

On-the-Go Prediction of Soil pH Using Generalized Additive Models in Mississippi Delta: A Case Study

Reginald S. Fletcher 

USDA-ARS Stoneville, MS, USA

Email: reginald.fletcher@usda.gov

How to cite this paper: Fletcher, R.S. (2025) On-the-Go Prediction of Soil pH Using Generalized Additive Models in Mississippi Delta: A Case Study. *Open Journal of Soil Science*, 15, 225-234.

<https://doi.org/10.4236/ojss.2025.154011>

Received: October 3, 2024

Accepted: March 30, 2025

Published: April 2, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Soil pH is a critical indicator of soil health and fertility, influencing nutrient availability and crop productivity. Leveraging real-time sensor technology allows for high-resolution data collection across diverse agricultural landscapes, yet effective modeling techniques are required to interpret these complex datasets. This study evaluated applying generalized additive models (GAMs) to predict soil pH in Mississippi using data collected from an on-the-go soil sensor. pH was the dependent variable, and apparent electrical conductivity shallow (EC_{as}) and deep readings (EC_{ad}), x and y coordinates, EC_a ratio, altitude, curvature, and slope were the independent variables used to develop GAMs capturing the nonlinear relationships between the predictors and soil pH. For the study site, a GAM derived from the x-coordinate and the x- and y-coordinate interaction was best for estimating pH. It achieved an r-squared value of 0.84 and a root mean squared error of 0.08 on the original testing dataset and an r-squared value of 0.87 and a root mean squared error of 0.07 on a bootstrap simulated dataset created from the original testing set. The model effectively exhibited the nonlinear dynamics of soil pH, providing insights into the relative contributions of individual predictors. This approach enhances prediction accuracy and offers interpretability, allowing agronomists to identify critical factors affecting soil pH. The findings support the potential of GAMs as a valuable tool for precision agriculture, facilitating informed decision-making for soil management and crop production.

Keywords

Machine Learning, Generalized Additive Models, Soil Chemical Properties, Mobile Sensors, Coordinates

1. Introduction

Digital soil mapping (DSM) involves using computers and data to create detailed maps of soil properties across different landscapes [1] [2]. It integrates various data sources, including remote sensing, geographic information systems (GIS), and soil property measurements, to produce maps for land use planning and environmental management [1] [3]-[8]. Overall, digital soil mapping is a powerful tool for sustainable land management and agricultural productivity, leveraging technology to enhance our understanding of soil resources [7] [8].

Soil pH, a critical factor influencing soil health, plant growth, and nutrient availability, measures the acidity or alkalinity of the soil with a scale ranging from 0 to 14 [9]-[12]. Values below 7 indicate acidity, while values above 7 indicate alkalinity. Most crops thrive in a slightly acidic to neutral pH range of 6 to 7.5, where essential nutrients like nitrogen, phosphorus, and potassium are readily available. Soil pH also affects microbial activity and the solubility of minerals. Acidic soils can lead to toxic levels of aluminum and manganese, inhibiting root development and overall plant growth [10]-[12]. Conversely, alkaline soils may limit the availability of essential micronutrients such as iron and zinc [10] [11]. Managing soil pH is necessary for sustainable agriculture.

Traditional soil sampling methods are tedious and time-consuming and often involve extensive fieldwork, including sampling from various locations and depths that requires thorough planning to ensure representative sampling [1] [3] [6]. Over the years, the commercial industry has developed on-the-go soil sensors that collect data which can serve as a proxy to estimate soil chemical and physical properties, revolutionizing agricultural practices by providing real-time data on soil conditions. These advanced tools measure soil properties, such as moisture content, pH, and apparent electrical conductivity. This technology has enabled precise and timely decision-making, thus optimizing crop management, enhancing yield, leading to cost savings, and promoting sustainable farming practices.

On-the-go soil sensor systems, like the Veris MSP3, incorporate multiple sensors and stand out as advanced tools for precision agriculture, offering comprehensive capabilities in collecting soil data. While traversing fields, this innovative system employs advanced sensor technology to measure fundamental soil properties, including EC_a , reflectance, and pH. Soil reflectance information has been used as a proxy to map organic matter and organic carbon concentrations in soil [13]. By capturing data in real-time, the MSP3 system enables farmers to create detailed soil maps illustrating spatial variability, essential for informed decision-making. One of the standout features of the Veris MSP3 is its ability to operate in various field conditions, ensuring consistent data collection regardless of terrain. The system is equipped with GPS technology, allowing for precise georeferencing of soil data, enhancing the resulting maps' accuracy. This on-the-go system represents a significant advancement in digital soil mapping, helping to bridge the gap between traditional agriculture and modern data-driven farming techniques.

