

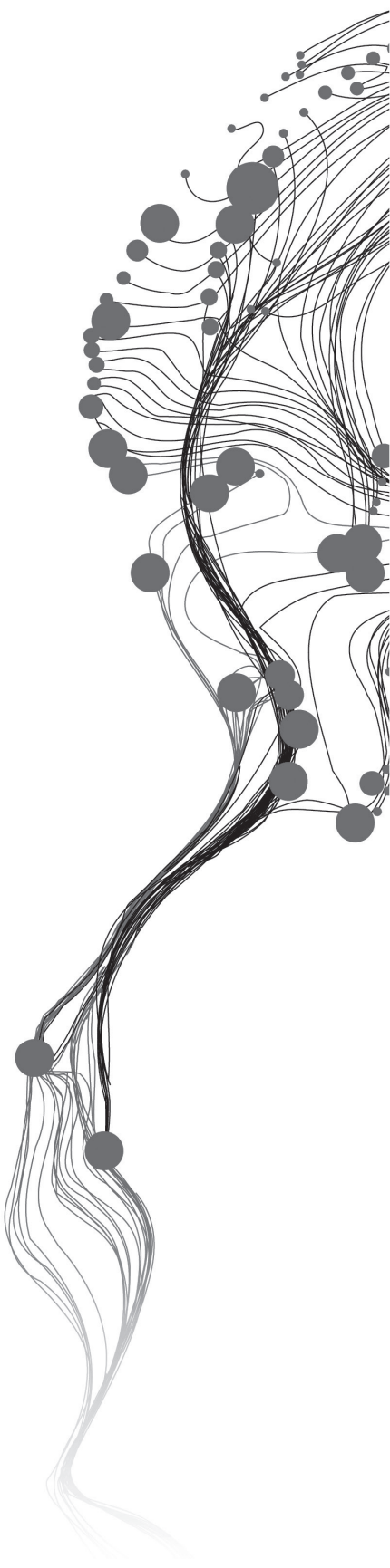
Comparison of a model based geo-statistical approach with a design-based approach to estimate carbon stock in Gorkha district, Nepal

TSIKAI SOLOMON CHINEMBIRI

February, 2012

SUPERVISORS:

**Ir. M.C. Bronsveld
Dr. D.G. Rossiter**



Comparison of a model-based geo-statistical approach with a design-based approach to estimate carbon stock in Gorkha district, Nepal

TSIKAI SOLOMON CHINEMBIRI

Enschede, the Netherlands, February 2012

Thesis submitted to the faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-Information Science and Earth Observation.

SUPERVISORS:

Ir. M.C. Bronsveld
Dr. D.G. Rossiter

THESIS ASSESSMENT BOARD:

Chair: Dr. Y. A. Hussin
External Examiner: Dr. N. Kerle (Faculty of Geo-Information Science and Earth Observation, ITC)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

Abstract

Two sampling protocols using a geo-statistical and a design-based framework were juxtaposed to evaluate their accuracy in the estimation of carbon stock in Ludikhola watershed, Nepal. The geo-statistical approach exploits the spatial dependencies in the sampled variable and may therefore be attractive over the design-based approach as it reduces the substantial costs of survey and effort required in the later. Scales of spatial variation for C stock which resulted in a grid resolution of 10 000m² were determined in a pre-field work exercise using a historical dataset sampled from the same study area. A Box-Cox transformation approach resulted in a square root transformation for the C stock. A criterion-based Akaike Information Criterion (AIC) was used for the selection of the best linear model of feature space for use in kriging with external drift (KED). Modelling of C stock using KED gave a range of spatial dependence of 762m, slightly higher than previously determined by the reconnaissance variogram. The Successful modelling of the regionalised variable is therefore found to depend on sampling density and sample size. Carbon stock was found to be moderately spatially dependent, an important milestone towards the adoption of promising C stock estimation methods. Out of the five tested covariates, distance, elevation and aspect were statistically significant, with the best model of feature space accounting for 87.7% variability of C stock. An analysis of variance (ANOVA) established significance differences in mean C stocks amongst the community forests ($F_{6, 103}=5.21$, $P=0.00017$). KED using the best model of feature space was found to be more accurate (9.89 ± 0.17) Sqrt MgCha⁻¹ than a pure model based approach of ordinary kriging and the design-based approach (9.91 ± 0.8) Sqrt MgCha⁻¹. Comparison of the confidence bounds of the two estimators showed that their confidence intervals will overlap 99.7% of the time, with both confidence intervals falling within the 95% confidence bounds of each other. The confidence interval of the model-based approach gave a narrower margin of error, implying less uncertainty around the estimated mean C stock and more accurate than the design-based approach. There is insufficient evidence to suggest that the model-based approach is significantly better than the design-based approach. However, the balance of probabilities on the best estimator to use favours the model-based approach as it can essentially answer more questions, especially on the spatial front than the design-based approach.

Key words: Nepal, Geo-statistics, Kriging with External Drift, Design-based, AIC, ANOVA, Community forest, Uncertainty.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to the Dutch government for providing me with a scholarship through the Netherlands fellowship Programme (NFP). I also wish to thank my employer, the Ministry of Lands and Rural Resettlement together with the Public Service Commission of Zimbabwe (PSC) for supporting and facilitating my application to study abroad. The successful completion of this work could not have been possible without the tremendous support from academics, organisations and friends named henceforth.

I am very grateful to my first supervisor Ir.M.C Bronsveld for providing the much needed impetus and buoyancy in order to keep the ship afloat. Your critical and constructive suggestions during the entire course of this study will go a long way in strengthening my academic rigour. I am equally grateful to my second supervisor Dr. D.G. Rossiter whose vast knowledge and experience in geo-statistics is second to none. Your informed suggestions and thorough review of my coursework during and after the proposal period gave shape and credence to this research.

A special thanks to the carbon and climate change theme leader, Dr. Y.A. Hussin and my course Director Dr. M.J.C. Weir for their dedicated engagement and professionalism during the entire period of this study.

I wish to acknowledge International Centre for Integrated Mountain Development (ICIMOD), Asia Network for Sustainable Agriculture and Bio-resources (ANSAB), Federation of Community Forest Users Nepal (FECOFUN), Forest Resources Assessment (FRA) and Arbonaut of Finland projects in Nepal for providing the necessary data, field itinerary and manpower during the course of my field work in Ludikhola watershed of Gorkha district, Nepal. My sincere gratitude goes to fellow NRM MSc students for providing the necessary suggestions and assistance before and after field-work.

Lastly, I wish to express my sincere and deepest gratitude to my family for being with me morally and spiritually in spite of the geographical barriers between us.

TABLE OF CONTENTS

Abstract.....	i
Acknowledgements.....	ii
List of Figures.....	iii
List of tables.....	iv
1. INTRODUCTION	11
1.1. Background.....	11
1.2. Above ground biomass estimation	13
1.3. Application of geo-statistics for estimating forest carbon stock.....	15
1.4. Kriging	16
1.5. Statement of the Problem	18
1.6. General Objective.....	19
1.7. Research questions	19
1.8. Assumptions.....	20
1.9. Hypotheses	20
1.10. Thesis outline	20
2. Methods and Materials	23
2.1. Study area.....	23
2.2. Materials	24
2.3. Above ground tree biomass (AGTB) Measurement.....	26
2.4. Biomass Calculation and carbon stock derivation.....	26
2.5. The model-based (geo-statistical) approach.....	27
2.6. Sampling design.....	27
2.7. Rationale for using kriging with external drift.....	28
2.8. Approach for the selection of covariates.....	29
2.9. Covariates Selected	29
2.10. Clipping of the Interpolation grid and extraction of covariates data.....	30
2.11. Principal Components analysis and data reduction	30
2.12. Exploratory data analysis.....	31
2.13. The Box-Cox transformation approach	31
2.14. Selection of the best linear model.....	32
2.15. Cross-validation.....	33
2.16. The design-based approach	33
2.17. Significance test for the assessment of confidence interval overlap.....	36
2.18. Description of the strata (community forests).....	37
3. Results	39
3.1. Non-spatial exploratory data analysis.....	39
3.2. Spatial exploratory data analysis.....	41
3.3. Response variable (C stock) transformation	42
3.4. Reconnaissance variogram.....	43
3.5. Variance explained by the principal components.....	45
3.6. Spatial structure of carbon stock	46
3.7. Feature space model selection criterion	47

3.8. Ordinary kriging.....	48
3.9. Block Ordinary kriging.....	49
3.10. Kriging with external drift (KED).....	50
3.11. Cross-validation Statistics and model diagnostics	51
3.12. The design-based approach.....	54
3.13. Carbon stock estimates: the design-based approach.....	56
3.14. Jackknifing.....	57
3.15. Comparison of mean confidence intervals	59
4. Discussion.....	61
4.1. The geo-statistical approach	61
4.2. Limitations of the study	70
5. Conclusions and Recommendations.....	71
5.1. Conclusions.....	71
5.2. Recommendations.....	72
References	74
6. Appendices.....	79

LIST OF FIGURES

Figure 2-1: Study area, Ludikhola watershed.	23
Figure 2-2: Methodological flowchart for the design and model-based approach for C stock estimation	25
Figure 2-3: Carbon Stock observation locations for the model-based (geo-statistical) approach.	28
Figure 2-4: Location of sampling plots for the design-based approach.	36
Figure 3-1: Distribution of carbon stock, a) . Before transformation and b) . After Square root transformation.....	40
Figure 3-2: Relationship between carbon stock and candidate predictor variables.	40
Figure 3-3: a) . The behavior of carbon stock density and distribution with respect to geography and b) . quartiles of the empirical distribution of measured carbon stock values.	41
Figure 3-4: Directional variograms for Carbon stock.....	42
Figure 3-5: Reconnaissance variogram for the sampling distance of the model-based approach.	43
Figure 3-6: Log-likelihood plot for the Box-Cox transformation of the Carbon stock response variable	43
Figure 3-7: Scree plot of the results of a Principal Component Analysis (PCA) for the C stock dataset.	45
Figure 3-8: Stage wise comparison of the effects of covariates on the spatial correlation structure of carbon stock.	46
Figure 3-9: Kriging results illustrating, a) . Square root kriging predictions and b) . Kriging variances of carbon stock (Mg C ha^{-1}).....	49
Figure 3-10: Ordinary block-kriging illustrating a) . Square root kriging predictions and b) . Kriging variances of C stock (Mg C ha^{-1}).....	49
Figure 3-11: Kriging with external drift (KED) results with log (distance) as a covariable, illustrating, a) . Square root kriging predictions and b) . Kriging variances of Carbon stock (Mg C ha^{-1})	50
Figure 3-12: a) . Diagnostic plots of the residuals of the kriging prediction variants and b) . Model diagnostics of the best linear model (Elevation + log (distance) for carbon stock.....	51
Figure 3-13: Kriging with external drift (KED) results with (elevation and log (distance)) as covariables, illustrating, a) . Square root kriging predictions and b) .Kriging variances of Carbon stock (Mg C ha^{-1})	52
Figure 3-14: C stock distribution for the design-based approach.....	54
Figure 3-15: Box and whisker plot of the mean carbon stock density within each of the seven sampled community forests.	56
Figure 3-16: Simultaneous 95% confidence limits, Turkey method with carbon stock density as the response variable.....	56
Figure 3-17: Diagnostic plots for the ANOVA model.....	58
Figure 3-18: Jackknife diagnostics for the design-based approach.	59
Figure 3-19: Overlapping of confidence internals	60
Figure 4-1: Successional forests and low biomass plots.	70

