Utilizing Machine Learning to map soil organic carbon in Rangelands: A case study of Umzimvubu, Eastern Cape, South Africa

Submitted by Milcah Cherono Kirinyet

to the University of Exeter as a thesis for the degree of

MSc Applied Data Science (Environment & Sustainability)

[AUGUST,2023]

Supervisor: Dr Andrew Cunliffe, Oppenheimer Senior Research Fellow, University of Exeter

Co-Supervisors: Dr Tom Powell, Oppenheimer Impact Fellow, University of Exeter

Dr Perushan Rajah, MERL Senior Manager at Conservation South Africa.

This dissertation is part of the Oppenheimer Programme in African Landscape Systems (OPALS) in partnership with Conservation South Africa (CSA).

I certify that all materials in this thesis which are not my own work, have been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Table of Contents

List of Figures

Table

Abstract

Regenerative agriculture has the potential to sequester atmospheric carbon and mitigate the effects of climate change, thereby safeguarding the provision of ecosystem services under changing climatic conditions. To inform various stakeholders on the commercial feasibility of soil management practices potential to sequester carbon through regenerative agriculture and the potential to attract carbon credit through climate finance, soil carbon stocks should be estimated using approaches that combine soil carbon models with soil samples to accurately quantify soil carbon sequestered due to interventions implemented. Measuring rangeland soil carbon in Africa is crucial for understanding the potential of carbon markets to support improved management. Previous studies have shown different performance of models in terms of accuracy and precision in quantifying soil organic carbon. This study aimed at exploring three machine learning models, namely Random Forests, support vector machines and XGBoost and evaluating their performance in predicting soil organic carbon over the Umzimvubu catchment in South Africa. Utilizing 87-point soil samples taken at near surface (0-20) cm, soil properties from legacy soil maps and environmental covariates that represent the terrain, bioclimate and vegetation indices from remotely sensed data, the models were trained and evaluated on their performances based on 3-fold cross validation and 80% - 20% training and test split ratio using metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE) and R squared. The random forest model demonstrated better performance in prediction accuracy compared to the other two models and was then used to generate a predictive map of SOC distribution. Among the variables used, mean EVI, elevation, soil clay percentage, mean NDVI and the precipitation of the wettest month were the major contributors to the variability in soil organic carbon. The predicted SOC map was compared with a global map of SOC and the predicted national level SOC map, revealing low SOC levels in the northeastern part of the study area.

List of Abbreviations

CSA - Conservation South Africa

DWS - Department of Water and Sanitation (DWS), South Africa

ERS - Environmental & Rural Solution

FAO - Food and Agriculture Organization of the United Nations

EVI - Enhanced Vegetation Index

NDVI - Normalized Difference Vegetation Index

SANBI - South Africa National Biodiversity Institute

SOC - Soil Organic Carbon

UCP - Umzimvubu Catchment Partnership

VCS - Verified Carbon Standard

1. Introduction

1.1. The Role of Soil Organic Carbon and Regenerative Agricultural Practices in climate change mitigation efforts

The pressing global need to tackle climate change necessitates effective strategies, particularly those that combine the reduction in greenhouse gas emissions with the enhancement of carbon storage (Griscom *et al*., 2017; Amelung *et al*., 2020). Soil Organic Carbon (SOC) is a major contributor to the global carbon budget and is a crucial component of terrestrial ecosystems. It regulates soil function, provides energy for soil microorganisms, and influences soil water retention and nutrient availability (Billings *et al*., 2021; Georgiou *et al*., 2022; Grunwald, 2022; Nyawira *et al*., 2021). Consequently, understanding and managing SOC have emerged as crucial elements in sustainable land management practices and strategies to mitigate climate change.

However, climate change and human activities associated with changes in land use and land cover have significantly impacted the decline in soil carbon levels over the past two centuries. Research indicates that alterations in land cover and use result in the depletion of natural vegetation and disruption of the balance between carbon input and output from the soil, consequently leading to a reduction in soil carbon storage (Deng *et al*., 2016; Padbhushan *et al*., 2022). These effects, contingent upon the extent and length of land use modification, can endure for extended periods, ranging from several decades to centuries. The quantity of SOC in a region is susceptible to these changes, which can disrupt SOC cycling and lead to considerable changes in terrestrial carbon storage (Paustian *et al*., 2019). SOC depletion can result in soil degradation, reduced agricultural productivity and heightened greenhouse gas emissions. Given these challenges, it becomes evident that promoting improved soil management practices that encourage carbon sequestration in stable carbon pools is crucial to counteract this negative trend and address climate change (Georgiou *et al*., 2022). Despite the widely recognized significance of SOC in ecosystem processes, further research is needed to accurately estimate the capacity for regional soil carbon sequestration. This need is particularly critical given the dynamic nature of land-use and climatic patterns. Regular monitoring of SOC is essential for a comprehensive understanding of the carbon cycle and the development of effective strategies to combat climate change (Poeplau and Dechow, 2023).

Recently, there has been increasing attention on carbon sequestration, especially in the context of regenerative agricultural practices. According to Olson (2010), soil carbon sequestration involves the "Process of transferring carbon dioxide from the atmosphere into the soil through plants, plant residues and other organic solids, which are stored or retained as part of the soil organic matter (humus)". One significant advantage of regenerative agriculture is its potential to absorb carbon dioxide from the atmosphere and store it as soil organic matter. These practices play a critical role in preserving and enhancing soil carbon content, thereby mitigating the effects of climate change, promoting biodiversity, and combating soil degradation (Vargas-Rojas *et al*., 2019; Rumpel and Chabbi, 2021; Wiltshire and Beckage, 2022; Khangura *et al*., 2023; Havens, 2021).

In the context of emerging carbon-based economies and markets, soil carbon plays a crucial role in counterbalancing a substantial portion of yearly fossil fuel emissions. The concept of soil carbon credits has been steadily gaining traction, and the convergence of voluntary carbon markets and agriculture has become increasingly significant in light of the ongoing global climate crisis. Numerous studies have proposed that the process of soil carbon sequestration has the potential to make a considerable contribution to the reduction of emissions (Lal, 2004; Zomer *et al*., 2017; Grace and Robertson, 2021). Therefore, promoting these practices could yield significant benefits to both the environment and agriculture.

1.2. Implementing SOC Sequestration in South Africa: Challenges and Opportunities

While the role of SOC and regenerative agriculture is a global concern, South Africa's unique landscape presents invaluable insights into the challenges and opportunities related to soil management and carbon sequestration. Rangeland degradation in South Africa, characterised by overgrazing, soil erosion, reduced vegetation cover, and declining biodiversity, has emerged as a significant environmental issue (Franke and Kotzé, 2022; Gusha *et al.*, 2023). To address these challenges, comprehensive strategies such as bush clearing, reseeding, erosion prevention, and optimised grazing practices are being employed (Mudau *et al*., 2022). Controlled grazing, in particular, has been recognized for its potential to enhance livestock productivity and increase carbon capture capacity (Balehegn *et al*., 2021; Franke & Kotzé, 2022; Wang *et al*., 2022). However, implementing and scaling up these regenerative techniques on a larger scale requires substantial investment, community support, technical assistance, and access to technology (Lunn-Rockliffe *et al*., 2020; Paul *et al*., 2023).

In South Africa, small-scale farming systems dominate the agricultural landscape (Flynn, 2019), and this brings unique challenges. Despite the gradual adoption of soil carbon farming techniques, barriers such as exclusion of carbon capture projects from certain compliance markets and lack of reliable methods to accurately measure and verify soil carbon sequestration for carbon markets or credits persist (Schilling *et al*., 2023). Overcoming these technical and financial obstacles is crucial for the successful initiation and sustainability of carbon projects in the region. These obstacles highlight the need for accurate, reliable, and cost-effective methods for quantifying the amount of carbon sequestered by different land management strategies. Such precision is pivotal for facilitating carbon sequestration and assessing its effectiveness (Paustian *et al*., 2019; Smith *et al*., 2020). Furthermore, understanding the interplay between cost, adoption rates, policy support, and other factors is essential for evaluating the feasibility of various agricultural practices like regenerative agriculture (Martin *et al*., 2021; Keesstra *et al*., 2020).

To address these challenges, various governmental and non-governmental organisations are actively working on solutions. One such organisation is Conservation South Africa (CSA), a non-governmental organisation that prioritises sustainable land management and plays a leading role in carbon sequestration initiatives in South Africa. As an affiliate of Conservation International, CSA is dedicated to preserving the country's ecosystems, biodiversity, and natural resources while promoting sustainable growth and improved living conditions for local communities. They strive to promote sustainable land management in South Africa, in areas such as the Eastern Cape, through collaborative conservation efforts, sustainable livelihood initiatives, climate change adaptation, and capacity-building (CSA, 2023). CSA actively collaborates with farmers, traditional authorities, community leaders, and local governments to identify, design, and implement appropriate land-use actions that promote carbon sequestration and sustainable land management practices, such as rotational grazing and on-farm tree planting, to sequester carbon in the soil (CSA & UCP, 2022; CSA,2023). By working closely with farmers and promoting sustainable land management practices, CSA aims to register carbon projects that contribute to carbon sequestration and reduce greenhouse gas emissions.

Recognizing the complexity of quantifying and managing carbon sequestration, CSA adheres to globally recognized standards and methodologies. One such methodology approved for use in the Verra carbon credit system is VM0032 “Methodology for the Adoption of Sustainable Grasslands through Adjustment of Fire and Grazing” (Verra, 2015). The VM0032 methodology presents a new and innovative approach that centres its attention on the removal of soil carbon. It places significant importance on sustainable management of grasslands and measurement of the reduction in greenhouse gas emissions that occur as a result of alterations in grazing density and fire regime (Verra, 2015). VM0032 aims to incentivise farmers to adopt a suite of such practises in order to improve soil health and mitigate climate change. The methodology establishes criteria for eligible project activities and provides guidance on setting an appropriate baseline and demonstrating additionality (Verra, 2015). It also sets out requirements for project design, management, and monitoring plans. The project design phase of VM0032 is essential for maintaining the integrity of carbon removal projects. It outlines the selection of suitable carbon removal activities and establishes a baseline against which the project's carbon removal performance will be evaluated. This methodology employs a comprehensive carbon accounting framework to quantify the amount of carbon dioxide removed by the project, which is crucial in determining the carbon credits that can be traded on the voluntary carbon market.

CSA's project is strongly committed to transitioning towards regenerative grazing practices in the Umzimvubu catchment area of the Eastern Cape. Hawkins *et al*. (2021) projected that effective management of grazing and burning could lead to an approximate 20% increase in soil organic carbon in the Umzimvubu catchment. However, the implementation of these practices requires significant funding. Therefore, the registration of carbon offset projects is crucial to provide the necessary financial support to facilitate this transition, creating a system where sustainable land management can be effectively supported by climate finance. The focus of these efforts is the Umzimvubu Catchment, a river system where overgrazing and land degradation pose threats to carbon sequestration potential and water quality. The Umzimvubu Catchment Partnership Program (UCPP), CSA included, is dedicated to conserving the river system and restoring the catchment area in a sustainable and economically supportive manner. This includes a focus on carbon sequestration and rangeland stewardship (SANBI and Wildlands Conservation Trust, 2015).

To actualize the projected gains in the Umzimvubu catchment and align with the VM0032 guidelines, CSA is now developing a structured plan in the form of the Project Description Document (PDD). The PDD is a critical component for establishing a carbon offset project, demonstrating its feasibility, sustainability, and compliance with carbon credit protocols such as the Verified Carbon Standard (VCS) supervised by Verra. The PDD outlines the project's approach to carbon dioxide sequestration and greenhouse gas emission reduction resulting from the project, detailing the project location, the stakeholders involved, the timeline, and the expected environmental and social benefits. CSA is doing this in partnership with other stakeholders like the Oppenheimer Program in African Landscape Systems (OPALS), an innovative initiative dedicated to supporting sustainable land management and environmental conservation in African landscapes through research, capacity building, and collaboration that this research is part of.

