Using Multiple Linear Regression Techniques to Quantify Carbon Stocks of Fallow Vegetation in the Tropics

E. M. Attua

Department of Geography and Resource Development, University of Ghana, Legon- Accra, Ghana

Abstract

Fallow ecosystems provide a significant carbon stock that can be quantified for inclusion in the accounts of global carbon budgets. Process and statistical models of productivity, though useful, are often technically rigid as the conditions for their application are not easy to satisfy. Multiple regression techniques have been applied to study some biophysical phenomena but yet to be applied to carbon stock estimation. Using ecological data from 28 sampling locations, the study applied the stepwise multiple regression technique to identify ecological variables that would explain carbon stock of fallow vegetation, aged between 3 and 8 years. The procedure generated three predictive regression models. The full model, could explain nearly 98% of variability of carbon stock ($R^2 = 0.979$), using cation exchange capacity and total nitrogen content of soil and leaf area index as the three predictor variables. Sampling inaccuracies could have contributed to the error component of models and sample size increase has been suggested for reduction of such errors. The advantage of the method is its simplicity. The paper suggests that the derived models be validated before broad application. Also, the cost-effectiveness of the approach should be tested against other approaches.

Introduction

The role of the carbon cycle in contributing to global climate change from greenhouse gas emissions continues to be an issue of considerable concern, especially in the context of global warming (Roberts, 1994; IPCC, 2001). Though there is much agreement on increasing atmospheric carbon dioxide concentration and the concomitant rise in global mean temperature, there is still much debate over the exact relationship between climate variability and changes in atmospheric carbon dioxide concentrations because of uncertainties over the exact figures and the spatial and temporal variations (Foody *et al*., 1996).

The carbon balance of terrestrial ecosystems is uncertain, in part due to discrepancies and errors in measurements, but, perhaps more importantly, due to methodological problems resulting in incomplete accounting (Houghton 2003). Lieth (1975) had long observed this problem and blamed the discrepancies in available datasets on lack of defensible methodologies for measurements of terrestrial primary production. This has contributed to a poor understanding of the role of the terrestrial biosphere, particularly vegetation, in carbon cycling at a range of spatial and temporal scales (Houghton & Skole, 1990; Jarvis & Dewar, 1993; Sampson *et al*., 1993).

Different types of models are now routinely applied to elucidate ecosystem dynamics, especially in the context of regional or global carbon budgets (e.g. Ruimy *et al*., 1996; Goetz & Prince, 1996; Coops *et al*., 2001; Reeves *et al*., 2001; Lobell *et al*., 2002). However, the available ground-based data that most of these models require to generate reliable estimates of primary production at the regional and continental scales are lacking for large parts of the earth's ecosystems (Seaquist *et al*., 2003).

Comparatively, collection of carbon stock data of terrestrial ecosystems has been relatively less vigorous in most parts of the developing world than in the developed. Constrained mainly by economic woes, primary productivity data of most ecosystems has remained scanty or non-existent in most developing countries and, for purposes of global carbon budget accounts, more need to be done to improve data availability. Even with some of the available data, uncertainties exist because productivity measurements of vegetation had frequently been deduced from changes in forest extent over time. The main source of information has been the series of tabulations from FAO production yearbooks and yearbooks of forest products, compiled by the

FAO (Woodwell *et al*., 1984; Grainger, 1993). The data from such sources are unreliable because, as succinctly explained by Hampicke (1979), they are often impregnated with errors.

Direct field-plots assessment of biomass, either through destructive or non-destructive sampling, is adjudged the most accurate and reliable since it is based on direct field sampling (Newbould, 1967). However, the approach is laborious, time-consuming and the associated field costs are prohibitive and impracticable for routine assessments, even at the farm scale. This has largely contributed to the dearth of ground data on primary productivity of many tropical terrestrial ecosystems.

One approach that has, in recent times, gained popularity is the application of multiple linear regression techniques to explain biophysical phenomena. For instance, linear regression models have been developed in this way by using satellite information and field data to study biomass (Häme *et al*., 1997), net primary production (Rasmussen 1998a) and crop yield (Rasmussen, 1998b). To the best of the author's knowledge, the technique is yet to find application in the study of carbon stocks, particularly of fallow vegetation in the tropics.

