AUTHORS:
Omosalewa Odebiri[1]
Onisimo Mutanga[1]
John Odindi[1]
Kabir Peerbhay[1,2]
Steven Dovey[2]
Riyad Ismail[1,2]

AFFILIATIONS:
[1]Department of Geography, School of Agriculture, Earth and Environmental Science, University of KwaZulu-Natal, Durban, South Africa
[2]Sappi Forests, Howick, South Africa

CORRESPONDENCE TO:
Omosalewa Odebiri

EMAIL:
odebsconstant@gmail.com

Estimating soil organic carbon stocks under commercial forestry using topo-climate variables in KwaZulu-Natal, South Africa

Commercial forests are expanding globally, with great potential to absorb carbon and mitigate climate change. However, whereas the role of natural forests in carbon sequestration has been widely investigated, there is a paucity in the literature on the role of commercial forests in carbon assimilation. Hence, understanding the role of commercial forests in carbon storage is essential for quantifying local, regional or global carbon balances, which is valuable for climate change mitigation. Soil carbon is known to be the largest pool within any forest landscape, and is controlled by a wide range of physical and climatic factors. However, the relationship between soil organic carbon (SOC) and topo-climatic variables controlling its distribution within commercial forests is still poorly understood. Due to the limitations encountered in traditional systems of SOC determination, particularly at large spatial extents, geospatial techniques have recently emerged as a viable alternative for mapping soil properties. Therefore, this study sought to map SOC stocks variability within the commercial forest landscape, using landscape topo-climatic variables. A total of 81 soil samples was analysed for SOC concentrations and 31 topographic and climatic variables were used as predictors to SOC variability. To reduce multicollinearity, these variables were reduced to 11 using stepwise backward elimination and the maximum entropy (Maxent) algorithm was used for regression analysis to determine the relationship between SOC and the selected topo-climatic variables. Good accuracies were obtained for both training (area under the curve = 0.906) and test (area under the curve = 0.885) data sets, and demonstrate the effectiveness of selected topo-climatic variables and the Maxent algorithm in predicting SOC stocks. This study provides a framework for monitoring the status of soil carbon in commercial forest compartments and provides a viable approach for local, national or regional carbon accounting – valuable for climate change mitigation.

Significance:

- Rainfall and temperature, as well as topographic variables (such as slope, elevation and topographic wetness index) are effective in mapping SOC distribution.

- The model developed is useful in predicting SOC occurrence and yielded an effective framework for continuous monitoring and assessment of SOC.

- The method developed in this study is cost-effective and suggests the use of other readily available climatic and topographic information for the prediction of SOC under commercial forestry in South Africa and indeed globally.

- Results from this study are important to achieve the national carbon accounting objective and are also valuable to forest managers, ecologists and relevant stakeholders in understanding the spatial distribution of SOC.

Introduction

Commercial forests represent a large carbon pool with the potential to reduce net greenhouse gas emissions.[1] The world's commercial forests increased by over 105 million ha between 1990 and 2015, while natural forests decreased by 234 million ha.[2] Federici et al.[3] notes that carbon held by commercial forests was comparable to that of natural forests (i.e. 1.08 vs. 1.44 gigatonne CO_2 per year) during this period. Recent projections have also shown that commercial forest plantations will increase by 20 to 50% by the year 2030.[4,5] Hence, commercial forests are increasingly becoming valuable environmental assets. Specifically, commercial forests could be used to reduce harvest pressure on remnant natural forests, restore degraded ecosystems, conserve natural resources and design climate change mitigation policies.[4]

Carbon in commercial forest ecosystems is stored in five different pools: above- and below-ground live tree biomass, dead wood, litter and soils.[6] These pools are dynamic and controlled by a wide range of environmental factors such as climate, topography, forest type, moisture, temperature, soil type and land use.[7] Within these pools, soil is the largest carbon reservoir and constitutes 50–80% of total carbon stocks.[6] As indicated by Liu et al.[8], the global soil organic carbon (SOC) stock has been evaluated to be about 1500 Pg carbon in the upper 100-cm soil layer, which is approximately double the measure of carbon in the atmosphere and thrice the amount stored in terrestrial vegetation.[9] However, whereas the above-ground biomass carbon pools have been widely investigated, the spatial variability of SOC within forest landscapes is still poorly understood.[10] Hence, an in-depth understanding of SOC and its variability in relation to topographical and environmental factors is crucial in quantifying regional and global carbon balances and in examining the responses and feedbacks of the terrestrial ecosystem to climate change.[8] Furthermore, knowledge of SOC variability is useful for developing suitable management strategies to improve carbon assimilation[11] and to achieve total annual national and global carbon accounting objectives as well as Intergovernmental Panel on Climate Change and Kyoto Protocol objectives.

Environmental variables are critical to the spatial distribution of SOC within a forest landscape.[12] Some of these environmental factors occur naturally while others are human-induced.[13] Variables such as topography (e.g. elevation, slope and aspect) and climate (e.g. temperature and rainfall) significantly influence SOC distribution within a forest landscape.[14] Studies have also shown that climate and topography exert a strong influence on the amount of vegetation density within a forest landscape, which in turn determines SOC distribution.[15,16] However, Chaplot et al.[11] and Liu et al.[8] note that large-scale investigations on the connection between SOC distribution and the influence of multiple topo-climatic variables at a regional and global level are not sufficient. Explicitly significant relationships and models that link topo-climatic variables with SOC processes at large spatial extents are still very necessary.[17]

Traditionally, SOC has been quantified through, among others, field surveys and wet soil analysis.[18] Although these strategies are highly accurate, they are difficult to conduct over large areas, expensive, time-consuming and labour-intensive, and may lead to the generation of toxic waste such as chromate oxidation, which requires careful and proper disposal.[19] Geospatial techniques, on the other hand, offer more practical and economical means of predicting and quantifying soil parameters at local, regional and even global scales.[8,20]

Recently, the adoption of spatial techniques in SOC studies has attracted significant attention[8,14,21,22], particularly because spatially continuous topographic metrics that affect SOC distribution (e.g. slope, aspect, elevation, curvatures, catchment area) can now be easily generated from digital elevation models (DEMs) and satellite imagery.[14] Climatic variables such as rainfall and temperature can also be derived from WorldClim database (global climate layers) at a spatial resolution of about 1 km^2.[23] The use of DEMs derived from satellite missions such as Light Detection and Ranging (LiDAR) and Shuttle Radar Topography Mission (SRTM) for terrain analysis has become particularly popular due to their relatively high resolutions.[24] Hence, the development of models with SRTM and LiDAR-derived topographic metrics in concert with selected bioclimatic variables could benefit studies characterised by limited observations and can be used to produce continuous SOC distribution.[21] Additionally, remotely sensed derived models (e.g. DEMs, Landsat and Sentinel sensors) are useful in determining regional and global land conditions, impacts and efficacy of land management and better at characterising forest landscapes.[14] Therefore, the objective of this study was to explore the use of remotely sensed SRTM DEM derived topographic metrics and other bioclimatic variables in the prediction of SOC stocks under commercial forestry in KwaZulu-Natal, South Africa.

