# Seamless, regionally-specific raster soil property maps to support interpretations

Colby Brungard[1], Travis Nauman[2], Michael Duniway[2],

[1]New Mexico State University, Las Cruces, NM
[2]US Geological Survey, Southwest Biological Science Center, Moab, UT

Research priority number: 2.2.2. New Investigations: Support production of raster soil property maps that support interpretations.

Justification – Supports Soil and Plant Science Division priorities for improving procedures for soil and ecological site inventory.

An important application of soil survey is to provide land owners, land managers, and policy makers with ratings on the suitability or limitations of soil for a particular application. These ratings are delivered as interpretations and have been solely delivered by polygon-based soil data to date. Digital soil mapping (DSM) presents intriguing possibilities for the display and development of soil interpretations (Dobos et al., 2017). There are currently two shortcomings that limit the precision and accuracy of interpretations: 1) the current system requires only data from the SSURGO database, which may not be entirely reasonable for climate or geomorphology data and more authoritative data sources may be more applicable, 2) the interpretive output can only be displayed as aggregated values from the original mapping (Dobos et al., 2017). Interpretations are generally scale-independent so using higher resolution input data with quantified uncertainty is likely to allow greater confidence (or identify areas where confidence is too low to be useful) in the spatial location of the results (Dobos et al., 2017). It is likely that DSM can contribute to these two limitations by providing higher resolution climate, geomorphology, and soil property data with quantified uncertainty.

Current DSM approaches for predicting soil property data for input into generating soil interpretations generally predict soil properties at global, continental or national extents. These top-down approaches use a single 'global' model, meaning that only one model is applied to predict a soil property across the entire spatial extent. For example Viscarra-Rossel et al., (2015); Ballabio et al., (2016); Mulder et al., (2016); and Ramcharan et al., (2018) used single models for predicting soil properties across Australia, France, continental Europe, and the conterminous USA, while Hengl et al., (2017) predicted soil properties across the globe using a single model. We hypothesize that a bottom-up approach using regionally-specific models, rather than a single 'global' model, may provide more accurate national-scale predictions than a single 'global' model. Our initial tests have generally supported this hypothesis (Figure 1). However; making seamless spatial predictions from regional models has proven highly computationally demanding (initial predictions of soil depth classes from nine MLRA-specific models for the entire Upper Colorado River Basin took 19 days using a 32 Core 32 GB RAM 2.8 GHz computer) and may introduce seam artifacts along model spatial boundaries.

Thus we propose to develop 1) a prototype interpretations engine that produces soil interpretations using digital soil maps of key soil properties and 2) a methodology that minimizes the computational cost and border artifacts of regional modeling. It is intended that these approaches will be applicable to help address Soils2026 goals, support the NRCS–DSM team, be applicable in other soil survey regions, and facilitate application of gridded DSM products by soil survey users, including land owners and federal land management agency partners.