The study, therefore, implored the techniques of multiple linear regression to identify explanatory variables of carbon stocks in fallow vegetation, using ecological parameters that are known to influence primary productivity. The objective was to empirically formulate simple predictive models for quantifying carbon stocks of fallow vegetation. The study was guided by a null hypothesis that predictor variables have no effect on carbon stock variability. The paper describes the procedures used and reports on the outcomes.

Materials and methods

Study area
The study area is located in the southern forest-savanna transition ecozone in the Eastern Region of Ghana (Fig. 1). The zone constitutes a broad ecological area between the dry coastal savanna in the south and the dry semi-deciduous forest zone in the north (Hall & Swaine, 1976). The area has undergone a dramatic reduction in forest cover and fallow lengths since the 1970s, important reductions in fallow lengths over the same period (Gyasi *et al*., 1994; Amanor, 1994). There is also evidence of soil deterioration and infestation by obnoxious weeds, particularly the virulent weed, *Chromolaena odorata* (Amanor, 1994).

The transitional zone is characterized by 1,260 mm mean annual bimodal rainfall, generally adequate to support crop production. Predominantly, the soils are acidic (*p*H 4.5–6.0), and are suitable both for perennial tree crops and annual food crops (Soil Research Institute, 1971; Dickson & Benneh, 1988). The human population of the area, estimated to be over 1 million, continues to increase. As a result of increasing land usage, the predominantly thick deciduous forest vegetation of the past is fast dwindling and being replaced by an expanding mosaic of forest-savanna vegetation (Gyasi *et al.,* 1994).

Sample plots selection
Field work was conducted from the first week of March to the end of October 2000. Twenty-eight 20×20 m^2 plots of different fallow lengths were selected for the study. These comprised 11 three-year old fallows, eight 4-year old fallows, four 5-year old fallows, and five 8-year old fallows. All plots were located on well-drained soils of gentle topography. Plots selection was non-random and unequally replicated as it depended on willingness of landowners to volunteer fallow lands for the study. The ecological data were collected through field work to include both vegetation and soil parameters that are thought to influence biomass accumulation (Newbould, 1967; Running 1990; Häme *et al.,* 1997; Le Houérou, 1989).

Field data collection

Field data comprised plant and soil measurements. Plant data consisted of plant biomass, stem volume (SV) and leaf area (LA). The soil data were percent soil moisture (SM), percent total nitrogen (TN), available phosphorus (AP) and cation exchange capacity (CEC). Direct field-plots assessment procedure was used to collect the biomass data (Newbould, 1967). Stem volume and leaf area were estimated as described by Anderson & Ingram (1989). Because tree felling was not permissible, biomass of canopy and sub-canopy trees were estimated from allometric formula as recommended by the Tropical Soil Biology and Fertility (TSBF) Programme (Anderson & Ingram, 1989) below; taking into account the annual rainfall regime of the study area, which is below 1500 mm (Dickson & Benneh, 1988; Adu & Asiamah, 1992; Attua, 1996):

$$y = 34.4703 - 8.0671D + 0.6589D^2 \quad \ldots\ldots\ldots\ldots \quad (n = 32; R^2 \text{ adj} = 0.67)$$

where y is the biomass in kilograms, D is the diameter of tree at breast-height in centimeters and n is the sample size. Soil samples were collected per sample plot with an augur to a depth of 20 cm. Five samples were taken per plot and composited.

Laboratory analyses

Laboratory analyses of biomass involved dry weight determination of all collected sub-samples of plant parts. The carbon equivalents of total dry matter per plot were obtained with a conversion factor of 45% (Ajtay *et al.*, 1979) and expressed in Megagram carbon per hectare (MgCha^{-1}). Soil samples were also analyzed for bulk density as recommended by Schlesinger (1984). Laboratory procedures for chemical analyses of soil samples followed those of Black (1965) and Kalra & Maynard (1991).

Data analyses

The Statistical Package for Social Sciences (SPSS 8.0) for Windows programme was used for statistical analyses of the data. Multiple linear regression methods were applied to analyse the data and develop regression models. This is considered the best approach for estimation of a continuous variable (Saukkola & Jaakkola 1983; Oza *et al.*, 1989).

