**Project Proposal**

Title: *Examining the extent which differing algorithms used in spatially summarizing soil components vary the reporting of soil water holding capacity*

Notice: Dr. Bryan Runck

Author: Michael Felzan

Date: Feb. 13th 2022

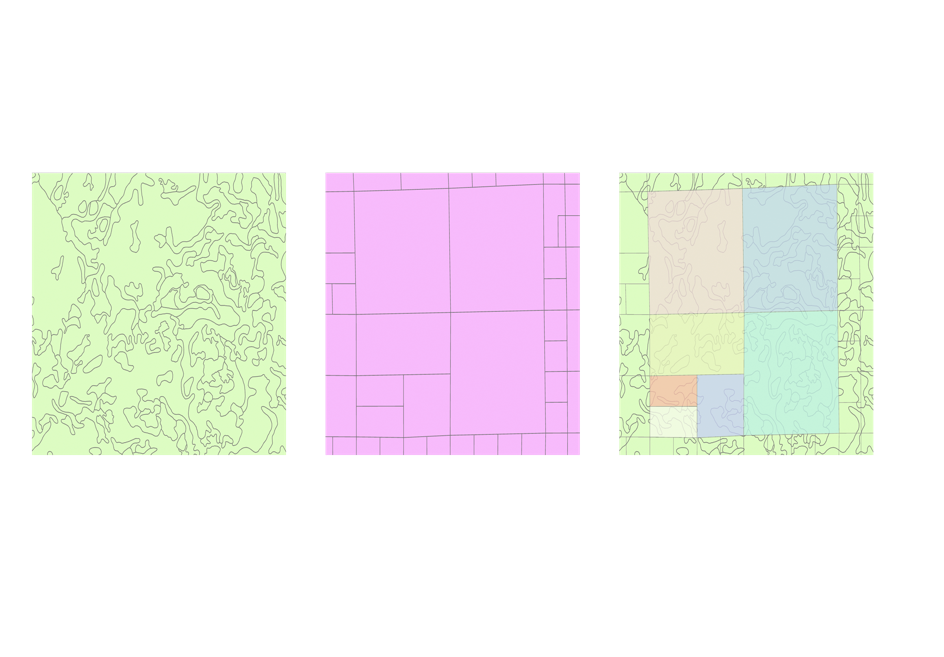
**Project Repository:**

*https://github.com/fezfelzan/soil-water-holding-capacity.git*

**Google Drive Link:** *https://drive.google.com/drive/folders/177BDi2wi\_lLcki5Tex4st2x8SwGpZHt4?usp=sharing*

**Problem Statement**

Soil water holding capacity can be calculated multiple different ways from individual gSSURGO polygons aggregated to the field-level. Our goal is to understand how the different ways of summarizing this within field spatial variability results in different field-level soil water holding capacity estimates (MAUP research objective).

 We are in particular interested in Soil Water Holding Capacity 0-12, 12-24, 24-36, and 36-48 inches (is this supposed to be centimeters?). This project aims to produce a map for each depth interval (total maps total) which depict the difference between minimum estimate and maximum estimate for each field. This study additionally concerns the development of Python Jupyter Notebook products to be used for accessing, querying, and manipulating SSURGO data, aggregating Available Water Storage (AWS) values from gSSURGO polygons to MN parcel divisions (via a spatial interpolation method designated by the user), and a tool which takes custom soil depth intervals (12-24cm) as an input to calculate AWS for SSURGO polygons and MN parcels.

Specific methods to consider include:

* If more than 80% of field area, use dominant soil type
* Area weighted average (most common spatial analysis technique)
* If the lowest soil water holding capacity area is >40%, use the weighted average

The two main research questions explored in this study are:

1. How does weighted averaging order impact field-level soil water holding capacity outcomes?
2. How does SSURGO calculate soil water holding capacity for each depth range (0-25cm, 0-50cm, 0-100cm, 0-150cm)? **Fig 1.** *Visualization of gSSURGO*

