Tanzania Journal of Science 46(3): 673-683, 2020
ISSN 0856-1761, e-ISSN 2507-7961

Sensitivity Analysis and Uncertainty Parameter Quantification in a Regression Model: The Case of Deforestation in Tanzania

Thadei Sagamiko[1*] , Nyimvua Shaban[2*] and Isambi S Mbalawata[3]
[1]*Department of Physics, Mathematics and Informatics, Dar es Salaam University College of Education,University of Dar es Salaam, P. O. Box 2329, Tanzania.*
[2]*Department of Mathematics, University of Dar es Salaam, P. O. Box 35062 Dar es Salaam, Tanzania.*
[3]*African Institute for Mathematical Sciences – Kigali, Rwanda.*
**Corresponding author, email: tsagamiko@gmail.com*
Co-authors emails: shabanmbare@gmail.com; mbalawata@yahoo.com
Received 13 May 2020, Revised 2 Sep 2020, Accepted 3 Sep 2020, Published Oct 2020
https://dx.doi.org/10.4314/tjs.v46i3.9

Abstract
In this paper a multiple regression model for the economic factors and policy that influence the rate of deforestation in Tanzania is formulated. Sensitivity analysis for parameters of explanatory variables using one-at-a time and direct methods is carried out and the model is fitted by classical least square (LSQ) and Markov Chain Monte Carlo (MCMC) methods. Uncertainty quantification of parameters by adaptive Markov Chain Monte Carlo methods is performed. The coefficient of determination indicates that 87% of deforestation rate is explained by explanatory variables captured in the model. Household poverty rate is found to be the most sensitive factor to deforestation, while purchasing power is the least sensitive in both methods. Model validation indicates a good agreement between the collected data and the predicted data by the model and Markoc Chain Monte Carlo method yielded a good sample mix. Thus, the study recommends that since economic activities tend to increase the rate of deforestation, then policy and decision-making processes should link the country's desire for economic growth and environmental management.

Keywords: deforestation, economic factors, Markov Chain Monte Carlo methods, regression model, sensitivity.

Introduction

Tropical deforestation has become an issue of global environmental and developmental concerns because of the benefits obtained from forests. Turner et al. (2006) suggested that global forest areas would decline by 477 million hectares between the year 1999 and 2030, largely in Asia and Africa. Kindermann et al. (2006) studied the potential effects of different financial mechanisms that motivate deforestation, and predicted the deforestation-trend under different carbon-prices from socio-economic land use model. They found that about 200 million hectares or around 5% of global forest areas will be lost between the year 2006 and 2025.

Also, poor income contributes to deforestation since people clear forests and use the areas for growing cash crops, charcoal production, firewood and plant residue collection for cooking and heating because they cannot afford alternative energy sources such as kerosene, electricity and liquefied petroleum gas. A study by Reetz (2012) analyzed the influence of poor income on tropical deforestation and the results showed that poor technology, economy, education, gender and geographical conditions significantly influenced the rate of deforestation.

In Tanzania, deforestation still exists due to multiple factors such as economic activities, expansion of land for agricultural activities, over-extraction of wood for logging (both legally and illegally), domestic fuel or charcoal making and other developmental projects like road construction and urbanization. The Ministry of Natural Resources and Tourism (1999) pointed out that market competitions and price trends for the period of the years 1998 to 2020 are expected to be inflexible based on economic factors such as declining resource base, poor domestic production and low purchasing power as a result of slow economic growth. Thus, the rate at which forests are declining prompts researchers to put more efforts in studying the processes of deforestation in order to understand the dynamics of forests disappearance and propose remedial measures.

Much work has been done on the rate of deforestation in Tanzania such as those by Luoga et al. (2000), Mitinje et al. (2007), Giliba et al. (2011) and Doggart et al. (2020) which focused mostly on social-economic factors. The factors considered are livelihood activities, period of residence close to forest, distance from homestead to the forest, farm land size, household size, awareness of the management of the reserve and its boundaries, education and the growing population. However, none of the works assessed the relative contribution of each factor to the rate of deforestation. In this paper, a study was carried out on the influence of other factors on deforestation, namely; per capita income over time, purchasing power over time, inflation rate over time, poverty rate over time, and electricity price over time on deforestation rate in Tanzania. A multiple linear regression model is developed to establish the regression parameters of the factors. Sensitivity analysis and uncertainty quantification of parameters of these factors are studied to establish the relative contribution of each factor to the rate of deforestation in Tanzania. Moreover, the statistical approach of Markov Chain Monte Carlo (MCMC) method which was less treated in previous studies is applied.