In searching for a statistical explanation of carbon stocks in fallows, a correlation matrix (Pearson's correlation coefficient) was first applied to the datasets (Kinnear & Gray, 1999). This is vital when sorting out suitable predictor variables for regression analysis as well as finding intercorrelated variables to be considered in the analysis (Bryman & Cramer, 1997; Kinnear & Gray, 1999).

Because Pearson's correlation coeffi-cient is very vulnerable to the leverage exerted by extreme values or atypical data points (Kinnear & Gray, 1999), normal probability plots were done to identify all extreme values or outliers in the data. Though logarithmic transformations were also tried, the results generally did not improve the distribution and, therefore, the original data was retained.

The stepwise regression techniques were used for estimation of carbon stocks. This method, according to Shaw & Wheeler (1996), is useful when trying to find the best subset of predictors. The SPSS programme used offered different kinds of procedures within this context: forward selection, backward elimination and standard stepwise regression. With the forward selection method, predictors are added one at a time provided they meet an entry criterion and cannot, subsequently, be removed while, with the backward elimination method, predictors are all present initially and are removed one at a time if they do not meet a retention criterion (Kinnear & Gray, 1999).

The standard stepwise regression procedure was applied. This procedure combined both the forward selection and backward elimination methods, and is considered comparatively superior to either the forward selection or the backward elimination procedures (Kinnear & Gray, 1999). The statistical criteria used for selection or removal of variables from regression are often based on

either the significance (probability) of the F value, or the F value itself (Bryman & Cramer, 1997; Kinnear & Gray, 1999). The commonly accepted minimum probability of F-to-enter and F-to remove, quoted, respectively, as d" 0.05 and e" 0.10, were used. This stopping rule selected variables with a high partial correlation coefficient and a minimum contribution to the mean squares (Afifi & Clark, 1984; Kinnear & Gray, 1999).

The criterion used for testing the significance of the regression algorithms was that a calculated *F*-ratio exceeded the tabulated *F* value, and, at the same time, the individual regression coefficients were found to be significant at the $p < 0.05$ level. Thus, the chance was better than 5% that an additional observation would support the regression model found (Afifi & Clark, 1984; Rasmussen, 1998a).

To indicate the reliability of a regression model, the residuals of the regression equations were analysed. The requirement was that the residuals should be a normal distribution around the regression line, and the degree of scatter around the line should not vary. This meant the residuals should have zero mean and unit variance over all the observed independent terms (Shaw & Wheeler, 1996; Kinnear & Gray, 1999). Also, to test for a possibility of autocorrelation in the regression residuals, and to accept or reject the null (H_o) hypothesis of absence of autocorrelation, the Durbin-Watson test (Kinnear & Gray, 1999) was used at critical bounds of 0.05 significance level.

Results and discussion

The descriptive statistics of all variables explored as explanatory variables of carbon stock of fallow vegetation are shown in Table 1. The distribution pattern of each variable was examined using probability plots and histograms.

The strength of association between variables, using Pearson's correlation is indicated in Table 2. All variables correlated positively with C_s. A consistently high positive correlation of more than 0.800 was found between C_s and variables SV, LA, SM, TN, and CEC. The highest correlation was with CEC ($r = 0.965$) and the lowest was with AP ($r = 0.674$).

TABLE 1

Descriptive statistics of all variables explored for model development.

Variables	Units	Minimum	Maximum	Mean	Std. Deviation
C_S	MgCha^{-1}	11.112	24.327	16.101	3.722
SV	m^{3}ha-1	12.56	45.11	22.03	8.84
LA	m^{2}kg^{-1}	7.11	8.26	7.58	0.32
SM	%	7.14	7.80	7.40	0.19
TN	%	0.08	0.21	0.14	3.00×10^{-2}
AP	mgkg^{-1}	4.14	5.81	5.02	0.49
CEC	Cmol$_c$kg^{-1}	10.56	18.51	14.60	2.34

n= 28; C_S = Carbon stock; SV= Stem volume; LA= Leaf area; SM= Soil moisture;

TN= Total nitrogen; AP= Available phosphorus; CEC= Cation exchange capacity.