Materials and methods

Study site

Commercial forest plantations in South Africa cover approximately 1.4 million hectares, representing approximately 1.1% of the country's landmass.[6,25] South African commercial forest plantations are mostly found in the eastern part of the country, stretching from the Western Cape through the Eastern Cape, to the KwaZulu-Natal, Mpumalanga and Limpopo Provinces.[6,25,26] South Africa's commercial forest plantations are of great economic and environmental value. Commercial forestry contributes about 1.27% to South Africa's gross domestic product, provides direct and indirect employment to about half a million people, supplies South African forest and fibre needs of approximately 17 million air dry tons annually, and sequestrates about 4.1 million tons of CO_2 per year.[6,27] The country's forest plantations are predominantly pine (49.8%), eucalyptus (42.7%) and wattle (7.1%) species.[25]

This study was conducted across commercial forest plantations in KwaZulu-Natal Province – a region located along the east coast of South Africa (Figure 1). The Province is an important agricultural and forestry production area and is characterised by different forest species. The dominant commercial forest species in the region are *Eucalyptus* (hardwood) and *Pinus* (softwood) species, which cover approximately 371 034 and 115 922 ha, respectively.[25] The area experiences sub-tropical climatic conditions, with the main rainfall from October all through to March. The mean annual temperature of the study region is approximately 21.7 °C while the mean annual rainfall ranges from 700 mm to 1500 mm; thus conditions are

favourable for commercial forestry farming.[20,28] The topography of the area is generally characterised by undulating plains and the altitude rises from 800 m to 1400 m above mean sea level. The zone is underlain by geological formations including granite, arenite, basalt, tillite, shale, sandstones and mudstones, which result in clay to pure sandy soils.[19]
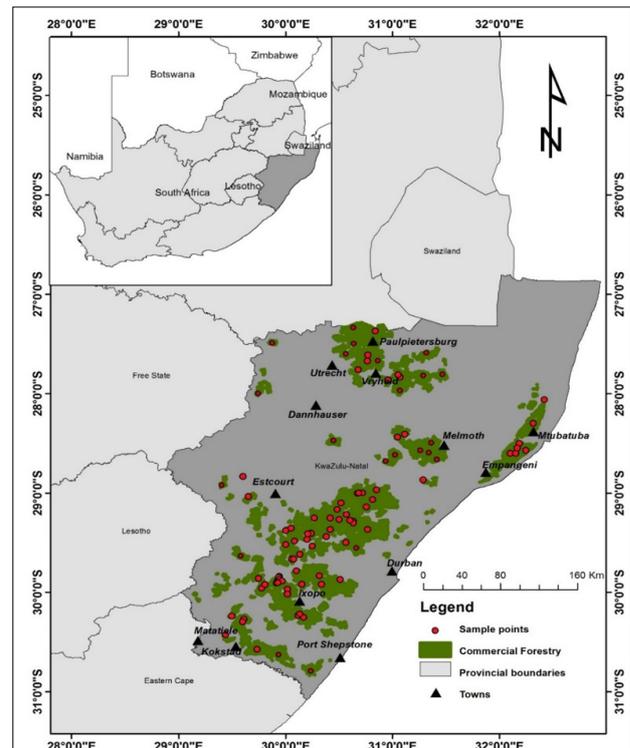

Figure 1: Commercial plantations located within KwaZulu-Natal Province, South Africa.

Field data collection and analysis

SOC analysis and quantification is an important part of forest resource management.[6] The procedure of soil sample collection in the field should be typical of the area being sampled because the utilisation of the obtained results from the laboratory analysis relies on the sampling accuracy.[29,30] Commercial forest sites are categorised into broad regions of site quality (in reference to productivity), i.e. poor, medium and good [26], hence soil samples must be taken to represent each identified site quality in the field. In this study, field data were collected during the rainy summer season (January and March 2013) – a period characterised by high above-ground biomass productivity. A stratified random sampling method was adopted, in which the area was divided into three different strata (i.e. low, medium and high productivity) based on prior knowledge of the commercial forest stands. The purpose of using a stratified random sample is to get a more representative sample of all the regions of commercial forest site qualities. A Hawth's Analysis tool in ArcGIS version 10.4.1 was used to generate random sample plots on a commercial forest cover map across all site qualities. These sample plots were allotted to the three various strata of homogeneous vegetation according to the stratified random sampling strategy and then uploaded into a handheld GPS (TRIMBLE GEO-7X) which was used to navigate to the field sites. Once the sample plots were located on the sites, soil was dug at each sample points to a depth of 30 cm using a soil auger, which is the recommended depth in spatial SOC inventories.[10,29,31] In total, 81 accessible soil samples were collected and transported to the laboratory where they were processed and analysed. SOC concentration was determined in the laboratory using the Walkley–Black[32] dichromate oxidation method.

Environmental variables

Topographic metrics

Previous studies have identified topographic variables as critical drivers to SOC distribution.[14,21] According to Li et al.[14], spatial topographic variables

are categorised into three main groups: local, non-local and combined topographical variables. Local topographical variables examine the surface geometry at a point on the land surface such as slope, elevation and curvatures while non-local attributes portray relative locations of selected points, such as relief, catchment area, openness and flow accumulation. Combined attributes are an integration of both the local and non-local topographic variables such as slope length factor, topographic wetness index (TWI) and stream power index.[14] In this study, 29 different topographical variables (Table 1) that cut across the three classes (i.e. local, no-local and combined) were selected. These variables were derived from a 30-m resolution DEM created from SRTM data in SAGA GIS (2.3.2) and ArcGIS 10.4 software.