1.3. Methodologies and Tools

To leverage carbon finance opportunities through certification programs like the Verified Carbon Standard (VCS) by Verra, CSA is currently utilizing process-based models of soil carbon dynamics such as Rothamsted Carbon Model (RothC), SNAP and SNAPGRAZE, as stipulated by the VCS VM0032 methodology to quantify soil organic carbon for their baseline projects. RothC is a model that primarily focuses on the turnover of organic carbon in non-waterlogged topsoil and takes into consideration various factors including soil type, temperature, soil moisture and plant cover (Coleman and Jenkinson, 1996). The SNAP model is employed to quantify the influence of grazing management on soil carbon dynamics. It provides a means to assess the effects of different grazing practises on soil carbon sequestration (Ritchie, 2014). SNAPGRAZE, which is derived from SNAP model allows for the evaluation of grazing management practises and their influence on soil carbon sequestration and forage production (Ritchie, 2020). The SNAP model and its extension SNAPGRAZE are known for their robust simulation of soil nitrogen and organic carbon dynamics over time. Notably, they have been successfully applied in the tropical grasslands of the Serengeti National Park (Ritchie, 2014, 2020). While these models offer valuable insights, their use in isolation may present limitations as they may not entirely capture the complex processes influencing soil carbon dynamics. Their over-reliance on simplified relationships and the need for extensive data for calibration could lead to inaccuracies or biases in their predictions.

Alternatively, VCS VM0032 methodology permits the utilisation of methodologies that have been published in peer-reviewed scientific literature. Given the potential limitations of the process-based models currently being utilised, it is prudent to explore alternative methodologies that may provide more robust and accurate predictions. In this regard, we are considering integrating recent advancements such as machine learning (ML) models to evaluate the changes in SOC stock resulting from the implementation of conservation initiatives in the Umzimvubu catchment. The accurate quantification, mapping, and estimation of future SOC sequestration rates heavily rely on the ability of these technologies to understand the effects of changes in land use and management practices on soil carbon storage. Machine learning models offer a promising alternative as detailed in the review by (Grunwald, 2022), they are well suited to handle complex, non-linear relationships and interactions among variables. They can capture high-dimensional, non-linear patterns in data, which makes them excellent at predicting complex variables like soil carbon. Their ability to learn from data reduces the need for extensive calibration that traditional models require. Furthermore, ML models can handle large datasets, making them ideal for integrating various data sources for improved predictions. Additionally, ML models can readily incorporate diverse types of data, including climate, soil, topography, land use, and management practices to predict SOC content (Meier *et al.*, 2018; Wadoux, Minasny and McBratney, 2020). This flexibility can allow them to capture the influence of a broader range of factors on soil carbon dynamics than traditional process-based models, potentially improving the accuracy and reliability of predictions.

Integrating machine learning models into CSA's existing methodological framework could potentially overcome the limitations of the current models. While the adoption of such models does come with its own challenges, including the need for high-quality and high-volume data, the benefits they offer in terms of improved prediction accuracy and capability to handle complex relationships among variables could significantly enhance CSA's understanding of soil carbon dynamics. However, it is important to note that the development of ML technologies in soil carbon quantification and mapping is still in its early stages, primarily due to limitations in data availability (Grunwald, 2022). Specifically, the Eastern Cape region of South Africa faces challenges in accurately quantifying soil carbon stocks. These challenges include linking SOC stocks with management practices, measuring SOC content, bulk density, and soil depth for all soil groups, and addressing anthropogenic disturbances that significantly alter the distribution of SOC across local biomes (Seboko *et al*., 2021; Odebiri *et al*., 2023; Paterson *et al*., 2015). These obstacles emphasize the need for extensive data collection efforts and the use of sophisticated models capable of capturing the complexity and diversity of these factors. Despite these challenges, the potential of AI and ML in enhancing our understanding and management of SOC sequestration remains significant.

Improved precision in SOC quantification enables reliable demonstration of various impacts of soil management practices, bolstering the credibility and validation of such projects in climate finance markets. In addition, accurate SOC estimates guarantee that the genuine benefits of regenerative agricultural practices are accounted for, potentially leading to an elevated valuation of carbon credit. Ultimately, this strategy, combining established carbon offset methodologies with the integration of machine learning innovations, provides a strong foundation for CSA's shift towards regenerative agricultural practices funded by climate finance. This could attract substantial investments from climate finance mechanisms, ensuring the long-term financial viability and sustainability of regenerative agriculture projects in the long run.

1.4. Literature review

The urgency of addressing climate change has brought attention to the practice of carbon sequestration in regenerative agriculture as a way to offset greenhouse gas emissions. However, accurately measuring the amount of organic carbon stored in the soil presents a challenge. Soil organic carbon dynamics are complex and influenced by various biogeochemical processes. While physical sampling and measurement techniques are accurate, they are expensive and time-consuming. Consequently, different models with varying levels of complexity and mathematical structures have been developed to understand and predict SOC dynamics. This literature review was aimed at identifying the optimal methodologies that enhance soil carbon mapping. Initially, SOC stocks were estimated using simple statistical models and process-based models. Simple linear and multivariate regression models used statistical analysis to simulate the relationship between soil organic carbon content and various soil characteristics such as pH, texture, and land use (McClean *et al*., 2015; Shelukindo *et al*., 2014; Offiong and Iwara, 2012). Despite their simplicity, these models provide a straightforward approach to modeling SOC changes by identifying key variables that affect SOC dynamics and their relationships. Process-based models, on the other hand, are based on a deep understanding of the ecological and physical processes that influence SOC dynamics. These models use equations to describe carbon inputs and outputs, as well as carbon transformations within the soil, enabling the prediction of SOC changes under different conditions. Some of the models developed include the Rothamsted Carbon (RothC), Denitrification Decomposition (DNDC), and Century Ecosystem (Century) models, each designed to simulate different aspects of the soil carbon cycle. For example, the DNDC model developed by Li *et al*. (1992) has emerged as a robust process-oriented tool for simulating soil carbon (C) and nitrogen (N) biogeochemistry in agricultural ecosystems. It has evolved over time and demonstrated significant capabilities in mimicking soil carbon dynamics and their interaction with nitrogen processes. The model has been modified and tested against various field studies and production frameworks (Giltrap, Li and Saggar, 2010; Gilhespy *et al*., 2014; Zhang and Niu, 2016; Macharia, Ngetich and Shisanya, 2021; Abdalla *et al*., 2022). However, the model has recognized limitations in simulating soil hydrology, which has a significant impact on biogeochemical processes (Smith *et al*., 2020).

The RothC model was initially developed to assess carbon turnover in arable soils of the Rothamsted Long-Term Field Experiments (Coleman and Jenkinson, 1996). The model considers factors such as soil type, temperature, moisture content, and other variables that may affect the breakdown rate of organic carbon in non-waterlogged soils (Jenkinson and Coleman, 2008). Over time, its use expanded to include various ecosystems. Despite its usefulness in predicting SOC stock changes under different climate change scenarios and soil characteristics (Morais *et al*., 2019; Afzali *et al*., 2019; Nemoto, 2014), the RothC model has limitations regarding different management practices and climatic variables (Jebari *et al*., 2021).

The Century model also plays a significant role in understanding SOC dynamics. This model simulates the long-term dynamics of carbon, nitrogen, phosphorus, and sulphur in different ecosystems and has been instrumental in predicting the impacts of management techniques and climate change on SOC dynamics (Kelly *et al*., 1997; Cong *et al*., 2014; Stergiadi *et al*., 2016). Similarly, the Pasture Simulation model estimates the effects of climate change on greenhouse gas emissions and livestock production. It simulates the flow of carbon, nitrogen, water, and energy within grassland ecosystems (Graux *et al*., 2011; Pulina *et al*., 2018). However, the simplicity of these models may limit their ability to capture the complex non-linear relationships between SOC and environmental factors. Consequently, ML algorithms are increasingly being applied for digital soil mapping as they can better handle these complexities. ML models, such as Random Forests (RF), XGBoost, Support Vector Machines (SVM), and Artificial Neural Networks (ANN), have shown improved SOC predictions compared to traditional statistical models. This is mainly due to their capacity to handle non-linear relationships, consider interactions among predictor variables, and model complex SOC dynamics (Hengl *et al*., 2017; Grunwald, 2022). The shift towards ML is evident in recent literature (Lamichhane, Kumar and Wilson, 2019). For example, studies conducted by Emadi *et al*. (2020) and Mahmoudzadeh *et al*. (2020) employed ML algorithms to predict and map SOC in various regions of Iran. These studies showed that SVM, ANN, and RF models provided more accurate predictions of SOC. Similarly, John *et al*. (2020) used ANNs, SVMs, cubist regression, random forests and regression trees to quantify the impact of environmental factors and soil nutrient indicators on SOC variability in alluvial soils. The results showed that the RF, SVM, ANN, and Cubist Regression models provided accurate estimations of SOC variability. Additionally, Zhang *et al*. (2022) compared three machine learning algorithms based on tree models, including Cubist, XGBoost and RF, in predicting SOC in the dryland of Northeast and North Plain China. The study found that RF and XGBoost models outperformed Cubist in terms of predicting SOC.

Additionally, the efficacy of the random forest model in predicting rates of soil organic carbon sequestration and stocks has been investigated by Chinilin and Savin (2023), Payen *et al*. (2021) and Nabiollahi *et al*. (2019). Comparisons between Boosted Regression Trees (BRT) and RF for predicting SOC stocks conducted by Wang *et al*. (2017) and Wang *et al*. (2018) revealed that RF demonstrated superior performance. Across various investigations into SOC modeling, RF, support vector machine, XGBoost and neural network models consistently provided more accurate predictions (Matinfar *et al.*, 2021; Salehi Hikouei, Kim and Mishra, 2021). These studies collectively highlight the potential of machine learning (ML) in overcoming the limitations of process-based models and improving SOC estimates.

The advancement of ML and artificial intelligence (AI) has introduced more sophisticated methods, such as deep learning algorithms. Grunwald (2022) discusses the potential of AI in predicting SOC storage and other properties of the global carbon cycle, citing improved performance compared to traditional modeling approaches. Furthermore, the integration of satellite remote sensing and ML has been demonstrated as a valuable tool for quantifying pasture biomass and supporting the monitoring of regenerative grazing practices (Ogungbuyi *et al*., 2023). Evidence indicates that ML and AI offer a substantial advantage over traditional process-based models in predicting and validating carbon sequestration. However, the predictive performance of ML models remains data dependent (Rolnick *et al*., 2023).

Despite the promising advancements AI and ML bring, the uneven global distribution and accessibility of high-quality soil data present challenges, particularly in developing nations with inadequate data collection infrastructure. This has sparked interest in the creation of comprehensive soil databases, such as the Soil Grids project by the International Soil Reference and Information Center (ISRIC), which aims to provide global soil information free of charge. In recent years, satellite remote sensing technology has been increasingly utilized to overcome data limitations, enabling extensive and continuous soil monitoring. Remote sensing, combined with powerful statistical methods, has revolutionized digital soil mapping and prediction. It provides information on surface characteristics that are representative of soil properties, which has been used to successfully estimate biomass and carbon sequestration in various ecosystems, including forests, grasslands, and blue-carbon ecosystems (Abbas *et al*., 2020; Lanceman *et al*., 2022).

