Predicting Soil Cation Exchange Capacity from Satellite Imagery Using Random Forest Models

Ivo Müller[1], Minan Li[2], Joby M. Czarnecki[1], and Brian K. Smith[2]

[1]Geosystems Research Institute, Mississippi State University, Box 9627, Mississippi State, Mississippi, 39762, USA

[2]Department of Industrial and Systems Engineering, Mississippi State University, Box 9542, Mississippi State, Mississippi, 39762, USA

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.

Crop yield variability is often attributed to spatial variation in soil properties. Remote sensing offers a practical approach to capture soil surface properties over large areas, enabling the development of detailed soil maps. This study aimed to predict cation exchange capacity (CEC), a key indicator of soil quality, in the agricultural fields of the Lower Mississippi Alluvial Valley using digital soil mapping techniques. Approximately 14,000 soil samples were collected from agricultural fields across the study area between fall 2019 and fall 2023, which provided ground reference information. A bare soil composite image was created from Landsat 9 imagery drawn from the months of October to April of each year from 2019 to 2023 using Google Earth Engine. Cation exchange capacity was predicted using a Random Forest model based on the reflectance of bands 2 (blue, 0.45 - 0.51 µm), 3 (green, 0.53 - 0.59 µm), 4 (red, 0.64 - 0.67 µm), 5 (near infrared, 0.85 - 0.88 µm), 6 (short-wave infrared 1, 1.57 - 1.65 µm), and 7 (short-wave infrared 2, 2.11 – 2.29 µm). The model's performance was evaluated using an 80/20 training/testing split. Overall, the predictive model performed successfully, achieving an R-squared value of 0.63 and a root mean squared error of 5.6 meq 100 g^{-1}. The most important spectral bands were short-wave infrared 2 and near infrared. These results suggest that digital soil mapping techniques can effectively map CEC in this region, reducing the need to extensively sample for future studies.

Keywords.
digital soil mapping, bare soil composite imagery, machine learning, management zones.

Introduction

Crop yield variability, often attributed to spatial variation in soil properties, can be addressed by exploring the relationship between remotely sensed properties and agronomic relevance. As a tool for precision agriculture, satellite-based sensors offer a practical approach to capture soil surface properties over large areas (Ben-Dor et al., 2009), enabling the development of detailed soil maps, a process referred to as “digital soil mapping” (DSM). An alternative to traditional methods (Lagacherie & McBratney, 2006), DSM uses readily available spatial data like remotely sensed imagery to develop models that can predict soil properties. Machine learning algorithms have been applied to DSM as early as the 1990s (Lagacherie, 2008). Brown et al. (2006) demonstrated the relative strength of decision trees for estimating soil parameters such as clay, soil carbon, and cation exchange capacity (CEC); decision trees also outperformed geostatistical techniques such as kriging for DSM (Vandana et al., 2024). While DSM requires bare soil (Dobos, 2006), techniques like time series composites from multiple satellite images can overcome this limitation (Demattê et al., 2016). The success of these techniques in other regions (Demattê et al., 2018; Rizzo et al., 2020; Rogge et al., 2018; Safanelli et al., 2020) suggests they may be appropriate for mapping the Lower Mississippi Alluvial Valley with its highly variable soils (Logan, 1916). Moreover, the window between crop cycles in the region offers a multi-month fallow period with bare soil imagery.

Heil et al. (2018) found that soil parameters offered a better basis for delineation of management zones than yearly variation in yield. Cation exchange capacity reflects a soil's capacity to retain exchangeable cations that could be absorbed or released at a defined pH (Huang et al., 2015). High CEC soils exhibit improved fertility, structure, and buffering capacity against chemical changes; the end result of which is that high CEC is a key indicator of soil quality (Charman & Murphy, 2007; Larson & Pierce, 1994; Masto et al., 2008). Precise spatial mapping of CEC allows for site-specific management practices, such as variable-rate application of fertilizers, seeds, and herbicides (Lacoste et al., 2014), and thus serves as a tool for farmers to optimize crop performance and improve resource management. However, conventional methods involving field data collection are recognized for being inefficient, requiring significant time, financial resources, and manual labor (Minasny & McBratney, 2016). Therefore, this study seeks to apply a Random Forest (RF) model to predict CEC across a large portion of Lower Mississippi Alluvial Valley.

