Resources Conservation Service. Rapid Carbon Assessment (RaCA) sampling in New York.

Citation: Soil Survey Staff and T. Loecke. 2016. Rapid Carbon Assessment: Methodology, Sampling, and Summary. S. Wills (ed.). U.S. Department of Agriculture, Natural Resources Conservation Service.

Rapid Carbon Assessment (RaCA) Report

Methodology, Sampling, and Summary

Soil scientist extracts a bulk density sample from a pastureland site using the core method. The sample is trimmed, bagged, labeled, and transported back to the office along with the other samples from this site. The bulk density values will give us a better idea of how much total carbon exists in the soil.

Introduction

The Natural Resources Conservation Service's (NRCS) Soil Science Division started the Rapid Carbon Assessment (RaCA) project to capture information on the carbon content of soils across the conterminous United States (CONUS) at a single point in time. A secondary goal was to capture organic carbon stocks in different kinds of soils and land uses. This document describes the project design and procedures. The attached appendices provide detailed information regarding the procedures, scripts used in data processing, and preliminary data outputs. A brief summary of the data gathered is included in this report.

Project Design

The RaCA project was designed to capture the range and total amount of soil carbon across the CONUS. The project initially emphasized soil organic carbon (SOC) stocks, or the amount of SOC in a volume (area and depth) of soil. Staff at the National Soil Survey Center (NSSC) developed the concept and NRCS soil scientists at the field soil survey offices executed the project. A multi-level stratified random sampling scheme was created to maximize geographical and spatial sample coverage, to maximize the number of conditions represented, and to give a framework for aggregating information into regional areas.

The complete relevant project population includes all lands for which SSURGO maps had been created in the conterminous United States as of January 2012 (Soil Survey Staff, 2012). The population was first defined by a snapshot of the Soil Data Mart (SDM) in October 2010 (Soil Survey Staff, 2012).

For logistical reasons, the RaCA regions that comprised the first-level strata were based on the Soil Science Division's major land resource area (MLRA) regions. These broad geographic regions were organized to manage soil survey data collection and to facilitate quality assurance for soil survey information (USDA, 2010). Soil survey activities in each MLRA region were managed by the MLRA regional offices (known at the time as the "MO's"). When the RaCA project was initiated, NRCS had 18 MLRA regions, 17 of which were in CONUS. The RaCA project, therefore, has 17 regions. The

MLRA regions were later changed as part of an NRCS reorganization. To avoid confusion, we refer to the regions at the time of sampling as RaCA regions. Within each RaCA region, sampling was further stratified by information related to soils and land use and land cover. Each RaCA region had one to three project coordinators who were responsible for project planning and for coordination of sample and data collection following guidance from the NSSC.

In order to cover the full range of possible carbon stocks across the most combinations of soil and land use, we developed an algorithm to group soil series by expected SOC stocks (total mass of carbon to a depth of 1 meter). The objective of the algorithm was to group, by region, soils that were likely to have similar SOC stocks and to respond similarly to changes in land use. For each mapped soil component, taxonomy and property information were compiled from the official series description (Soil Survey Staff, 2010a). This information was supplemented with additional information from the soil data access portal (Soil Survey Staff, 2010b). The combined information was translated into scores that related to the expected content of SOC in the soil. The scores were then used in a statistical clustering algorithm to create 8 to 20 groups for each RaCA region.[1] The soil groups were created independently for each RaCA region. See Wills et al. (2013) for a complete description of scoring and clustering methods. Appendix A lists all of the groups and their soil series components.

Land use-land cover (LULC) classes were developed to coordinate with classes and definitions of the National Resources Inventory (NRI) (USDA, 2007). These definitions include both land uses and land covers. NRI refers to the combination as “land cover/use.” We use the term LULC to convey the same meaning but avoid confusion with information collected by the NRI. 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 LULC classes to correspond with the classes used in the NRI (Table 1). There was some geographic variation in the correspondence of the NRI and NLCD classes. In the eastern CONUS (regions 11, 12, 14, and 18), the grassland/herbaceous NLCD class was assigned to pastureland instead of rangeland to better match NRI assessments.

The spatial distribution of soil group strata and LULC strata was represented as a raster of polygons from the Soil Data Mart snapped to the 2006 NLCD grid. NSSC staff developed the raster. Current versions are available as “gSSURGO” (Soil Survey Staff, 2013). The gSSURGO pixels represent both soil map units and NLCD classes on a 30-meter grid. Each 30-meter pixel had one unique identifier assigned for soil group by map unit key and one for LULC by NLCD code. The map unit key can be joined to a relational database that includes information about each of the components. The dominant soil series component of each map unit was assigned to each pixel. Map units with other types of dominant components (miscellaneous areas, badlands, rock outcrop, etc.) were not evaluated. Table 1 was used to convert NLCD codes to RaCA LULC classes.

Soil group strata and LULC strata were linked together into a LULC-Soil Group Combination, which is designated as “LUGR.” This term includes the RaCA region, the soil group within the region, and the LULC class. Pixel counts were converted to acreages, and an algorithm was used to assign approximately 400 sites per region. The assignments were weighted by LUGR. A minimum extent was required before sampling sites were assigned to a LUGR, and all LUGRs for which sites were assigned had a minimum of five sites. LUGRs of greater extent were assigned additional sites. The minimum acreage for sampling and for additional sampling was adjusted manually for each region. An additional LULC category was added for sampling at this point. Any soil group for which cropland LULC sites were assigned was also assigned three Conservation Reserve Program (CRP) sites. CRP is a well-known program administered by USDA’s Farm Service Agency (USDA, 2013). There is great interest in quantifying the program’s impact on carbon stocks.

---

[1] In region 2, component information from NASIS was used instead of the official series description information due to incongruences between areas mapped by different agencies. In region 1, the statistical clustering algorithm was not used. Instead, a logic tree was used to group soils largely by taxonomy and partially by other series properties.

Table 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*
81	Pasture hay	Pastureland
82	Cultivated crop	Cropland
90	Woody wetlands	Wetland
95	Emergent herbaceous wetlands	Wetland

* In the eastern conterminous United States (RaCA regions 11, 12, 14, and 18), the grass land/herbaceous NLCD class was assigned to pastureland instead of rangeland to better match NRI assessments.

The NRI sampling framework (Nusser et al., 1998) was used to distribute samples across the CONUS. The primary sampling units of NRI are arranged randomly within geographic strata in a way that provides complete coverage of the CONUS. One point was randomly generated to be a potential RaCA site within each NRI primary sampling unit. For each of these points, the soil group was assigned by performing a spatial join with the Gridded Soil Survey Geographic Database (gSSURGO). The land use-land cover of the nearest NRI point was used to assign LULC class. In cases where using the NRI land use-land cover was not possible (i.e., 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 points were randomized, and only those needed for analysis were retained. The exact routine for placing the points will not be made public due to confidentiality requirements associated with the NRI program.

A randomized list of potential sites was supplied to the RaCA coordinators for each region as x-y coordinates with an attached RaCA site identification number. The site ID included information about the region, soil group, land use-land cover, and randomization order. An excess of sites were supplied so that rejected sites could be replaced by the next random site on the list. Coordinators were instructed to evaluate the locations using GIS layers, including aerial imagery, topography, and digital soil maps. The instructions, as distributed, are in Appendix B. If the list of sites was exhausted for a particular LUGR, another set of sites was generated by the NSSC. If no suitable replacements were found, coordinators were instructed to replace the sites with other LULCs from the same soil group.

RaCA Site Data Collection and Pedon Sampling

Once a site had been remotely verified as being viable, a team of NRCS soil scientists was dispatched to collect site information and soil samples. The RaCA field data and sample collection protocols were given to each team, and the teams were instructed to collect information in an Excel file. The protocols and

Excel file are available in Appendix C. Sites were assessed to ensure that the target sample conditions (soil group and LULC) were met and that no acute disturbances or safety hazards prevented sampling as intended. Each site was sampled as a plot. Vegetation information was collected over a 60 square meter area. Five pedons were sampled in a fixed arrangement; although, the arrangement could be modified somewhat based on conditions. The sampling arrangement was designed to maximize the amount of information gained with each sampled site while limiting the logistical costs of obtaining independent pedons.

Because sites were randomly selected, access commonly required a great deal of logistical effort (securing access, obtaining a vehicle, and crossing difficult terrain). Most sites were accepted as found and sampled in the standard manner. Some had localized features, such as a fence or road that required moving individual pedons or the entire site a short distance. When that was not feasible, crews were asked to select a site at another location in the same map unit delineation or at a location outside the map unit but less than 250 meters away. In both cases, the new location was required to be within the LUGR. A relatively few LUGRs were dominated by sites (large forests, tidal marshes) that were completely inaccessible due to a lack of roads or to an inability to secure permission to sample. For those situations, a protocol was developed to move sites to an accessible location. The protocol is in Appendix D. Due diligence was applied to select new sites that replicated the soil, landscape features, and expected vegetation of the original, randomly assigned sites.

Once a site was accepted and the sample layout determined, information about the site or plot as a whole was recorded. The GPS coordinates of the central pedon were recorded (including any offset distance and azimuth from the originally assigned location). If applicable, the ecological site ID, state, and phase were recorded. A measure of exposed (bare) soil was taken, and any observed disturbances (such as trails or watering tanks) were noted. The observed dominant vegetation species were listed, and measurements were taken of tree height and diameter at breast height as applicable. Pictures were taken of the plot center and in each cardinal direction.