Materials and Methods

The current study used secondary data from the National Bureau of Statistics (NBS) and National Forest Management Assessment (NAFORMA). Twenty one (21) observations were collected for each variable in 21 years. The data included deforestation rate in Tanzania (ha/year) as response variable and five predictor variables, namely per capita income over time measured in US$ at market price, per capita purchasing power over time (constant at 2011 US$), inflation rate over time (% change in consumer price index), poverty rate over time (measured as household consumption % per capita) and electricity consumption (as per electricity price) per population over time (KWh per capita).

The response variable is assumed to be directly related to a linear combination of explanatory variables. Multiple linear regression model was used to fit the data. The model aimed at establishing the relationship between the rate of deforestation as a response variable and the explanatory variables. Adaptive Markov Chain Monte Carlo methods were used to understand the behaviour of the parameters . Moreover, MATLAB software was used for data analysis.

The model

A regression model that establishes the relationship between the rate of deforestation over time in Tanzania as the response variable denoted by Y and five explanatory variables (denoted as $X_i, i = 1, 2, ..., 5$), namely per capita income over time X_1, purchasing power over time X_2, inflation rate over time X_3, poverty rate over time X_4 and electricity price over time X_5 is as presented in equation (1);

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4 + \theta_5 X_5 + \varepsilon, \quad (1)$$

where; ε is a noise factor. Data for the factors influencing deforestation in Tanzania are shown in Table 1 and are measured in their units as follows: *Y* is the rate of deforestation in Tanzania (ha/year) and is the reduction in total area of the forest cover per year and has a negative sign, X_1 is per capita income over time measured in US$ at market price, X_2 is per capita purchasing power over time (constant at 2011 US$), X_3 is inflation rate over time (% change in consumer price index), X_4 is the poverty rate over time (measured as household consumption % per capita) and X_5 is the electricity consumption (as per electricity price) per population over time (KWh per capita).

Table 1: Data for economic factors influencing deforestation in Tanzania from 1994 to 2014

Year	X_1	X_2	X_3	X_4	X_5	Y
1994	170.260	1363.304	34.100	81.240	49.259	–400400
1995	171.310	1370.840	29.400	83.120	57.184	–400320
1996	200.480	1394.497	28.000	82.810	60.715	–400300
1997	250.440	1407.133	18.100	87.260	55.211	–400280
1998	277.210	1423.740	15.800	80.280	60.079	–400211
1999	301.200	1456.200	9.900	81.150	55.651	–400210
2000	308.410	1489.650	9.900	78.270	58.250	–400206
2001	306.240	1538.050	9.150	74.970	61.979	–400101
2002	320.210	1604.370	10.550	71.930	67.298	–400090
2003	325.250	1667.920	9.550	69.730	67.162	–400073
2004	338.050	1747.900	9.150	66.940	78.634	–400062
2005	456.160	1836.150	9.350	66.800	78.355	–400060
2006	495.210	1864.770	10.450	64.020	64.206	–400043
2007	503.170	1961.280	9.050	61.360	80.126	–400035
2008	607.230	2006.640	8.250	63.960	84.351	–400024
2009	685.340	2048.890	12.150	66.290	70.553	–400001
2010	788.220	2111.200	7.200	68.370	93.869	–372816
2011	790.380	2206.910	7.650	65.220	84.651	–372701
2012	897.230	2247.860	10.000	65.580	94.602	–372670
2013	930.380	2335.960	3.900	64.770	89.478	–372231
2014	945.140	2421.210	3.130	64.570	89.111	–372000

Source: National Bureau of Statistics 2013, NAFORMA 2015

The regression coefficients θ_i's for $i = 0,...,5$ represent a change of forest cover (for $i = 0$) and the change in the rate of deforestation for one unit change in the respective explanatory variable, while holding other explanatory variables in the model constant (for $i = 1,...,5$). Substituting the values of unstandardized coefficients θ_i from Table 2 in the regression equation, the following model is obtained:

$$Y = -527262.514 + 33.625X_1 + 10.231X_2 + 87.329X_3 + 999.71X_4 + 369.245X_5 \quad (2)$$

Thus, the constant term (–527262.514) is described as predicted “mean rate of deforestation” and is obtained when all explanatory variables are set to zero, whereas the negative sign implies “a change of forest cover”, that is deforestation.

