Geospatial Analysis of Crop Production and Publicly-Funded Research Stations to Scale Carbon Markets

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Madeline D. Ginsberg
August 2024

DEDICATION

To my parents, who always have been and will always be my biggest supporters.

ACKNOWLEDGEMENTS

I would first like to acknowledge and thank Dr. Lori Duncan and Dr. Andrea Ludwig for being the greatest major professors I could have had the fortune of working with. Both Dr. Duncan and Dr. Ludwig embody the kind of engineer and mentor I can only hope to become one day. Thank you to Dr. Liem Tran and Dr. Daniel Yoder for serving on my committee and for being such dedicated, influential professors during my time as a student at the University of Tennessee. I am greatly appreciative for all four of my committee members and their unwavering investments in the success of their students.

I am also incredibly grateful for my parents, family, friends, coworkers, fellow students, and fellow faculty and staff in the Department of Biosystems Engineering and Soil Science who have supported me throughout my academic career, but especially throughout my time as a graduate student. As they say, it takes a village, and I would not be where I am today without them.

ABSTRACT

Agricultural carbon markets compensate farmers for the implementation of climate smart practices to increase soil organic carbon (SOC) sequestration in agricultural land. To incentivize farmers to enroll in carbon credit contracts, it is necessary to identify model farms with historical crop and land management data from which carbon markets can base their estimations of SOC sequestration. This research utilizes row crop production, SOC sequestration estimations, and Soil Conditioning Index (SCI) values to evaluate the potential for scaling measurement, reporting, and verification (MRV) efforts for carbon markets in Tennessee. Relationships between University of Tennessee Research and Education Centers (RECs) and cropland in surrounding regions across the state were analyzed by developing a geospatial model to relate SOC sequestration potential of fields on private agricultural land to fields at the RECs. The model was developed using ArcGIS Pro software and an iterative self-organizing clustering technique. Model inputs included factors that affect agricultural SOC sequestration: crop system type (corn, soybeans, and cotton), soil texture, average daily temperature, average daily precipitation, and elevation. Model outputs included clusters of geospatially-correlated regions determined by the model that covered the entire state. To determine the model with the best representation of the geospatially-correlated regions based on the input layers, the silhouette coefficient was calculated and compared for each model output. SOC sequestration estimates from COMET-Farm and Soil Conditioning Index values from the Field to Market Fieldprint Platform were also evaluated for current management and future sustainable management scenarios for each field. The comparisons of these scenarios were used to assess the application of the model results to relate SOC sequestration potential on private agricultural land to publicly-funded RECs in West Tennessee. Future work intends to expand this modeling approach to analyze effects of climate smart land management practices across the Southeastern United States. This work also seeks to encourage further utilization of publicly-funded RECs for use in developing standards for carbon market MRV, as well as other conservation efforts.

TABLE OF CONTENTS

CHAPTER I: INTRODUCTION AND LITERATURE REVIEW 1
1.1 Background 2
1.2 Agricultural Carbon Budgets 3
1.3 Quantifying Carbon 5
1.4 Techniques for Measuring Carbon Sequestration 6
1.4.1 Direct Measurement 7
1.4.2 Indirect Measurement 7
1.4.3 Modeling 9
1.4.4 Hybrid Approach 10
1.5 Reporting and Verifying Carbon Sequestration 10
1.6 MRV Efforts for Carbon Markets in the United States 12
1.7 Geospatial Regional Modeling Approach for Carbon Market MRV 14
1.7.1 Overview of Clustering Techniques 15
1.7.2 Iterative Self-Organized Clustering and Unsupervised Classification 17
1.7.3 Clustering Model of Geospatially-Correlated Regions for MRV Efforts 18
1.8 Model Evaluation Metrics 19
1.8.1 Silhouette Coefficient 19
1.8.2 COMET 20
1.8.3 Soil Conditioning Index 21
1.9 Research Objectives 22
1.9.1 Overall Research Objective 22
1.9.2 Development of Geospatial Model 22
1.9.3 Evaluation of Geospatial Model 22
References 24
CHAPTER II: GEOSPATIAL ANALYSIS OF CROP PRODUCTION AND PUBLICLY-FUNDED RESEARCH STATIONS TO SCALE CARBON MARKETS . 31
2.1 Abstract 32
2.2 Introduction 33
2.3 Materials and Methods 36
2.3.1 Geographical Area of Interest 36
2.3.2 Geographic Information System 38
2.3.3 Data Sources and Descriptions 38
2.3.3.1 Crop Production 39
2.3.3.2 Soil Texture 39
2.3.3.3 Precipitation 43
2.3.3.4 Temperature 45
2.3.3.5 Elevation 45
2.3.4 Input Optimization 45
2.3.5 Model Development 48
2.3.6 Model Evaluation 49
2.3.6.1 Fields on UTIA RECs and Private Farmland 49
2.3.6.2 Silhouette Coefficient 49

2.3.6.3 COMET-Farm Scenarios 52
2.3.6.4 Fieldprint Calculator for Soil Conditioning Index Values........................ 53
2.4 Results and Discussion 54
2.4.1 ICUC Model Results.. 54
2.4.2 Correlations within Modeled Regions 58
2.4.3 COMET-Farm Results.. 62
2.4.4 Field to Market Fieldprint Calculator SCI Results 66
2.4.5 Applications of Model Results.. 66
2.5 Conclusions.. 70
References... 72
CHAPTER III: CONCLUSIONS AND FUTURE WORK... 75
3.1 Introduction.. 76
3.2 Research Findings.. 76
3.2.1 Conclusions for Objective 1: Model Development 76
3.2.2 Conclusions for Objective 2: Model Evaluation.. 77
3.3 Research Limitations 77
3.4 Research Implications.. 78
3.5 Future Work.. 79
VITA 80

LIST OF TABLES

Table 2.1. Data Sources. 40
Table 2.2. Silhouette Coefficient (SC) and Clusters (k) for each model output[1].............. 61
Table 2.3. Summary of COMET-Farm estimations[1] of soil carbon [tons CO_2-e ac^{-1} yr^{-1}] for the current baseline (base) scenario, future cover crop (CC) scenario, and change in sequestration (change) scenario for all fields in the dataset. 63
Table 2.4. Summary of COMET-Farm estimations[1] of average change in soil carbon [tons CO_2-e ac^{-1} yr^{-1}] per crop scenario and per cluster for the best-performing model (k=3, produced without crop frequency data layers). 63
Table 2.5. Summary of Field to Market Fieldprint Calculator Soil Conditioning Index (SCI) results for the sustainable management (best-case) scenario and non-sustainable management (worst-case) scenario. 67
Table 2.6. Summary of Field to Market Fieldprint Calculator Soil Conditioning Index (SCI) results per crop scenario and per cluster for the best-performing model (k=3, produced without crop frequency data layers) for the sustainable management (best-case) scenario and non-sustainable management (worst-case) scenario................... 67

LIST OF FIGURES

Figure 2.1. University of Tennessee AgResearch and Education Centers. 37
Figure 2.2. Crop frequency layers for (a) corn, (b) cotton, and (c) soybeans in Tennessee, 2008-2022. 41
Figure 2.3. Simplified crop frequency layers for (a) corn, (b) cotton, and (c) soybeans in Tennessee, 2008-2022... 42
Figure 2.4. Soil texture component percentage for (a) sand, (b) silt, and (c) clay, 2022.. 44
Figure 2.5. Average daily precipitation on annual basis, 2012-2022. 44
Figure 2.6. Average daily temperature on annual basis, 2012-2022. 46
Figure 2.7. Digital Elevation Model for Tennessee, 2022. 46
Figure 2.8. Milan REC and associated fields used in COMET-Farm and Field to Market scenarios for estimating soil organic carbon sequestration and Soil Conditioning Index values. 50
Figure 2.9. Jackson (West Tennessee) REC and associated fields used in COMET-Farm and Field to Market scenarios for estimating soil organic carbon sequestration and Soil Conditioning Index values.. 50
Figure 2.10. Ames REC and associated fields used in COMET-Farm and Field to Market scenarios for estimating soil organic carbon sequestration and Soil Conditioning Index values. 51
Figure 2.11. Private fields used in COMET-Farm and Field to Market scenarios for estimating soil organic carbon sequestration and Soil Conditioning Index values. . 51
Figure 2.12. ArcGIS Pro Iso Cluster Unsupervised Classification model outputs without crop input layers with the parameter for maximum number of classes set to (a) k=3, (b) k=5, (c) k=7, and (d) k=10. 55
Figure 2.13. ArcGIS Pro Iso Cluster Unsupervised Classification model outputs for corn with the parameter for maximum number of classes set to (a) k=3, (b) k=5, (c) k=7, and (d) k=10.. 56
Figure 2.14. ArcGIS Pro Iso Cluster Unsupervised Classification model outputs for cotton with the parameter for maximum number of classes set to (a) k=3, (b) k=5, (c) k=7, and (d) k=10. 57
Figure 2.15. ArcGIS Pro Iso Cluster Unsupervised Classification model outputs for soybeans with the parameter for maximum number of classes set to (a) k=3, (b) k=5, (c) k=7, and (d) k=10. 59
Figure 2.16. Box plot visualization of COMET-Farm estimations of carbon sequestration for the baseline (base), future cover crop (CC), and difference between base and CC (change) scenarios (n=125) for cotton (red), corn (yellow), and soybeans (green).. 64
Figure 2.17. Box plot visualization of COMET-Farm estimations of carbon sequestration for the difference between the baseline and cover crop scenarios (change) per cluster, as determined by the best-performing model. 65
Figure 2.18. Box plot visualizations of Soil Conditioning Index (SCI) values from the Field to Market Fieldprint Calculator for the sustainable management (best-case) scenario and non-sustainable management (worst-case) scenario. 68

Figure 2.19. Box plot visualization of Soil Conditioning Index (SCI) values from the Field to Market Fieldprint Calculator per cluster as determined by the best-performing model (k=3, produced without crop frequency data layers). 69

CHAPTER I:
INTRODUCTION AND LITERATURE REVIEW

1.1 Background

Atmospheric concentrations of greenhouse gases such as carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O) have increased by 50% since the Industrial Revolution (NASA, 2022; NOAA, 2022). Greenhouse gas concentrations in the atmosphere have increased largely as a result of burning coal and oil to produce energy for the transportation sector, electric power sector, and industrial sector (EPA, 2022). Livestock production, nitrogen fertilizers, and equipment powered by fossil fuels are key sources of greenhouse gas emissions produced by agricultural practices (EPA, 2022; Hayes et al., 2018; Hristov et al., 2018). Clearing land for agriculture, industry, residential land, and other human activities has also increased greenhouse gas concentrations, though to a lesser extent than the burning of fossil fuels and plant biomass (Lal, 2004; NASA, 2022). Once emitted, these greenhouse gases trap thermal energy in the earth's atmosphere and cause the surface temperature of the Earth to warm. If current levels of global warming persist, a global temperature increase of 1.5°C above pre-industrial levels could be exceeded in 2032 (Friedlingstein et al., 2022; Gillespie, 2022). Although greenhouse gas concentrations have been decreasing in the past decade, the amount of warming over the United States is still greater than the global average (USGRCP, 2023). The effects of this warming are not only occurring in the atmosphere but also in aquatic, forested, and agricultural ecosystems (NASA, 2022).

Production of agricultural commodities contributes over $300 billion to the United States' economy each year, and agricultural production is already being impacted by climate change (Melillo, Richmond, & Yohe, 2014). Warming temperatures, increasing greenhouse gas concentrations, and more frequent and intense extreme weather events are impacting water availability and the environmental conditions necessary for crop growth (USGRCP, 2023). These impacts also affect crop development, soil health, water availability, lengths of growing seasons, and crop sensitivity to climate stressors. The agricultural sector must regularly adapt to climate change by strategically altering crop rotations and management practices, or even by choosing new areas to utilize for crop production in response to climate stresses (Melillo, Richmond, and Yohe, 2014). The adoption of sustainable management practices can allow agricultural producers to limit their greenhouse gases emissions and increase crop resiliency to global warming and climate change while maintaining desirable crop yields (Hristov et al., 2018; Lal, 2015; USGRCP, 2023).

Agricultural land stores about 7.2 Gt of organic carbon in just the top 20 cm of soil, nearly one-third of the estimated 22.5 Gt of organic carbon stored across the contiguous United States (Zhu et al., 2011; Zhu & Reed, 2012). Due to the ability of agricultural ecosystems to sequester carbon from the atmosphere, utilize that carbon for plant growth, and store that carbon in plant matter and organic matter in the soil, agricultural soils are a critical component of the global carbon cycle (Hristov et al., 2018). Adopting conservation tillage or no-till practices to reduce soil erosion and carbon emissions, applying compost and livestock manure to utilize waste streams, breeding drought-resistant crop varieties, and implementing sustainable irrigation techniques are all approaches the agricultural sector is taking to restore soil organic carbon (SOC) and reduce carbon emissions in agricultural ecosystems (Lal, 2015). Cover crops can also be used to slow the rate of erosion from fields, improve soil health and water availability, and maintain nutrient content in soils between primary crop growing seasons (USDA, 2023). Cover crops have had lower adoption rates than reduced or no-tillage practices in the United States. It is estimated adopting cover cropping on the 88 Mha of cropland growing the five primary crops – corn, soybeans, wheat, rice, and cotton – not already using cover cropping could sequester upwards of 100 MMt CO_2-e each year (Fargione et al., 2018). Assessing carbon sequestration potential and the spatial distribution of carbon storage across agricultural ecosystems in the United States will further the understanding of how sustainable management practices can be used to influence agricultural SOC sequestration.

1.2 Agricultural Carbon Budgets

Carbon budgets in agricultural ecosystems are influenced by numerous environmental factors, including climate, weather conditions, soil characteristics, topography, the types of land management practices utilized, and the use of perennial vegetation or cover crops to improve soil health between primary crop growing seasons (Hristov et al., 2018; USGRCP, 2023). These factors that influence SOC sequestration also influence the growth of the crops within these agricultural ecosystems. Climate change introduces more variability in weather patterns, including the frequency of rainfall, fluctuating air temperatures, and an increased number of generally shorter duration yet greater intensity storm events (IPCC, 2023). Changes to the duration of adequate temperature and precipitation patterns necessary for crop growth can cause growing seasons to change. In order for farmers to continue to meet desired yields and maintain profit margins each year, it becomes increasingly necessary for farmers to provide irrigation, pest management, fertilization, and the use of sustainable management practices to limit erosion rates and protect soil health (Hristov et al., 2018).

Carbon cycling in agricultural systems can be measured in numerous ways: the amount of carbon in harvested plant biomass, the net primary production (NPP) of the agricultural region in terms of the net amount of carbon sequestered from the atmosphere into above-ground plant biomass, the net ecosystem exchange (NEE) of carbon between the agricultural ecosystem and the atmosphere through emissions or sequestration, and the total amount of SOC storage (Huang et al., 2009; X. Zhang et al., 2015). While these measurements can be taken and related to agricultural carbon budgets at the site scale, one of the major challenges the agricultural modeling community is currently facing is taking those site-scale observations and extrapolating them to regional-scale geospatial models. Most crop management data for tillage, fertilization, and irrigation practices are generally not spatially available at the regional scale, which is necessary in order to model the distribution of carbon across a given region in relation to management practices at larger scales.

Carbon budgets for agricultural ecosystems have been extensively estimated and modeled for the United States (US) Midwest (Grace & Robertson, 2021; Lu et al., 2018; Prince, Haskett, Steininger, Strand, & Wright, 2001; X. Zhang et al., 2015; Zhou et al., 2021), as the Midwest is a highly productive region that produces over 80% of the corn, 80% of the soybeans, and 50% of the wheat grown in the US each year (X. Zhang et al., 2015). The cropping systems most commonly modeled in the US are major field crops including corn, soybean, and wheat, as compared to limited studies in cotton cropping systems (He & Mostovoy, 2019). From 1990 to 2015, the total acreage utilized for corn and soybean production increased in the US, while acreage utilized for cotton and wheat production decreased (Hristov et al., 2018). Carbon budgets and climate change impacts have also been modeled for the Great Plains region (Kukal & Irmak, 2018) and the North Central region (Hollinger, Bernacchi, & Meyers, 2005). Compared to the aforementioned regions in the US, fewer studies have modeled agricultural carbon budgets and SOC sequestration in the Southeastern region.

1.3 Quantifying Carbon

Estimating and modeling agricultural carbon cycling through SOC change and land-atmosphere carbon exchange is a common technique used by researchers through the incorporation of inventory data, statistics, and data obtained via remote sensing. Taking on-site samples can assist with fine-tuning of parameters used in model development, but processing of samples increases time and costs substantially, depending on the scale of the site. Agricultural data that can be collected through remote sensing methods can include, but are not limited to: crop rotation patterns, air temperatures, precipitation patterns, soil types, and land use/land cover types. Remote sensing data can also be used to create empirical relationships for SOC change in response to land management practices using data from site-scale experiments. When using remote sensing data, there must be assumptions made based on the spatial and temporal resolution of the data, though remote sensing data helps to reduce the reliance on physical sampling across the entire region of interest.

Soil is a critical component of the global carbon cycle, as it can sequester carbon from the atmosphere, utilize it for plant growth, and store it in plant biomass and organic matter. To restore and increase SOC sequestration, many sustainable land management options are promoted by conservation efforts and legislation to reduce global carbon emissions. Widespread efforts to measure the SOC content and the spatial distribution of carbon across ecosystems can help to further the understanding of how sustainable management practices influence carbon budgets (Hristov et al., 2018). SOC storage is not stagnant; the rate of SOC accumulation is slow compared to the amount of time it takes for SOC sequestration to be reversed as a result of deforestation, wildfires, land use change, and erosion. It is crucial to be able to reliably measure and track existing SOC storage and its change over time in order to establish standards for baselines used in carbon credit systems known as carbon markets. Assessing current legislation and carbon market participation in the United States can help identify regions with under-allocated resources and increase funding for restoration efforts and implementation of sustainable management practices included in carbon markets, or carbon credit trading systems.

Measurement, reporting, and verification (MRV) techniques are needed to validate SOC sequestration and CO_2 emission reduction across different project scales. MRV efforts in carbon credit systems involve tracking specific measurement methods and their results, transparently reporting SOC information, and utilizing third-party verifiers to validate the accuracy of the data being collected and reported. It is becoming increasingly common for agricultural, forested, wetland, and stream ecosystems to utilize credit protocols to encourage sustainability and nature-based solutions for a multitude of ecosystem services (Altland et al., 2020; Canning et al., 2021; Oldfield et al., 2022). These services include local climate regulation, stormwater management, soil retention, nutrient regulation, pollutant removal, and efficient food and fiber production. In agricultural and forest ecosystems, optimizing biomass production not only helps to preserve existing soil carbon stocks and protect ecosystem health, but also influences the global carbon budget and helps countries work towards their international pledges for reducing greenhouse gas emissions as required by Intergovernmental Panel on Climate Change regulations (IPCC, 2023).

1.4 Techniques for Measuring Carbon Sequestration

Measurement techniques for SOC sequestration and carbon emissions are commonly categorized into four main approaches: direct, indirect, modeling, and hybrid (Oldfield et al., 2021). The subsequent synthesis of the measurement component of MRV efforts is organized into sections for each of these four measurement categories. The resources reviewed in this synthesis do not capture all of the significant contributions to this area of research, but rather were chosen to compare a variety of SOC measurement and estimation techniques that are currently being utilized in the development of MRV efforts.

1.4.1 Direct Measurement

To directly measure SOC sequestration, common techniques include soil coring at varying depths, analyzing the carbon composition in samples using laboratory equipment, and using portable sensors to collect data in the field. One study by Upson, Burgess, and Morison (2016) examined the effects of woodland, pasture, and silvopasture land use treatments on the soil bulk density and organic content on a 35-acre site in England. SOC content and bulk density measurements from across the site were compared. The silvopasture treatment had the highest bulk density, while the pasture treatment had the lowest bulk density. The pasture also had the greatest SOC sequestration at the surface level, though as the depth of samples increased, the SOC content across all three treatments were not significantly different from one another.

Near infrared reflectance (NIR) spectroscopy is another type of direct measurement approach that has been utilized to make predictions on soil properties at the field scale. Tang, Jones, and Minasny (2020) compared the use of four portable NIR sensors of different sizes and wavelength capacities on cropland in Southeastern Australia using two research-grade sensors and two miniature sensors. The NIR sensors collected three replicates of spectra for each sample to predict ten soil properties, including CEC, total carbon, and soil texture. Despite the differences in spectral wavelengths able to be detected by each NIR sensor, results of the soil property predictions showed comparable accuracy using both types of sensors. The Cubist regression model was used to calculate the coefficient of determination (R^2) for the calibration and validation datasets. The R^2 values for the four NIR sensors ranged from 0.73-0.81 for total carbon, 0.69-0.79 for CEC, 0.63-0.71 for sand content, and 0.69-0.79 for clay content. These results suggest advancements in the capabilities of direct in-field soil sampling with the use of handheld sensors to predict soil characteristics rather than solely using standard laboratory analyses.

1.4.2 Indirect Measurement

The eddy covariance (EC) method is commonly used to indirectly measure components of carbon budgets at the ecosystem scale. The EC method utilizes flux towers equipped to measure the net ecosystem exchange (NEE) of CO_2 entering and leaving a given ecosystem. The EC method is used to estimate gross primary production (GPP) in cropland and forests, which is a measure of the total uptake of CO_2 in the ecosystem through photosynthesis at night and can be estimated by subtracting the respiration rate in the daytime from the total NEE. One limitation of the EC method is that it only quantifies CO_2 and water vapor fluxes at relatively short timescales (Smith et al., 2010); however, the EC method is valuable for estimating NEE over large land areas where substantial direct measurement techniques may not be feasible.