Table 1: Topo-climatic variables used in predicting the occurrence of soil organic carbon

S/n	Variable	Description	Unit
1	Topographic wetness index (TWI)	Steady state wetness index	–
2	Elevation (DEM)	Ground height	m
3	Slope	Steepness of the ground	radian
4	Aspect	Slope direction	radian
5	Mass balance index (MBI)	Mass balance index	m
6	Normalised height	Relative height and slope position	m
7	Standardised height	Relative height and slope position	m
8	Direct insolation	Potential incoming insolation	kw/m^2
9	Diffuse insolation	Solar radiation	kw/m^2
10	Visible sky	Lightening and sky view factor	kw/m^2
11	Sky view factor	Visibility	kw/m^2
12	Negative openness	Drainage features, soil water content	–
13	Positive openness	Drainage features, soil water content	–
14	LS Factor	Slope length factor effect on erosion	–
15	Catchment area	Run-off velocity and volume	m^2
16	Catchment height	Elevation above ground	m
17	Terrain ruggedness index (TRI)	Quantifies topographic heterogeneity	–
18	Plan curvature	Horizontal (contour) curvature	degrees/m
19	Profile curvature	Vertical rate of change of slope	degrees/m
20	Longitudinal curvature	Morphometric features	degrees/m
21	Cross-sectional curvature	Morphometric features	degrees/m
22	Minimum curvature	Lowest curvature	degrees/m
23	Maximum curvature	Highest curvature	degrees/m
24	General curvature	Horizontal and vertical curvature	degrees/m
25	Valley depth	Relative height	m
26	Terrain surface texture (TST)	Terrain texture of the surface	–
27	Topographic position index (TPI)	Location higher or lower than the average of the surroundings	–
28	Wind effect	Effect of wind on the surface	m/s
29	Convergence index	Shows structure of relief as a set of convergent (channels) and divergent (ridges) areas	m
30	Temperature (mean annual)	Temperature	°C
31	Rainfall (mean annual)	Rainfall	mm

Bioclimatic data

Bioclimatic data sets including rainfall and temperature are critical determinants of SOC.[13,23] In this study, mean temperature and rainfall bioclimatic variables were used in concert with topographic variables to predict SOC occurrence. The bioclimatic variables were obtained from the 1-km² 30 arc-seconds spatial resolution of the WorldClim archives (http://www.worldclim.org/) of the global climate conditions. The WorldClim climatic data sets are long-term (30-year) mean annual measurements (containing grids including rainfall and temperature as well as other climatic layers summaries such as the wettest, driest, coldest and hottest quarters and months of the year. The derived temperature (Bio01) and rainfall (Bio12) bioclimatic layers used in this study were resampled in order to match the SRTM-derived DEM spatial resolution (i.e. from 1 km² to 30 m).

Statistical analysis

To determine the relationship between SOC and the derived environmental variables, SOC within the commercial forest stands was predicted using the maximum entropy (Maxent) regression algorithm software (version 3.3.3).[33]

Maximum entropy model

The maximum entropy model (Maxent) is a machine-learning algorithm proposed by Phillips et al.[33] Maxent models the probability of species presence based on environmental constraints and estimates the likelihood distribution with the maximum entropy, i.e. the most spread out distribution. Typically, Maxent produces an estimate of a probability of occurrence that ranges from 0 to 1, with 1 being the highest and 0 the lowest. It is a concise mathematical definition, and therefore can be adjusted to analyse data with efficient deterministic algorithms that are certain to give optimal probability distribution. Maxent performs efficiently, even with small sample sizes, is resistant to errors in occurrence data, and applies the presence-only data sets to estimate the suitability of habitat or the likelihood of target occurrence.[34] At the point where absence data exist for the species, a conditional model can be used to enable presence/absence modelling.[33] Maxent uses background/pseudo-absence and presence points that evaluate the environmental space for model testing. Environmental variables (continuous and categorical) and species-presence data are used to run the model and the influence of each variable can be determined from the jackknife tool in Maxent.[33]

In SOC modelling, the Maxent model begins with equal distribution and performs a number of repetitions based on the most important predictor variable until no further improvements in the spatial estimation of SOC are made.[35] The Maxent model aims to identify the maximum likelihood variability of SOC occurrences within the commercial forest stands in the study area.[36] The most uniform spread of the SOC occurrences is consequently identified by the model and selected from among many possible distributions.[37]

Predictor variables selection

One common limitation with regression is the issue of multicollinearity, which occurs when two or more predictor variables are highly correlated. Hence, it is often advisable to use the best and fewest number of predictor variables useful in building a model.[38] Stepwise regression was adopted to solve any possible multicollinearity and to select the best and fewest predictive variables for the Maxent model. Stepwise regression identifies the statistical importance of a subset of predictors through forward selection, backward elimination or a combination of both.[39] In this study, the stepwise backward elimination method was conducted using the 'stepAIC' function in the 'MASS' package of R statistic software 3.5.1 to select the best predictor variables which were then used for the Maxent final model. Backward elimination works by removing predictor variables ($n=31$) that are not significant to the model until the ideal number of predictor variables is obtained. The ideal number of variables was selected after the backward elimination process. These selected variables were then used to predict SOC using Maxent (version 3.3.3).

Model calibration, evaluation and validation

Settings are an integral part of building any model in order to get the best results. The SOC data ($n=81$) used in this study were partitioned into training (70%) and test (30%) data sets.[28] The training data ($n=54$) were used in the model building while the test data ($n=23$) were used for model validation as proposed by Phillips et al.[35] The default Maxent

settings were used to run the model. Prior to obtaining final results, visual inspection of the response curve and the difference between test and training area under the curve (AUC) values of the model were assessed to determine overfitting. The default regularisation was adopted as there was no overfitting detected. For the probabilistic model output, the 10-percentile train occurrence logistics limit command in Maxent was utilised to generate a SOC variability map. This threshold makes certain that 90% of the occurrence data have been predicted as present and that the omission error does not surpass 10%.[40]

Evaluation and validation are critical steps in any model building process. In this study, we used the random test percentage settings in Maxent, i.e. 70% of the data set was used to train and 30% to test the performance of the model. The receiver operating characteristic (ROC) curve was used to evaluate and validate the model. The ROC curve is a method that describes the performance level of probabilistic and deterministic detection of forecast systems.[36] It shows the likelihood that presence (sensitivity) is correctly ordered by the classifier (in our case predictors) as compared to the absence (specificity) of SOC. A two-dimensional space is used to generate the ROC curve by plotting the sensitivity as Y and the specificity as X. Generally, models with high accuracy have an AUC value close or equal to 1, whereas a value equal or less than 0.5 shows a model that performs no better than random.[33] Previous studies have broadly and successfully demonstrated the application of the ROC curve to quantitatively evaluate and validate the effectiveness of probability modelling.[14,21,36]

Results

Predictor variables selection

Table 2 shows the results of the stepwise backward elimination method conducted to remove redundant and highly correlated predictor variables. The procedure to select the best predictor variables was done using the 'stepAIC' function in the 'MASS' package of the R statistic software (version 3.5.1). The Akaike information criterion (AIC) acts like an examiner of the relative quality of models for a given set of data by assigning a value to the model (in this case 5.61), while the stepwise backward elimination method eliminates predictor variables that are less significant at each stage of the model until the lowest AIC value is attained (in this case -15.85). The general rule is that the lower the AIC value, the better the model. As shown in the table, 11 out of the 31 predictor variables used produced the lowest AIC value of -15.85 after the elimination procedure. Table 2 additionally shows the AIC value attached to each of the eleven selected variables, indicating the degree to which the lowest AIC value (-15.85) will increase should any of the variables be removed. Hence these selected variables are regarded as the best subset of predictors to be used in the prediction of SOC variability using the Maxent algorithm.

