Modeling Soil Organic Carbon in the Easter Allegheny Plateau and Mountains from Legacy Data and Spatial Covariates

K.M. Yoast[1,2] and J.A. Thompson[2]

[1]USDA Natural Resources Conservation Service and [2]Division of Plant & Soil Sciences, West Virginia University

INTRODUCTION

The deeply dissected topography and diverse climate of the Eastern Allegheny Plateau and Mountains (Major Land Resource Area (MLRA) 127) (Fig. 1) create challenges for dynamic ecological and pedogenic modeling. Soil organic carbon (SOC), one of the most dynamic soil properties, has been previously modeled using multiple methods and multiple data sources, with estimates for the upper 1 meter ranging from 2.2 to 32.0 kg m^{-2} (Table 1). Estimates derived from legacy soil survey databases have been shown to underestimate true carbon stock due to unpopulated organic horizons and inconsistencies within the databases (Jenkins, 2001; Minasny 2013). Between 1960 and 2009, the Kellogg Soil Survey Lab (KSSL) sampled and characterized 243 pedons within MLRA 127 based on soil survey needs. Each pedon has a site description and associated chemical and physical lab analyses to support its taxonomic classification. Data mining revealed that 12.6% of these 243 pedons lack SOC data for one or more horizons and 49.7% lack bulk density values. Random forest (RF) and median and mean techniques were assessed, validated, and then used to populate missing bulk density (BD) and SOC data. Geographically weighted regression kriging (GWRK) with KSSL pedons and environmental covariates will be used to model SOC stock across MLRA 127. The resulting SOC estimates will be independently validated with Rapid Carbon Assessment (RaCA) samples and uncertainty will be assessed. The methodology used in this study will serve as the foundation for estimating SOC stock across other MLRA throughout the United States. Improving SOC stock estimates across MLRA 127 will enhance the understanding of dynamic soil properties and will provide guidance for better land management practices to benefit biological communities.

Table 1. SOC estimates for MLRA 127 (adapted from Jenkins, 2001).

Source	SOC (kg m^{-2})
Kern (1994)	
Ecosytem Approach	28.1 - 32.0
Soil Taxonomy	13.6 - 15.0
Eswaran et al. (1993)	
World Soils Map	10.6 - 12.1
Bliss et al. (1995)	
Soil Interpretation Database	2.2
Jenkins (2001)	
Mesic Series	8.9 - 11.7
Frigid Hardwood Sites	12.9 - 18.8
Frigid Spruce Sites	11.3 - 13
Mishra et al. (2010)	10.3
NRCS Products:	
STATSGO2	2.6
SSURGO	4.4
Rapid Carbon Assessment	13.3 - 15.6

