Scientific
Research
Publishing

Machine Learning Mapping of Soil Apparent Electrical Conductivity on a Research Farm in Mississippi

Reginald S. Fletcher ⓘ

Crop Production Systems Research Unit, Agricultural Research Service, United States Department of Agriculture, Stoneville, USA
Email: reginald.fletcher@usda.gov

Abstract

Open-source and free tools are readily available to the public to process data and assist producers in making management decisions related to agricultural landscapes. On-the-go soil sensors are being used as a proxy to develop digital soil maps because of the data they can collect and their ability to cover a large area quickly. Machine learning, a subcomponent of artificial intelligence, makes predictions from data. Intermixing open-source tools, on-the-go sensor technologies, and machine learning may improve Mississippi soil mapping and crop production. This study aimed to evaluate machine learning for mapping apparent soil electrical conductivity (EC_a) collected with an on-the-go sensor system at two sites (*i.e.*, MF2, MF9) on a research farm in Mississippi. Machine learning tools (support vector machine) incorporated in Smart-Map, an open-source application, were used to evaluate the sites and derive the apparent electrical conductivity maps. Autocorrelation of the shallow (EC_{as}) and deep (EC_{ad}) readings was statistically significant at both locations (Moran's I, p 0.001); however, the spatial correlation was greater at MF2. According to the leave-one-out cross-validation results, the best models were developed for EC_{as} versus EC_{ad}. Spatial patterns were observed for the EC_{as} and EC_{ad} readings in both fields. The patterns observed for the EC_{ad} readings were more distinct than the EC_{as} measurements. The research results indicated that machine learning was valuable for deriving apparent electrical conductivity maps in two Mississippi fields. Location and depth played a role in the machine learner's ability to develop maps.

Keywords

Spatial Variability, Machine Learning, Electrical Conductivity, Mapping, Data Mining

1. Introduction

In agricultural fields, spatial variability of soil chemical and physical properties affects crop growth and productivity [1] [2]. Understanding soil spatial patterns enhance the producer's ability to improve crop production. Various tools have been developed for growers to use for sensing the spatial variability of soils, including electromagnetic, electrochemical, mechanical, optical and radiometric, airflow, and acoustic and pneumatic sensors. On-the-go sensors that measure apparent electrical conductivity (electromagnetic) have gained popularity for measuring soil spatial patterns because they can cover large areas quickly. Researchers have used information from on-the-go apparent electrical conductivity (EC_a) sensors as proxies for estimating other soil parameters. Electromagnetic induction and electric resistivity are standard approaches for measuring EC_a [3].

Researchers have tested various computer algorithms to map soil spatial patterns [4]. Standard algorithms include random Forest [5] [6], support vector machine [5] [7], Cubist [5] [8], kriging [4], k-nearest neighbors [4], and artificial neural networks [5]. There has been no one fits all approach for using computer algorithms for digital soil mapping.

Over the years, geostatistical techniques have been popular for developing digital soil maps. It provides a statistically sound model for soil spatial variation, minimizes sampling bias, measures spatial autocorrelation, and provides an error layer. Nevertheless, several disadvantages exist in using geostatistical methods for soil mapping [9] [10]: the overall concept can be challenging to understand and implement for users that do not have a background in geostatistics, large datasets can be computationally intensive to analyze, the residuals must be normally distributed, stationary, and not affected by and change in direction. Geostatistical models can be affected by outliers, spatial data clustering, and data collection errors.

Computer algorithms based on the machine learning concept are classified as a method of data analysis and artificial intelligence. They are becoming a popular alternative to geostatistical methods because machine learning techniques make no assumptions about the data distribution, can process large datasets containing cross-correlated covariates as predictors, and can function with little intervention [10] [11]. Machine learning successes for digital soil mapping purposes include soil organic carbon concentration [8] [12] [13] and associated stocks [13] [14], soil texture [15] [16], pH [17], cation exchange capacity [18], nitrogen [18] [19], phosphorus [19] [20], potassium, calcium, and magnesium [20], bulk density [21], and soil pollutants [22]. Finally, machine learning models also can be challenging to interpret and visualize.