Five pedons were sampled at each site: one at the center of the plot and one 30 meters from the center in each cardinal direction (except where the arrangement was altered using the guidance in Appendix C). Each pedon was described according to the “Field Book for Describing and Sampling Soils” (Schoenberger et al., 2002) and assigned to the most likely soil series given the information available. Minimum required information for each horizon included: horizon designation, depths, color, texture, rock fragment modifier (percent coarse fragments by volume), redoximorphic features, and structure (where possible). Small pits were excavated to a depth of 50 centimeters or to a root-limiting layer, such as bedrock or cemented soil. Samples were collected from the surface to a depth of 5 centimeters and from 5 to 50 centimeters by genetic horizon. Probes or augers were used to sample genetic horizons from 50 to 100 centimeters. Volumetric samples were collected for samples from the surface to a depth of 50 centimeters in the most appropriate manner. Samples were labeled, sealed in air-tight bags, and transported to the soil survey regional office for processing.

At the regional office, RaCA coordinators analyzed the samples using specific protocols. The protocols are in shown in Appendix E. Samples were air-dried and sieved to a size of less than 2 millimeters. A sub-sample was oven-dried to obtain an air-dry weight that could be used for bulk density calculation. All mineral samples were scanned using a LabSpec 2500 Visible Near Infrared Spectormeter (VNIR). Each region had an identical VNIR and measured both reference samples and high/low quality control samples to maintain consistency and comparability across regions. Organic horizon samples, samples from central pedons, were sent to the Kellogg Soil Survey Laboratory. Those samples were scanned with another LabSpec 2500, and carbon was measured as described on page 12.

After sample data was collected, the information in the Excel file and the VNIR scans were uploaded to the National Soil Information System (NASIS). All further data processing and analysis were done by NSSC staff with assistance from university cooperators. A list of information about each sample was compiled from all description information uploaded into NASIS (based on SQL database queries for RaCA site ID). Information in the list included RaCA site, pedon number within the site, horizon sequence number, horizon nomenclature, texture, sample dimensions (as available), and a volumetric estimate of coarse fragments for each horizon. Figure 1 shows the location of all RaCA sites. Table 2 lists the number of sites sampled by region and LULC.

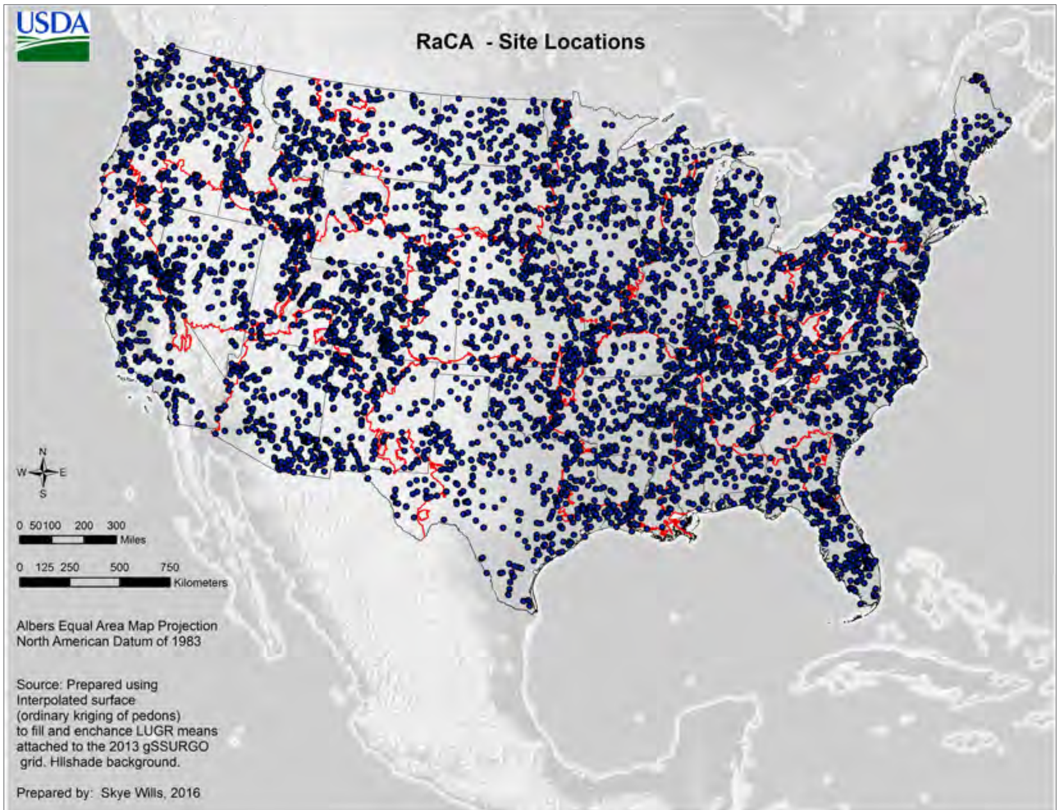

Figure 1.—RaCA site locations.

Table 2. Total Number of Sites Sampled for the Rapid Carbon Assessment (RaCA) Project

RaCA Region	Land use-land cover					
	Cropland	Forestland	Pastureland	Rangeland	Wetland	CRP*
01	64	104	59	103	28	28
02	74	89	43	145	19	9
03	70	94	55	136	40	13
04	43	69	53	79	30	18
05	87	47	49	108	38	33
06	41	89	59	136	53	5
07	98	46	50	145	28	37
08	67	50	30	236	8	1
09	70	73	58	137	28	36
10	119	90	84	33	44	30
11	179	73	79	—	34	30
12	49	200	64	—	46	15
13	69	184	76	—	33	3
14	42	160	93	—	69	12
15	66	121	61	41	73	33
16	69	123	71	28	35	28
18	56	152	104	—	23	16

* Conservation Reserve Program (CRP)

Sample and Pedon Data and Calculations

Bulk Density

When possible, samples occurring between 0 and 50 centimeters in depth were collected by volume for determination of bulk density. The method used was determined by a flowchart that considered the nature of the pedon and horizons (coarse fragments, etc.). Methods for determining bulk density included: volumetric scoop, compliant cavity, constant-volume, and variable-volume cores. The decision flowchart and a full description of each method, including calculations, are given in Appendix C. Calculated bulk density values were checked for quality by first relabeling horizons nomenclature (described according to Schoenenberger et al. 2002) and comparing them to the 1st and 99th percentile (Table 3) of similarly labeled horizons from the NCSS database. Only values within realistic levels were maintained in the dataset for further modeling and calculation.

Table 3. Range of Acceptable Bulk Density Values (g cm^{-3}) by Horizon from the NCSS Database that Are Used to Screen for Outliers of the Measured RaCA Bulk Density

Horizon	1st percentile	99th percentile
A	0.44	1.73
AB	0.36	1.72
Ap	0.76	1.77
B	0.43	1.87
BA	0.68	1.74
Bd	1.07	2.02
Bhs	0.32	1.85
Bk	0.91	1.83
Bm	0.51	2.27
Bq	0.56	2.35
Bo	0.70	1.78
Bt	0.98	1.85
Bv	1.03	1.98
Bx	1.24	2.00
By	0.91	1.72
C	0.71	1.99
Cd	0.94	2.08
Cr	0.58	2.30
E	0.65	1.89
O	0.06	1.39
R*	0.96	2.61

* All values for bedrock (R) were set to null and not used in carbon stock calculations.

For samples that were not collected volumetrically (including all samples collected below a depth of 50 centimeters), pedotransfer functions (PTFs) were developed to predict bulk density using the methodology of Sequeira et al. (2014a). The model was developed using RaCA data with bulk density measurements supplemented with sample data from the National Cooperative Soil Survey database (NCSS, 2011). From a database of 200,000 samples from more than 40,000 pedons, a suite of 58,000 samples from 13,000 pedons with SOC and bulk density measurements were selected. All bulk density measurements in the NCSS dataset were determined at -33 kPa matric potential using the clod method where natural soil clods are dipped in wax and the volume is determined by water displacement (Burt, 2004). The measurement of SOC is discussed below. The randomForest package (Breiman, 2001; Liaw and Wiener, 2002) in the statistical software program R (R Core Team, 2016) was used to develop bulk density PTFs. Random forest algorithms are a type of machine learning technique where a large number of observations and variables are used to cluster data and form decision trees to make predictions. Graphs of the information are presented using the R package ggplot2 (Wickam, 2009).

The PTFs developed for RaCA used information available in both the RaCA and NCSS samples. Sequeira et al (2014a) used generalized horizon designation, textural class, depth for each sample and neighboring samples to build a suite of properties. They reported prediction accuracies of 0.10 to 0.15 g cm^{-3}. In this implementation, two models were created: one for organic samples (horizon master nomenclature 'O') and another 'overall' model for mineral and organic samples. The factors in Table 4 were used in both models; however, the importance of each variable were slightly different. The general horizon designations were expanded for this use to better represent O and L horizons (Table 5).

Table 4. Variables Used in the Overall (organic and mineral horizon) Bulk Density Model

The source refers to whether the data came from the sample being modeled, the sample below the sample being modeled, or above the sample being modeled. The percent increase in mean square error (MSE) and node purity are indicators of the importance of that variable in the accuracy of the overall model.

Source	Variable	Percent increase in MSE	Percent increase in node purity
Sample	Mean bulk density of pedon	78.43	3436.08
Below	Bulk density	63.57	1809.46
Sample	Soil organic carbon	41.81	1027.23
Sample	Generalized horizon	38.56	486.46
Above	Bulk density	38.52	1090.75
Below	Texture	35.55	450.89
Below	Bottom	31.95	373.09
Below	Generalized horizon	31.05	462.87
Sample	Texture	29.49	228.76
Below	Soil organic carbon	28.84	485.28
Sample	Bottom	28.83	253.9
Above	Texture	28.64	224.78
Below	Top	22.36	188.7
Above	Soil organic carbon	19.74	158.69
Above	Top	17.43	128.69
Sample	Top	16.61	232.71
Above	Generalized horizon	15.98	201.38
Above	Bottom	11.61	198.26
Sample	Horizon master	10.63	66.19

Table 5. Generalized Horizon Designation: Strategy Used for Grouping Individual Horizon Designations in the Dataset into Horizon Designations Used for Modeling Bulk Density

Additional horizon information (such numerical prefixes and suffixes) were ignored and horizons were grouped into the most similar match. See the generalized horizon table in Appendix F (modified from Sequeira et al., 2014a).

