National Soil Carbon Network: Database calculations and quality assessment.

4 October 2012

By: Luke Nave (Network Coordinator), Kris Johnson (AK Deep Soil Carbon Project Working Group Leader)

This document contains information to help users understand how the NSCN Database performs SOC content (pool size) calculations, and is designed to assist users in their assessments of data quality. This information is presented in 3 overarching sections, covering the following topics:

1) How the database computes SOC contents (g cm^{-2}) from the available data
2) Guidelines for assessing the quality of SOC data
3) Guidelines for assessing the quality of spatial data (coordinates and overlays)

1. Database SOC calculations

1a. SOC content calculations for individual soil layers or horizons

For any given soil layer, whether it is a genetic horizon or a uniformly sampled depth increment, the SOC content of that layer is calculated by the database as:

$$SOC = (\%C * BD * Th) / 100$$

Where, for each layer, *SOC* is the soil organic C content (g cm^{-2}); *%C* is the concentration (per cent by mass) of C in the sample, *BD* is the bulk density (g cm^{-3}), and *Th* is the thickness (cm). Depending on the methods used by the data contributor, *%C* may be either the concentration of organic carbon or total carbon in the sample, and bulk density may have been measured or estimated by a number of different methods (see Section 1c for more information about these variants). Also, note that when a data contributor provides a dataset with her/his own calculated layer SOC contents, these values supersede internal database calculations.

1b. SOC calculations for soil profiles

A soil profile is the sum of its component layers, which may have been sampled by horizon or depth. The maximum sampled depth of any given profile in the database may be determined by a host of factors, such sampling/experimental design considerations or depth to bedrock. Regardless of the reason, the bottom sampled depth for each profile is noted in the relevant database reports, which provide several distinct, profile-level SOC values:

1. User-contributed SOC contents. When a data contributor provides a dataset with calculated profile-level SOC contents, these values override the internal database computation.
2. *SOC Profile Total.* SOC content through the full extent of the profile. If not supplied by the data contributor, the database calculates this value by summing the SOC contents of the profile's component layers.
3. *SOC to 1 m Depth.* For profiles equal to 1 m, this variable is the simple sum of the calculated layer SOC contents to 1 m. For profiles >1 m depth, the database sums the calculated layer SOC contents through the extent of the layer most closely approaching (but not exceeding) 1 m

depth, then applies the following equation to calculate an adjusted SOC content for the layer intersected by 1 m:

$SOC_{adj} = SOC * ((100 - layer_top) / (layer_bot - layer_top))$

Where SOC_{adj} is the SOC content of the layer that is intersected by 1 m, adjusted to 1 m depth; *SOC* is the SOC content of the layer that is intersected by 1 m; *layer_top* and *layer_bot* are the top and bottom of the layer that is intersected by 1 m. SOC_{adj} is then added to the sum of the layer SOC contents through the extent of the layer most closely approaching 1 m, producing an estimated SOC content to a standardized depth of 1 m. Note that this adjustment assumes a homogenous vertical distribution of SOC within the layer intersected by 1 m.

4. *Alaska SOC (Profile Total or to 1 m Depth)*. In certain reports containing profile SOC data, users will find variant profile SOC calculations identified with an 'AK SOC' or 'soc_AKfill' column heading, or a value of 'AKfill' in a column called 'Carbon Flag' (see Section 2b for more information about Carbon Flags). Whether expressed on a profile total basis or to a standardized 1 m depth, these Alaska SOC values are based on special rules, relationships, and gap-filling equations developed by the Alaska Deep Soil C Project and detailed in the Appendix at the end of this document.

1c. Variables used for database SOC calculations

As noted in section 1a, SOC content computations for the individual soil layers in the database encompass both types of %C reporting: % organic carbon (*OC*) and % total carbon (*C_tot*). In performing these computations, the database uses *C_tot* as the default because this is the most reproducible and commonly reported method. Furthermore, a large fraction of the *OC* data are based on obsolete wet oxidation methods (e.g., 6A1c; Burt, 2004) with significant bias in certain soils. However, in soils with high inorganic C contents, using *C_tot* as an estimate of *OC* in the layer SOC content calculations will overestimate SOC; users may simply acknowledge this as a limitation (potential bias) in the dataset, address this problem by using %$CaCO_3$ data to calculate *OC* values for samples reporting both *C_tot* and $CaCO_3$, or use other approaches.