Smith et al (2010) discussed the multitude of ways to measure aspects of the carbon budgets of agricultural ecosystems not directly measured when using the EC method. The net primary production (NPP) of an agricultural ecosystem can be estimated by adding the NPPs of the above-ground plant material that is not harvested (estimated from the proportion of above ground biomass to the harvested biomass), the harvested crop yield (estimated from harvested weight using known specific dry matter and carbon contents of the crop), and the below-ground plant material (estimated using either fixed or country-specific factors determined by soil coring or the use of rhizotrons). To estimate respiration from microbial communities, soil chambers and gas analyzers can be used at the site scale to measure CO_2 flux from the soil, and carbon content can then be determined using the mass fraction of carbon in CO_2. Using isotope detection methods can also help determine whether autotrophic or heterotrophic microbes are contributing to the soil respiration. The estimations of microbial respiration, NPP, and NEE are all components of a carbon budget, but these components do not account for long-term changes in carbon through crop harvesting, fertilizer application, and other land management practices that can affect cropland carbon cycling.

Another example of a method for indirectly measuring SOC sequestration is the use of aerial imagery of field sites (Gelder, Anex, Kaspar, Sauer, & Karlen, 2011). The range of SOC values for three agricultural sites in central Iowa were derived from the USDA-NRCS Soil Survey Geographic (SSURGO) database. SOC was estimated using the brightness value in each pixel, and these estimates were compared to SSURGO data for the region and ground truthing SOC measurements taken from soil coring and dry combustion. The correlations between the SSURGO data and the SOC estimates were between 0.60 and 0.82, suggesting the use of aerial images with SOC ranges supplemented by existing databases can assist with SOC analyses at the site-scale.

1.4.3 Modeling

Rather than taking field measurements, creating models of carbon budgets using machine learning techniques and remotely sensed data has become a novel way to predict and analyze ecosystem characteristics quickly as compared to the time required for direct measurement techniques. In a study by Padarian, Minasny, and McBratney (2019), a convolutional neural network (CNN) was trained using data from the Land Use and Coverage Area frame Survey (LUCAS) raw spectra database for soils in bare, cropland, grassland, and forest ecosystems across Europe. The multitasking CNN was able to make simultaneous predictions of six soil characteristics in one model: organic carbon, CEC, clay content, sand content, pH, and nitrogen content. When compared to two other machine learning techniques commonly used to predict soil properties from spectra, partial least squares regression and cubist regression trees, the multitask CNN significantly reduced the predictive error for all six soil properties examined.

In addition to soil spectroscopy, geospatial analysis of carbon cycling is another common modeling approach. Zhang et al (2015) compared two cropland carbon budget models for the US Midwest region: the Inventory Method suggested by West et al (2008) and the process-based Environmental Policy Integrated Climate (EPIC) model. EPIC was originally developed by the USDA for the Soil and Water Resources Conservation Act of 1980 to estimate erosion but has since been applied worldwide for modeling agricultural cropping systems and analyzing emission fluxes between ecosystems and the atmosphere. The results of the study showed similarities in the estimations of harvested carbon, NPP, NEE, and SOC sequestration by the Inventory Method and GAMS. When compared to inventory statistics, the GAMS estimations of crop yields were less accurate and subject to limitations based on the spatial resolution of the data, as well as assumptions made about crop management practices where management data was not spatially available. These limitations are frequently encountered in geospatial modeling.

1.4.4 Hybrid Approach

Some SOC measurement protocols rely on periodic direct soil sampling to assist with the development and validation of process-based models of carbon budgets. One hybrid approach compared NEE measurements by the EC method with soil carbon stock data taken over 13 years at the FluxNet site CH-Oe2 in Switzerland (Emmel et al., 2018). The Proof of Ecological Performance framework began managing this cropland in 2003, requiring wheat and cover crop rotations, efficient use of fertilizers and pesticides, and validation of neutral nitrogen and phosphorus budgets from the site. Cropland carbon budgets were estimated by incorporating the EC method, direct SOC stock measurements in 2004 and 2017, and land management data. Statistical analyses were performed to determine whether there were significant differences in soil carbon, soil nitrogen, and bulk density between the two sampling years and in the depths of the soil samples. Rather than increasing the amount of SOC sequestered, the results showed a net loss of carbon over the 13-year study period. However, the Proof of Ecological Performance framework limited the amount of carbon that would have been lost as compared to what would have been lost under a bare field scenario modeled using the Soil-Plant-Atmosphere Crop Model (Sus et al., 2010).

In an effort to assess the changes in SOC storage within tropical pastures created by deforestation in the Amazon basin, a similar hybrid approach used the eddy covariance (EC) method to compare SOC inventory data from tropical pastures with NEE measurements (Stahl et al., 2017). This study only considered pastures that were managed without the use of fires for pasture burning to reduce dense vegetation, a common management technique used in Amazonia. After comparing the soil characteristics to the NEE estimations, this study confirmed that the sustainable management techniques applied to this tropical pasture – avoiding overgrazing of livestock, rotational grazing, and grassland biodiversity – partially restored the SOC that had been lost through deforestation.

1.5 Reporting and Verifying Carbon Sequestration

Hybrid approaches are currently being used by multiple carbon market registries to quantify greenhouse gas emissions and simulate carbon and nitrogen biogeochemistry in agricultural ecosystems. Many carbon market protocols utilize direct measurements from on-site soil samples to help determine appropriate parameters for process-based models. These models are used to estimate the productivity and sequestration potential for the baseline, or business-as-usual, scenario for a given site for a carbon credit project. The baseline sequestration can then be compared to estimates for additional sequestration obtained throughout the life of the project. Examples of such process-based models are the DayCent model, the DeNitrification-DeComposition (DNDC) model, the System Approach to Land Use Sustainability (SALUS) model, and the Ecosys model (USDA, 2023). These process-based models incorporate remote sensing data and publicly-available data to reduce dependency on physical soil samples and to lessen the amount of management or site-specific historical data needed from landowners and producers. The USDA has invested in improving measurement techniques and long-term monitoring of SOC and carbon emissions from agricultural land to further improve the amount of geospatial data available for use in process-based models. With more data comes more opportunities to ground-truth these models and validate their estimations of SOC changes in agricultural ecosystems after implementing sustainable management practices. Better validation of modeling estimates also provides more opportunities to validate the potential volume of credits that could be contributed to carbon markets at both site-specific and regional scales.

As shifts in carbon budgets and greenhouse gas emissions from management practices are estimated and verified, policy-makers can enact incentive programs through carbon markets to encourage corporations to reduce or offset their greenhouse gas emissions. Carbon markets also encourage farmers to increase carbon sequestration into their fields by assigning a monetary value to carbon sequestered through sustainable management practices. By the United Nations' standards, one tradable carbon credit equals one metric ton of carbon dioxide, or the equivalent amount of a different greenhouse gas, that is reduced, sequestered, or avoided (United Nations Development Programme, 2022). The price of carbon credits can differ by region, by country, or by the third-party entity that is providing the funding to compensate farmers and landowners for implementing sustainable practices.

It is important to be able to quantify how much of an impact site-scale and regional-scale cropland carbon cycling can have towards sequestering atmospheric CO_2 and to understand how this impact can vary spatially across different agricultural regions, not only from an ecological perspective but also for use in allocating a monetary value to sequestered carbon. That said, there is currently no universally-accepted standard for allocating monetary values within carbon markets. MRV efforts for carbon credit systems also vary between regulating bodies, third-party verifiers, regions, and countries. This presents the question of how to scale estimations of carbon sequestration for entire regions, since it is often not economically or logistically feasible to take measurements for every acre of land enrolled in a carbon market contract, while also incentivizing carbon credit trading in the industrial and agricultural sectors. Carbon credit systems may have different methods of accounting for additionality (the emission reduction and SOC sequestration achieved by the project are distinct from the baseline scenario), permanence (the SOC stock is maintained over an extended time period after the life of the project contract), leakage (reduced emissions under the crediting program are not offset by increased emissions elsewhere), and reversals (potential losses of SOC sequestration due to unforeseen climate impacts) (Oldfield et al., 2022). Differences in MRV protocols, including how a given carbon market may account for defining and proving additionality, permanence, leakage, and reversals, can potentially create credits that are not equivalent between markets (USDA, 2023). There is a need for consistency across carbon market systems to ensure the integrity of credits and MRV efforts (Oldfield et al., 2022). There is also a need to generate and trade carbon credits that have equivalent value between countries, so carbon markets are not constrained to given countries and can move towards standardization.

1.6 MRV Efforts for Carbon Markets in the United States

Current techniques for reporting and verifying SOC sequestration vary by region and by country. Countries participating in the Paris Climate Agreement are required to establish standards for MRV efforts based on governmental regulations in public markets in order to meet their annual sustainable development goals and nationally determined contributions. In the United States, many private carbon markets exist. Third-party companies and carbon registries act as intermediaries, working with farmers and landowners to develop contracts for compensation and for proving additionality on a case-by-case basis. The Growing Climate Solutions Act (GCSA) passed the US Senate in June 2021 and was signed into law as part of the Consolidated Appropriations Act in December 2022 (S. 1251 - 117th Congress, 2021-2022). The GCSA authorized the USDA to establish the Greenhouse Gas Technical Assistance Provider and Third-Party Verifier Certification Program (Program) in the interest of reducing barriers for farmers and to move towards standardizing verification efforts. The Program is designed to provide technical assistance from qualified entities to help increase participation in voluntary carbon markets. As required by the GCSA, the USDA must publish a list of protocols and qualifications to inform eligible entities how to assist farmers, ranchers, and private forest landowners with implementing sustainable practices and gaining compensation from markets. The USDA published its first deliverable for the GCSA in October 2023, which provided a comprehensive assessment of carbon market activity in the US, barriers to producer participation in carbon markets, and suggestions for strategies to reduce the barriers identified (USDA, 2023).

The Verified Carbon Standard and American Carbon Registry are two of the largest third-party carbon registry programs in the US (USDA, 2023). Carbon registries set standards for the MRV of sequestration, emissions, and the valuation of carbon credits. The buyers in the market may purchase carbon credits once the credits have been reported and verified. The buyers are often large corporations desiring to offset their carbon dioxide emissions to meet state regulations or voluntary sustainability pledges. These can be used as carbon offsets, since the companies are investing in sustainable or regenerative practices which encourage sequestration, or as carbon insets, if the practices being funded through the purchase of carbon credits are within the value chain of the company.

California has a compliance carbon market, often referred to as a cap-and-trade program, enforced by the California Air Resources Board (ARB). This cap-and-trade program is the first multi-sector program of its kind in North America (Center for Climate and Energy Solutions, 2021). California's program regulates electric utilities, industrial facilities, and fuel distributors on emissions of multiple greenhouse gases (California EPA

ARB, 2015). Companies who produce more emissions than their allowable cap can trade emission credits with companies who are producing emissions below their cap at quarterly credit auctions. In this way, a net maximum amount of allowable carbon emissions can be restricted as companies sell and trade credits while funding sustainable land management and energy production practices, which results in a net reduction of carbon emissions over time.

In the US, carbon market regulations have not yet become universal, and states can determine the level of regulation for markets within their jurisdiction, though the steps put in place by the GCSA and the USDA will help towards standardizing carbon markets MRV efforts in the US. In addition to California, some other states have implemented programs to move towards compliance markets. Washington has a cap-and-invest program, and the Regional Greenhouse Gas Initiative includes eleven states in the Northeast who have pledged to put a joint cap on emissions from the power sector (Center for Climate and Energy Solutions, 2021). That said, there are currently no widespread cap-and-trade programs or carbon market systems spanning the entire US. A regional approach to MRV of carbon emission reductions and SOC sequestration, as suggested by Oldfield et al (2022), would enable policy changes at a larger scale to create regional frameworks based on similar land classifications across the US which could then be related to these similar land classifications in other countries. Standardization across these similar regions could be achieved by providing regional standards for the carbon market registries and project developers to meet. Utilizing process-based models and geospatial data to assist with MRV efforts for carbon markets at the regional scale could improve confidence in SOC estimates as compared to estimating SOC changes for specific projects at the site-scale (Oldfield et al., 2022).

1.7 Geospatial Regional Modeling Approach for Carbon Market MRV

The verification component of MRV efforts often requires direct sampling over the life of the contract, though many verification programs also utilize empirical models to estimate carbon stock baselines and effects on greenhouse gas emissions after the implementation of sustainable management practices. Examples of such models include the DayCent model, the DeNitrification-DeComposition (DNDC) model, and the System Approach to Land Use Sustainability (SALUS) model (USDA, 2023). Direct sampling is time intensive and expensive; costs increase with the scale of the project and the diversity of practices being used. Verification approaches based only on direct sampling often take longer to generate credits, anywhere between three to five years, while model-based estimates avoid the need for intensive direct sampling and can issue credits faster, within the same year (Oldfield et al., 2022). There is some inherent uncertainty with model estimates which can be influenced by the precision of the model inputs and scale. For these reasons, modeling is commonly supplemented with representative soil and greenhouse gas sampling to verify model results. It would be advantageous to utilize existing soil monitoring programs and networks of agricultural research sites for this representative sampling across entire regions and across entire countries in the future, since these existing research sites are often publicly-funded and provide unbiased, historical data to better estimate SOC stocks over extended periods of time.

1.7.1 Overview of Clustering Techniques

In recent decades, machine learning has become a popular tool to accurately model and predict phenomena across diverse fields of research and for data spanning across varying spatial scales. Machine learning is a useful method for creating models like the ones used in the measurement and verification of carbon emissions and sequestration for carbon market systems, due to the amount of data required to develop these process-based models. Clustering models are one common and effective machine learning technique used to understand correlations in geospatial data, as clustering can handle numerous input features and can output clusters of data based on similar characteristics and patterns in the input data. In the initialization step of a clustering model, the cluster means may be arbitrarily chosen by an algorithm or may be defined by the user. Each data point is then assigned to a cluster based on a set of criteria as applicable for the research purpose. The minimum distance between a data point and the nearest cluster mean is a commonly used method for assigning points in the input dataset to the clusters. This assignment process is repeated for each input data point in the dataset. Another common initialization technique is to define the number of clusters desired in the model output and to arbitrarily assign all data points to clusters. The mean of each cluster is then calculated, the distances from each data point in a given cluster to the cluster mean are measured, and the location of each mean is updated. This process is then repeated until the assignment of all data points within clusters is no longer changing or until some other stopping criteria is met, such as the total number of iterations run by the model.

There are two types of clustering techniques, supervised and unsupervised. Supervised machine learning utilizes information from both the training (or input) dataset and the testing (or target) dataset. The attributes for the test dataset must already be known, and these attributes are used to train the clustering model to accurately predict target characteristics when presented with training or validation data. The ratio of training to testing data is commonly split unevenly, with the size of the training set greater than the testing set, to ensure the model is trained on enough data to be able to make accurate predictions when presented with new data not used to develop the model. Alternatively, unsupervised machine learning does not utilize information from the testing dataset, and instead analyzes statistical patterns between attributes in the training dataset to predict their target characteristics for the model output. Unsupervised machine learning techniques are applicable for research purposes where testing data is not available. Ground-truth data can then be collected and used to validate unsupervised model results.

When working with geospatial data, clustering techniques can be used to study and compare spatial and temporal characteristics of given spatial input layers. Among other environmental phenomena, clustering algorithms have been used to assess and predict land use/land cover changes (Lv, Liu, Shi, Benediktsson, & Du, 2019; Navin & Agilandeeswari, 2019; Nijhawan, Srivastava, & Shukla, 2017; Y. Zhang & Zhao, 2020), flood vulnerability (Ahmed & Kranthi, 2018; Feloni, Mousadis, & Baltas, 2020; Fernandez, Mourato, Moreira, & Pereira, 2016; Lin et al., 2020; Röthlisberger, Zischg, & Keiler, 2017), and deforestation trends (Carvalho, Adami, Amaral, Bezerra, & de Aguiar, 2019; Duveiller, Defourny, Desclée, & Mayaux, 2008; Harris et al., 2017; Sathler, Adamo, & Lima, 2018). Using environmental data layers as inputs into a clustering model is a potentially suitable technique for developing a model of geospatially-correlated regions based on geospatial data for use in MRV efforts.

1.7.2 Iterative Self-Organized Clustering and Unsupervised Classification

The Spatial Analyst toolbox in Environmental Systems Research Institute (ESRI) ArcGIS Pro software contains clustering tools for use in analyzing correlations and relationships between geospatial data from multiple rasters. The Iterative Self-Organizing (ISO) Cluster tool is most commonly used to prepare for unsupervised classification of raster data, and unsupervised clustering of multivariate data is performed based on the input raster bands provided using an iterative clustering approach known as migrating means or k-means (ESRI, 2024a). In addition to the desired input rasters, the following input parameters must be specified: the number of classes into which to group the data, the minimum class size to define the minimum number of cells in each class, and the sample interval to define the number of cells in each n-by-n block in the input rasters to analyze when determining the clusters (ESRI, 2024c). Since the optimal number of classes to group the cells into is typically not known, a conservatively high number should be initially specified in the input parameter and the ISO Cluster tool should be rerun, then rerun as needed while reducing the input number of classes, and the outputs compared. The ISO Cluster tool outputs a signature file that can then be utilized in ArcGIS Pro classification tools to produce a classified raster using unsupervised classification.

The Maximum Likelihood Classification (MLC) tool in the Spatial Analyst toolbox is one such classification tool that can create a classified raster using the signature file output from the ISO Cluster tool (ESRI, 2024b). MLC assumes the cells within each class are normally distributed and follows Bayes Theorem to use the mean vector and covariance matrix of each class to assign each cell to one of the classes represented in the signature file, based on the statistical probability for each class and the specified A priori probability weighting. For unsupervised classification problems, the A priori probability weighting for each class is often not known, as there are not testing data to use to validate the classification results, and in some cases there is not prior knowledge about the geospatial area to be classified to assist with interpreting classification results. When the A priori probability weighting is not known, the default Equal weighting can be used, which allows each cell to be assigned to the class to which it has the highest likelihood of being a member (ESRI, 2024b).

The ISO Cluster Unsupervised Classification (ICUC) tool utilizes the ISO Cluster and MLC tools directly. The ICUC tool performs unsupervised ISO clustering on the multivariate input rasters, uses the resulting signature file to perform MLC on the cells in the clustered raster, and then outputs a classified raster. The convention used to provide a numerical value for each class in the output raster is arbitrary; the range of class identification values starts at one and increases as additional clusters are identified. The ICUC tool has been used to classify rasters containing numerous types of remotely-sensed data, including but not limited to: land cover types using Landsat imagery (Lemenkova, 2021), trees on a palm oil plantation using LiDAR imagery (Norzaki & Tahar, 2019), benthic habitat mapping using bathymetric and backscatter maps (Calvert, Strong, Service, McGonigle, & Quinn, 2014), deforestation trends using Landsat imagery (Duveiller et al., 2008), and soil fertility prediction in agricultural fields (Gulhane, Rode, & Pande, 2023). The ICUC tool is a potentially suitable machine learning technique for analyzing and visualizing geospatial correlations between multivariate inputs for the development of a regional model for use in carbon market MRV efforts.

1.7.3 Clustering Model of Geospatially-Correlated Regions for MRV Efforts

Spatial data for crop management techniques such as tillage, irrigation, and fertilizer application are not widely available for use in developing a regional geospatial model, but utilizing other spatial data for factors that influence SOC sequestration from publicly-available datasets would allow for standardization when developing models for entire regions in the US. There are nearly 700 publicly-funded research and education centers (RECs) operated at land grant universities and by the USDA Agricultural Research Service (ARS). Each REC studies a multitude of research objectives, including ways in which SOC sequestration and greenhouse gas emissions are affected by sustainable management practices. Due to the widespread diversity of the research and historical site-specific data that can be collected at these publicly-funded RECs across the country, it would be advantageous to gain insight into how the locations of the RECs may relate to geospatial similarities and conditions found in agricultural land. If the potential compensation for additionality from a carbon credit project implemented on a plot of land at a REC was estimated, based on historical SOC data under various sustainable land management practices, these estimations could be used to assist farmers within the REC's spatially-correlated region with estimating the amount of SOC sequestration they could obtain and the profit they could earn over time after enrolling in a carbon market. Farmers could compare the SOC sequestration potential of their farm to historical SOC sequestration trends at their most correlated REC and with RECs in other correlated regions. A clustering model that produces maps for visualization of clusters representing geospatially-correlated regions, based on environmental conditions and crop production trends, would assist farmers with identifying their most correlated REC(s) and with estimating the SOC sequestration potential of their farmland. Farmers then might be more inclined to enroll in a carbon market and report the data required by the contract and third-party verifiers to prove SOC additionality throughout the life of the contract. Data from publicly-funded RECs could help improve the quality of direct and indirect measurements used to validate model estimates of SOC sequestration potential and emission reduction potential. RECs could be used as model farms for carbon credit programs when developing and updating their standards, and this could allow for more reliable estimates of baseline SOC sequestration levels within a given region. Using data from RECs could also allow for more reliable estimates of compensation for farmers enrolled in carbon credit programs, as these estimations could be based on the historical trends in SOC storage observed at research plots on the research stations.