*‘mukey’ polygons and MN Parcel*

*polygons. The aggregated AWS*

*values for each ‘mukey’ polygon*

*will be summarized to the parcel*

*polygon boundaries using a variety*

*of spatial interpolation algorithms.*

**Input Data**

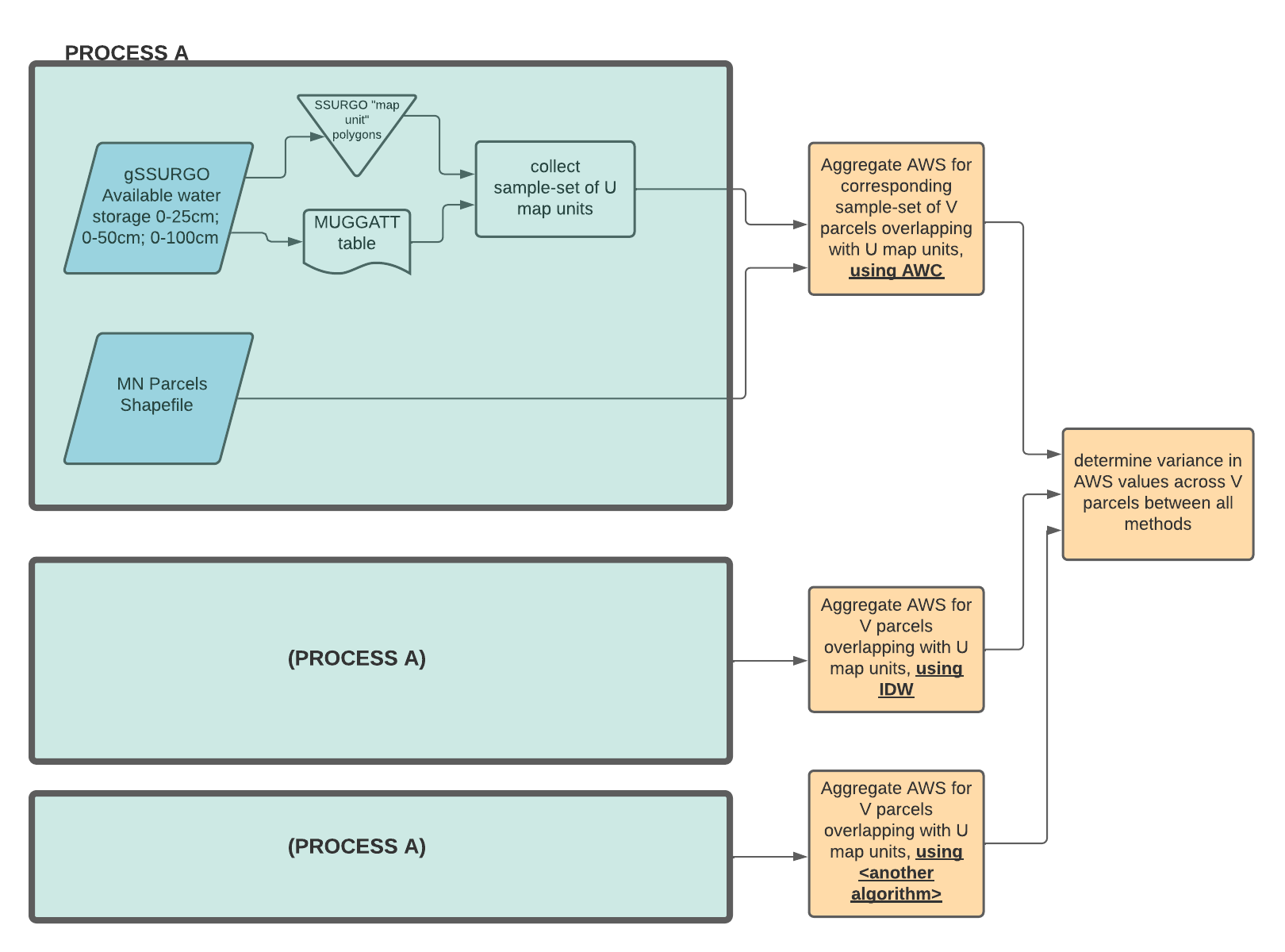
The NRCS gSSURGO dataset is the main input data for this study. This dataset is made up of a series of tables which contain unique soil statistics for a series of map polygons comprising MN. Tables of particular include the ‘muggatt’ table (contains summarized AWS values for each polygon at 0-25, 0-50, 0-100, 0-150cm), the ‘chorizon’ table (contains ‘hztept’/‘hzdepb,’ distance from top/bottom of soil horizon to surface), and the ‘component’ table (contains ‘comppct,’ % of component with gSSURGO map unit). Additionally, the MN County Parcel dataset (available via the University of Minnesota) will be used to examine the effects of various algorithms in summarizing gSSURGO map unit AWS values to larger parcel divisions.