The MSP3 system samples pH at a lower frequency than its other sensors, re-

sulting in fewer data points and less detailed pH mapping. This design choice for pH data collection reflects a deliberate engineering trade-off rather than a flaw. Currently, it is unclear whether the data from the other sensors could be used to estimate pH at locations lacking direct measurements. If such estimation proves feasible, a more detailed soil sampling map could be developed for calibration with field samples. This study speculates that machine learning may offer the necessary predictive capability by utilizing the additional data captured by the MSP3 system to estimate pH. Information is lacking in that concept in the Mississippi Delta, one of the country's premier agricultural areas.

Machine learning has emerged as a transformative tool in digital soil mapping because it can leverage complex concepts to identify patterns and relationships that traditional methods might overlook [1] [4]-[6] [14]. For example, random forests and support vector machines are standard machine-learning tools used to predict key soil properties like pH, organic matter, and nutrient levels [5] [9]. These methods utilize vast amounts of data to build robust predictive models, allowing for the spatial interpolation of soil characteristics across large areas. Advanced tools such as artificial neural networks can capture nonlinear relationships, making them particularly effective in diverse landscapes [5] [6]. Moreover, machine learning facilitates the integration of various data sources, enabling comprehensive assessments of soil variability. Machine learning algorithms have been commonly used to estimate pH [1] [9].

GAMs are a flexible machine-learning approach for modeling complex relationships between variables. They extend traditional linear models by allowing users to model the nonlinear effects of predictor variables through smooth functions [15]-[17]. This adaptability increased its popularity as a tool for sciences such as ecology [17] and agriculture [16], where relationships are often not linear. The additive nature of GAMs enables the incorporation of various types of predictors, making it easier to interpret the contributions of each variable, which is often vital for soil science. Additionally, GAMs provide a robust framework for analyzing data with varying distributions, enhancing predictive accuracy while maintaining interpretability, essential for informed decision-making. More information is needed on using GAMs to predict soil properties, specifically in Mississippi, USA. It is hypothesized that GAMs effectively capture the nonlinear relationships between soil sensor measurements and soil pH, leading to accurate predictions that can enhance understanding of soil health and inform precision agriculture practices. This study aims to investigate the effectiveness of GAMs in capturing the nonlinear relationship between pH and soil parameters measured by on-the-go sensors. The expectation is that GAMs will provide enhanced predictive performance and greater interpretability, particularly in complex agricultural environments where soil properties exhibit intricate interactions.

2. Materials and Methods

The study was conducted at the United States Department of Agriculture, Agri-

cultural Research Service Farm in Stoneville, Mississippi, USA. The data were collected May 5, 2015 (**Figure 1**). The study area was approximately 3.7 ha and consisted of the following soil types [18]: Commerce silty clay loam (Ch, 0% to 2% slopes); Newellton silty clay (Ng, 0% to 2% slopes, occasionally flooded); Sharkey clay (Sb, 0.5% to 2% slopes); and Tunica clay (Ta, 0% to 2% slopes).