Similar variation exists in the forms of bulk density reported to the database by the data contributors. Whenever possible, the database calculates layer SOC contents using sample bulk densities, *i.e.*, the grams of dried, 2mm-sieved, root free soil per cm^{-3} (*bd_samp*; sample bulk density; the fine earth fraction). However, in some cases, bulk density is only available for unsieved soils or soils that have not been corrected for the contributions of roots or particles >2mm to their measured bulk density values (*bd_tot*; total bulk density). Still other SOC contents in the database reports are calculated using whole soil bulk densities (*bd_whole*), which are derived, not measured, bulk density values. Some of these whole soil bulk densities are derived from measured mass and volume that have been corrected for in-the-field visual estimates of coarse fragment content, while others are simply gap-filled by the contributor according to averaged or theoretical values. The end result of this variation in bulk density measurement and estimation techniques is the introduction of unquantified error into the calculated layer SOC contents in the database reports. Users with strict requirements may consider filtering the

data and using only layers that report *bd_samp*, or performing a sensitivity analysis using several bulk density variants. For some purposes, using all SOC contents regardless of the type of bulk density measurement may be sufficient; furthermore, the sheer volume of data in the database will often enable users to develop their own modeled relationships among bulk density variants and other parameters.

Users of Alaska SOC data should consult the Appendix at the end of this document for information about the variables and approaches used to calculate layer SOC contents for soils from Alaska.

2. Guidelines for assessing SOC data quality

The NSCN database provides several metrics that the user may employ to assess the quality of data, or to restrict the data used for a particular analysis to only those observations passing a certain quality filter. As with the SOC content calculations described in Section 1, these quality metrics apply variously to soil profiles or their constituent, individual soil layers. Detailed information about these quality metrics appears in Sections 2a and 2b below, with guidelines describing the application of these metrics to data filtering in Section 2c.

2a. Quality flags for soil layers

Data for individual soil layers are found in the special table called NSCNLayerData. The individual soil layers in this table are flagged according their depth measurements and SOC calculations, using the variables called 'layer_flag' and 'soc_flag', respectively. Layer flags denote the distribution of layers within their parent profiles according to the rubric in Table 1 below. SOC flags, described in Table 2, identify how the layer SOC content was computed.

Table 1. Definitions of Layer Flags.

Layer_flag	Definition
Contiguous	Well-behaved layer. Has ≤5cm of gap or overlap with any adjacent layer.
Discontiguous	Problematic layer. Has >5cm of gap or overlap with an adjacent layer.
MissingInfo	Bad layer. Is missing layer_top or layer_bot, or layer_top = layer_bot

Table 2. Definitions of SOC Flags.

SOC_flag	Definition
nofill	SOC content computed by the database without gap-filling equations
con	SOC content computed by data contributor, no database SOC calculation
con:nofill	SOC content computed by data contributor, database SOC available
AKfill	SOC content computed by the database using Alaska gap-filling equations
(blank)	SOC calculation not possible (MissingInfo layer_flag)

2b. Quality flags for soil profiles

Data for soil profiles are found in the special tables called NSCNProfileData and NSCNCarbonByHorizon, and in the database reports called Carbonto1M and Carbonto1M_AK. The profiles in these database files are flagged according to their completeness and SOC calculations, using the variables called Profile Flag and Carbon Flag, respectively. A profile flag denotes the depth of the profile, as well as whether any of the constituent layers are missing or have gaps or overlaps with adjacent layers (see Table 3 below).

Carbon Flags have the same nomenclature as the soc_flags of individual layers; as applied to profiles these denote how the SOC values were calculated for the layers that make up each profile (see Table 2).