**Table 1.** *Input data*

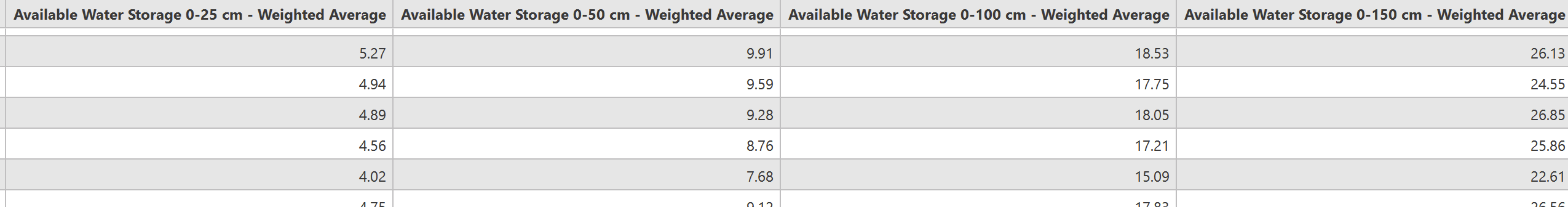
|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 | Statewide soil composition (XY&Z dimensions) (via gSSURGO dataset) | Parsing out various soil component percentages across the extent of each gSSURGO “map unit” and at varying depths within units (0-25, 0-50cm, etc.) to address the question of how SSURGO arrives at their available water storage calculations aggregated by map units (*research question 2)*. | [USDA-NRCS Soil Survey Geographic Database](https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/home/?cid=nrcs142p2_053628) |
| 2 | Available Water Storage 0-25cm; 0-50cm; 0-100cm; 0-150cm (via gSSURGO dataset) | SSURGO provides Available Water Storage (AWS) values for each of their map units, aggregated at the depths of 0-25cm, 0-50cm, 0-100cm, and 0-150cm in their ‘*muggatt*’ table. This table will be used to assess if the use of different algorithms in summarizing map unit AWS values to parcel boundaries results in substantial variance between results. | [USDA-NRCS Soil Survey Geographic Database](https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/home/?cid=nrcs142p2_053628) |
| 3 | MN County Parcel Data (Statewide) | Aggregating gSSURGO soil component data to agricultural fields | (Requested and received from [UMN](https://docs.google.com/forms/d/e/1FAIpQLSfJzXRaEpbbUptZuv-ngLZgCYMHZJJdEdm8_LL7etorTzpxjw/viewform)) |