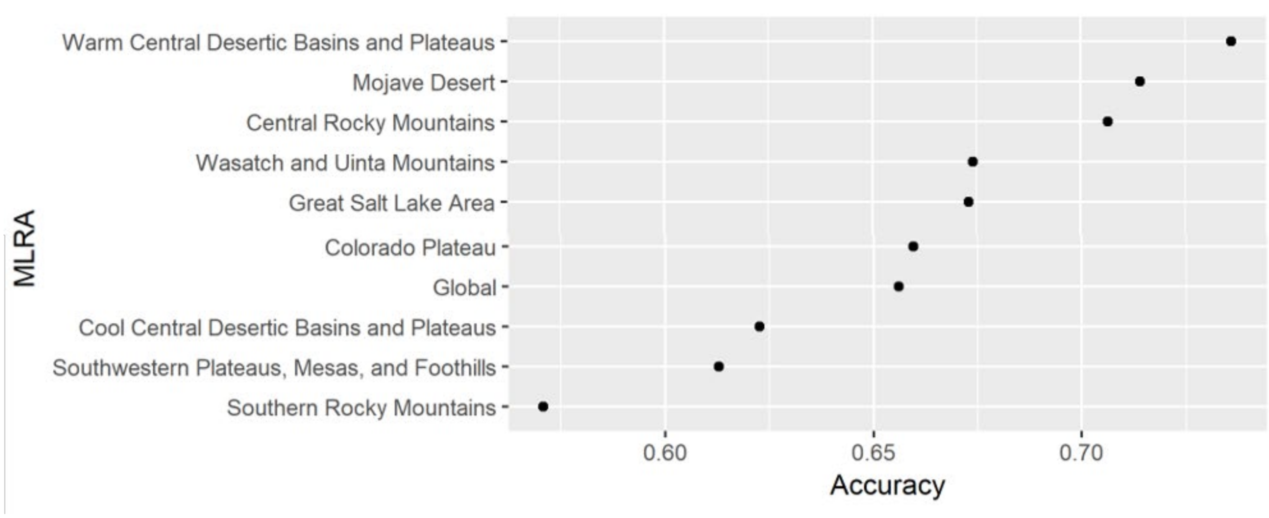


Figure 1. Predictive accuracy for modeling soil depth classes using a global vs. regional approach. 'Global' on the y-axis indicates the global model. This plot shows that five of the nine MLRA-specific models within the Upper Colorado River Basin (UCRB) were modeled more accurately than the global model and three were modeled with less accuracy. Soil depth class observations (n=15,317) were collected for the UCRB. Forty-seven environmental covariates were derived from a 30m digital elevation model and satellite imagery. Soil depth class observations were split into separate training (90%) and validation (10%) sets. The training set was used for modeling, the validation set used for validating model predictions. A 'global' random forest model was first used to model soil depth classes using all observations. Secondly, all observations were subset by MLRA and random forest models were refit for each MLRA. Models fit by MLRA are referred to as regional models. Including MLRA's as a predictor decreased global accuracy (data not shown) suggesting that including MLRAs as a predictor in a global model is less accurate than modeling by region.

**D. Deliverables**

| Deliverable | Contribution to Soil Survey mission |
|---|---|
| 1. Documented R code demonstrating a prototype soil interpretations engine that can integrate digital soil property maps. | Scientific assistance to NRCS users and Technical soil services |
| 2. Documented R-code demonstrating a computationally efficient method to mosaicking regionally specific soil property maps into nation-wide seamless soil property maps. | Contributing to Soil and Ecological Site inventory |

**E. Literature Review**

An important application of soil survey is to provide stakeholders, including conservation planners, with information on the limitations or suitability (i.e., interpretations) of soil for regional land use and management planning, erosion prediction, timber and energy management, urban planning, crop yield estimation, and the ability of a soil to perform certain ecosystem services (Dobos et al., 2017). However; the literature is relatively sparse regarding how DSM products can be used as input data for such assessment. Indeed, only recently has DSM progressed to a point that digital soil property maps are being used to understand soil condition for agriculture (Kidd et al., 2015; Okonkwo et al., 2018), soil security (Yang et al., 2018), and salinity loading into surface waters (Nauman et al., *In Press*) although the need to do so was recognized more than a decade ago (Carré et al., 2007).

There is also surprisingly little in the literature about using a regional modeling approach for digital soil mapping, although the original intent was to use a regional approach (McBratney et al., 2003). Hewitt et al., (2010) recognized the need for a regional approach and both Hewitt et al., (2010) and Schmidt et al., (2010) demonstrated a method to create homogenous physiographic-modeling regions, but neither implemented a regionally-specific DSM workflow. Although not specifically investigating regional modeling, Mulder et al., (2016) found that national models for France were more accurate for predicting soil organic carbon than were global models because the national model used more soil observations.

Although there is little in the DSM literature about regional modeling, USDA-NRCS soil survey has always used a regional approach, although regions have often been defined by administrative units (e.g., counties, national parks). This has resulted in some sharp boundary artifacts along region (i.e., county) boundaries and the need to harmonize mapping between regions. This same challenge of boundary artifacts is likely to be encountered when using a regional DSM approach. However, the recent conterminous DSM product produced by Chaney et al., (2016) successfully dealt with mosaicking artifacts by training fully overlapping predictive models and then only rendering predictions for a smaller central target within those spatial domains. Overlapping models were trained and rendered in an optimized parallel workflow that also enabled fast computing.