Table 3. Definitions of Profile Flags

Profile_flag	Definition
Complete	Profile is ≥1m deep and comprised of contiguous layers with no gaps or overlaps in the first 1m
Complete(5cm)	Profile is ≥1m deep and comprised of contiguous layers, up to 5cm of gaps or overlaps in the first 1m
Short	Profile is <1m deep and comprised of contiguous layers with no gaps or overlaps
Short(5cm)	Profile is <1m deep and comprised of contiguous layers, up to 5cm of gaps or overlaps between the layers
GapOverlap	Profile depth varies, may be comprised of contiguous or discontiguous layers, >5cm of gaps or overlaps
MissingInfo	Profile consists entirely of layers with missing depth measurements
NoSampleData	Profile has no depth measurements, carbon or bulk density data

2c. Using quality flags and other information to filter data

Soil layers

Data users performing analysis with raw data from **individual layers** may periodically want to use only layers passing certain quality criteria. Filtering downloaded data according to their layer_flags enables users to exclude data from layers that do not form perfect sequences with adjacent layers (i.e., Discontiguous or MissingInfo layers). For example, consider a case where the user wishes to aggregate data from individual layers into profiles—in such a case it is logical to exclude layers with a layer_flag of 'MissingInfo' because these lack the necessary depth information to construct profiles (conversely, if the user is interested strictly in the chemical properties of the layers themselves, depth measurements may be irrelevant).Users who wish to construct their own profiles from raw layer data may have different tolerances for overlapping layers, such that some would prefer to exclude layers flagged as 'Discontiguous' while others would include them. SOC_flags as applied to individual layers are most likely useful only in cases where the user wants to restrict analysis to SOC contents that have been calculated by original data contributors vs. calculated by the database according to its preferences for specific %C and bulk density parameters. Lastly, it should be noted that the availability of multiple %C and bulk density variants (see Section 1c) provides the opportunity for the user to filter datasets according to which of these variants are best applied to the question at hand. For example, if a user is calculating SOC content of layers with high coarse fragment content, (s)he may wish to exclude layers that do not have rock-corrected bulk density values (i.e., use only layers with bd_samp or bd_whole).

Soil profiles

Data users performing analysis of **soil profiles** will find many helpful uses for the Profile Flag. By filtering to include or exclude profiles falling within the various profile flag categories, and using additional information about profile properties (e.g., profile depth and layer counts), it is possible to conveniently

refine datasets from the large data tables and reports available on the NSCN website. While >80% of the profiles in this report have <5cm of gaps or overlaps between their component layers (>60% are complete to 1m), the profile depth and layer count parameters elucidate missing layers where they do occur. This section considers use of profile flags in the context of the various reports and tables where these flags appear.

Carbonto1M: Interpret Profile Flags alongside the values in the Profile Top, Profile Bot, SOC Total Layer Count, and Total Layer Count columns. A difference between the number of layers in the SOC Total Layer Count and the Total Layer Count indicates that profile does not have SOC content calculations for all of the layers in the profile. Such a profile may still be Complete (i.e., have a perfectly contiguous sequence of layers ≥1m deep), but nonetheless lack a SOC to 1m value due to a missing SOC content calculation in one or more of the layers within the first 1 m (accordingly, this will have a (blank) Carbon Flag).

NSCNProfileData: Use Profile Flags, profile top and bottom depths, and layer counts in the same manner as in the Carbonto1M report. Note that the 'soc' column in this table contains the calculated profile SOC contents submitted by data contributors (for profiles with a Carbon Flag of 'con' in the Carbonto1M report).

NSCNLayerData: For users wishing to assemble their own profiles from raw layer data, the layers in this table have Profile Flags, allowing convenient filtering to eliminate layers from Short, GapOverlap, or other profiles that may not be appropriate for stringent quality criteria.

CarbonByHorizon: This table contains SOC contents for the master horizons, as well as the profile total SOC contents, for all profiles for which SOC calculation is possible. Filter by profile_flag to refine the profiles under consideration for analysis.