Despite its potential, remote sensing data for SOC estimation faces challenges related to data quality, resolution, and sensitivity to meteorological and environmental conditions. The geographical and temporal resolutions of different remote sensing platforms may not align with the scale of interest for a specific study, and existing legacy soil data have been criticized for being outdated and lacking crucial uncertainty data for practical application at local scales (Ahmad *et al*., 2023; Buenemann *et al*., 2023). These challenges, among others, can impact the sensitivity and precision of carbon estimation. Nevertheless, remote sensing data remains a practical and affordable method for measuring carbon sequestration across various habitats.

Despite these challenges, the future of soil carbon research appears promising as it moves towards an integrated strategy that combines established process-based models with cutting-edge ML approaches, supported by remote sensing and continually improving soil databases (Dannenberg *et al*., 2023). However, to fully leverage these evolving techniques, investment in high-quality soil data collection, particularly in developing countries, is crucial. Improvements in soil monitoring techniques using remote sensing are also essential. Considering the potential influence of local weather conditions and soil characteristics on soil organic carbon (Venter *et al*., 2021), more region-specific research is necessary. This would help refine these techniques and maximize their effectiveness in specific local contexts, such as the Umzimvubu Catchment.

In summary, the modeling of SOC sequestration has evolved significantly over the years, transitioning from process-based to regression models and now embracing ML and AI techniques. The effectiveness of these models is influenced by various factors, including their adaptability to diverse environmental conditions, the quality and quantity of available soil data, their interpretability, and their ability to account for complex SOC dynamics. Therefore, coordinated global and regional efforts to improve data collection and analysis techniques, along with region-specific research, holds the key to advancing SOC modeling and consequently, climate change mitigation efforts.

For this research, three ML algorithms were chosen for simulation of soil organic carbon based on their performance in similar studies in different regions. The random forest algorithm has been shown to provide accurate predictions of SOC in arid regions, considering the spatial variability and heterogeneity of soil properties (Sodango *et al.*, 2021; Fathizad *et al.*, 2022; Mousavi *et al.*, 2022; Zhu *et al.*, 2022; Pouladi *et al.*, 2023). For example, a study in an arid and semi-arid region of Iran used random Forest to model and map SOC at high spatial resolution, demonstrating its effectiveness in capturing the spatial distribution of SOC (Mousavi *et al.*, 2022). Similarly, the XGBoost and the support vector machine have shown promising results in accurately mapping SOC in these challenging environments (John *et al*., 2020; Lamichhane, Kumar and Wilson, 2019; Mahmoudzadeh *et al.*, 2020; Wang *et al.*, 2021, 2023; Xie *et al.*, 2022; Zhang *et al*., 2022; Zeraatpisheh *et al.*, 2022).

1.5. Aims and objectives.

The overall aim of this study is to contribute towards the establishment of a robust rangeland carbon verification framework in the Eastern Cape by exploring approaches that enhance the accuracy of mapping the spatial distribution of soil organic carbon stocks in the Umzimvubu catchment. It had three specific objectives:

1. Utilize point estimates of SOC stocks with other remotely sensed data that influence soil carbon dynamics to predict the spatial distribution of near-surface (0-20 cm) SOC stocks for the current period (2022) comparing random forest, support vector machine and XGBoost machine learning algorithms.

2. Evaluate and compare individual learning algorithms performance in predicting the spatial distribution of SOC stocks.

3. Compare the new predicted SOC map against the existing global SOC maps (Hengl *et al.*, 2017) and the national scale SOC map produced by Venter *et al*. (2021) for South Africa, assessing similarities and differences.

2. Methods

2.1. The Study Area and CSA's Interventions

The Umzimvubu Catchment, located in the Eastern Cape Province of South Africa, encompasses a vast area of approximately 2,874,800 hectares (CSA, 2019; DWS, 2016). This catchment is part of the Maputaland–Pondoland-Albany Hotspot, which is internationally recognised as one of the thirty-six biodiversity hotspots worldwide. The topography of the Umzimvubu Catchment is diverse, including high-altitude grasslands, steep valleys, and a coastal plain. Within the catchment, there are four designated Key Biodiversity Areas (KBAs) that are globally acknowledged for their significant contribution to biodiversity conservation (CSA & UCP, 2022). However, these fragile ecosystems are facing increasing threats from human activities, the spread of non-native species, improper management of rangelands, and associated issues of overgrazing (UCPP, 2015). Building on its expertise in biodiversity and livelihoods, CSA recognises the importance of these ecosystems and has developed a strategy and programme to support the preservation and restoration of a thriving Umzimvubu river system, ensuring the continued provision of vital ecosystem services (ERS - CSA, 2012).

CSA has collaborated with various stakeholders to formulate a strategy for conserving the Umzimvubu catchment (ERS-CSA,2012). This long-term vision enables comprehensive and sustainable conservation efforts in the area, ensuring the responsible management of the catchment for future generations. By incorporating initiatives to restore grasslands, such as controlled grazing, replanting native grass species, and removing invasive plants, it is possible to enhance carbon storage. These interventions not only contribute to the rehabilitation of degraded land but also improve water quality and support wildlife habitats. Through the implementation of projects that focus on soil carbon, CSA can make a significant contribution to climate change mitigation efforts in the Umzimvubu Catchment. Simultaneously, this approach promotes biodiversity conservation, improves soil health, and fosters sustainable livelihoods.

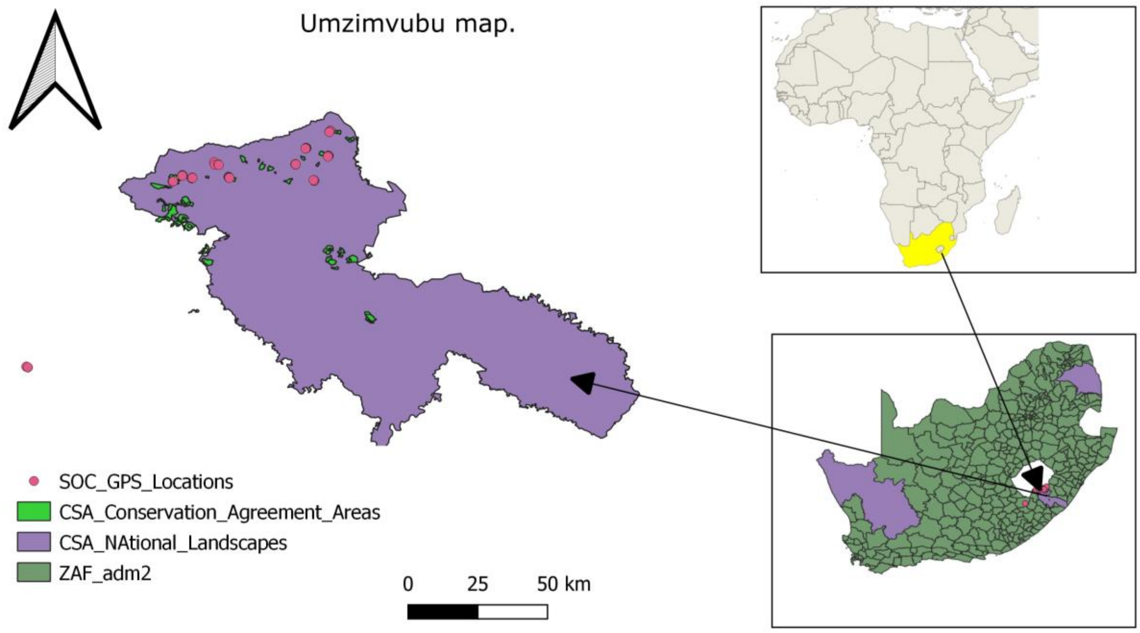

Figure 1. CSA's Intervention areas and soil sample's locations map.

2.2. Field observations of soil properties

A total of eighty-seven soil samples were collected by CSAs field team using an Edelman Auger and small spades at 20 cm depth in December 2022. Purposive sampling was done following a period of about five years controlled grazing interventions implemented through grazing associations in the area. Even though the sampling locations chosen may be representative of the areas where conservation agreements have been implemented and are more efficient in terms of time and resources and could potentially provide insights into the effectiveness of these agreements in improving soil carbon, this can limit the generalizability of the findings to a broader context, for example it may introduce bias where the data may not accurately represent areas without such interventions. The collected soil samples underwent chemical analysis to determine the SOC content using the Walkley-Black method. This method involves oxidising organic matter with a specific amount of potassium dichromate and concentrated sulphuric acid at high temperatures (Gerenfes, Giorgis and Negasa, 2022). The Black Walkley method has advantages such as affordability and minimal equipment requirements, making it accessible for many laboratories (Chatterjee *et al.*, 2009). Its simplicity and widespread availability also contribute to its common use in soil analysis. However, there are limitations to this method, the method relies on calculating organic carbon based on organic matter containing 58% carbon, this ratio may vary depending on the type of soil, vegetation, and other factors (De Vos *et al*., 2007; FAO,2020). Therefore, the method may not be accurate in some cases. Additionally, the method's environmental impact concern is the use of hazardous chemicals and production of heavy metal wastes (Gerenfes, Giorgis, & Negasa, 2022). Despite these limitations, the Black Walkley method remains a popular technique for measuring organic carbon.

Soil organic carbon was calculated using the following equation:

$$\text{SOC stock } (\mathrm{kg/m} \wedge 2) = \mathrm{C(g/kg)} * \text{ bulk density } * (\mathrm{l} - \mathrm{CF}) * \text{ soil depth } (\mathrm{m})$$

Where:

-C refers to carbon concentration, expressed in grams per kilogram $(\mathrm{g/kg})$

-Bulk density refers to the bulk density of the fine soil component, expressed in kilograms per cubic meter $(\mathrm{kg/m} \wedge 3)$

-CF is the volumetric coarse fraction, expressed as a decimal.

-Soil depth refers to the depth of the soil layer being measured, expressed in meters (m).

Other soil properties like silt, sand and clay contents were quantified using the Malvern laser particle size analyser (Miller and Schaetzl, 2012). Bulk density was measured using Eijelkamp rings with a volume of 100 cm3 at a depth of 8-13 cm from field measurements. However, where field measurements were unreliable, SoilGrids data were also used.

To preprocess the data for training the ML models, wrangling was done where the point soil samples recorded in latitude and longitude degrees (D), minutes (M), and seconds (S) were converted to decimal degrees (DD) using the formula below.

$$DD = D + \frac{M}{60} + \frac{S}{3600}$$

The missing longitude and latitude point values were omitted and the data converted to spatial point data frame with a desired coordinate reference system (EPSG:4326) for further spatial analysis. Exploratory data analysis including the density distribution of the soil samples and the relationship between SOC and other variables measured was also conducted.

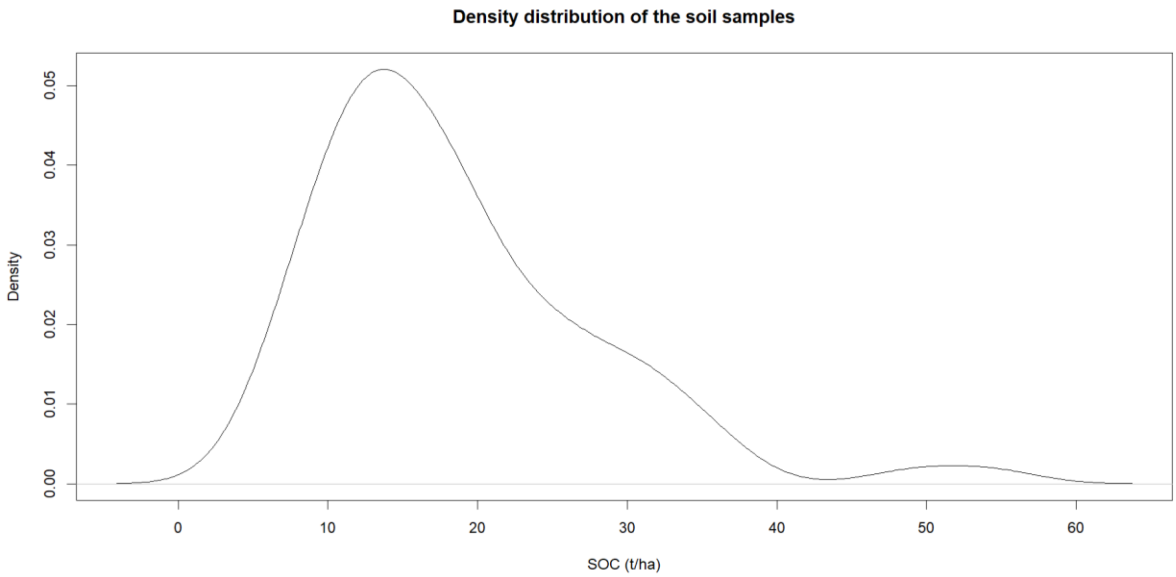

Figure 2.Density distribution of the soil samples.