Generalized horizon designation	Common individual horizon designation during pedon description
A	A, A/E, A/O, AE, Ad Ay
AB	A/B, Ab, ABd, ABgb, ABt, Agb
Ap	Ap, Abp, Agp, Ap/C
B	B, B', BE, B/E, B/C, B and E, Bg, Bj, Bn, Bw, Bwg
BA	BA, B/A
Bd	Bd, Bad, BCd,
Bhs	Bhs, Bh, Bs
Bk	Bk, Bkb, Bkg, Bkk, Bkb, Bkg, Bkny, Bkqy
Bm	Bm, Bkm, Bkqm, Bsm,
Bq	Bq
Bt	Bt, Btg, Btk, Btn, BCt, Bss, Btss, Btb, Btz
Bv	Btv, Bv, BCv
Bx	Bx, Btx, Btgx, B/Ex, BCx,
By	By, Byy, Byz
C	C, CB, C/B, Cg, Ck Ct, Cy,
Cd	Cd
Cr	Cr
E	E, EB, E/B, E and B, E/Bt, E and Bt, E/A, EA, EC
O	Oi, Oa, Oe, Oeb, Oajjb
OC	OA, OC, O/C, O/C, Oajj, B/Oejjf
L	L, Lco, Ldi, Lk, Lma
R	R, R/C, R/Cr

The relationship between measured and predicted bulk density was generally linear (fig. 2); however, organic horizons (O) were sometimes over or under predicted. Figure 3 shows the relationship between modeled and measured values for each generalized horizon designation. When adequate measured samples were present, the modeled values for each generalized horizon fell within the range of measured values, indicating that the model is representing typical situations well. Model bulk density residuals (Table 6) were low (average absolute value of <0.01). Organic horizons had greater differences between modeled and measured values with a slight positive (overestimation) bias. All files and final models used in bulk density prediction for this dataset is available in Appendix F.

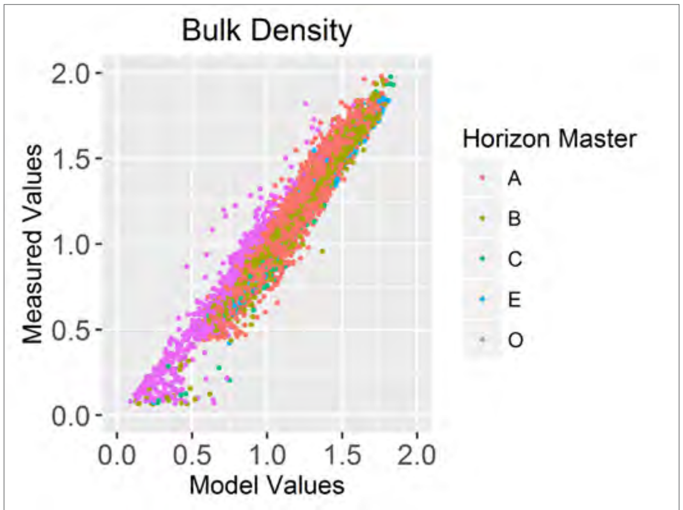

Figure 2.—Scatterplot of measured and modeled bulk density of RaCA samples by horizon master.

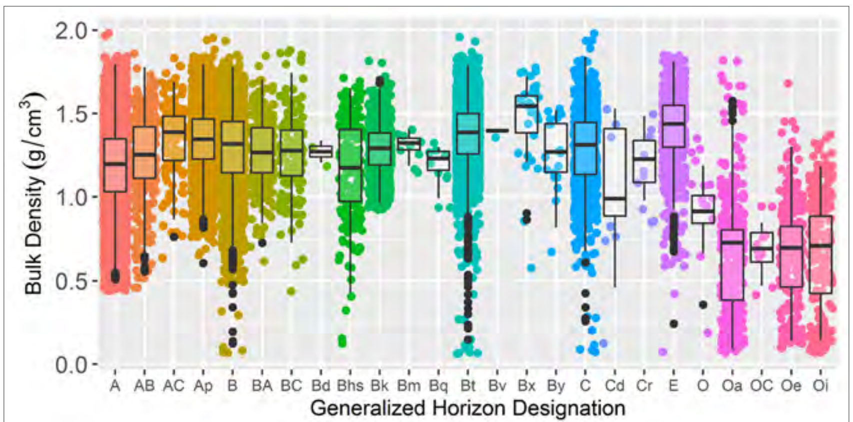

Figure 3.—Bulk density by generalized horizon designation. Points represent measured values; boxplots represent the 25th, 50th, and 75th percentiles of modeled bulk density. Whiskers extend to 1.5 the interquartile range (75th – 25th percentile) and black points represent outliers.

Table 6. Summary of Bulk Density (g cm^{-3}) and Residuals for Measured and Modeled Values

Master horizon	Samples (N)	Bulk density (g cm^{-3})		Residual (measure-model)			
		Measure	Model	Minimum	Mean	Maximum	Standard deviation
A	9267	1.24	1.24	-0.41	0.00	0.36	0.07
B	6008	1.32	1.32	-0.49	0.00	0.28	0.05
C	837	1.27	1.28	-0.54	-0.02	0.16	0.06
E	623	1.41	1.41	-0.33	0.00	0.23	0.05
O	794	0.67	0.66	-0.58	0.01	0.56	0.12

Soil Organic Carbon

Soil organic carbon (SOC) was the central focus of the RaCA project. All samples were to have SOC measured or predicted. Initially, a Visible-Near Infrared (VNIR) spectra was used to predict SOC for all samples and a small subset of samples was sent to the Kellogg Soil Survey Laboratory (KSSL) for laboratory analysis. After initial results indicated inadequate accuracy and noticeable bias, all central pedon samples were sent to KSSL for analysis.

Laboratory Analysis

Soil organic carbon was taken as the difference between total carbon (measured by combustion) and inorganic carbon (measured as calcium carbonate calcimeter equivalence according to the Soil Survey Laboratory Methods Manual (Burt et al., 2004). For the NCSS dataset, some older samples used the Walkley-Black (WB) procedure to measure organic carbon which was converted to equivalent SOC as SOC = 0.25 + WB*.86 as reported by Wills et al. (2014). As expected, the distribution of SOC concentration is skewed with many values near zero and a few relatively large values. Wetlands and forestland have bimodal distribution when plotted on a log scale (fig. 4). This reflects the higher likelihood of organic horizons in those land uses (fig. 5). The depth distribution of all land uses shows that SOC is concentrated near the surface, with wetlands having slightly greater depth of moderate SOC concentration (fig. 6).

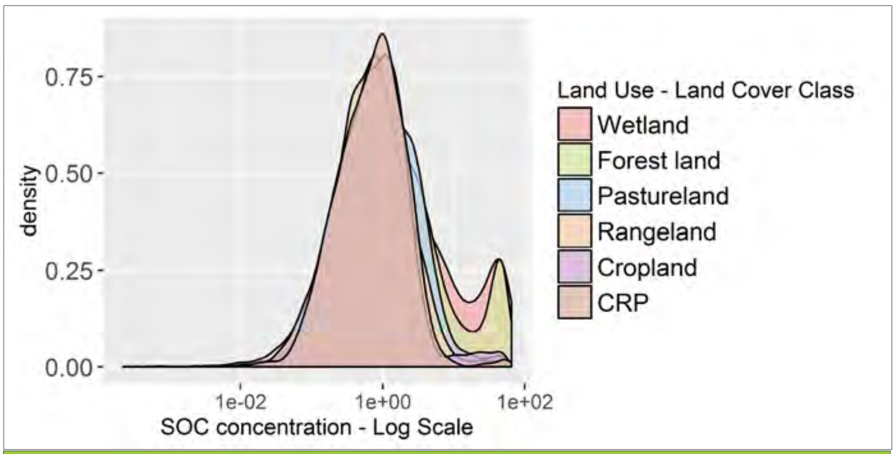

Figure 4.—Density distribution of SOC concentration for individual samples by LULC class.

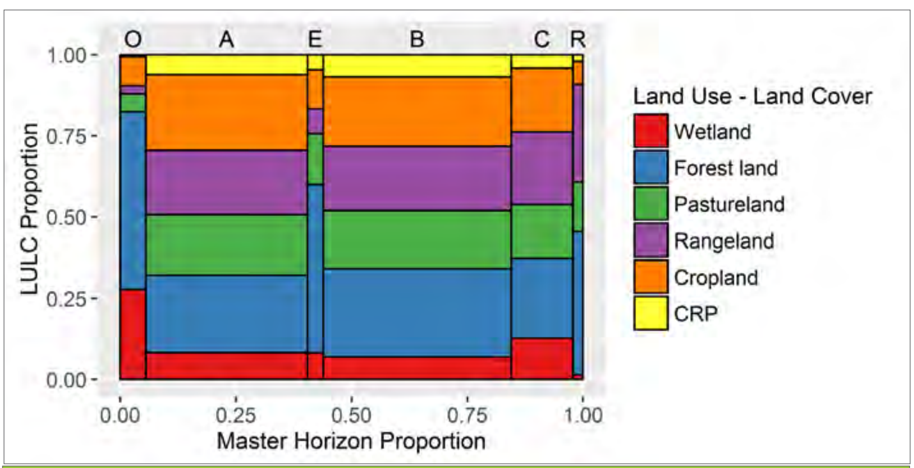

Figure 5.—Mosaic plot with the frequency of horizon masters by LULC class.