3. Guidelines for assessing spatial data quality

The NSCN receives from data contributors, and provides to users, the geospatial coordinates of each Site in the database. The preferred format for geocoordinates is latitude and longitude in decimal degrees; when a contributor provides geocoordinates in degrees, minutes, and seconds, these values are converted to decimal degrees by the database using standard functions. The preferred datum for geocoordinates is WGS84, although an inspection of database reports and tables will reveal that the datum of measurement is NAD27, NAD83, or unknown for some Sites. Thus, while many of the Sites in the database have latitude and longitude reported by the contributor to 5 decimal degrees (~1m resolution), it is typically not appropriate to consider these geocoordinates to be accurate to such a fine degree of spatial resolution. Indeed, this is why latitude and longitude are treated as Site-level variables in the database, and why overlay (GIS-derived) data such as NLCD covertypes are provided as Site-level parameters. Users with strict spatial requirements may wish to use only data from WGS84 Sites, or to use their own tools to convert Sites reporting latitude/longitude relative to another datum to a standard of WGS84. On average, there is not likely to be any directional bias in the geocoordinates stored in the NSCN database, but a sensitivity analysis and full quality assessment has not yet been performed to

validate these data. Users who discover problems with geocoordinates are asked to visit the Forum section of the NSCN website and report them to the thread on known database quality problems.

Appendix. Calculations specific to the AK Deep Soil C Project reports

The database reports produced by the Alaska Deep Soil C Project are populated by some data that were derived from modeled relationships between reported variables. In most cases, these modeled relationships were applied in order to derive bulk density from %C for individual samples. Other equations were applied to make minor adjustments so that %C and bulk density measurements from different methods, and their SOC calculations, were comparable. The propagated error from using these equations and adjustments for calculating SOC at the profile level remains unquantified. These calculations are described below, with text adapted from Appendix 1 of Johnson et al. (2011).

Gap-filling procedure

Negative exponential models that predict bulk density from %C were applied for missing data in all mineral soil horizons except arctic soils (Table 4 below). There were some rare cases when samples had bulk density data available but not %C and were gap-filled using a modified equation (Eq. 3b). Models of SOC or bulk density were better fit when the horizon designation was known (Eq. 4-6). When there was no horizon designation, and the horizon was only known to be organic (SOC_O) or mineral (BD_{min}), then general models were applied (Eq's. 3a, 9). Frozen mineral soil bulk density of mainly boreal profiles was predicted separately from the relation found from the %C and bulk density relation of similar soils, but was not distinguished by horizon designation (Eq. 7). In contrast to bulk density measurements of mineral soils, bulk density in organic soils was not well-predicted by non-linear models of %C. The best approach in this case proved to be the direct prediction of SOC content from horizon thickness, *Th*, using a weighted least squares regression and by horizon designation (Eq. 10-12).

Adjustment equations were applied to bulk density and organic carbon concentration measurements from the USDA-NRCS in order to make them comparable to other datasets. Bulk density measurements by the USDA-NRCS were done by the clod method, BD_{clod} (method 3B1; Burt, 2004) whereas all the other bulk densities in the AK database reports were measured by the cylinder method, BD_{core}. The clod method yields consistently higher values than the cylinder method (Van Remortel and Shields, 1993; Calhoun et al., 2001). To correct for this difference in mineral soils, the same equation used in VanRemortel and Shields (1993) was applied (Eq. 1). A similar correction equation has not been published for organic soils to our knowledge. Yet, we found that organic layer bulk densities measured by the clod method were between 1.4 and 5 times greater than by the core method (using a subset of data including black spruce stands only). Therefore, organic horizon bulk density measurements by the clod method were excluded and treated as if they were missing data. For organic carbon concentration, some NRCS data (26% of the total dataset) was measured only for organic carbon, $\%C_{org}$ *(e.g.*, method 6A1c; Burt, 2004). The rest of the dataset was measured for total carbon, $\%C_{tot}$ (methods 4H2a or 6A2d; Burt, 2004).Therefore, a relation was found so that in cases where only $\%C_{org}$ data was available, it was adjusted to more closely match $\%C_{tot}$ (Eq. 2).

In the arctic tundra many profiles were highly cryoturbated which requires specialized methods of calculating SOC content (*e.g.*, Michaelson et al., 1996). The 1-m SOC estimates for highly cryoturbated

profiles in this study included only those with published values (Michaelson et al., 1996; Ping et al., 1997; Bockheim et al., 1999; Bockheim, 2007a,b) and therefore no bulk density predictions were necessary. Non-cryoturbated soils whether organic or mineral, frozen or unfrozen, were predicted by a separate relation specific to arctic soils (Eq. 13; see also Bockheim et al., 2003 for a similar equation).

