Quantification of outcomes generated using multi-scale geomorphic classification systems in predictive and update modes of digital soil mapping

Karen Vaughan[1], Robert Vaughan[2], Jay Noller[3], Taylor Cullum[4], and Mackenzie Taggart[4]

[1]Ecosystem Science and Management Department, University of Wyoming; [2]Red Castle Resources, USDA-USFS Remote Sensing Applications Center, Salt Lake City, UT; [3]Crop and Soil Science Department, Oregon State University; [4]Natural Resources Management and Environmental Sciences Department, California Polytechnic State University, San Luis Obispo

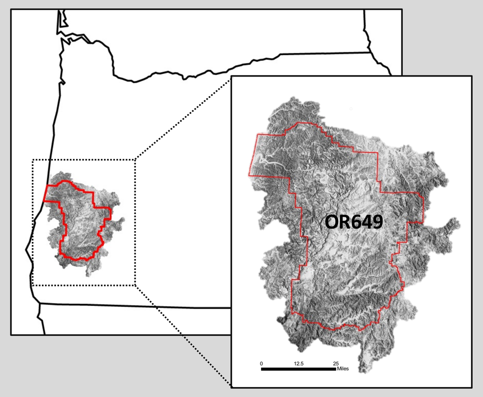
Abstract

We assessed terrain attributes and data mined soil survey databases in an effort to develop a robust, multi-scale, multi-process geomorphological database to be tested and used in predictive soil mapping. The "R" factor of relief (topography, terrain, etc) was isolated by ignoring or holding the other factors (climate, organisms, parent material, time) as constants to predict the soil at any given point purely as a function of the community of the underlying terrain attributes. Information on the geological nature of the interior of the landform, however, was used, for populating the geomorphic process or environment of the geomorphic matrix for said landform. **We are attempting to improve spatial disaggregation of soil-area class maps by incorporating multi-scale geomorphic information derived from existing SSURGO data sources.** Our approach involves the spatial disaggregation of soil map units into their identified components. Geomorphology data and terrain attributes derived from terrain attribute generators will be used as dependent and independent layers, respectively, in implementation of Random Forests. We hypothesize that (i) the application of a geomorphic classification system enhances the predictability of natural soil bodies in landscapes with complex landforms and (ii) outcomes of digital soil mapping are significantly improved where the intrinsic variability of a geomorphic landscape classification system is properly matched to the appropriate spatial scale and complexity of the environment which it is attempting to classify.

Objectives and Goals

1. Develop a workflow to data mine multi-scale geomorphic information from Official Series Descriptions (OSD).
2. Enhance existing SSURGO component level geomorphic information with smaller- scale geomorphic information to enhance soil map unit component disaggregation.
3. Involve undergraduate students in research.

Study Area



- NRCS Soil Survey Area OR649
- Douglas County, Oregon, United States
- 2.25 million acres (911,000 ha)
- 939 map units | association (n= 6) | complex (n=353) | consociation (n= 574) | undifferentiated group (n=1) | water (n = 4)
- Diversity of land uses: production forest, agriculture, and developed areas.

Preliminary Outcomes and Lessons Learned

1. Official Series Descriptions (OSD) are variable in their format (HTML) complicating automated parsing of important information.
2. OSDs are inconsistent with respect to their information content.
3. Soil scientist bias in identifying the correct geomorphic processes through OSDs may continue to limit the usefulness of these data in map unit disaggregation activities. While not perfect, more quantitative and consistent geospatial products may prove more useful.
4. A soil scientist's interpretation of the "genetic" geomorphic processes identified by the originator of the soil series, does not appear to improve on the bias introduced through database entry errors and different interpretations within adjacent survey areas.
5. Regardless of the method used to develop multi-scale geomorphic maps (SSURGO or OSD data) the outcome will still be a reflection of the quality and intensity of the parent SSURGO mapping.

Methods

Geomorphic Data Preparation Steps

Method key: NRCS Soil Data Access Query | Python | R | Microsoft Word or Excel | Manual Process | ArcGIS

1. Create a list of all map unit components identified as a "series" in project area using NRCS Soil Data Access (SDA) Query.
2. Download all Official Series Descriptions (OSD) using "series list."
3. Parse each OSD description into individual sections (e.g. description, taxonomic class, etc.) and combine into single file.
4. Create list of NRCS Geomorphic Description System terms for Process Environment, Landscape, and Landform.
5. Search each OSD description for existence of any term found in NRCS Geomorphic Description System and populate table.
6. Table data clean up, (e.g. sort, order, and format all terms discovered in list).
7. Students review and attribute each OSD for Process Environment, Landscape, and Landform.
8. Query all landform and landscape attributes for all map unit components including those identified as other than "series."
9. Merge student compiled geomorphic settings data with existing SSURGO geomorphic setting data .
10. Data clean up and review. Fill in missing student geomorphic data with SSURGO where available.
11. Subset into individual component level datasets by dominance (by map unit component percent composition).
12. Merge component geomorphic attribute information back to gSSURGO FY15 MUKEY grid.