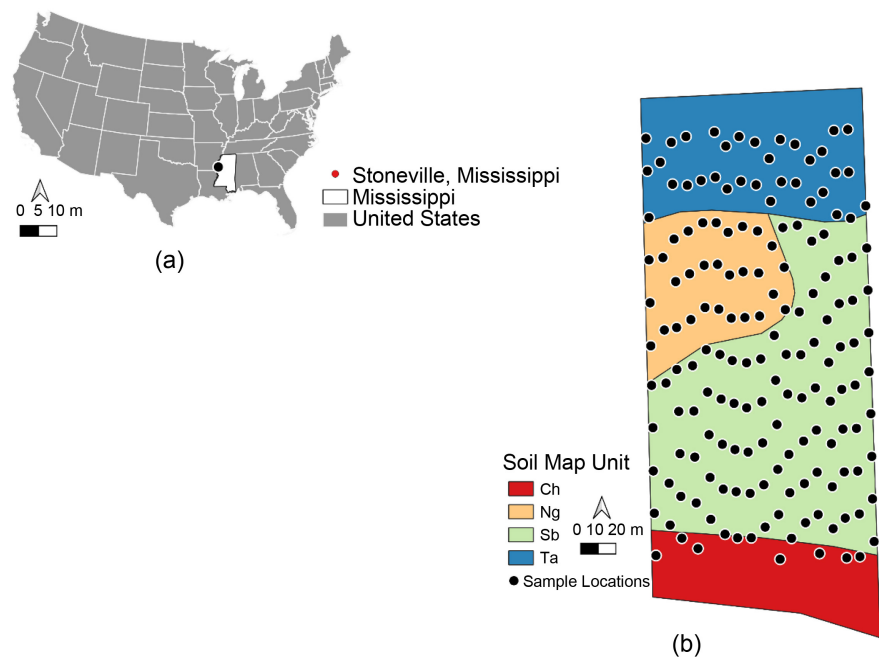


Figure 1. (a) General location of the study site. (b) Study plot and location of sampling points and the different soil types. Commerce silty clay loam (Ch, 0% to 2% slopes); Newellton silty clay (Ng, 0% to 2% slopes, occasionally flooded); Sharkey clay (Sb, 0.5% to 2% slopes); Tunica clay (Ta, 0% to 2% slopes).

The Veris MSP3 on-the-go soil mapping system was employed to collect soil data; it acquired apparent electrical conductivity, near-infrared and red soil reflectance readings, latitude, longitude, pH, slope, and altitude as a tractor pulled the system across the field. The tractor was driven along 17 predetermined transects at approximately 11 km/hr (**Figure 1**). EC_{as} and EC_{ad} readings were collected from 0 - 30 cm and 0 - 90 cm respectively [19] [20]. Soil reflectance readings were collected using optic mapper, which measures reflectance from 2.5 to 5 cm below the soil surface while being pulled through the field. The pH system employs a sampling shoe that collects soil approximately 8 cm beneath the surface. This sample is then brought into contact with pH electrodes for measurement. After the pH is recorded, the system automatically rinses the electrodes, and the sampling shoe is repositioned back into the soil to collect the next sample. Notably, all these actions occur while the vehicle is in motion.

The near-infrared and red soil reflectance values recorded by the on-the-go system were not recorded in a format available for quantitative analysis. The user must upload the data to the Field Fusion web data site to obtain the reflectance

values. Veris customer support will convert the near-infrared and red digital count data into a readable format. The customer support staff will also clean the raw data (removing outliers and noise) and provide additional information such as the EC_a ratio, slope, and curvature.

After the initial data evaluation, Veris Technologies concluded that the near-infrared and red measurements were questionable for further analysis. Therefore, for this study, pH estimates were based on the following parameters: x-coordinate, y-coordinate, EC_{as} , EC_{ad} , altitude, EC_a ratio, curvature, and slope. 2267 samples were acquired for x-coordinate, y-coordinate, EC_{as} , EC_{ad} , altitude, EC_a ratio, curvature, and slope, and 202 samples were obtained for the pH. The pH measurements and the other measurement locations are not exactly aligned; therefore, the QGIS software was used to assign the pH information to the closest point in space for the other measurements. After further evaluation, 166 samples were available for further analysis (**Figure 1**) to evaluate the relationships between the pH and other soil measurements.

The data were analyzed using R studio (version 2024.12.1 Build 563, 2009-2025) and the R software (version 4.4.3; “Trophy Case”; [21]). RStudio is an integrated development environment (IDE) specifically designed for working with the R programming language. Its user-friendly interface simplifies coding, especially for tasks like statistical analysis, data visualization, and machine learning. Data analyses were broken down into the following steps: Separating data into training and testing sets, creating simulation data from the testing set, training the model using the original training set, testing the model on the simulated testing and original testing datasets. The training dataset consisted of 114 samples; the original testing and simulated testing consisted of 52 samples. The simulated testing set was derived using bootstrapping with replacement [22]. Base r was used to create the bootstrap simulation dataset consisting of 52 samples.

The generalized additive model algorithm located in the `mgcv` package [17] [23]-[26] was employed to establish the relationships between the dependent and independent variables based on the restricted maximum likelihood (REML) criteria as the basis for the smoothing parameter. The formula was specified to include both linear and smooth terms to capture non-linear relationships. The general form of the generalized additive model is described as follows:

$$g(Y) = b + f_1(x_1) + f_2(x_2) + \dots + f_n(x_n) \quad (1)$$

where Y is the expected response value, b is the model intercept, and f_1 , f_2 , f_n are the smooth functions of the predictors x_1 , x_2 , x_n [17]. It does not make any assumptions about a linear agreement between independent and dependent variables. For this study, the independent variables were the x and y coordinates, EC_{as} , EC_{ad} , the ratio of EC_{as} and EC_{ad} , altitude, and slope. Model diagnostics were performed using the summary function to evaluate the significance of predictors, along with the plot function to visualize smooth terms. Residuals were checked for homoscedasticity and normality. R-squared (R^2 , coefficient of determination) [22] and the root mean squared error [22] were used to evaluate the accuracy of

the final model on testing datasets.