Figure 6 illustrates the depth distribution by LULC class using the slab function in the R package aqp (Beaudette, et al., 2013).

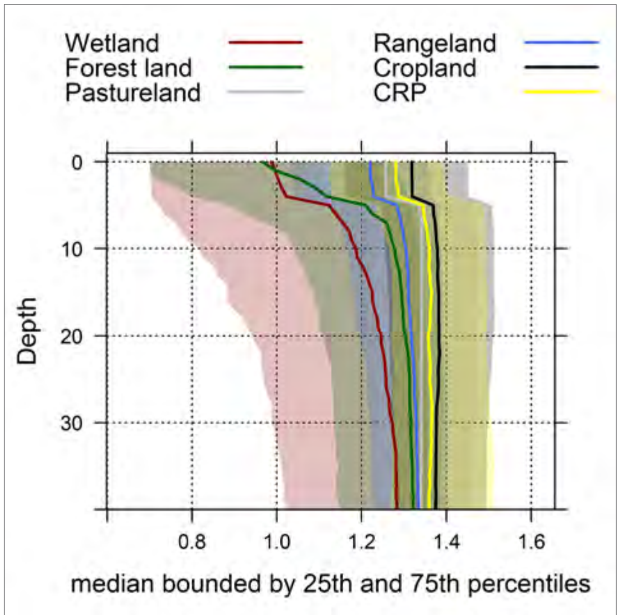

Figure 6a.—Soil organic carbon (SOC, %) concentration.

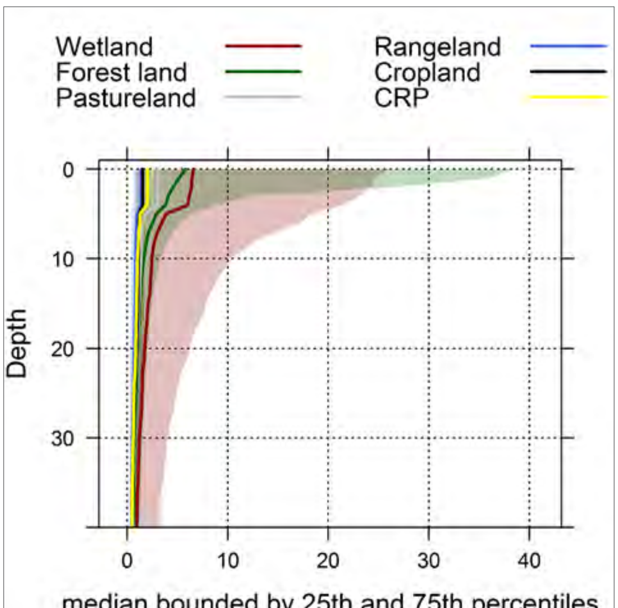

Figure 6b.—Bulk density (BD, g cm^{-3}).

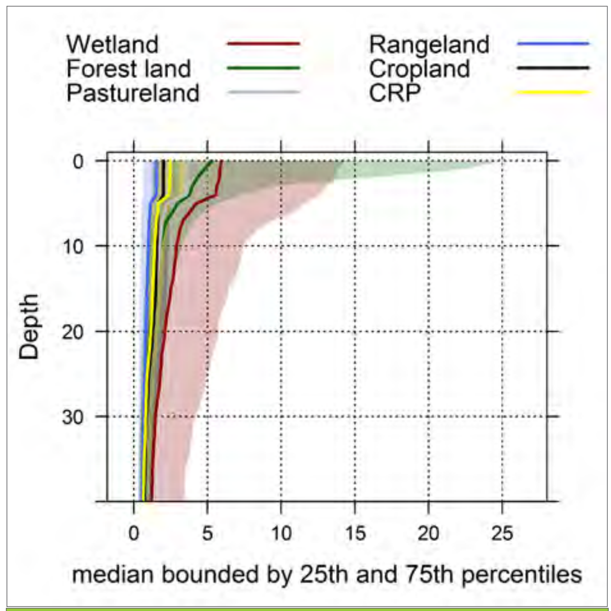

Figure 6c.—Soil organic carbon density (SOC*BD) by depth for each LULC.

Spectra Collection

VNIR scans were performed on air-dry samples of the fraction smaller than 2 millimeters using the LabSpec® 2500 spectrometer (Analytical Spectral Devices, Boulder, Colorado) with spectral range of 350 to 2,500 nanometers, acquired at 1-nanometer increments. Both Sequeira et al. (2014b) and Wijewardane et al. (2016) presented models that can be used to predict SOC from the RaCA scans. In the future, the samples from satellite pedons will have SOC values predicted and used in stock calculation and site level variance calculations.

SOC Stock Calculation

Initial SOC stock calculations were only completed for those pedons with complete sample information and laboratory-measured carbon (available in Appendix G). The complete sample list used in sampling SOC stocks were calculated by multiplying horizon bulk density and coarse-fragment-adjusted carbon concentration (SOC or SIC) and then summing the horizon stocks to fixed depth increments (e.g., depths of 5, 30, or 100 centimeters) for each pedon.

$$SOCstock_{pedon} = \sum_i^n (SOCi * \overline{BD_i} * Dep_i * (1 - CFRAG_i/100))$$

where $\boldsymbol{SOCstock_{pedon}}$ is the soil carbon stock for a pedon in Mg C ha^{-1} , $\boldsymbol{i}$ is the soil horizon, $\boldsymbol{n}$ is the total soil horizons measured, $\boldsymbol{SOC_i}$ is the soil organic carbon concentration in % or g/g soil, $\boldsymbol{BD_i}$ is soil bulk density in g cm^{-3}, $\boldsymbol{Dep_i}$ is the thickness of the horizon within the depth of interest in cm, and $\boldsymbol{CFRAG_i}$ is percent coarse fragments (> 2mm) estimated during pedon description.

Where horizon depth did not match fixed depth increments, the within-horizon bulk density and carbon concentration were assumed to be constant. All calculations were done with the statistics software R (R core team, 2016). Scripts and output files are available in Appendix H and Appendix I.

Figure 7 shows the density of SOC stocks by LULC. Distribution of SOC pedon stocks is skewed with a few very high values. While wetlands have more high value pedons (fig. 7c) than other LULC classes, there are a few pedons that have much higher values than others.

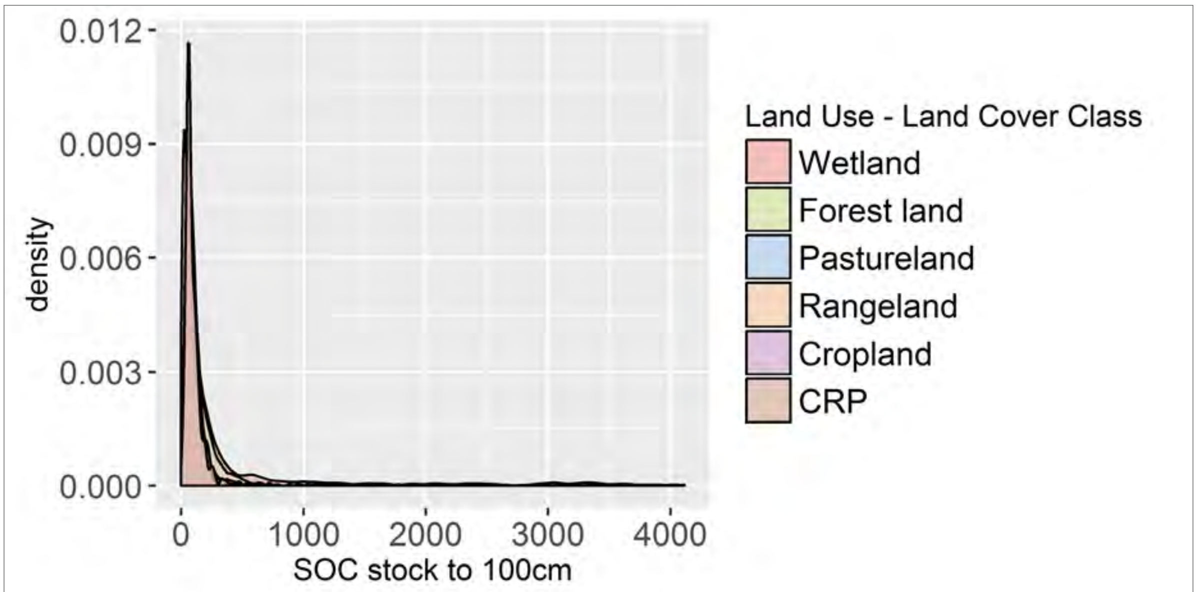

Figure 7a.—Standard density curve.

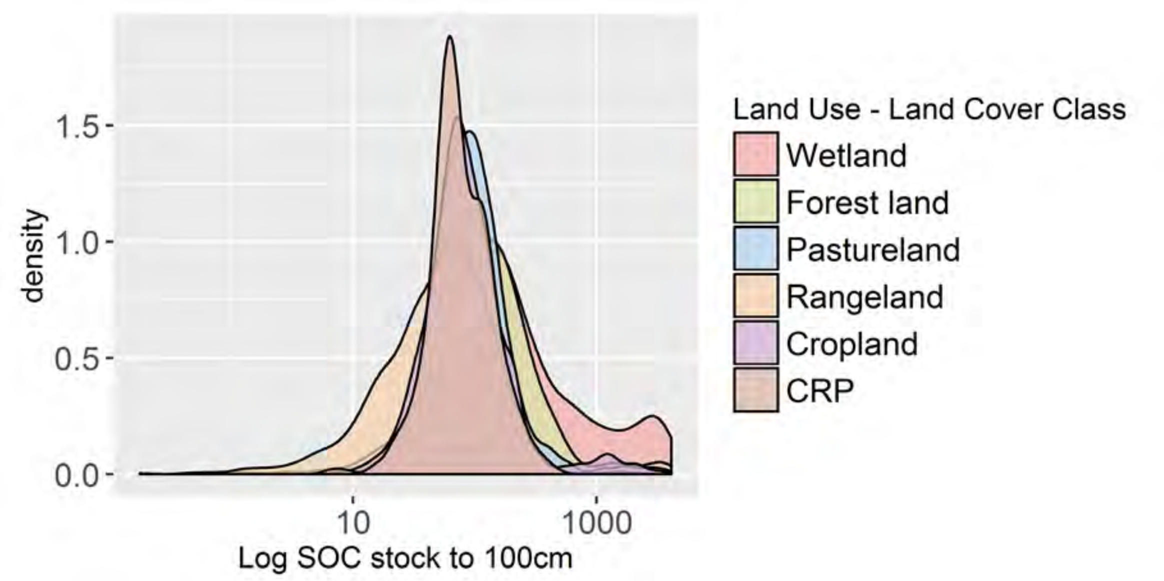

Figure 7b.—Standard density curve on a log scale.