Commercial and open-source tools allow users to process data collected by on-the-go sensor systems. Open-source or freeware technology provides excellent opportunities for users to evaluate data at minimal or no cost for using the software. Open-source technology is growing and is supported by numerous communities worldwide. More research is needed on using open-source tools

and machine learning technologies at the farm level for making management decisions. With its vast, diverse agricultural landscapes, growers in Mississippi may benefit from these technologies. In Mississippi, Fletcher [23] has demonstrated, using unsupervised machine learning in cluster analysis, that EC_a spatial patterns are similar for at least five years. The study focused on keeping the data in its native point form and not deriving interpolated maps. The current study was conducted to build on the previous research initiative of using open-source software and machine-learning tools for digital mapping EC_a.

The objective of this study was to evaluate machine learning as a tool for mapping EC_a of soils located on a research farm in Mississippi, to determine if measurements collected at different depths affected the accuracy of the algorithm, and to determine if measurements collected from a field containing multiple soil types impacted the results. The study focused on using open-source software available to the public (Smart-Map) and using it to derive interpolated maps.

2. Materials and Methods

2.1. Study Site

The experiments were conducted at the United States Department of Agriculture, Agriculture Research Service Farm (−90.872157 Longitude, 33.446486 Latitude, elevation—38 m above sea level), near Stoneville, Mississippi, USA. The average precipitation and temperature were approximately 133 cm and 17.5°C, respectively [24]. Two study sites were evaluated and were referred to as MF2 and MF9. MF2 was 4 ha and contained the following soil mapping units: Commerce very fine sandy loam, 0% to 2% slopes; Newellton silty clay, 0% to 2% slopes, occasionally flooded; Commerce silty clay loam, 0% to 2% slopes; and Tunica clay, 0% to 2% slopes [25]. MF9 was 2.0 ha and consisted of the following soil types: Sharkey clay, 0% to 2% slopes, and Tunica clay, 0% to 2% slopes [25]. Both sites were in a continuous soybean (*Glycine max* L.) and corn (*Zea mays* L.) rotation. The farm manager used standard agricultural practices of the area for irrigation, weed treatment, and fertilization.

2.2. Data Collection

EC_a readings were collected from MF2 and MF9 with the Veris MSP 3 (Veris Technologies, Salina, KS, USA) on-the-go sensor system. The device used six coulters to collect shallow (0 - 30 cm) and deep (0 - 90 cm) EC_a measurements as the tractor pulled it through the fields. Coulters two and five injected an electrical current into the soil; coulters three and four obtained the EC shallow readings; coulters one and six recorded the deep readings [26]. Data output was in millisiemens (mS) per meter. The location of each measurement was recorded in latitude and longitude coordinates (WGS84) with a Garmin global positioning system. It recorded location information when receiving differential global positioning data. A laptop (HP Pavillion TouchSmart Notebook, Windows 8.1) inside the tractor's cab was used to log the readings of each measurement. The da-

ta were collected from MF2 on May 1, 2015, and MF9 on April 28, 2022, before the growing season.

2.3. Data Preprocessing

Each measurement was assigned an identification number, converted from longitude and latitude coordinates to the UTM coordinates (UTM 15N, WGS84) system, and cleaned (*i.e.*, removal of negative values, duplicated x-y coordinates, and outliers). The data cleaning step resulted in 1573 and 741 sampling points for further processing at MF2 and MF9, respectively. For comparison purposes, the mean, median, standard deviation, and coefficient of variation were calculated for each variable. The preprocessing and summary statistics were completed with the QGIS software (3.22.8-Białowieża) [27].