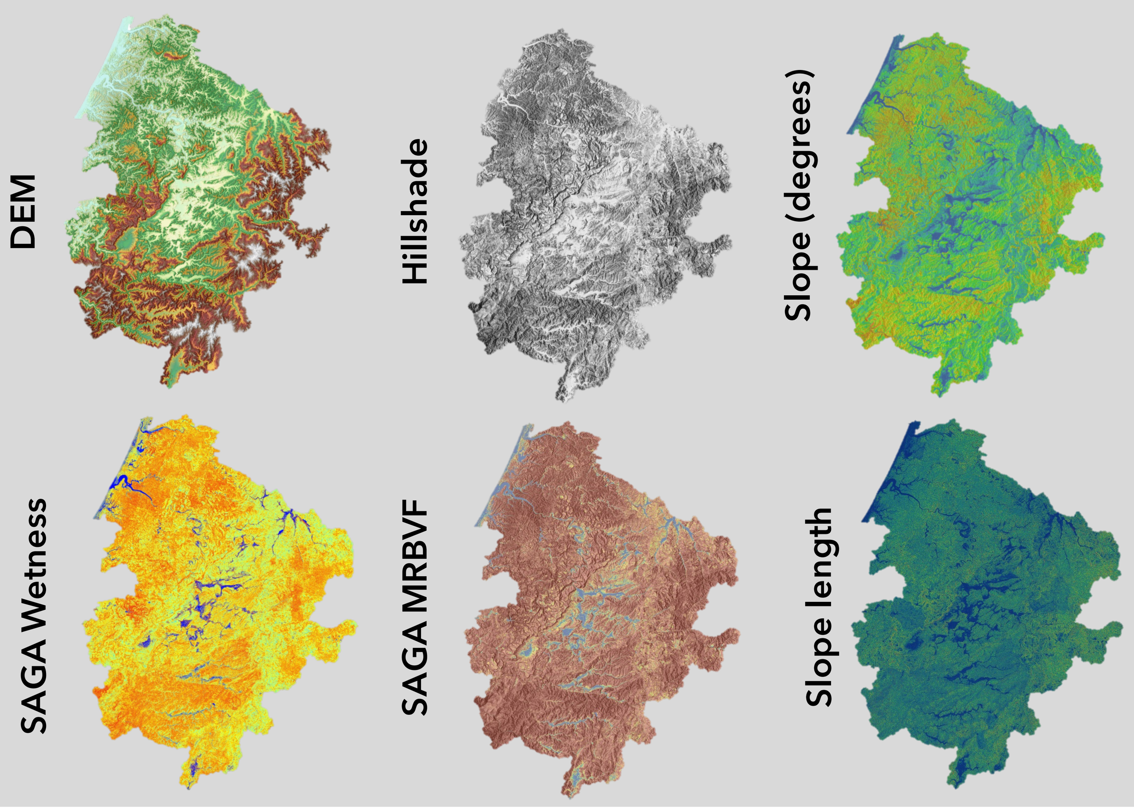
Geospatial Data Preparation

Environmental Covariates

Topographic environmental covariate data were developed using Geographic Information System (GIS) software such as ESRI ArcGIS, SAGA, and GRASS.

A 10 m USGS National Elevation Dataset (NED) digital elevation model (DEM) was used as the primary base dataset. The data processing extent was set to the boundaries of the USGS HUC 10 digit watersheds that intersected Soil Survey Area OR649. For hydrologically-based derivatives (e.g. wetness), the DEM was hydrologically conditioned using SAGA's deepen channels routine.

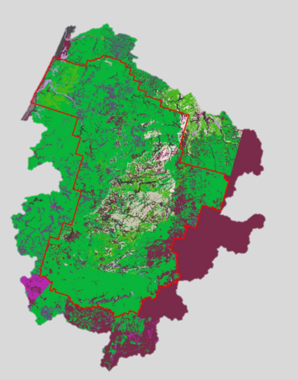
Brief sample of topographic derivatives



Preliminary Results and Discussion

**Legends are omitted due to length. The authors are intending to show the spatial variety and patterns in the data. The red line reflects Survey Area OR649 boundaries within overall study area.*

The geomorphic classification reflects that of the dominant map unit component. Where the dominant map unit component is not a series, the SSURGO geomorphic classification is used (only in the Official Series Description (OSD) data).



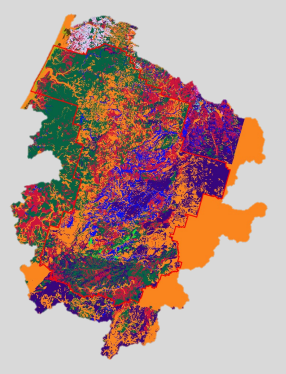
NRCS SSURGO Landscapes

- Broad, consistent geomorphic landscapes are identified in the SSURGO data.
- The lack of landscape diversity identified at this geomorphic scale may not yield much benefit to disaggregation.
- The large *NoData* area (burgundy area) to the east is due a lack of SSURGO Soil Survey data on the Umpqua National Forest.