1.8 Model Evaluation Metrics

1.8.1 Silhouette Coefficient

The silhouette coefficient was first proposed by Rousseeuw (1987) for comparing the within-cluster and between-cluster dissimilarities, also referred to as the intra- and inter-cluster distances. The values for the silhouette coefficient range from [-1,1]. When the within-cluster dissimilarity is much smaller than the smallest between-cluster dissimilarity, the silhouette coefficient s(*i*) is close to 1, and it is assumed the sample *i* within the cluster is appropriately assigned. A silhouette coefficient of 0 indicates the sample *i* is close to the decision boundary between clusters A and B. A silhouette coefficient less than 0 indicates the sample likely was assigned to the wrong cluster, since the between-cluster dissimilarity is less than the within-cluster dissimilarity for that sample. Silhouette coefficients can be calculated for each model output and compared to determine the model with the best representation of the correlations between the input data layers. The silhouette coefficient metric has been used to assess the performance of clustering models on various agricultural research applications: zoning corn based on corn canopy characteristics (Di & Wang, 2023), determining soybean fertilization zones (Chen et al., 2023), characterizing management zones based on soil properties (Yuan et al., 2022), and identifying farm susceptibility to waterlogging using remote sensing data (Bukombe, Csenki, Szlatenyi, Czako, & Láng, 2023). The silhouette coefficient can give an indication of how well the clustering model differentiated the data into groups based on cluster characteristics, and it is a useful performance metric for assessing the performance of unsupervised models.

1.8.2 COMET

The Carbon Management and Emissions Tool (COMET) was developed by the USDA NRCS and Colorado State University to provide estimations of SOC sequestration and greenhouse gas emissions over a 10-year period from an agricultural production field. COMET-Farm provides estimates at the site-scale while COMET-Planner, a simplified version, provides estimates on a county scale. COMET is already being used by current carbon market registries to estimate sequestration rates for their contracts with farmers. The COMET-Farm, COMET-Planner, and associated tools are resources for supporting MRV efforts. The COMET tools give farmers and policymakers a quantitative way to estimate SOC sequestration potential of recent or planned changes to field management practices. Baseline and future management scenarios can be modeled in COMET-Farm Version 4.1 (USDA NRCS and Colorado State University, 2023) for private farms in Tennessee and for farms at University of Tennessee RECs. The results from the COMET-Farm scenarios can then be compared for the farms within the geospatially-correlated regions determined by the model to assess the model's ability to accurately represent trends based on the environmental input layers.

1.8.3 Soil Conditioning Index

The Field to Market (FTM) Fieldprint Calculator allows users to input a delineation of a field, along with information about crop production and management practices used at that site. The FTM platform utilizes COMET-Planner to determine estimations of SOC sequestration based on the delineated farm and the specified input parameters for fertilizer use, tillage practices, soil conservation practices, and organic matter inputs to the soil. The Soil Carbon Metric in the Fieldprint Calculator then calculates the Soil Conditioning Index (SCI) to estimate a directional measure of SOC sequestration for the delineated area. The SCI was first published by Wayne Austin in 1964 focusing on the Western US; a national version of the SCI was introduced in 2002 as an Excel-based spreadsheet (USDA NRCS, 2002). The Fieldprint Calculator utilizes the USDA Water Erosion Prediction Project (WEPP) Model v2.0, USDA Wind Erosion Prediction System (WEPS) Simulation v5.0, and USDA SSURGO 2.0 to compile data for estimating the erosion rates and soil properties needed to calculate the SCI.

The Fieldprint Calculator estimates SCI values ranging from [-1,1] based on the effects of organic matter returned to the soil, field operations, and erosion over a ten-year period. The SCI value associated with each scenario is qualitative; the magnitude represents directionality, where a positive value indicates the amount of SOC in the field is increasing and a negative value indicates SOC is being depleted. The SCI has been used to relate SOC sequestration data from published studies in various regions of the US, including the Midwest and Southeast (Franzluebbers, Causarano, & Norfleet, 2011) and the Southern High Plains (Zobeck et al., 2007). Another study by Komé, Andrews, and Franzluebbers (2013) compared the outputs from the COMET tool for voluntary reporting of greenhouse gases (COMET-VR) and SCI for 18 sites across the US, each with varying soil characteristics, tillage systems, and cropping systems. This study found the SCI scores and COMET-VR predictions for SOC change were positively related.

1.9 Research Objectives

1.9.1 Overall Research Objective

The overarching objective of this research is to create a geospatial model of carbon sequestration potential of agricultural land in Tennessee for future implementation in US carbon markets. This objective will be accomplished by (1) developing a geospatial model to relate SOC sequestration potential of private agricultural land to data collected at publicly-funded University of Tennessee Institute of Agriculture (UTIA) RECs, and (2) evaluating the potential of the geospatial model to relate SOC sequestration at UTIA RECs to spatially-correlated agricultural regions across Tennessee.

1.9.2 Development of Geospatial Model

Using a regional approach, the regional unit for quantification of SOC sequestration was made up of biophysically- and geospatially-defined agroecological regions based on correlated soils, climate trends, and crop production, as suggested by Oldfield et al (2022). The model was developed using ESRI ArcGIS Pro software and input layers containing spatial data for environmental factors influencing SOC sequestration. Model outputs were compared for corn, cotton, and soybeans to better understand how cropping system might impact the geospatially-correlated regions. Carbon credits could be quantified using the model output for a given crop type once the model estimates have been verified with data collected from a network of research sites and participating farms within the geospatially-correlated regions for the given crop. This modeling approach could also be used to study sustainable management scenarios specific to a given region and to compare management scenarios across multiple regions. The delineation of the agroecological regions could account for variation in climate and soil characteristics that can affect crop productivity and SOC sequestration potential (Oldfield et al., 2022) based on the data provided in the model input layers.

1.9.3 Evaluation of Geospatial Model

The silhouette coefficient defined by Rousseeuw (1987) is a common way to assess the results of data clustered using a k-means algorithm like that of the ICUC tool in ArcGIS Pro. The class labels in the ICUC model output were used to calculate the silhouette coefficient for the model to assess its ability to recognize correlations between the input layers and generate class assignments to characterize regions with similar soil and climate conditions. The silhouette coefficient was calculated for each model to compare the outputs for each crop type. Similarities between estimated SOC sequestration potential and SCI values for farms located at RECs and for farms located in the same geospatially-correlated regions can also be used to evaluate the correlations of the clusters in the model outputs. COMET-Farm is a tool used by current US carbon market registries to estimate SOC sequestration rates for their contracts with farmers. Estimates of SOC sequestration potential for baseline management scenarios and future scenarios were calculated, depending on the sustainable management practice(s) specified in the COMET-Farm input parameters. SCI values were calculated for the same baseline and future scenarios using the Fieldprint Calculator in the Field to Market Fieldprint Platform. The SCI considers the effects of organic matter returned to the soil, field operations, and erosion on net gains and losses of SOC sequestration for a given site.

If the estimated SOC sequestration and SCI data for a farm and its spatially-correlated REC, determined by the model output, are similar, then the data at this REC could be used to accurately estimate the SOC sequestration potential of the farm if it were to adopt the sustainable management practices researched at that REC. It was anticipated that farms in spatially-correlated regions would have similar SOC values as their correlated REC(s) if they have the same production and management practices used over similar periods of record as compared to the historical data from the REC(s). An opportunity exists to use geospatial modeling to better understand how the environmental conditions observed at the RECs can be related to their surrounding areas. This geospatial modeling approach would assist farmers with understanding how research plots at their most correlated REC(s) might be compensated in a carbon credit contract. The historical, non-biased data from the RECs could also be used to increase transparency around carbon credit verification and to further the development of carbon market MRV standards in the US.

References

AgResearch and Education Center at Milan. (2024). 33rd Milan No-Till Crop Production Field Day. University of Tennessee Institute of Agriculture. https://milannotill.tennessee.edu/wp-content/uploads/sites/100/2024/06/Final-2024_0620-Milan-No-TIll-Program.pdf.

Ahmed, C. F., & Kranthi, N. (2018). Flood Vulnerability Assessment using Geospatial Techniques: Chennai, India. *Indian Journal of Science and Technology, 11*(6), 1-13. doi:https://doi.org/10.17485/ijst/2018/v11i6/110831

Altland, D., Becraft, C., Berg, J., Brown, T., Burch, J., Clearwater, D., . . . Geratz, J. (2020). Consensus recommendations to improve protocols 2 and 3 for defining stream restoration pollutant removal credits. *Report to the Water Quality Goal Implementation Team of the Chesapeake Bay Program*. Retrieved from https://chesapeakestormwater.net/wp-content/uploads/dlm_uploads/2022/07/10032-3.pdf

Boyer, C., Cavasos, K., Smith, A., & Walker, F. (2023). *2023 Tennessee Nutrient Management Report*. Retrieved from University of Tennessee Department of Agricultural and Resource Economics and Tennessee Department of Agriculture: https://www.tn.gov/content/dam/tn/environment/water/nutrient-management/ntf/wr_ntf_2023-nutrient-management-report-ut-ag.pdf

Bukombe, B., Csenki, S., Szlatenyi, D., Czako, I., & Láng, V. (2023). Integrating Remote Sensing, Proximal Sensing, and Probabilistic Modeling to Support Agricultural Project Planning and Decision-Making for Waterlogged Fields. *Water, 15*(7), 1340. Retrieved from https://www.mdpi.com/2073-4441/15/7/1340

California EPA ARB. (2015). Overview of ARB Emissions Trading Program. Retrieved from https://ww2.arb.ca.gov/sites/default/files/cap-and-trade/guidance/cap_trade_overview.pdf

Calvert, J., Strong, J. A., Service, M., McGonigle, C., & Quinn, R. (2014). An evaluation of supervised and unsupervised classification techniques for marine benthic habitat mapping using multibeam echosounder data. *ICES Journal of Marine Science, 72*(5), 1498-1513. doi:https://doi.org/10.1093/icesjms/fsu223

Canning, A. D., Jarvis, D., Costanza, R., Hasan, S., Smart, J. C., Finisdore, J., . . . Beck, M. W. (2021). Financial incentives for large-scale wetland restoration: Beyond markets to common asset trusts. *One Earth, 4*(7), 937-950. doi:https://doi.org/10.1016/j.oneear.2021.06.006

Carvalho, R., Adami, M., Amaral, S., Bezerra, F. G., & de Aguiar, A. P. D. (2019). Changes in secondary vegetation dynamics in a context of decreasing deforestation rates in Pará, Brazilian Amazon. *Applied Geography, 106*, 40-49. doi:https://doi.org/10.1016/j.apgeog.2019.03.001

Center for Climate and Energy Solutions. (2021). Market Based State Policy. *Carbon Pricing Policy Hub*. Retrieved from https://www.c2es.org/content/market-based-state-policy/

Chen, H., Wang, X., Zhang, W., Wang, X. Z., Di, X. D., & Qi, L. Q. (2023). APPLYING NDVI FROM DIFFERENT NUMBER OF SPECTRAL SENSORS IN DELIMITING SOYBEAN FERTILIZATION MANAGEMENT ZONES.

APPLIED ECOLOGY AND ENVIRONMENTAL RESEARCH, 21(6), 5175-5187. doi:https://doi.org/10.15666/aeer/2106_51755187
Di, X., & Wang, X. (2023). Method for Zoning Corn Based on the NDVI and the Improved SOM-K-Means Algorithm. *Journal of the ASABE, 66*(4), 943-953. doi:https://doi.org/10.13031/ja.15081
Duveiller, G., Defourny, P., Desclée, B., & Mayaux, P. (2008). Deforestation in Central Africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed Landsat extracts. *Remote Sensing of Environment, 112*(5), 1969-1981. doi:https://doi.org/10.1016/j.rse.2007.07.026
Emmel, C., Winkler, A., Hörtnagl, L., Revill, A., Ammann, C., D'Odorico, P., . . . Eugster, W. (2018). Integrated management of a Swiss cropland is not sufficient to preserve its soil carbon pool in the long term. *Biogeosciences, 15*(17), 5377-5393. doi:https://doi.org/10.5194/bg-15-5377-2018
EPA. (2022). Sources of Greenhouse Gas Emissions. Retrieved from https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions
ESRI. (2024a). ArcGIS Pro Tool Reference: How Iso Cluster Works. Retrieved from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-iso-cluster-works.htm
ESRI. (2024b). ArcGIS Pro Tool Reference: How Maximum Likelihood Classification works. Retrieved from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-maximum-likelihood-classification-works.htm
ESRI. (2024c). ArcGIS Pro Tool Reference: Iso Cluster (Spatial Analyst). Retrieved from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/iso-cluster.htm
Fargione, J. E., Bassett, S., Boucher, T., Bridgham, S. D., Conant, R. T., Cook-Patton, S. C., . . . Griscom, B. W. (2018). Natural climate solutions for the United States. *Science Advances, 4*(11), eaat1869. doi:https://doi.org/10.1126/sciadv.aat1869
Feloni, E., Mousadis, I., & Baltas, E. (2020). Flood vulnerability assessment using a GIS-based multi-criteria approach—The case of Attica region. *Journal of Flood Risk Management, 13*(S1), e12563. doi:https://doi.org/10.1111/jfr3.12563
Fernandez, P., Mourato, S., Moreira, M., & Pereira, L. (2016). A new approach for computing a flood vulnerability index using cluster analysis. *Physics and Chemistry of the Earth, Parts A/B/C, 94*, 47-55. doi:https://doi.org/10.1016/j.pce.2016.04.003
Franzluebbers, A. J., Causarano, H. J., & Norfleet, M. L. (2011). Soil conditioning index and soil organic carbon in the Midwest and southeastern United States. *Journal of Soil and Water Conservation, 66*(3), 178-182. doi:https://doi.org/10.2489/jswc.66.3.178
Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., . . . Zheng, B. (2022). Global Carbon Budget 2022. *Earth Syst. Sci. Data, 14*(11), 4811-4900. doi:https://doi.org/10.5194/essd-14-4811-2022
Gelder, B. K., Anex, R. P., Kaspar, T. C., Sauer, T. J., & Karlen, D. L. (2011). Estimating Soil Organic Carbon in Central Iowa Using Aerial Imagery and Soil Surveys. *Soil*

Science Society of America Journal, 75(5), 1821-1828. doi:https://doi.org/10.2136/sssaj2010.0260
Gillespie, A. (2022). Global atmospheric carbon dioxide levels continue to rise. Retrieved from https://research.noaa.gov/2022/11/15/no-sign-of-significant-decrease-in-global-co2-emissions/
Grace, P., & Robertson, G. P. (2021). Soil carbon sequestration potential and the identification of hotspots in the eastern Corn Belt of the United States. *Soil Science Society of America Journal, 85*(5), 1410-1424. doi:https://doi.org/10.1002/saj2.20273
Gulhane, V. A., Rode, S. V., & Pande, C. B. (2023). Correlation Analysis of Soil Nutrients and Prediction Model Through ISO Cluster Unsupervised Classification with Multispectral Data. *Multimedia Tools and Applications, 82*(2), 2165-2184. doi:https://doi.org/10.1007/s11042-022-13276-2
Harris, N., Goldman, E., Gabris, C., Nordling, J., Minnemeyer, S., Ansari, S., . . . Potapov, P. (2017). Using spatial statistics to identify emerging hot spots of forest loss. *Environmental Research Letters, 12*(2), 024012. doi:https://doi.org/10.1088/1748-9326/aa5a2f
Hayes, D. J., Vargas, R., Alin, S. R., Conant, R. T., Hutyra, L. R., Jacobson, A. R., . . . Woodall, C. W. (2018). Chapter 2: The North American carbon budget. *Second State of the Carbon Cycle Report (SOCCR2): A Sustained Assessment Report,* U.S. Global Change Research Program, Washington, DC, USA, pp. 71-108. doi:https://doi.org/10.7930/SOCCR2.2018.Ch2
He, L., & Mostovoy, G. (2019). Cotton yield estimate using Sentinel-2 data and an ecosystem model over the southern US. *Remote Sensing, 11*(17), 2000. doi:https://doi.org/10.3390/rs11172000
Hollinger, S. E., Bernacchi, C. J., & Meyers, T. P. (2005). Carbon budget of mature no-till ecosystem in North Central Region of the United States. *Agricultural and Forest Meteorology, 130*(1), 59-69. doi:https://doi.org/10.1016/j.agrformet.2005.01.005
Hristov, A. N., Johnson, J. M. F., Rice, C. W., Brown, M. E., Conant, R. T., Del Grosso, S. J., . . . Shrestha, G. (2018). Chapter 5: Agriculture. *Second State of the Carbon Cycle Report (SOCCR2): A Sustained Assessment Report*, U.S. Global Change Research Program, Washington, DC, USA, pp. 229-263. doi:https://doi.org/10.7930/SOCCR2.2018.Ch5.
Huang, Y., Yu, Y., Zhang, W., Sun, W., Liu, S., Jiang, J., . . . Yang, Z. (2009). Agro-C: A biogeophysical model for simulating the carbon budget of agroecosystems. *Agricultural and Forest Meteorology, 149*(1), 106-129. doi:https://doi.org/10.1016/j.agrformet.2008.07.013
IPCC. (2023). *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Core Writing Team, H. Lee and J. Romero (eds.). IPCC, Geneva, Switzerland, pp. 35-115.). Retrieved from https://doi.org/10.59327/IPCC/AR6-9789291691647

Komé, C. E., Andrews, S. S., & Franzluebbers, A. J. (2013). Soil organic carbon assessment using the Carbon Management Evaluation Tool for Voluntary Reporting and the Soil Conditioning Index. *Journal of Soil and Water Conservation, 68*(4), 296. doi:https://doi.org/10.2489/jswc.68.4.296

Kukal, M. S., & Irmak, S. (2018). Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production. *Scientific Reports, 8*(1), 3450. doi:https://doi.org/10.1038/s41598-018-21848-2

Lal, R. (2004). Soil carbon sequestration to mitigate climate change. *Geoderma, 123*(1), 1-22. doi:https://doi.org/10.1016/j.geoderma.2004.01.032

Lal, R. (2015). Sequestering carbon and increasing productivity by conservation agriculture. *Journal of Soil and Water Conservation, 70*(3), 55A-62A. doi:10.2489/jswc.70.3.55A

Lemenkova, P. (2021). ISO Cluster classifier by ArcGIS for unsupervised classification of the Landsat TM image of Reykjavík. *Bulletin of Natural Sciences Research, 11*(1), 29-37. doi:https://doi.org/10.5937/bnsr11-30488

Lin, K., Chen, H., Xu, C.-Y., Yan, P., Lan, T., Liu, Z., & Dong, C. (2020). Assessment of flash flood risk based on improved analytic hierarchy process method and integrated maximum likelihood clustering algorithm. *Journal of Hydrology, 584*, 124696. doi:https://doi.org/10.1016/j.jhydrol.2020.124696

Lu, C., Yu, Z., Tian, H., Hennessy, D. A., Feng, H., Al-Kaisi, M., . . . Arritt, R. (2018). Increasing carbon footprint of grain crop production in the US Western Corn Belt. *Environmental Research Letters, 13*(12), 124007. doi:https://doi.org/10.1088/1748-9326/aae9fe

Lv, Z., Liu, T., Shi, C., Benediktsson, J. A., & Du, H. (2019). Novel Land Cover Change Detection Method Based on k-Means Clustering and Adaptive Majority Voting Using Bitemporal Remote Sensing Images. *IEEE Access, 7*, 34425-34437. doi:https://doi.org/10.1109/ACCESS.2019.2892648

Melillo, J. M., Richmond, T. C., & Yohe, G. W. (2014). Climate Change Impacts in the United States: The Third National Climate Assessment. doi:https://doi.org/10.7930/J0Z31WJ2

NASA. (2022). The Causes of Climate Change. *Global Climate Change: Vital Signs of the Planet.* Retrieved from https://climate.nasa.gov/causes/

Navin, M. S., & Agilandeeswari, L. (2019). Land use land cover change detection using k-means clustering and maximum likelihood classification method in the javadi hills, Tamil Nadu, India. *International Journal of Engineering and Advanced Technology (IJEAT)*. doi:https://doi.org/10.35940/ijeat.A1011.1291S319

Nijhawan, R., Srivastava, I., & Shukla, P. (2017). Land cover classification using supervised and unsupervised learning techniques. *2017 International Conference on Computational Intelligence in Data Science (ICCIDS)*, 1-6. doi:https://doi.org/10.1109/ICCIDS.2017.8272630

NOAA. (2022). Trends in Atmospheric Carbon Dioxide. *Global Monitoring Laboratory - Carbon Cycle Greenhouse Gases.* Retrieved from https://gml.noaa.gov/ccgg/trends/

Norzaki, N., & Tahar, K. N. (2019). A comparative study of template matching, ISO cluster segmentation, and tree canopy segmentation for homogeneous tree counting. *International journal of remote sensing, 40*(19), 7477-7499. doi:https://doi.org/10.1080/01431161.2018.1524182

Oldfield, E., Eagle, A., Rubin, R., Rudek, J., Sanderman, J., & Gordon, D. (Producer). (2021). Agricultural soil carbon credits: Making sense of protocols for carbon sequestration and net greenhouse gas removals. [PowerPoint slides] Retrieved from https://edf.org/sites/default/files/content/agricultural-soilcarbon-credits-protocolsynthesis.pdf

Oldfield, E., Eagle, A., Rubin, R., Rudek, J., Sanderman, J., & Gordon, D. (2022). Crediting agricultural soil carbon sequestration. *Science, 375*(6586), 1222-1225. doi:https://doi.org/10.1126/science.abl7991

Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional, 16*, e00198. doi:https://doi.org/10.1016/j.geodrs.2018.e00198

Prince, S. D., Haskett, J., Steininger, M., Strand, H., & Wright, R. (2001). Net primary production of U.S. Midwest croplands from agricultural harvest yield data. *Ecological Applications, 11*(4), 1194-1205. doi:https://doi.org/10.1890/1051-0761(2001)011[1194:NPPOUS]2.0.CO;2

Röthlisberger, V., Zischg, A. P., & Keiler, M. (2017). Identifying spatial clusters of flood exposure to support decision making in risk management. *Science of The Total Environment, 598*, 593-603. doi:https://doi.org/10.1016/j.scitotenv.2017.03.216

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65. doi:https://doi.org/10.1016/0377-0427(87)90125-7

S. 1251 - 117th Congress. (2021-2022). Growing Climate Solutions Act of 2021. Retrieved from https://www.congress.gov/bill/117th-congress/senate-bill/1251/all-info. https://www.congress.gov/bill/117th-congress/senate-bill/1251/all-info

Sathler, D., Adamo, S. B., & Lima, E. E. C. (2018). Deforestation and local sustainable development in Brazilian Legal Amazonia: an exploratory analysis. *Ecology and Society, 23*(2). Retrieved from https://www.jstor.org/stable/26799094

Smith, P., Lanigan, G., Kutsch, W. L., Buchmann, N., Eugster, W., Aubinet, M., . . . Jones, M. (2010). Measurements necessary for assessing the net ecosystem carbon budget of croplands. *Agriculture, Ecosystems & Environment, 139*(3), 302-315. doi:https://doi.org/10.1016/j.agee.2010.04.004

Stahl, C., Fontaine, S., Klumpp, K., Picon-Cochard, C., Grise, M. M., Dezécache, C., . . . Blanfort, V. (2017). Continuous soil carbon storage of old permanent pastures in Amazonia. *Global Change Biology, 23*(8), 3382-3392. doi:https://doi.org/10.1111/gcb.13573

Sus, O., Williams, M., Bernhofer, C., Béziat, P., Buchmann, N., Ceschia, E., . . . Wattenbach, M. (2010). A linked carbon cycle and crop developmental model: Description and evaluation against measurements of carbon fluxes and carbon stocks at several European agricultural sites. *Agriculture, Ecosystems & Environment, 139*(3), 402-418. doi:https://doi.org/10.1016/j.agee.2010.06.012

Tang, Y., Jones, E., & Minasny, B. (2020). Evaluating low-cost portable near infrared sensors for rapid analysis of soils from South Eastern Australia. *Geoderma Regional, 20*, e00240. doi:https://doi.org/10.1016/j.geodrs.2019.e00240

United Nations Development Programme. (2022). What are carbon markets and why are they important? Retrieved from https://climatepromise.undp.org/news-and-stories/what-are-carbon-markets-and-why-are-they-important

Upson, M. A., Burgess, P. J., & Morison, J. I. L. (2016). Soil carbon changes after establishing woodland and agroforestry trees in a grazed pasture. *Geoderma, 283*, 10-20. doi:https://doi.org/10.1016/j.geoderma.2016.07.002

USDA. (2023). *A General Assessment of the Role of Agriculture and Forestry in the U.S. Carbon Markets*. Retrieved from https://www.usda.gov/sites/default/files/documents/USDA-General-Assessment-of-the-Role-of-Agriculture-and-Forestry-in-US-Carbon-Markets.pdf

USDA National Agricultural Statistics Service. (2024). 2023 Cotton Yield. In: https://www.nass.usda.gov/Charts_and_Maps/graphics/cotnmap.pdf.

USDA NRCS. (2002). *Soil conditioning index for cropland management systems*. Retrieved from NRCS National Agronomy Manual, Directive No. 190-V-NAM. Washington, DC: USDA Natural Resources Conservation Service: https://efotg.sc.egov.usda.gov/references/Agency/TX/Archived_National_Agronomy_Manual_Oct_2002_160822.pdf

USDA NRCS and Colorado State University. (2023). COMET-Farm. Retrieved from http://cometfarm.nrel.colostate.edu. http://cometfarm.nrel.colostate.edu

USGRCP. (2023). *Fifth National Climate Assessment*. Washington, DC, USA: U.S. Global Change Research Program.

West, T. O., Brandt, C. C., Wilson, B. S., Hellwinckel, C. M., Tyler, D. D., Marland, G., . . . Nelson, R. G. (2008). Estimating Regional Changes in Soil Carbon with High Spatial Resolution. *Soil Science Society of America Journal, 72*(2), 285-294. doi:https://doi.org/10.2136/sssaj2007.0113

Yuan, Y., Shi, B., Yost, R., Liu, X., Tian, Y., Zhu, Y., . . . Cao, Q. (2022). Optimization of Management Zone Delineation for Precision Crop Management in an Intensive Farming System. *Plants, 11*(19), 2611. Retrieved from https://www.mdpi.com/2223-7747/11/19/2611

Zhang, X., Izaurralde, R. C., Manowitz, D. H., Sahajpal, R., West, T. O., Thomson, A. M., . . . Williams, J. R. (2015). Regional scale cropland carbon budgets: Evaluating a geospatial agricultural modeling system using inventory data. *Environmental Modelling & Software, 63*, 199-216. doi:https://doi.org/10.1016/j.envsoft.2014.10.005

Zhang, Y., & Zhao, H. (2020). Land–Use and Land-Cover Change Detection Using Dynamic Time Warping–Based Time Series Clustering Method. *Canadian Journal of Remote Sensing, 46*(1), 67-83. doi:https://doi.org/10.1080/07038992.2020.1740083

Zhou, W., Guan, K., Peng, B., Tang, J., Jin, Z., Jiang, C., . . . Mezbahuddin, S. (2021). Quantifying carbon budget, crop yields and their responses to environmental variability using the ecosys model for U.S. Midwestern agroecosystems. *Agricultural and Forest Meteorology, 307*, 108521. doi:https://doi.org/10.1016/j.agrformet.2021.108521

Zhu, Z., Bouchard, M., Butman, D., Hawbaker, T., Li, Z., Liu, J., . . . Zhu, Z. (2011). *Baseline and projected future carbon storage and greenhouse-gas fluxes in the Great Plains region of the United States: U.S. Geological Survey Professional Paper 1787*. Retrieved from https://pubs.usgs.gov/pp/1787/

Zhu, Z., & Reed, B. C. (2012). *Baseline and projected future carbon storage and greenhouse-gas fluxes in ecosystems of the Western United States: U.S. Geological Survey Professional Paper 1797*. Retrieved from https://pubs.usgs.gov/pp/1797/

Zobeck, T. M., Crownover, J., Dollar, M., Van Pelt, R. S., Acosta-Martinez, V., Bronson, K. F., & Upchurch, D. R. (2007). Investigation of soil conditioning index values for southern high plains agroecosystems. *Journal of Soil and Water Conservation, 62*(6), 433-442.

CHAPTER II:
GEOSPATIAL ANALYSIS OF CROP PRODUCTION AND PUBLICLY-FUNDED RESEARCH STATIONS TO SCALE CARBON MARKETS

2.1 Abstract

Agricultural carbon markets compensate farmers for the implementation of climate smart practices to increase soil organic carbon (SOC) sequestration in agricultural land. To incentivize farmers to enroll in carbon credit contracts, it is necessary to identify model farms with historical crop production and land management data from which carbon markets can base their estimations of SOC sequestration. Since direct soil measurements are often required to verify SOC sequestration throughout the life of the contract and to validate the quantity of carbon credits earned, it would be advantageous to utilize publicly-funded research centers and their historical, unbiased datasets for these model farms. This research utilizes production data and SOC estimates for fields on private farmland and University of Tennessee AgResearch and Education Centers (RECs) to evaluate the potential for scaling measurement, reporting, and verification (MRV) efforts for carbon markets in Tennessee. Relationships between the RECs and cropland in surrounding regions across the state were analyzed by developing a geospatial model using ArcGIS Pro software and an iterative self-organizing clustering technique. Model inputs were factors affecting agricultural SOC sequestration: soil texture, average daily temperature, average daily precipitation, and topography. Models were executed with and without the addition of crop production frequency data for corn, cotton, and soybeans, the three main crops produced in West Tennessee, to compare the geospatial regions for each crop. Models were also executed with varying k-values representing the number of cluster means for the model to initialize randomly across the state for use in determining the output clusters. Model outputs represented clusters of geospatially-correlated regions determined based on crop input and k-value. The silhouette coefficient was calculated for each model output to evaluate the model abilities to determine correlated regions for each cropping system. COMET-Farm and Field to Market (FM) scenarios were also executed for cropland on private farms and at three RECs located in close proximity to corn, cotton, and soybean production in West Tennessee. SOC sequestration values from COMET-Farm were recorded for baseline scenarios -- where corn, cotton, and soybeans were each grown as a continuous crop -- and future sustainable management scenarios. The future scenarios utilized cereal rye, one of the most common cover crops in Tennessee, as a cover crop

between cash crop growing seasons. Soil Conditioning Index (SCI) values were also calculated using the FM Fieldprint Calculator for best-case and worst-case management scenarios representative of common management practices in Tennessee. SOC and SCI values were then compared by cluster for each model output. The model executed without any crop production information and with a k-value of three had a silhouette coefficient of 0.667, the highest score out of all evaluated models. SOC sequestration for this highest-performing output ranged from -0.07 to -0.65 tons CO_2-e ac^{-1} yr^{-1}, and SCI values ranged from 0.39 to -0.61 for the best-case and worst-case management scenarios respectively. Future work intends to expand this modeling approach to include direct soil samples for ground-truthing and validating model outputs to further assess how climate smart land management practices can affect SOC sequestration and SCI values based on cropping system. Additionally, this work could encourage further utilization of publicly-funded RECs for developing standards to estimate SOC sequestration potential on public farmland, and for implementing sustainable management practices for compensation in carbon markets.

2.2 Introduction

Agricultural land stores about 7.2 Gt of organic carbon in the top 20 cm of soil, nearly one-third of the estimated 22.5 Gt of organic carbon stored across the contiguous United States (Zhu et al., 2011; Zhu & Reed, 2012). Due to the ability of agricultural ecosystems to sequester carbon from the atmosphere, utilize it for plant growth, and store it in plant matter and organic matter in the soil, agricultural soils are a critical component of the carbon budget. Agricultural carbon budgets are influenced by numerous factors, including climate, weather conditions, soil characteristics, topography, the types of soil and nutrient management practices utilized, and the use of perennial vegetation or cover crops to maintain and improve soil health between primary crop growing seasons. Climate change introduces more variability in the duration and frequency of rainfall events, air temperatures, and crop growing seasons in response to changes in climate patterns. In addition to implementing sustainable management practices like reduced till and no-till, cover crops can be used to slow the rate of erosion from fields and improve soil health and water availability between primary crop growing seasons (USDA, 2023). Cover crops have had lower adoption rates than reduced or no-tillage practices in the United States, though it is estimated the adoption of cover crops on the 88 Mha of cropland growing the five primary crops – corn, soybeans, wheat, rice, and cotton – not already using cover cropping could advance emission reduction efforts by sequestering upwards of 100 MMt CO_2-e each year (Fargione et al., 2018). Assessing carbon sequestration potential and the spatial distribution of carbon storage across agricultural ecosystems in the United States will further the understanding of how sustainable management practices can be used to influence agricultural SOC sequestration.

While direct soil organic carbon (SOC) measurements can be analyzed to assess agricultural carbon budgets and their responses to sustainable management practices at the site scale, one of the major challenges the agricultural modeling community is currently facing is utilizing site-scale observations to extrapolate for use in regional-scale geospatial models. Crop management spatial data for tillage, fertilization, and irrigation practices are generally not available at the regional scale, presenting a challenge when evaluating the distribution of carbon across a given region in relation to said management practices. Carbon budgets for agricultural ecosystems and management practices have been extensively estimated and modeled for the entire US Midwest (Grace & Robertson, 2021; Lu et al., 2018; Prince et al., 2001; X. Zhang et al., 2015; Zhou et al., 2021), as the Midwest is a highly productive region that produces over 80% of the corn, 80% of the soybeans, and 50% of the wheat grown in the US each year (X. Zhang et al., 2015). Carbon budgets and climate change impacts have also been modeled for the Great Plains region (Kukal & Irmak, 2018) and the North Central region of the US (Hollinger et al., 2005). Fewer studies have modeled agricultural carbon budgets and SOC sequestration in the Southeastern United States, where 53% of US cotton was produced in 2023 (USDA National Agricultural Statistics Service, 2024).

It is crucial to be able to reliably measure and track SOC storage trends in agricultural ecosystems in order to establish standards for baselines used in carbon credit projects within carbon markets. Measurement, reporting, and verification (MRV) techniques are needed to validate SOC sequestration and carbon dioxide (CO_2) emission reductions across different project scales. These MRV efforts involve tracking specific measurement methods and their results, transparently reporting SOC information, and utilizing third-party verifiers to validate the accuracy of the data being collected and reported. Process-based models supplemented with remote sensing data are currently being used by multiple carbon market registries to quantify greenhouse gas emissions and simulate biogeochemistry in agricultural ecosystems. Examples of such models include the DayCent, DeNitrification-DeComposition, System Approach to Land Use Sustainability, and Ecosys models (USDA, 2023). Use of models such as these helps reduce dependency on direct soil sampling and lessens the amount of site-specific historical data that is often required from agricultural landowners and producers. The verification component of MRV efforts often requires direct sampling over the life of the contract, though many verification programs also utilize empirical models to estimate carbon stock baselines and effects on greenhouse gas emissions after the implementation of sustainable management practices. Verification approaches based only on direct sampling often take longer to generate credits, anywhere between three to five years, while model-based estimates avoid the need for intensive direct sampling and can issue credits faster, often within the same year (Oldfield et al., 2022). Differences in MRV protocols, including how a given carbon market may account for defining and proving additionality, permanence, leakage, and reversals, can potentially create credits that are not equivalent between markets (USDA, 2023). There is a need for consistency across carbon market systems to ensure the integrity of credits and MRV efforts (Oldfield et al., 2022). There is also a need to generate and trade carbon credits that have equivalent value between countries, so carbon markets are not constrained to given countries and can move towards international standardization.

There is some inherent uncertainty with model-based estimates as compared to direct soil measurements, and this uncertainty can be influenced by the precision of the model inputs and the scale. Utilizing existing soil monitoring programs and networks of agricultural research and education centers (RECs) for this representative sampling to validate model-based estimates could be advantageous, since these existing RECs are often publicly-funded and provide unbiased historical data to better estimate SOC stocks over extended periods of time. This data could then be used to validate results from machine learning models that use environmental inputs for a given region to create geospatially-correlated regions using a clustering algorithm. In the interest of creating a framework for developing this type of model, the scope was narrowed from the Southeastern US to the state of Tennessee. A model of Tennessee with clusters representing geospatially-correlated regions to the RECs based on correlated environmental conditions and crop production frequency would assist farmers with identifying their most correlated REC(s) and with estimating the SOC sequestration potential of their farmland. Farmers then might be more inclined to enroll in a carbon market and report the data required by the contract and third-party verifiers to prove SOC additionality throughout the life of the contract.

2.3 Materials and Methods

2.3.1 Geographical Area of Interest

In Tennessee, forests and agricultural land used to produce soybeans, corn, and cotton make up over 17 million acres, or over 60% of the state's land area (USDA NASS 2024). There are ten University of Tennessee Institute of Agriculture (UTIA) RECs across the state, each with varying research focuses and cropping systems (fig. 2.1). Corn and soybeans are grown and studied at the Ames, West Tennessee, Milan, Highland Rim, Middle Tennessee at Spring Hill, East Tennessee, and Northeast Tennessee research centers. Cotton is grown primarily in West Tennessee and researched at the Ames, West Tennessee, and Milan research centers. Forestry is studied at the Ames, West Tennessee, Forest Resources and UT Arboretum, and Northeast Tennessee research centers.

Due to the widespread diversity of research and historical site-specific data at the RECs, an opportunity exists to gain insight into how the location of each REC relates to geospatial similarities and growing conditions found in agricultural land across Tennessee. For example, if a farmer is growing corn in the region between the West Tennessee and the Spring Hill research centers, it could be beneficial for the farmer to know which research center might have more relevant data and resources they could obtain to improve production on their fields. The farmer might be able to compare their data with one research center over the other, or both. If the farmer could gain insight on how research plots at their

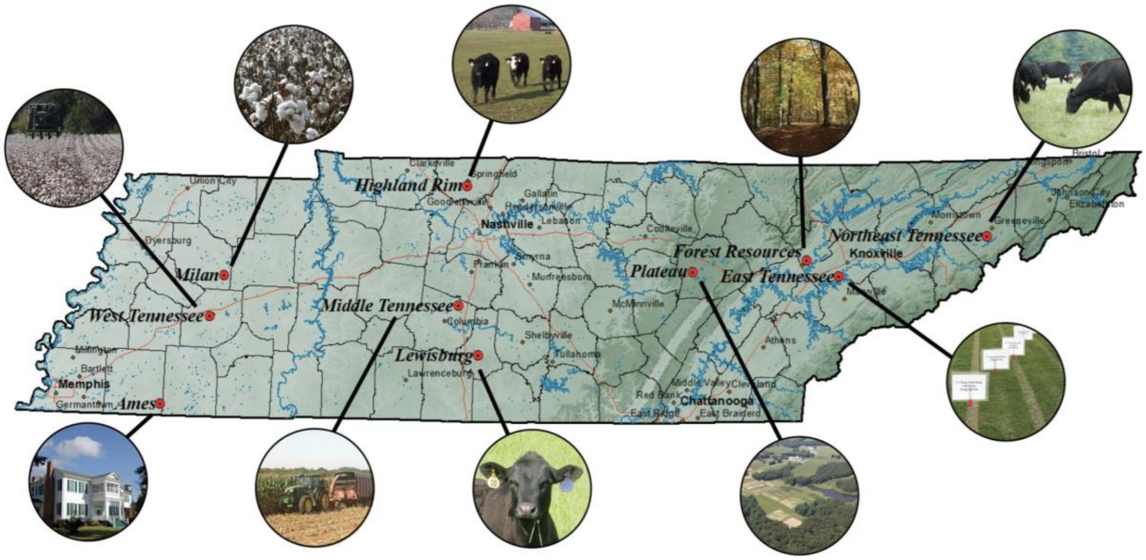

Figure 2.1. University of Tennessee AgResearch and Education Centers.

Reprinted from UT AgResearch (2024). Retrieved from https://agresearch.tennessee.edu/centers.

most similar REC(s) might be compensated for additionality in a carbon credit project, the farmer might be more inclined to enroll in a carbon market system and report their data since they could estimate the amount of SOC sequestration they might yield and the profit they could earn from enrolling in a carbon market.

2.3.2 Geographic Information System

The Geographic Information System (GIS) used for this research was Environmental Systems Research Institute (ESRI) ArcGIS Pro 3.1.3 software. ArcGIS Pro is one of the most common geospatial modeling softwares because of its user interface and the multitude of analyses that can be performed. Projects in ArcGIS Pro can be developed to map and analyze geospatial problems relating to geosciences, environmental studies, and engineering using spatial data in the form of rasters and vectors. Raster data is formatted as a grid of cells with attributes and values associated to each cell. Rasters are useful for displaying data such as satellite images, land use classifications, and digital elevation models. Vector data can be formatted as points, lines, and polygons. This type of data is useful for displaying discrete points and boundaries, such as a county map for a state or a layer of water bodies located in a given area. Raster and vector data layers can be used in many geoprocessing tools available in ArcGIS Pro.

2.3.3 Data Sources and Descriptions

SOC sequestration capacity depends on soil depth, clay and mineral content, plant available water holding capacity, nutrient reserves, landscape position, and crop type (Huang et al., 2009; Lal, 2015). Climatic factors including rainfall, maximum and minimum air temperatures, and evapotranspiration rates are important to consider for crop growth (Ogle, Swan, and Paustian, 2012), which affects the amount of above ground and below ground carbon associated with crop and root biomass. Cash crops are harvested each year and do not provide long-term carbon storage in the form of above ground plant biomass, though the addition of organic residues and cover cropping can help create a positive SOC budget to overcome SOC lost through mineralization -- the conversion of organic carbon to inorganic carbon, occurring largely due to organic matter decomposition -- and erosion from water or wind (Chenu et al., 2019; Huang et al., 2009). A positive SOC budget can be achieved by applying carbon-rich amendments like biochar, compost, and manure, and by allowing crop residue to remain on the soil surface after harvest (Chenu et al., 2019; Lal, 2015). Due to their importance in influencing the amount of SOC stored in agricultural soils, data layers utilized in the model will include crop production locations, average air temperature, average precipitation, soil texture characteristics, and topography (table 2.1).

2.3.3.1 Crop Production

US Department of Agriculture (USDA) Crop Data Layers (CDL) were used for the crop production input layers (USDA NASS, 2024). Crop production frequency data was collected for 2008-2022 for Tennessee for corn, cotton, and soybeans. The crop frequency data is provided at 30 m resolution; layers for corn, cotton, and soybeans were clipped to Tennessee using the nearest neighbor resampling method for categorical data (fig. 2.2). These three row crops were chosen as they are the crops primarily produced in Middle and West Tennessee. The crop frequency data layers were converted to simplified layers using the Reclassify tool to produce layers with raster cell values of 1 where the crop had been grown at least once within 2008-2022 and raster cell values of 0 where the crop had not been grown within that time period (fig. 2.3).