2.4. Data Analysis

The EC_{as} and EC_{ad} readings were processed with Smart-Map [7], an open-source QGIS plugin created to complete digital soil mapping. Its machine learning tools were used to evaluate spatial patterns of the EC_{as} and EC_{ad}. The data were processed based on the protocol described in [7]: 1) loading the data into the plugin, 2) selecting a target variable to interpolate, 3) setting the grid size for the map, 4) choosing the machine learning interpolation method, 5) evaluating the model using cross-validation, and 6) creating the map based on model development. Support vector machine is the machine learning method offered by Smart-Map. The software automatically fits the hyperparameters required by support vector machines. It uses the radial base function kernel because it is non-linear and can be fitted to most data. The grid size used for the map interpolation was 8 m × 8 m.

The software automatically selected the x and y coordinates as covariables. The user has the option to add other covariables if needed. The only covariables used was the default x and y coordinates. Model accuracy was determined by leave-one-out cross-validation with root mean squared error and the coefficient of variation (R^2) being the accuracy measures. Additionally, Moran's I [28] statistic was offered as a means to determine the autocorrelation for a specific variable. The value ranges from −1 to +1. The closer the value is to 1, the more clustered the data values. Values going towards −1 were more dispersed, and values close to zero were random. Moran's I was used to compare the autocorrelation of the EC_{as} and EC_{ad} deep readings. Maps of the final predicted measurements were created to evaluate within-field spatial patterns of EC_{as} and EC_{ad}. The maps were created with the QGIS software.

3. Results and Discussion

Summary statistics of the study sites are summarized in Table 1. For MF2 and MF9, the EC_{as} mean, median, minimum, and maximum readings were less than EC_{ad}, mean, median, minimum, and maximum values, indicating a more con-

ductive soil component in the lower soil depth. Similar trends in EC_{as} and EC_{ad} summary statistics have been observed at this farm [23]. The EC_{as} coefficient of variation of MF2 was greater than the EC_{ad} coefficient of variation. The opposite was observed at MF9.

Moran's I statistics values for MF2 EC_{as} and EC_{ad} readings were greater than 0.90 (Table 2), indicating a statistically significant autocorrelation for EC_{as} and E_{ad} readings. The EC_{as} readings at MF9 also had statistically significant autocorrelation but to a lesser extent when compared with MF2 results (Table 2).

Figure 1 shows EC_{as} and EC_{ad} maps of MF2. The spatial patterns were similar between the EC_{as} and EC_{ad} readings, and the transitioning of the soil to higher EC_a values horizontally and vertically was evident in the map comparisons. For MF2, the lowest EC_{as} values were observed in the southwest section of the field, whereas the higher EC_{as} and EC_{ad} values occurred in the northern portion of the plot. For this dataset, moderate values were observed in the middle of the field. The leave-one-out cross-validation accuracy for model selection was higher for the EC_{as} readings than for the EC_{ad} readings (Table 3).

Table 1. Summary statistics for study sites MF2 and MF9.

Location	n	Variable	Mean (mS/m)	Median (mS/m)	Minimum (mS/m)	Maximum (mS/m)	CV (%)
MF2	1573	EC_{as}	55.1	57.1	22.5	90.4	27.9
		EC_{ad}	66.9	66.0	29.4	114.6	25.2
MF9	741	EC_{as}	91.4	91.8	74.2	107.9	7.0
		EC_{ad}	108.0	108.0	81.5	133.0	8.8

CV—Coefficient of variation.

Table 2. Moran's I measurement of autocorrelation.

Location	Variable	Moran Value	p
MF2	EC_{as}	0.970	0.001
	EC_{ad}	0.959	0.001
MF9	EC_{as}	0.739	0.001
	EC_{ad}	0.800	0.001

EC_{as}—shallow apparent electrical conductivity readings; EC_{ad}—deep apparent electrical conductivity readings.

Table 3. Leave-one-out cross-validation results of model used to derive maps of the study sites.