NRCS SSURGO Landforms

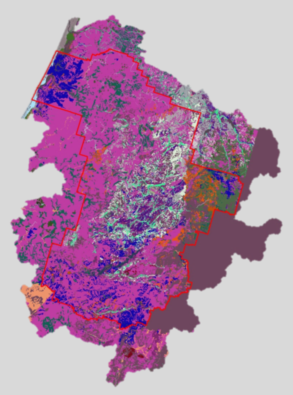
- A greater diversity of landforms found in the central part of the project area is reflective of the finer-scale mapping.
- While not a specific focus of this study, stark differences in the interpretation of landforms can be seen when crossing into adjacent SSURGO survey areas.

Decreasing Spatial Scale



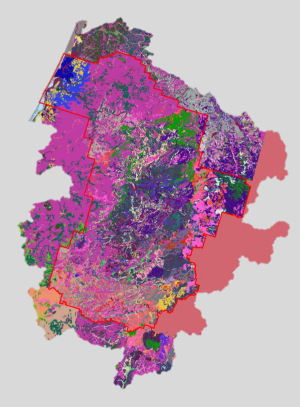
OSD Landscape Process

- Landscape process environments were manually inferred from OSD descriptions. Many OSDs lacked any geomorphic process environment information or were too vague and ambiguous to develop one (orange areas).
- Nevertheless, spatial patterns are present, and appear to be reflective of broad scale geomorphic processes.



OSD Landscapes

- Geomorphic landscapes derived from OSDs appear to show more spatial variety than SSURGO landscapes. However, most likely this is due to the number of unique classes developed during the interpretation of geomorphic setting information within the OSD.
- The appearance of geomorphic setting differences between survey areas appears to be diminished.

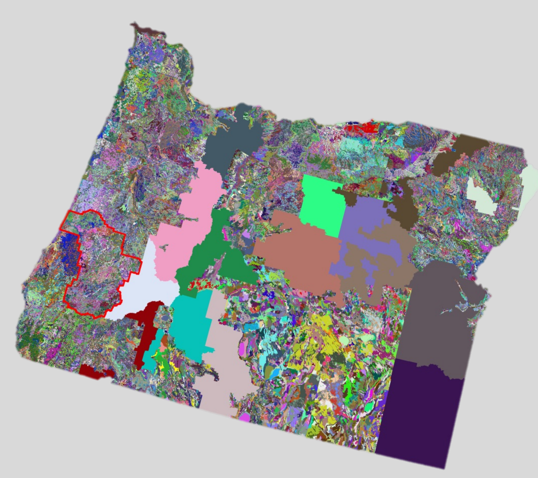


OSD Landforms

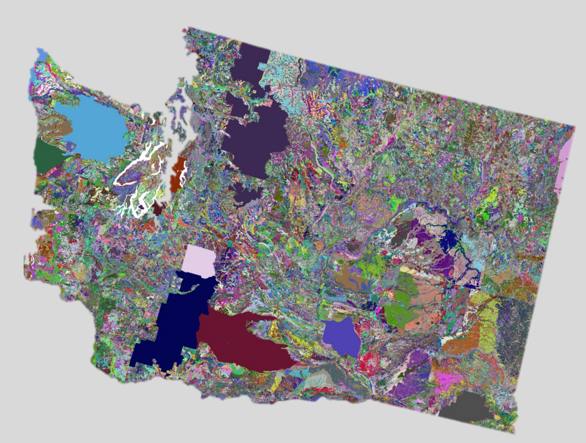
- Large continuous areas of the same landform both within and outside of the survey can be easily identified. This pattern is also evident within the SSURGO landform data. This would suggest that the most important factor in developing geomorphic setting maps has more to do with the underlying map unit polygons than the source of the geomorphic data.

Future Directions

- Soil map unit component disaggregation will be performed using machine learning tools such as Random Forest and Support Vector Machines.
- Environmental covariates along with the multi-scale geomorphic classification maps will be used as model inputs. Variable importance and out-of-bag error (Random Forest) will be used in an effort to quantify the contribution of multi-scale geomorphic information to soil map unit disaggregation results.
- Time permitting, the process will be scaled up to larger areas such as the State of Oregon and Washington. OSDs have been parsed for geomorphic classification data for all series occurring in both Oregon and Washington (approximately 3,000 soil series).



Oregon MUKEY gSSURGO FY15



Washington MUKEY gSSURGO FY15

Funding for this project has been provided by the United States Department of Agriculture, Natural Resources Conservation Service, Soil Survey Division, Soil Survey Research Grants