2.3.3.2 Soil Texture

Soil texture data was obtained using the protocol developed for the USDA Natural Resources Conservation Service (NRCS) Soil Survey Geographic (SSURGO) Database Bulk Downloader for ArcGIS Pro toolbox (SSURGO, 2024). This protocol served as a guide for batch downloading soil surveys for all counties in Tennessee, creating a geospatial database, aggregating SSURGO data based on specified properties, and creating soil texture layers. The SSURGO Portal 0.1.0.37 Beta application was used to create a geospatial SQLite SSURGO template database, and the Tennessee soil surveys were imported into the template database. The SSURGO On Demand (SOD) tool was installed from the SSURGO Portal website and the SQLITE template database was selected in the Properties pop-up window. The Numeric Dominant Component aggregation method and the representative values for the Total Sand, Total Silt, and Total Clay soil properties were selected. The soil depth ranges selected were a top depth of 0 cm and a bottom depth of 25 cm. The SOD tool was executed, and the output was added to the SSURGO template

Table 2.1. Data Sources.

Data Type	Source	Year	Original Resolution	Final Resolution
Cropland Area	USDA Crop Frequency Layers	2008-2022	30 m	30 m
Soil Texture	NRCS gSSURGO Database[1]	2022	10 m	30 m
Average Daily Precipitation	Daymet ORNL DAAC SDAT[2]	2012-2022	1 km	30 m
Average Daily Temperature	Daymet ORNL DAAC SDAT[2]	2012-2022	1 km	30 m
Elevation	USGS 3DEP[3]	2022	30 m	30 m

[1] gridded Soil Survey Geographic Database

[2] Oak Ridge National Laboratory Distributed Active Archive Center Spatial Data Access Tool V1

[3] 3D Elevation Program

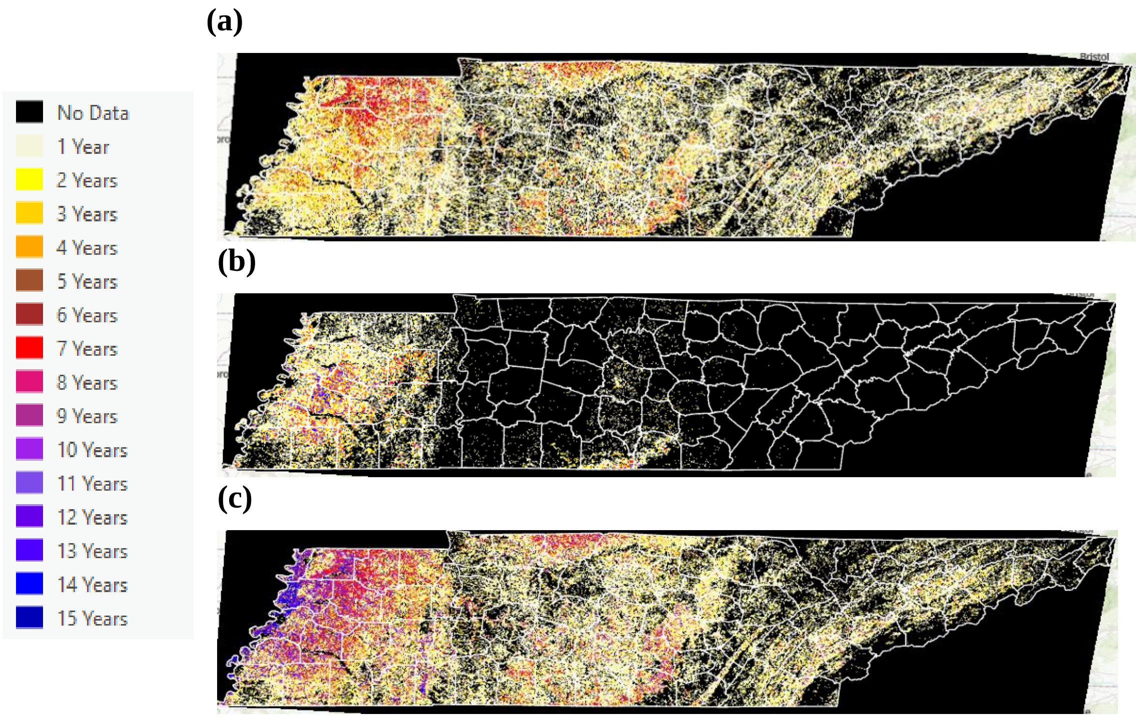

Figure 2.2. Crop frequency layers for (a) corn, (b) cotton, and (c) soybeans in Tennessee, 2008-2022.

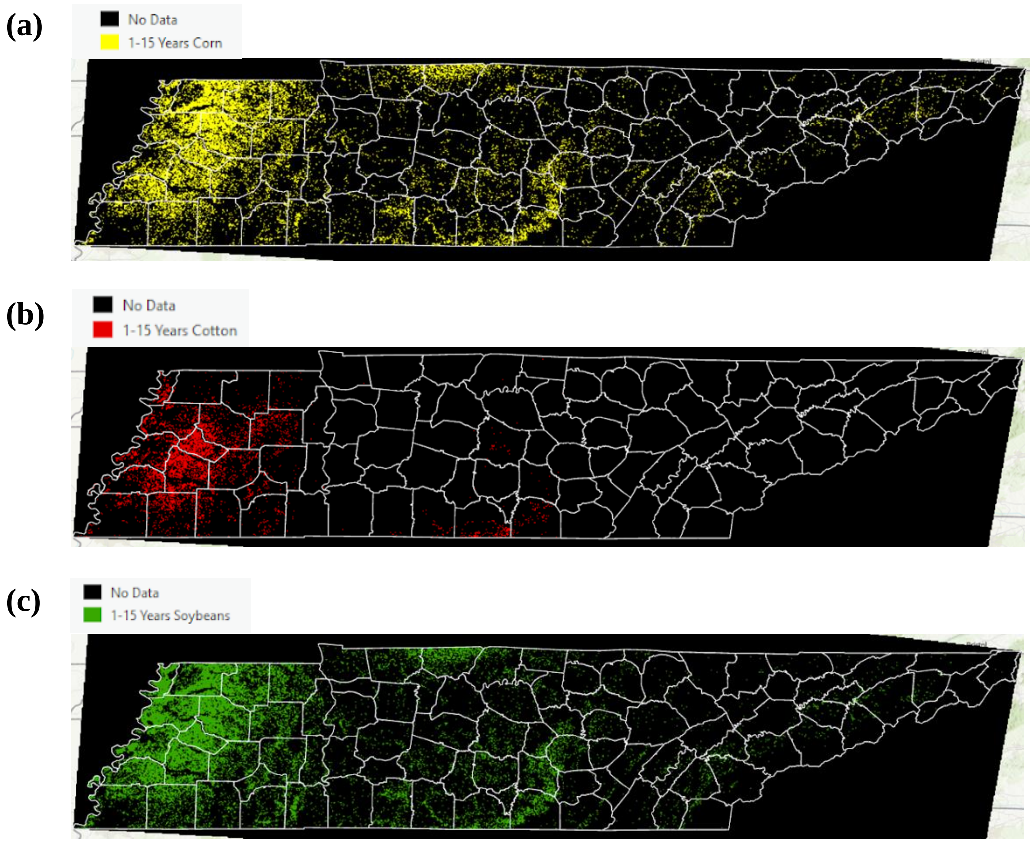

Figure 2.3. Simplified crop frequency layers for (a) corn, (b) cotton, and (c) soybeans in Tennessee, 2008-2022.

database. This database was then added to the ArcGIS Pro project. The main.SOD sand, silt, and clay tables from the database were added to the project. The main.polygon feature layer from the database was added to the project, and the sand table was joined to the main.polygon table using the common map unit key (MUKEY) field. The Export Features tool was then used to create a new feature class layer symbolized by the total representative sand feature. The Feature to Raster tool then converted this feature class layer to a raster, with the CDL used both as the snap raster and to define the 30 m resolution for the output sand raster. This process was repeated to create a silt raster and a clay raster (fig. 2.4).

2.3.3.3 Precipitation

For precipitation, open-source data was obtained from the Daymet data product through the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) Spatial Data Access Tool (SDAT) V1. The SDAT (ORNL DAAC, 2017) provides access to Daymet Version 4 R1 monthly climate summaries at 1 km resolution for North America. When accessing the Daymet data through SDAT for 2012 through 2022, the year, monthly total precipitation, and Continental North America placename were selected. A bounding box for the region of interest was then drawn to define the spatial extent for downloading the Daymet data for the region surrounding Tennessee. The Daymet data was downloaded in the GeoTIFF (FLOAT32) raster file format for all 12 months in the specified year. The GeoTIFF rasters were added to the ArcGIS Pro project and the Raster Calculator tool was used to compute the average daily precipitation for each year by summing all 12 monthly total precipitation rasters and dividing by 365 (or 366 in the case of a leap year). Once the average daily precipitation was calculated for 2012 through 2022 and the resulting layers were projected and resampled, the overall average daily precipitation layer for the study period was created using Raster Calculator. The resulting layer (fig. 2.5) was clipped to Tennessee using the Clip Raster tool, in which a shapefile of the Tennessee border was selected for the clipping geometry to define the extent of the output raster. The Project Raster and Resample tools were then used to create average daily precipitation rasters using the North American Datum 1983 (NAD83) Contiguous USA Albers as the projected coordinate system (PCS) at 30 m resolution to match the resolution of the crop production frequency layers.

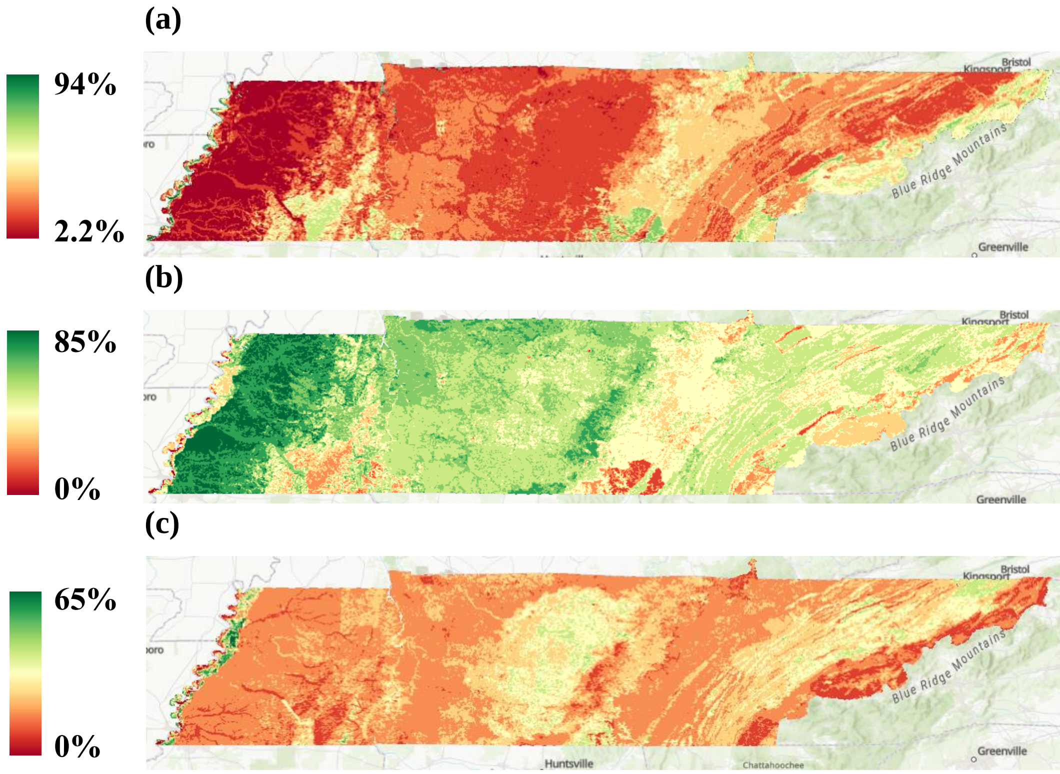

Figure 2.4. Soil texture component percentage for (a) sand, (b) silt, and (c) clay, 2022.

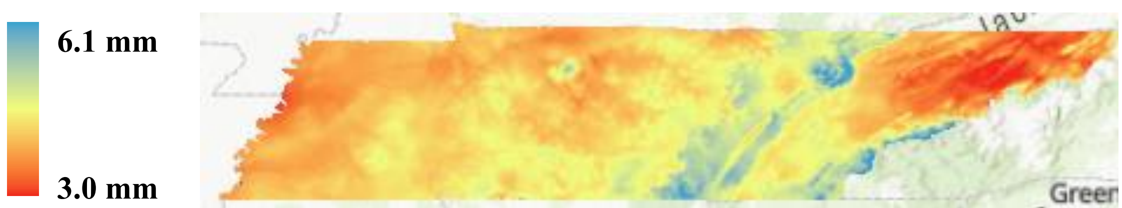

Figure 2.5. Average daily precipitation on annual basis, 2012-2022.

2.3.3.4 Temperature

To obtain average temperature data from Daymet, a similar approach was taken using the SDAT (ORNL DAAC, 2017). Monthly average daily maximum and minimum temperature GeoTIFF rasters were downloaded at 1 km resolution for all months in each year from 2012 to 2022. The GeoTIFF rasters were added to the ArcGIS Pro project, the Raster Calculator tool was used to compute the maximum and minimum average daily temperatures for each year, and each resulting temperature layer was projected using the NAD83 Contiguous USA Albers PCS and resampled to 30 m resolution to match the resolution of the crop frequency layers. The average daily temperature for each year was calculated from the maximum and minimum temperature layers for the respective year before the overall average temperature for 2012 through 2022 was calculated. The Clip Raster tool was used to clip the resulting layer (fig. 2.6) to the Tennessee border. The Project Raster and Resample tools were then used to create average daily temperature rasters using the NAD83 Contiguous USA Albers as the PCS at 30 m resolution to match the resolution of the crop production frequency layers.

2.3.3.5 Elevation

Topography data was obtained from the USGS 3D Elevation Program National Elevation Dataset. Digital Elevation Model (DEM) tiles were downloaded at 30 m resolution using The National Map (TNM) Download Client v2.0 (U.S. Geological Survey, 2019) for the entirety of Tennessee. The TIF tiles were then added to the ArcGIS Pro project, and the Mosaic tool was used to stitch the tiles together to create one raster. The Clip Raster tool was used to clip the mosaic raster to Tennessee, and the raster was then projected using the NAD83 Contiguous USA Albers PCS and bilinear interpolation resampling for continuous data (fig. 2.7).

2.3.4 Input Optimization

It is common to use geospatial modeling softwares to analyze existing SOC sequestration capacity and predict future trends for policy planning purposes or to better understand overall carbon budgets in a given region (Grace & Robertson, 2021; He & Mostovoy, 2019; Hollinger, Bernacchi, & Meyers, 2005; Kukal & Irmak, 2018; Zhang et al., 2015; Zhou et al., 2021). When working with multiple geospatial layers in an ArcGIS Pro model, the

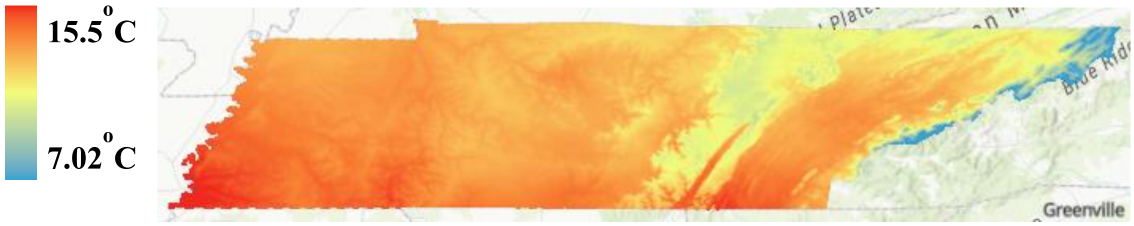

Figure 2.6. Average daily temperature on annual basis, 2012-2022.

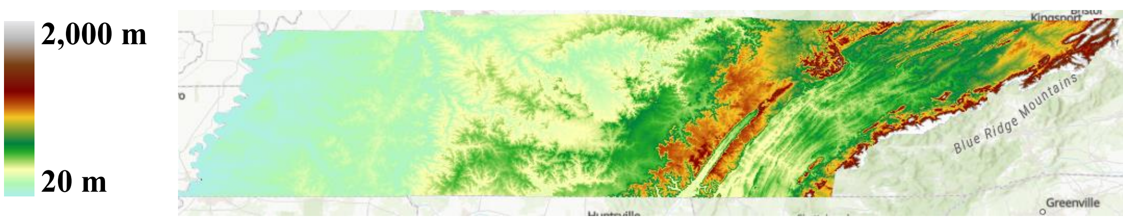

Figure 2.7. Digital Elevation Model for Tennessee, 2022.

coordinate systems of each input layer should be considered. There are two types of coordinate systems in geospatial layers: geographic coordinate system (GCS) and PCS. A GCS defines where data is located on the earth’s surface and utilizes three-dimensional information to appropriately display the data. GCS data locations are commonly reported in angular units such as degrees. A PCS is used to visualize data on a flat surface such as a paper map or a computer screen. GCS data locations are commonly reported in linear units such as feet or meters. If data layers each have a different PCS, the data may be skewed based on the datum utilized by each PCS. It is best practice to convert all data layers to the same PCS for optimal data visualization and for use in geospatial modeling. For these reasons, all input layers were converted to the NAD83 Contiguous USA Albers PCS.

It is also important to compare the resolution of each input layer in relation to the size of the area being modeled. The spatial resolution of the model output is limited to the input layer with the lowest resolution, and therefore the data resolution is directly correlated with the accuracy of the model. If the spatial resolution is low, generalizations will be made as the model generates its predicted outputs. As the spatial resolution of the data is increased and the area encompassed by each cell in the input layers is refined, the accuracy of the model predictions will improve. This increased accuracy comes at the cost of increasing the overall processing time required to run the ArcGIS model, because the software will have to process more information as the spatial resolution increases. Assumptions must be made when resampling a lower-resolution raster to a higher resolution, and the results of the model will assume a greater degree of accuracy than the resampled data actually represents, since the value of one pixel in the original layer would be applied to the smaller pixels making up the same area in the resampled layer. The resampling of the Daymet temperature and precipitation data from 1 km to 30 m was performed so the model outputs would not be restricted to 1 km and could provide information for each 30 m pixel based on the crop frequency rasters. When considering the intended application of the model results, it would be more meaningful to have classified outputs for a smaller spatial unit rather than being limited to providing cluster assignments for a 1 km by 1 km square pixel, equivalent to approximately 250 acres, as fields in West Tennessee can be smaller than this acreage.

The ranges of values within each input layer are important to assess since there are often different units and ranges across layers, depending on the data type. For example, when working with a study region as large and as diverse as Tennessee, the DEM input layer will have broader ranges of values compared to the input layer with average daily temperature data. While average daily temperature does vary across the state, the difference between the maximum and minimum value is not as large as that of the elevation ranges across the state. For these reasons, it is necessary to normalize the input layers before running analyses and generating model results. The normalization approach suggested by ESRI for the Iterative Self-Organizing (ISO) Cluster Unsupervised Classification (ICUC) method is to transform the range of values in all input layers using the normalization equation. Normalization minimizes the effects of different units and ranges of values across input variables and ensures the inputs with larger ranges do not overpower the model output compared to smaller scale inputs, since the ranges of all layers are transformed to be between [0,1], as shown in Equation 2.1:

$$Z = \frac{(X - min_{old})(max_{new} - min_{new})}{(max_{old} - min_{old})} + min_{new} \tag{2.1}$$

where,

Z = output raster with normalized data ranges

X = original raster

$max_{new} = 1$

$min_{new} = 0$

2.3.5 Model Development

Once appropriate input data layers were collected and preprocessed, the ICUC tool was initialized. Input layers were ordered based on their impacts on carbon sequestration in agricultural ecosystems. Per ArcGIS Pro documentation, the minimum class size was set as 10 times the number of input layers for the ICUC model. The default value of ten was used for the sample interval. The parameter for the maximum number of classes was compared for 3, 5, 7, and 10, first without any crop inputs to determine the geospatially-correlated regions across Tennessee based primarily on the environmental inputs. These

ICUC model outputs run were then used as a baseline for comparing ICUC outputs created with the crop frequency layer for corn, cotton, and soybeans.

2.3.6 Model Evaluation

2.3.6.1 Fields on UTIA RECs and Private Farmland

Five fields on each of the RECs in West Tennessee (Milan, Jackson, and Ames) were selected for use in the model evaluation step. The majority of corn, cotton, and soybean production in Tennessee occurs in the Western grand division. These three RECs were chosen due to their proximity to row crop production, their research on these three crops, and their research on sustainable management practices compensated for in carbon markets. The coordinates for the RECs and their associated fields were formatted as a table of comma separated values and added to the ArcGIS Pro project. The XY Table to Point tool was then used to create a point layer for the fields at the Milan (fig. 2.8), Jackson (fig. 2.9), and Ames (fig. 2.10) RECs. Fields on farmland totaling over 6,500 acres in West Tennessee (fig. 2.11) were also used in the model evaluation steps in order to compare SOC and SCI estimations to the RECs based on the geospatially-correlated regions defined by the model.