LIST OF TABLES

Table 2-1: Data sets used.....	24
Table 2-2: Lists of instruments used	24
Table 2-3: List of software used.	24
Table 2-4: Allometric equations for <i>Shorea robusta</i> and other species.	27
Table 2-5: Proportions and required sample sizes for the sampled community forests. .	35
Table 2-6: Summary of the characteristics of the sampled community forests.....	38
Table 3-1: Descriptive statistics for the target and potential predictors. MEAN- mean, STDV- standard deviation, MED- median, MIN- minimum, MAX- maximum, SKEW- skewness, KURT- kurtosis.....	39
Table 3-2: Correlation matrix between the predictor variables and the response, carbon stock Sqrt (Mg C ha ⁻¹).	44
Table 3-3: Proportion of variances explained by the principal components.....	44
Table 3-4: Significance of the eigenvectors of each predictor variable with respect to the response, carbon stock.....	45
Table 3-5: Parameters for the ordinary variogram and residuals for the distance and elevation covariates	46
Table 3-6: Summary of the selection criterion for the best candidate linear model of carbon stock prediction using kriging with an external drift.	47
Table 3-7: summary of the coefficients of the best linear model of feature space.....	47
Table 3-8: Efficiencies of the different kriging variants in comparison to the design-based approach.	48
Table 3-9: Summary of cross-validation statistics.	51
Table 3-10: summary of total C stock estimates for KED	53
Table 3-11: Homogeneity of variance test for the community forests strata.	55
Table 3-12: Summary of mean and total C stock for the community forests.....	57
Table 3-13: Comparison of confidence intervals.....	60

1. INTRODUCTION

1.1. Background

Forests ecosystems embody an important potential for carbon storage over other terrestrial ecosystems, though proportionally smaller in extent compared to soils and oceans. This study juxtaposes two sampling designs deriving from different theoretical bases on their accuracy for estimating forest carbon stock using a geo-statistical and design-based framework. The research focuses on humid tropical forest containing the most substantial biomass carbon forest stocks (Malhi, 2010).

The difference between a model-based and design-based approach is that, in the former, the assessed values of a variable are assumed to be a realization of a random process whereas in the later, the variable is assumed fixed and the origin of the randomness is in the design process (Brus & de Gruijter, 1997). An important difference that is highlighted in this definition is in the manner the sample data used by each one of these designs are weighted to obtain estimates or predictions.

There is compelling evidence that the average temperature of the earth is increasing and it is this potentially adverse phenomenon which has stimulated research into the dynamics of carbon sequestration by forest ecosystems (Wysowski, 2010). The effects of this rising global temperature are manifested in ecological disruptions that has led to the disappearance of certain forest types and species to become extinct (Keeling & Phillips, 2007). Furthermore, climate change could result in a decline of agricultural production, expansion of deserts and a rise in sea level (Falkowski et al., 2000).

Carbon dioxide (CO₂) is believed to play a major role in global warming. Therefore, solutions to the problem of climate change are aimed at reducing the levels of carbon dioxide in the atmosphere (UNFCCC, 1998). The Reduction of Emissions from Deforestation and forest Degradation (REDD) is part of the clean development mechanism (CDM) under the United Nations Framework Convention on Climate Change (UNFCCC). This mechanism establishes bilateral investment corporation between the developing and developed countries in reducing carbon emissions. The REDD policy is part of an international framework to halt deforestation in the developing world (UNFCCC, 1998).

The growth in population and expansion of human activities such as settlements, agriculture and construction are threatening forests globally. This trend is particularly predominant in the developing countries, and it is this part of the tropical belt of these countries that has the best potential to sequester carbon. The Kyoto Protocol subsequently placed heavier burden on developed economies under the principle of “common but differentiated responsibilities” (Fletcher, 2011). This strategy is aimed at cutting carbon dioxide emissions of developed countries by investing in REDD projects in the developing countries. Informed decision towards the goal of curbing the devastating effects of climate change requires elimination of uncertainty in the estimates of the rate of deforestation (Friedlingstein et al., 2006). Hence, quantification of carbon stocks is important in the determination of the extent of carbon sequestration by forest.

Even though much emphasis has been laid on the importance of understanding the global carbon budget and climate change, much controversy still revolve around the sustainability of the REDD policy and its perceived benefits to the rural poor families in the developing world (Kaipainen et al., 2004). It is a common trend in the greater part of the developing world that forest resources are integral to the survival and livelihoods of rural communities. The implementation of REDD without due consideration of how exactly thousands of forest dependent communities will benefit sustainably from the REDD initiatives, will likely face resistance (Binkley et al., 2002). Without securing the rights of indigenous people and traditional users, REDD is a risk undertaking (Killeen et al., 2010).

The pace at which the preparedness for REDD is currently being worked on is not consistent with the complex processes and decisions that a new climate forest system would involve (Dhakal & Raut, 2010). As a result, a lot remains to be seen on the successes of the market-oriented REDD system as a climate policy. A market-oriented system is deemed dubious due to its basis in offsets that will likely create divisions between indigenous people and hence, could bring about great resistance (Jaenicke, Englhart, & Siegert, 2011).

Nepal commands 40% of its land mass covered by forest ecosystems and it is this substantial potential that can lend it an important target for the REDD project (Dhital, 2009). As a developing country, the risk of deforestation and forest degradation, largely due to the dependence of the local community on forest resources remains high (Panta et al., 2008). The 1970's has seen widespread deforestation and in response to this, forest resources have been put under the custody of community forest groups, which to date, own more than 25% of the total forest (Dhital, 2009). As a consequence of the realization of the importance of forest resources

on global carbon sequestration and livelihoods of forest community groups, Nepal joined the REDD project through the UNFCCC in 2008 (Dahal & Banskota, 2009). It was subsequently granted forest carbon partnership facility for implementation of the REDD project the same year.

A community forest is a unit of land under the custodianship of a common group of community people specifically set aside for the conservation and management of forest ecosystems (ANSAB, 2010). Due to the extensive and heterogeneity that characterise the tropical forests, accurate and reliable carbon stock estimation can be greatly increased through the establishment of management units. Stratification or blocking of the forests into homogenous units using geographical or management regimes usually result in improved accuracy of carbon estimation (ANSAB, 2010). Hence, much of the variation that could come from the heterogeneity of the forests is eliminated since estimates will be based on the strata (community forests) and this is important for carbon accounting under the REDD initiatives.

1.2. Above ground biomass estimation

The main carbon pools in the tropical forest ecosystems are the living biomass of trees, understory vegetation, dead mass of litter, woody debris and soil organic matter (Brown, 2002). The above ground biomass carbon storage pool typically forms the largest pool and it more than doubles the amount of carbon in the atmosphere and it is the most directly impacted by degradation and deforestation (Fox et al., 2010). Hence, estimation of above ground forest biomass carbon is a critical step in the global carbon cycle.

The most direct way to quantify the carbon stored in aboveground living forest biomass is through harvesting of all the trees in a known area, weigh the biomass and then apply a conversion factor to carbon stock (Leemans et al., 1996). This method could be accurate for a particular area but it's not environmentally sustainable and moreover, time consuming and expensive. The greater part of the tropical forest typically contain over 300 species and research has shown that species-specific allometric relationships are not necessary for the generation of reliable estimates of forest carbon stocks (Ravindranath & Ostwald, 2007). Instead, Brown (2002) demonstrated that generalized allometric equations, stratified by broad forest types are found to be highly effective for the tropics. This is because diameter at breast height alone explains more than 95% of the variation in above ground tropical forest carbon stock.

Significant headway has been made towards the quantification and estimation of forest carbon in the various tropical forest of the world as a consequence of the signing and ratification of the

Kyoto protocol by various member countries (Lehtonen et al., 2004). Researchers and forest practitioners including Houghton (2005) have estimated the amount of carbon sequestered in forest using allometric equations, all with a goal of providing unbiased and accurate estimates of carbon stock in forest ecosystems. Sales et al., (2007) compared the performance of a geo-statistical method with a simple average for mapping forest biomass as a step towards the estimation of CO₂ emissions due to land use and land cover changes in the Brazilian Amazon. The research demonstrated the superiority of geo-statistics in improving estimates of carbon emission in the Amazon forest, one of the world's biggest CO₂ reservoir and sinks.

Above ground biomass (ABG) spatial distribution using kriging interpolation is demonstrated to be a reliable method by Du et al., (2010). It is shown that the carbon storage of moso bamboo in China is 68.40Tg C, accounting for 1.65% of the total forest C storage. Phachomphon, Dlamini and Chaplot (2010) demonstrated that tree height explains 85% of carbon stock with a standard error of 15.1kg for the Japanese red pine and 68% variability with 21.0kg standard error for the Japanese cedar. On the other hand, Brown (2002) has also demonstrated that DBH alone explains more than 95% of the variation in above ground tropical forest and hence, provides effective means of estimating C stocks in the tropical forests.

In addition, biomass and carbon stocks in the major tropical forest of India have been estimated using deterministic and geo-statistical models, giving an estimated average of 84t/ha of carbon stocks (Shijo et al., 2010). Rossi et al., (2009) applied sampling strategies for estimating soil organic carbon using conventional statistical methods and geo-statistics for five forest types of southern Tanzania (coastal dry forest, miombo woodlands, teak plantation, pine plantation and cashew plantation).

Variability of carbon stock and pools on different spatial scales as a result of interactions between climate, parent material, topography and site management creates a challenge for efficient sampling (Leemans, et al., 1996).

Sampling methods are largely based on design-based (classical) statistical techniques, central to which is randomized sampling that has its domains in probability theory (Keller, Palace, & Hurtt, 2001). An alternative and promising technique is that of a model-based (geo-statistical) analysis that is independent of randomization. For example, Sales et al, (2007), demonstrated a significant reduction in uncertainty (RMSE) of biomass estimates in Rondonia, Brazil by using kriging with an external drift compared with the use of a simple average. However, given the spatial variability

of biomass due to changing soil catena and topography in tropical forest, a model-based approach might be more appropriate over other techniques in improving carbon stock estimates in these landscapes (Bryan et al., 2010). This is because the derivation of site specific estimates and predictions of carbon stock in inaccessible areas using the model-based framework is possible which is useful for planning and management. Furthermore, the classical approach falls short of capturing the variability of the variable that occasionally responds to the behaviour of the surface due to changing geography in both space and time.