## F. Research objective(s)

| | |
|---|---|
| Objective I | Develop prototype interpretations engine for using digital soil property maps to generate soil interpretations |
| Objective II | Develop methodology to join regional digital soil property maps into national seamless digital soil property maps |

## G. Hypothesis to be tested

While the proposed research is not strictly hypothesis driven, but rather a demonstration of a methodological approach, it is expected that this research will result in the development of workflow for producing nation-wide seamless digital soil maps from regional models as well as a prototype interpretations generator which demonstrates how digital soil property maps can generate interpretations.

## H. Research methods and procedures

*Description of Project Area*

The proposed project will be conducted using data from the Upper Colorado River Basin (UCRB) above Lake Mead, a 432,000 km$^2$ area spanning portions of eight different MLRAs: Great Salt Lake Area (28A), Mojave Desert (30), Cool Central Desertic Basins and Plateaus (34A), Warm

Central Desertic Basins and Plateaus (34B), Colorado Plateau (35), Southwestern Plateaus, Mesas, and Foothills (36), Central Rocky Mountains (43B), Wasatch and Uinta Mountains (47). This region is characterized by a large and diverse set of mountain ranges as well as large areas of dissected sedimentary lithologic landscapes, with soils derived from sandstones, silt/mudstones, limestones, and shales (Nauman and Duniway, 2019). Landforms consist of large mountain systems, complex canyon systems, bottoms and flats receiving run-on moisture, large bedrock controlled plateau uplands, slopes and badlands derived from shales and other saline geologic parent materials all with varying degrees of soil depths and development (Nauman et al., *In Press*; Nauman and Duniway, 2019). This project area, although not perfectly aligned with MLRA boundaries is proposed because the necessary data for this area already exists, thus building on previous work.

*Data Sources*

The spatial data framework for the proposed prototype area has already been established, including a suite of climatic, topographic, and geographic data layers at 30m resolution (Nauman and Duniway, 2019) and an extensive database of soil properties from multiple data sources including the NRCS National Soil Information System (NASIS), NRCS Kellogg Soil Survey Laboratory 2017 snapshot, Forest Service NRM Inventory and Mapping database, and multiple independent research projects. Additionally, global models of fifteen soil properties (pH, texture fractions, coarse fragment content, electrical conductivity, gypsum, $CaCO_3$, sodium adsorption ratio, available water capacity, bulk density, erodibility, and organic matter at 7 depths as well as depth to restrictive layer, and surface rock size and cover) have also already been developed and will be used to test regional versus global approaches Nauman et al., (*In Press);* accessible at: https://github.com/usgs/Predictive-Soil-Mapping/tree/master/SoilSurvReconstrProperties.
Additionally, Dr. Brungard has developed R-scripts for regional modeling. Thus, the majority of data preparation work has been completed.