2.3.6.2 Silhouette Coefficient

The silhouette coefficient was first developed by Rousseeuw (1987) for comparing within-cluster and between-cluster dissimilarities. Values for the silhouette coefficient range from [-1,1]. When the within-cluster dissimilarity is much smaller than the smallest between-cluster dissimilarity, the silhouette coefficient $s(i)$ is close to 1, and it is assumed the sample i within the cluster is appropriately assigned. A silhouette coefficient of 0 indicates the sample i is close to the decision boundary between clusters A and B. A silhouette coefficient less than 0 indicates the sample likely was assigned to the wrong cluster, since the between-cluster dissimilarity is less than the within-cluster dissimilarity for that sample. The silhouette coefficient equation (Rousseeuw, 1987) is shown in Equation 2.2:

Figure 2.8. Milan REC and associated fields used in COMET-Farm and Field to Market scenarios for estimating soil organic carbon sequestration and Soil Conditioning Index values.

Figure 2.9. Jackson (West Tennessee) REC and associated fields used in COMET-Farm and Field to Market scenarios for estimating soil organic carbon sequestration and Soil Conditioning Index values.

Figure 2.10. Ames REC and associated fields used in COMET-Farm and Field to Market scenarios for estimating soil organic carbon sequestration and Soil Conditioning Index values.

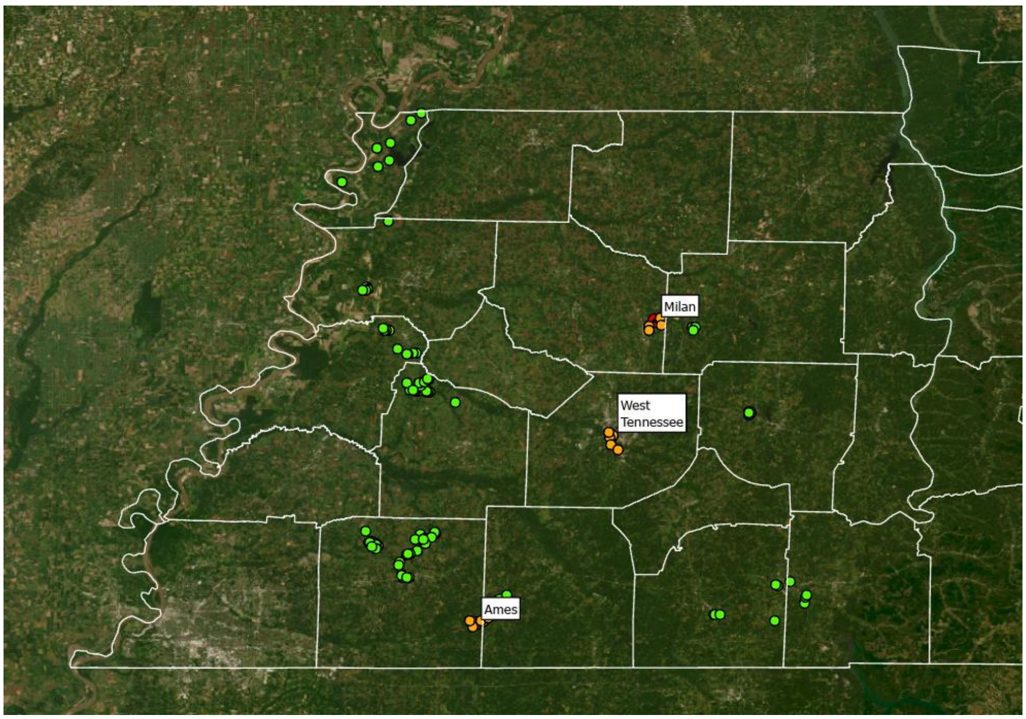

Figure 2.11. Private fields used in COMET-Farm and Field to Market scenarios for estimating soil organic carbon sequestration and Soil Conditioning Index values.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2.2)$$

where,

s = silhouette coefficient

i = location

a = mean dissimilarity of i to all other locations in cluster A

b = mean dissimilarity of i to all other locations in nearest cluster B

The silhouette coefficient is a common way to assess the results of data clustered using a k-means algorithm, (Bunkombe, Scenki, Szlatenyi, Czako, & Láng, 2023; Di & Wang, 2023; Pedregosa et al, 2011; Yuan et al., 2022) like that of the ICUC algorithm in ArcGIS Pro. The Scikit-Learn machine learning Python library contains a silhouette analysis metric referred to as silhouette score (Pedregosa et al., 2011) which can be used to calculate the average silhouette coefficient for each model output based on the samples within each cluster. The Extract Multi Values to Points tool in ArcGIS Pro was used to record the cluster ID for each field location on private farmland and the three RECs in West Tennessee based on each model output (Section 2.4.1). These labels were then utilized to assess the ability of the model to recognize correlations between the input layers and generate class assignments to characterize areas of similar soil and climate conditions using the silhouette score for each model. The k-means algorithm in the scikit-learn Python library was also executed directly on the input dataset for the field points to compare to the clustering results from the ICUC models. Each k-means algorithm was executed with randomly-initialized cluster means and with the number of clusters compared for k=3, 5, 7, and 10. The silhouette coefficients were calculated using the same silhouette analysis metric with the labels determined by the Scikit-Learn k-means models.

2.3.6.3 COMET-Farm Scenarios

COMET-Farm Version 4.1 (USDA NRCS and Colorado State University, 2023) was also utilized in the model evaluation step to return site-specific estimations for carbon sequestration and emissions over a ten-year period for all fields. COMET-Farm reports were generated for the fields on private farmland and the fields located at each of the three

UTIA RECs in West Tennessee. The results from COMET-Farm were then compared based on the cluster assignments of the geospatially-correlated regions determined by the ICUC models for each crop in order to assess how similar the SOC sequestration estimations were within the correlated regions. It was anticipated that farms in the spatially-correlated regions would have similar SOC values as their correlated REC(s) if they have comparable crop production and management practices used over similar periods of record. The most common practices compensated for in carbon markets are reduced tillage, no tillage, and cover cropping. The first Milan No-Till Crop Production Field Day occurred in 1981, and since then Tennessee has historically been a leader in the adoption of conservation tillage practices used in agricultural production (AgResearch and Education Center at Milan, 2024). A survey conducted by the Tennessee Department of Agriculture in 2023 reported 90% of surveyed crop producers use no-till, while only 60% of surveyed crop producers use cover crops, although from 2014-2022 Tennessee ranked second in cover crop enrollment in the US (Boyer, Cavasos, Smith, & Walker, 2023).

All COMET-Farm scenarios for fields on private farmland and at Milan, Jackson, and Ames were modeled with the following assumptions: the parcel was lowland non-irrigated before 1980, the parcel was not enrolled in the Conservation Reserve Program at any time before 2000, and the management for 1980-2000 was non-irrigated, no-till, with annual crops in rotation. The baseline scenario for 2000-2023 was defined as a non-irrigated field with no tillage, and three different baseline scenarios were modeled for continuous cropping of each of corn, cotton, and soybeans. These baseline scenarios were modeled to compare SOC sequestration estimates for each crop with its associated future management scenario. The three future management scenarios for the ten-year period of 2024-2034 were defined as a crop rotation with each of the cash crops and cereal rye as a cover crop between cash crop growing seasons. Estimated values of SOC sequestration for the baseline period, future period, and difference between baseline and future were recorded for each of corn, cotton, and soybeans for each field in the dataset.

2.3.6.4 Fieldprint Calculator for Soil Conditioning Index Values

The Soil Conditioning Index (SCI) is a Microsoft Excel-based model which utilizes data on soil texture, crop rotation, crop yield, field operations, and erosion rates to estimate a directional measure of SOC for a given defined cropland area. The SCI is based on the effects of organic matter returned to the soil, field operations, and erosion on net gains and losses of SOC sequestration. The Field to Market (FTM) Fieldprint Platform, first developed in 2006 as a collaborative effort to assess agricultural production and supply chains, utilizes the SCI directly through the FTM Soil Carbon Metric in the FTM Fieldprint Calculator. The Fieldprint Calculator utilizes the USDA Water Erosion Prediction Project (WEPP) Model v2.0, USDA Wind Erosion Prediction System (WEPS) Simulation v5.0, and USDA SSURGO 2.0 to compile data for estimating the erosion rates and soil properties needed to calculate the SCI. The SCI value associated with each Field to Market scenario is qualitative; SCI values estimated by the Fieldprint Calculator can range from [-1,1]. The magnitude of the value represents directionality, where a positive value indicates the amount of SOC in the field is increasing and a negative value indicates SOC is being depleted.

Two scenarios were developed for each field on private farmland and on the three RECs in West Tennessee: a sustainable management scenario with conservation practices and a tillage scenario without conservation practices. The two scenarios were executed for each field in order to represent the best- and worst-case management scenarios that are realistic for West Tennessee and to better compare the SCI values estimated for the entire dataset of fields. The inputs and assumptions used for all best-case scenarios ran in the Fieldprint Calculator included: crop rotation with cover crop; no-till; precision application of nutrients per the 4R's of nutrient stewardship; integrated pest management; and conservation practices including crop residue management, field borders, and grassed waterways. The inputs and assumptions used for all worst-case scenarios included: continuous single crop; no cover crop; conventional tillage; no precision application of nutrients; and no conservation practices.

2.4 Results and Discussion

2.4.1 ICUC Model Results

The ICUC model outputs produced without crop inputs are shown in Figure 2.12 and can be compared with the model outputs produced when the corn (fig. 2.13), cotton (fig. 2.14),

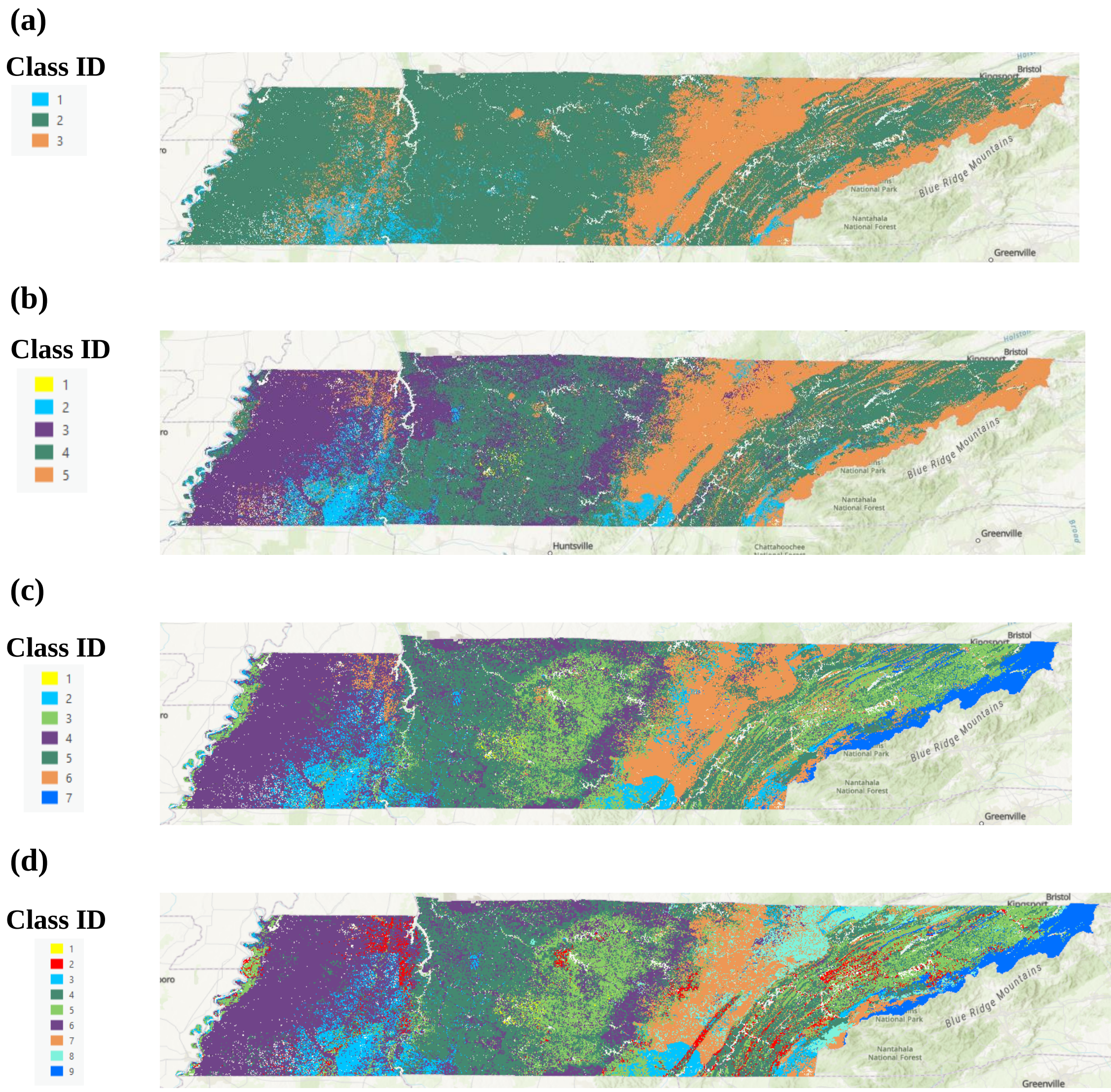

Figure 2.12. ArcGIS Pro Iso Cluster Unsupervised Classification model outputs without crop input layers with the parameter for maximum number of classes set to (a) k=3, (b) k=5, (c) k=7, and (d) k=10.

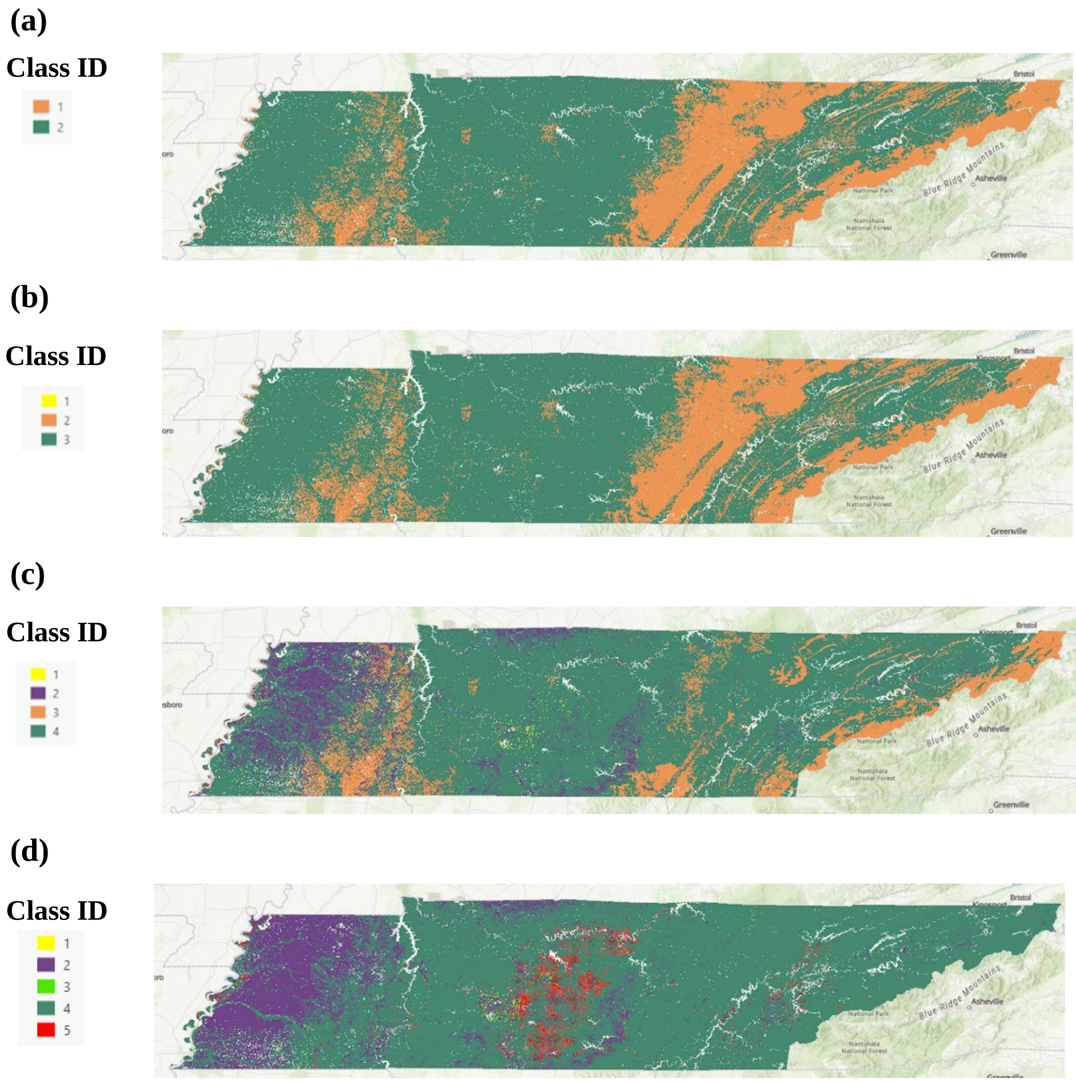

Figure 2.13. ArcGIS Pro Iso Cluster Unsupervised Classification model outputs for corn with the parameter for maximum number of classes set to (a) k=3, (b) k=5, (c) k=7, and (d) k=10.

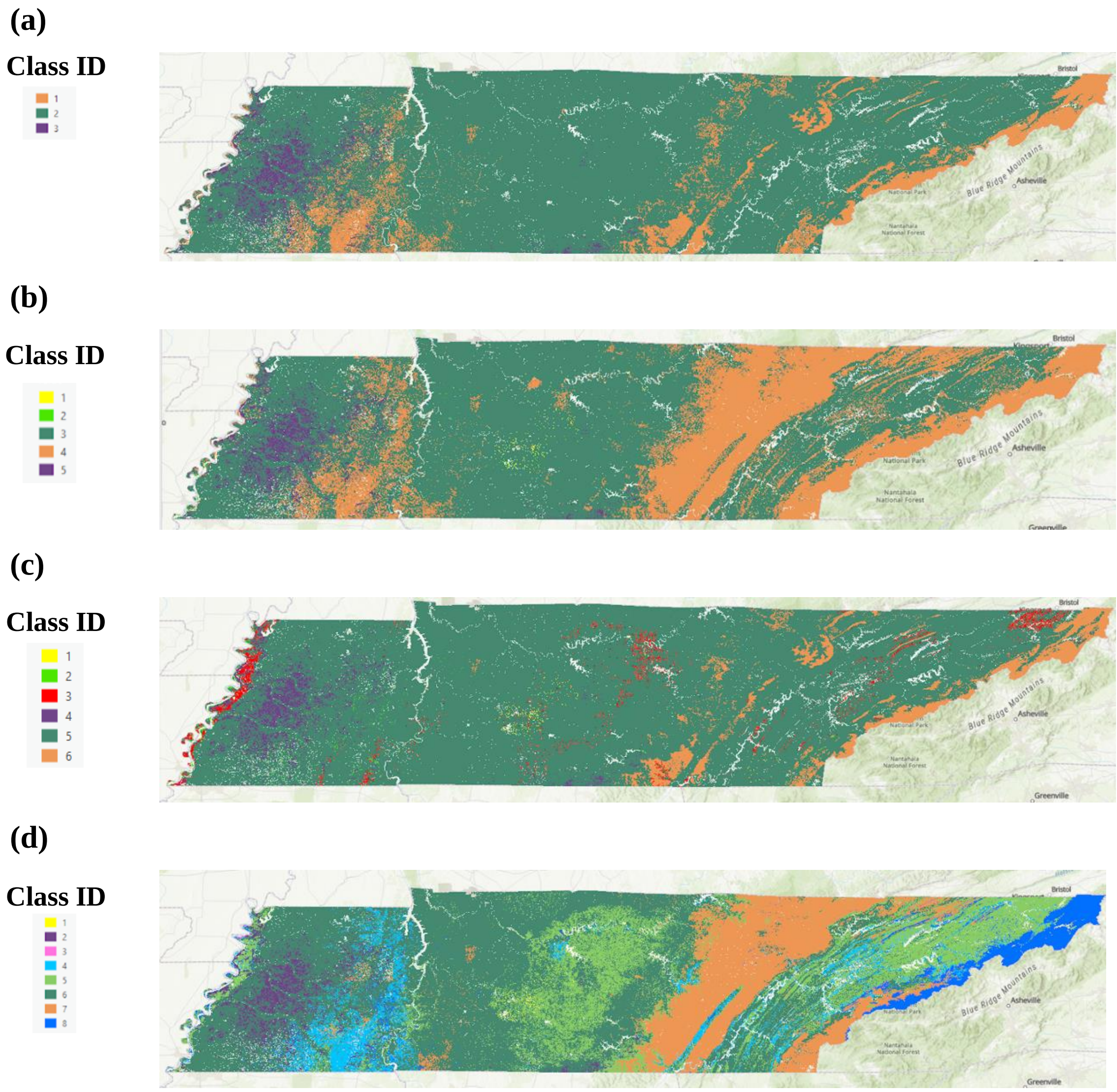

Figure 2.14. ArcGIS Pro Iso Cluster Unsupervised Classification model outputs for cotton with the parameter for maximum number of classes set to (a) k=3, (b) k=5, (c) k=7, and (d) k=10.

and soybean (fig. 2.15) crop frequency layers were used as the first input layer for each respective model. The ICUC tool was executed with the input parameter for the number of clusters set at k=3, 5, 7, and 10 to assess how changes to the number of initialized cluster means affected the output rasters. As the value of this parameter increased, the number of classes differentiated in the output raster generally increased. The values given to each class in each output raster were arbitrarily assigned and do not indicate rank or weighting. The class ID for a given class visible in multiple outputs may be different since the class IDs are arbitrarily assigned, but colors were chosen to represent each class for visual differentiation and comparison of model outputs. Similarities were observed in the model outputs even as the number of classes increased. The class represented in purple generally remained visible in the models executed without a crop input layer (fig. 2.12 b-d) as well as some of the corn (fig. 2.13 d) and cotton (fig. 2.14 d) outputs, though the size and spread of the class varied between the outputs. The same was seen for the light blue class visible in the outputs without crops (fig. 2.12 b-d) and the cotton output with the greatest number of classes (fig. 2.14 d). All outputs had a similar class 1 group, visible in yellow, for the models executed with k=5, k=7, and k=10 (fig. 2.12 b-d, fig. 2.13 b-d, fig. 2.14 b-d, fig. 2.15 b-d). The corn and soybean models generally produced the least number of classes when comparing each set of outputs based on the k parameter. The classes in West Tennessee were likely heavily influenced by soil texture and topography, as this part of the state has the lowest elevations and the highest silt content. As the value of the k parameter increased, more classes were identified in the ICUC outputs in Middle Tennessee and East Tennessee, where less dense crop frequency data in the crop layers likely led the ICUC model to assign classes based on the other input layers. Topography seemed to be most influential in determining the classes in East Tennessee along the Cumberland Plateau and in the Great Smoky Mountains. In addition to topography, the classes in Middle Tennessee seem most influenced by the sand and clay soil texture layers.

