**Implementation of SHAPE scoring**

Soil Health assessment value - % soil organic carbon

Sample - %SOC and geographic location (lat lon) – 5 to 10 other properties to be added using the same framework

|  |  |  |
| --- | --- | --- |
| “Intake” process for DSP Hub and parameters – this needs some “data planning” with the customer of the data to get expectations, needs, and plan out | Value | Notes |
| Dynamic Soil Property | Soil organic carbon, Percent | Raw data may come in from different methods |
| Data collection parameters and methodology |  | SHD has excel spreadsheet  SPSD will bring in data through Lab and Pedon Data Mart  Need to convert from input to consistent property value (for instance SOM to SOC, or between methods)  Pre-analysis need to have rules to merge datasets  Need to maintain source and metadata throughout merge |
| Data analysis notes |  | Join to covariates – through spatial and tabular |
|  | Soil Health Division |  |
| Data need- questions that need to be answered |  |  |
| Planned Use of data | Soil Health Assessment  Aggregations by soil, mlra, CMZ, political bourndry |  |
| Desired scale of decision making | Data value to data value  Future use at multiple scales | Field and individual interpretation  Shape score is focused on transforming one SOC% to one score |
| Baseline and delta plan |  |  |
| Method | Laboratory analysis (need more specifics) | We need a lot more detail here to align to various data tables and still parse out the methods for decision making on quality and consistency |
|  |  |  |
| How the data needs to be served up (delivery – Tableau?) |  |  |
| For Office Use Only Section – the detailed data review needed to load into the DSP Hub Covariate Databases or Reference Tables – these items will need to be communicated back to the customer in the form of a required data standard and minimum metadata, along with any technical requirements that they need to develop (e.g. potentially a cross-walk between the SH Min Data Set and CR-LMOD tables) | | |
| Data interoperability/data standard needed  (this is the fetch and join stuff we need to start pushing “upstream” by using a data standard so it shows up at the doorstep in the right format) | Polygon for either project site or the treatment/practice polygon, depending on the experimental design/hypothesis and sampling plan | Align to PLU or practice polygon concept depending on what is determined from the needs of the experimental design |
| CR-LMOD crosswalk tables to conservation practices and any key parameters needed | We need some analysis on “lumping” tables that become master reference data on the conservation practice parameters from CR-LMOD to align to CPDES. Once we figure out what these lumping categories are for the “treatments” we can determine where the values need to reside (CPDES, CR-LMOD, other data tables from ARS, etc) |
| Any land unit or site covariates needed | Irrigation, cropping history, land use |
| Climate variables at that site: | Extract MAP and MAT from PRISM 30 year average rasters (or equivalent) by location  It would be nice to pull in local data from SCAN  Eventually, we should be able to pull in and use local information for context (all soil samples, climate data etc.) – weighted |
| Land cover variables at that site: | This should align to the land unit tables with some parameters – seems like we will need to develop and own some reference tables on this |
| Other soil data - SDA or Intersects |  |
| Rules for acceptable entry | - %, 1 decimal pt, 0 – 60, layer depth top = 0 |
| Required metadata for parameterization | value method, source, any local overrides (like texture) of generalized data |
| Required detail on data collection plan to parameterize and design correct assumptions | Process is for surface samples – need depths of sample collection |
| Transformations from other methods and outputs | SOC = Organic matter/ 1.72 |
| Select/develop authoritative processing and statistical methodology | Perform spatial join | This seems like we are creating a new product – we just need to document it and then see what level of “peer review” needed |
| Reclassify covariates | Which of the above is this and when can it be done on the workbench/process – upstream with master reference tables or is this something that needs to be processed each time?  Could be loaded, but should be updated at least each time SSURGO is refreshed – these are fairly simple but may require that data be clipped/cleaned etc.  Soil suborder groups:  S1: Fribists, Folists, Hemists, Histels, Saprists, Wassists  S2: Aquands, Aquents, Aquepts, Aquods, Aquoxs, Cryods, Humods, Orthels, Peroxs,  Torrands, Tropepts, Turbels, Udands, Udoxs, Ustands  S3: Albolls, Andepts, Aquolls, Aquults, Cryands, Cryepts, Cryolls, Gelepts, Gelolls,  Humults, Rendolls, Umbrepts, Ustoxs, Vitrands, Wassents, Xerands  S4: Aqualfs, Aquerts, Boralfs, Borolls, Cryalfs, Ochrepts, Orthods, Orthoxs, Udalfs, Udepts,  Uderts, Udolls, Usterts, Ustolls, Xeralfs, Xerepts, Xerolls, Xerults  S5: Arents, Argids, Calcids, Cambids, Cryerts, Cryids, Durids, Fluvents, Gypsids, Orthents,  Orthids, Psamments, Salids, Torrerts, Torroxs, Udults, Ustalfs, Ustepts, Ustults, Xererts  Texture Groups:  Class T1 sand, loamy sand, sandy loam (with <8% clay)  Class T2 sandy loam (with clay >8%), sandy clay loam, loam  Class T3 silt loam, silt  Class T4 sandy clay, clay loam, silty clay loam, silty clay, clay (<60%)  Class T5 clay (>60%) |
| Apply statistical function: Currently an R script with a shiny app for singular use: cooperators are currently working to make this into a batch mode | Currently under peer review: https://paparker.shinyapps.io/ISHI\_app/ |
| Evaluation/validation | Evaluate results of processing (is the results of the process a reference table of values by treatment/practice and land cover?) | We need to figure out the details of this step during our MVP process |
|  | Use scores to evaluate the outcomes/effects of management systems, land cover and conservation practices | We will need to develop caveats and confidence “levels” and appropriate scale of use, decisions-support limitations, etc |
|  | Initially – we need to ‘report’ back the score results to submitter.  An interactive interface could be developed. – I think this might be on a 1:1 report | SHD really needs to outline how they want these data to come back.  Rules for acceptable output – 0 – 1, 2 decimal points |
|  | Then want to aggregate by metadata classes from CR-LMOD and conservation practices |  |
|  | Then we might want to aggregate spatially (mentioned above |  |
|  | Then we might want to apply geospatial modeling and statistical analysis (anova, mixed models, linear regression, random forest etc.) |  |