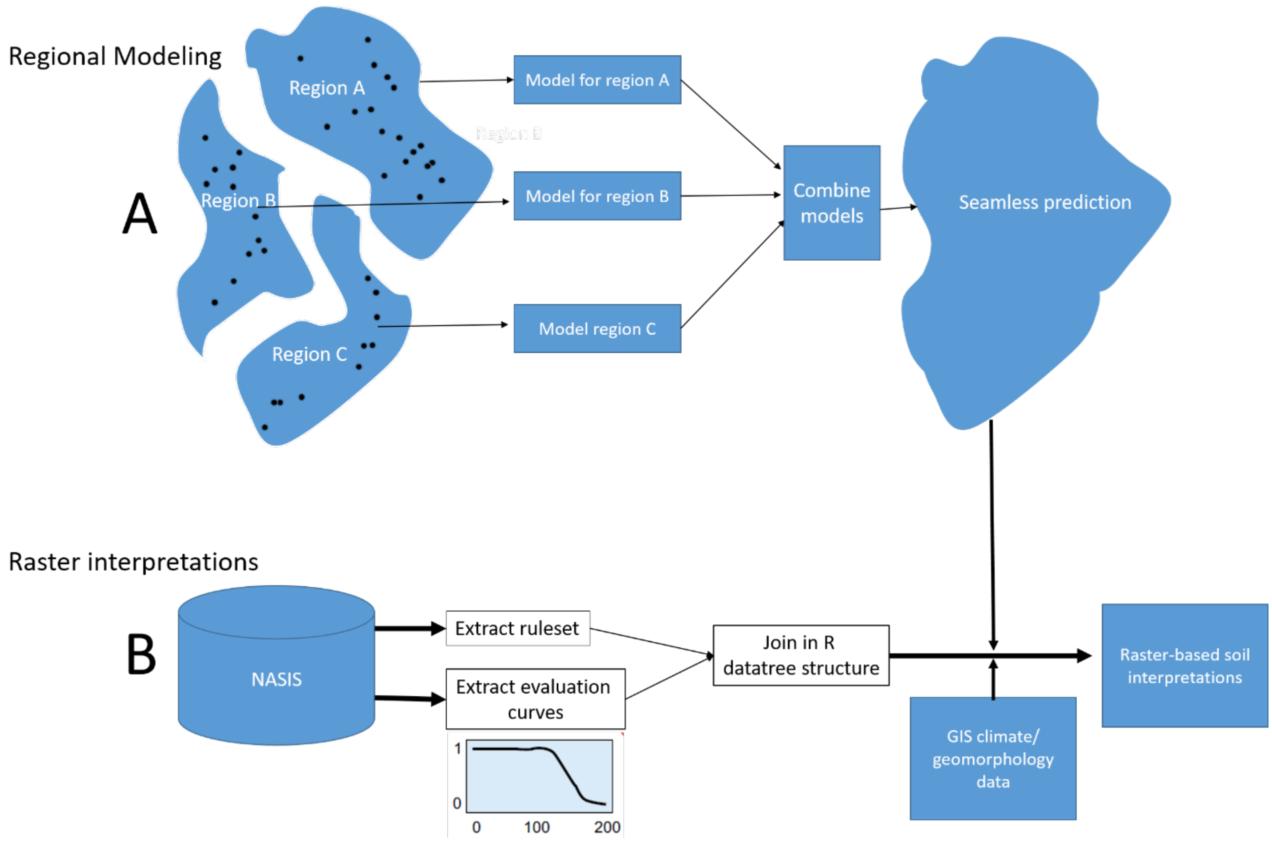


Figure 2. Diagram of proposed workflow. Regional modeling (A) for a soil property is accomplished by dividing the point soil property dataset into individual regions (each region = MLRA), modeling the soil property by region, and combining the models to produce a seamless soil property map. Raster interpretations (B) are developed by extracting the rulesets and evaluation curves from NASIS, joining these in an R software datatree structure, then using the seamless soil property maps (and related uncertainty) from A along with GIS climate/geomorphology data to derive raster-based soil interpretations.

*Data Analysis*

*DSM interpretations*

Analytical workflow will follow Figure 2.

1. Build an interpretations generator outside of NASIS.
    a. Extract existing interpretation rules, evaluations, and properties from NASIS. This workflow builds exclusively on, and in cooperation with, the work done by Dr. Dylan Beaudette. The initial code to be used for development is available at the national cooperative soil survey github version control and software development platform (https://github.com/ncss-tech/interpretation-engine).
        i. Load rules and evaluations from NASIS into R via open database connections
        ii. Load rules into a data.tree object
        iii. Load evaluation functions into each terminal node of data.tree object
        iv. Combine fuzzy values via operators and hedges to generate a final fuzzy rating.
2. Use DSM of soil properties (and uncertainties) as input to the evaluation functions to generate fuzzy values for each interpretation.