2.4.2 Correlations within Modeled Regions

The silhouette coefficient (Rousseeuw, 1987) was used to determine the model with the best representation of the correlations between the input data layers. The silhouette

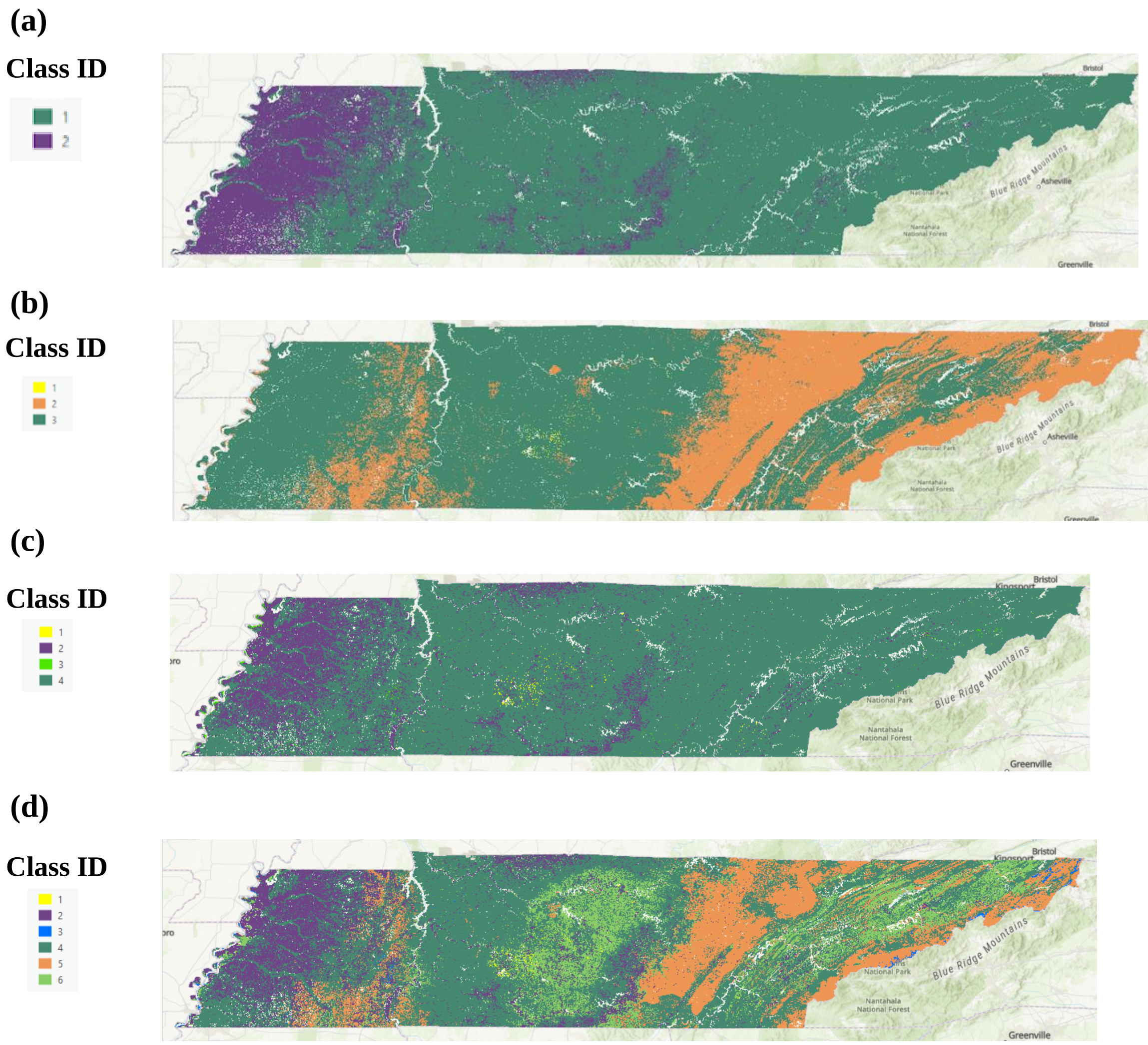

Figure 2.15. ArcGIS Pro Iso Cluster Unsupervised Classification model outputs for soybeans with the parameter for maximum number of classes set to (a) k=3, (b) k=5, (c) k=7, and (d) k=10.

coefficient is only defined when the number of clusters is greater than or equal to two. Since three of the model outputs labeled all fields with the same label, their silhouette coefficients were not defined. These silhouette score results can be compared with those calculated from the labels determined by the ICUC models in Table 2.2. For the ArcGIS Pro ICUC model outputs and the Scikit-Learn k-means model outputs, the models with the highest silhouette scores were generally those with three clusters in the output. The ICUC models produced without crop frequency layers had higher silhouette scores and can be said to have performed better at the clustering task than those produced with corn, cotton, and soybean input layers, as they labeled the fields in the dataset such that there were greater dissimilarities between clusters as compared to the dissimilarity within each cluster. The highest silhouette score was 0.667 for the ICUC model with k=3, produced without a crop frequency layer.

As silhouette coefficients approach 1, the within-cluster dissimilarity is much smaller than the smallest between-cluster dissimilarity, indicating the samples within each cluster have been labeled appropriately. As silhouette coefficients approach 0, the samples are generally close to the decision boundary between clusters, making it more difficult for the clustering model to differentiate and properly label the data for each cluster. Since the corn models either labeled all fields in the dataset with the same cluster label (as was the case for k=3 and k=5) or had a silhouette score close to 0, the corn models did not produce as representative of results as the models produced without crops and with cotton. The ICUC model outputs for soybeans all produced slightly negative silhouette coefficients. This could be influenced by the acreage utilized in Tennessee for each respective crop: about 1.55 million acres for soybeans, about 862,000 acres for corn, and just over 300,000 acres for cotton as of 2022 (USDA National Agricultural Statistics Service, 2022). There is less variability in the simplified crop frequency layer for soybeans since more cells are labeled 1 to represent locations of soybean production (Section 2.3.3.1). For this reason, the soybean ICUC model outputs likely produced clusters with a higher degree of overlap than the models without crops and the models with corn and cotton input layers, as the model had less information to obtain from the crop input layer to assist with differentiating the data into the resulting clusters.

Table 2.2. Silhouette Coefficient (SC) and Clusters (k) for each model output[1].

k_i[2]	Iso Cluster Unsupervised Classification								Scikit-Learn k-Means
	No Crop		Corn		Cotton		Soybeans		
	k_o[3]	SC	k_o	SC	k_o	SC	k_o	SC	SC
3	3	0.667	2	-	3	0.199	2	-0.102	0.615
5	5	0.409	3	-	5	0.199	3	-	0.436
7	7	0.338	4	0.004	6	0.252	4	-0.125	0.384
10	9	0.163	5	0.104	8	0.072	6	-0.125	0.391

[1] For blank SC values, all fields had the same cluster label in the corresponding model output.
[2] k_i = model input parameter for maximum number of clusters
[3] k_o = clusters defined in ICUC model output

2.4.3 COMET-Farm Results

The baseline COMET-Farm scenario was executed for each field with corn, cotton, and soybeans, under the assumption that each field was managed without tillage or irrigation. The future scenario was executed with the assumption that cereal rye was planted one week after harvesting the corn, cotton, or soybean cash crop, respectively. The results of the COMET-Farm scenarios for all fields showed an overall net increase in SOC sequestered through the adoption of the cereal rye cover crop practice in the future management scenario, as shown by the negative values in the change columns for corn, cotton, and soybeans in Table 2.3. For the model with the best performance and the highest silhouette score (produced with k=3 and without crop input layers), the average value of the change in carbon per crop scenario is shown in Table 2.4. Comparing these values with the values in Table 2.3, similar trends are shown: soybean scenarios had the greatest change in SOC sequestration as a result of cover cropping with cereal rye, followed by cotton, with corn scenarios exhibiting the lowest average change in SOC sequestration. Figure 2.16 shows a box plot visualization of the COMET-Farm results for the corn, cotton, and soybean baseline management scenario, cover crop scenario, and the difference (change) induced by the addition of the cover crop in the future management scenario. Figure 2.17 shows a box plot visualization of the COMET-Farm results for the change scenario for each crop, organized per cluster as determined by the best performing model.

Negative values in Table 2.3, Table 2.4, Figure 2.16, and Figure 2.17 indicate a net increase in the amount of carbon sequestered, while a positive value would have indicated a net loss of SOC from the fields as a result of the management practices in each respective scenario. Corn was estimated to have the highest overall amount of SOC sequestration in both the baseline scenario and the future management scenario. Soybeans were estimated to have the overall highest change in SOC sequestration as a result of incorporating a cereal rye cover crop in the future management scenario. These results are consistent with literature for Tennessee which suggests cereal rye to be a well-suited cover crop to plant between soybean growing seasons, while legumes may be better-suited for crops like corn and cotton which require higher nitrogen inputs (McClure et al., 2017).

Table 2.3. Summary of COMET-Farm estimations[1] of soil carbon [tons CO_2-e ac^{-1} yr^{-1}] for the current baseline (base) scenario, future cover crop (CC) scenario, and change in sequestration (change) scenario for all fields in the dataset.

	Corn			Cotton			Soybeans		
Statistic	Base	CC	Change	Base	CC	Change	Base	CC	Change
Minimum	-7.55	-9.09	-1.54	-0.82	-8.27	-2.41	-1.90	-7.49	-5.59
Maximum	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00
Average	-0.82	-0.98	-0.17	-0.69	-0.89	-0.26	-0.21	-0.84	-0.63

[1] Negative values represent sequestered soil organic carbon.

Table 2.4. Summary of COMET-Farm estimations[1] of average change in soil carbon [tons CO_2-e ac^{-1} yr^{-1}] per crop scenario and per cluster for the best-performing model (k=3, produced without crop frequency data layers).

	Corn	Cotton	Soybeans
Cluster 1 (n=3)	-0.15	-0.24	-0.55
Cluster 2 (n=117)	-0.17	-0.26	-0.65
Cluster 3 (n=5)	-0.07	-0.12	-0.30
Average	-0.13	-0.21	-0.50

[1] Negative values represent sequestered soil organic carbon.

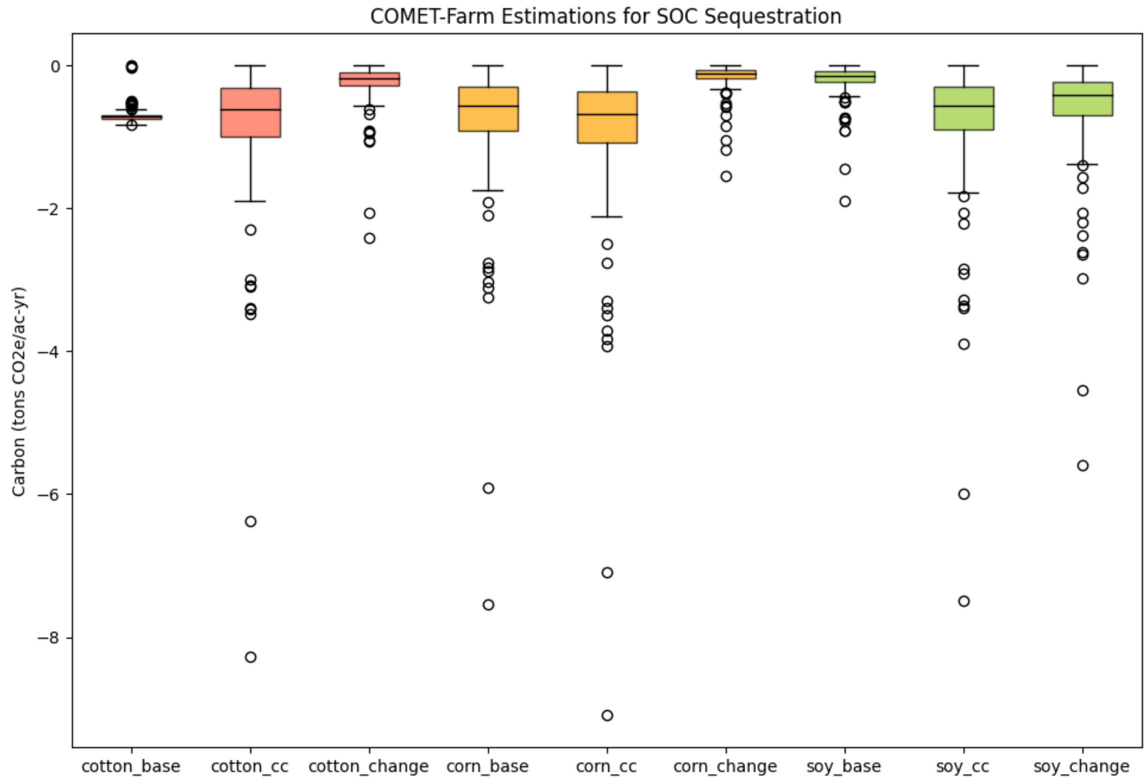

Figure 2.16. Box plot visualization of COMET-Farm estimations of carbon sequestration for the baseline (base), future cover crop (CC), and difference between base and CC (change) scenarios (n=125) for cotton (red), corn (yellow), and soybeans (green).

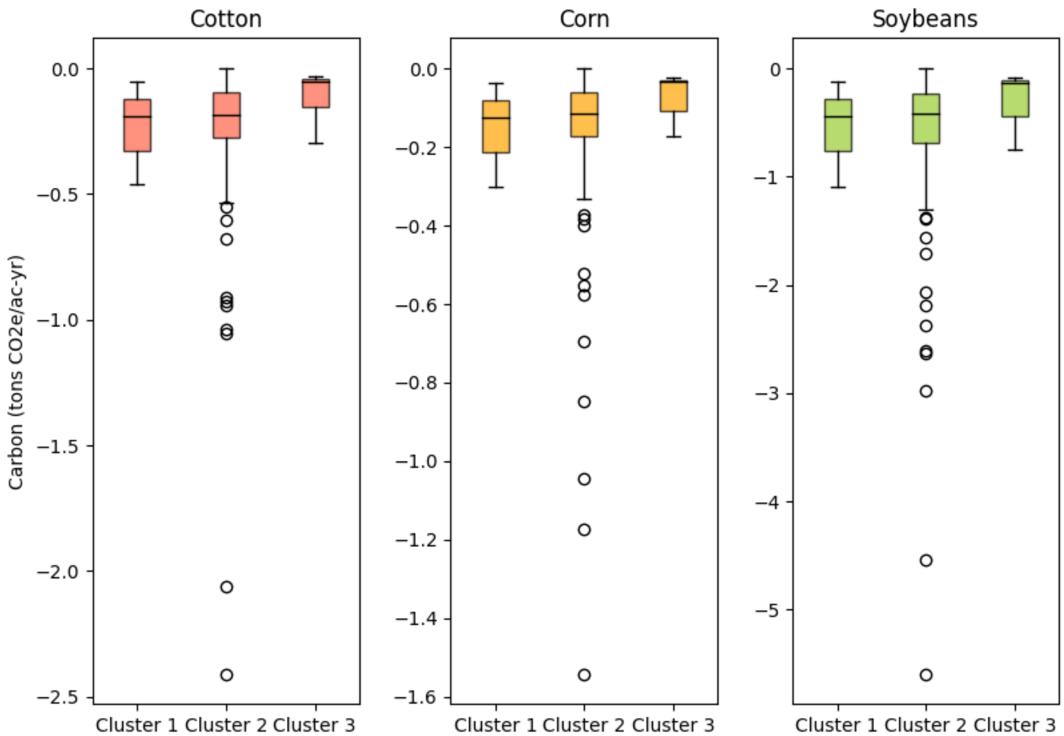

Figure 2.17. Box plot visualization of COMET-Farm estimations of carbon sequestration for the difference between the baseline and cover crop scenarios (change) per cluster, as determined by the best-performing model.

2.4.4 Field to Market Fieldprint Calculator SCI Results

The SCI represents a directional measure of SOC sequestration, and its value is qualitative. A positive value indicates SOC sequestration in the field is increasing, and a negative value indicates SOC is being depleted from the soil. The sustainable management (best-case) scenarios, which had continuous organic matter crop residue input into the soil, precision application of nutrients, and cover cropping between growing seasons, allowed for a net increase in SOC and all SCI values were positive. The worst-case scenarios, which had single cropping systems without conservation management practices and with conventional tillage, resulted in a net decrease in SOC and all SCI values were negative.

The same approach for calculating the average change in SOC sequestration for the entire field dataset using COMET-Farm data (Table 2.3) was used to calculate the average SCI value for the best-case and worst-case management scenarios (Table 2.5). The average SCI value was calculated for comparison with the overall averages for the model with the best performance based on the highest silhouette score (Table 2.6). For the entire field dataset, the average SCI value under the sustainable management scenario was 0.40, and the average SCI value under the tillage scenario without any soil conservation practices was -0.63. The average SCI values for each class defined by the best-performing model (produced with k=3 and without crop input layers) are very similar to those averages calculated for the entire dataset for the best-case and worst-case scenarios for Cluster 2. Fields covering a smaller acreage between 3-30 acres were generally the fields with an SCI value of -1 for the worst-case management scenario, which represents the highest magnitude of SOC loss as calculated by the Fieldprint Calculator. The field with the highest best-case scenario SCI value of 0.53 also had the highest worst-case scenario SCI value of -0.31. Figure 2.18 shows a box plot visualization of the SCI values for the best-case and worst-case management scenarios. Figure 2.19 shows a visualization of the SCI values for the best performing model.

2.4.5 Applications of Model Results

The results of the ICUC model with the highest silhouette score suggest nearly all of the fields in the dataset are correlated with each other, based on the input data provided to the model. Cluster 2 comprised of 117 out of the 125 fields, including those fields located at

Table 2.5. Summary of Field to Market Fieldprint Calculator Soil Conditioning Index (SCI) results for the sustainable management (best-case) scenario and non-sustainable management (worst-case) scenario.

Statistic	SCI best-case	SCI worst-case
Minimum	0.31	-1.00
Maximum	0.54	-0.31
Average	0.40	-0.63

Table 2.6. Summary of Field to Market Fieldprint Calculator Soil Conditioning Index (SCI) results per crop scenario and per cluster for the best-performing model (k=3, produced without crop frequency data layers) for the sustainable management (best-case) scenario and non-sustainable management (worst-case) scenario.

	SCI best-case	SCI worst-case
Cluster 1 (n=3)	0.35	-0.62
Cluster 2 (n=117)	0.40	-0.63
Cluster 3 (n=5)	0.41	-0.57
Average	0.39	-0.61

(a) (b)

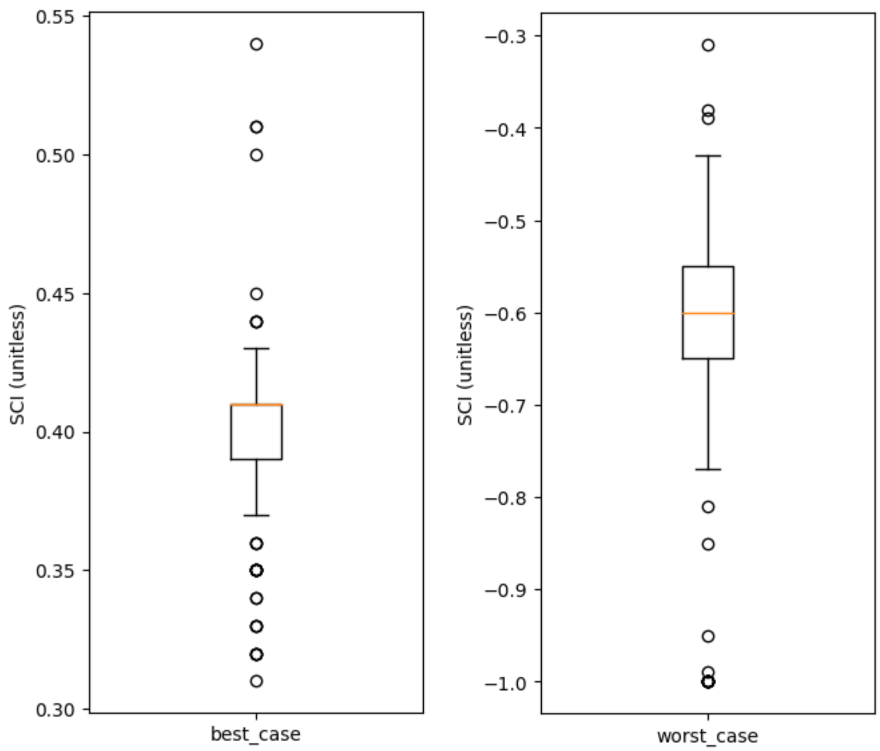

Figure 2.18. Box plot visualizations of Soil Conditioning Index (SCI) values from the Field to Market Fieldprint Calculator for the sustainable management (best-case) scenario and non-sustainable management (worst-case) scenario.