Table 2: Regression coefficients for economic factors influencing deforestation

	Coefficients	Standard Error	t Stat	P-value
Intercept	–527262.514	76737.79	–6.87096	5.31E–06
X1	33.6246466	35.37673	0.950474	0.35693
X2	10.2309177	34.56151	0.296021	0.77127
X3	87.3299223	211.9411	0.412048	0.68613
X4	999.710319	502.988	1.987543	0.06543
X5	369.245100	203.0001	1.818941	0.08893

Sensitivity analysis

This study used two methods for determining local sensitivity so as to increase confidence in the ranking of key inputs. One-at-a-time sensitivity measure and direct method were used because they assume linear relationship between variables and univariate model outputs.

One-at-a-time sensitivity measure

One-at-a-time sensitivity analysis is an approach which is done by changing one-factor-at-a-time independently while keeping other input variables constant to see the effects that the factor produces on the output. Sensitivity measures are determined by monitoring changes in the output from the linear regression model. Sensitivity ranking is obtained by increasing each input parameter by 20% while leaving all others constant and quantifying the change in the model output (Hamby 1994). Since the regression model (2) is linear, unstandardized regression coefficients were directly used to measure sensitivity by varying the input parameter values by 20% of their central value. When varying one input variable at a time, all other variables are fixed to their central values and all effects are computed with reference to the same central point in space.

Consider the model equation (2), $\Delta Y = Y_c - Y$ where Y_c is the value of Y obtained after 20% increase of the input parameter $\theta_i (i = 1,...,5)$. When all inputs are kept constant ($X_i = 1$), that is, $X_1 = X_2 = X_3 = X_4 = X_5 = 1$, the value of output becomes $Y =$ –525762.374.

20% increase in $\theta_1 = 40.35$, $Y_c = -525755.649$, $\nabla Y = 6.725$

20% increase in $\theta_2 = 12.2772$, $Y_c = -525760.3278$, $\nabla Y = 2.0462$

20% increase in $\theta_3 = 104.7948$, $Y_c = -525744.9085$, $\nabla Y = 17.4658$

20% increase in $\theta_4 = 1199.652$, $Y_c = -525562.432$, $\nabla Y = 199.942$, and

20% increase in $\theta_5 = 443.094$ $Y_c = -525688.525$, $\nabla Y = 73.849$

From Table 3, X_4 (poverty rate over time) is more sensitive than other factors. This implies that the increase in poverty in households contributes more in the rate of deforestation. The factor X_2 (per capita purchasing power over time) is least sensitive to the rate of deforestation.

Table 3: Rank of sensitivity indices using one-at-a-time method

Factor	X_1	X_2	X_3	X_4	X_5
Output sensitivity	6.725	2.0462	17.4658	199.942	73.849
Ranking	4	5	3	1	2

Direct method

The direct method of sensitivity analysis is based on partial derivatives of the aggregated model. From the explicit algebraic equation describing the regression model (2), sensitivity coefficient could be used to measure how much change produced in the output is caused by changes in the input quantity or parameter (Hegazy and Mohamad 2013). The variance in the rate of deforestation Y is used to measure the uncertainty in model predictions, while the variance $X_i (i = 1,...,5)$ weighted by the first-order partial derivative of Y with respect to $X_i (i = 1,...,5)$ provides a measure of the model sensitivity (Hamby 1994). The sensitivity coefficient (C_i) is obtained from the partial derivatives of the model function with respect to the input quantities. Therefore, the sensitivity coefficient C_i of the economic factors X_i is given by

$$C_i = \frac{\partial Y}{\partial X_i}$$

Thus, the values of the C_i's in the regression model (2) are;

$$C_1 = \frac{\partial Y}{\partial X_1} = 33.625, \quad C_2 = \frac{\partial Y}{\partial X_2} = 10.231,$$

$$C_3 = \frac{\partial Y}{\partial X_3} = 87.32, \quad C_4 = \frac{\partial Y}{\partial X_4} = 999.71,$$

and $C_5 = \frac{\partial Y}{\partial X_5} = 369.245$

The sensitivity coefficients describe how the value of Y varies with small changes in the values of the variables X_i (Hegazy and Mohamad 2013). From Table 4, X_4 (poverty rate over time) is more sensitive than other factors, which implies that increase in poverty to households contributes more in the rate of deforestation. The factor X_2 (per capita purchasing power over time) is least sensitive to the rate of deforestation, so it causes little variation on deforestation.