3. Results and Discussion

The initial GAM model (**Table 1**) was constructed to examine the relationship between pH levels and various predictors captured by the on-the-go system, including spatial coordinates (x_{utm} , y_{utm}), altitude, curvature, slope, and soil conductivity measures (ec_{sh} , ec_{dp} , ec_{ratio}). The model highlighted significant non-linear effects of $x_{coordinate}$ ($P < 0.001$) and slope ($P < 0.05$) on pH, suggesting spatial variability in the east-west direction and a moderate influence of slope. However, the other measurements showed minimal contributions ($P > 0.05$). The overall model explained 40.7% of the deviance (adjusted $R^2 = 0.37$), indicating a reasonable fit but leaving considerable room for improvement. These results suggested a need for refinement, particularly to address insignificant predictors and enhance the model's explanatory power.

Table 1. Summary information for the starting point generalized additive model.

Family: gaussian					
Link function: identity					
Formula: $pH_{avg} \sim s(x_{coordinate}) + s(y_{coordinate}) + s(altitude) + s(ec_{sh}) + s(ec_{dp}) + s(ec_{ratio}) + s(slope) + s(curvature)$					
Parametric coefficients:		Estimate	Std. Error	t value	Pr (>t)
Intercept		6.54	0.02	404.8	<0.001***
Approximate significance of smooth terms:					
	edf	Ref. df	F	P-value	
$s(x_{coordinate})$	4.53	9	5.65196	<0.001***	
$s(y_{coordinate})$	0.64	9	0.19583	0.10	
$s(altitude)$	0.000001	9	0	0.66	
$s(ec_{sh})$	0.000001	9	0	0.83	
$s(ec_{dp})$	0.000001	9	0	0.60	
$s(ec_{ratio})$	0.000001	9	0	0.63	
$s(slope)$	0.79	9	0.42662	0.03*	
$s(curvature)$	0.000003	9	0	0.40	
$R^2(\text{adj}) = 0.37$					
Deviance explained = 40.7%					
REML = -27.21					
n = 114					

n = number of samples, DE = deviance explained, edf = effective degrees of freedom, and Ref. df = reference degrees of freedom, Std. = standard. Statistical significance: ***P < 0.001, **P < 0.01, *P < 0.05.

The refined GAM model (**Table 2**) provided strong explanatory power, with 86.3% of the deviance explained and an adjusted R^2 of 0.81. The response variable

(pH) was modeled using a single smoother for the x-coordinate and a tensor interaction between x- and y-coordinates [te(x_utm, y_utm)]. The smoother for the x-coordinate was statistically significant ($P < 0.001$), highlighting substantial variability in the east-west direction. The tensor interaction term was also statistically significant ($P < 0.001$), capturing meaningful spatial dependencies between x- and y-coordinates. These results indicated that pH levels were influenced primarily by spatial patterns, with strong contributions from both the east-west direction and the combined interaction of location. Jiang *et al.* (2018) observed from GAMs output that zinc concentrations in soils were influenced by their location on the landscape which agreed with the current study. The model highlights the importance of considering spatial correlations for accurate pH prediction while maintaining a meaningful structure by focusing on the most variables.

Table 2. Summary information for the final generalized additive model.

Family: gaussian				
Link function: identity				
Formula: pHavg ~ s(xcoordinate, k = 50) + te(xcoordinate, Ycoordinate)				
Parametric coefficients:	Estimate	Std. Error	t value	Pr (>t)
Intercept	6.54	0.009	738.33	<0.001***
Approximate significance of smooth terms:				
	edf	Ref. df	F	p-value
s(xcoordinate)	27.73	49	5.38	<0.001***
te(xcoordinate, Ycoordinate)	3.17	23	1.46	<0.001***
R-sq.(adj) = 0.81				
Deviance explained = 86.3%				
REML = -42.989				
n = 114				

n = number of samples, DE = deviance explained, edf = effective degrees of freedom, and Ref. df = reference degrees of freedom, Std. = standard. Statistical significance: *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$.