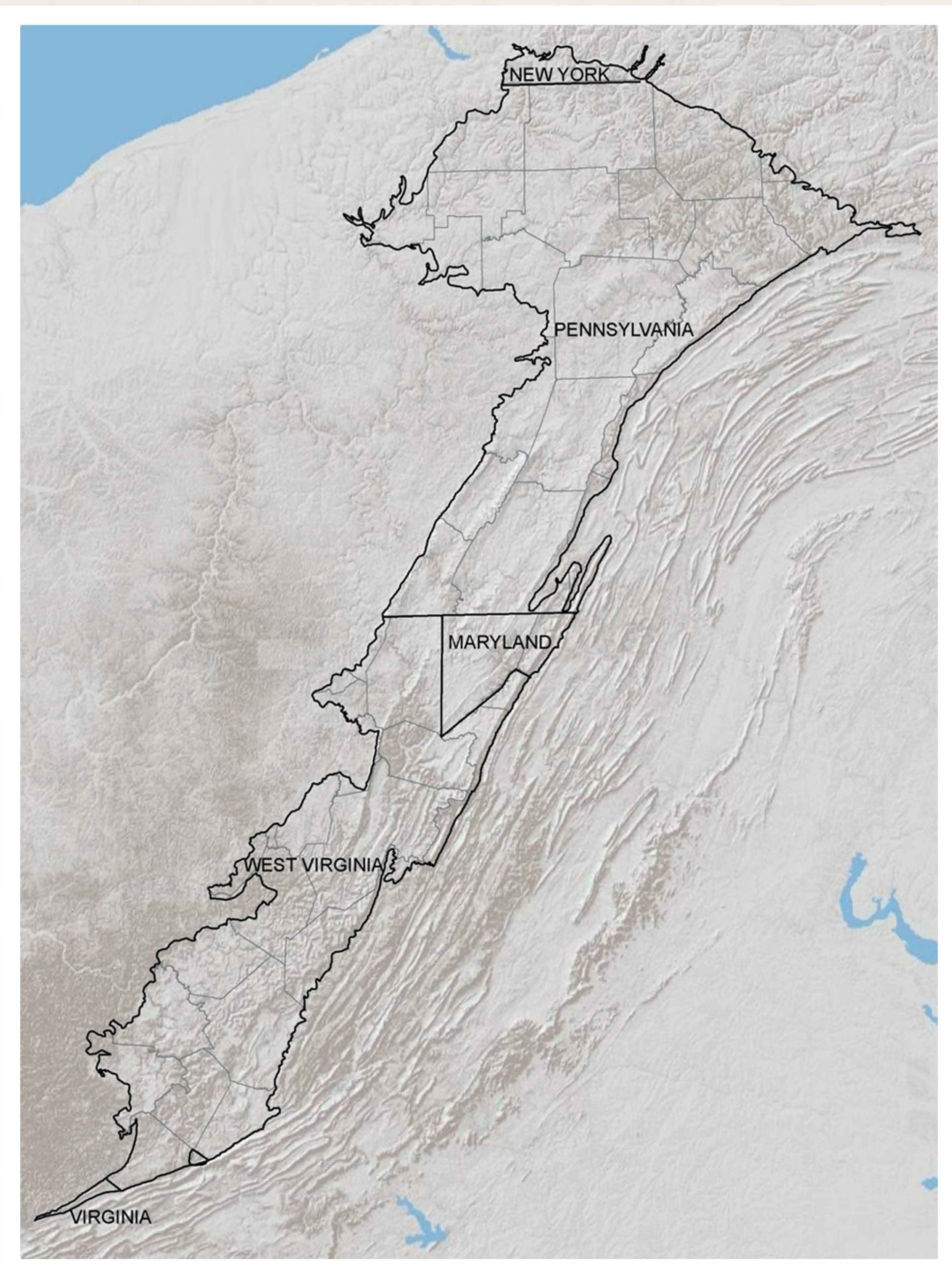

Figure 1. MLRA 127 includes parts of five states: Maryland, New York, Pennsylvania, Virginia, and West Virginia that encompass 50,370 km2. The physiography of this area consists of the Kanawha and Allegheny Mountain Sections of the Appalachian Plateaus Province of the Appalachian Highlands, and is located just west of the Allegheny Front

HYPOTHESES

1. SOC calculated from the SSURGO database underestimates true carbon stock within MLRA 127.
2. Missing soil legacy BD and OC laboratory data can be estimated using nonparametric pedotransfer functions (PTF).
3. SOC can be modeled through GWRK as a function of harmonized legacy soil point data and environmental covariates to more accurately reflect carbon stock in MLRA 127.

METHODS

- Site and pedon data were downloaded and assembled from the KSSL online database within MLRA 127 (NCSS, 2013)
- Pedon taxonomic classification and series correlation were updated to current Soil Taxonomy (USDA, 2010)
- A carbon-equivalent regression factor was used to convert SOC by wet combustion to total carbon by dry combustion (Wills et al., 2013)
- Normality was assessed by the D'Agostino statistics (USEPA, 2002) and outliers were identified through box and whisker plots (Sequeira et al., 2013)
- Known KSSL one-third bar BD and SOC data were split into 70-30 training and testing sets for analysis (Sequeira et al., 2014)
- Three methods were used to predict missing SOC and BD data: mean and median by horizon designation and texture (Buell and Markewich, 2004) and RF (Sequeira et al., 2013). These methods were performed with and without outliers.