2.3. Covariates selection and preprocessing

The differences in soil organic carbon stocks between geographical areas can be attributed to various environmental factors. These factors include topography, climate, vegetation, soil characteristics, time, and human activities (McBratney *et al*., 2003; Mishra *et al*., 2021). Understanding the existence of complex and nonlinear associations between soil organic carbon and environmental factors is crucial for accurate predictions and mapping of soil properties. To achieve this, most researchers have employed an approach introduced by McBratney *et al*. (2003), which categorises environmental covariates into several categories: soil (S), climate (C), organisms (O), relief (R), parent material (P), age (A), and other factors (N) (SCORPAN) as predictive factors for soil attributes.

The digital soil mapping approach is used to quantify the relationships between SOC content and environmental covariates. These covariates comprise both quantitative and qualitative variables derived from digital terrain models (DTM), remote sensing indices and thematic maps of landform, land cover and soil types. Additionally, other soil properties, including pH, bulk density, and the distribution of sand, silt and clay contents, have been demonstrated to account for variation in SOC stocks and other dependent soil attributes across landscapes (Zeraatpisheh et al., 2019). By leveraging various machine learning techniques, researchers can predict SOC content based on these soil-forming factors.

Based on literature and as detailed in the review by Xia, McSweeney, and Wander (2022) there is limited correlation between the number of environmental predictors and the performance of SCORPAN models employed to predict SOC, the authors suggest comparing different model performances based on a selected set of covariates and also comparing process based models to SCORPAN based models. Several studies, such as Chinilin and Savin (2023), Kaya *et al*. (2022, 2023) and Khanal *et al*. (2023) have utilized environmental covariates to predict and map SOC, utilising machine learning. These studies highlight the importance of selecting relevant environmental predictors and using appropriate modeling techniques to improve the accuracy of SOC predictions.

To incorporate the environmental factors utilized in SCORPAN framework, terrain data such as slope, aspect, elevation, and topographical wetness index (TWI) were generated from the digital terrain model (DTM) using the "elevatr" package in R (Hollister et al., 2023). This data, acquired at a resolution of ten meters, was then stored for further preprocessing and modeling. The terrain predictors provide information about the topography and hydrological characteristics of the soil, which are known to influence soil formation processes.

The climatic variables used in the study were obtained from the WorldClim data version 2., which represents the long-term averages for the period 1970-2000 (Fick and Hijmans, 2017). Previous research has demonstrated that climatic variables, such as precipitation and temperature, are significant predictors of soil organic carbon variability (Flathers and Gessler, 2018). These variables impact the decomposition of organic matter, nutrient availability, and microbial activity in the soil, ultimately affecting soil carbon stocks (Khanal et al., 2023). Nineteen bioclimatic indicators were utilized to represent biologically meaningful factors relevant to soil organic carbon dynamics. For instance, the "BIO16" indicator represents the Precipitation of Wettest Quarter, accounting for seasonal variability in SOC. Some of these climatic variables, like the maximum temperature of the warmest month ("BIO5") and the precipitation of the wettest month ("BIO13"), exhibited strong correlations with SOC, as shown in the variable importance chart in Figure 6.

Apart from climatic variables, soil parameters also play a crucial role in understanding SOC. The ISRIC data hub provides soil parameter estimates, terrain, and unit composition data for Southern Africa at a scale of 1:2M, ranging from (0 – 1) m depth (Batjes, 2004). This dataset includes spatial and soil attribute data for eight Southern African countries. The "SOTERSAF" data were utilized to obtain soil parameter estimates for the study area, which were then Preprocessed to select only the relevant variables, such as bulk density (BULK), total nitrogen (TOTN), coarse fragments (SDTO), and soil pH (PHAQ), among others (Batjes, 2004). These properties significantly influence the storage and cycling of carbon in soils. For example, coarse fragments are important for understanding soil structure and porosity, which impact the storage and cycling of carbon in soils, while total nitrogen affects the decomposition of organic matter (Batjes *et al*., 2007). The selected variables were extracted for model training.

To capture the relationship between vegetation and SOC, MODIS vegetation index data, specifically the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), were downloaded for the year 2022 at a resolution of 250m using the "MODIStsp" package in R (Busetto and Ranghetti, 2016). Since SOC is influenced by the input of organic matter from vegetation, incorporating NDVI and EVI as covariates helps to capture this relationship. The sixteen-day data were then used to calculate the annual mean values for these indices. All the raster data were subsequently reprojected to the WGS84 geographic coordinate system (CRS) as necessary, ensuring that they have a similar spatial extent, making them ready for model fitting.

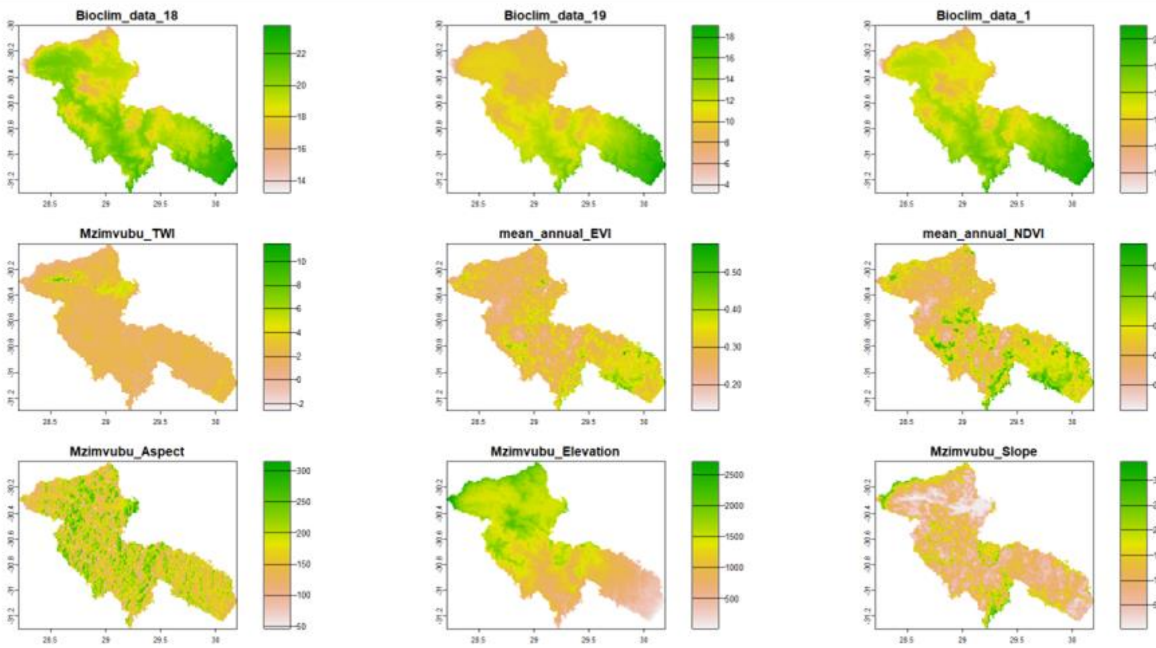

Figure 3. Sample Covariates prepared for modelling.

2.4 Machine learning framework

The methodology employed in this study utilized spatial data handling tools along with machine learning algorithms in the R programming language. The initial step involved extracting the raster values of Preprocessed covariates at specific point locations of the soil samples. Additionally, soil classes and legacy parameters were also extracted at these points. These extracted values were then combined to create a comprehensive dataset containing both soil properties and environmental covariates as predictor variables.

Subsequently, a spatial regression task was established using the "mlr3 package" in R (Lang *et al*., 2019). The predictor variables consisted of the covariates, while the target variable was the amount of soil organic carbon measured in tons per hectare. To assess the performance of the machine learning algorithms, a 3-fold cross-validation approach was implemented. Specifically, three algorithms, random Forest, XGBoost, and support vector machines, were trained and compared based on their performance on the regression task based on performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared. The evaluated metric results are as shown in Figure 4 below.

```
> print(bmr_aggr)
   nr      task_id   learner_id resampling_id iters regr.mse regr.mae  regr.rsq
1:  1 spatial_task  regr.ranger            cv     3 52.24913 5.375829 0.2684298
2:  2 spatial_task regr.xgboost            cv     3 59.00837 5.625589 0.1694110
3:  3 spatial_task     regr.svm            cv     3 60.49626 5.340814 0.1488186
```

Figure 4. benchmark results of the three learners based on 3-fold cross validation.

Given the small sample size and the poor performance of the models on cross validation, alternative methods to train the model on the 80%-20% train-validation split ratio to check the difference in performance was also implemented. The models were trained on 80% of the total data and predictions were made on the remaining 20% and their performance metrics MSE, MAE, and R-squared were computed. Two of the models: random forest and XGBoost showed a significant improvement on their performance compared to their performance on 3-fold cross validation above. The performance metrics of the three models on (80 - 20) train and test set are as shown in table 1 below.

Table 1. learners' performance metrics based on 80%-20% train and test split ratio.

LEARNER	MSE	MAE	R SQUARED
random forest	25.2399	3.831604	0.5562988
XGBoost	28.96004	4.211441	0.4909011
SVM	42.09079	4.889049	0.2600709

The random forest continually showed better performance on both 3-fold cross-validation and on the (80- 20) train and test set. Based on the model performance it was chosen for further predictive mapping. A set of new random locations within the study area were generated, for these new locations, covariate values were extracted and predictions of soil organic carbon were then made using the trained random forest model. Subsequently, variable importance was extracted from the trained random forest model and a bar plot visualizing the importance of each variable was generated (refer to Figure 6). The variable importance chart helps identify which among the variables used are most influential in predicting soil organic carbon. Finally, a grid based Inverse Distance Weighting (IDW) interpolation was used to predict soil carbon for a continuous region based on point values, this method was chosen based on its performance in other similar studies like the study conducted by (Van Huynh *et al.*, 2022). The data was then masked to the study area boundaries.

3. Results

The predicted soil organic carbon map is shown in Figure 5 below.