Various interpolation techniques capitalize on the spatial correlation between observations to predict resource values at unvisited locations (Clark & Clark, 2000). It is therefore the intention of this study to investigate the efficiency and effectiveness of a model-based (geo-statistical) technique contrasted with a design-based (classical) technique, for estimating forest carbon stock.

1.3. Application of geo-statistics for estimating forest carbon stock

Like many other environmental variables, the distribution of carbon stock can be regarded as a continuous spatial process which is described by a model of spatial covariance e.g., the variogram model as follows:

$$\gamma(h) = \frac{1}{2} \{Z(x) - Z(x+h)\}^2 \quad \dots\dots\dots (1-1)$$

Where:

$Z(x)$ and $Z(x+h)$ are random variables separated by h

x is the sampled location

h is the distance between sampled locations

The variogram gives a description of the spatial dependence as a function of the separation vector h . The variogram can be estimated for any distance or direction class h by the equation below (Matheron, 1971):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{Z(x_i) - Z(x_i+h)\}^2 \quad \dots\dots\dots (1-2)$$

Where:

$N(h)$ is the number of pairs used to compute the experimental variogram at distance h

$Z(x_i)$ is the measured C stock at location x

$Z(x_i+h)$ is the C stock at location (x_i+h) which is within distance h of location x .

The variogram, which is a form of computing the variance of a population under study, taking into account the spatial position of the samples, is applied as a descriptor of the spatial structure of the C stock.

Geo-statistics exploits the presence of spatial autocorrelation and joint dependences in space and time that occur in most natural resource variables (Oliver & Webster, 2008). Thus, geo-statistics is used to predict values of a sampled variable over the entire area of interest, especially for vast forest areas. The need for extensive and expensive field surveys, as in the case of traditional methods like satellite imagery and forest inventories is therefore eliminated (Longford, 2008b). Amongst the common models used for fitting variograms is the spherical model that is shown in equation 1-3 and 1-4:

$$\gamma(h) = C_0 + C_1 \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right), \quad h \leq a \quad \dots\dots\dots (1-3)$$

$$\gamma(h) = C_0 + C_1, \quad h > a \quad \dots\dots\dots (1-4)$$

Where;

C_0 is the nugget variance

C_1 is the structural variance

a is the range of influence of the target variable

h is the separation distance between successive pairs.

1.4. Kriging

Kriging is a method of interpolation that makes use of the spatial autocorrelation structure of the variable, in this case, distribution of carbon stock over the area of study (de Gruijter et al., 2006). The distribution of carbon stock can be considered as a regionalized variable defined throughout a region D, so that Kriging can be used to explore the way carbon stock varies in space (Diggle & Ribeiro, 2007). Under the second-order stationarity hypotheses, the mathematical expectation of the first moment of $Z(x)$ and its variance is assumed to be constant over the field of study (Kangas, 2006). However, this assumption does not usually apply in heterogeneous forests, where there are mixed tree species and under different management regimes.

Amongst the various techniques used for the mapping of environmental resources are the inverse distance interpolation and thiessen polygons. However, the Best Linear Unbiased Prediction (BLUP) model of spatial data is the one given by the universal model of spatial variation

(Hengl, 2007). Thus, the model can be decomposed into a deterministic and stochastic component as shown in equation 1-5:

$$Z(s) = m(s) + \varepsilon'(s) + \varepsilon'' \quad \dots\dots\dots (1-5)$$

Where:

$m(s)$ is the deterministic component

$\varepsilon'(s)$ is the stochastic component

ε'' is the pure noise due to measurement error

It is therefore the stochastic component that is supplied as predictors and explains the variability of the deterministic (mean) component with respect to location of the variable in addition to other sources of variation. The incorporation of such explanatory information (drift) is therefore used as an aide for the prediction and distribution of the regionalized variable at different locations across the sampled field.

1.4.1. Ordinary kriging

Ordinary kriging is a variant of kriging that assumes stationarity of the modelled regionalised variable. It therefore predicts the value of a variable using the spatial dependences between the sampled values of the target variable. It may be described by the acronym BLUE- Best Linear Unbiased Estimator (Oliver & Webster, 2008). It is linear in the sense that the estimate is build from a weighted linear sum of the observed values and unbiased since it aims at making the average difference between the estimated and the observed values equate to zero.

The unknown value $Z(x)$ at any location is typically decomposed into a mean (drift) μ and a residual component $\varepsilon(x)$ (equation 1-5):

$$Z(x) = \mu + \varepsilon(x) \quad \dots\dots\dots (1-6)$$

The predictor is therefore formed as a weighted sum of the sampled values (equation 1-6):

$$\hat{Z}(x_0) = \sum_{i=1}^N \lambda_i Z(x_i) \quad \dots\dots\dots (1-7)$$

Where:

$\hat{Z}(x_0)$ is the value to be predicted at location x_0

N is the size of the sampled observations

λ_i are the kriging weights

1.4.2. Kriging with an external drift (KED)

KED represents a variant of kriging for non-stationary geo-statistics. In a spatial field, spatial variation can be categorised into large-scale and small-scale variation. The large spatial variation in the target variable is provided by the KED trend whereas the small-scale variation is derived from the residuals of the trend. A linear relationship between the primary variable and the auxiliary variable is assumed and therefore, the secondary variable should smoothly vary in space and it is the variability of the secondary variable that is assumed to explain the distribution of the target variable (Hengl, 2007). The algorithm limits stationarity within each search neighbourhood, thereby offering more local detail than when ordinary kriging is used (Deutsch & Journel, 1998). The KED estimator is given by:

$$Z_{KED}^*(\mu) = \sum_{\alpha=1}^{n(\mu)} \lambda^{KED}(\mu) Z(\mu_{\alpha}) \quad \dots\dots\dots (1-8)$$

Where:

$Z_{KED}^*(\mu)$ is the KED estimate at location μ

$\lambda^{KED}(\mu)$ are the KED weights relating to n samples at location μ

$Z(\mu_{\alpha})$ are the sample values within the search neighbourhood

1.5. Statement of the Problem

Accurate estimation of carbon stocks in forest is central to the understanding of the global carbon budget and climate change, as forests form one of the most important and biggest carbon sinks in the earth's surface (Fang, Xue, & Tang, 2007).

To date, a lot of advances have been made towards the estimation and quantification of above ground carbon (AGB) in tropical forests using techniques ranging from forest inventories to remote sensing (Gibbs, Brown, & Niles, 2007).

Conventional remote sensing and inventories as methods of estimating AGB present major limitations in terms of both resource and technical commitments (Simpson, Botkin, & Nisbet, 1993). Remote sensing could go a long way in alleviating the problem of limitations in coverage of extensive forest lands, but still suffer heavily from the risk of error propagation that becomes integral to the data processing chain, from data acquisition to data analysis (Kiyono et al., 2010). On the other hand, these limitations could be addressed by novel technologies like light detection

and ranging (LIDAR) laser scanning but are prohibitively expensive and consequently, cannot be operated on a larger scale in a cost-effective manner (Wulder et al., 2007).

However, to date, to the best of our knowledge, little has been done in comparing the accuracy of a design-based approach compared with a model based approach using geo- statistics on their suitability in estimating forest carbon in the sub-tropical forests of Nepal. Typically, forest resources cover extensive tracts of land and therefore pose great challenges in adopting technologies that cover the spatial variability of these resources in a representative, cost effective and sustainable way. Inaccessibility of forest in steeping landscapes, typically, Nepal, can pose challenges in producing accurate estimates of carbon stock (Gautam, 2006). It is this complexity in the topography of most tropical forest regimes that leads to the estimation of the aforementioned methods falling short of capturing the necessary variability and heterogeneity that characterize the distribution of above ground forest biomass. It is physically impossible to sample everywhere due to the costs of sampling survey and inaccessibility of some locations. Some areas are even inaccessible to high resolution remote sensing satellites or in the event of their application, cloud cover and forests canopy still pose a great challenge to acquire error free data (Fuchs et al., 2009).

In light of this background, the inevitable limitation regarding the aforementioned technologies opens avenues for geo-statistics and hence, the application of a geo-statistical approach to estimate carbon stock may provide a solution to this shortfall.

1.6. General Objective

The overall objective of this study aims at improving the estimates of carbon stock in community managed forest ecosystems of Ludikhola watershed, Gorkha district, Nepal.

1.6.1. Specific Objectives

1. To select an approach that gives the most accurate estimates of C stock between a model-based and design-based approach.

1.7. Research questions

This research endeavoured to answer the following research questions

- What spacing distance/ interval is required to model the spatial correlation structure of C stock?
- How effective is the model-based design at generating Carbon stock estimates?

- What is the gain in efficiency for Carbon stock estimation between a model-based and design-based approach?
- Are there any covariables that can be used in kriging with an external drift (KED) that improve the spatial correlation structure of carbon stock? What proportion of the apparent spatial correlation structure (ordinary variogram) is removed by covariables?
- What strata should be used to give accurate estimates of C stock for a design-based approach?
- What are the accuracies of the different methods in estimating C stock?

1.8. Assumptions

It is assumed that the samples for the model and design-based approaches are representative of the entire population (community forests) from which the samples were taken.

1.9. Hypotheses

H₀: There is no statistically significant relationship between carbon stock and the covariates of elevation, aspect, distance from the nearest human settlements, slope or normalised difference vegetation index (NDVI)

H₁: There is a statistically significant relationship between carbon stock and the covariates of elevation, aspect, distance from the nearest human settlements, slope or NDVI.

H₀: There is no significant difference amongst the mean carbon stock of the community forests ($\mu_1 = \mu_2 = \dots = \mu_7$)

H₁: At least one pair of the sampled community forests is different from the others.

H₀: $\sigma^2_1 = \sigma^2_2 = \dots = \sigma^2_7 = (\sigma^2)$

H₁: At least one variance of the community forests is different from the others.

H₀: There is no significant difference between the mean C stock estimates of the model-based and design-based approaches.

H₁: The model based approach is significantly better than the design-based approach

1.10. Thesis outline

Chapter 1 gives a general background and introduction, literature review, statement of the problem, research objectives, research questions, hypothesis and assumptions. Chapter 2 gives an outline of the study area and a detailed description of the methods and materials employed in the course of the study. Chapter 3 presents the results and findings of the study. Chapter 4 discusses

the results of the research. Chapter 5 gives the conclusions and recommendations, followed by a list of references and appendices.