Location	Variable	RMSE	R^2
MF2	EC_{as}	4.86	0.901
	EC_{ad}	6.44	0.863
MF9	EC_{as}	4.22	0.576
	EC_{ad}	6.75	0.505

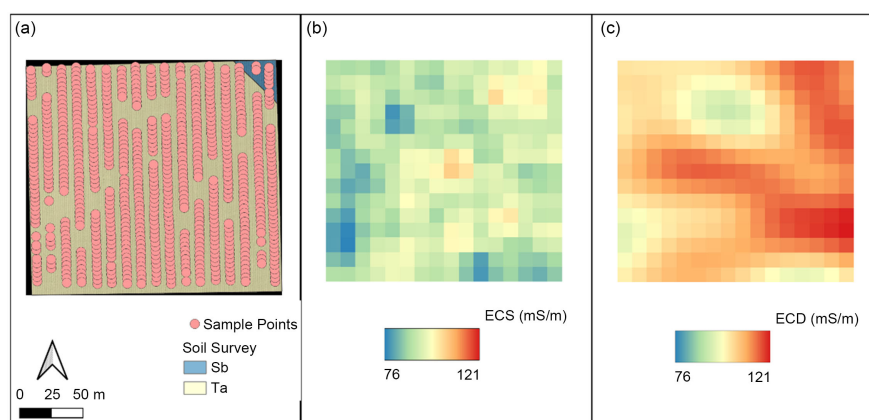RMSE—root mean square error, R^2—coefficient of variation.
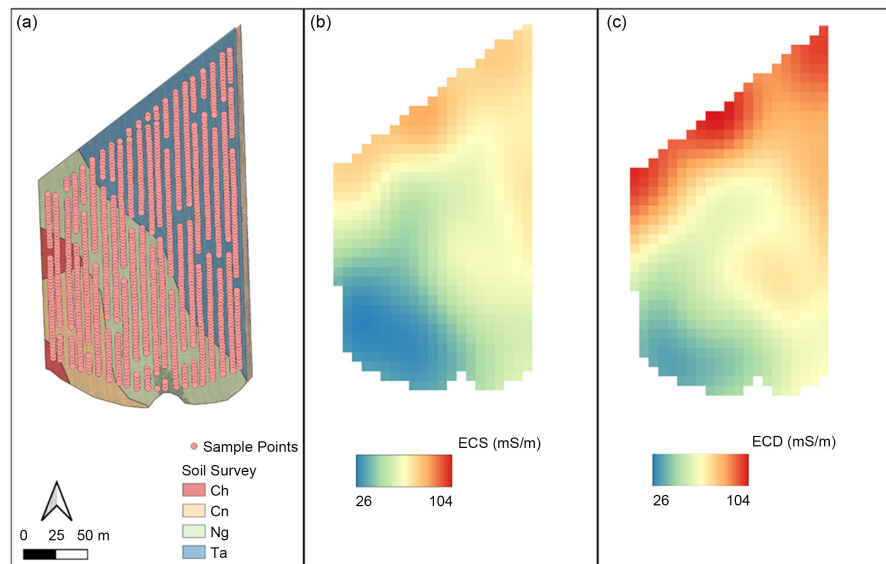
Figure 1. (a) Study site MF2, sampling points, and soil mapping units, (b) apparent electrical conductivity shallow (ECS) readings, and (c) apparent electrical conductivity deep (ECD) readings. Ch—Commerce silty clay loam, 0% to 2% slopes; Cn—Commerce very fine sandy loam, 0% to 2% slopes; Ng—Newellton silty clay, 0% to 2% slopes, occasionally flooded; and Ta—Tunica clay, 0% to 2% slopes.

Figure 2 illustrated the apparent electrical conductivity readings at MF9. The EC_{as} readings were more variable and showed weaker patterns than the EC_{ad} readings. A distinct pattern was observed in the EC_{ad} readings with the highest readings occurring in the middle and eastern sections of the field. *Pereira et al.* [7] indicated that the plugin was not a one-fits-all soil mapping software. Khaledian and Miller's [5] review of machine learning tools for digital soil mapping also stressed that there is no ideal protocol for developing models for digital soil mapping. The EC_{as} results at MF9 support that concept; however, future research must be conducted to determine why higher accuracies could not be achieved for model development at MF9.