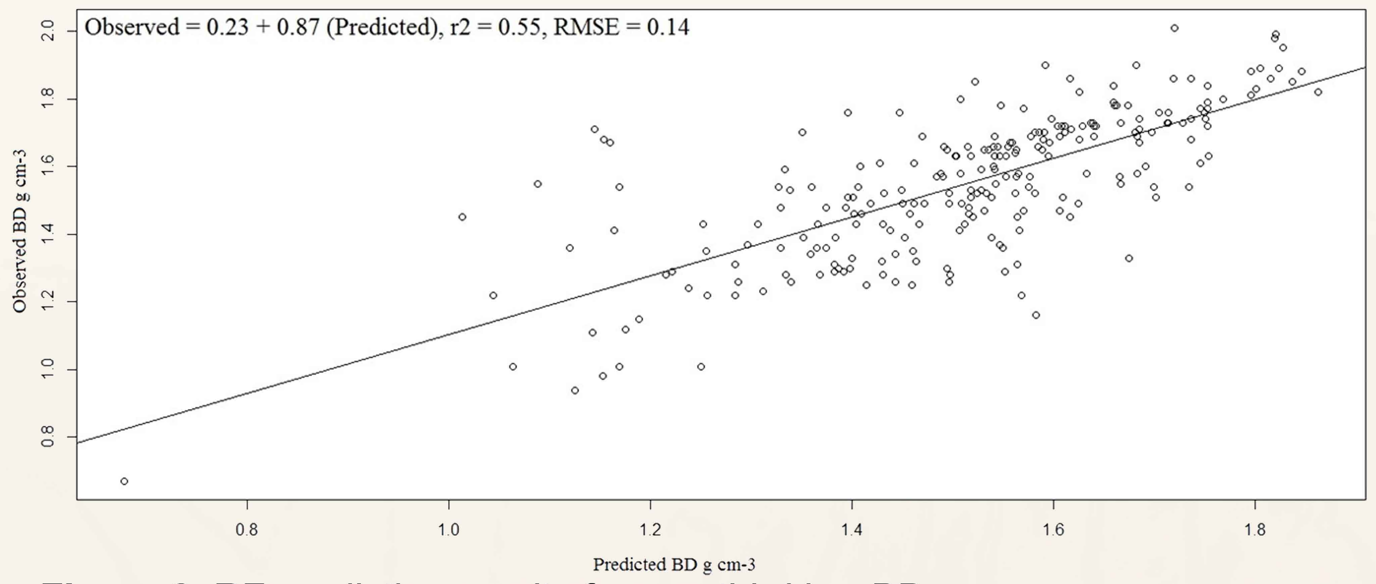

Figure 2. RF prediction results for one-third bar BD.