Table 2: Change in the Akaike information criterion (AIC) value with removal of each variable

S/N	Selected predictor variable	AIC = -15.8546
1	Longitudinal curvature	-15.7907
2	Aspect	-15.6924
3	Direct insolation	-14.6213
4	Digital elevation model (DEM)	-14.5480
5	Positive openness	-13.7504
6	Catchment area	-13.5806
7	Profile curvature	-13.3244
8	Topographic wetness index (TWI)	-10.5943
9	Rainfall (mean annual)	-7.1520
10	Slope	-5.9165
11	Temperature (mean annual)	8.6211

Table 3: Contribution (%) of each variable in predicting the occurrence of soil organic carbon

Variable	Contribution (%)
Rainfall (mean annual)	30.2
Temperature (mean annual)	22.9
Slope	16.1
Digital elevation model (DEM)	11.6
Topographic wetness index (TWI)	8.9
Direct insolation	3.1
Catchment area	2.9
Positive openness	1.8
Aspect	1.1
Profile curvature	0.8
Longitudinal curvature	0.6

Maxent model

Analysis of Maxent model omission/commission and ROC

The omission/commission and the ROC analysis is useful in examining the performance of the Maxent model. The omission/commission rate is calculated on both the training and test data sets. The general rule of Maxent is that a good model will produce omission rates that are in close proximity with the predicted omission due to the definition of the cumulative threshold. The predicted omission rate is demonstrated by a straight black line on the cumulative threshold result output of the Maxent model. As shown in Figure 2a, the omission on test data sets (turquoise line) is in close proximity with the predicted omission rate (black line) from the Maxent distribution, which signifies a good Maxent model for SOC occurrence.

Figure 2b shows the model evaluation for predicting SOC occurrence using the ROC of the randomly selected training and test data sets. The AUC ranges between 0 and 1. A good model will be closer or equal to 1 while a model with an AUC less than or equal to 0.5 signifies a poor model and was no better than random. As shown in Figure 2b, the Maxent model accomplished a high accuracy, giving AUC values of 0.906 and 0.885 for training and test data sets, respectively, and thus indicating that the developed model performed better than a random model ($p < 0.005$).

Predictor variables contribution

A major strength of the Maxent algorithm is that it permits the assessment of individual predictor variables in order of their influence in the model. Table 3 shows the estimated percentage contribution of each predictor variable to the Maxent model. The increase in regularised gain is added to the contribution of the corresponding variable or subtracted from it in order to determine the estimates in each iteration of the training algorithm. The higher the percentage contribution of any of the variables, the higher its impact in predicting the SOC occurrence. The percentage contribution table generated by the Maxent model shows that 'rainfall (mean annual)' with a percentage contribution of 30.2% had the highest predictive contribution and was the most influential in predicting SOC occurrence. Also, variables such as temperature (22.9%), slope (16.1%), elevation (11.6%), TWI (8.9%) and direct insolation (3.1%) were contributors to SOC occurrence and accounted for over 60% of the total SOC occurrence. Predictor variables such as aspect, profile and longitudinal curvature had little influence on the Maxent model.

Figure 3 shows the jackknife test results; the jackknife test is used to assess each predictor variable contribution in order of importance to the Maxent model. The turquoise bars indicate the overall model accuracy when each of the predictor variables is excluded from the model, while the blue bars show the individual performance and accuracy of predictors when used in isolation. The red bars depict the overall gain of the model when all the variables are used together. As shown in Figure 3, 'rainfall (mean annual)' is the environmental predictor variable that has
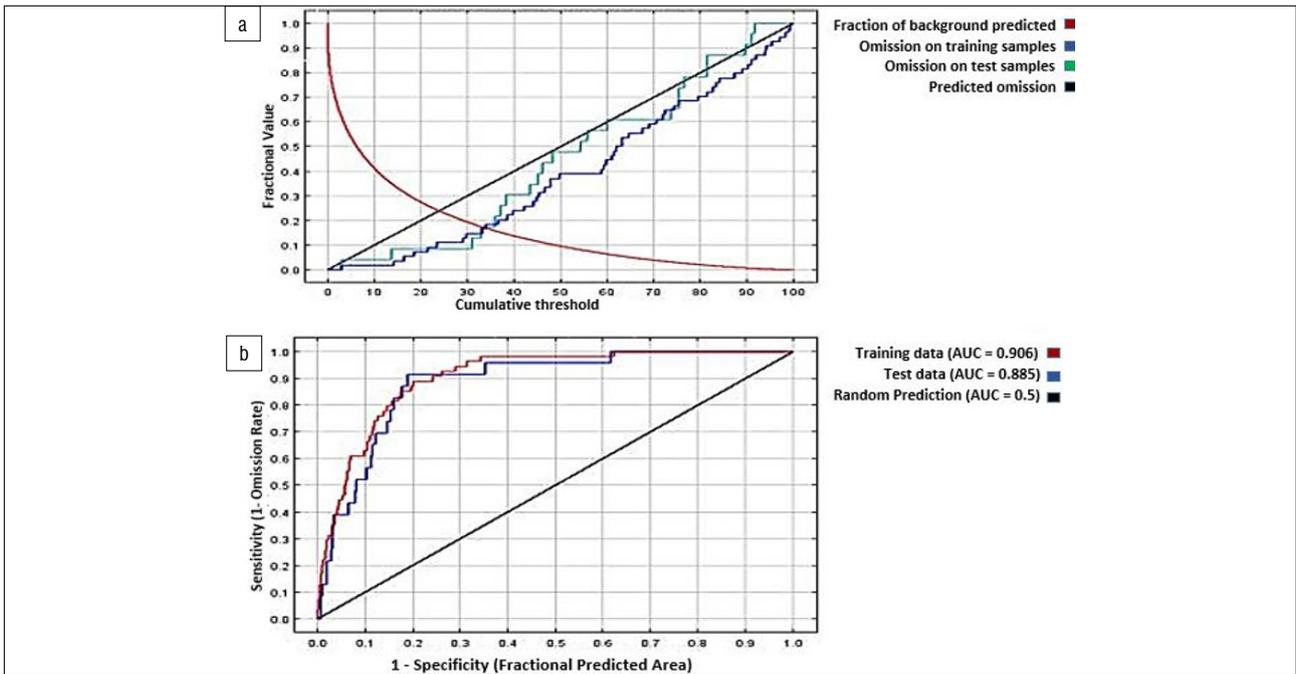
Figure 2: (a) Maxent test and training data sets omission analysis, as well as the predicted region for SOC occurrence. (b) Receiver operating characteristic (ROC) for training and test data sets using the area under the curve (AUC).