Table 4. Gap-filling equations for the AK Deep Soil C Project database reports.

Equations	**adj. R2**
adjustment equations	
1. $BD_{adj} = BD_{core} = \frac{BD_{clod} - 0.068}{1.011}$	0.98
2. $\%C_{adj} = \%C_{tot} = 0.2107 + 0.8830 * \%C_{org}$	0.98
prediction equations for mineral soils	
3a. $BD_{\min} = 0.4189 + e^{-0.1868*\%C_{tot}}$	0.64
3b. $\%C_{pred} = -\frac{\log(BD_{core} - 0.4223)}{0.1890}$	0.54
4. $BD_A = 0.3417 + e^{-0.1712*\%C_{tot}}$	0.59
5. $BD_B = 0.4671 + e^{-0.1915*\%C_{tot}}$	0.52
6. $BD_C = 0.6560 + e^{-0.2466*\%C_{tot}}$	0.49
7. $BD_{frzn} = 0.3105 + e^{-0.1400*\%C_{tot}}$	0.48
prediction equations for organic soils	
9. $SOC_O = 0.0085 + 0.0334 * Th$	0.47
10. $SOC_{Oi} = 0.0109 + 0.0183 * Th$	0.38
11. $SOC_{Oe} = 0.0269 + 0.0411 * Th$	0.59
12. $SOC_{Oa} = 0.0743 + 0.0354 * Th$	0.63
prediction equation for all arctic soils frozen or unfrozen; mineral, organic, or cryoturbated	

13. $BD = 0.0577 + e^{-0.0694*\%C_{tot}}$ 0.60

Data quality assessment for the AK Deep Soil C Database reports is based on the gap-filling equations described above, which were used to calculate the SOC values in the reports. See the rubric below:

SOC pool size computed using	Quality score
No gap-filling or adjustments*	A
Equations 1-2	B
Equations 3-13	C
Equations 1-2 and 3-13	C

*includes highly cryoturbated arctic tundra 1-m SOC estimates from publications

References

Bockheim, J.G., Everett, L., Hinkel, K.M., Nelson, F.E., Brown, J., 1999. Soil Organic Carbon Storage and Distribution in Arctic Tundra, Barrow, Alaska. Soil Science Society of America Journal 63, 934-940.

Bockheim, J.G., 2007a. Importance of Cryoturbation in Redistributing Organic Carbon in Permafrost-Affected Soils. Soil Science Society of America Journal 71(4), 1335.

Bockheim, J.G., Hinkel, K.M., 2007b. The Importance of "Deep" Organic Carbon in Permafrost-Affected Soils of Arctic Alaska. Soil Science Society of America Journal 71(6), 1889.

Burt, R., 2004. Soil survey laboratory methods manual. NRCS Soil Survey Investigations Report No. 42.

Calhoun, F., Smeck, N., Slater, B., Bigham, J., Hall, G., 2001. Predicting bulk density of Ohio soils from morphology, genetic principles, and laboratory characterization data. Soil Science Society of America Journal 65(3), 811.

Johnson, K.D., Harden, J., McGuire, A.D., Bliss, N.B., Bockheim, J.G., Clark, M., Nettleton-Hollingsworth, T., Jorgenson, M.T., Kane, E.S., Mack, M., O'Donnell, J., Ping, C., Schuur, E.A.G., Turetsky, M.R., Valentine., D.W. 2011. Soil carbon distribution in Alaska in Relation to Soil-Forming Factors. Geoderma 167-168, 71-84.

Michaelson, G.J., Ping, C.L., Kimble, J.M., 1996. Carbon storage and distribution in tundra soils of arctic Alaska, U.S.A. Arctic and Alpine Research 28(4), 414-424.

Ping, C.L., Michaelson, G.J., Kimble, J.M., 1997. Carbon storage along a latitudinal transect in Alaska. Nutrient Cycling in Agroecosystems 49, 235-242.

VanRemortel, R.D., Shields, D.A., 1993. Comparison of clod and core methods for determination of soil bulk density. Communications in soil science and plant analysis 24(17-18), 2517-2528.