Furthermore, another type of predictor may have been better for interpolating the EC_a layers, such as kriging. For example, *Veronesi and Schillaci* [4] showed in a comparison study of kriging to machine learning algorithms to predict topsoil organic carbon that ordinary and universal kriging were the best predictors, followed by random forest. According to a review by [10], random forest is the most popular machine learning tool used for regression purposes related to digital soil mapping. The Smart-Map tool does offer an option to complete kriging; thus, it will be explored in future research studies. The software uses support vector machine for the machine learning approach, which has been used less than other machine learning tools for digital soil mapping studies [10] [29].

Strong spatial contrasts were evident at MF2 compared to MF9. The results also indicated the strength of the patterns was depth dependent on the fields studied. Generally, the lower EC_a values at MF2 occurred in areas consisting of

Figure 2. (a) Study site MF9, sampling points, and soil mapping units, (b) apparent electrical conductivity shallow (ECS) readings, and (c) apparent electrical conductivity deep (ECD) readings. Sb—Sharkey clay, 0% to 2% slopes; and Ta—Tunica clay, 0% to 2% slopes.

silty clay loam, very fine sandy loam, and silty clays soils based on soil survey results (Figure 2). Additionally, this field was irrigated from south to north, which could have contributed to the higher EC_a values observed in the northern section of the field. Fletcher [23] has also observed similar results based on cluster analysis of a field at the same farm.

This study was conducted on a research farm with 5 - 10 ha plots. MF2 was approximately double the size of MF9, which probably led to better spatial patterns observed in the former compared with the latter. To improve the mapping of the EC_{as} layer for MF9, the distance between the mapping transects may need to be decreased from the 8 m to possibly 4 m. Also, the transects were collected along the rows for each field. Collecting the transects along and across rows should also be explored. However, that change would result in more time to collect and analyze the data. Finally, it is essential to determine what sampling design is optimal for machine learning tools used in digital soil mapping [9] [10]. Model choice and sample design can influence final outputs.

4. Conclusion

The research results indicated that machine learning was valuable for deriving EC_{as} maps in two Mississippi fields located on a research farm. Open-source software and machine learning based on support vector machine was used to derive the maps. Autocorrelation of EC_{as} and EC_{ad} measurements was site-specific. Location and depth played a role in the machine learner's ability to derive the maps. Overall spatial patterns in EC_a were evident in both fields; these maps can aid in developing strategies to collect soil and plant samples. Future research will focus on using the other tools provided by the software to establish management zones for the fields located at the research facility, evaluate the effect of sample design on machine learning tools, and compare different algorithms at the field scale.

Acknowledgements

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Bourennane, H., Nicoullaud, B., Couturier, A. and King, D. (2004) Exploring the Spatial Relationships between Some Soil Properties and Wheat Yields in Two Soil Types. *Precision Agriculture*, **5**, 521-536. https://doi.org/10.1007/s11119-004-5323-z

[2] Corwin, D.L., Lesch, S.M., Shouse, P.J., Soppe, R. and Ayars, J.E. (2003) Identifying Soil Properties That Influence Cotton Yield Using Soil Sampling Directed by Apparent Soil Electrical Conductivity. *Agronomy Journal*, **95**, 352-364. https://doi.org/10.2134/agronj2003.3520

[3] Corwin, D.L. and Lesch, S.M. (2005) Apparent Soil Electrical Conductivity Measurements in Agriculture. *Computers and Electronics in Agriculture*, **46**, 11-43. https://doi.org/10.1016/j.compag.2004.10.005