2. METHODS AND MATERIALS

2.1. Study area

2.1.1. Climate and Geographical location

Nepal is located in southern Asia and lies between 26°N and 30°N latitude and 80°E and 88°E longitude with a total area of 147 181km² (Dahal & Banskota, 2009). Ludikhola watershed is found in the Gorkha district of the western development region of Nepal. The watershed has a characteristic hill physiographic region with altitude ranging from 318m to 1714m (ANSAB, 2010). The area lies between 27°55'02.85" latitude and 27°59'43.58" longitude with an average annual precipitation of 1972-2000mm/year and an average temperature of 23.1°. The watershed commands a total area of 5 750ha with 31 community forests covering a total area of 1 888ha with an estimated population of 23 197 (ANSAB, 2010).

The watershed falls under the sub-tropical ecological zone with *Schima wallichii*, *Shorea robusta*, pine and *Castanopsis indica* as predominant species. The winter season extends from November to March with temperatures varying from 19°C to 13°C and summer lasts through April and June with temperatures ranging from 21° to 28°C. The landscape is generally hilly inhabited by diverse social groups such as the Magar, gurung, tamang, dalit, Brahmin and chetin people. It is within these social communities that the community forests are also found.

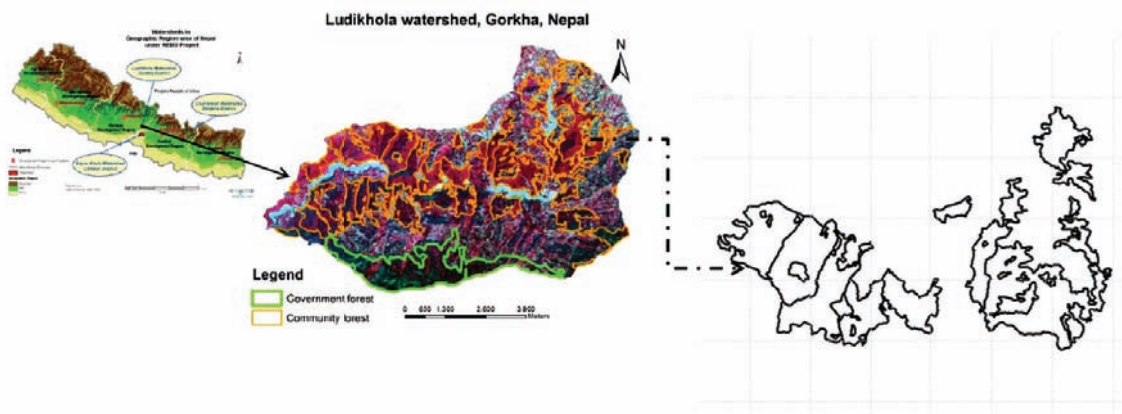


Figure 2-1: Study area, Ludikhola watershed.

As shown in Figure 2-1, seven strata (community forests) were sampled out of a possible total of 31 community forests under different management regimes of community forests user groups. The subsequent data analysis of the present study for both sampling designs (model and design-based approaches) relied on observations derived from the same community forests.

2.2. Materials

2.2.1. Data, instruments and software used

In addition to the primary data collected from the field for the model and design-based approaches, datasets from different sources and organisations as listed in Table 2-1 were used for the analysis.

Table 2-1: Data sets used

Data	Source
Topography	DEM (20m) of the Ludikhola watershed availed by ICIMOD
Management of CFs	Community forests representatives and field observations and Interviews
Biomass measurements	Field measurements

A list of the instruments used for the study is given in Table 2-2.

Table 2-2: Lists of instruments used

Instruments	Application
30 meter measuring tape	Plot radius measurements
iPAQ and GPS	Field navigation and plot location
Diameter tape (5m)	DBH (1.3m) measurements
Slope meter	Gradient determination
Haga altimeter	Tree height measurements
Datasheets and clipboards	Fieldwork database

A list of different software used for the different stages of processing and analysis for the study is given in

Table 2-3: List of software used.

Software	Application
R	Data analysis and geo-statistical computing
Erdas Imagine 10	Image processing
ArcGIS 10	GIS analysis
Endnote	Thesis reference building
Microsoft office	Thesis writing

A detailed outline of the approach followed and the various stages of analysis and processing together with the research questions addressed at different stages are presented in the flowchart shown in Figure 2-2

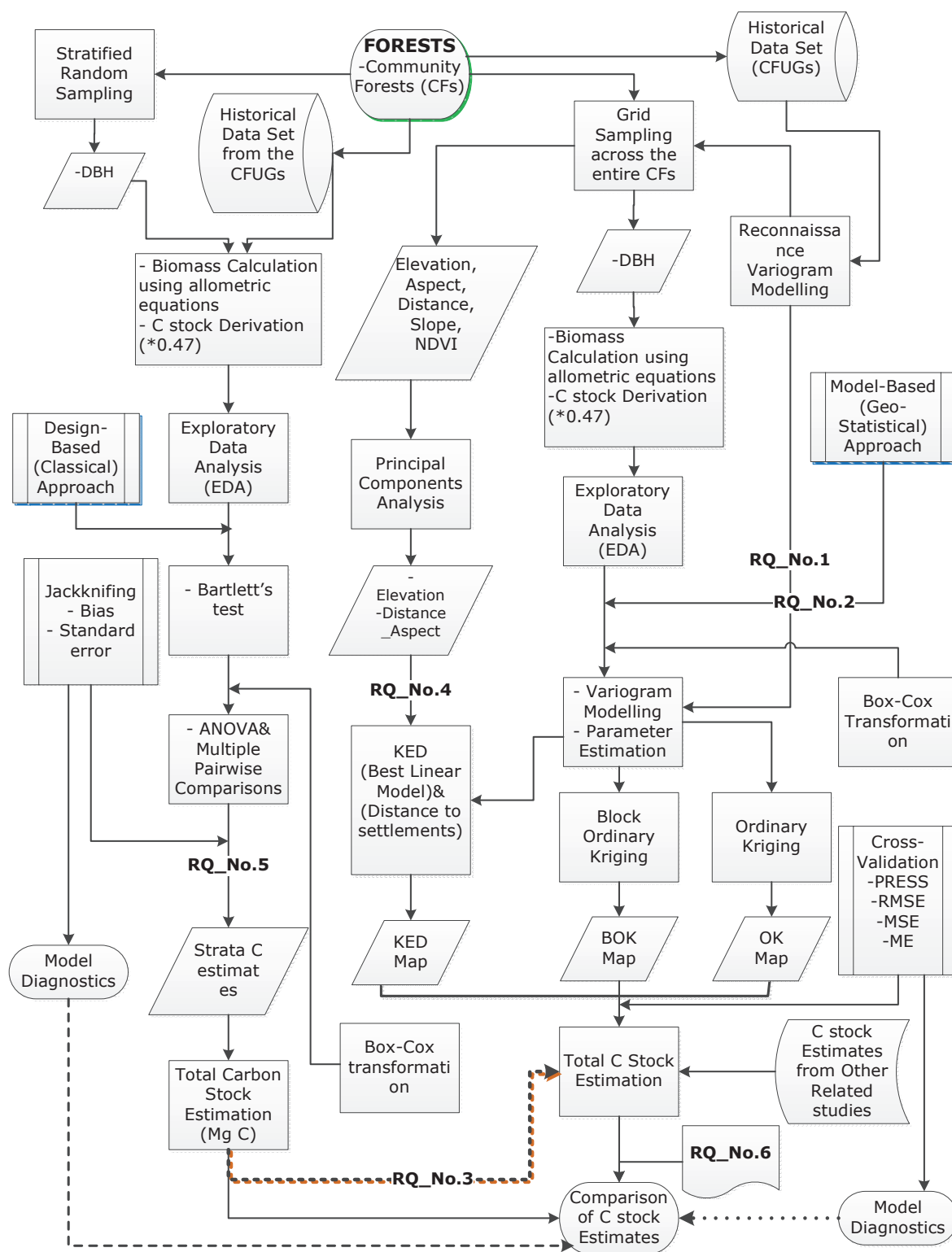


Figure 2-2: Methodological flowchart for the design and model-based approach for C stock estimation

2.3. Above ground tree biomass (AGTB) measurement

Diameter at breast height (at 1.3m) and height of trees of greater than or equal to 10cm DBH were measured in each of the circular 500m² supports using diameter tape, clinometers and linear tape or vertex-IV/ transponder. Trees with DBH less than 10cm generally have insignificant carbon stocks (Gibbs, et al., 2007). Trees were first marked starting from the edge and working inwards to prevent accidental double counting. All the trees marked were then numbered from inwards to the edges, starting from the North in a clockwise fashion. Each tree was recorded individually, together with its species name. Trees on the borders of the plot were included if >50% of their basal area fell within the plot and excluded if <50% of their basal area fell outside the plot (Mugo et al., 2011). Trees overhanging into the plot were excluded, but trees with their trunk inside of the sampling plot and branches out were included. For trees of unusual shape, a standard forestry practice was adopted.

On average, slopes of at least 30% characterize most of the plots in the study area and a slope correction was therefore necessary (Ravindranath & Ostwald, 2007). A slope correction table was carried in the field (see Appendix B) for the determination of the appropriate plot for a given angle.

2.4. Biomass Calculation and carbon stock derivation

Due to the lack of local allometric equations, general IPCC, (2007) equation was used for estimating the above ground biomass while an equation developed by Basuki et al., (2009) for tropical forest was used for *Shorea robusta*. The basis for the application of this allometric equation for this species in particular, emanates from the fact that the mean annual rainfall (2000mm) and temperature range (21°-34° C) is similar to the climatic conditions prevailing in the study area. Similarly, the equation used for the other species was also formulated using DBH ranges from 5-148cm and the rainfall (2000mm) and temperature were similar to the study area. Table 2-4 gives an illustration of these equations.

Table 2-4: Allometric equations for *Shorea robusta* and other species.

Species	Allometric equation	Coefficients			Correction Factor	Source
		a	b	c		
<i>Shorea robusta</i>	$\ln(TAGB) = a + b \ln(DBH)$	-2.193	2.371		1.034	Basuki et al., (2009)
Other species	$Y = e^{(a+b\ln(DBH)+c\ln(DBH)^2)}$	-2.289	2.649	-0.021		IPCC, (2007)

The above ground biomass for each individual species was subsequently converted to carbon stocks per species using a conversion factor of IPCC (2007). Per plot (support) values were then expanded to unit area, in this case, a hectare (Mg C ha^{-1}). The conversion formula is shown as follows:

$$C = B * CF \quad \dots\dots\dots (2-1)$$

Where;

C = carbon stock per species

B = dry biomass

CF = carbon fraction of biomass

2.5. The model-based (geo-statistical) approach

The procedures and methods of analysis followed before and after field work for the geo-statistical approach are described in the paragraphs outlined below.