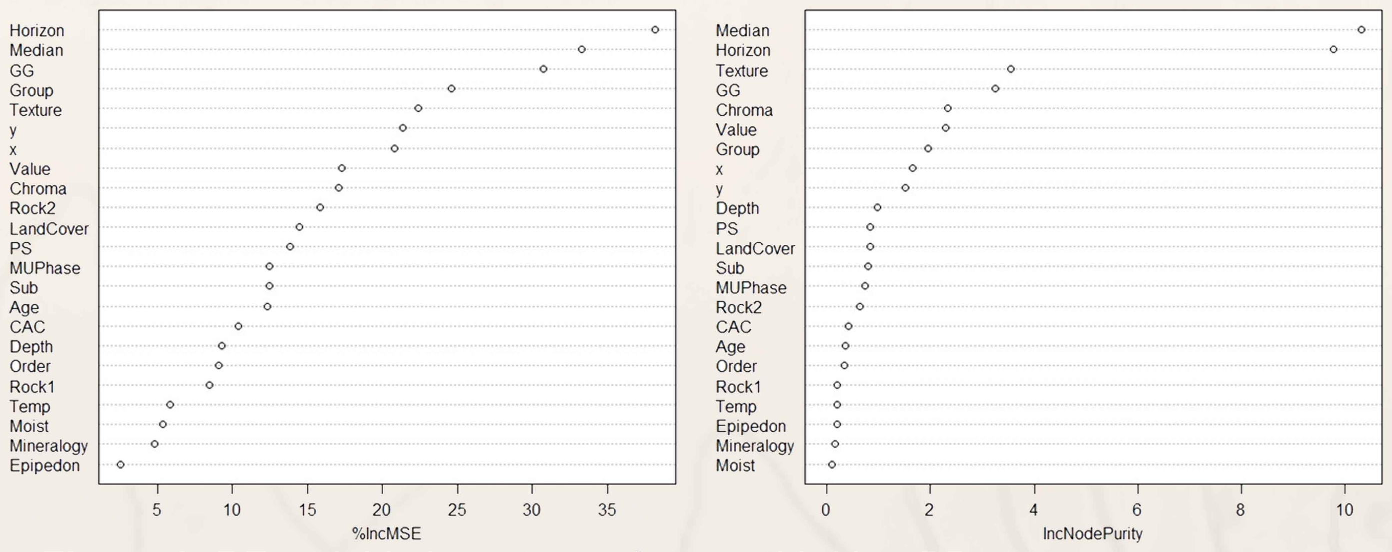

Figure 3. RF variable importance for one-third bar BD.

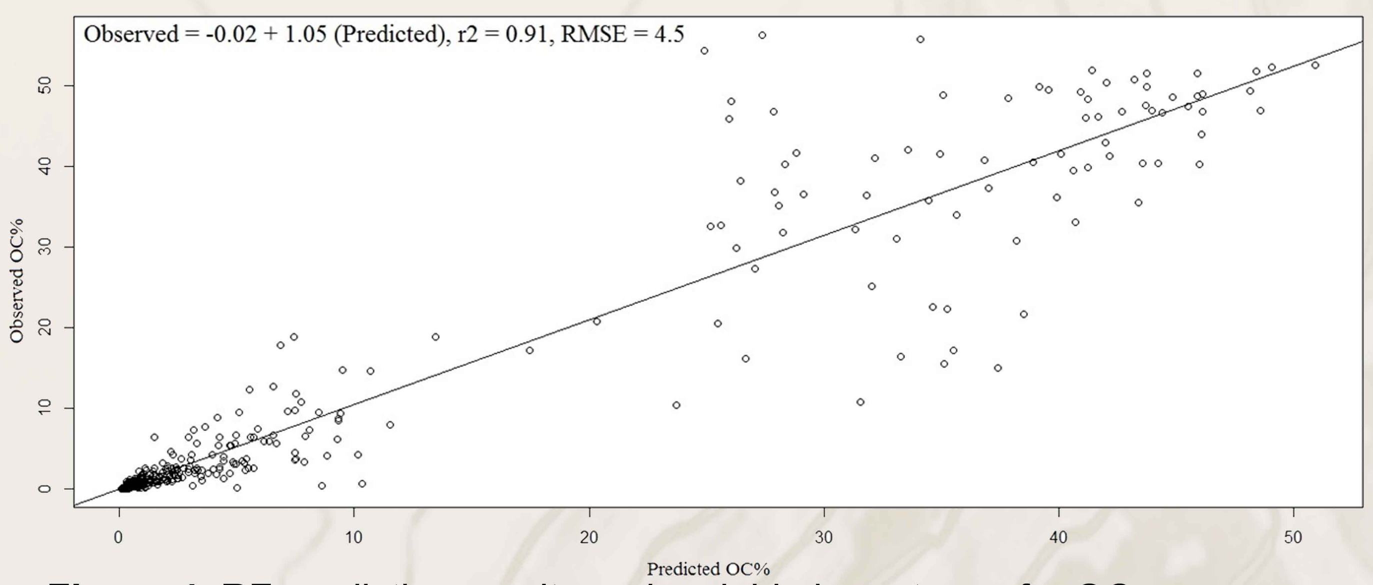

Figure 4. RF prediction results and variable importance for OC.