**Current process:**

**Step 1.**

**Get the input data**

1. Transform input data into common SOC units (if a commercial lab reports OM we can make a mathematical assumption).
2. Integrate metadata categories with CR-LMOD lists (conservation practices?)
3. Gather scoring covariates (depending on how the values are retrieved this may
   1. Climate and land cover
      1. Go to restricted BOX account (some layers are from outside sources, we can’t serve the) - [(2) 100m\_covariates | Powered by Box](https://nrcs.app.box.com/folder/125444425517?s=p3dcg9lyw6ocmmlhtobm04ticax4thj5&utm_campaign=collab%20auto%20accept%20user&utm_medium=email&utm_source=trans)
      2. ID relevant coverages in SoilGrids\_USA48\_COvs100m.csv
         1. P01PRI5.tif (from PRISM)
         2. T01PRI5.tif (from PRISM)
      3. ID relevant Landcover\_lagend.csv – Go to restricted BOX account (some layers are from outside sources, we can’t serve the)
   2. Other Soil Info
      1. SDA query or GIS intersection
         1. Suborder
         2. Surface texture

**Process Data**

1. Perform spatial join – add covariate values to sample point (done with python by others, or in ArcGIS by me)
2. Reclassify covariates (excel or SQL)
   1. Mean Annual PPT – leave as continuous
   2. Mean Annual Temp – leave as continuous
   3. Surface Texture – apply reclassification into 5 groups
   4. Suborder – apply reclassification into groups

**Apply Function**

1. Apply statistical function: Currently an R script with a shiny app for singular use: cooperators are currently working to make this into a batch mode

Currently under peer review: [https://paparker.shinyapps.io/ISHI\_app/](https://gcc02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fpaparker.shinyapps.io%2FISHI_app%2F&data=04%7C01%7C%7Cd9a83eb235b74295906d08d89c4b8043%7Ced5b36e701ee4ebc867ee03cfa0d4697%7C0%7C0%7C637431195368594437%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=gzdg6UTWQdoMZmPyhofQ3CkYnDwcQth0b7GiFV7yOt0%3D&reserved=0)

**Evaluation**

1. Use scores to evaluate the outcomes/effects of management systems, land cover and conservation practices

**Additional Thoughts:**

* This process will be similar for all soil health properties that need to be scored.
* The process needs to both be automated and allow for testing and flexibility
* The statistical function will be unique for each property – the development of the function could be assisted by streamlining the data processing step
  + Fetching and joining data is really cumbersome