Methods and Materials

The study area covers 44,590 km^2 in the Lower Mississippi Alluvial Valley, including portions of northwest Mississippi, southeast Arkansas, and northeast Louisiana. Centuries of seasonal flooding from the Mississippi River and its tributaries have deposited deep alluvial soils, rich in nutrients but exhibiting significant spatial heterogeneity in texture, structure, depth, and drainage due to variations in overflow frequency (Snipes, 2005). Approximately 14,000 soil samples were collected from agricultural fields representing 0-15 cm depth on a ~0.4-ha grid across the study area from 2019 to 2023. Cation exchange capacity was determined for all soil samples based on the Mehlich III extraction of calcium, magnesium, potassium, sodium, and an estimate of exchangeable hydrogen obtained from the buffer pH. A multi-temporal bare soil composite image representing the median reflectance for 2021-2024 was generated utilizing the Google Earth Engine platform (Hird et al., 2017) using atmospherically corrected Landsat 9 images drawn from the fallow season (October to April). Bare areas were isolated by applying masks (as outlined in Table 1). Reflectance bands in the composite image included blue (band 2, 450-510 nm), green (band 3, 530-590 nm), red (band 4, 640-670 nm), near infrared (band 5, 850-880 nm), and short-wave infrared (band 6, 1570-1650 nm; band 7, 2110-2290 nm) (Demattê et al., 2020; Demattê et al., 2018; Sorenson et al., 2021). Finally, a RF model was developed utilizing the 'ranger' package (Wright & Ziegler, 2015) in program R to predict soil CEC across the study area based on reflectance values from a bare soil composite image and soil test CEC. To ensure robust evaluation and minimize farm-specific biases, stratified K-fold cross-validation (K=5) was employed. Data were first stratified by farm, and within each stratum, an 80/20 split was used for

training and testing. For each combination of train and test data, a model was generated, and the average result was extracted. The model performance was evaluated using several metrics including: the square root of the mean squared error (RMSE), the coefficient of determination (R^2), the ratio of performance to deviation [RPD, (Williams, 1987)], and bias.

Table 1. Vegetation indices and their associated thresholds used for generating the bare soil composite image

Vegetation Index	Equation	Threshold
Normalized difference vegetation index (NDVI)	$NIR - Red/NIR + Red$	0.3
Normalized difference water index (NDWI)	$Green - NIR/Green + NIR$	0.5
Normalized burn ratio 2 (NDR_2)	$SWIR1 - SWIR2/SWIR1 + SWIR2$	0.1
Normalized different index 7 (NDI7)	$NIR - SWIR2/NIR + SWIR2$	0

Abbreviations: NIR, near infrared; SWIR, short-wave infrared. Reflectance bands: green (530-590 nm); red (640-670 nm); NIR (850-880 nm), SWIR1 (1570-1650 nm); SWIR2 (2110-2290 nm).

Results and Discussion

Cation exchange capacity values ranged from 3.7 to 50.4 milliequivalents per 100 grams of soil (meq $100g^{-1}$), with a mean of 22.5 meq $100g^{-1}$ and a median of 22.2 meq $100g^{-1}$. The standard deviation for CEC was 9.25 meq $100g^{-1}$. This suggests that the sampling captured significant variability within the study area, providing a good representation of the region's soil characteristics. The CEC model exhibited an R^2 of 0.64, a RMSE of 5.54 meq $100g^{-1}$, and a RPD of 1.67 (Fig. 1). The most important spectral bands were short-wave infrared 2 and near infrared. Since clay share absorption features within the short-wave infrared band (Rossel & Behrens, 2010), the model captures the CEC variability indirectly through its relationship as observed in other studies (Sorenson et al., 2022). This study achieved performance that was comparable or better than similar previous research (Gallo et al., 2018; Rizzo et al., 2020; Safanelli et al., 2020), but the need for improvement is clear. Results from other studies support improvements to model performance through the addition of covariates such as topography, climate, and soil texture (Bouslihim et al., 2024; Cambule et al., 2013; McBratney et al., 2003; Xia et al., 2022). However, there are some limitations inherent in the study area to meaningful inclusion of some covariates (e.g., the widespread prevalence of laser land leveling, which effectively removes the influence of topography), and therefore careful consideration of potential covariates is warranted.