TABLE 2

Pearson's correlation coefficients for variables included in analysis (n = 28, *p < 0.05; **p < 0.01, 2-tailed)

	Variable		
C_s	0.000		
SV	0.825**	0.000	
LA	0.864**	0.722**	0.000

SM	0.901**	0.854**	0.797**	0.000			
TN	0.927**	0.774**	0.751**	0.817**	0.000		
AP	0.674**	0.539**	0.419*	0.631**	0.571**	0.000	
CEC	0.965**	0.764**	0.830**	0.871**	0.850**	0.721**	0.000

Table 3 shows a summary of linear regression algorithms generated through the approach. Each regression equation was statistically significant at the *p*-value of 0.05. They are represented, respectively, by equations 1, 2 and 3. Table 4 shows the descriptive statistics of the analysis of variance of the generated regression models.

<div align="center">TABLE 3</div>

<div align="center">*Descriptive statistics of generated regression models*</div>

No.	Model parameters	R^2	R^2adjusted	Standard error of estimates	F-statistic	Sig. F change
1	a CEC + b	0.930	0.927	1003.880	346.725	0.000
2	a CEC + b TN + c	0.972	0.969	654.062	429.582	0.000
3	a CEC + b TN + c LA + d	0.979	0.976	579.890	368.929	0.015

$$C_S = 1532.40\ CEC - 6266.079\ (R^2 = 0.930;\ R^2adj. = 0.927) \qquad (1)$$
$$C_S = 1010.938\ CEC + 47910.577\ TN - 5170.666\ (R^2 = 0.972;\ R^2adj. = 0.969) \quad (2)$$
$$C_S = 843.728\ CEC + 44815.002\ TN + 1787.314\ LA - 15859.052\ (R^2 = 0.979;\ R^2adj. = 0.976) \qquad (3)$$

A high level of explained C_S variance ($R^2 > 0.90$) was obtained with the inclusion of either CEC or TN in the regression. The joint use of CEC and TN as predictor variables improved the regression further by 0.042 (Equation 2). The inclusion of LA as the third predictor variable improved the accuracy of prediction of the dependent variable by a further 0.007 (Equation 3). The H_0 hypothesis of absence of autocorrelation in the regression residuals was also accepted since the Durbin-Watson test used at critical bounds of 0.05 significance level was found to be significant.

The results (Table 3) indicate that the full model (Equation 3) explained approximately 98% variability in C_s ($R^2 = 0.979$). The reduced models (Equations 1 and 2) comparatively had reduced R^2 values of 0.930 and 0.972, respectively. Analysis of variance (ANOVA) results (Table 4) indicated that R^2 of all three models were significant at *p*-value of 0.05. The null hypothesis of no linear regression of C_s on the predictor variables (CEC, TN and LA) was, therefore, rejected and the alternate hypothesis accepted in all cases. Thus, a significant portion of variability in C_s is explained by CEC, TN and LA. Compara-tively, the error terms (error sum of squares) of the models probably attributable to sampling, reduced greatly in the full model because it had the minimum contribution to the mean squares (Table 4). These errors could probably have been reduced by increasing the sample size and/or including other statistically eligible predictor variables.

<div align="center">TABLE 4</div>

<div align="center">*Analysis of variance (ANOVA) of derived models (n = 28, p < 0.05, 2-tailed)*</div>

Model No.	Source of variation	Sum of squares	df	Mean square	F-statistic	Significance F

1	Regression	309. 279	1	309.279	346.725	0.000
	Error	23.179	26	0.892		
	Total	332.458	27			
2	Regression	323.046	2	161.523	429.582	0.000
	Error	9.412	25	0.376		
	Total	332.458	27			
3	Regression	325.396	3	108.465	368.929	0.000
	Error	7.062	24	0.294		
	Total	332.458	27			

Conclusion

The results of the study support the use of multiple linear regression for estimation of carbon stocks in fallow vegetation. The advantage of the method is its simplicity. However, because model parameters are derived from field plot measurements, they already embody the response of a particular landscape to C_s change and, therefore, the use of the models must be done with care, or an unrealistic prediction could result. At this stage, derived models are site-specific and have limited application. They, thus, require validation for reliable broad application. Also, the cost-effectiveness of the approach need to be tested against other approaches. A first step would be to check that models correctly predict carbon stocks on new fallow plots under similar soil and climatic conditions. Also, the current study could probably gain from increasing sampling size and exploring other ecological data. Nevertheless, this study could serve as an initial step for improving carbon estimation in tropical fallow vegetation.