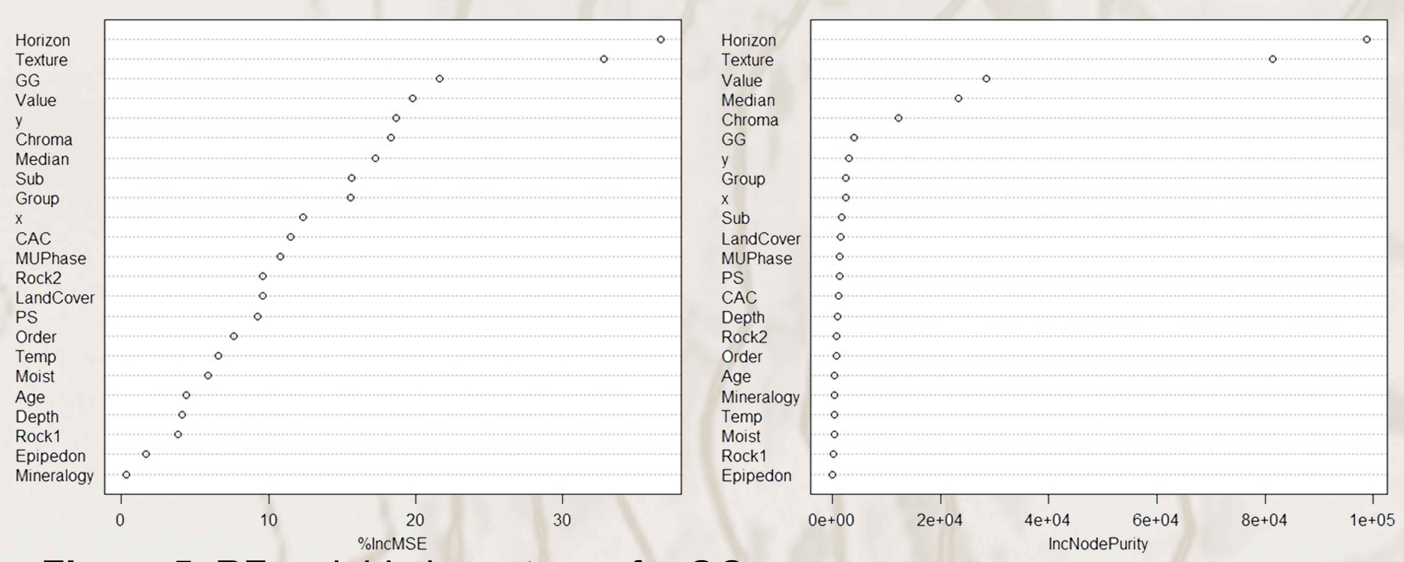

Figure 5. RF variable importance for OC.

RESULTS AND DISCUSSION

- 53% of the KSSL sites within MLRA 127 were not correlated to any soil series. When updated to current Soil Taxonomy, 68% of the pedons were correlated as series taxadjuncts. 31% of the pedons correlated in the KSSL database were correlated to the incorrect soil series.
- Both SOC and BD with and without outliers were not normally distributed even with logarithmic and square root transformations.
- RF predicted missing legacy data with the least amount of error when compared to mean and median of horizon designations and texture classes (R^2 = 0.55, RMSE = 0.14 for BD and R^2 = 0.91, RMSE = 4.50 for SOC) (Fig. 2-5).
- Outliers did not influence the predictive power of RF.
- Given the stability of prediction and simplicity of training and validation, RF is the recommended method for populating missing soil legacy data.

SUMMARY AND CONCLUSIONS

- Data preparation through data download and assemblage, updating series correlation, methodological correction factors, normality testing, outlier assessment, and estimation and validation of missing data is essential for the extrapolation (Sulaeman et al., 2013).
- RF performs similarly to other machine learning techniques, but is more interpretable, faster to train, has fewer parameters, can handle missing data, readily handles larger numbers of predictors, and doesn't require cross-validation (Nemes et al., 2006; Hudak et al., 2008; Myers et al. 2009; Strobl et al., 2009; Ghehi et al., 2011).
- GWRK models will be built from the updated KSSL pedons and scorpan covariates (Fig. 6) (Minasny et al., 2013).
- These models will be validated from an independent dataset of RaCA pedons and available data from the U.S. Forest Service.
- Once these models are validated and uncertainty is assessed, they will be compared with the low, representative, and high SOC ranges populated for each SSURGO map unit in MLRA 127. At that point, there will be a better understanding of true SOC stocks across MLRA 127.