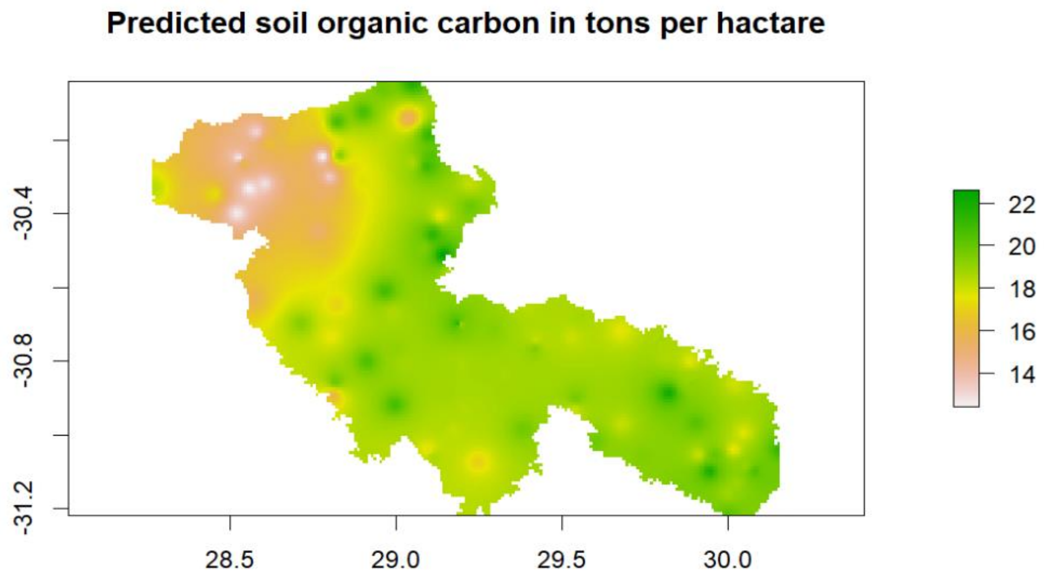

Figure 5. Predicted soil organic carbon map.

The covariates importance chart is as shown in Figure 6 below. This shows the relative importance of the environmental covariates used in estimating soil organic carbon by the random forest model. Mean annual EVI is shown to be the most significant variable that impact SOC followed by precipitation of the wettest month ("BIO13").

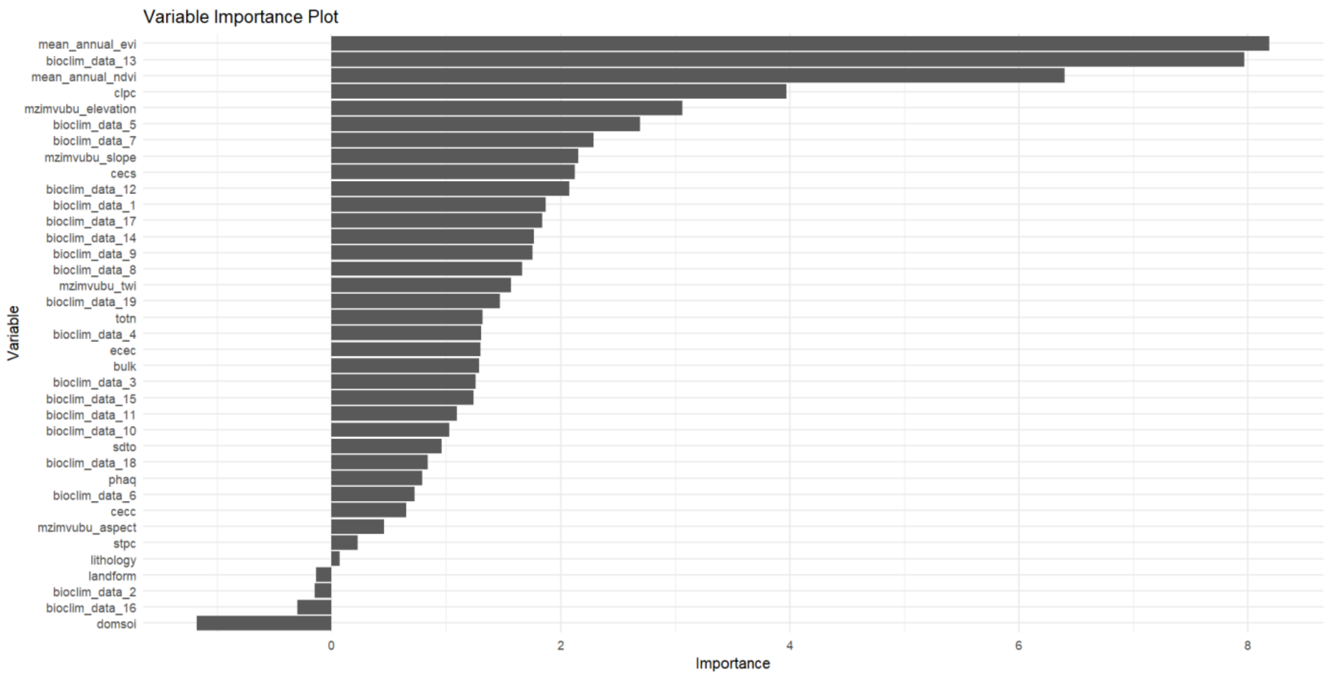

Figure 6. Variable importance chart.

The subset map of estimates from (Venter *et al.,* 2021) and ISRIC global data (Hengl *et al.*, 2017) are shown in Figure 7 and Figure 8 respectively.

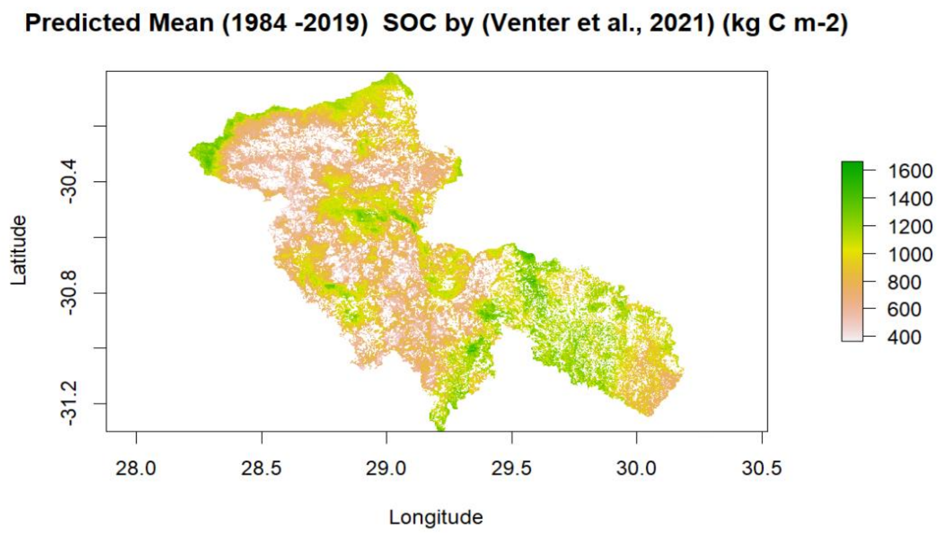

Figure 7. Subset of Predicted Mean (1984 -2019) SOC by (Venter et al., 2021).

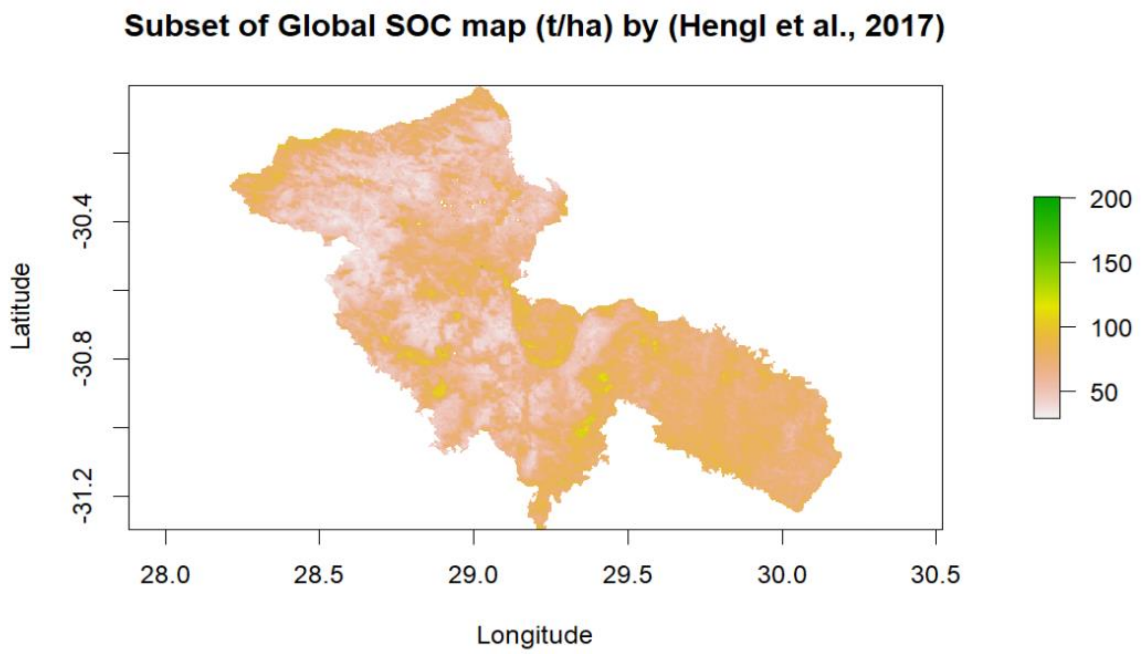

Figure 8. Subset of Global SOC map (t/ha) by (Hengl et al., 2017).

4. Discussion

Utilizing point soil samples and environmental covariates, a map depicting the predicted SOC was generated (refer to Figure 4). It should be noted that the sampling strategy employed in this study may limit the generalizability of the results within the study area, however, it does provide insights into the distribution of SOC. Specifically, the northeastern part of the study area demonstrates consistently low levels of SOC. This observation aligns with the prediction maps created by Venter et al. (2021) and Hengl et al. (2017), despite the use of different covariates in their respective predictions. Therefore, targeted interventions should be considered in this region to enhance carbon sequestration rates.

The performance of the three models: random forest, SVM and XGBoost explored in this research was evaluated using both 3-fold cross validation and 80% - 20% train-test split ratio. Performance metrics: R-squared, mean squared error (MSE), and mean absolute error (MAE) were calculated and the results are as shown in Figure 4 and Table 1 respectively. The random forest model exhibited lower MSE and MAE values, indicating a better predictive applicability compared to the two other models. Additionally, it displayed a higher R-squared value, suggesting a better fit to the data in terms of explaining the variability in SOC. The evaluation metrics shows that repeated resampling of the data using cross- validation helps in obtaining an average estimate of the model's performance, which is more reliable estimate of the model's performance compared to evaluating it on a single validation set to avoid under or overfitting. These results are however not generalisable based on the small sample size and the potential bias presented by the sampling strategy. Overall, the objective of this study to explore methodologies that have been shown to enhance the measurement of soil organic carbon (SOC) using a case study of the Umzimvubu catchment was achieved. The limitations of this study including but not limited to a small sample size can be improved in further studies by Incorporating data from more locations with or without interventions, this will provide more insights in the soil organic carbon dynamics in the mzimvubu catchment and support the accurate quantification of soil organic carbon both as an internal monitoring framework for CSA and other stakeholders or in compliance with requirements of carbon markets.

5. Conclusion

The accuracy of current soil maps is often uncertain, particularly in countries with limited soil surveys (Buenemann *et al*., 2023). Consequently, relying on these existing maps for decision-making and policy making is highly unreliable. Recent advancements in climate finance and research programs are providing support for carbon sequestration in rangelands (Elavarthi, 2014). To encourage and reward well-managed rangelands for their carbon sequestration efforts through climate finance opportunities, it is crucial to improve the estimation of soil properties at local levels and enhance the accuracy of soil map predictions. This can be accomplished through collaborative research and policy efforts involving various stakeholders to develop effective strategies for quantifying carbon sequestration and establishing carbon markets in rangeland contexts. In the rangeland areas of South Africa's Eastern Cape, soil plays a significant role as a potential carbon sink. However, accurately measuring and quantifying the amount of carbon sequestered in rangelands is a complex undertaking that requires robust monitoring and measurement techniques which can be costly. It is essential to collaborate with stakeholders who are implementing carbon sequestration efforts, such as CSA, to explore the viability of carbon sequestration techniques in offsetting the effects of climate change at the local level. This collaboration enhances the value of research by seamlessly bridging the gap between academia and real-world practical application of theories and methods which can be challenging. There is a need to increase the number of soil observations in the study area in order to gain a more to a deeper understanding of the primary factors that govern soil organic carbon concentration at the local scale in areas with or without intervention within the Umzimvubu catchment. Moreover, it is crucial to evaluate the efficacy of management strategies and consider the unique advantages and disadvantages of each site. This will ensure the swift adoption and rollout of effective soil carbon sequestration practices in areas with a high potential for carbon storage.