the highest gain when utilised in isolation and, as a result, appears to have the most valuable information independent of any other variable. The environmental variable that most reduced the Maxent model overall gain when omitted was 'temperature (mean annual)', which as a result appears to have the bulk of information that is absent in other variables. Variables such as slope, DEM and TWI also had an impact on the Maxent model gain as there was a significant decrease in the final model gain (red bar) when they were excluded. Other variables such as aspect, catchment area, curvatures, direct insolation and positive openness, had little or no significant contribution to the final model and hence are regarded as unimportant in predicting SOC occurrence in this study.
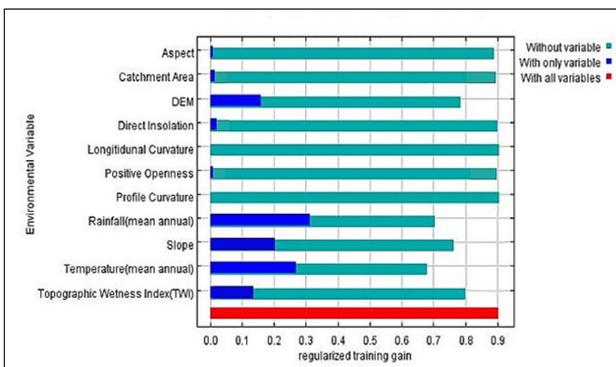

Figure 3: The jackknife of variable importance in modelling the spatial distribution of soil organic carbon.

Map generation of SOC distribution

The Maxent model generated a map that shows the spatial distribution of SOC occurrence within the study area. Figure 4 shows the likely occurrence of SOC based on the field observation points using topographic metrics and bioclimatic data as predictor variables. Areas with darker colours indicate better-predicted conditions of SOC occurrence while areas with lighter colours indicate moderate or low predicted conditions for SOC occurrence. A general assessment of the resultant SOC distribution map shows that the possible occurrence of SOC within the study region is relatively high in the southern and central plantations as compared to plantations located at the northernmost and the far eastern regions of the Province.
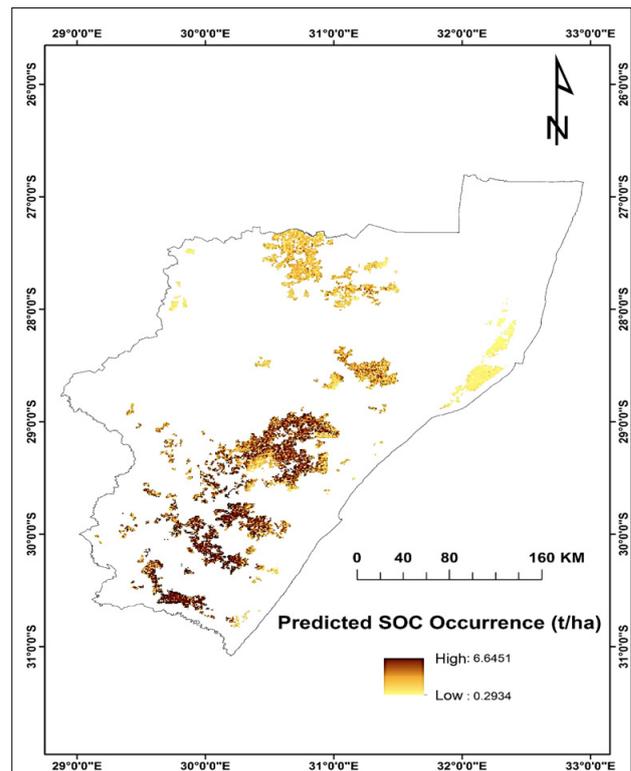

Figure 4: Map showing the spatial distribution of soil organic carbon occurrence produced by the Maxent model.

Discussion

The global expansion of commercial forestry and its potential to sequestrate carbon is increasingly becoming important in climate change mitigation.[6] Hence, there is a growing interest in commercial forest management practices to further enhance the ability to sequestrate carbon because of the carbon assimilation efficiency of commercial forests.[41] Although commercial forest soils sequestrate and store more carbon than other carbon pools, their role in carbon sequestration remains largely unexplored.[6,12,41] Therefore, we sought to estimate SOC

within commercial forestry by utilising a range of environmental variables that include topographic metrics and bioclimatic variables in a Maxent environment. Results in this study show that environmental variables such as topography and climate can be used to effectively model SOC spatial variability in a Maxent environment (AUC 0.885). The Maxent model's percentage contribution and jackknife results showed that 5 of the 11 predictor variables contributed significantly to the accuracy of the Maxent model, and hence could be used to determine SOC occurrence, distribution and variability in commercial forestry. These variables in order of importance were: rainfall, temperature, slope, elevation and TWI.

Rainfall, which contributed 30.2% to the Maxent model, was the most significant determinant of SOC occurrence in the study area. Forest stands with higher SOC occurrence are characterised by higher mean annual rainfall. This finding is consistent with that of Meier and Leuschner[42] who observed a substantial reduction (25%) in the SOC pool in forests stands that received less than 600 mm rainfall annually compared with stands that received more than 900 mm rainfall annually. Other studies, such as those of Jobbágy and Jackson[43], Bhandari and Bam[7], Chen et al.[2], Ramifehiarivo et al.[13], Hewins et al.[23] and Soucémarianadin et al.[12], also noted that mean annual rainfall within a forest landscape strongly influences the amount of SOC. The relationship between rainfall and SOC occurrence can be attributed to the influences of rainfall on soil moisture and hydrological processes such as surface run-off and groundwater infiltration.[44] Zhou et al.[45] and O'Brien et al.[46] noted that soil moisture determines SOC accumulation as it affects the vegetation density and decomposition. Furthermore, rainfall influences the presence of decomposers.[47]