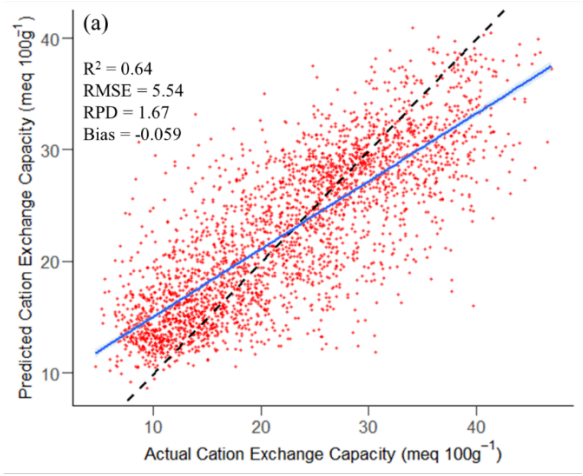

Figure 1. The relationship between predicted and observed cation exchange capacity (meq $100g^{-1}$) in the validation dataset. The dashed line represents the 1:1 line for perfect agreement, while the solid line represents the linear regression line.

Summary

This study utilized DSM techniques with RF models to predict CEC in the Lower Mississippi Alluvial Valley. The model achieved fair accuracy, comparable to or exceeding the results of previous remote sensing applications for similar purposes. While this study demonstrates the effectiveness of DSM, the model could have benefitted from additional data, and future efforts should seek to include such parameters as soil moisture, topography, and climate (where such factors exhibit sufficient variability or where utility is not limited by external factors).

Acknowledgments

This research was funded by USDA NIFA CARE award #2022-68008-36356.

References

Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Taylor, G., Hill, J., Whiting, M., & Sommer, S. (2009). Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment*, *113*, S38–S55.

Bouslihim, Y., John, K., Miftah, A., Azmi, R., Aboutayeb, R., Bouasria, A., Razouk, R., & Hssaini, L. (2024). The effect of covariates on Soil Organic Matter and pH variability: A digital soil mapping approach using random forest model. *Annals of GIS*, 1–18.

Brown, D. J., Shepherd, K. D., Walsh, M. G., Mays, M. D., & Reinsch, T. G. (2006). Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma*, *132*(3–4), 273–290.

Cambule, A., Rossiter, D., & Stoorvogel, J. (2013). A methodology for digital soil mapping in poorly-accessible areas. *Geoderma*, *192*, 341–353.

Charman, P. E., & Murphy, B. W. (2007). Soils: Their properties and management. *Oxford University Press 2nd Ed.*

Demattê, J. A. M., Alves, M. R., Terra, F. da S., Bosquilia, R. W. D., Fongaro, C. T., & Barros, P. P. da S. (2016). Is it possible to classify topsoil texture using a sensor located 800 km away from the surface? *Revista Brasileira de Ciência Do Solo*, *40*.

Demattê, J. A. M., Fongaro, C. T., Rizzo, R., & Safanelli, J. L. (2018). Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sensing of Environment*, *212*, 161–175.

Demattê, J. A., Safanelli, J. L., Poppiel, R. R., Rizzo, R., Silvero, N. E. Q., Mendes, W. de S., Bonfatti, B. R., Dotto, A. C., Salazar, D. F. U., & Mello, F. A. de O. (2020). Bare earth's surface spectra as a proxy for soil resource monitoring. *Scientific Reports*, *10*(1), 4461.

Dobos, E. (2006). *Digital soil mapping: As a support to production of functional maps*. Office for Official Publication of the European Communities.