**Methods**

**Research Question 1.**

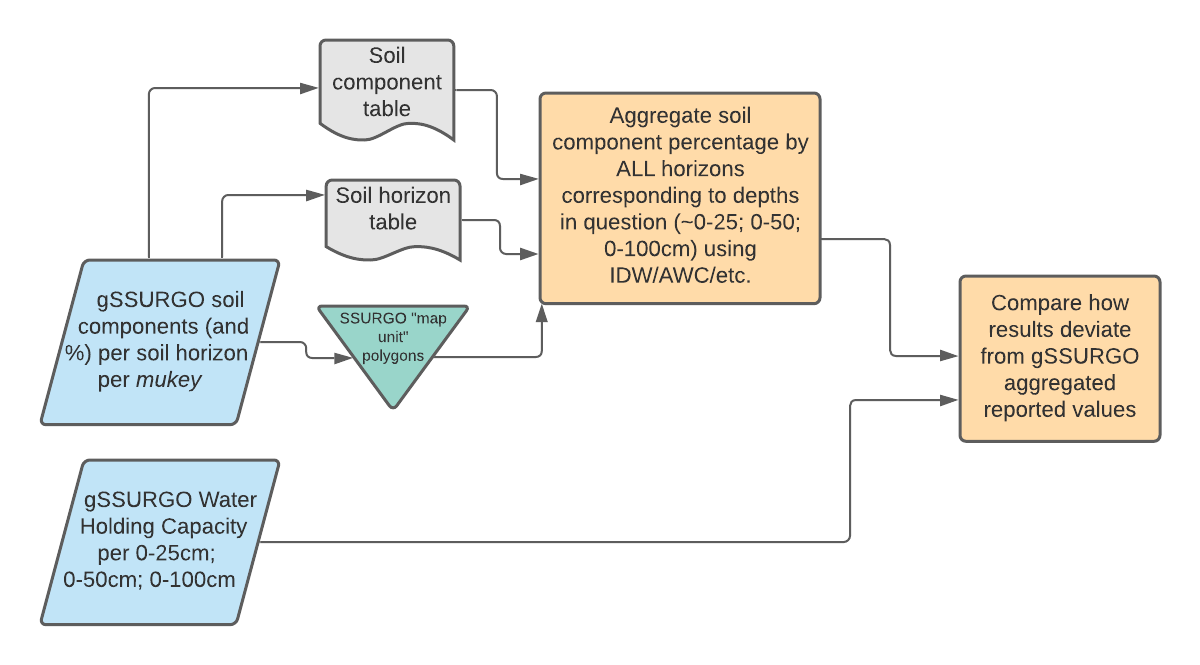
**

The first research question explored in this study concerns the variance between differing algorithms that may be applied to summarize total available water storage (AWS) for MN parcels via SSURGO polygons. The gSSURGO data available from the NRCS includes a multipolygon shapefile of “map units,” and each of these polygons has a unique key (*‘mukey’*) which relates to a set of tables within the dataset. One of these tables, tiled *‘muggatt,’* provides summarized AWS values for each *mukey* polygon, aggregated by the depth ranges of 0-25cm, 0-50cm, 0-100cm, and 0-150cm. Using the state of MN as a study area, this study will quantify the extent of variance between spatial interpolation algorithms (Inverse Distance Weighting, Kriging, Nearest Neighbor, etc.) used to summarize AWS from SSURGO polygons to MN parcels *at each of these four depth intervals.* **Fig. 2** Flow diagram depicting methodology for Research Question I.

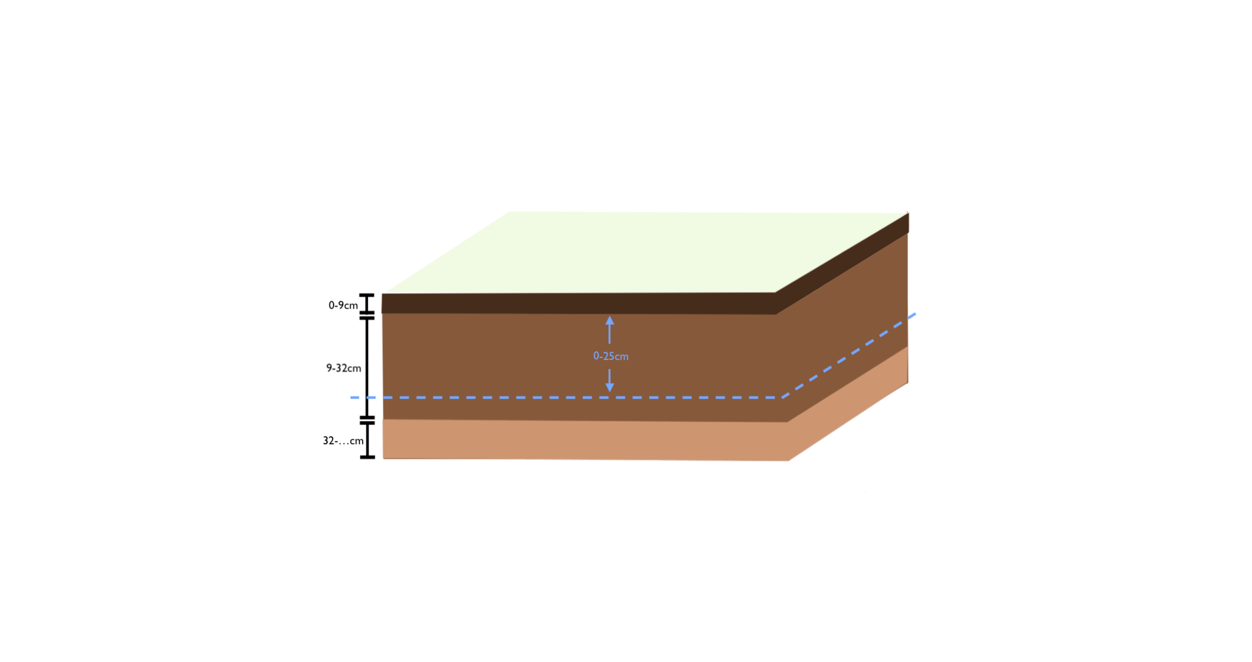
To accomplish this task, a Python Jupyter Notebook will be produced to automate the summarization of AWS per parcel polygon by each algorithm at each of the four depth intervals. This script will first join all of SSURGO’s AWS values for each depth interval to all of the SSURGO map polygons. The following process will involve using either the “Identity” or “Union” tool from the ArcPy toolbox to determine which ‘mukey’ polygons underlay each parcel polygon, and writing this information to the layer’s attribute table. After the ‘mukey’ polygons have been linked to the AWS values at each of the four depth ranges, and the ‘mukey’ polygons have been spatially associated with the MN parcel polygons, the script will then aggregate ‘mukey’ polygon AWS to the parcel divisions, documenting the values achieved by way of each spatial interpolation algorithm for each of the four depth intervals, and the associated standard deviations among algorithm choices within each depth category. The algorithms for IDW, Kriging, and Nearest Neighbor built into the ArcPy toolbox will be used to perform these ****calculations.