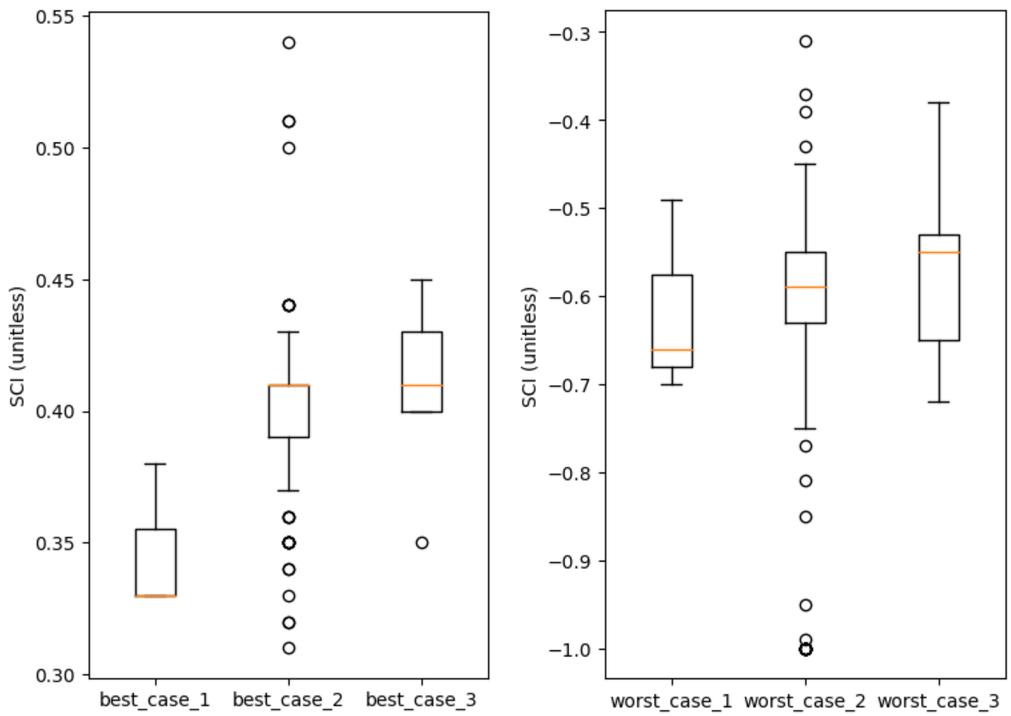

Figure 2.19. Box plot visualization of Soil Conditioning Index (SCI) values from the Field to Market Fieldprint Calculator per cluster as determined by the best-performing model (k=3, produced without crop frequency data layers).

the three RECs. In contrast, only three fields were labeled Cluster 1 and only five fields were labeled Cluster 3. This suggests historical data from the research plots at the Milan, Jackson, and Ames RECs would all be relevant to farms in West Tennessee which produce corn, cotton, and soybeans Future research following this modeling approach intends to include direct SOC measurements in conjunction with COMET-Farm estimates of SOC sequestration for ground-truthing of model outputs to further assess how climate smart land management practices can affect SOC sequestration and SCI values based on cropping system. Additionally, this work could encourage further utilization of publicly-funded RECs for developing standards to estimate SOC sequestration potential on public farmland and for implementing sustainable management practices for compensation in carbon markets. Before results would be applicable on a larger scale than Tennessee, the use of higher resolution inputs, specifically for the temperature and precipitation data, would produce outputs at greater confidence and could potentially improve correlations observed in the model results.

2.5 Conclusions

The purpose of this research was to create a geospatial model of SOC sequestration potential of agricultural land in Tennessee. This research involved developing a geospatial model using environmental, topographic, and agricultural production data layers to relate SOC sequestration potential of private agricultural land to that of publicly-funded UTIA RECs. Geospatial models were developed for corn, cotton, and soybeans using the Iso Cluster Unsupervised Classification model in ArcGIS Pro. The models were then evaluated by comparing silhouette coefficients, a relationship between the average inter- and intra-cluster distances for each model output as a measure of how differentiated and representative each cluster was in the model outputs. Baseline estimations of SOC sequestration and potential future SOC sequestration with the implementation of a cereal rye cover crop were obtained by executing COMET-Farm scenarios for 6,700 acres of private farmland and 430 acres of public farmland at the Milan, Jackson, and Ames UTIA RECs in West Tennessee. SCI values were also estimated for each field by executing Field to Market scenarios to compare ranges in SCI for sustainable management practices to tillage management scenarios which introduced soil disturbance and a net loss of SOC. The model output with the highest silhouette coefficient was produced without a crop frequency input layer and with k=3. COMET-Farm scenarios for all fields showed an overall net increase in SOC sequestration through the implementation of cereal rye as the cover crop in the future management scenario as compared to the baseline scenario. SCI values from the Fieldprint Calculator ranged from 0.32 to 0.54 for the sustainable management scenario and -1.0 to -0.3 for the tillage scenario without conservation practices. The results of the best performing ICUC model suggest nearly all fields in the dataset are geospatially-correlated with one another, as the majority of all fields were labeled within the same cluster. This suggests historical data from the research plots at the Milan, Jackson, and Ames RECs would all be relevant to farms in West Tennessee which produce corn, cotton, and soybeans. Suggestions for future research include the collection of site-specific soil data for use in ground-truthing model results, and in the identification of a greater number of sampling points in agricultural fields on public and private land to better evaluate the model's ability to represent geospatial correlations across the modeled regions.

References

AgResearch and Education Center at Milan. (2024). 33rd Milan No-Till Crop Production Field Day. University of Tennessee Institute of Agriculture. https://milannotill.tennessee.edu/wp-content/uploads/sites/100/2024/06/Final-2024_0620-Milan-No-TIll-Program.pdf.

Boyer, C., Cavasos, K., Smith, A., & Walker, F. (2023). *2023 Tennessee Nutrient Management Report*. Retrieved from University of Tennessee Department of Agricultural and Resource Economics and Tennessee Department of Agriculture: https://www.tn.gov/content/dam/tn/environment/water/nutrient-management/ntf/wr_ntf_2023-nutrient-management-report-ut-ag.pdf

Bukombe, B., Csenki, S., Szlatenyi, D., Czako, I., & Láng, V. (2023). Integrating Remote Sensing, Proximal Sensing, and Probabilistic Modeling to Support Agricultural Project Planning and Decision-Making for Waterlogged Fields. *Water, 15*(7), 1340. Retrieved from https://www.mdpi.com/2073-4441/15/7/1340

Chenu, C., Angers, D. A., Barré, P., Derrien, D., Arrouays, D., & Balesdent, J. (2019). Increasing organic stocks in agricultural soils: Knowledge gaps and potential innovations. *Soil and Tillage Research, 188*, 41-52. doi:https://doi.org/10.1016/j.still.2018.04.011

Di, X., & Wang, X. (2023). Method for Zoning Corn Based on the NDVI and the Improved SOM-K-Means Algorithm. *Journal of the ASABE, 66*(4), 943-953. doi:https://doi.org/10.13031/ja.15081

Fargione, J. E., Bassett, S., Boucher, T., Bridgham, S. D., Conant, R. T., Cook-Patton, S. C., . . . Griscom, B. W. (2018). Natural climate solutions for the United States. *Science Advances, 4*(11), eaat1869. doi:https://doi.org/10.1126/sciadv.aat1869

Grace, P., & Robertson, G. P. (2021). Soil carbon sequestration potential and the identification of hotspots in the eastern Corn Belt of the United States. *Soil Science Society of America Journal, 85*(5), 1410-1424. doi:https://doi.org/10.1002/saj2.20273

He, L., & Mostovoy, G. (2019). Cotton yield estimate using Sentinel-2 data and an ecosystem model over the southern US. *Remote Sensing, 11*(17), 2000. doi:https://doi.org/10.3390/rs11172000

Hollinger, S. E., Bernacchi, C. J., & Meyers, T. P. (2005). Carbon budget of mature no-till ecosystem in North Central Region of the United States. *Agricultural and Forest Meteorology, 130*(1), 59-69. doi:https://doi.org/10.1016/j.agrformet.2005.01.005

Huang, Y., Yu, Y., Zhang, W., Sun, W., Liu, S., Jiang, J., . . . Yang, Z. (2009). Agro-C: A biogeophysical model for simulating the carbon budget of agroecosystems. *Agricultural and Forest Meteorology, 149*(1), 106-129. doi:https://doi.org/10.1016/j.agrformet.2008.07.013

Kukal, M. S., & Irmak, S. (2018). Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production. *Scientific Reports, 8*(1), 3450. doi:https://doi.org/10.1038/s41598-018-21848-2

Lal, R. (2015). Sequestering carbon and increasing productivity by conservation agriculture. *Journal of Soil and Water Conservation, 70*(3), 55A-62A. doi:https://doi.org/10.2489/jswc.70.3.55A

Lu, C., Yu, Z., Tian, H., Hennessy, D. A., Feng, H., Al-Kaisi, M., . . . Arritt, R. (2018). Increasing carbon footprint of grain crop production in the US Western Corn Belt. *Environmental Research Letters, 13*(12), 124007. doi:https://doi.org/10.1088/1748-9326/aae9fe

McClure, A., Steckel, L., Raper, T., Sykes, V., Montgomery, G., Kelly, H., . . . Lee, J. (2017). Cover Crop Quick Facts. UT Extension, W 417. Retrieved from https://utia.tennessee.edu/publications/wp-content/uploads/sites/269/2023/10/W417.pdf

Ogle, S. M., Swan, A., & Paustian, K. (2012). No-till management impacts on crop productivity, carbon input and soil carbon sequestration. *Agriculture, Ecosystems & Environment, 149*, 37-49. doi:https://doi.org/10.1016/j.agee.2011.12.010

Oldfield, E., Eagle, A., Rubin, R., Rudek, J., Sanderman, J., & Gordon, D. (2022). Crediting agricultural soil carbon sequestration. *Science, 375*(6586), 1222-1225. doi:https://doi.org/10.1126/science.abl7991

ORNL DAAC. 2017. Spatial Data Access Tool (SDAT). ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1388

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research, 12*, 2825-2830.

Prince, S. D., Haskett, J., Steininger, M., Strand, H., & Wright, R. (2001). Net primary production of U.S. Midwest croplands from agricultural harvest yield data. *Ecological Applications, 11*(4), 1194-1205. doi:https://doi.org/10.1890/1051-0761(2001)011[1194:NPPOUS]2.0.CO;2

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65. doi:https://doi.org/10.1016/0377-0427(87)90125-7

SSURGO. (2024). Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database for Tennessee.

USDA. (2023). *A General Assessment of the Role of Agriculture and Forestry in the U.S. Carbon Markets*. Retrieved from https://www.usda.gov/sites/default/files/documents/USDA-General-Assessment-of-the-Role-of-Agriculture-and-Forestry-in-US-Carbon-Markets.pdf

USDA NASS. (2024). Cropland Data Layer: USDA NASS, USDA NASS Marketing and Information Services Office, Washington, D.C. https://croplandcros.scinet.usda.gov/

USDA National Agricultural Statistics Service. (2024). 2023 Cotton Yield. In: https://www.nass.usda.gov/Charts_and_Maps/graphics/cotnmap.pdf.

USDA NRCS and Colorado State University. (2023). COMET-Farm. Retrieved from http://cometfarm.nrel.colostate.edu.

U.S. Geological Survey, 2019, 3D Elevation Program 30-Meter Resolution Digital

Elevation Model (published 20230804), accessed at URL https://www.usgs.gov/the-national-map-data-delivery

Yuan, Y., Shi, B., Yost, R., Liu, X., Tian, Y., Zhu, Y., . . . Cao, Q. (2022). Optimization of Management Zone Delineation for Precision Crop Management in an Intensive Farming System. *Plants, 11*(19), 2611. Retrieved from https://www.mdpi.com/2223-7747/11/19/2611

Zhang, X., Izaurralde, R. C., Manowitz, D. H., Sahajpal, R., West, T. O., Thomson, A. M., . . . Williams, J. R. (2015). Regional scale cropland carbon budgets: Evaluating a geospatial agricultural modeling system using inventory data. *Environmental Modelling & Software, 63*, 199-216. doi:https://doi.org/10.1016/j.envsoft.2014.10.005

Zhou, W., Guan, K., Peng, B., Tang, J., Jin, Z., Jiang, C., . . . Mezbahuddin, S. (2021). Quantifying carbon budget, crop yields and their responses to environmental variability using the ecosys model for U.S. Midwestern agroecosystems. *Agricultural and Forest Meteorology, 307*, 108521. doi:https://doi.org/10.1016/j.agrformet.2021.108521

Zhu, Z., Bouchard, M., Butman, D., Hawbaker, T., Li, Z., Liu, J., . . . Zhu, Z. (2011). *Baseline and projected future carbon storage and greenhouse-gas fluxes in the Great Plains region of the United States: U.S. Geological Survey Professional Paper 1787*. Retrieved from https://pubs.usgs.gov/pp/1787/

Zhu, Z., & Reed, B. C. (2012). *Baseline and projected future carbon storage and greenhouse-gas fluxes in ecosystems of the Western United States: U.S. Geological Survey Professional Paper 1797*. Retrieved from https://pubs.usgs.gov/pp/1797/

Zobeck, T. M., Crownover, J., Dollar, M., Van Pelt, R. S., Acosta-Martinez, V., Bronson, K. F., & Upchurch, D. R. (2007). Investigation of soil conditioning index values for southern high plains agroecosystems. *Journal of Soil and Water Conservation, 62*(6), 433-442.

CHAPTER III:
CONCLUSIONS AND FUTURE WORK

3.1 Introduction

This research used ESRI ArcGIS Pro software to develop a geospatial model of defined geospatially-correlated agricultural regions across Tennessee based on factors influencing SOC sequestration potential for future implementation in US carbon markets. The first objective was to identify appropriate input data and develop a geospatial model to relate SOC sequestration potential on publicly-funded research stations to private farmland. The second objective was to evaluate the geospatial model's ability to predict geospatially-correlated regions across Tennessee based on the provided environmental inputs and to relate SOC sequestration potential at fields within the defined regions. The intended application of the model results could allow farmers within a spatially-correlated region to a given REC to utilize the historic data collected at that research center to estimate the SOC sequestration potential of their farm, rather than being required to sample the soils in each acre of their cropland. The model could also be used to determine which sustainable management practices should be enacted at their site to increase SOC sequestration, which would be driven by the practices utilized at their most correlated REC. In this way, this model can act as a tool to guide management decisions and assess the potential monetary value to be earned if a farm were to join a carbon credit system after implementing sustainable management practices. Summaries of research findings for the development of the geospatial model and the evaluation of the geospatial model results are presented in this chapter.

3.2 Research Findings

3.2.1 Conclusions for Objective 1: Model Development

Using ArcGIS Pro, a geospatial model was developed using input layers based on the selected spatially-available environmental data. Input parameters were optimized and pre-processed to have the same spatial resolutions, projected coordinate systems, and ranges of values normalized from [0,1] to ensure the inputs with larger ranges would not overpower the model output. When resampling the temperature and precipitation raster layers to the higher resolution of the crop production frequency layers, the value of one pixel in the original layer was applied to the smaller pixels making up the same area in the resampled raster. The Iso Cluster Unsupervised Classification (ICUC) model was utilized in ArcGIS Pro with and without crop production frequency layers for corn, cotton, and soybeans in order to better understand how the production information for each crop influenced model results. Models were developed with varying numbers of initialized cluster means (k), with k=3, 5, 7, and 10 in order to compare clustered regions between outputs.

3.2.2 Conclusions for Objective 2: Model Evaluation

The silhouette coefficient was calculated for each ICUC model output to evaluate the average within-cluster and between-cluster dissimilarities for each cropping system and each k-value. The model with the highest level of performance would have a silhouette coefficient close to 1, as this value would indicate low within-cluster dissimilarity and high between-cluster dissimilarity. The best-performing model output using this metric had a silhouette coefficient of 0.667 and was produced with k=3, without a crop frequency layer. SOC sequestration for this model ranged from -0.07 to -0.65 tons CO_2-e ac^{-1} yr^{-1}. SCI values for this model ranged from 0.39 to -0.61 for the best-case and worst-case management scenarios. The results of the best performing ICUC model suggest nearly all fields in the dataset are geospatially-correlated with one another, as the majority of all fields were labeled within the same cluster. These findings suggest the data from the research plots and crop variety trials managed at the Milan, Jackson, and Ames RECs would all be relevant to the farms in West Tennessee producing corn, cotton, and soybeans. These findings also suggest the utilization of the historical data from the Milan, Jackson, and Ames RECs to estimate SOC sequestration potential of agricultural land in Tennessee could assist with further development of MRV efforts in this region.

3.3 Research Limitations

Assumptions were made when resampling input raster layers from their original resolutions to 30 m to match the crop frequency input layers. The single value of a given larger pixel in the original, lower-resolution raster would not be fully representative of the smaller pixels within the same area in the resampled, higher-resolution raster. Lower-resolution input layers were resampled to allow the model results to determine clustered regions at the same 30 m resolution as the crop data layers. The model results at this higher resolution would provide farmers with more detailed information at the regional scale when determining the most correlated REC to their specific field and therefore more detailed information about the estimated SOC sequestration potential of their fields in relation to the historical data at the REC. That said, the results of the model outputs assume a greater degree of accuracy than the resampled data actually represents. Additionally, there are more inputs that would benefit the development and evaluation of this geospatial model, including site-specific management data on tillage, cover cropping, and nutrient management practices, though these datasets are not currently spatially available at the scale required for this modeling approach.

The lack of site-specific ground-truth data to validate model results and the COMET-Farm outputs for SOC sequestration, estimated for the baseline and future change scenarios, were additional limitations to this research. The SCI results from the Fieldprint Calculator cannot be used to quantify the amount of organic matter in the soil for the specified area of interest, or to determine a target amount of SOC sequestration, but rather represent the directional value of organic matter accrual/depletion and the significance of the increase/decrease based on site characteristics and management. Model evaluation could also be benefitted by including fields located at RECs in Middle Tennessee and East Tennessee, which would provide a more robust field dataset and assist in determining the correlations between private farmland and fields at the RECs across the entire state.

3.4 Research Implications

In addition to allowing farmers within a spatially-correlated region to a given REC to utilize its historical data to estimate the SOC sequestration potential of their farms, the historical data could help farmers identify other potentially suitable agricultural practices. Data pertaining to cover crops, variety trials, and management strategies for crop pests, weeds, diseases, irrigation rates, and fertilizer application rates are among the additional categories of research currently investigated at the RECs. In this way, the model results can be applied to a multitude of benefits for farmers and are not limited to understanding SOC sequestration and appropriate sustainable management practices currently compensated for in carbon markets.

3.5 Future Work

This methodology for the development and evaluation of a geospatial model for use in carbon market MRV efforts serves to encourage increased use of publicly-funded research stations as model farms for carbon market standards and for comparison when allocating carbon credits for sustainable management practices implemented on private farmland. This methodology could benefit from the addition of model inputs containing more specific management data as technologies advance the development and display of these datasets at appropriate spatial resolutions. Average daily temperature and average daily precipitation data layers averaged for each crop growing season were considered during the model development stage of this research, since the length of the growing season varies for the row crops of interest in this study. Daily temperature and daily precipitation layers were instead averaged on an annual basis due to the use of cover crop scenarios in COMET-Farm and the Fieldprint Calculator, in which crop production was maintained throughout the entire year, whether for the main cash crop or the winter cover crop between main growing seasons. Temperature and precipitation data averaged by growing seasons could allow for future iterations of this model to be further refined to determine geospatially-correlated regions of SOC sequestration potential for the business-as-usual crop production scenarios without the utilization of additional sustainable management practices. These outputs for business-as-usual scenarios could then be compared with monocropped systems, cropping systems with crop rotations, and the utilization of other cover crops to maintain SOC sequestered throughout the entire year in conjunction with additional sustainable management practices to gain a larger picture of the effects of these agricultural practices on SOC sequestration potential. Future work using this methodology also intends to expand this approach to analyze effects of climate smart land management practices in other Southern states and to encourage further utilization of publicly-funded research for developing standards for estimating SOC sequestration potential and compensation for implementing sustainable management practices in US carbon markets.

VITA

Madeline is originally from Philadelphia, Pennsylvania. She obtained her Bachelor of Science in Biosystems Engineering from the University of Tennessee, Knoxville in May 2022. She continued her studies at the University of Tennessee in pursuit of a Master of Science degree in Biosystems Engineering in August 2022. After graduate school, she will begin a full-time position as an environmental engineer, where she will continue to work with environmental data and will strive to improve water quality through stormwater and wastewater treatment applications.