2.5.1. Pre-field work (Reconnaissance variogram)

A historical dataset (year 2010), Rachina (2011), of 60 observations, collected using a stratified random scheme with a support of 500m^2 was used to model a reconnaissance variogram for the determination of scales of spatial variability and the subsequent derivation of the sampling interval (Oliver & Webster, 2008). Thus, the range of influence and the grid distance for estimating the resource of interest in question using geo-statistics were derived from the variogram parameters of the spherical-model variogram computed using this dataset. The results of the reconnaissance variogram are shown in Figure 3-5.

2.6. Sampling design

186 observations with a support of 500m^2 as the unit of replication were sampled from the 19th of September to the 13th of October 2011. The study area is found within Ludikhola watershed (Figure 2-1) which has 31 community forests (Gautam, 2006). Seven community forests covering

a total area of 687.9 ha were sampled to make up the 186 observations. The 186 observations were collected on a 100 x 100 m regular grid (Figure 2-3), the grid nodes being the location of the sampling points. Assuming that the UTM grid intersection is random with respect to the study area, de Gruijter et al., (2006) asserts that the optimal location of a sampling point is the grid node or intersection.

An initial location was randomly chosen within the study area using ArcGIS sampling facility. Establishment of this point paved the way for the rest of the sampling locations, which were defined so that all the locations were at regular 100 x 100 m over the grid. Sampling was therefore undertaken and done on every node of the regular grid overlaid over the 7 community forest.

Because the precision was unknown, classical formulae for determining confidence intervals could not be applied to come up with the required sample size. As a rule of thumb, Webster and Oliver (1992) recommends computation of variograms with sample data between 100 and 200 data.

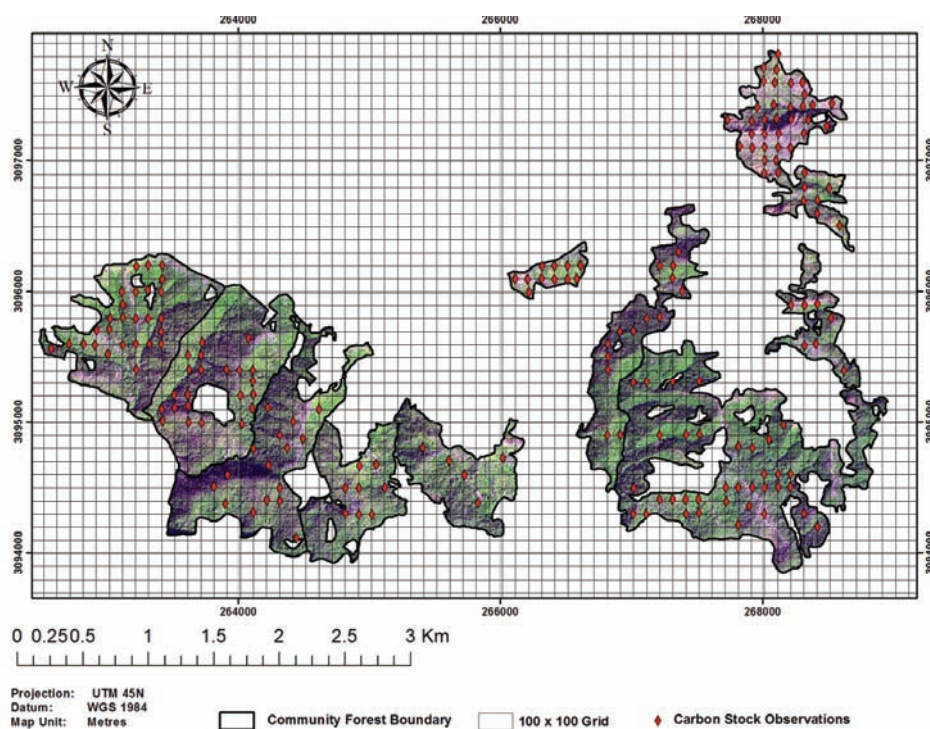


Figure 2-3: Carbon Stock observation locations for the model-based (geo-statistical) approach.

2.7. Rationale for using kriging with external drift

KED was preferred in this study to model the spatial correlation structure of carbon stock since it allows the incorporation of many covariates. Again, comparative studies, Sales (2007), have

shown KED interpolation to perform better than classical statistical approaches in estimating carbon stock.

2.8. Approach for the selection of covariates

Knowledge of the relationship between biomass and topography guided the choice of drift variables for use in kriging with an external drift in this study. In order to predict total forest biomass more accurately, it is necessary to include at a minimum, a plot level characteristic such as elevation (Miettinen & Liew, 2009). Consequently, covariates which included elevation, normalized difference vegetation index (NDVI), slope and terrain aspect were measured at the support level as an attempt to improve the prediction of carbon stock. For areas such as Nepal, which are characterized by rapidly changing elevations and climatic zones, topographic variables exert important control on the spatial distribution of aboveground biomass and subsequently, carbon stocks (Goetz et al., 2009).

In addition to covariates relating to physical factors, anthropogenic factors also play a role in the distribution and density of carbon stock (Kiyono, et al., 2010). For the study at hand, such factors were hypothesized to be the approximate distance of the sampled plots within the community forests from the nearest human settlements. Field assessment through interviews together with visualization of the study area using high resolution satellite imagery (Geo eye, 0.5m resolution) were used to determine the approximate distance of each sampled plot from the nearest human settlements. An assessment of the distribution of the tree species across the community forest was made during the sampling survey. *Shorea robusta* predominated over the other recorded species in all the sampled community forest. From field interviews with community forest personnel, it was established that *Shorea robusta* has an economic importance as it has application in various areas, ranging from furniture to construction. The status of each of the seven community forests was assessed during sampling and the status was established as either being a disturbed or an undisturbed forest as shown in table 2-6.

2.9. Covariates Selected

Elevation

20 m Digital Elevation Model (DEM) was processed in ArcGIS to give elevation data for both observation and interpolation points used in kriging with external drift.

Slope

Slope was derived from elevation data referred above using ArcGIS software.

Aspect

Aspect was derived from elevation data referred above using ArcGIS. This variable is measured on a circular scale. To avoid problems modelling with a circular variable, a trigonometric transformation used by (Richard et al., 2007) was applied as shown in the formula below.

$$Aspect = 0.5 \left(\cos \left(\frac{\pi}{180} Asp - 30 \right) + 1 \right) \dots \dots \dots (2-2)$$

Approximate distance from the nearest human settlements

The approximate straight distance of each sampled plot of the community forest from the nearest human habitations was determined using high resolution Geo-Eye (0.5m) satellite imagery and overlaying of the sampled Carbon stock observations on the Google image.

Normalised difference vegetation Index (NDVI)

NDVI was derived from the high resolution Geo-Eye imagery (0.5m resolution) using Erdas Imagine 2011 software. Digital numbers (DN) were converted to reflectance values of vegetation using the conversion formula (Heo & FitzHugh, 2000).

$$Re\ flectance = a * (DN) + b \dots \dots \dots (2-3).$$

Where;

a and b are coefficients with values 0.0005 and 0.00 respectively

2.10. Clipping of the Interpolation grid and extraction of covariates data

The created grid was imported into ArcMap and converted into a point shapefile with XY coordinates and a projection defined, WGS 1984 to match the other map layers carrying covariates data. It was then clipped in order to confine the grids to the study area. The clipped grid was overlaid with the extracted layers of the covariates in order to extract the data on every grid point. The created layer was then saved as a dbf, changed to a tab delimited text file. At this point, a table for the observation points and the interpolation grid with all the necessary covariates had been created.

2.11. Principal Components analysis and data reduction

A correlation matrix of the response variable (C stock) and five potential predictor variables namely elevation, transformed terrain aspect, normalised difference vegetation index (NDVI), slope and the approximate distance of the sampled plots within the community forests from the

nearest human settlements was made. Any correlations between the covariates with a correlation of more than 50% ($r \geq 0.5$) warrants investigation for multicollinearity (Mansfield & Helms, 1982). Multicollinearity can render the partial regression coefficients of independent variables imprecise, resulting in the t-values becoming less significant (Nelson et al., 1999). Consequently, a principal component analysis was carried out on all the predictors in order to make them independent.

In order to provide insight into the combination of predictors to use for the modelling of carbon stock density and variability in the study area, a principal component analysis was carried out. Thus, principal component analysis provides a selection criteria of candidate predictor variables to retain (or eliminate) which consequently results in data reduction (Mansfield & Helms, 1982).

2.12. Exploratory data analysis

A spatial and non-spatial exploratory analysis of the response and predictor variables was made in order to shape the road map for the subsequent analysis. This is important as it establishes the strength between the independent variable and the predictors and for assessing the adequacy of modelling assumptions before making a decision to transform. For KED, the assumption of normality was tested with the residuals of the best linear model. A spatial plot of carbon stock observations with respect to their locations was done using geoR geo-statistical library (Ribeiro & Diggle, 2001).

2.13. The Box-Cox transformation approach

Because of the non-normality in the error distribution of the residuals of the model-based linear models (kriging with an external drift), a Box-Cox transformation was used as an objective way to find an optimal transformation that befitted the data requirements (Box & Cox, 1964).

An estimate of λ was obtained by finding the value of λ that maximizes the log-likelihood function (shown in Equation. 2-4), which is proportional to the probability of observing untransformed carbon stock data, when a normal independent model properly describes the transformed carbon stock observations (Rahman & Pearson, 2009).

$$\log(L(\lambda / y_1, y_2, \dots, y_n)) = -\frac{n}{2} \log(s^2) + (\lambda - 1) \sum_{i=1}^n \log(y_i) \dots \dots \dots (2-4)$$

Where;

$\log(L(\lambda / y_1, y_2, \dots, y_n))$ is the log-likelihood function

n is the number of carbon stock observations

s^2 is the estimated variance (using n as the divisor) of the transformed carbon stock observations

y_i denotes the original carbon stock observations

λ is the interim estimate of the unknown transformation parameter.