Acknowledgement

References

Adu S. V. and **Asiamah R. D.** (1992). *Soils of the Ayensu/Densu Basin: Central, Eastern and Greater Accra Regions*, Soil Research Institute Memoir No. 9. Advent Press, Kwadaso-Kumasi, Ghana.

Afifi A. A. and **Clark V.** (1984). *Computer-aided Multivariate Analysis*, Life Learning Publications, Belmont, California.

Ajtay G. L., Ketner P. and **Duvigneaud P.** (1979). Terrestrial Primary Production and Phytomass, *Global Carbon Cyc.* **SCOPE 13**. John Wiley and Sons, New York.

Amanor K. S. (1994). *The New Frontier: Farmer's Response to Land Degradation. A West African Study.* UNRISD, Geneva.

Anderson and **Ingram** (ed.) (1989). *Tropical Soil Biology and Fertility: A Handbook of Methods*. C.A.B. International, Wallingford, UK.

Attua E. M. (1996). *Ecological Studies of Agricultural Land Use and Cover Change in Ghana: A case study of the mid-Densu basin.* (MPhil. Thesis.). Botany Department, University of Ghana, Legon.

Black C. A. (ed.) (1965). *Methods of Soil Analysis.* Agronomy No. 9, Part 2. American Society of Agronomy, Madison, Wisconsin.

Bryman A. and **Cramer D.** (1997). *Quantitative Data Analysis with SPSS for Windows. A Guide for Social Scientists.* Routledge, London.

Coops N. C., Waring R. H., Brown S. R. and **Running S. W.** (2001). Comparisons of predictions of net primary production and seasonal patterns in water use derived with two forest growth models in Southwestern Oregon. *Ecol. Model.* **142**: 61–81.

Dickson K. B. and **Benneh G.** (1988). *A New Geography of Ghana.* Longman, London.

Foody G. M., Palubinskas G., Lucas R. M., Curran P. J. and **Honzak M.** (1996). Identifying Terrestrial Carbon Sinks: Classification of Successional Stages in Regenerating Tropical Forest from Landsat TM Data. *Rem. Sens. Envir.* **55**: 205-216.

Goetz S. J. and **Prince S. D.** (1996). Remote Sensing of Net Primary Production in Boreal Forest Stands. *Agric. For. Met.* **78**(3): 149–179.

Grainger A. (1993). Rates of Deforestation in the Humid Tropics: Estimates and Measurements, *Geography J.* **159**: 33–44.

Gyasi E. A., Agyepong G. T., Ardayfio-Schandorf E., Enu-Kwesi L., Nabila J. S. and **Owusu-Bennoah E.** (1994). *Environmental Endangerment in the Forest-Savanna Zone of Southern Ghana.* The United Nations University (UNU), Tokyo, Japan.

Hall J. B. and **Swaine M. D.** (1976). Classification and ecology of closed canopy forest in Ghana. *J. Ecol.* **64**: 913–51.

Häme T., Salli A., Andersson K. and **Lohi A.** (1997). A New Methodology for Estimation of Biomass of Conifer-dominated Boreal Forest using NOAA AVHRR data, *Int. J. Rem. Sens.* **18**: 3211–3243.

Hampicke U. (1979). Net Transfer of Carbon between Land Biota and the Atmosphere, Induced by Man, *Glob. Carbon Cyc.* **SCOPE 13**. John Wiley and Sons, New York.

Houghton J. T. and **Skole D. L.** (1990). *Carbon*, *The Earth as Transformed by Human Action.* Cambridge University Press, Cambridge.

Houghton R. A. (2003). Why are estimates of terrestrial carbon balance so different? *Glob. Change Biol.* **9**: 500–509.

IPCC (2001). *Climate Change (2001). The Scientific Basis.* Intergovernmental Panel on Climate Change, Geneva.

Jarvis P.G. and **Dewar R. C.** (1993). *Forests in the Global Carbon Balance: From Stand to Region, Scaling Physiological Processes: Leaf to Globe*, Academic, Press, San Diego.

