

Title: Protocol for merging SSURGO and SCD data
Prepared by: Adam Skibbe
Date: 7/1/2007

Intro: This protocol arose from a need to create a soils dataset that had a selection of the data found in the NRCS SSURGO (<http://www.ncgsc.nrcs.usda.gov/products/datasets/ssurgo/>) dataset as well as appending several other properties not supported by SSURGO. For our purposes, these other data were found in the Soils Characterization Data (SCD), from the NRCS Soil Survey Laboratory (SSL) (<http://ssldata.nrcs.usda.gov/default.htm>). In addition to appending these other data, we also required that the SSURGO data be broken down into a more manageable subset of its original database.

This protocol describes how to ...

- 1) Reduce the SSURGO dataset to a simplified selection showing only each soil series' primary soil type and associated data.
- 2) Obtain SCD for each of the primary soil types to fill in gaps for chemical properties not already included in the SSURGO dataset.
- 3) Append supplementary SCD data to the reduced SSURGO dataset.

Part 1:

SSURGO data is delivered as a comprehensive, spatial soils dataset. SSURGO contains data on soil series, referenced by a unique id called a Map Unit Symbol (MUSYM). Depending on the specific series, the MUSYM may include data on more than one soil type. The Map Unit Key (MUKEY) is the code that identifies where the data originated from.

There are several issues with the SSURGO dataset which impact how it can be used in light of our objectives. Because there is not yet a single statewide dataset, SSURGO is provided for, and is specific to each county. In studies with a more regional approach this will cause problems in the area of redundancy. If more than one county has a particular soil series, there will be multiple MUKEYs for a given MUSYM from which to choose.

SSURGO data is presented as both spatial (a shapefile) and tabular (a Microsoft Access database) data. To avoid getting lost in the large dataset a Query was created in Access to pull from the database the specific data of interest. This process creates a subset of the database which contains those portions of interest to the user so that the following steps are less convoluted.

To select an example of the main soil type in each series we looked at each series MUKEY to identify one which was both a) complete, in the sense of a complete profile and b) preference for one whose horizons were named (ex: A, BA, Bt, C as opposed to Hz1, Hz2, etc.).

Once a MUKEY was selected, the main soil type was identified by selecting the horizons representing the highest component percent (Comp %). For example; for the Chase silty clay loam, rarely flooded, MUSYM 4350, soil series, the MUKEY of 1472302, contained a soil with the horizons Ap, AB, Bt, BC, and C and which constituted 84% of the composition of the soil series. This was chosen as it was the single soil type in the series which was most representative of that particular soil series.

The data was exported to a Microsoft Excel document so that other soil series data could later be added.

Part 2:

The SCD contains many of the chemical properties not found in SSURGO. The database consists of survey information on soil pedons, or individual sites, which have gone through laboratory testing to determine specific properties and to scientifically determine soil type.

There are several things to consider when using these data. First, these data reflect specific pedons, points, or soil information reflecting a specific location. This is in contrast to SSURGO which represents an average of values over a spatially defined area. Secondly, there are a limited number of pedons that have been surveyed, so there are not many to choose from. Third, data are not standardized across all surveys. Some of the methods used for data collection have become obsolete over the years and replaced one or more times. Data is left in its original form and thus there may be a variety of methods used depending on specific pedons. Fourth, data are sometimes incomplete. Not all readings are taken for all pedons. Also, many pedon surveys contain “null” values. And finally, SCD data does not always exist for a soil type; several have not yet been analyzed by the SSL.

When choosing the dataset for a particular soil type there is a few things to note. Included in the search criteria is an drop down box to select the “type” of record. The options include “sampled as”, “field”, “unchanged”, “SSL” and “correlated”. Of these, if possible, “correlated” is the best option as it reflects a soil that has been identified through the lab as being the soil type listed. In comparison, a soil which is “sampled as” may have been thought to be a particular soil type, but is not necessarily confirmed to be

so. In addition, due to changing soil survey methods, the newest “correlated”, if there is more than one, should be selected.

Part 3:

Our data needs required that we append SCD data for a series of chemical properties to the existing SSURGO data. SSURGO was chosen as the primary data source because it is already spatially explicit and contains the majority of necessary data.

The merging process is based first on soil horizon name, then by horizon depth. Soil horizons are named to represent their physical and chemical make up, therefore an “A” horizon is by definition different from a “C” horizon. Because the two soil datasets are different, there are often discrepancies in how horizons line up. Sometimes horizons will be omitted from one dataset or the other, or one horizon in one dataset may represent multiple horizons in another.

When horizons line up perfectly between datasets, for example Ap and Ap, both with roughly similar horizon depths, the SCD data is simply imported into the SSURGO table for that horizon.

If horizons do not perfectly line up, a more qualitative approach is necessary for merging. Horizons should be placed together in like pairs if possible, if this is not possible... Where 2 (or more) SCD horizons fit to 1 SSURGO horizon the average of those data is used. If one of the two (or more) is not available (null, no data, ---, or tr.) then the value that is present is used alone. In situations where 1 SCD horizon fits 2 (or more) SSURGO horizons the SCD data will be used more than once. Finally, if one set (either SCD or SSURGO) has its deepest horizon that does not match the depth or name

of the other data set... it was excluded. No data was left as such and was not weighted in averaging.

Example of one to one merging

<i>SSURGO Table</i>					
MUSYM	Mapunit Name	Horizon	Top Depth	Bottom Depth	MUKEY
4350	Chase...	Ap	0	16	1472302

<i>SCD Table</i>					
Pedon	Soil Type	Horizon	Top Depth	Bottom Depth	Total N
83P0729	Chase	Ap	0	16	.14

<i>Merged Table</i>					
MUSYM	Mapunit Name	Horizon	Top Depth	Bottom Depth	Total N
4350	Chase...	Ap	0	18	.14

Example of multiple SCD horizons being merged

<i>SSURGO Table</i>					
MUSYM	Mapunit Name	Horizon	Top Depth	Bottom Depth	MUKEY
8683	Dennis...	A1	0	23	1454604

<i>SCD Table</i>					
Pedon	Soil Type	Horizon	Top Depth	Bottom Depth	Total N
86P0869	Dennis	Ap, A	0	27	.1395*

* Indicates average of multiple horizons

<i>Merged Table</i>					
MUSYM	Mapunit Name	Horizon	Top Depth	Bottom Depth	Total N
4350	Chase...	Ap	0	18	.1395*

* Indicates average of multiple horizons