Data and code availability

The bioclimatic variables used were generated from http://worldclim.org/version2. The soil legacy data are available at https://data.isric.org/. The other variables were dynamically generated in R. The code used for data analysis in this study can be accessed on https://github.com/TESS-Laboratory/Kirinyet-development

Appendix

The 19 bioclimatic indicators from WorldClim (Fick and Hijmans, 2017) used are coded as follows:

BIO1 = Annual Mean Temperature

BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))

BIO3 = Isothermality (BIO2/BIO7) ($\times$100)

BIO4 = Temperature Seasonality (standard deviation $\times$100)

BIO5 = Max Temperature of Warmest Month

BIO6 = Min Temperature of Coldest Month

BIO7 = Temperature Annual Range (BIO5-BIO6)

BIO8 = Mean Temperature of Wettest Quarter

BIO9 = Mean Temperature of Driest Quarter

BIO10 = Mean Temperature of Warmest Quarter

BIO11 = Mean Temperature of Coldest Quarter

BIO12 = Annual Precipitation

BIO13 = Precipitation of Wettest Month

BIO14 = Precipitation of Driest Month

BIO15 = Precipitation Seasonality (Coefficient of Variation)

BIO16 = Precipitation of Wettest Quarter

BIO17 = Precipitation of Driest Quarter

BIO18 = Precipitation of Warmest Quarter

BIO19 = Precipitation of Coldest Quarter

The selected soil properties from legacy soil maps from ISRIC (Batjes, 2004) are coded as follows:

PHAQ - pH measured in water

TOTC - Organic carbon content (g kg-1)

TOTN - Total nitrogen (g kg-1)

STPC - Silt (mass %)

CLPC - Clay (mass %)

BULK - Bulk density (kg dm-3)

CECS - Cation exchange capacity (cmolc kg-1) of fine earth fraction

ECEC - Effective CEC (cmolc kg-1)

SDTO - Sand (mass %)

DOMSOI - dominant soils

References:

Abbas, S. *et al*. (2020) 'Approaches of Satellite Remote Sensing for the Assessment of Above-Ground Biomass across Tropical Forests: Pan-tropical to National Scales', *Remote Sensing*, 12(20), p. 3351.Available at: https://doi.org/10.3390/rs12203351.

Abdalla, M. *et al*. (2022) 'Evaluation of the DNDC Model to Estimate Soil Parameters, Crop Yield and Nitrous Oxide Emissions for Alternative Long-Term Multi-Cropping Systems in the North China Plain', *Agronomy*, 12(1), p. 109. Available at: https://doi.org/10.3390/agronomy12010109.

Afzali, S.F. *et al*. (2019) 'Using RothC Model to Simulate Soil Organic Carbon Stocks under Different Climate Change Scenarios for the Rangelands of the Arid Regions of Southern Iran', *Water*, 11(10), p.2107. Available at: https://doi.org/10.3390/w11102107.

Ahmad, N. *et al*. (2023) 'Comparative Analysis of Remote Sensing and Geo-Statistical Techniques to Quantify Forest Biomass', *Forests*, 14(2), p. 379. Available at: https://doi.org/10.3390/f14020379.

Amelung, W. *et al*. (2020) 'Towards a global-scale soil climate mitigation strategy', *Nature Communications*, 11(1), p. 5427. Available at: https://doi.org/10.1038/s41467-020-18887-7.

Balehegn, M. *et al*. (2021) 'Livestock sustainability research in Africa with a focus on the environment', *Animal Frontiers*, 11(4), pp. 47–56. Available at: https://doi.org/10.1093/af/vfab034.x.

Batjes NH. (2004) SOTER-based soil parameter estimates for Southern Africa. Wageningen, The Netherlands: ISRIC—World Soil Information.

Batjes, N.H. *et al*. (2007) 'Preparation of consistent soil data sets for modelling purposes: Secondary SOTER data for four case study areas', *Agriculture, Ecosystems & Environment*, 122(1), pp. 26–34. Available at: https://doi.org/10.1016/j.agee.2007.01.005.

Buenemann, M. *et al*. (2023) 'Errors in soil maps: The need for better on-site estimates and soil map predictions,' *PLOS ONE*. Edited by S.P. Aldrich, 18(1), p. e0270176. Available at: https://doi.org/10.1371/journal.pone.0270176.

Byrne, K. and Kiely, G., (2012). Evaluation of models (PaSim, RothC, CENTURY and DNDC) for simulation of grassland carbon cycling at plot, field and regional scale.

Chatterjee, A. *et al*. (2009) 'Evaluation of Different Soil Carbon Determination Methods', *Critical Reviews in Plant Sciences*, 28(3), pp. 164–178. Available at: https://doi.org/10.1080/07352680902776556.

Chinilin, A. and Savin, I. Yu. (2023) 'Combining machine learning and environmental covariates for mapping of organic carbon in soils of Russia,' *The Egyptian Journal of Remote Sensing and Space Science,* 26(3), pp. 666–675. Available at: https://doi.org/10.1016/j.ejrs.2023.07.007.

Coleman, K. and Jenkinson, D.S. (1996) 'RothC-26.3 - A Model for the turnover of carbon in soil', in D.S. Powlson, P. Smith, and J.U. Smith (eds) Evaluation of Soil Organic Matter Models. Berlin, Heidelberg: *Springer Berlin Heidelberg*, pp. 237–246. Available at: https://doi.org/10.1007/978-3-642-61094-3_17.

Cong, R. *et al*. (2014) 'Evaluation of the CENTURY Model Using Long-Term Fertilization Trials under Corn-Wheat Cropping Systems in the Typical Croplands of China', *PLOS ONE*. Edited by J. Vera, 9(4), p.e95142. Available at: https://doi.org/10.1371/journal.pone.0095142.

Conservation South Africa. (2019) *Conservation South Africa Annual Report 2018-2019*. Available at: https://www.conservation.org/docs/default-source/south-africa-documents/csa-report-2018-2019.pdf . [Accessed 9 August 2023].

Conservation South Africa, IIED (2021). Policy brief: *The importance of ecosystem-based adaptation and mitigation actions in the Nationally Determined Contributions*. Conservation South Africa, Cape Town Available at: https://www.iied.org/20641g. [Accessed 9 August 2023].

Conservation South Africa. (2023). *Conservation South Africa Impact Report 2020-2022*. Available at: https://www.conservation.org/docs/default-source/south-africa-documents/csa-impact-report-11-04-23_hi-res.pdf?sfvrsn=5084245d_10. [Accessed 8 August 2023].

CSA & UCP (2022). *UMZIMVUBU RIVER CATCHMENT, SOUTH AFRICA*. Available at: https://www.conservation.org/docs/default-source/south-africa-documents/umzimvubufhireport.pdf?Status=Master&sfvrsn=266a351c_1 [Accessed 9 August 2023].

Conservation South Africa. (2022). *Kruger to Canyons Carbon Project* [Online video]. Retrieved from https://youtube.com/watch?v=xf00CgVl46U. [Accessed 9 August 2023].

Dannenberg, M.P. *et al*. (2023) 'Upscaling dryland carbon and water fluxes with artificial neural networks of optical, thermal, and microwave satellite remote sensing', *Biogeosciences*, 20(2), pp. 383–404. Available at: https://doi.org/10.5194/bg-20-383-2023.

Deng, L. *et al*. (2016) 'Global patterns of the effects of land-use changes on soil carbon stocks', *Global Ecology and Conservation*, 5, pp. 127–138. Available at: https://doi.org/10.1016/j.gecco.2015.12.004.

Department of Water and Sanitation (DWS), South Africa. 2016. Determination of Water Resource Classes and Resource Quality Objectives for the Water Resources in the Mzimvubu Catchment: Inception Report. Prepared by Scherman Colloty & Associates cc. Report no. WE/WMA7/00/CON/CLA/0116.

De Vos, B. et al. (2007) 'Walkley? Black analysis of forest soil organic carbon: recovery, limitations and uncertainty', *Soil Use and Management*, 23(3), pp. 221–229. Available at: https://doi.org/10.1111/j.1475-2743.2007.00084.x.

Elavarthi S, M.C. (2014) 'Rangelands as Carbon Sinks to Mitigate Climate Change: A Review', Journal of Earth Science & Climatic Change, 05(08). Available at: https://doi.org/10.4172/2157-7617.1000221.

Emadi, M. *et al*. (2020) 'Predicting and Mapping of Soil Organic Carbon Using Machine Learning Algorithms in Northern Iran', *Remote Sensing*, 12(14), p. 2234. Available at: https://doi.org/10.3390/rs12142234.

ERS - CSA (2011). *UMZIMVUBU CATCHMENT OVERVIEW*. Available at: https://umzimvubu.files.wordpress.com/2014/09/umzimvubu-summary-report-dec-2011.pdf. [Accessed 7 August 2023].

ERS -CSA (2012) *Umzimvubu 5-Year Strategy Phase 1*. Available at: https://umzimvubu.files.wordpress.com/2014/09/vubu-5-year-strategy-phase-1-jan-2012.pdf. [Accessed 7 August 2023].

Fathizad, H. et al. (2022) 'Spatiotemporal Assessment of Soil Organic Carbon Change Using Machine-Learning in Arid Regions,' *Agronomy*, 12(3), p. 628. Available at: https://doi.org/10.3390/agronomy12030628.

Flathers, E. and Gessler, P.E. (2018) 'Building an Open Science Framework to Model Soil Organic Carbon,' Journal of Environmental Quality, 47(4), pp. 726–734. Available at: https://doi.org/10.2134/jeq2017.08.0318.

Food and Agriculture Organization of the United Nations. (2020). Standard operating procedure for soil organic carbon. Walkley-Black method: titration and colorimetric method. Retrieved from https://www.fao.org/publications/card/en/c/CA7471EN/.

Fick, S.E. and R.J. Hijmans, (2017). WorldClim 2: new 1 km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 (12): 4302-4315. [Rainfall, Temperature] Available at: http://worldclim.org/version2.[Accessed 1 June 2023]

Flynn, T. (2019) 'Digital soil mapping techniques across multiple landscape scales in South Africa'.

Le Noë, J. *et al*. (2023) 'Soil organic carbon models need independent time-series validation for reliable prediction', *Communications Earth & Environment*, 4(1), p. 158. Available at: https://doi.org/10.1038/s43247-023-00830-5.

Grace, P. and 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), pp.1410–1424. Available at: https://doi.org/10.1002/saj2.20273.

Georgiou, K. *et al*. (2022) 'Global stocks and capacity of mineral-associated soil organic carbon', *Nature Communications*, 13(1), p. 3797. Available at: https://doi.org/10.1038/s41467-022-31540-9.

Gerenfes, D., Giorgis, A.G. and Negasa, G. (2022) 'Comparison of organic matter determination methods in soil by loss on ignition and potassium dichromate method', *International Journal of Horticulture and Food Science*, 4(1), pp. 49–53. Available at: https://doi.org/10.33545/26631067.2022.v4.i1a.85.