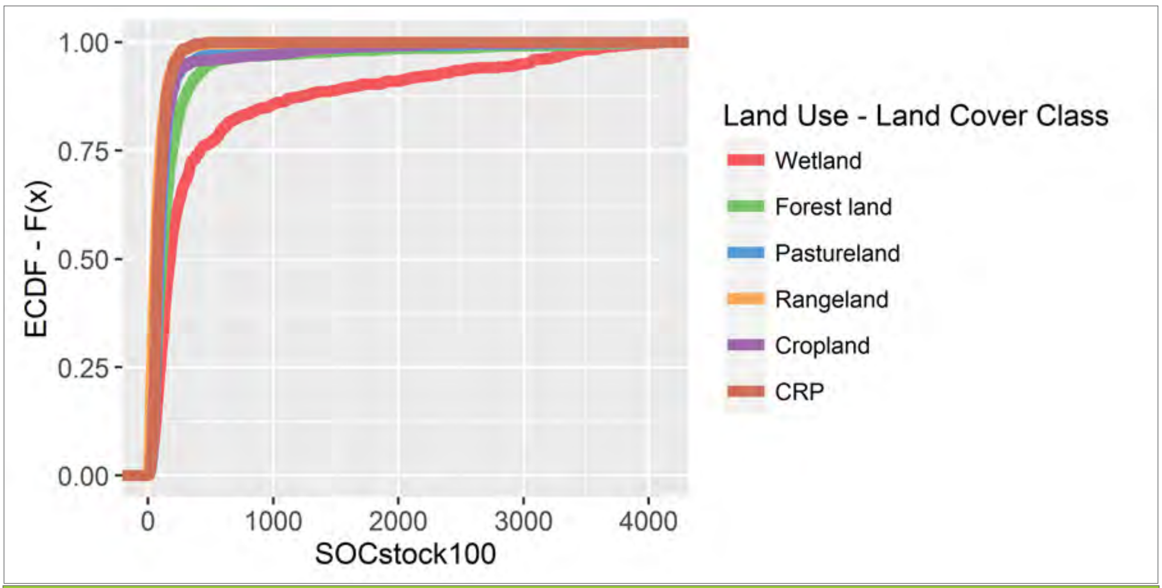

Figure 7c.—Empirical cumulative distribution function.

Site locations and SOC stocks were plotted using ArcGIS (ESRI Inc., Redlands, CA) in figure 8. All maps were made with an Albers equal area projection. The cumulative distribution and map both highlight the skewed distribution of SOC stocks with a few sites containing most of the SOC stocks.

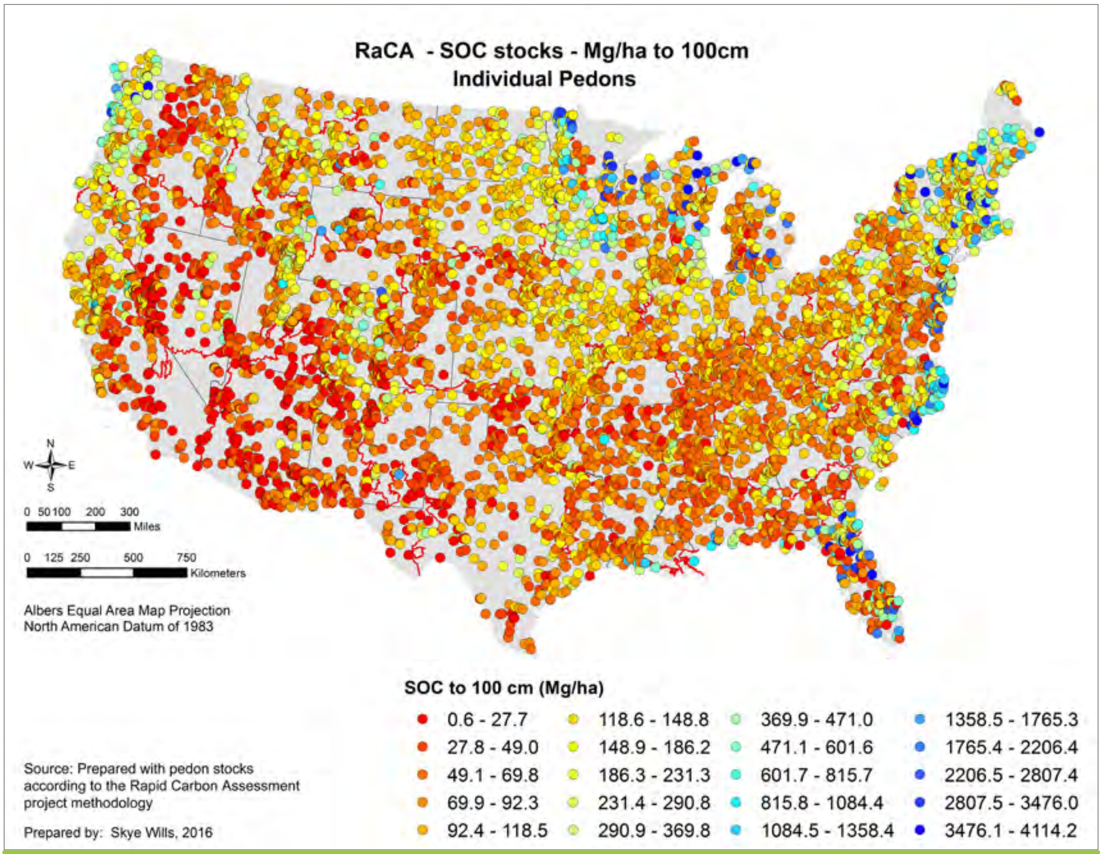

Figure 8.—Soil organic carbon stocks to 100cm for all RaCA sites.

SOC Stock Summaries

Calculating accurate stock summaries is complicated by the distribution of SOC stocks across pedons. The overall distribution of SOC pedon stocks is skewed with very few high values. Plotted on a log scale, wetlands have consistently higher levels of SOC stocks across thicknesses (fig. 9). For most LULC classes, the majority of carbon is stored within the upper 0–5cm of the soil (fig. 10). For each depth, wetlands store more carbon with depth than other land use-land cover types. Near the surface (0–5cm), forestlands have SOC stocks nearly as great as wetlands.

The first method of calculating summaries was a straight-forward transformation, summarization and back transformation. Within LUGR classes, the distribution of SOC stocks was closer to normal so no transformation was done for the initial summary and mapping of SOC stocks by LUGR. For summaries by land use classes and regions, pedon stocks were first-natural-log transformed so that the values were close to a normal distribution and to avoid the skew of extremely large values. The number of pixels assigned to each LUGR was calculated from the January 2012 SSURGO–NLCD raster grid. Only assessed areas were considered (i.e., pixels with both an assigned soil group and a relevant NLCD class). The relative contribution of each LUGR to LULC and region was also calculated. The LUGR contribution to each category (pixel count) was used to calculate weighted averages for CONUS as well as for each region and for wetland, rangeland, pastureland, and forestland. Because CRP is a program and constitutes a combination of LULC types, it is not represented on the NLCD map. Averages for CRP sites, therefore, are simple averages of site stocks (there are generally more pedons in areas with more CRP sites).