[4] Veronesi, F. and Schillaci, C. (2019) Comparison between Geostatistical and Machine Learning Models as Predictors of Topsoil Organic Carbon with a Focus on Local Uncertainty Estimation. *Ecological Indicators*, **101**, 1032-1044. https://doi.org/10.1016/j.ecolind.2019.02.026

[5] Khaledian, Y. and Miller, B.A. (2020) Selecting Appropriate Machine Learning Methods for Digital Soil Mapping. *Applied Mathematical Modelling*, **81**, 401-418. https://doi.org/10.1016/j.apm.2019.12.016

[6] Adhikari, K., Smith, D.R., Collins, H., Hajda, C., Acharya, B.S. and Owens, P.R. (2022) Mapping Within-Field Soil Health Variations Using Apparent Electrical Conductivity, Topography, and Machine Learning. *Agronomy*, **12**, Article No. 1019. https://doi.org/10.3390/agronomy12051019

[7] Pereira, G.W., Valente, D.S.M., Queiroz, D.M. de, Coelho, A.L. de F., Costa, M.M. and Grift, T. (2022) Smart-Map: An Open-Source QGIS Plugin for Digital Mapping Using Machine Learning Techniques and Ordinary Kriging. *Agronomy*, **12**, Article No. 1350. https://doi.org/10.3390/agronomy12061350

[8] Pouladi, N., Møller, A.B., Tabatabai, S. and Greve, M.H. (2019) Mapping Soil Organic Matter Contents at Field Level with Cubist, Random Forest and Kriging. *Geoderma*, **342**, 85-92. https://doi.org/10.1016/j.geoderma.2019.02.019

[9] Wadoux, A.M.J.-C., Brus, D.J. and Heuvelink, G.B.M. (2019) Sampling Design Optimization for Soil Mapping with Random Forest. *Geoderma*, **355**, 113913. https://doi.org/10.1016/j.geoderma.2019.113913

[10] Wadoux, A.M.J.-C., Minasny, B. and McBratney, A.B. (2020) Machine Learning for Digital Soil Mapping: Applications, Challenges and Suggested Solutions. *Earth-*

Science Reviews, **210**, Article ID: 103359.
https://doi.org/10.1016/j.earscirev.2020.103359

[11] Bishop, C.M. (2006) Pattern Recognition and Machine Learning. Springer, Berlin.
https://link.springer.com/book/9780387310732

[12] Henderson, B.L., Bui, E.N., Moran, C.J. and Simon, D.A.P. (2005) Australia-Wide Predictions of Soil Properties Using Decision Trees. *Geoderma*, **124**, 383-398.
https://doi.org/10.1016/j.geoderma.2004.06.007

[13] Grimm, R., Behrens, T., Märker, M. and Elsenbeer, H. (2008) Soil Organic Carbon Concentrations and Stocks on Barro Colorado Island—Digital Soil Mapping Using Random Forests Analysis. *Geoderma*, **146**, 102-113.
https://doi.org/10.1016/j.geoderma.2008.05.008

[14] Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A. and Liu, D.L. (2018) High Resolution Mapping of Soil Organic Carbon Stocks Using Remote Sensing Variables in the Semi-Arid Rangelands of Eastern Australia. *The Science of the Total Environment*, **630**, 367-378. https://doi.org/10.1016/j.scitotenv.2018.02.204

[15] Chagas, C.S., de Carvalho Junior, W., Bhering, S.B. and Calderano Filho, B. (2016) Spatial Prediction of Soil Surface Texture in a Semiarid Region Using Random Forest and Multiple Linear Regressions. *Catena*, **139**, 232-240.
https://doi.org/10.1016/j.catena.2016.01.001

[16] Vaysse, K. and Lagacherie, P. (2017) Using Quantile Regression Forest to Estimate Uncertainty of Digital Soil Mapping Products. *Geoderma*, **291**, 55-64.
https://doi.org/10.1016/j.geoderma.2016.12.017