3. Climate and site data necessary for interpretations (e.g., mean annual temperature or slope) will be derived from corresponding GIS layers such as the PRISIM climate summaries (www.prism.oregonstate.edu) or calculated from a digital elevation model.
4. We anticipate demonstrating one soil interpretation from each of the following areas: building site development (e.g., corrosion of concrete), construction materials (e.g., sand source), and military operations (e.g., helicopter landing zone) as these are expected to cover a range of soil properties for developing interpretations and demonstrate the prototype interpretations engine.

*Regional modeling*
Analytical workflow will follow Figure 2.

1. Subset soil properties dataset by MLRA. MLRA's with only a small areal extent on the boundary of the project area boundary may be removed or combined with other MLRA's.
2. Model key soil properties by MLRA. These are the regional models. Modeling is anticipated to use quantile random forests to compare with existing global models and to provide estimates of uncertainty (Vaysse and Lagacherie, 2017).
   *a.* Key soil properties will include the 15 soil properties that already exist as global models (pH, texture fractions, coarse fragment content, electrical conductivity, gypsum, $CaCO_3$, sodium adsorption ratio, available water capacity, bulk density, erodibility, and organic matter at 7 depths as well as depth to restrictive layer, and surface rock size and cover) as well as other soil properties identified as necessary for soil interpretations (see preceding section).
3. Produce seamless digital soil maps from the regional models.
   a. There are three potential options for generating seamless soil property maps from the regional models. The solution may involve a combination several of these options and each will be tested.
      i. Train overlapping predictive models of adjacent MLRAs and then only render predictions for a single MLRA. Repeat for each MLRA. Evaluate this approach for opportunities to parallelize processing steps.
      ii. Combine regional models into a single supra-regional model then use this supra-regional model to make predictions.
      iii. Compute a membership/similarity vector to each region for each raster cell. Use this membership/similarity vector to choose which regional models make a prediction at each raster cell. Also use this membership/similarity vector to weight prediction uncertainty.
   b. We will compare results of these efforts to the global outputs created by Nauman et al. (*In Review).* https://github.com/usgs/Predictive-Soil-Mapping/tree/master/SoilSurvReconstrProperties).
      i. Comparison will use accuracy metrics (root mean square error, etc.) using independent validation data from individual research projects (data provided by Dr. Brungard) and cross validation where validation data did not measure all soil properties (e.g., bulk density). The spatial patterns of each prediction will also be reviewed to see if they match expected soil-landscape relationships.
      ii.

*Involvement of NRCS personnel*

Data collection, processing, and analysis will all be conducted in coordination with NRCS staff to identify potential interpretations most salient to NRCS end-users. All tools and code will be housed in the NCSS github repository with detailed documentation, allowing for transparency, portability and extensibility. Dr. Dylan Beaudette (Digital Soil Mapping Specialist, Soil Survey Region II) has agreed to work with the PI's in implementing the soil interpretations. The involvement of Dr. Beaudette is critical to the success of this work because the proposed objectives are ambitious and Dr. Beaudette has extensive experience in both of the proposed objectives. All involved NRCS staff will be co-authors on publications that result from this effort.

**I. Project timetable**

| Task | Estimated Completion Date |
|---|---|
| Develop prototype interpretations engine using digital soil property maps as input | February 2020 |
| Develop methodology for efficient computation of regionally-specific seamless digital soil property maps | June 2020 |
| Manuscript writing and webinar delivery | July 2020 |

**J. Information transfer plan**

We acknowledge that the proposed work overlaps with NRCS goals and we anticipate that this project will provide a foundation for deriving raster-based interpretations and seamless digital soil property maps. We anticipate the following to facilitate information transfer from this project:

1. A webinar presentation to NRCS personnel on the results of the regional modeling approach and the development of raster-based interpretations.
2. Detailed webpage and R-code supplementing the webinar documenting analytical methodologies as well as outputs. This will be hosted on the NCSS-tech github website which will make the code available to anyone.
3. Presentation of the core findings at a professional conference
4. Peer-reviewed papers(s) in an open-access journal.