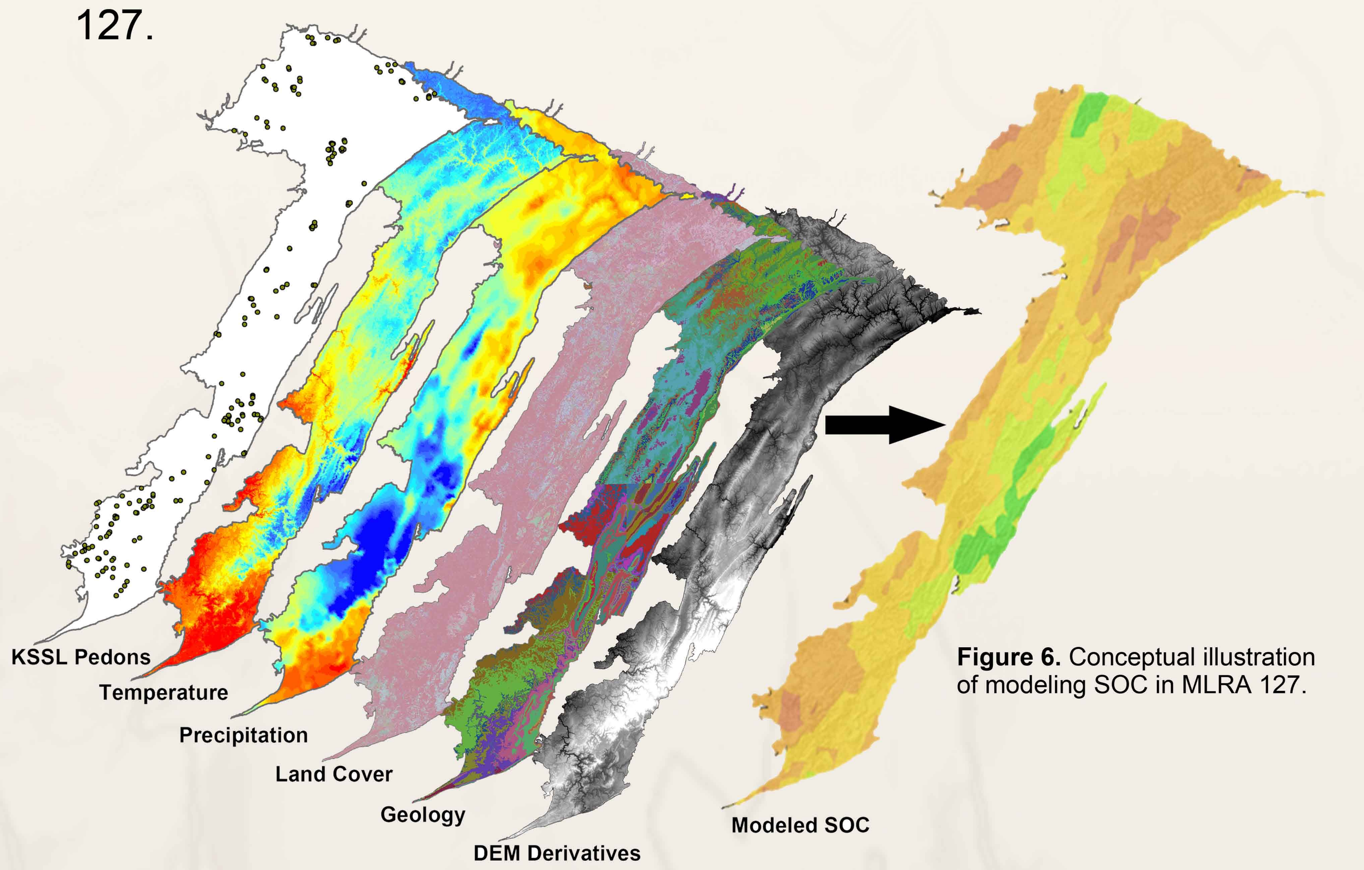

Figure 6. Conceptual illustration of modeling SOC in MLRA 127.