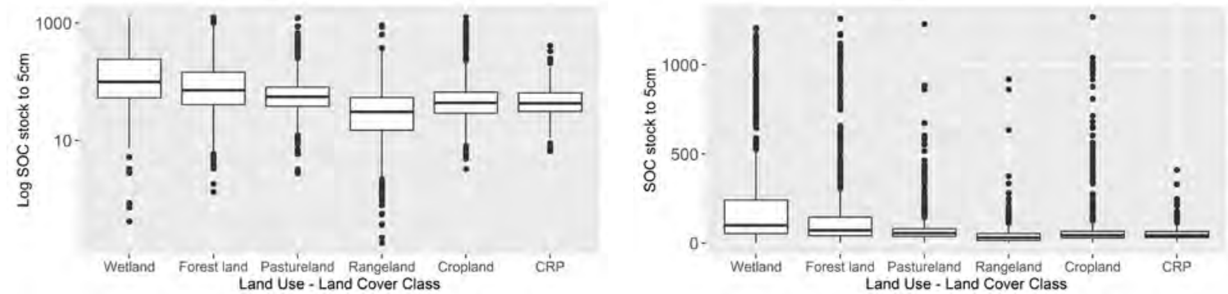

Figure 9a.—5cm thickness

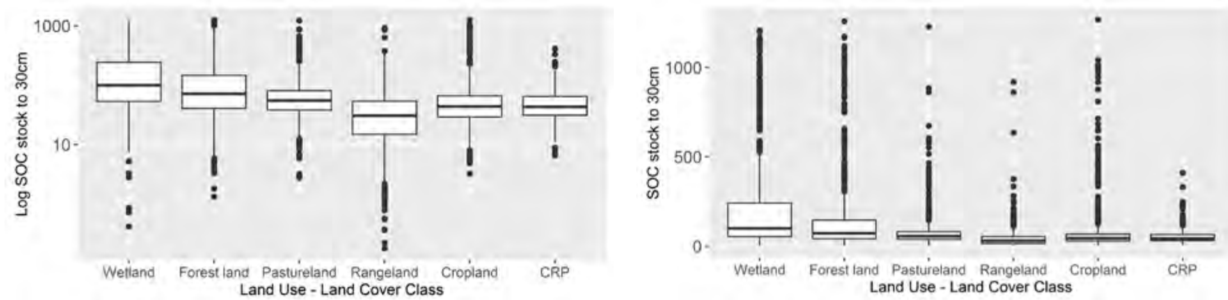

Figure 9b.—30cm thickness

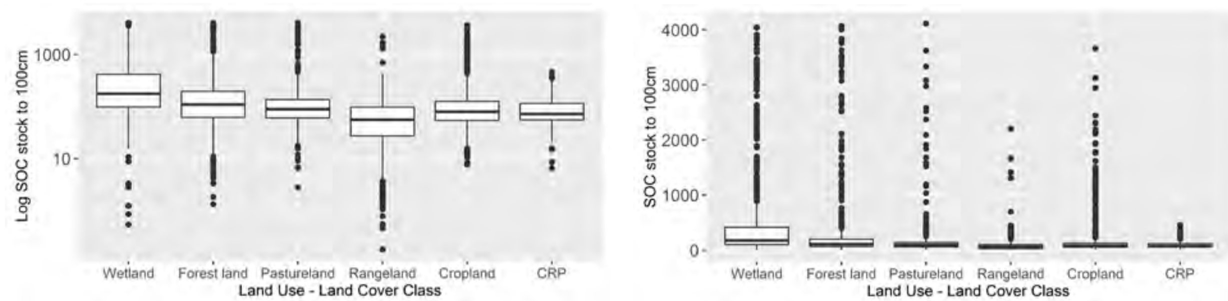

Figure 9c.—100cm thickness

Figure 9.—SOC stock by LULC class on a log and Mg/ha scale: a) to 5cm, b) to 30cm, and c) to 100cm.

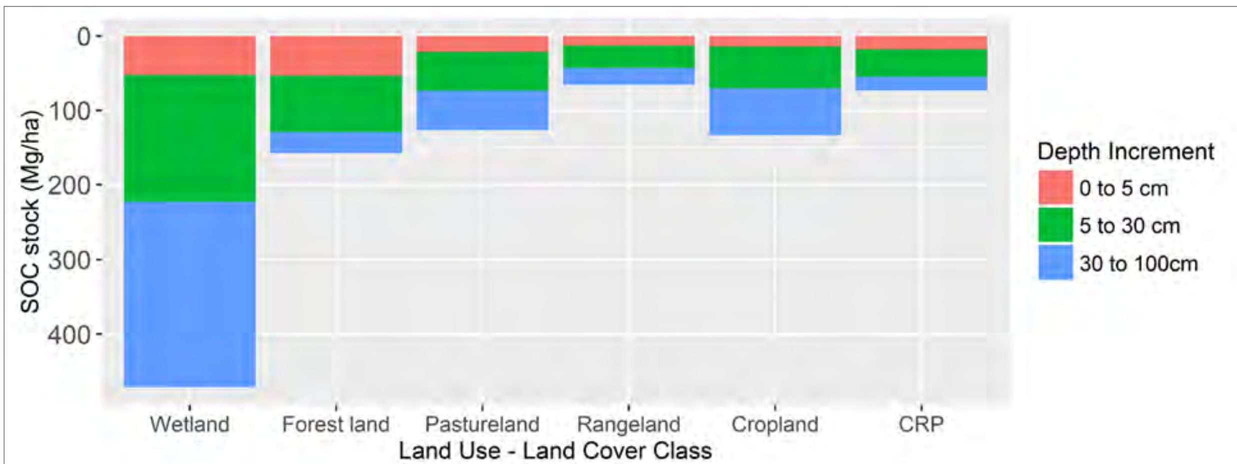

Figure 10.—Stacked bar chart of carbon stock in three depth increments, by land use.

Averages for CRP sites do not contribute to overall regional weighted means. Standard deviations, standard errors, and confidence intervals were calculated on transformed values. Values were back transformed from natural log to whole stock values before being reported and mapped; therefore, reported means represent geometric means as described by Crawley (2013). Table 7 shows the weighted average SOC stocks by land use, region, and across the evaluated area of CONUS.

Table 7. SOC Stock (geometric)* Means Weighted by LUGR Pixels across LULC Classes and RaCA Region

RaCA region[1]	Land use-land cover						Region mean[2]
	Cropland	Forestland	Pastureland	Rangeland	Wetland	CRP[4]	
	--Mg SOC/ha--						
01	77.9	215.1	129.9	49.4	201.3	77.9	87.0
02	69.7	147.3	90.4	28.4	87.5	69.7	57.4
03	68.1	51.6	90.1	25.4	100.7	68.1	33.2
04	113.3	92.1	99.6	69.5	215.8	113.3	86.3
05	81.6	90.1	107.9	66.8	113.9	81.6	79.8
06	61.3	74.1	92.4	40.4	81.2	61.3	56.2
07	106.5	97.5	136.0	85.7	290.2	106.5	105.5
08	60.6	60.9	71.9	23.2	25.4	60.6	31.0
09	53.9	85.7	86.1	48.8	201.0	53.9	64.1
10	121.3	116.0	120.1	116.7	1029.4	121.3	150.7
11	95.3	90.1	87.3	NA	424.1	95.3	103.2
12	86.1	202.1	103.0	NA	508.2	86.1	182.3
13	59.7	140.4	81.6	NA	405.3	59.7	121.0
14	52.2	116.5	77.6	NA	295.3	52.2	119.1
15	61.4	59.5	72.0	55.3	127.9	61.4	71.8
16	61.3	64.2	60.1	71.2	177.4	61.3	72.2
18	77.6	88.6	84.5	NA	100.9	77.6	87.1
LULC mean[3]	91.7	102.7	91.0	50.6	260.6	82.1	78.2

* SOC pedon stocks were natural log transformed. Calculations were completed then back-transformed.
[1] RaCA regions are MLRA office regions prior to 2013.
[2] Area weighted geometric mean based on proportion of each LUGR (land use–soil group) in each region.
[3] Area weighted geometric mean based on proportion of LUGR in each LULC.
[4] CRP values are not reflected in the NLCD maps, and thus are not included in weighted averages.

A secondary method of summarization relied on permutations to both incorporate measurement errors and to estimate mean and quantile values without transformation. For error modeling purposes, samples were designated as organic (including O and L horizons) or mineral (A, B, C, and E horizons) and only samples in pedons with complete pedon information were kept for further calculations. For measurement error estimates of lab-measured values, the KSSL Laboratory Information Management System (LIMS) was queried for samples that had been measured more than once for total carbon (TC), calcium carbonate equivalent (CCE), and bulk density (BD). These replicates would be combined before the information was released as the NCSS characterization database. For each property, the standard deviation was calculated between replicates, and then those values were averaged for mineral and organic horizons. The error of coarse fragments was based on expert knowledge of the error associated with visual estimates of coarse fragment values and assumed to be zero for when no coarse fragments are estimated, small for small values, and larger for large volumes of coarse fragments. An additional error term was added to those samples with modeled bulk density. We assume that this error is in addition to measurement error and could move the estimate closer or further from the 'true' value.

The error terms were generated using the rnorm() function in base R statistics program for bulk density and with the rtriangle() function for carbon and coarse fragments with distributions skewed towards zero. The triangle package (Carnell, 2016) limits the distribution to a minimum and maximum value. For each property, the distribution was limited to plus or minus 2*SD (or error term) with a mean of zero. These error terms were introduced to the summary by permuting 500 times for each sample. With a new random error term added to each value, each time. For each replication, pedon stocks were calculated as above. From that, LUGR mean was calculated for each replication. Both pedon means and LUGR summaries were calculated. Further summarization of LULC and regions were done on LUGR means because this was thought to be most applicable to the landscape being summarized and mapped. The scripts and additional inputs are found in Appendix J. Table 8 shows both overall weighted means and error terms and quantiles using this error propagation and LUGR weight approach. These values are larger than those produced by geometric means. This is largely due to the lack of transformation of extremely large pedon stock values and the bias introduced by error terms of near zero values. The best values to use will depend on the use of the data. For mapping general trends, the back-transformed values will provide the most likely accurate value at any given point. For quantifying the range and error of total stocks, the permutation approach (Table 8) better approximates the entire set of observations for areas within each region.