**K. Education**

Throughout the project we will include interested field soil scientists from the targeted MLRAs. We anticipate their involvement will include evaluating model outcomes and assessing interpretations based on their local knowledge and expertise. The project will also support the efforts of early-career scientists on this proposal, including C. Brungard and T. Nauman.

**L. References**

Ballabio, C., P. Panagos, and L. Monatanarella. 2016. Mapping topsoil physical properties at European scale using the LUCAS database. Geoderma 261: 110–123.

Carré, F., A.B. McBratney, T. Mayr, and L. Montanarella. 2007. Digital soil assessments: Beyond

DSM. Geoderma 142(1–2): 69–79.
Chaney, N.W., E.F. Wood, A.B. McBratney, J.W. Hempel, T.W. Nauman, C.W. Brungard, and N.P. Odgers. 2016. POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. Geoderma 274: 54–67.
Dobos, R., C. Seybold, J. Chiaretti, S. Southard, and M. Levin. 2017. Ch. 8. Interpretations: The Impact of Soil Properties on Land Use. *In* Ditzler, C., Scheffe, K., Monger, H.C. (eds.), Soil Survey Manual. USDA Handbook 18. Government Printing Office, Washington, D.C.
Hengl, T., J. Mendes De Jesus, G.B.M. Heuvelink, M.R. Gonzalez, M. Kilibarda, A. Blagotí, W. Shangguan, M.N. Wright, X. Geng, B. Bauer-Marschallinger, M.A. Guevara, R. Vargas, R.A. Macmillan, N.H. Batjes, J.G.B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, and B. Kempen. 2017. SoilGrids250m: Global Gridded Soil Information Based on Machine Learning. PLoS One 12(2).
Hewitt, A.E., J.R.F. Barringer, G.J. Forrester, and S.J. McNeill. 2010. Soilscapes Basis for Digital Soil Mapping in New Zealand. p. 297–307. *In* Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (eds.), Digital Soil Mapping Bridging Research, Environmental Application, and Operation. Springer Netherlands.
Kidd, D., M. Webb, B. Malone, B. Minasny, and A. McBratney. 2015. Digital soil assessment of agricultural suitability, versatility and capital in Tasmania, Australia. Geoderma Reg. 6: 7–21.
McBratney, A.., M.. Mendonça Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma 117(1–2): 3–52.
Mulder, V.L., M. Lacoste, A.C. Richer-de-Forges, M.P. Martin, and D. Arrouays. 2016. National versus global modelling the 3D distribution of soil organic carbon in mainland France. Geoderma 263: 16–34.
Nauman, T.W., and M.C. Duniway. 2019. Relative prediction intervals reveal larger uncertainty in 3D approaches to predictive digital soil mapping of soil properties with legacy data. Geoderma 347(October 2018): 170–184.
Nauman, T.W., C.P. Ely, M.P. Miller, and M.C. Duniway. Salinity yield modeling of the Upper Colorado River Basin using 30-meter resolution soil maps and random forests. Water Resour. Res.
Okonkwo, E., R. Corstanje, and G. Kirk. 2018. Digital soil assessment for quantifying soil constraints to crop production: a case study for rice in Punjab, India. Soil Use Manag.
Ramcharan, A., T. Hengl, T. Nauman, C. Brungard, S. Waltman, S. Wills, and J. Thompson. 2018. Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. Soil Sci. Soc. Am. J. 82(1): 0.
Schmidt, K., T. Behrens, K. Friedrich, and T. Scholten. 2010. A method to generate soilscapes from soil maps. J. Plant Nutr. Soil Sci. 173(2): 163–172.
Vaysse, K., and P. Lagacherie. 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291: 55–64.
Viscarra-Rossel, R.A., C. Chen, M.J. Grundy, R. Searle, D. Clifford, and P.H. Campbell. 2015. The Australian three-dimensional soil grid: Australia 's contribution to the GlobalSoilMap project. Soil Res. 53(September): 845–864.
Yang, R.-M., B. Minasny, Y. Ma, D. Field, A. Mcbratney, and C. Wu. 2018. A preliminary soil security assessment of agricultural land in middle-eastern China.