[17] Dharumarajan, S., Hegde, R. and Singh, S.K. (2017) Spatial Prediction of Major Soil Properties Using Random Forest Techniques—A Case Study in Semi-Arid Tropics of South India. *Geoderma Regional*, **10**, 154-162.
https://doi.org/10.1016/j.geodrs.2017.07.005

[18] Forkuor, G., Hounkpatin, O.K.L., Welp, G. and Thiel, M. (2017) High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLOS ONE*, **12**, e0170478. https://doi.org/10.1371/journal.pone.0170478

[19] Song, Y.-Q., Zhao, X., Su, H.-Y., Li, B., Hu, Y.-M. and Cui, X.-S. (2018) Predicting Spatial Variations in Soil Nutrients with Hyperspectral Remote Sensing at Regional Scale. *Sensors*, **18**, Article No. 3086. https://doi.org/10.3390/s18093086

[20] Hengl, T., Leenaars, J.G.B., Shepherd, K.D., Walsh, M.G., Heuvelink, G.B.M., Mamo, T., Tilahun, H., Berkhout, E., Cooper, M., Fegraus, E., Wheeler, I. and Kwabena, N.A. (2017) Soil Nutrient Maps of Sub-Saharan Africa: Assessment of Soil Nutrient Content at 250 m Spatial Resolution Using Machine Learning. *Nutrient Cycling in Agroecosystems*, **109**, 77-102. https://doi.org/10.1007/s10705-017-9870-x

[21] Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D. and Campbell, P.H. (2015) The Australian Three-Dimensional Soil Grid: Australia's Contribution to the GlobalSoilMap Project. *Soil Research*, **53**, 845-864.
https://doi.org/10.1071/SR14366

[22] Bou Kheir, R., Greve, M.H., Abdallah, C. and Dalgaard, T. (2010) Spatial Soil Zinc Content Distribution from Terrain Parameters: A GIS-Based Decision-Tree Model in Lebanon. *Environmental Pollution*, **158**, 520-528.
https://doi.org/10.1016/j.envpol.2009.08.009

[23] Fletcher, R.S. (2022) Temporal Comparisons of Apparent Electrical Conductivity: A Case Study on Clay and Loam Soils in Mississippi. *Agricultural Sciences*, **13**, 936-946.

https://doi.org/10.4236/as.2022.138058

[24] WorldClimate. Stoneville, Mississippi Climate—38756 Weather, Average Rainfall, and Temperatures. http://www.worldclimate.com/climate/us/mississippi/stoneville

[25] Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey.
https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcseprd1464818

[26] Veris Technologies. Optic Mapper Operating Instructions. Manual, Veris Technologies, 61.

[27] QGIS Development Team (2021) QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

[28] Moran, P. (1948) The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society*: *Series B* (*Methodological*), **10**, 243-251.
https://doi.org/10.1111/j.2517-6161.1948.tb00012.x

[29] Guevara, M., Olmedo, G.F., Stell, E., Yigini, Y., Aguilar Duarte, Y., Arellano Hernández, C., Arévalo, G.E., Arroyo-Cruz, C.E., Bolivar, A., Bunning, S., Bustamante Cañas, N., Cruz-Gaistardo, C.O., Davila, F., Dell Acqua, M., Encina, A., Figueredo Tacona, H., Fontes, F., Hernández Herrera, J.A., Ibelles Navarro, A.R., Loayza, V., Manueles, A.M., Mendoza Jara, F., Olivera, C., Osorio Hermosilla, R., Pereira, G., Prieto, P., Ramos, I.A., Rey Brina, J.C., Rivera, R., Rodríguez-Rodríguez, J., Roopnarine, R., Rosales Ibarra, A., Rosales Riveiro, K.A., Schulz, G.A., Spence, A., Vasques, G.M., Vargas, R.R. and Vargas, R. (2018) No Silver Bullet for Digital Soil Mapping: Country-Specific Soil Organic Carbon Estimates across Latin America. *Soil*, **4**, 173-193. https://doi.org/10.5194/soil-4-173-2018