Table 8. Means and Quantiles by LULC Class, Regions, and Overall Sampled Areas Using the Error Permutation and LUGR Weighting Scheme

Class	Weighted mean	Error (SD of weighted mean across replications)	Percentiles of weighted means (across replications)				
			5th	25th	50th	75th	95th
LULC							
Cropland	106.1	0.7	104.9	105.7	106.1	106.6	107.1
Forestland	141.8	0.8	140.4	141.3	141.8	142.3	143.1
Pastureland	106.2	0.8	105.0	105.7	106.2	106.7	107.5
Rangeland	65.8	0.5	65.0	65.5	65.8	66.1	66.5
Wetland	628.8	12.4	608.4	620.8	628.4	636.9	650.6
Region							
1	129.1	1.6	126.5	128.1	129.1	130.3	131.7
2	94.2	1.4	92.0	93.2	94.1	95.0	96.6
3	49.1	1.0	47.5	48.4	49.1	49.8	50.7
4	99.5	1.5	97.0	98.5	99.5	100.4	101.8
5	88.0	1.2	86.1	87.2	88.0	88.7	89.8
6	76.8	1.2	74.6	75.9	76.7	77.6	78.7
7	123.9	1.2	121.9	123.1	123.9	124.7	125.9
8	42.3	1.0	40.8	41.7	42.3	43.0	43.9
9	79.0	1.1	77.1	78.2	79.0	79.7	80.8
10	318.2	7.7	305.4	313.1	318.0	323.6	331.4
11	156.6	3.1	151.6	154.5	156.7	158.5	161.6
12	261.2	2.2	257.4	259.5	261.2	262.7	264.9
13	159.1	1.3	156.8	158.3	159.2	160.0	161.3
14	190.3	3.0	185.3	188.3	190.2	192.3	195.2
15	151.0	2.5	146.8	149.2	151.0	152.9	155.1
16	94.5	1.9	91.5	93.1	94.4	95.7	97.7
18	95.3	1.0	93.4	94.6	95.3	95.9	97.0
Overall	124.8	0.7	123.7	124.4	124.9	125.3	126.0

SOC stocks graphs were constructed in the R statistics software (R Core Team, 2016) with the packages ggplot2 (Wickam, 2014) and aqp (Beaudette et al., 2013). SOC stocks were mapped in ArcGIS (ESRI Redlands, CA) by site, region, and LUGR. First, pedon site values were plotted with x-y coordinates in the NAD 83 datum obtained at the time of sampling (as in fig.5). Means for each region (weighted by LUGR pixel counts) were attached to regional polygons (fig. 11).

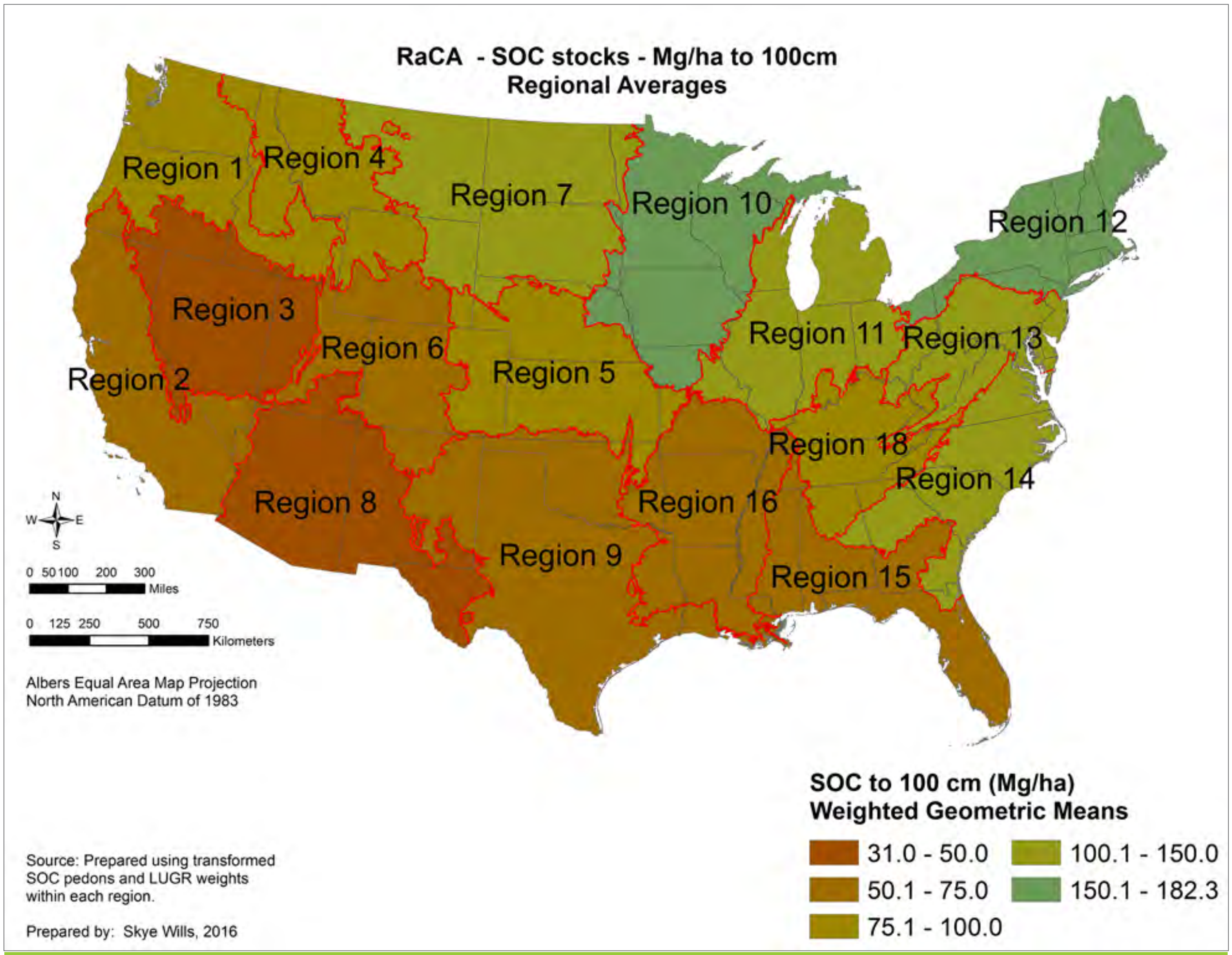

Figure 11.—Regional averages (weighted mean by pixel count of LUGR classes) of SOC stocks to 100cm. The scale is consistent with values in figures 12 and 13.

More detailed raster maps were created using grids of January 2013 gSSURGO with NLCD codes and map unit keys, which were assigned to RaCA LULC and soil series group. The maps were then used to define LUGR classes and to link each pixel to a SOC value based on LUGR mean (fig. 12). An interpolated map was created from the RaCA sites using the ordinary kriging (OK) tool in ArcGIS. The final RaCA map was produced by overlaying the LUGR grid and (OK) maps over a CONUS hillshade (fig. 13).

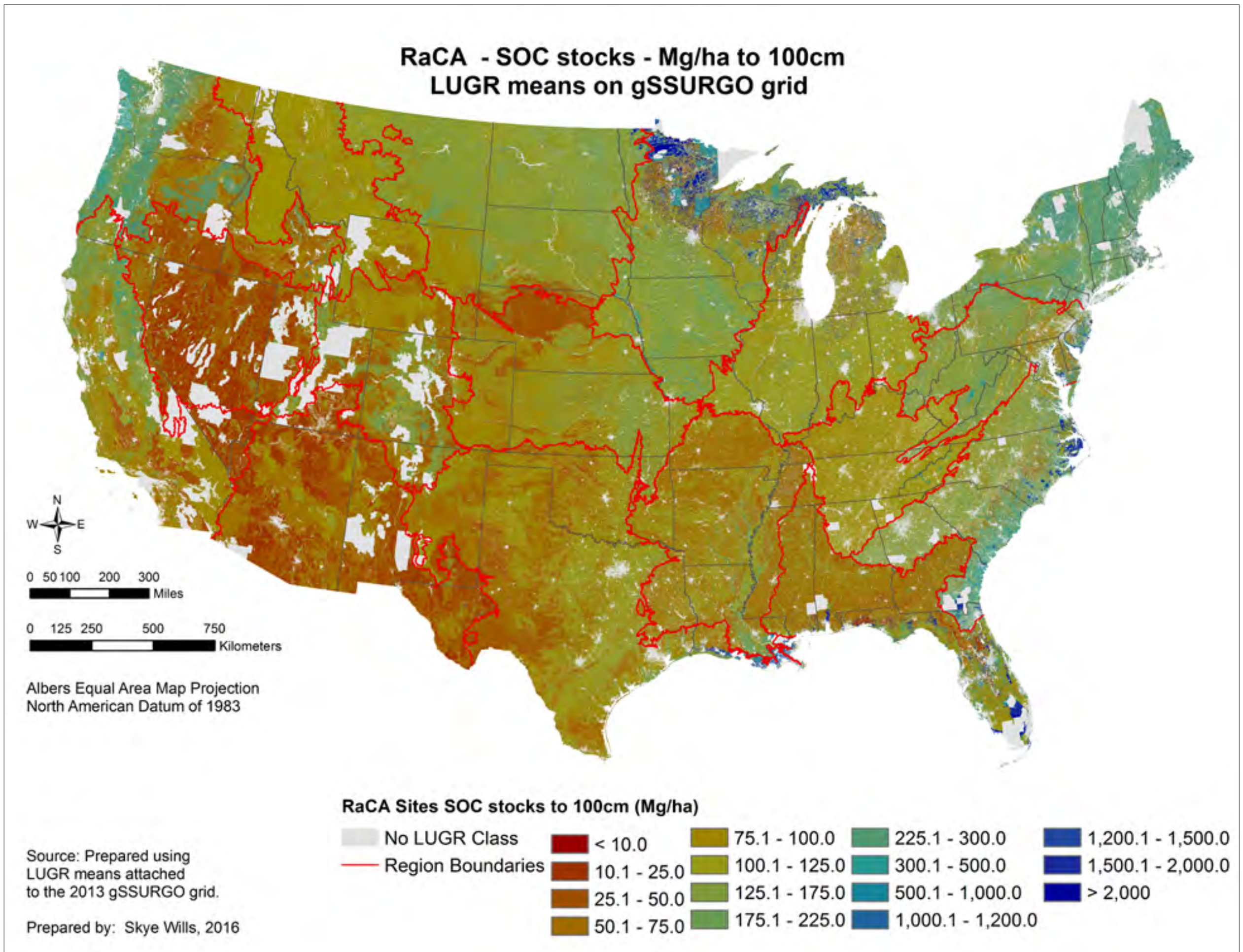

Figure 12.—Summary map of RaCA SOC stocks to 100cm. Created with LUGR means only on lands with 2013 gSSURGO grids.