Kalra Y. P. and **Maynard D. G.** (1991). *Methods Manual for Forest Soils and Plant Analysis.* Information Report Nor-x-319, Canada.

Kinnear P. R. and **Gray C. D.** (1999). *SPSS for Windows Made Simple*, third edition. Psychology Press Ltd, Hove, United Kingdom.

Le Houérou H. N. (1989). *The Grazing Land Ecosystems of the African Sahel,* Spring-Verlag, Berlin.

Lieth H. (1975). *Modelling the Primary Productivity of the World, Primary Productivity of the Biosphere*. Spring-Verlag, Berlin.

Lobell D. B., Hicke J. A., Asner G. P., Field C. B., Tucker C. J. and **Los S. O.** (2002). Satellite estimates of productivity and light use efficiency in United States agriculture 1982–1998. *Glob. Chem. Biol.* **8**: 722–735.

Newbould P. J. (1967). *Methods for Estimating the Primary Production of Forests.* IBP Handbook No. 2. International Biological Programme, Oxford.

Oza M. P., Srivastava V. K., Pariswad B. S. and **Setty K. R. V.** (1989). Relationship between Landsat MSS data and forest tree parameters. *Int. J. Rem. Sens.* **10**: 1813–1819.

Rasmussen M. S. (1998a). Developing Simple, Operational, Consistent NDVI-Vegetation Models by Applying Environmental and Climatic Information, Part I. Assessment of net primary production. *Int. J. Rem. Sens.* **19**: 97–117.

Rasmussen M. S. (1998b). Developing simple, operational, consistent NDVI-vegetation models by applying environmental and climatic information, Part II. Crop Yield Assessment. *Int. J. Rem. Sens.* **19**: 119–139.

Reeves M. C., Winslow J. C. and **Running S. W.** (2001). Mapping weekly rangeland vegetation productivity using MODIS algorithms. *J. Range Mgmt* **54**: 90–105.

Roberts N. (1994). *The Global Environmental Future, The Changing Global Environment* (N. Roberts, ed.) Blackwell, Oxford.

Ruimy A., Dedieu G. and **Saugier B.** (1996). TURC – Terrestrial Uptake and Release of Carbon by vegetation: a diagnostic model of continental gross primary productivity and net primary productivity, *Glob. Biogeochem. Cycl.* **10:** 269–285.

Running S. W. (1990). *Estimating Terrestrial Primary Productivity by combining Remote Sensing and Ecosystem Simulation, Remote Sensing of Biosphere Functioning.* Springer-Verlag, New York Inc, Berlin Hiedelberg.

Sampson R. N., Apps M. and **Brown S.** (1993). *Workshop Summary Statement: Terrestrial Biospheric Carbon Fluxes – quantification of sinks and sources of CO2. Wat. Air Soil Pollut.* **70**: 3-15.

Saukkola P. and **Jaakkola S.** (1983). *Numerical image interpretation in forest inventory and mensuration.* Research Report 151, Technical Research Centre of Finland, Tutkimuksia.

Schlesinger W. H. (1984). *Soil Organic Matter: A Source of Atmospheric CO2, The Role of Terrestrail Vegetation in the Global Carbon Cycle: Measurement by Remote Sensing.* **SCOPE 23**, Wiley, New York.

Seaquist J. W., Olsson L. and **Ardo J.** (2003). A remote sensing-based primary production model for grassland biomes. *Ecological Model.* **169**: 131–155.

Shaw G. and **Wheeler D.** (1996). *Statistical Techniques in Geographical Analysis.* David Fulton Publishers, London.

Soil Research Institute (1971). *Soil Suitability Map of Import Substitution Crops in Ghana.* Soil Research Institute, Kumasi and Survey Department, Accra.

Woodwell G. M., Hobbie J. E., Houghton R. A., Mellilo J. M., Moore B., Park A. B. Peterson B. J. and **Shaver G. R.** (1984). *Measurement of Changes in the Vegetation of the Earth by Satellite Imagery, The Role of Terrestrial Vegetation in the Global Carbon Cycle. Measurement by Remote Sensing.* **SCOPE 23**. Wiley, Chichester.