Gilhespy, S.L. *et al*. (2014) 'First 20 years of DNDC (DeNitrification DeComposition): Model evolution', *Ecological Modelling*, 292, pp. 51–62. Available at: https://doi.org/10.1016/j.ecolmodel.2014.09.004.

Giltrap, D.L., Li, C. and Saggar, S. (2010) 'DNDC: A process-based model of greenhouse gas fluxes from agricultural soils', *Agriculture, Ecosystems & Environment*, 136(3–4), pp. 292–300. Available at: https://doi.org/10.1016/j.agee.2009.06.014.

Gosnell, H., Charnley, S. and Stanley, P. (2020) 'Climate change mitigation as a co-benefit of regenerative ranching: insights from Australia and the United States', *Interface Focus*, 10(5), p. 20200027. Available at: https://doi.org/10.1098/rsfs.2020.0027.

Graux, A.-I. *et al*. (2011) 'Development of the Pasture Simulation Model for assessing livestock production under climate change', *Agriculture, Ecosystems & Environment*, 144(1), pp. 69–91. Available at: https://doi.org/10.1016/j.agee.2011.07.001.

Griscom, B.W. *et al*. (2017) 'Natural climate solutions', *Proceedings of the National Academy of Sciences*, 114(44), pp. 11645–11650. Available at: https://doi.org/10.1073/pnas.1710465114.

Grunwald, S. (2022) 'Artificial intelligence and soil carbon modeling demystified: power, potentials, and perils', *Carbon Footprints*, 1(1), p. 6. Available at: https://doi.org/10.20517/cf.2022.03.

Gusha, B., Palmer, A.R. and Zondani, T.C. (2023) 'Assessing Livestock Grazing Distribution in

Communal Rangelands of the Eastern Cape, South Africa: Towards Monitoring Livestock Movements in Rangelands,' *Land*, 12(4), p. 760. Available at: https://doi.org/10.3390/land12040760.

Havens, Drew, "Agricultural Carbon Markets: How Might They Work?" (2021). *Presentations, Working Papers, and Gray Literature: Agricultural Economics*. 51. https://digitalcommons.unl.edu/ageconworkpap/51.

Hawkins, H.J. *et al*. (2021) 'Modelling Grazing and Burning in Communal Rangelands to Help Understand Trade-offs between Production, Carbon, and Water'.

Hollister J, Shah T, Robitaille A, Beck M, Johnson M (2023). elevatr: Access Elevation Data from Various APIs. doi:10.5281/zenodo.5809645, R package version 0.4.5, URL: https://github.com/jhollist/elevatr/.

Huang, H. *et al*. (2022) 'A review on digital mapping of soil carbon in cropland: progress, challenge, and prospect', *Environmental Research Letters*, 17(12), p. 123004. Available at: https://doi.org/10.1088/1748-9326/aca41e.

Jebari, A. *et al*. (2021) 'Estimating soil organic carbon changes in managed temperate moist grasslands with RothC', *PLOS ONE*. Edited by R. Paradelo Núñez, 16(8), p. e0256219. Available at: https://doi.org/10.1371/journal.pone.0256219.

Jenkinson, D.S. and Coleman, K. (2008) 'The turnover of organic carbon in subsoils. Part 2. Modelling carbon turnover,' *European Journal of Soil Science*, 59(2), pp. 400–413. Available at: https://doi.org/10.1111/j.1365-2389.2008.01026.x.

Karra, Kontgis, *et al*. "Global land use/land cover with Sentinel-2 and deep learning." *IGARSS 2021-2021 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2021.

Kaya, F. *et al*. (2022) 'Assessing Machine Learning-Based Prediction under Different Agricultural Practices for Digital Mapping of Soil Organic Carbon and Available Phosphorus,' *Agriculture*, 12(7), p. 1062. Available at: https://doi.org/10.3390/agriculture12071062.

Kaya, F. et al. (2022) 'Assessing Machine Learning-Based Prediction under Different Agricultural Practices for Digital Mapping of Soil Organic Carbon and Available Phosphorus,' *Agriculture*, 12(7), p. 1062. Available at: https://doi.org/10.3390/agriculture12071062.

Kaya, F. et al. (2023) 'Combining Digital Covariates and Machine Learning Models to Predict the Spatial Variation of Soil Cation Exchange Capacity', *Land*, 12(4), p. 819. Available at: https://doi.org/10.3390/land12040819.

Keesstra, S, Franke, A, Wösten, H & Mashingaidze, N (2020). *Potential role of Conservation Agriculture in South Africa for carbon sequestration for climate mitigation: A provisional research agenda*. Report / Wageningen Environmental Research, no. 3024, Wageningen Environmental Research, Wageningen. Available at: https://doi.org/10.18174/526557.

Kelly, R.H., Parton, W.J., Crocker, G.J., Graced, P.R., Klir, J., Körschens, M., Poulton, P.R. and Richter, D.D., 1997. Simulating trends in soil organic carbon in long-term experiments using the century model. *Geoderma*, 81(1-2), pp.75-90.

Khanal, S. et al. (2023) 'Mapping soil organic carbon stocks in Nepal's forests', *Scientific Reports*, 13(1), p. 8090. Available at: https://doi.org/10.1038/s41598-023-34247-z.

Khangura, R. *et al*. (2023) 'Regenerative Agriculture—A Literature Review on the Practices and Mechanisms Used to Improve Soil Health', *Sustainability*, 15(3), p. 2338. Available at: https://doi.org/10.3390/su15032338.

Kim, M.-J. *et al*. (2022) 'Development of a Soil Organic Matter Content Prediction Model Based on Supervised Learning Using Vis-NIR/SWIR Spectroscopy', *Sensors*, 22(14), p. 5129. Available at: https://doi.org/10.3390/s22145129.

L. Busetto, L. Ranghetti (2016) MODIStsp: An R package for automatic preprocessing of MODIS Land Products time series, Computers & Geosciences, Volume 97, Pages 40-48, ISSN 0098-3004, https://doi.org/10.1016/j.cageo.2016.08.020, URL: https://github.com/ropensci/MODIStsp.

Lal, R. *et al*. (2021) 'The role of soil in regulation of climate', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1834), p. 20210084. Available at: https://doi.org/10.1098/rstb.2021.0084.

Lal, R. (2010) 'Managing Soils and Ecosystems for Mitigating Anthropogenic Carbon Emissions and Advancing Global Food Security', *BioScience*, 60(9), pp. 708–721. Available at: https://doi.org/10.1525/bio.2010.60.9.8.

Lal, R. (2004) 'Soil Carbon Sequestration Impacts on Global Climate Change and Food Security', Science, 304(5677), pp. 1623–1627. Available at: https://doi.org/10.1126/science.1097396.

Lamichhane, S., Kumar, L. and Wilson, B. (2019) 'Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review', *Geoderma*, 352, pp. 395–413. Available at: https://doi.org/10.1016/j.geoderma.2019.05.031.

Lanceman, D. *et al*. (2022) 'Blue carbon ecosystem monitoring using remote sensing reveals wetland restoration pathways', *Frontiers in Environmental Science*, 10, p. 924221. Available at: https://doi.org/10.3389/fenvs.2022.924221.

Lang M, Binder M, Richter J, Schratz P, Pfisterer F, Coors S, Au Q, Casalicchio G, Kotthoff L, Bischl B (2019). "mlr3: A modern object-oriented machine learning framework in R." *Journal of Open Source Software*. doi:10.21105/joss.01903, https://joss.theoj.org/papers/10.21105/joss.01903.

Lunn-Rockliffe, S., Davies, M.I., Willman, A., Moore, H.L., McGlade, J.M. and Bent, D. (2020). Farmer Led Regenerative Agriculture for Africa. London, Institute for Global Prosperity.

Macharia, J.M., Ngetich, F.K. and Shisanya, C.A. (2021) 'Parameterization, calibration and validation of the DNDC model for carbon dioxide, nitrous oxide and maize crop performance estimation in East Africa', *Heliyon*, 7(5), p. e06977. Available at: https://doi.org/10.1016/j.heliyon.2021.e06977.

Mahmoudzadeh, H. et al. (2020) 'Spatial prediction of soil organic carbon using machine learning techniques in western Iran,' *Geoderma Regional*, 21, p. e00260. Available at: https://doi.org/10.1016/j.geodrs.2020.e00260.

Martin, M.P. *et al*. (2021) 'Feasibility of the 4 per 1000 aspirational target for soil carbon: A case study for France', *Global Change Biology*, 27(11), pp. 2458–2477. Available at: https://doi.org/10.1111/gcb.15547.

Matinfar, H.R. *et al*. (2021) 'Evaluation and Prediction of Topsoil organic carbon using Machine learning and hybrid models at a Field-scale', *CATENA*, 202, p. 105258. Available at: https://doi.org/10.1016/j.catena.2021.105258.

McBratney, A.B., Mendonça Santos, M.L. and Minasny, B. (2003) 'On digital soil mapping', *Geoderma*, 117(1–2), pp. 3–52. Available at: https://doi.org/10.1016/S0016-7061(03)00223-4.

McClean, G.J. *et al*. (2015) 'An empirical model approach for assessing soil organic carbon stock changes following biomass crop establishment in Britain', *Biomass and Bioenergy*, 83, pp. 141–151. Available at: https://doi.org/10.1016/j.biombioe.2015.09.005.

Meier, M. *et al*. (2018) 'Digital Soil Mapping Using Machine Learning Algorithms in a Tropical Mountainous Area,' *Revista Brasileira de Ciência do Solo*, 42(0). Available at: https://doi.org/10.1590/18069657rbcs20170421.

Mgalula, M.E. *et al*. (2021) 'Greenhouse gas emissions and carbon sink potential in Eastern Africa rangeland ecosystems: A review', *Pastoralism*, 11(1), p. 19. Available at: https://doi.org/10.1186/s13570-021-00201-9.

Miller, B.A. and Schaetzl, R.J. (2012) 'Precision of Soil Particle Size Analysis using Laser Diffractometry,' *Soil Science Society of America Journal*, 76(5), pp. 1719–1727. Available at: https://doi.org/10.2136/sssaj2011.0303.

Mishra, U. *et al*. (2020) 'Ensemble Machine Learning Approach Improves Predicted Spatial Variation of Surface Soil Organic Carbon Stocks in Data-Limited Northern Circumpolar Region', *Frontiers in Big Data*, 3, p. 528441. Available at: https://doi.org/10.3389/fdata.2020.528441.

Mondal, A. *et al*. (2017) 'Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data', *The Egyptian Journal of Remote Sensing and Space Science*, 20(1), pp. 61–70. Available at: https://doi.org/10.1016/j.ejrs.2016.06.004.

Mousavi, S.R. et al. (2022) 'Three-dimensional mapping of soil organic carbon using soil and environmental covariates in an arid and semi-arid region of Iran,' *Measurement*, 201, p. 111706. Available at: https://doi.org/10.1016/j.measurement.2022.111706.

Mudau, H.S. *et al*. (2022) 'Veld restoration strategies in South African semi-arid rangelands. Are there any successes? —A review,' *Frontiers in Environmental Science*, 10, p. 960345. Available at: https://doi.org/10.3389/fenvs.2022.960345.

Nemoto, R. (no date) 'Soil organic carbon (SOC) now and in the future. Effect of soil characteristics and agricultural management on SOC and model initialisation methods using recent SOC data.'

Nyawira, S.S. *et al*. (2021) 'Enhancing Soil Carbon in East Africa: The biophysical evidence, socioeconomic incentives, and policy implications. Available at: https://doi.org/10.13140/RG.2.2.12204.80009.

Odebiri, O. *et al*. (2023) 'Mapping soil organic carbon distribution across South Africa's major biomes using remote sensing-topo-climatic covariates and Concrete Autoencoder-Deep neural networks', *Science of The Total Environment*, 865, p. 161150. Available at: https://doi.org/10.1016/j.scitotenv.2022.161150.