Table 4: Rank of sensitivity indices using direct method

Factor	X_1	X_2	X_3	X_4	X_5
Sensitivity coefficient	33.625	10.231	87.329	999.71	369.245
Rank	4	5	3	1	2

Model validation

Model validation is the process of determining the degree to which a model and its data are accurate representation of the real life situation surrounding the problem (Petty 2010). Model validation involves comparison of results from the model equation (2) to the observed data from Table 1. Figure 1 shows the behaviour of the factors influencing deforestation from actual data. Predicted and actual data plotted in Figure 2 show their good agreement with a correlation between observed data of rate of deforestation and the predicted data from the model of 0.93.

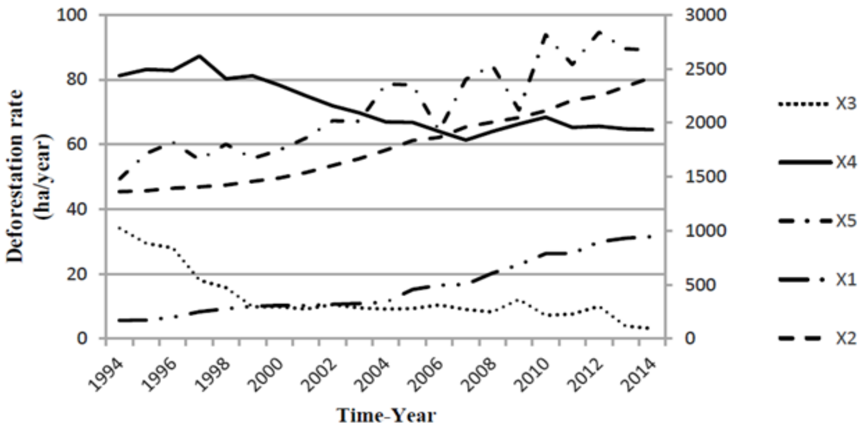

Figure 1: Effect of predictor variables on deforestation over time.

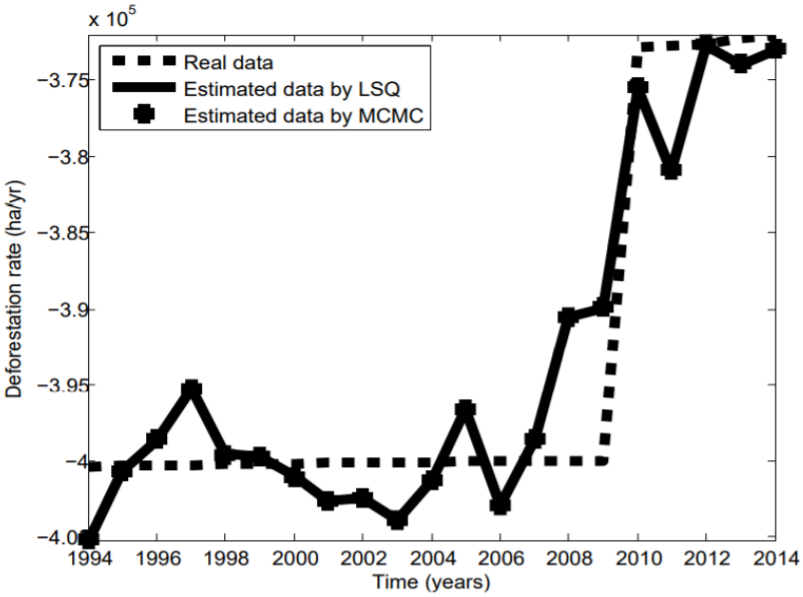

Figure 2: Model fitting by LSQ and MCMC – describe LSQ before using it.

Uncertainty quantification of parameters

To understand the behaviour of parameters, the adaptive MCMC methods are used (Gilks et al. 1996, Haario et al. 2001, Haario et al. 2006, Andrieu and Thomas 2008). In this paper, the MCMC method is used for parameter indentifiability. Generally, the MCMC method is used to compute the multidimensional integral where parameter samples are drawn from the posterior distribution and the integral approximated by the expectation as the sample average.