The mean annual temperature also significantly influenced the prediction of SOC occurrence as it contributed 22.9% in the Maxent model. Forest stands characterised by warmer temperatures accelerate SOC mineralisation, compared with stands with lower temperatures.[48] A similar finding was reported by Wang et al.[49] who observed that soil samples incubated at different temperatures (15 °C and 75 °C) indicated higher SOC mineralisation with increasing incubation temperature. Previous investigations[13,23] have also reported temperature to be strongly linked with SOC, as it can promote plant productivity, which in turn promotes the presence of SOC. Furthermore, the influence of temperature as one of the major drivers of SOC can also be attributed to its direct relationship with rainfall, because rainfall may influence soil moisture, and by extension the surface temperature, by controlling the partitioning between sensible and latent heat fluxes.[50] In addition, the amount of rainfall within a forest landscape can also be influenced by temperature because higher temperatures signify more water vapour in the atmosphere, thus increasing the chances of heavy downpours.[51]

Slope (16.1%), elevation (11.9%) and topographic wetness index (8.9%) also showed influence on SOC. Areas with relatively steep slopes – such as the northernmost and the far eastern parts of the study area – had lower SOC occurrence than areas with relatively gentle slopes, such as the southernmost and central parts of the commercial forest areas. This finding can be attributed to higher erosion rates, which are commonly observed in areas of steeper slope as opposed to areas with less steep or gentle slopes. Erosion transports soil properties such as SOC from the upland to low-lying areas.[52] Previous studies by Li et al.[14], Fissore et al.[21], Young et al.[52] and García-Ruiz[53], among others, all noted slope as a strong determinant of SOC distribution. Elevation was also a significant influence on SOC occurrence in commercial forestry. Low likelihood SOC occurrence was predicted for areas with higher altitude (>1300 m) while high probabilities were predicted for low lying areas (<1300 m). These findings are consistent with another study[54] in which it was noted that with increasing altitude, the amount of soil and vegetation became less abundant, which was also reflected in the amount of SOC. The variability of SOC due to elevation can be attributed to the fact that low-lying areas favour vegetation growth due to optimal soil development conditions that may include erosion of nutrient-rich topsoil from higher grounds that are deposited in low-lying areas. Additionally, most low-lying areas are characterised by higher soil moisture content, nutrients and deeper depth as opposed to higher grounds characterised by extreme environmental conditions, hindering vegetation growth due to limited soil microorganisms.[55] Furthermore, the influence of elevation on microclimate

could also be the reason for SOC occurrence in our study as altitude influences temperature, wind flows and soil moisture peculiar to a region, which in turn impacts the amount of vegetation and by extension SOC.

Although TWI contributed less than 10% in the Maxent model, it can still be regarded as an important driver of SOC occurrence due to its influence on the overall gains shown in the jackknife results (Figure 3). TWI determines soil moisture distribution along slopes[14], hence areas with higher TWI (soil moisture) indicate higher SOC density than areas with low TWI. Previous studies[56-58] have also reported the significance of TWI in SOC distribution by observing a strong correlation between TWI and soil organic matter. For instance, Li et al.[14] identified TWI as the most significant topographic predictor variable in SOC variability.

Results indicated relatively little impact of longitudinal and profile curvatures, catchment area, aspect, direct insolation and positive openness to SOC variability. This finding is in contrast to other investigations that demonstrated their importance to SOC formation and distribution.[14] One possible reason for the significance of these environmental factors in other studies could be due to landscape variability in relation to our study area.[21,22,39] For instance, curvatures are generally more sensitive to places of higher relief than relatively moderate landscape and have been broadly used to depict flow acceleration.[14] As a result, curvature plays a significant role in the variability patterns of SOC in regions of higher landscapes.

The present study showed that the utility of Maxent based on key topographic and bioclimatic variables provides a useful and effective methodology for predicting SOC under commercial forest landscapes. Maxent also demonstrated the ability to show the percentage contribution of each predictor variable and their influence through the analysis of the jackknife results (Table 3; Figure 3). It also automatically generated a visually appealing SOC distribution map (Figure 4) and accuracy assessments by producing the ROC curve, which is used to determine its predictive performance. Regardless of the benefits of the Maxent algorithm, it has some limitations. Maxent's logistic output relies on an assumption and not an estimation of SOC prevalence. It is also hard to compare the outputs of Maxent with other regression algorithms as it gives environmental suitability rather than predicted likelihood of SOC occurrence.

Conclusion

In this study, we investigated the impact of topographic and bioclimatic data on SOC distribution under commercial forestry using the Maxent algorithm. Results indicate that rainfall and temperature, as well as topographic variables such as slope, elevation and TWI, are effective in mapping SOC distribution. Our model was useful in predicting SOC occurrence and yielded an effective framework for continuous monitoring and assessment of SOC. The method developed in this study is cost-effective and suggests the use of other readily available climatic and topographic information for the prediction of SOC under commercial forestry in South Africa and indeed globally. Results from this study are important to achieve the national carbon accounting objective and are also valuable to forest managers, ecologists and relevant stakeholders in understanding the spatial distribution of SOC. We recommend further experiments be conducted using higher-resolution data sets to assess the performance of the Maxent algorithm in predicting SOC.

Acknowledgements

Authors' contributions

O.O.: Conceived and conceptualised the study, participated in field data collection and subsequent visits to the site for ground-truthing, developed the theoretical formalism, performed the analytic calculations as well as the numerical simulations, wrote the initial manuscript draft including substantive translation. O.M.: Verified the analytical methods as well as supervised the findings of this work, contributed to results interpretation and discussion as well as participated in writing the final

manuscript, provided critical feedback and helped shape the research, analysis, and manuscript. J.O.: Participated in the formulation of the aim and objective of the study, co-supervised and contributed to the writing of every section of the manuscript, aided in interpreting the overall replication/reproducibility of results/experiments and other research outputs. K.P.: Designed field sampling procedures and techniques, coordinated the field protocol and data collection, assisted in laboratory analysis of the collected field data, participated in the overall results and discussion. S.D.: Spearheaded the laboratory analysis of collected field data, participated in the methodology section of the manuscript as well as the overall results and discussion. R.I.: Assisted in site design, coordination, and supervision, participated in subsequent site visits for ground-truthing, coordinated site mobility and equipment movement, participated in the interpretation and discussion of results.

References

1. Eggleston HS, Buendia L, Miwa K, Ngara T, Tanabe K. IPCC Guidelines for National greenhouse gas inventories. Japan: N. p.; 2006. Available from: https://www.ipcc-nggip.iges.or.jp/public/2006gl/

2. Chen X, Zhang D, Liang G, Qiu Q, Liu J, Zhou G, et al. Effects of precipitation on soil organic carbon fractions in three subtropical forests in southern China. J Plant Ecol. 2015;9(1):10–19. https://doi.org/10.1093/jpe/rtv027

3. Federici S, Tubiello FN, Salvatore M, Jacobs H, Schmidhuber J. New estimates of CO$_2$ forest emissions and removals: 1990–2015. Forest Ecol Manag. 2015;352:89–98. https://doi.org/10.1016/j.foreco.2015.04.022

4. Payn T, Carnus J-M, Freer-Smith P, Kimberley M, Kollert W, Liu S, et al. Changes in planted forests and future global implications. Forest Ecol Manag. 2015;352:57–67. https://doi.org/10.1016/j.foreco.2015.06.021

5. Penna I. Understanding the FAO's 'Wood Supply from Planted Forests' Projections. Ballarat, Victoria: Centre for Environmental Management, University of Ballarat; 2010.