Gallo, B. C., Demattê, J. A., Rizzo, R., Safanelli, J. L., Mendes, W. de S., Lepsch, I. F., Sato, M. V., Romero, D. J., & Lacerda, M. P. (2018). Multi-temporal satellite images on topsoil attribute quantification and the relationship with soil classes and geology. *Remote Sensing*, *10*(10), 1571.

Heil, K., Heinemann, P., & Schmidhalter, U. (2018). Modeling the effects of soil variability, topography, and management on the yield of barley. *Frontiers in Environmental Science*, *6*, 146.

Hird, J. N., DeLancey, E. R., McDermid, G. J., & Kariyeva, J. (2017). Google Earth Engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. *Remote Sensing*, *9*(12), 1315.

Huang, J., Barrett-Lennard, E. G., Kilminster, T., Sinnott, A., & Triantafilis, J. (2015). An error budget for mapping field-scale soil salinity at various depths using different sources of ancillary data. *Soil Science Society of America Journal*, *79*(6), 1717–1728.

Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., & Walter, C. (2014). High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. *Geoderma*, *213*, 296–311.

Lagacherie, P. (2008). Digital soil mapping: A state of the art. *Digital Soil Mapping with Limited Data*, 3–14.

Lagacherie, P., & McBratney, A. (2006). Spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping. *Developments in Soil Science*, *31*, 3–22.

Larson, W. E., & Pierce, F. J. (1994). The dynamics of soil quality as a measure of sustainable management. *Defining Soil Quality for a Sustainable Environment*, *35*, 37–51.

Logan, W. (1916). The soils of Mississippi. Tech. *Bull*, *7*.

Masto, R. E., Chhonkar, P. K., Singh, D., & Patra, A. K. (2008). Alternative soil quality indices for evaluating the effect of intensive cropping, fertilisation and manuring for 31 years in the semi-arid soils of India. *Environmental Monitoring and Assessment*, *136*, 419–435.

McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, *117*(1–2), 3–52.

Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma*, *264*, 301–311.

Rizzo, R., Medeiros, L. G., de Mello, D. C., Marques, K. P., de Souza Mendes, W., Silvero, N. E. Q., Dotto, A. C., Bonfatti, B. R., & Dematte, J. A. (2020). Multi-temporal bare surface image associated with transfer functions to support soil classification and mapping in southeastern Brazil. *Geoderma*, *361*, 114018.

Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., & Heiden, U. (2018). Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). *Remote Sensing of Environment*, *205*, 1–17.

Rossel, R. V., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, *158*(1–2), 46–54.

Safanelli, J. L., Chabrillat, S., Ben-Dor, E., & Demattê, J. A. (2020). Multispectral models from bare soil composites for mapping topsoil properties over Europe. *Remote Sensing*, *12*(9), 1369.

Snipes, C. E. (2005). *Current agricultural practices of the Mississippi Delta*.

Sorenson, P., Shirtliffe, S., & Bedard-Haughn, A. K. (2021). Predictive soil mapping using historic bare soil composite imagery and legacy soil survey data. *Geoderma*, *401*, 115316.

Sorenson, P. T., Kiss, J., Bedard-Haughn, A. K., & Shirtliffe, S. (2022). Multi-Horizon Predictive Soil Mapping of Historical Soil Properties Using Remote Sensing Imagery. *Remote Sensing*, *14*(22), 5803.

Vandana, N., Suresh, G., Mitran, T., & Mahadevappa, S. (2024). Digital Mapping of Soil pH and Electrical Conductivity Using Geostatistics and Machine Learning. *International Journal of Environment and Climate Change*, *14*(2), 273–286.

Williams, P. (1987). Variables affecting near-infrared reflectance spectroscopic analysis. *Near-Infrared Technology in the Agricultural and Food Industries*, 143–167.

Wright, M. N., & Ziegler, A. (2015). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *arXiv Preprint arXiv:1508.04409*.

Xia, Y., McSweeney, K., & Wander, M. M. (2022). Digital Mapping of Agricultural Soil Organic Carbon Using Soil Forming Factors: A Review of Current Efforts at the Regional and National Scales. *Frontiers in Soil Science*, *2*, 890437.