The MCMC samples are generated from proposal distributions. The proposal distributions for MCMC algorithms should always lead to a well mixing of samples and in a suitable acceptance rate. In most cases, the proposal distribution used is the Gaussian distribution. Hence, it is needed to find a suitable covariance matrix for better mixing and good acceptance rate. It is for this reason that the adaptive MCMC algorithm is used where, given an initial/starting covariance matrix, the algorithm tends to automatically adapt the proposal distribution during the run (Haario et al. 2001, Haario et al. 2006, Andrieu and Thomas 2008). The adaptive MCMC algorithm developed by Haario et al. (2001) is as presented below.

Adaptive MCMC algorithm

i. Initialization: start with initial values of parameters and covariance matrix,
ii. At each step, propose a new parameter value from a Gaussian distribution,
iii. Accept/reject the proposed value according to the MCMC accepting probability,
iv. Adapt the proposal covariance matrix per preference,
v. Iterate the algorithm above until you get enough samples.

In this paper, 100000 parameter samples are generated. Parameter values from classical least squares as the MCMC initial parameters' values are used, while the initial covariance is set to be a 6 by 6 identity matrix. The identifiability of parameters is studied by examining the trace plots, scatter plots, marginal distributions, and sample autocorrelation functions.

Results and Discussion

The MATLAB outputs for fitting multiple regression model for response variable Y with five explanatory variables (without interaction) X_1, X_2, X_3, X_4 and X_5 are presented in Tables 2, 5 and 6, with correlation analysis data presented in Table 7. In addition, an adaptive Markov Chain Monte Carlo MATLAB output in studying the behaviour of the model parameters are presented in Figures 3, 4, 5 and 6 together with Tables 8 and 9. The results in Table 2 show that the combinations of explanatory variables significantly contribute to the rate of deforestation in Tanzania. Results from Table 5 indicate strong relationship among the variables involved in the model with Multiple R of 0.93. Moreover, the coefficient of determination ($R^2 = 0.87$) shows that 87% of deforestation rate is caused by explanatory variables captured in the model. That means only 13% is explained by other factors not considered in the study.

Table 5: Summary output

Regression Statistics	
Multiple R	0.933339201
R square	0.871122064
Adjusted R square	0.828162753
Standard error	5006.118236
Observations	21

The results presented in Table 6 for analysis of variance [F (5, 15) = 20.2778] suggest that the developed multiple regression model is useful for predicting the rate of deforestation in Tanzania and the results are statistically significant with P-value = 0.00000344, which is less than 5% of recommended condition. Both methods for sensitivity measures (Table 3 and Table 4) revealed that the poverty rate over time is the most sensitive factor that contributes to deforestation, while purchasing power takes the least sensitive case.

Table 6: Analysis of variance (ANOVA)

	Df	SS	MS	F	P-Value
Regression	5	2.54E+09	5.08E+08	20.27784	3.44E-06
Residual	15	3.76E+08	25061220		
Total	20	2.92E+09			

Figure 3 represents MCMC sample parameters for 100,000 iterations. The vertical axis represents samples' values and horizontal axis represents number of iterations. From Figure 3, all samples 'mix well', indicating that the parameters are indentifiable.

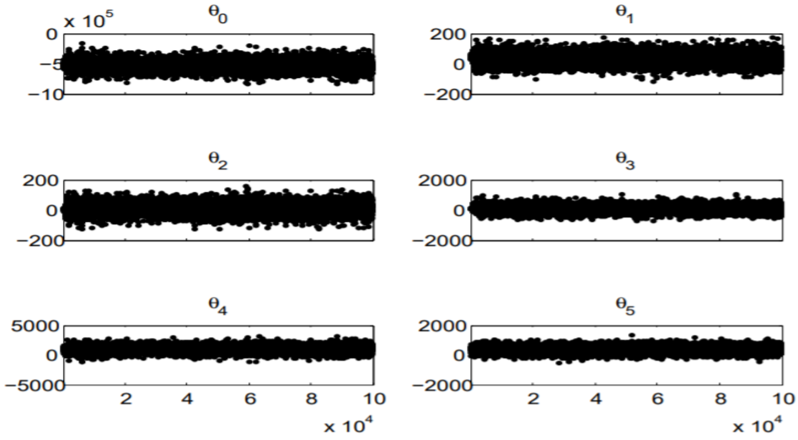

Figure 3: Markov Chain Monte Carlo plots.