6. Dovey SB. Current carbon stock estimation capability for South African commercial forest plantations. Report produced for the Department of Environmental Affairs. Pietermaritzburg: Institute for Commercial Forestry Research; 2014. Available from: https://www.environment.gov.za/sites/default/files/reports/carbonstocks_report2015_southafrica.pdf

7. Bhandari S, Bam S. Comparative study of soil organic carbon (SOC) under forest, cultivated and barren land: A case of Chovar village, Kathmandu. Nepal J Sci Technol. 2013;14(2):103–108. https://doi.org/10.3126/njst.v14i2.10422

8. Liu Z, Shao M, Wang Y. Effect of environmental factors on regional soil organic carbon stocks across the Loess Plateau region, China. Agric Ecosyst Environ. 2011;142(3–4):184–194. https://doi.org/10.1016/j.agee.2011.05.002

9. Smith P. How long before a change in soil organic carbon can be detected? Glob Chang Biol. 2004;10(11):1878–1883. https://doi.org/10.1111/j.1365-2486.2004.00854.x

10. Muchena R. Estimating soil carbon stocks in a dry miombo ecosystem using remote sensing. Forest Res. 2017;6(1), Art. #198, 6 pages. https://doi.org/10.4172/2168-9776.1000198

11. Chaplot V, Bouahom B, Valentin C. Soil organic carbon stocks in Laos: Spatial variations and controlling factors. Glob Chang Biol. 2010;16(4):1380–1393. https://doi.org/10.1111/j.1365-2486.2009.02013.x

12. Soucémarianadin LN, Cécillon L, Guenet B, Chenu C, Baudin F, Nicolas M, et al. Environmental factors controlling soil organic carbon stability in French forest soils. Plant Soil. 2018;426:267–286. https://doi.org/10.1007/s11104-018-3613-x

13. Ramifehiarivo N, Brossard M, Grinand C, Andriamananjara A, Razafimbelo T, Rasolohery A, et al. Mapping soil organic carbon on a national scale: Towards an improved and updated map of Madagascar. Geoderma Regional. 2017;9:29–38. https://doi.org/10.1016/j.geodrs.2016.12.002

14. Li X, McCarty GW, Karlen DL, Cambardella CA. Topographic metric predictions of soil redistribution and organic carbon in Iowa cropland fields. Catena. 2018;160:222–232. https://doi.org/10.1016/j.catena.2017.09.026

15. Bhunia GS, Shit PK, Pourghasemi HR. Soil organic carbon mapping using remote sensing techniques and multivariate regression model. Geocarto Int. 2019;34(2):215–226. https://doi.org/10.1080/10106049.2017.1381179

16. Szakács GGJ, Cerri CC, Herpin U, Bernoux M. Assessing soil carbon stocks under pastures through orbital remote sensing. Sci Agric. 2011;68(5):574–581. http://dx.doi.org/10.1590/S0103-90162011000500010

17. Fiener P, Dlugoß V, Van Oost K. Erosion-induced carbon redistribution, burial and mineralisation – Is the episodic nature of erosion processes important? Catena. 2015;133:282–292. https://doi.org/10.1016/j.catena.2015.05.027

18. Wang G, Luo Z, Han P, Chen H, Xu J. Critical carbon input to maintain current soil organic carbon stocks in global wheat systems. Sci Rep. 2016;6, Art. #19327, 8 pages. http://doi.org/10.1038/srep19327

19. Mzinyane T, Van Aardt J, Gebreslasie MT. Soil carbon estimation from *Eucalyptus grandis* using canopy spectra. S Afr J Geomatics. 2015;4(4):548–561. http://dx.doi.org/10.4314/sajg.v4i4.12

20. Dube T, Mutanga O. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. ISPRS J Photogramm Remote Sens. 2015;101:36–46. https://doi.org/10.1016/j.isprsjprs.2014.11.001

21. Fissore C, Dalzell B, Berhe A, Voegtle M, Evans M, Wu A. Influence of topography on soil organic carbon dynamics in a Southern California grassland. Catena. 2017;149:140–149. https://doi.org/10.1016/j.catena.2016.09.016

22. Ließ M, Schmidt J, Glaser B. Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches. PLoS ONE. 2016;11(4), e0153673, 22 pages. https://doi.org/10.1371/journal.pone.0153673

23. Hewins DB, Lyseng MP, Schoderbek DF, Alexander M, Willms WD, Carlyle CN, et al. Grazing and climate effects on soil organic carbon concentration and particle-size association in northern grasslands. Sci Rep. 2018;8(1), Art. #1336, 9 pages. http://doi.org/10.1038/s41598-018-19785-1

24. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 2005;25(15):1965–1978. https://doi.org/10.1002/joc.1276

25. Forestry South Africa. South African forestry and forest products industry facts [webpage on the Internet]. c2017 [cited 2019 Nov 01]. Available from: https://forestry.co.za/south-african-forestry-and-forest-products-industry-facts-1980-2016-news/

26. Smith C, Pallett R, Kunz R, Gardner R, Du Plessis M. A strategic forestry site classification for the summer rainfall region of southern Africa based on climate, geology and soils. ICFR Bull Ser. 2005;3(05).

27. South African Department of Agriculture, Forestry and Fisheries (DAFF). Report on commercial timber resources and primary roundwood processing in South Africa. Pretoria: DAFF; 2012.

