Regenerative DesignGroup
a worker-owned cooperative

Progress Report

Soil Organic Carbon Mapping
Ref# C140

March 18, 2025

Dear Tom Anderson,

Thank you for the opportunity to develop the Soil Organic Carbon Mapping Project. We have made significant progress and our analysis is in its final stage. In particular, we have made great strides in developing and implementing a strategy to maximize the utility of SSURGO data while minimizing its shortcomings. Please find a table below a table with the status of each deliverable, followed by a more detailed description of our analytical workflow. We are preparing to complete our analysis in the next few weeks, and begin working on any revisions suggested by ongoing feedback. At the same time we will begin organizing the workshop where we will explain our process and the best use of project outputs, once our outputs are finalized.

Table 1. Deliverables and Status

Deliverable	**March 18, 2025 Status**
1. A table of land cover SOC impact factors that represent conversion factors between the land covers classes in the 2016 High Resolution Land Cover dataset.	1. This table will be produced as a summary/synthesis of deliverable #3, below. It will consist of a table of area-weighted proportional differences in SOC between pairs of landcovers. These proportions will constitute estimates of SOC impact by land conversion.
2. A GIS layer covering the extent of MA that represents the SOC under the land cover with the highest SOC potential. In upland soils in MA this is forest and in wetland soils this is wetland.	2. This analysis, as with deliverable #3 below, is in its final stage.We have calculated landcover-based SOC values for every mapunit and are ready to apply them to the intersection of the SSURGO mapunit geometry and the MA 2016 Landcover dataset. Our strategy is described in more detail below.
3. A GIS layer covering the extent of MA that represents the estimated SOC under the land cover identified by the 2016 High Resolution Land Cover dataset.	3. This deliverable is being produced in tandem with deliverable #2.

4. A workshop on how these layers were produced and how to utilize them for planning and analysis purposes. This workshop will be recorded and an edited version of the recording will be provided and made available to the public.	4. We are in the process of creating the documentation and detailed explanations of our process that will be the foundations of these workshops.

Analytical workflow

Bootstrapping SSURGO data

The central strategy of this project is to extend the usefulness of SSURGO data by intersecting it with the 2016 Land Cover dataset to produce a fine-grained, spatially explicit, landcover-adjusted estimate of SOC.

In our discussions with NRCS representatives (chiefly Maggie Payne and David Zimmerman), we identified two key factors in our efforts to make the best use of available SSURGO data.

1. Data for *major components* are generally regarded as more accurate and reliable than that for minor components.
2. Data for components that were updated in the Major Land Resource Area (MLRA) Data Harmonization project in the 2010-2013 era are regarded as significantly more accurate and reliable than others, which suffer from inconsistencies in data acquisition, estimation, and entry, across different epochs, offices, methods, and NRCS officers.

In response to this information, we decided to shape our strategy around two tactics:

1. Using major component data exclusively to determine SOC density rates.
2. Extend the upgraded SOC density data from MLRA components to other, non-MLRA components when they share (1) component name or (2) taxonomic subgroup; and cover.

Additionally, we identified a third tactic to extend the usefulness of the SSURGO data. Use of SSURGO for landcover- based analysis of SOC is complicated by the fact that 11% of the rows in the component table have no entry in the primary land cover variable, earthcovkind1. These rows represent ~30% of the area described by the dataset. We therefore developed several approaches to impute cover data. Primarily, we analyzed cover by component name. We created a lookup table based on the variable compname, containing an entry for every compname that (based on available cover data) consisted of 75% or more of a single cover type. For every component with a missing cover, we checked the lookup table and assigned the imputed cover value, when available. Additionally, for components still lacking cover data, we checked for the terms “Pits” or “Outcrop” in compname, and assigned the cover value “Barren land” if present, and checked for the word “Water” in compname, and assigned the cover value “Water cover” if present. Together, these imputation techniques reduced missing cover by 70%. Most importantly for our analysis, they increased the area with crop cover by 5.6%, tree cover by 18%, and grass/herbaceous cover by 28%.

Developing Mapunit-level SOC Values for 2016 Landcover Categories

Component data are spatially vague and low precision, with spatial data (including cover) offered only as a blanket percent coverage applied across multiple distinct and varied mapunits. Especially when using only dominant component data, each mapunit is only associated with a small number (1-3) of landcover types. However when intersected with 2016 Landcover, each mapunit will be broken into a large number of 1m pixels assigned to any of the 19 NLCD landcover classes. After dropping the aquatic landcover categories (Open Water Palustrine Aquatic Bed, and Estuarine Aquatic Bed) that are not used in this analysis, this leaves 16 potential landcover classes for each mapunit.

Our approach is therefore to assume that every mapunit may be crossed with any of these 16 landcover categories, and to assign each mapunit its own most plausible value for each class. To do this we developed crosswalk categories corresponding to a 2016 Landcover/NLCD class on one hand, and a set of SSURGO identifiers on the other. We assigned a crosswalk category to every SSURGO component based on its characteristics – primarily but not only cover. See **Table 2,** below.