Figure 4 represents scatter plots of model parameters which shows the relationship among them, where θ_0, θ_1, θ_2, θ_3, θ_4 and θ_5 represent estimated coefficients of mean deforestation rate, per capita income over time, per capita purchasing power over time, inflation rate, poverty rate and electricity consumption, respectively. For representation purpose, four (4) pairs of parameters are considered as shown in Figure 4. From Figure 4, it is observed that there is positive correlation between estimated mean rate of deforestation and estimated coefficient of per capita income, negative correlation between estimated coefficient of per capita purchasing power and that of per capita income, the result which differs from their corresponding variables as shown in Table 7 of correlation analysis. This is due to the suppressor effect (Frank and Miller 1992). Estimated coefficient of inflation rate is negatively correlated with that of poverty rate, while there is no correlation between estimated coefficient of poverty rate (household consumption% per capita) and that of electricity consumption.

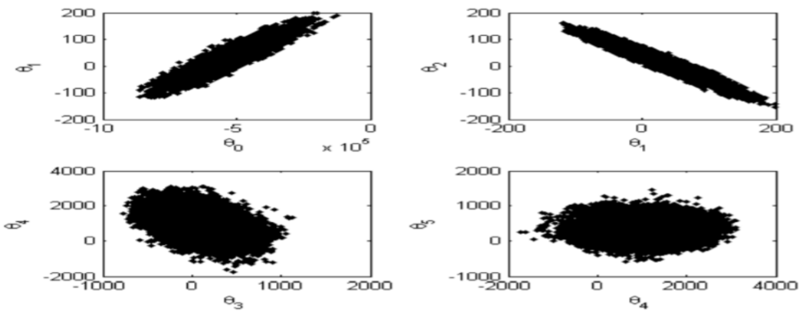

Figure 4: Correlation between model parameters.

Table 7: Correlation analysis for deforestation data set

	Per capita income	Per capita purchasing power	Inflation rate	Poverty rate	Electricity consumption	Deforestation rate
Per capita income	1					
Per capita purchasing power	0.977	1				
Inflation rate	–0.664	–0.694	1			
Poverty rate	–0.752	–0.859	0.717	1		
Electricity Consumption	0.889	0.919	–0.674	–0.814	1	
Deforestation rate	0.861	0.791	–0.443	–0.444	0.769	1

Marginal distributions are presented in Figure 5 in which parameters θ_1, θ_2 and θ_3 are close to normal distributions. This observation agrees with their corresponding coefficients of skewness which are close to zero as shown in Table 8 and coefficients of kurtosis which are supposed to be close to 3 as shown in Table 9.

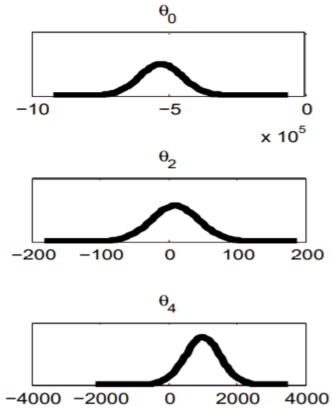

Figure 5: Parameter marginal distributions.

Table 8: Sample skewness

Sample	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5
Skewness	0.0248	–0.0109	–0.0056	0.0414	–0.0466	–0.0248

Table 9: Sample kurtosis

Sample	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5
Kurtosis	3.2832	3.2474	3.2789	3.2246	3.2536	3.1656

All autocorrelation functions follow standard form as they start from 1 and decay to zero at some point and start stabilizing to zero throughout lags as shown in Figure 6.

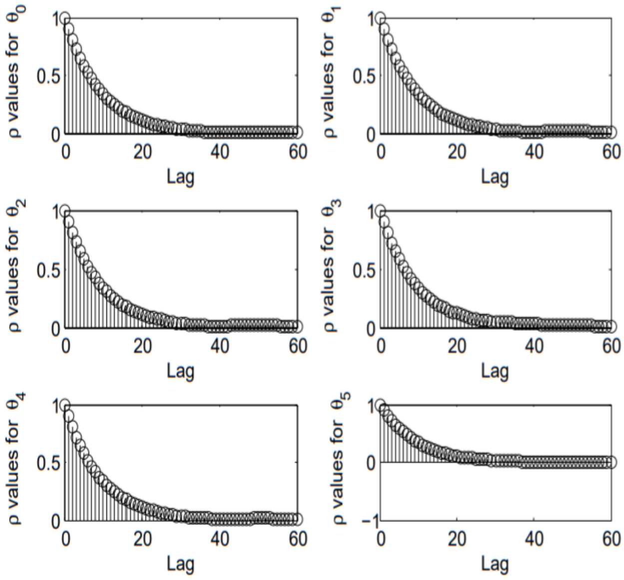

Figure 6: Autocorrelation functions.