2.14. Selection of the best linear model

A criterion based approach introduced by Akaike and Schwarz (1978) as defined by the Akaike Information Criterion (AIC) in equation 2-5 was used to select the best candidate model for use in kriging with an external drift.

$$AIC = -n[\ln(SS_{Residual})] + 2(p+1) - n\ln(n) \dots\dots\dots(2-5)$$

Where;

p is the number of predictors

n is the number of observations

$SS_{Residual}$ is the residual sum of squares

A related statistic to the AIC is the Bayesian Information Criterion (BIC) that works by imposing heavy penalties to larger models, resulting in the selection of smaller models in comparison to AIC and its algorithm is shown in equation 2-6.

$$BIC = -2\log\text{--}likelihood + p\log n \dots\dots\dots(2-6)$$

Criterion-based methods employ a wider search for the best model and compare models in a preferable manner (Faraway, 2002). Hence, the selection of this approach for linear model selection justifies from the inability of the stepwise procedures in considering the increased probability of Type I error due to multiple testing. Furthermore, the forward, backward and stepwise approaches for the inclusion or exclusion of predictors is inconsistent, resulting in different final models, even from the same dataset (James & McCulloch, 1990; Mac Nally, 2000). The number of potential predictors from the principal components regression were 3 and by the criterion-based rule, $2^3 (=8)$ candidate models were fitted and the model with the lowest AIC was chosen for subsequent modeling using KED.

2.15. Cross-validation

In order to assess the validity of the predictions, cross validation statistics as outlined in Rossiter (2004) and Webster and Oliver (2001) were calculated for each prediction method. The formulae shown below give a summary of how the cross-validation statistics of mean error (ME), mean squared error (MSE), root mean squared error (RMSE) and the predicted residual sum of squares (PRESS) were calculated in order to juxtapose the three kriging algorithms:

$$ME = \sum_{i=1}^n \frac{(Z^* - Z)}{n} \dots\dots\dots (2-7)$$

$$MSE = \sum_{i=1}^n \frac{(Z^* - Z)^2}{n} \dots\dots\dots (2-8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Z(x_i) - \hat{Z}(x_i))^2}{n}} \dots\dots\dots (2-9)$$

Where;

Z^* are the prediction values

Z are the mean values for carbon stock

$Z(x_i)$ are the observed carbon stock values

$\hat{Z}(x_i)$ are the estimated carbon stock values

n is the size of prediction of each validation case.

$$PRESS = \sum_i \hat{\epsilon}^2 \dots\dots\dots (2-10)$$

Where:

ϵ is the residual difference between the predicted and the observed values

The mean error gives an indication of the bias and the mean squared error gives the prediction accuracy while the root mean squared error gives an indication of how precise the predictions are (Utset, Lopez, & Diaz, 2000).

2.16. The design-based approach

2.16.1. Sampling Design

Under this plan, the population of interest was divided into L mutually exclusive and exhaustive strata and a simple random sample of n_L elements was taken within each strata L (Cochran, 1977). Letting $N_1, N_2, \dots, N_L$ to represent the number of units within each strata and $n_1, n_2, \dots, n_L$ to represent the number of randomly selected elements within each strata, then the total

possible number of possible stratified random samples that could be drawn within the study area is given by the following inequality (Neyman, 1978):

$$\binom{N_1}{n_1} \times \binom{N_2}{n_2} \times \dots \times \binom{N_L}{n_L} \leq \binom{N}{n} \dots \dots \dots (2-11)$$

Where;

N_L is the size of stratum L

n_L is the sample size for stratum L

N is the total population size

n is the overall sample size from all the strata.

Data derived from a historical dataset from the year 2010 (Rachina, 2011) from the same sampled community forests were integrated with the current design-based samples to make up the total required sample size to 115. This was done in order to make the two sampling designs comparable in terms of the study area covered. The following formula was used to calculate the required sample size and the respective stratum sample sizes:

$$\frac{\left(\sum_{i=1}^7 N_i \cdot S_i \right)^2}{\frac{N^2 \cdot E^2}{t^2} + \left(\sum_{i=1}^7 N_i \cdot S_i \right)^2} \dots \dots \dots (2-12)$$

Where;

n = Total number of sample plots (total number of sample plots required) in the project area

i = Project strata number from 1 to the last strata sampled

N_i = maximum possible number of sample plots in stratum i

S_i = Standard deviation for each stratum i

N = Maximum possible number of sample plots in the project area.

The following equation, (2-10) was used to distribute the total number of sample plots over the seven community forest.

$$n_i = \frac{n \cdot N_i \cdot S_i}{\left(\sum_{i=1}^7 N_i \cdot S_i \right)^2} \dots \dots \dots (2-13)$$

Where;

n_i = Number of sample plots for stratum i

A summary of the area and proportions of the community forest and the required sample sizes is shown in table 2-2.

Table 2-5: Proportions and required sample sizes for the sampled community forests.

Community Forest (CF)	Area (ha)	Proportion	Required sample size (n_i)
Berinchok	83.5	0.12	14
Chisapani	50.1	0.073	8
Kkarkopakho	51.1	0.074	9
Kuwadi	92.3	0.13	16
Ludi damgade	270.7	0.43	47
Shikhar	50.8	0.074	8
Taksartari	89.3	0.13	13
Total	687.9	1.00	115

The number of samples could be assigned to the strata equally, proportionally (i.e., proportional to the stratum sizes), optimal (i.e., by strata sizes and variances) or by Neyman allocation that in addition to strata sizes and variance, includes the assessment costs per stratum (Miranda & Souto de Miranda, 2011). However, in this study, proportional allocation of samples for monitoring and payment of REDD credits was most appropriate and feasible. This is because there was no information on the stratum variances available to enable calculation of samples sizes for optimal allocation (de Gruijter, et al., 2006). Again, changes in stratum (community forests) assignments over time do not affect the probabilities of selection, thus complicating estimation. Figure 2-4 shows a map of the location of the sampling plots for the design-based approach.

2.16.2. Test of the homogeneity of variance of the sampled community forests

The test of the analysis of variance explicitly assumes that the considered population (community forests) were generated by normal distributions with identical variances. It was therefore tested to check whether this assumption is plausible with regard to the samples at hand using Bartlett's statistic (Kutner et al., 2005) as shown in equation 2-14:

$$Q = \frac{A}{B} = \frac{(\sum_{i=1}^k v_i) \log S^2 - \sum_{i=1}^k v_i \log S_i^2}{1 + \frac{1}{3(k-1)} \left[\sum_{i=1}^k \left(\frac{1}{v_i} \right) - \left(\frac{1}{\sum_{i=1}^k v_i} \right) \right]} \quad \dots\dots\dots (2-14)$$

Where:

$$v_i = n_i - 1$$

S_i^2 is the unbiased estimate of σ_i^2

$$S^2 \text{ is the "pooled variance" estimate of the common variance } \sigma^2 = \frac{\sum_{i=1}^k v_i S_i^2}{\sum_{i=1}^k v_i}$$

2.16.3. Jackknifing

A jackknifing procedure was applied in order to calculate the standard error and bias estimates of the fitted ANOVA model. Jackknifing provides a parametric statistical inference of the dataset by applying re-sampling without replacement to the original dataset (Wells, 1994). Letting the statistic to be calculated to be denoted by $[\theta^*]$, the successive pseudo values of θ^* with the i^{th} observation removed is given by:

$$\tilde{\theta}_i = n\theta^* - (n-1)\theta_{-i} \quad \dots\dots\dots (2-15)$$

Where;

θ^* is the statistic calculated from the sample with observation i omitted.

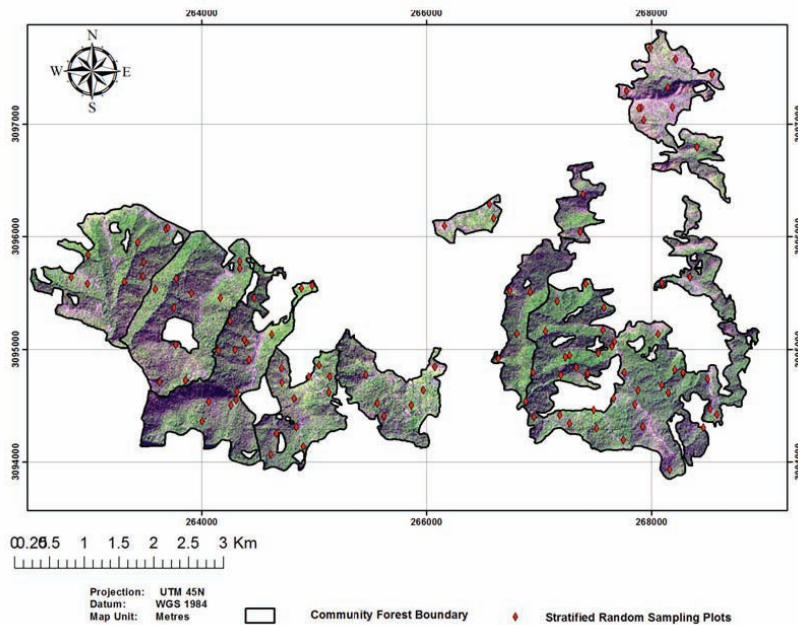


Figure 2-4: Location of sampling plots for the design-based approach.

2.17. Significance test for the assessment of confidence interval overlap

A test of significance to establish the extent of overlap between the two estimators (model and design-based approach) was carried out using a method applied by Goldstein and Healy (1995).

Letting the mean of the two estimators be denoted by $\bar{X}_1$ and $\bar{X}_2$, independently and normally distributed with standard errors σ_1 and σ_2 respectively, the confidence intervals do not overlap if the following inequality is satisfied:

$$\bar{X}_1 - \bar{X}_2 > Z_{\beta}(\sigma_1 + \sigma_2) \quad \dots\dots\dots (2-16)$$

Where:

Z_{β} Is the (positive) normal quartile with two-tailed probability β

The probability that the inequality in 2-16 is satisfied under the null hypothesis of equal means of C stocks of the aforementioned estimators is therefore given by equation 2-17 as follows:

$$\gamma_{12} = 2 \left[1 - \phi \left(\frac{Z_{\alpha} (\sigma_1 + \sigma_2)}{\sqrt{2}} \right) \right] \dots\dots\dots (2-17)$$

Where:

γ_{12} Denotes the comparison of estimator 1 (model-based) to estimator 2 (design-based)

ϕ Is the normal integral

2.18. Description of the strata (community forests)

Each of the seven community forest in which the samples for the model and design based approaches were taken was described according to information obtained from interviewing community forest staff and also observations recorded during the sampling exercise. A description of these community forests is given in table 2-6.

Table 2-6: Summary of the characteristics of the sampled community forests.