For each SSURGO polygon, we calculated SOC values for each crosswalk category using the closest available values for that category. In order of preference, values for each polygon were taken from:

1. **SSURGO components** from the **SSURGO mapunit** to which that **SSURGO polygon** belongs;
2. **SSURGO components** associated with **SSURGO mapunits** overlapping with the **STATSGO polygon** in which that **SSURGO polygon** is found;
3. **SSURGO components** associated with **SSURGO mapunits** overlapping with the **STATSGO mapunit** in which that **SSURGO polygon** is found.

At any point when multiple SSURGO components are included in the crosswalk category, the area-weighted median SOC value is calculated. For impervious cover, SOC values were derived from the SOC for forest cover, with the SOC value for the top 30cm subtracted from all depths (thereby setting values for the top 30cm to zero).

The above operations are conducted at the SSURGO polygon level because the multiple polygons of a SSURGO mapunit may fall into different STATSGO areas, and therefore may be linked with different sets of nearest available values for any given crosswalk category. After calculating polygon-level SOC values for each crosswalk category, mapunit-level SOC values are calculated as the area-weighted median for all polygons in each mapunit.

The result is the Current SOC Values table, with SOC values (at 30cm, 100cm, and 999cm) for most crosswalk categories for each mapunit. The method does invariably produce NAs for some crosswalk categories in some mapunits, e.g. the 'beach' category for inland mapunits. In the next step, when this table is joined to the intersection of SSURGO geometry and the 2016 Landcover raster, we will identify whether any of the missing values are needed, i.e. whether they represent actually existing landcover pixels for the given mapunit. In that case, the 3-level scale of nearest available values can be expanded to include more general values (i.e. moving from vpd_tree to an all-landcover pooled_vpd value) or expanded to include a broader region (i.e. moving from STATSGO mapunit to Land Resource Region) to generate the closest, most plausible available value for any given mapunit.

Table 2. Crosswalk Categories

2016 Landcover/NLCD Class	Crosswalk category	SSURGO Identifiers (including imputed cover)
Impervious	impervious	(Forest minus top 30 cm)
Developed Open Space	other_herb	Grass/herbaceous cover:Other grass/herbaceous
Cultivated Land	cultivated	Grassland/herbaceous cover:Row crop with Shrub cover:Crop vines
Pasture/Hay	pasture_hay	Grassland/herbaceous cover:Tame pastureland with Grassland/herbaceous cover:Hayland
Grassland	pooled_herb	Grass/herbaceous (all)
Deciduous Forest	forest	Tree cover & not Very Poorly Drained
Evergreen Forest	forest	Tree cover & not Very Poorly Drained
Scrub/Shrub	shrub	mean of forest and pooled_herb
Palustrine Forested Wetland (C-CAP)	vpd_tree	Tree cover & Very Poorly Drained
Palustrine Scrub/Shrub Wetland (C-CAP)	vpd_shrub	Shrub cover & Very Poorly Drained
Palustrine Emergent Wetland (C-CAP)	vpd_herb	Grass/herbaceous cover & Very Poorly Drained
Estuarine Forested Wetland (C-CAP)	vpd_tree	Tree cover & Very Poorly Drained
Estuarine Scrub/Shrub Wetland (C-CAP)	vpd_shrub	Shrub cover & Very Poorly Drained
Estuarine Emergent Wetland (C-CAP)	vpd_herb	Grass/herbaceous cover & Very Poorly Drained
Unconsolidated Shore	beach	Beach (from compname), cover is Barren:Sand and Gravel
Bare Land	barren	Barren, except for identified as beach from compname, above
Open Water	dropped	None

Palustrine Aquatic Bed (C-CAP)	dropped	None
Estuarine Aquatic Bed (C-CAP)	dropped	None

Estimating Potential SOC by Adjusting Non-forest Vegetative Cover to Forest Levels of SOC

Using Rapid Carbon Assessment to Produce Forest Adjustment Factors

In order to produce Forest Adjustment Factors for the SOC Potential map (deliverable #2), we first identified the Land Resource Regions (LRR) that intersect with Massachusetts: the Northeastern Forage and Forest Region and the Northern Atlantic Slope Diversified Farming Region. We then extracted SOC and landuse data for all the Rapid Carbon Assessment (RaCA) sample points that lie within those regions, and assigned them simplified landuse categories:

- non-forest (C, P, X)
- forest (F)
- wetland (W) - excluded from analysis

We then calculated median SOC stock, at 30cm and 100cm, for each simplified landuse category, within each LRR. At each depth and within each LRR, we divided forest SOC stock by non-forest, to calculate the forest normalization factor for adjusting non-forest vegetative cover to forest levels.

Adjusting to forest

We assigned Forest Adjustment Factors to mapunits by calculating area-weighted mean adjustment factors for each mapunit, based on the distribution of their polygons across the two Land Resource Regions. The Potential SOC Values table can then be produced by taking the Current SOC Values table, described above, and multiplying all non-forest upland vegetation values by the Forest Adjustment Factor for that mapunit.

Producing the Current and Potential SOC Maps and Statewide Estimates

In the next and final step of the analytical process, each of the two SOC Values tables will be joined to the intersection of the SSURGO geometry with the 2016 Landcover raster. For each map (Current and Potential SOC), each raster cell will be assigned the value corresponding to landcover and mapunit in its respective SOC Values table.

This process should be complete within the next two weeks.