The estimated degrees of freedom (edf) value illustrated a highly complex relationship (**Table 2**); an edf greater than 1 [17] [27] [28] indicated the relationship between pH and the east-west direction was nonlinear, validating the appropriateness of using GAMs in this study. Looking at the map in **Figure 1** showing the soil mapping units, one could get a general idea of why the pH values were changing across the field.

The k value in GAMs defines the basis dimension, specifying the number of basis functions used to model smooth terms and control the flexibility of nonlinear relationships [17] [26]. For the final model (**Table 2**), a value of 50 was used for the x-coordinate. For a study estimating organic matter, [28] observed that k values between 50 and 60 worked well.

Performance metrics further supported the model's effectiveness, with $R^2 = 0.84$ and root mean squared error (RMSE) = 0.08 for the original testing dataset, indicating that a substantial proportion of variance in soil pH was explained. Similar values were observed for the test dataset created using the bootstrap technique: $R^2 = 0.87$ and RMSE = 0.07. Generally, one could conclude the model was robust in that it achieved values comparable for two different testing datasets. Keep in mind that this data was collected from a case study and model was tested on an original test set and a simulated dataset created from bootstrapping the original test dataset. Therefore, more research needs to be done in other fields. However, the analysis provided valuable insights into the contributions of individual predictors and allowed for a clearer understanding of how these variables interact and affect pH levels in real time. Overall, the findings emphasized the utility of GAMs in precision agriculture, enabling agronomists to make data-driven decisions regarding soil management and crop production. The interpretability of GAMs is a significant advantage, as it allows agronomists to discern the relative contributions of individual predictors to soil pH.

4. Conclusions

This case study demonstrated the potential of Generalized Additive Models (GAMs) in modeling soil pH, a critical factor influencing agricultural production. By capturing non-linear relationships, GAMs provide a flexible approach to understanding how environmental variables measured by an on-the-go soil measuring system affect pH levels. For this study, location played a major role in changes of pH values.

As noted earlier, pH sampling was conducted at a lower resolution compared to the coordinate data by the on-the-go system. The derived model performed strongly on two testing datasets, demonstrating its capability to estimate pH at unsampled locations and supporting its use in creating a more detailed pH map. In real world situations, it is important to note that each field functions as an independent system, meaning factors influencing pH may vary significantly between fields. Future research should explore the integration of additional variables and the applicability of this approach to larger fields in Mississippi to advance precision agriculture.

Acknowledgements

The author wishes to thank Milton Gaston, Jr., for his assistance in data collection. Funding was received from the United States Department of Agriculture. Statements made in this article were the author's opinion and did not represent the opinion of the United States Department of Agriculture.