REFERENCES CITED

Bliss, N.B., S.W. Waltman, and G.W. Peterson. 1995. Preparing a soil carbon inventory for the United States using geographic information systems. In: Lal, R, J. Kimble, E. Levine, and B.A. Stewart (eds). Soils and Global Change. CRC Press Inc, Boca Raton, FL. 275-295.

Buell, G.R. and H.W. Markewich. 2004. Data compilation, synthesis, and calculations used for organic-carbon storage and inventory estimates for mineral soils of the Mississippi River Basin. U.S. Geological Survey Professional Paper 1686-A. <http://pubs.er.usgs.gov/publication/pp1686A>.

Ghehi, N.G., A. Nemes, A. Verdoodt, E. Van Ranst, W.M. Cornelis, and P. Boeckx. 2011. Nonparametric techniques for predicting soil bulk density of tropical rainforest topsoils in Rwanda. Soil Science Society of America Journal. 76 (4): 1172-1183.

Hudak, A.T., N.L. Crookston, J.S. Evans, D.E. Hall and M.J. Falkowski. (2008) Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sensing of Environment. 112(5): 2232-2245.http://www.treesearch.fs.fed.us/pubs/30691

Jenkins, A. 2001. Organic Carbon and Fertility of Forest Soils on the Allegheny Plateau of West Virginia. M.S. thesis, West Virginia University, Morgantown, WV. http://etd.wvu.edu/templates/showETD.cfm?recnum=2486

Kern, J.S. 1994. Spatial patterns of soil organic carbon distribution in the contiguous United States. Soil Science Society of America Journal. 58: 439-455.

Minasny, B., A.B. McBratney, B.P. Malone, and I. Wheeler. 2013. Digital mapping of soil carbon. Advances in Agronomy. 118: 1-47.

Mishra, U., R. Lal, D. Liu, and M. Van Meirvenne. 2010. Predicting the spatial variation of the soil organic carbon pool at a regional scale. Soil & Water Management & Conservation. Soil Science Society of American Journal. 74 (3): 906-914.

Myers D.B., S. Grunwald, G.M. Vasques and W.G. Harris. 2009. Pedotransfer functions for carbon methods and bulk density estimation in Florida soils. ASA-CSSA-SSSA Meeting, Pittsburgh, PA, Nov. 1-4, 2009.

National Cooperative Soil Survey (NCSS) National Cooperative Soil Characterization Database. 2013. <http://ncsslabdatamart.sc.egov.usda.gov>.

Nemes, A. W.J. Rawls, and Y.A. Pachepsky. 2006. Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. Soil Science Society of America Journal. 70 (2): 327-336.

Sequeira, C.H., S.A. Wills, C.A. Seybold, and L.T. West. 2013. Predicting soil bulk density from incomplete datasets. Geoderma. 213: 64-73.

Strobl, C., J. Malley, and G. Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods. 14, 323–48.

Sulaeman, Y., B. Minasny, A.B. McBratney, M. Sarwani, and A. Sutandi. 2013. Harmonizing legacy soil data for digital soil mapping in Indonesia. Geoderma, 192: 77-85.

USDA Natural Resources Conservation Service . 2006. Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin. USDA Handbook 296, 669p.

USDA Natural Resources Conservation Service. 2010. Keys to Soil Taxonomy. 11th edition. Soil Survey Staff, 338p.

U.S. Environmental Protection Agency (EPA) – Office of Emergency and Remedial Response. 2002. Guidance for comparing background and chemical concentrations in soil for CERCLA sites. EPA 540-R-01-003. OSWER 9285.7-41. < http://www.epa.gov/oswer/riskassessment/pdf/background.pdf>.

Wills, S., C. Seybold, J. Chiaretti, C. Sequeira, and L. West. 2013. Quantifying tacit knowledge about soil organic carbon stocks using soil taxa and official soil series descriptions. Soil Science Society of America Journal. 77: 1711-1723.