Name of Community Forest	Management style	Tree Density	Forest status	Accessibility	Dominant Species
Birechwork	- Mainly grazing	High	Undisturbed and dense	Difficult	<i>Shorea robusta</i>
Chisapani	- Grazing with mild cutting	Moderate	Some signs of disturbance	Difficult	<i>Shorea robusta</i>
Kharkopakho	- Recovering from degradation	Low	Mild disturbance, dense	Easily accessible	<i>Shorea robusta</i>
Kuwadi	- Little cutting	High	Dense and undisturbed	Difficult	<i>Shorea robusta</i>
Ludi damgade	- Land use conversion high, forest recovering from degradation	Low and patchy	A disturbed forest, secondary succession forest	Easily accessible	<i>Schima wallichii</i>
Taksartari	- Grazing with moderate encroachment	Moderate	Forest in secondary succession	Easily accessible	<i>Shorea robusta</i>
Shikhar	- Little encroachment and moderate cutting of undergrowths	Moderate	Maturing forest.	Difficult	<i>Shorea robusta</i>

3. RESULTS

3.1. Non-spatial exploratory data analysis

Table 3-1: Descriptive statistics for the target and potential predictors. MEAN- mean, STDV- standard deviation, MED- median, MIN- minimum, MAX- maximum, SKEW- skewness, KURT- kurtosis.

	Target variable			Predictors		
	Carbon stock	Elevation	Distance	Aspect	Slope	NDVI
MEAN	70.86	794.17	1.60	0.45	30.01	0.37
STDV	49.85	175.20	1.30	0.37	12.45	0.25
MED	63.05	797.20	1.44	0.58	34.41	0.29
MIN	1.98	471.3	0.01	0.001	1.98	0.02
MAX	317.60	1141	3.99	0.99	49.63	0.89
SKEW	1.09	0.08	0.37	-0.13	-0.85	0.61
KURT	2.55	-1.07	-1.38	-1.68	-0.08	-0.94
Shapiro test ($\alpha = 0.05$)	1.2×10^{-8}	9.14×10^{-3}	1.2×10^{-10}	1.8×10^{-4}	1.8×10^{-9}	1.8×10^{-10}

Table 3-1 gives a display of the summary statistics of the raw data for both covariates and the response. A Shapiro test to establish the probability with which the hypotheses of a normally distributed population could be rejected with was carried out on the primary and secondary variables. Serious deviations from normality in the response, distance, slope and NDVI covariables were noticed. Scatter plots of the response and secondary variables (Figure 3-2) were made in order to assess the strength of the relationship between the primary variable and covariates. A scatter plot of carbon stock and elevation, distance and slope covariables appeared strong and approximately linear whereas a scatter plot of the response and aspect and NDVI covariates did not suggests any meaningful linear relationship.

A log transformation was therefore applied to the distance of the sampled CF from the nearest human settlements before it could be used for the linear modeling of carbon stock data. On the other hand, a histogram of the distribution of the primary variable (Figure 3-1a) suggested the predominance of young individuals (young forest) within the sampled population as the graph displayed a negatively skewed distribution.

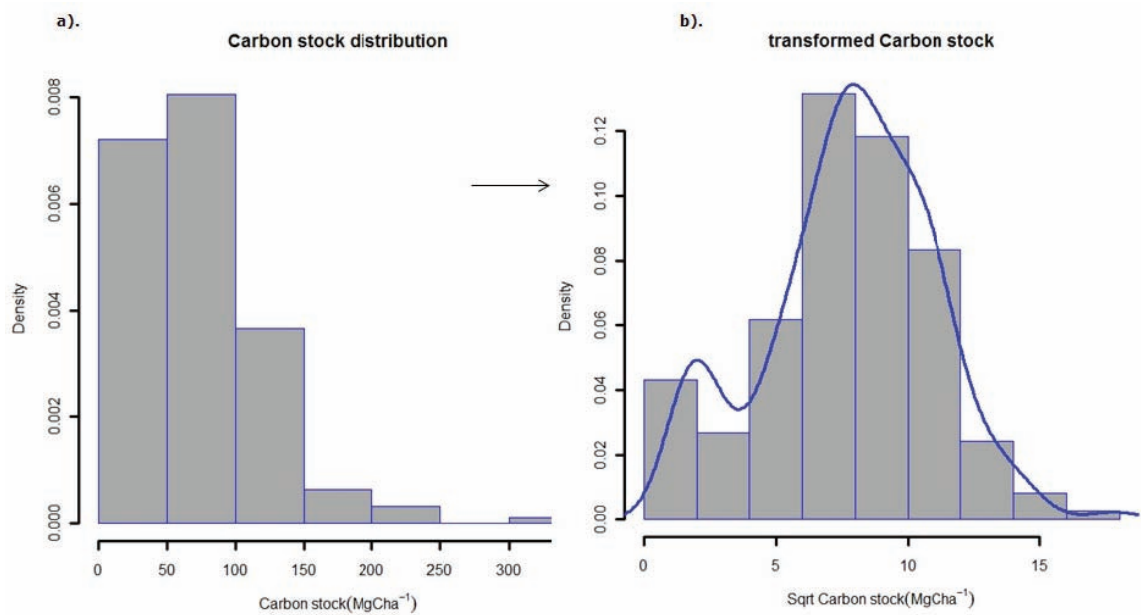


Figure 3-1: Distribution of carbon stock, a). Before transformation and b). After Square root transformation.

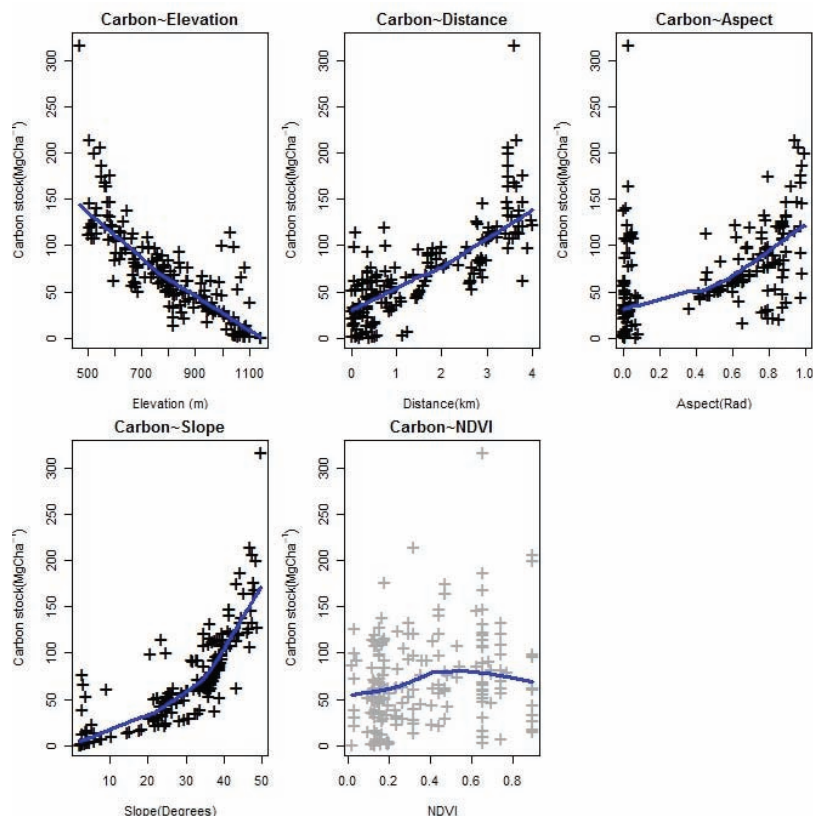


Figure 3-2: Relationship between carbon stock and candidate predictor variables.

3.2. Spatial exploratory data analysis

Figure 3-3 shows the sampling intensity across the entire study area (community forests) showing no evidence of discordance of observations with respect to their spatial neighbours. In particular, the display in Figure 3-3b shows the distribution and concentration of the target variable with respect to geography. Concentration of sampled carbon stock was high in the southern region, suggesting the existence of a spatially varying mean due to a trend surface (Ribeiro & Diggle, 2001). It is therefore evident from this plot that the behaviour of carbon stock density is not the same across the different geographical regions of the sampled field. A quartile plot (Figure 3-3b) of the empirical distribution of carbon stock revealed a predominance of third quartiles (> 75%) carbon stock values in the south east and south western regions of the sampled community forests.

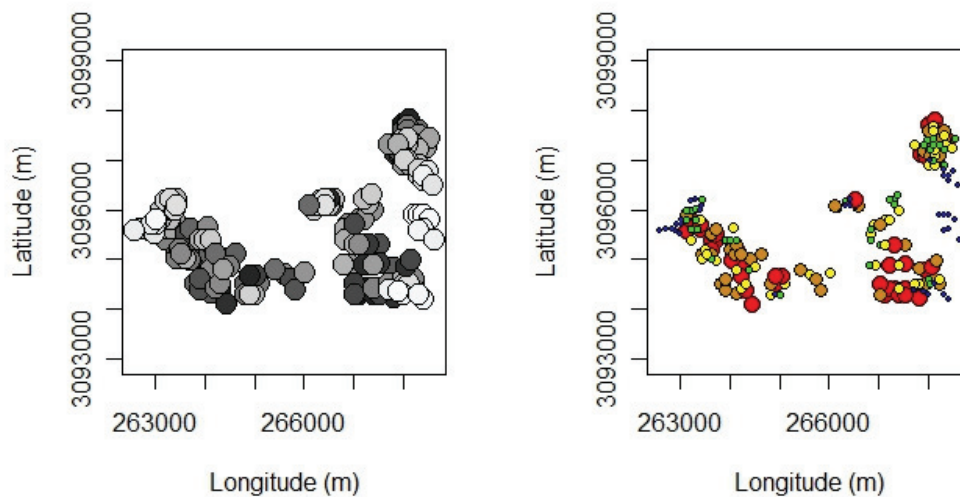


Figure 3-3: **a).** The behavior of carbon stock density and distribution with respect to geography and **b).** quartiles of the empirical distribution of measured carbon stock values.

3.2.1. Directional variograms and anisotropy

A clear indication of the lack of independence of the primary variable with respect to direction (directional dependency) was shown when a plot of directional variograms in four angles, namely north, north east, east and south east direction, (Figure 3-4) was done. The directional variograms exhibit different nugget effects and total sills, a hint for anisotropy.