Conflicts of Interest

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

References

- [1] Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., *et al.* (2022) Digital Mapping of Globalsoilmap Soil Properties at a Broad Scale: A Review. *Geoderma*, **409**, Article ID: 115567. <https://doi.org/10.1016/j.geoderma.2021.115567>
- [2] Piikki, K., Wetterlind, J., Söderström, M. and Stenberg, B. (2021) Perspectives on Validation in Digital Soil Mapping of Continuous Attributes—A Review. *Soil Use and Management*, **37**, 7-21. <https://doi.org/10.1111/sum.12694>
- [3] McBratney, A.B., Mendonça Santos, M.L. and Minasny, B. (2003) On Digital Soil Mapping. *Geoderma*, **117**, 3-52. [https://doi.org/10.1016/s0016-7061\(03\)00223-4](https://doi.org/10.1016/s0016-7061(03)00223-4)
- [4] Minasny, B. and McBratney, A.B. (2016) Digital Soil Mapping: A Brief History and Some Lessons. *Geoderma*, **264**, 301-311. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- [5] Padarian, J., Minasny, B. and McBratney, A.B. (2020) Machine Learning and Soil Sciences: A Review Aided by Machine Learning Tools. *Soil*, **6**, 35-52. <https://doi.org/10.5194/soil-6-35-2020>
- [6] Zhang, G., Liu, F. and Song, X. (2017) Recent Progress and Future Prospect of Digital Soil Mapping: A Review. *Journal of Integrative Agriculture*, **16**, 2871-2885. [https://doi.org/10.1016/s2095-3119\(17\)61762-3](https://doi.org/10.1016/s2095-3119(17)61762-3)
- [7] Poppiel, R.R., Demattê, J.A.M., Rosin, N.A., Campos, L.R., Tayebi, M., Bonfatti, B.R., *et al.* (2021) High Resolution Middle Eastern Soil Attributes Mapping via Open Data and Cloud Computing. *Geoderma*, **385**, Article ID: 114890. <https://doi.org/10.1016/j.geoderma.2020.114890>
- [8] Auzzas, A., Capra, G.F., Jani, A.D. and Ganga, A. (2024) An Improved Digital Soil Mapping Approach to Predict Total N by Combining Machine Learning Algorithms and Open Environmental Data. *Modeling Earth Systems and Environment*, **10**, 6519-6538. <https://doi.org/10.1007/s40808-024-02127-8>
- [9] Nussbaum, M., Zimmermann, S., Walthert, L. and Baltensweiler, A. (2023) Benefits of Hierarchical Predictions for Digital Soil Mapping—An Approach to Map Bimodal Soil pH. *Geoderma*, **437**, Article ID: 116579. <https://doi.org/10.1016/j.geoderma.2023.116579>
- [10] Khaled, F. and Sayed, A. (2023) Soil Ph and Its Influence on Nutrient Availability and Plant Health. *International Journal of Advanced Chemistry Research*, **5**, 68-70. <https://doi.org/10.33545/26646781.2023.v5.i2a.204>
- [11] Anderson, A., Khaleel, A., Dutter, C., Blauwet, M., Flores, A., Miller, B., Burras, C.L. and Fidel, R. (2023) Introduction to Soil Science. Iowa State University, Digital Press. <https://iastate.pressbooks.pub/introsoilscience/>
- [12] O’Kennedy, S. (2022) Soil pH and Its Impact on Nutrient Availability and Crop Growth. *International Journal of Geography, Geology and Environment*, **4**, 236-238.
- [13] Kweon, G. and Maxton, C. (2013) Soil Organic Matter Sensing with an On-the-Go Optical Sensor. *Biosystems Engineering*, **115**, 66-81. <https://doi.org/10.1016/j.biosystemseng.2013.02.004>
- [14] Wadoux, A.M.J., 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>
- [15] Clark, N.J. and Wells, K. (2022) Dynamic Generalised Additive Models (DGAMs) for Forecasting Discrete Ecological Time Series. *Methods in Ecology and Evolution*, **14**, 771-784. <https://doi.org/10.1111/2041-210x.13974>

- [16] Jiang, Y., Gao, W., Zhao, J., Chen, Q., Liang, D., Xu, C., *et al.* (2018) Analysis of Influencing Factors on Soil Zn Content Using Generalized Additive Model. *Scientific Reports*, **8**, Article No. 15567. <https://doi.org/10.1038/s41598-018-33745-9>
- [17] Wood, S.N. (2017) Generalized Additive Models: An Introduction with R, Second Edition. 2nd Edition, Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>
- [18] 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>
- [19] Fletcher, R.S. (2023) Machine Learning Mapping of Soil Apparent Electrical Conductivity on a Research Farm in Mississippi. *Agricultural Sciences*, **14**, 915-924. <https://doi.org/10.4236/as.2023.147061>
- [20] Veris Technologies (2017) Optic Mapper Operating Instructions. Manual, Veris Technologies, 61.
- [21] R Core Team. (2025) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.r-project.org/>
- [22] Kelleher, J., Namee, B. and D'Arcy, A. (2015) Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. The MIT Press.
- [23] Wood, S.N. (2010) Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **73**, 3-36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>
- [24] Wood, S.N. (2004) Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. *Journal of the American Statistical Association*, **99**, 673-686. <https://doi.org/10.1198/016214504000000980>
- [25] Wood, S.N. (2003) Thin Plate Regression Splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **65**, 95-114. <https://doi.org/10.1111/1467-9868.00374>
- [26] Wood, S.N., Pya, N. and Säfken, B. (2016) Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, **111**, 1548-1563. <https://doi.org/10.1080/01621459.2016.1180986>
- [27] Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A. and Smith, G.M. (2009) Mixed Effects Models and Extensions in Ecology with R. Springer. <https://doi.org/10.1007/978-0-387-87458-6>
- [28] Cheng, L., Yan, M., Zhang, W., Guan, W., Zhong, L. and Xu, J. (2024) Interpretable Digital Soil Organic Matter Mapping Based on Geographical Gaussian Process-Generalized Additive Model (GGP-GAM). *Agriculture*, **14**, Article 1578. <https://doi.org/10.3390/agriculture14091578>.