Using RaCA Data

Initial summaries of SOC stocks are appropriate for visualizing general trends and making broad comparisons. Current stock calculations do not include a full accounting of error propagation. The permutation summaries include sample measurement error, but do not include site variability. Any direct manipulation of this data is likely to underestimate the error associated with estimates. A more complete assessment would include using the VNIR scans of all samples to estimate multi-scale variance. More explicit and extensive geographic predictions should be made using extensive environmental covariates and complex geostatistical algorithms.

All sample data, including site vegetation information, pedon descriptions, laboratory values, and scans are available through the fetchRaCA() command of the soilDB R package (Beaudette et al., 2016) and online through the official RaCA website at http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_054164.

You may also request the data via email. Send your request to soilshotline@lin.usda.gov.

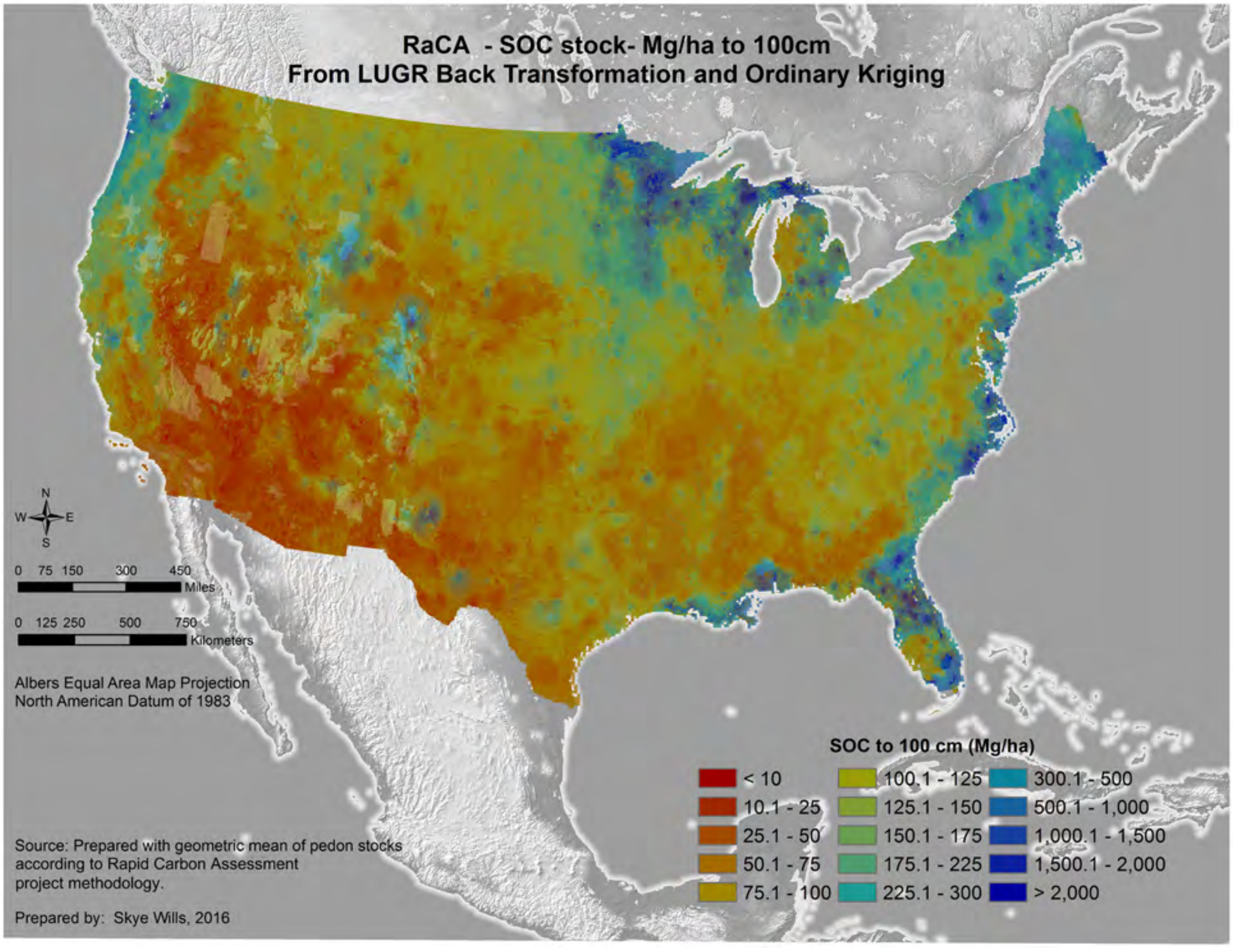

Figure 13.—Summary map of RaCA SOC stocks to 100cm. Created with LUGR means and ordinary kriging (OK) of SOC pedon values.

References

Beaudette, D.E., P. Roudier, and A.T. O'Geen. 2013. Algorithms for quantitative pedology: A toolkit for soil scientists. Computers & Geosciences 52:218-226. doi:10.1016/j.cageo.2012.10.020

Beaudette, D.E., J. Skovlin, and S. Roecker. 2016. soilDB: Soil Database Interface. R package version 1.8. https://CRAN.R-project.org/package=soilDB

Breiman, L. 2001. Random forests. Machine Learning 45:5−32. doi:10.1023/A:1010933404324

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/ (accessed Dec. 2010).

Carnell, R. 2016. Triangle: Provides the standard distribution functions for the triangle distribution. R package, Version 0.10. https://CRAN.R-project.org/package=triangle

Crawley, M.J. 2013. The R Book. 2nd ed. John Wiley & Sons, Ltd. West Sussex, United Kingdom. p. 116–118.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J. 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States, Photogrammetric Engineering and Remote Sensing 77(9):858–864.

Liaw, A. and M. Wiener. 2002. Classification and Regression by randomForest. R News 2(3):18–22.

National Cooperative Soil Survey (NCSS). 2011. National Cooperative Soil Characterization Database. http://ncsslabdatamart.sc.egov.usda.gov

Nusser, S.M., F.J. Breidt, and W.A. Fuller. 1998. Design and estimation for investigating the dynamics of natural resources. Ecological Applications 8(2):234–245.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Schoeneberger, P.J., D.A. Wysocki, E. C, Benham, and W. D. Broderson, editors. 2002. Field book for describing and sampling soils. Version 2.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.

Sequeira, Cleiton H., Skye A. Wills, Cathy A. Seybold, and Larry T. West. 2014a. Predicting soil bulk density for incomplete databases. Geoderma 213: 64–73.

Sequeira, Cleiton H., Skye A. Wills, Sabine Grunwald, Richard R. Ferguson, Ellis C. Benham, and Larry T. West. 2014b. Development and update process of VNIR-based models built to predict soil organic carbon.Soil Science Society of America Journal 78(3): 903–913.

Soil Survey Staff. 2010a. Official Soil Series Descriptions. United States Department of Agriculture, Natural Resources Conservation Service. http://soils.usda.gov/technical/classification/osd/index.html (accessed Dec. 2010).

Soil Survey Staff. 2010b. Soil Survey Geographic (SSURGO) Database. United States Department of Agriculture, Natural Resources Conservation Service. http://soildatamart.nrcs.usda.gov (accessed Dec 2010).

Soil Survey Staff. 2012. Soil Survey Geographic (SSURGO) Database, vector maps. United States Department of Agriculture, Natural Resources Conservation Service. http://soildatamart.nrcs.usda.gov (accessed Jan. 2012).

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

Wijewardane, Nuwan K., et al. 2016. Prediction of Soil Carbon in the Conterminous United States: Visible and Near Infrared Reflectance Spectroscopy Analysis of the Rapid Carbon Assessment Project. Soil Science Society of America Journal 80(4): 973–982.

Wills, Skye, Cathy Seybold, Joe Chiaretti, Cleiton Sequira, and Larry West. 2013. Quantifying tacit knowledge about soil organic carbon stocks using soil taxa and official soil series descriptions. Soil Science Society of America Journal 77(5): 1711–1723.

Wickham, H. 2014. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

United States Department of Agriculture, Farm Service Agency. 2013. Fact Sheet: Conservation Reserve Program. http://www.fsa.usda.gov/Internet/FSA_File/crpfactsheet0213.pdf (accessed March 2013).

United States Department of Agriculture, Natural Resources Conservation Service. 2007. National Resources Assessment. National Resources Inventory. http://www.nrcs.usda.gov/wps/portal/nrcs/main/national/technical/nra/nri/ (accessed March 2013).

United States Department of Agriculture, Natural Resources Conservation Service. 2010. National Soil Survey Handbook, title 430-VI. http://soils.usda.gov/technical/handbook/ (accessed Dec. 2010).

Appendix (Click on the filenames to download the documents.)

Appendix A: Soil Groups and Soil Components by Region (Region_soil components GROUPS.xlsx)

Appendix B: Site Location Guide (Location Instructions.pdf)

Appendix C: RaCA Field Data and Sample Collection Protocols (Rapid Carbon Assessment Workbook 2 15.xlsx)

Appendix D: Considerations and Procedures for Altering Inaccessible Site Locations (Alter Site Location.pdf)

Appendix E: Field Laboratory Instructions (Rapid Carbon Assessment Field Laboratory Instructions.pdf)

Appendix F: Final Models Used in Bulk Density Prediction (BD_model.zip)

Appendix G: Sample Information and Laboratory-Measured Carbon (RaCA_sample.csv)

Appendix H: SOC Stock Calculation (RaCA_SOCstock_calc_final.R)

Appendix I: SOC Pedons (RaCA_SOC_pedons.csv)

Appendix J: Permutations

National Cooperative Soil Survey • November 2016