Conclusion

A linear regression model on the economic factors and policy that influence the rate of deforestation in Tanzania is formulated. The model was numerically analysed by least square and Markov Chain Monte Carlo methods. A sensitivity analysis for the parameters was carried out to assess their influence on deforestation. As a means of understanding the behaviour of parameters, the adaptive Markov Chain Monte Carlo method was used to study the identifiability of these parameters by examining the trace plots, scatter plots, marginal distributions, and sample autocorrelation functions. In fitting the model to the data, the MCMC and classical least square methods were used to estimate the model parameters. The estimated deforestation values were computed and compared to the real deforestation data. It was found that the two data sets superimpose each other showing that the methods can be used for indentifiability of the model parameters.

Acknowledgements

Authors acknowledge the National Bureau of Statistics for granting acces to the data.

Declaration: Authors have no conflict of interest.

References

Andrieu C and Thomas J 2008 A tutorial on adaptive MCMC. *Statis. Comput.* 18(4): 343-373.

Doggart N, Morgan-Brown T, Lyimo E, Mbilinyi B, Meshack CK, Sallu SM and Spacklen DV 2020 Agriculture is the main driver of deforestation in Tanzania. *Environ. Res. Lett.* 15(3): p. 034028.

Frank RF and Miller NB 1992 A Primer for soft Modelling, 1st ed, The University of Akron, Akron, Ohio.

Giliba RA, Boon EK, Kayombo CJ, Chinjere LI and Musamba EB 2011 The influence of socio-economic factors on deforestation: a case study of the Bereku

Forest Reserve in Tanzania. *J. Biodiv.* 2(1): 31-39.

Gilks WR, Richardson S and Spiegelhalter DJ 1996 Markov Chain Monte Carlo in Practice, Chapman and Hall/CRC.

Haario H, Saksman E and Tamminen J 2001 An adaptive metropolis algorithm. *Bernoulli* 7: 223-242.

Haario H, Laine M, Mira A and Saksman E 2006 DRAM: efficient adaptive MCMC. *Statistics and Computing* 16(4): 339-354.

Hamby DM 1994 A review of techniques for parameter sensitivity analysis. *Environ. Monit. Assess.* 32(2): 135-154.

Hegazy R and Mohamad MI 2013 The role of the sensitivity coefficient on the authenticity of the Unc. *Int. J. Adv. Res. Electric. Electron. Instrument. Engin.* 2(8): 4107-4110.

Kindermann GE, Obersteiner M, Rametsteiner E and McCallum 2006 Predicting the deforestation-trend under different carbon-prices. *Carbon Balance Manage.* 1(1): 1-17.

Luoga EJ, Witkowisk ETF and Balkwill K 2000 Subsistence use of wood products and shifting cultivation within miombo woodland of eastern Tanzania, with some notes on commercial uses. *S. Afr. J. Bot.* 66: 72-85.

Mitinje E, Kessy JF and Mombo F 2007 Socio-economic factors influencing deforestation on the Uluguru Mountains, Morogoro, Tanzania. *Discov. Innov.* 19: 139-148.

Ministry of Natural Resources and Tourism 1999 Tanzania Forestry Action Plan 1994/95– 2007/08, Update.

National Bureau of Statistics 2013 Household Budget Survey 2011-2012, Sixth Round, National Bureau of Statistics, Tanzania.

National Forestry Resources Monitoring and Assessment (NAFORMA) 2015 Draft Report on NAFORMA, Tanzania Forest Services.

Petty MD 2010 Verification, validation, and accreditation. In Sokolowski JA and Banks CM (Eds) *Modeling and simulation fundamentals: Theoretical underpinnings and practical domains*, ch.10, pp. 325-372.

Reetz SWH 2012 *Effects of land use, market integration and poverty on tropical deforestation: evidence from forest margins areas in central Sulawesi, Indonesia*. PhD thesis, Georg-August-University of Göttingen.

Turner JA, Buongiorno J and Zhu S 2006 An economic model of international wood supply, forest stock and forest area change. *Scand. J. Forest Res.* 21: 73-86.