28. Odindi J, Mutanga O, Rouget M, Hlanguza N. Mapping alien and indigenous vegetation in the KwaZulu-Natal Sandstone Sourveld using remotely sensed data. Bothalia. 2016;46(2), Art. #a2103, 9 pages. http://dx.doi.org/10.4102/abc.v46i2.2103

29. Ringrose S, Matheson W, Vanderpost C. Analysis of soil organic carbon and vegetation cover trends along the Botswana Kalahari Transect. J Arid Environ. 1998;38:379–396. https://doi.org/10.1006/jare.1997.0344

30. Zomer RJ, Bossio DA, Sommer R, Verchot LV. Global sequestration potential of increased organic carbon in cropland soils. Sci Rep. 2017;7(1), Art. #15554, 8 pages. https://doi.org/10.1038/s41598-017-15794-8

31. Sommer R, Bossio D. Dynamics and climate change mitigation potential of soil organic carbon sequestration. J Environ Manage. 2014;144:83–87. https://doi.org/10.1016/j.jenvman.2014.05.017

32. Walkley A, Black IA. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci. 1934;37(1):29–38. http://dx.doi.org/10.1097/00010694-193401000-00003

33. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. Ecol Modell. 2006;190(3–4):231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

34. Graham CH, Elith J, Hijmans RJ, Guisan A, Townsend Peterson A, Loiselle BA, et al. The influence of spatial errors in species occurrence data used in distribution models. J Appl Ecol. 2008;45(1):239–247. https://doi.org/10.1111/j.1365-2664.2007.01408.x

35. Phillips SJ, Dudík M, Schapire RE, editors. A maximum entropy approach to species distribution modeling. In: Greiner R, Schuurmans D, editors. Proceedings of the Twenty-First International Conference on Machine Learning; 2004 July 4–8; Banff, Alberta, Canada. Menlo Park, CA: AAAI Press; 2004. http://doi.org/10.1145/1015330.1015412

36. Rahmati O, Pourghasemi HR, Melesse AM. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. Catena. 2016;137:360–372. https://doi.org/10.1016/j.catena.2015.10.010

37. Phillips SJ, Dudík M. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography. 2008;31(2):161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x

38. Ismail R, Mutanga O. A comparison of regression tree ensembles: Predicting *Sirex noctilio* induced water stress in *Pinus patula* forests of KwaZulu-Natal, South Africa. Int J Appl Earth Obs. 2010;12:S45–S51. https://doi.org/10.1016/j.jag.2009.09.004

39. Forkuor G, Hounkpatin OK, Welp G, Thiel M. High resolution mapping of soil properties using remote sensing variables in South-Western Burkina Faso: A comparison of machine learning and multiple linear regression models. PLoS ONE. 2017;12(1), e0170478, pages. https://doi.org/10.1371/journal.pone.0170478

40. Pearson TR, Brown SL, Birdsey RA. Measurement guidelines for the sequestration of forest carbon. Gen Tech Rep NRS-18. Newtown Square, PA: US Department of Agriculture, Forest Service, Northern Research Station; 2007. https://doi.org/10.2737/NRS-GTR-18

41. Ndalowa D. Evaluation of carbon accounting models for plantation forestry in South Africa. Stellenbosch: Stellenbosch University; 2014.

42. Meier IC, Leuschner C. Variation of soil and biomass carbon pools in beech forests across a precipitation gradient. Glob Chang Biol. 2010;16(3):1035–1045. https://doi.org/10.1111/j.1365-2486.2009.02074.x

43. Jobbágy EG, Jackson RB. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol Appl. 2000;10(2):423–436. https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2

44. Aanderud ZT, Richards JH, Svejcar T, James JJ. A shift in seasonal rainfall reduces soil organic carbon storage in a cold desert. Ecosystems. 2010;13(5):673–682. http://doi.org/10.1007/s10021-010-9346-1

45. Zhou G, Guan L, Wei X, Tang X, Liu S, Liu J, et al. Factors influencing leaf litter decomposition: An intersite decomposition experiment across China. Plant Soil. 2008;311(1–2), Art. #61. https://doi.org/10.1007/s11104-008-9658-5

46. O'Brien SL, Jastrow JD, Grimley DA, Gonzalez-Meler MA. Moisture and vegetation controls on decadal-scale accrual of soil organic carbon and total nitrogen in restored grasslands. Glob Chang Biol. 2010;16(9):2573–2588. https://doi.org/10.1111/j.1365-2486.2009.02114.x

47. Young I, Ritz K. Tillage, habitat space and function of soil microbes. Soil Till Res. 2000;53(3–4):201–213. https://doi.org/10.1016/S0167-1987(99)00106-3

48. Ci E, Al-Kaisi MM, Wang L, Ding C, Xie D. Soil organic carbon mineralization as affected by cyclical temperature fluctuations in a karst region of southwestern China. Pedosphere. 2015;25(4):512–523. https://doi.org/10.1016/S1002-0160(15)30032-1

49. Wang G, Zhou Y, Xu X, Ruan H, Wang J. Temperature sensitivity of soil organic carbon mineralization along an elevation gradient in the Wuyi Mountains, China. PLoS ONE. 2013;8(1), e53914, 7 pages. https://doi.org/10.1371/journal.pone.0053914

50. Small EE, Kurc S, editors. The influence of soil moisture on the surface energy balance in semiarid environments. AGU Spring Meeting Abstracts; 2002.

51. Cong R-G, Brady M. The interdependence between rainfall and temperature: Copula analyses. Sci World J. 2012;2012, Art. #405675, 11 pages. http://dx.doi.org/10.1100/2012/405675

52. Young CJ, Liu S, Schumacher JA, Schumacher TE, Kaspar TC, McCarty GW, et al. Evaluation of a model framework to estimate soil and soil organic carbon redistribution by water and tillage using 137Cs in two US Midwest agricultural fields. Geoderma. 2014;232:437–448. https://doi.org/10.1016/j.geoderma.2014.05.019

53. García-Ruiz JM. The effects of land uses on soil erosion in Spain: A review. Catena. 2010;81(1):1–11. https://doi.org/10.1016/j.catena.2010.01.001

54. Liu L, Wang H, Dai W, Lei X, Yang X, Li X. Spatial variability of soil organic carbon in the forestlands of northeast China. J Forestry Res. 2014;25(4):867–876. https://doi.org/10.1007/s11676-014-0533-3

55. Hijmans RJ, Graham CH. The ability of climate envelope models to predict the effect of climate change on species distributions. Glob Chang Biol. 2006;12(12):2272–2281. https://doi.org/10.1111/j.1365-2486.2006.01256.x

56. Lang M, McCarty G, Oesterling R, Yeo I-Y. Topographic metrics for improved mapping of forested wetlands. Wetlands. 2013;33(1):141–155. https://doi.org/10.1007/s13157-012-0359-8

57. Lang MW, McCarty GW. LiDAR intensity for improved detection of inundation below the forest canopy. Wetlands. 2009;29(4):1166–1178.

58. Pei T, Qin C-Z, Zhu A-X, Yang L, Luo M, Li B, et al. Mapping soil organic matter using the topographic wetness index: A comparative study based on different flow-direction algorithms and kriging methods. Ecol Indic. 2010;10(3):610–619. https://doi.org/10.1016/j.ecolind.2009.10.005