Offiong, R. and Iwara, A. (2012) 'Quantifying the Stock of Soil Organic Carbon using Multiple Regression Model in a Fallow Vegetation, Southern Nigeria,' *Ethiopian Journal of Environmental Studies and Management*, 5(2), pp. 166–172. Available at: https://doi.org/10.4314/ejesm.v5i2.7.

Ogungbuyi, M.G. *et al*. (2023) *Quantifying Grassland Biomass and Regenerative Grazing Using Satellite Remote Sensing and Machine Learning*. preprint. Environmental and Earth Sciences. Available at: https://doi.org/10.20944/preprints202304.1131.v1.

Olson, K.R., 2010. Impacts of tillage, slope, and erosion on soil organic carbon retention. *Soil science*, *175*(11), pp.562-567.

Padbhushan, R. *et al*. (2022) ‘Impact of Land-Use Changes on Soil Properties and Carbon Pools in India: A Meta-analysis’, *Frontiers in Environmental Science*, 9, p. 794866. Available at: https://doi.org/10.3389/fenvs.2021.794866.

Paterson, G. *et al*. (2015) ‘Spatial soil information in South Africa: Situational analysis, limitations and challenges’, *South African Journal of Science*, 111(5/6), pp. 1–7. Available at: https://doi.org/10.17159/sajs.2015/20140178.

Paul, C. *et al*. (2023) ‘Carbon farming: Are soil carbon certificates a suitable tool for climate change mitigation?’, *Journal of Environmental Management*, 330, p. 117142. Available at: https://doi.org/10.1016/j.jenvman.2022.117142.

Payen, F.T. *et al*. (2021) ‘Predicting the abatement rates of soil organic carbon sequestration management in Western European vineyards using random forest regression’, *Cleaner Environmental Systems*, 2, p.100024. Available at: https://doi.org/10.1016/j.cesys.2021.100024.

Paustian, K. *et al*. (2019) ‘Quantifying carbon for agricultural soil management: from the current status toward a global soil information system’, *Carbon Management*, 10(6), pp. 567–587. Available at: https://doi.org/10.1080/17583004.2019.1633231.

Poeplau, C. and Dechow, R. (2023) ‘The legacy of one hundred years of climate change for organic carbon stocks in global agricultural topsoils’, *Scientific Reports*, 13(1), p. 7483. Available at: https://doi.org/10.1038/s41598-023-34753-0.

Pouladi, N. et al. (2023) ‘Digital mapping of soil organic carbon using remote sensing data: A systematic review’*, CATENA*, 232, p. 107409. Available at: https://doi.org/10.1016/j.catena.2023.107409.

Pulina, A. *et al.* (2018) ‘Modelling pasture production and soil temperature, water and carbon fluxes in Mediterranean grassland systems with the Pasture Simulation model’, *Grass and Forage Science*, 73(2), pp. 272–283. Available at: https://doi.org/10.1111/gfs.12310.

Rie Nemoto. Soil organic carbon (SOC) now and in the future. Effect of soil characteristics and agricultural management on SOC and model initialisation methods using recent SOC data. Agricultural sciences. Université Blaise Pascal - Clermont-Ferrand II, 2013. English. NNT: 2013CLF22430. tel-00973853.

Ritchie, M.E. (2020) ‘Grazing Management, Forage Production and Soil Carbon Dynamics’, *Resources*,9(4), p. 49. Available at: https://doi.org/10.3390/resources9040049.

Rolnick, D. *et al*. (2023) ‘Tackling Climate Change with Machine Learning’, *ACM Computing Surveys*,55(2), pp. 1–96. Available at: https://doi.org/10.1145/3485128.

Rose, C., Majola, M. S., Msomi, T., Matandela., F. and McLeod., N. (2021), A grassroots approach to spring protection in rural South Africa. South Africa: UCP. Available at: https://www.conservation.org/docs/default-source/south-africa-documents/spring-protection-guide-toolkit.pdf?Status=Master&sfvrsn=4fef1044_1

Rumpel, C. and Chabbi, A. (2021) 'Managing Soil Organic Carbon for Mitigating Climate Change and Increasing Food Security,' *Agronomy*, 11(8), p. 1553. Available at: https://doi.org/10.3390/agronomy11081553.

Salehi Hikouei, I., Kim, S.S. and Mishra, D.R. (2021) 'Machine-Learning Classification of Soil Bulk Density in Salt Marsh Environments', *Sensors*, 21(13), p. 4408. Available at: https://doi.org/10.3390/s21134408.

SANBI. (2019) *National Biodiversity Assessment 2018: Supplementary Material Compendium of Benefits of Biodiversity.* Compiled by Carol J. Poole. South African National Biodiversity Institute, Pretoria. Handle: http://hdl.handle.net/20.500.12143/6491

SANBI and Wildlands Conservation Trust. 2015. *Case Study: Biodiversity Partnership Area: Umzimvubu Catchment Partnership Programme*. Compiled by Botts, E.A. for the South African National Biodiversity Institute, Pretoria.

Seboko, K.R. *et al*. (2021) 'Characterization of Soil Carbon Stocks in the City of Johannesburg', *Land*, 10(1), p. 83. Available at: https://doi.org/10.3390/land10010083.

Shelukindo, H.B., Semu, E., Singh, B.R. and Munishi, P.K.T., 2014. Predictor variables for soil organic carbon contents in the Miombo woodlands ecosystem of Kitonga forest reserve, Tanzania. Available at: http://www.taccire.sua.ac.tz/handle/123456789/343.

Smith, W. *et al*. (2020) 'Development of the DNDC model to improve soil hydrology and incorporate mechanistic tile drainage: A comparative analysis with RZWQM2', *Environmental Modelling &Software*, 123, p. 104577. Available at: https://doi.org/10.1016/j.envsoft.2019.104577.

Sodango, T. et al. (2021) 'Modeling the Spatial Dynamics of Soil Organic Carbon Using Remotely-Sensed Predictors in Fuzhou City, China,' *Remote Sensing*, 13(9), p. 1682. Available at: https://doi.org/10.3390/rs13091682.

Stergiadi, M. *et al*. (2016) 'Effects of climate change and land management on soil organic carbon dynamics and carbon leaching in northwestern Europe', *Bio geosciences*, 13(5), pp. 1519–1536. Available at: https://doi.org/10.5194/bg-13-1519-2016.

Umzimvubu Catchment Partnership Programme. (2015). Biodiversity Stewardship Case Study 1: Umzimvubu Catchment Partnership. Available at: https://www.cepf.net/sites/default/files/06-2015_10_02-biodiversity-stewardship-casestudy1-ucpp.pdf. [Accessed 9 August 2023].

Van Huynh, C. et al. (2022) 'Application GIS and remote sensing for soil organic carbon mapping in a farm-scale in the hilly area of central Vietnam', *Air, Soil and Water Research*, 15, p. 117862212211147. Available at: https://doi.org/10.1177/11786221221114777.

Vargas-Rojas, R. *et al*. (2019) 'Unlocking the Potential of Soil Organic Carbon: A Feasible Way Forward,' in H. Ginzky *et al*. (eds) International Yearbook of Soil Law and Policy 2018. Cham: *Springer International Publishing (International Yearbook of Soil Law and Policy)*, pp. 373–395. Available at: https://doi.org/10.1007/978-3-030-00758-4_18.

Venter, Z.S. *et al*. (2021) 'Mapping soil organic carbon stocks and trends with satellite-driven high resolution maps over South Africa', *Science of The Total Environment*, 771, p. 145384. Available at: https://doi.org/10.1016/j.scitotenv.2021.145384.

Verra. (2015). VM0032 Methodology for the Adoption of Sustainable Grasslands through Adjustment of Fire and Grazing, v1.0. Retrieved from https://verra.org/methodologies/vm0032-methodology-for-the-adoption-of-sustainable-grasslands-through-adjustment-of-fire-and-grazing-v1-0/.

Wadoux, A.M.J.-C. et al. (2021) 'Spatial cross-validation is not the right way to evaluate map accuracy', *Ecological Modelling*, 457, p. 109692. Available at: https://doi.org/10.1016/j.ecolmodel.2021.109692.

Wadoux, A.M.J.-C., Minasny, B. and McBratney, A.B. (2020) 'Machine learning for digital soil mapping: Applications, challenges and suggested solutions', *Earth-Science Reviews*, 210, p. 103359. Available at: https://doi.org/10.1016/j.earscirev.2020.103359.

Wang, S. *et al*. (2023) 'Temporal and spatial changes in soil organic carbon and soil inorganic carbon stocks in the semi-arid area of northeast China,' *Ecological Indicators*, 146, p. 109776. Available at: https://doi.org/10.1016/j.ecolind.2022.109776.

Wang, G. *et al*. (2022) 'Effects of climate and grazing on the soil organic carbon dynamics of the grasslands in Northern Xinjiang during the past twenty years', *Global Ecology and Conservation*, 34, p.e02039. Available at: https://doi.org/10.1016/j.gecco.2022.e02039.

Wang, B. *et al*. (2018) 'Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia', *Ecological Indicators*, 88, pp. 425–438. Available at: https://doi.org/10.1016/j.ecolind.2018.01.049.

Wang, B., Waters, C., Orgill, S., Clark, A., Liu, D.L., Simpson, M., Cowie, A., McGowen, I. and Sides, T., (2017). Estimating soil organic carbon stocks using machine learning methods in the semi-arid rangelands of New South Wales. In *22nd International Congress on Modelling and Simulation: Hobart, Tasmania, Australia* (Vol.3). Available at: https://www.researchgate.net/publication/323212758.

Wiltshire, S. and Beckage, B. (2022) 'Soil carbon sequestration through regenerative agriculture in the U.S. state of Vermont', *PLOS Climate*. Edited by K. Byrne, 1(4), p. e0000021. Available at: https://doi.org/10.1371/journal.pclm.0000021.

Xia, Y., McSweeney, K. and Wander, M.M. (2022) 'Digital Mapping of Agricultural Soil Organic Carbon Using Soil Forming Factors: A Review of Current Efforts at the Regional and National Scales', *Frontiers in Soil Science*, 2, p. 890437. Available at: https://doi.org/10.3389/fsoil.2022.890437.

Zeraatpisheh, M. *et al*. (2022) 'Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates,' *CATENA*, 208, p. 105723. Available at: https://doi.org/10.1016/j.catena.2021.105723.

Zhang, Y. et al. (2022) 'Identifying the scale-controlling factors of soil organic carbon in the cropland of Jilin Province, China', *Ecological Indicators*, 139, p. 108921. Available at: https://doi.org/10.1016/j.ecolind.2022.108921.

Zhang, X. et al. (2022) 'Digital Mapping of Soil Organic Carbon with Machine Learning in Dryland of Northeast and North Plain China', *Remote Sensing*, 14(10), p. 2504. Available at: https://doi.org/10.3390/rs14102504.

Zhang, Y. and Niu, H. (2016) 'The development of the DNDC plant growth sub-model and the application of DNDC in agriculture: A review,' *Agriculture, Ecosystems & Environment*, 230, pp.271–282. Available at: https://doi.org/10.1016/j.agee.2016.06.017.

Zhu, C. et al. (2022) 'Digital Mapping of Soil Organic Carbon Based on Machine Learning and Regression Kriging,' *Sensors*, 22(22), p. 8997. Available at: https://doi.org/10.3390/s22228997.

Zomer, R.J. *et al*. (2017) 'Global Sequestration Potential of Increased Organic Carbon in Cropland Soils', *Scientific Reports*, 7(1), p. 15554. Available at: https://doi.org/10.1038/s41598-017-15794-8.