**Fig. 2**

**Research Question 2.**

** The second objective of this study is to determine the specific spatial aggregation methods SSURGO used to arrive at the AWS values shown in Fig X. The various components which comprise soils (and their associated presence in the soil by volume) influence the soils ability to retain water (Saxton & Rawls, 2006). There are multiple soil horizons which underlay a single SSURGO map unit polygon, and multiple soil components (series) within a single horizon (need to confirm understanding on this). Additionally, soil horizon compositions are not uniform across space, and there exist gradients between discreetly classified horizons. The matter is further complicated by non-uniform distances from the tops and bottoms of discretely classified soil horizons to the soil surface across geographic space. **Fig. 3** Flow diagram depicting methodology for Research Question 2.

Because soil horizons and the various soil series that comprise them are not perfectly segmented at the depth intervals of 0-25cm, 0-50cm, 0-100cm, and 0-150cm, these soil components must be aggregated among the Z-dimension before summarizing AWC at these rigid depth intervals. (fig v). To constrain the likely methods used by SSURGO to aggregate AWS at each of the four depth intervals (0-25cm, 0-50cm, 0-100cm, 0-150cm), this study will explore the results of applying common spatial interpolation algorithms (IDW, AWC, etc) to the tables provided in the SSURGO dataset, and make inferences about SSURGO’s initial methodology based on the results of our own that closest coincide with SSURGO’s. The determination of SSURGO’s methodology may require the consideration of soil gradients and possible aggregation of component data across XY-space (still need to gain better understanding here).

** Once the method used by SSURGO to calculate AWS at a given depth interval is determined (or approximated), then the method may be applied to calculating soil water holding capacity at any custom depth interval. A Python Jupyter Notebook will be constructed which automates the process of calculating soil water holding capacity for each ‘*mukey’* polygon at the specific depth intervals of 0-12cm, 12-24cm, 24-36cm, and 36-48cm (confirm not inches). After these values have been written to an attribute table, the same script used in *Research Question 1.* will be used to aggregate *mukey* AWS values to MN parcel polygons, and the difference between minimum estimates and maximum estimates, based on differing algorithms used (IDW, etc.), will be determined. Map visualizations of these findings will be produced in Esri ArcGIS Pro.

**Fig. 4** Illustration depicting a possible example where soil component/horizon data would need to be aggregated in the Z-dimension to be summarized by rigid depth interval classification of 0-25cm

**Expected Results**

**Research Question 1.**

The expected results from the first research objective are a table of values documenting the variance of aggregated parcel AWS values between spatial interpolation methods. Additional deliverables will include a Python Jupyter Notebook allowing the replication of methods used and application to other research questions.

**Research Question 2.**

The expected results from the second research objective include documentation of the (probable) processes used by SSURGO to summarize AWS of *mukey* polygons at given soil depth intervals via the SSURGO “component” and “horizon” tables. Secondary results should include a Python Jupyter Notebook which utilizes this methodology to calculate AWS at any depth interval specified by the Notebook user. This study additionally anticipates the calculation of minimum and maximum AWS values (between various interpolation algorithms) specifically at the depth intervals of 0-12cm, 12-24cm, 24-36cm, and 36-48cm (confirm not inches), along with associated map visualizations of these findings.

**Expected Methods for Results Verification**

**Research Question 1.**

To verify the quality and consistency of results produced in RS1, apply methodology (script) to other map areas (eg. Iowa) and assess variance values between study areas?

(note reproducibility via jupyter notebook)

**Research Question 2.**

(Ideally, if results matched with SSURGO’s AWS reporting at each depth interval, enough verification. But can measure deviation between values for each map unit)

(Are there confidence intervals intrinsic to the algorithms used? Is this something this study should develop?)

**References**

Saxton, K. E., & Rawls, W. J. (2006). Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil science society of*

*America Journal*, *70*(5), 1569-1578.