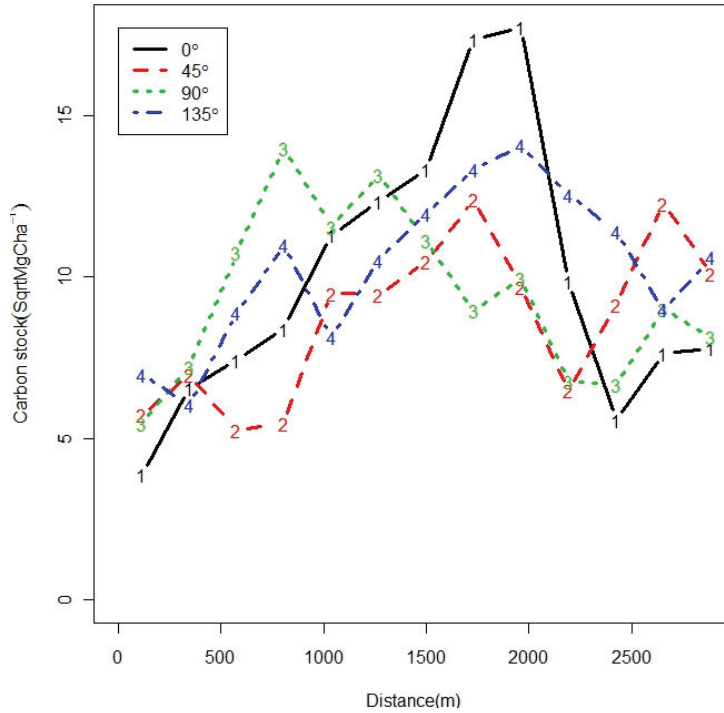


Figure 3-4: Directional variograms for Carbon stock

This outcome justifies the modeling of carbon stock variability and distribution with a trend. From the display of the plot, the existence of anisotropy was evident and this suggested a non-stationary mean in the primary variable and the stationarity assumption was therefore not valid in this study (Oliver & Webster, 2008). Subsequent modelling of C stock therefore relied on the information regarding the displayed trend.

3.3. Response variable (C stock) transformation

A maximum likelihood estimate of λ was estimated to be 0.5 with estimated 95% confidence interval of $(0.4 \leq \lambda \leq 0.6)$ which entailed a square root transformation to the carbon stock dataset (Figure 3-6). Thus, the square root transformation proved useful for the dataset at hand and the optimality and appropriateness of this transform is demonstrated by the perfect error distribution of the residuals of the best linear model of feature space used in the prediction of C stock (Figure 3-12b and Figure 3-13).

Transformation of C stock improved its correlation with the predictor variables as shown in Table 3-2 with elevation and distance from the nearest human settlements having the highest pair wise correlation with the transformed C stock

3.4. Reconnaissance variogram

This analysis showed a high nugget/ total sill ratio (0.62), indicating a weak spatial structure and high short-range variability, even after averaging over the 500m² support. The analysis of spatial structure from the sample collected for this study showed ranges of 750 to 1514m (Figure 3-5), considerably longer than the preliminary analysis. Thus, the grid distance could have been wider than the lag distance subsequently used for the modelling of the spatial structure of carbon stock.

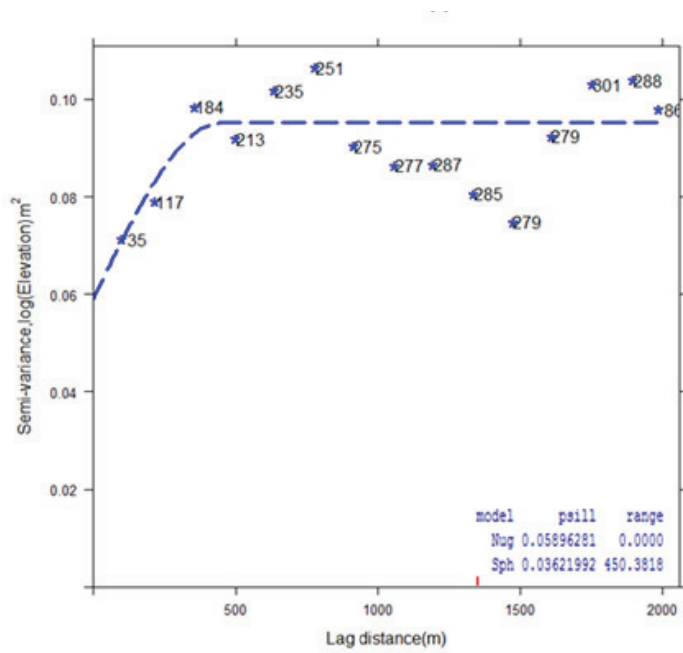


Figure 3-5: Reconnaissance variogram for the sampling distance of the model-based approach.

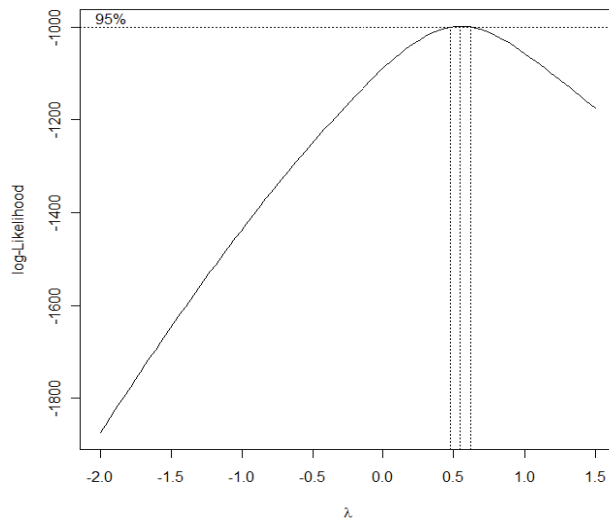


Figure 3-6: Log-likelihood plot for the Box-Cox transformation of the Carbon stock response variable

Table 3-2: Correlation matrix between the predictor variables and the response, carbon stock Sqrt (Mg C ha⁻¹).

	Carbon	Aspect	Distance	Elevation	NDVI	Slope
Carbon	1.00					
Aspect	0.54	1.00				
Distance	0.76	0.34	1.00			
Elevation	-0.89	-0.47	0.69	1.00		
NDVI	0.23	0.38	-0.12	-0.20	1.00	
Slope	0.77	0.81	0.47	-0.84	0.23	1.00

Figure 3-7 shows the variance that each of the PCs explained. A decision on the principal components to omit is critical for principal components regression (Conerly & Mansfield, 1988). Hence, simple deletion of those small eigenvalues (little relative contribution to the total variation in the X-predictor variables) can result in loss of important variables (Lecomte, 1990). The strength of the relationship of each component with the response was therefore tested as shown in Table 3-4. The eigenvectors of the first three components were highly significant and therefore used as predictor candidates for the modelling of C stock using KED.

Table 3-3 gives a breakdown of the proportion of variances explained by the different principal components alongside the predominant predictors for each of the components.

Table 3-3: Proportion of variances explained by the principal components.

	Comp1	Comp2	Comp3	Comp4	Comp5
Dominant variable	Elevation	Distance	Aspect	Slope	Elevation
Variance	3.00	1.05	0.56	0.26	0.12
Proport. of Variance	0.60	0.21	0.11	0.053	0.025
Cumulative Proport.	0.60	0.81	0.92	0.97	1.00

An objective way of filtering out insignificant principal components is to test the significance their respective eigenvectors in predicting the response (C stock) and that test demonstrated elevation, distance to nearest human settlements and aspect to be highly significant predictors as shown in Table 3-4.

The range of 1541 m (Table 3-5) shows that the spatial dependence of C stock is increased from the estimated range of 450 m to 1541m. Thus, in order to match the sampling density of the variable in question and the scales of variability, a grid cell resolution of 10 000m² was subsequently used in the kriging methods (Hengl, 2007). The ratio of nugget (C_0) to sill (C_1) was used as a characteristic of the strength of spatial correlation structure in the carbon stock data set. Cambardella et al., (1994) described different classes of spatial dependence using the ratio between the nugget and total sill variance. Values less than 25% are categorized as strongly

spatially dependent, 25-75% as moderately spatially dependent and values more than 75% as weakly spatially dependent. In line of this argument, a nugget to sill ratio for carbon stock of 0.30, Figure 3-8, corresponded to a moderately spatially dependent variable implying that 30% of the carbon stock variability consisted of unexplained, short distance random variation. It is therefore logical to anticipate good and relatively precise C stock predictions.

3.5. Variance explained by the principal components

Figure 3-7 shows a scree plot that guided the judgement on the number of principal components that were worth considering. As shown (Figure 3-7 and Table 3-3), most of the variation was explained by the first three components.

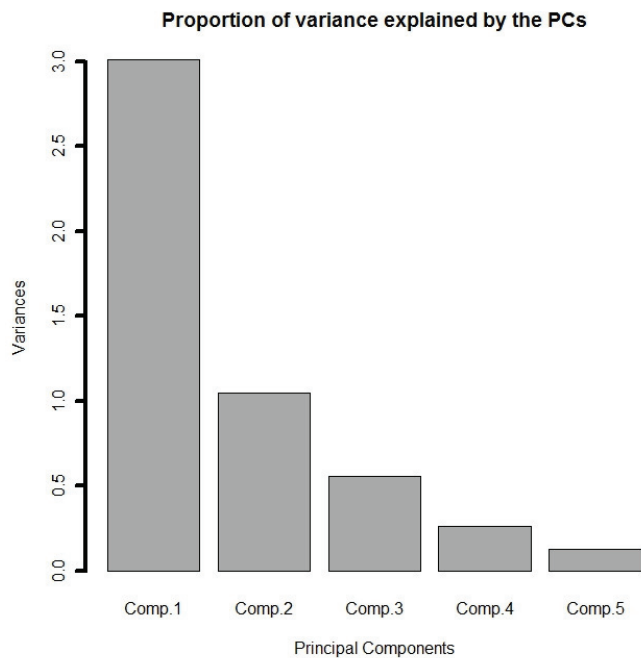


Figure 3-7: Scree plot of the results of a Principal Component Analysis (PCA) for the C stock dataset.

Table 3-4: Significance of the eigenvectors of each predictor variable with respect to the response, carbon stock.

Coefficient	Estimate	Standard error	t	p ($\alpha = 0.05$)
Intercept	70.80	1.49	47.32	2.0 x 10 ⁻¹⁶ ***
Elevation	-25.59	0.86	-29.56	2.0 x 10 ⁻¹⁶ ***
Distance	-8.8	1.46	-6.04	8.55 x 10 ⁻⁹ ***
Aspect	-6.71	2.01	-3.37	0.000919 ***
Slope	-2.70	2.92	-0.92	0.35
NDVI	-1.1	4.26	-0.26	0.79

***Highly significant

3.6. Spatial structure of carbon stock

Ordinary kriging, which relies on the spatial dependence in the primary variable (sampled carbon stock values) was used as a reference to assess the actual gain of accounting for covariates. A spherical model as defined in equation 1-3 and 1-4 was fitted to the variogram.

The nugget variance shows variance at scales where measurements of carbon stock could not be taken. The sill is the structured variance (autocorrelation components) and the range is the limit of spatial dependence after which the sampled variable is spatially independent (Lopez-Granados et al., 2002.).

Table 3-5: Parameters for the ordinary variogram and residuals for the distance and elevation covariates

Prediction Method	Model	Nugget (C_0)	Partial Sill (C_1)	Range
Ordinary Variogram	Spherical	3.77	8.45	1541
Log (Distance)	Spherical	2.02	3.95	1035
Elevation + Log (Dist)	Spherical	1.04	0.59	762

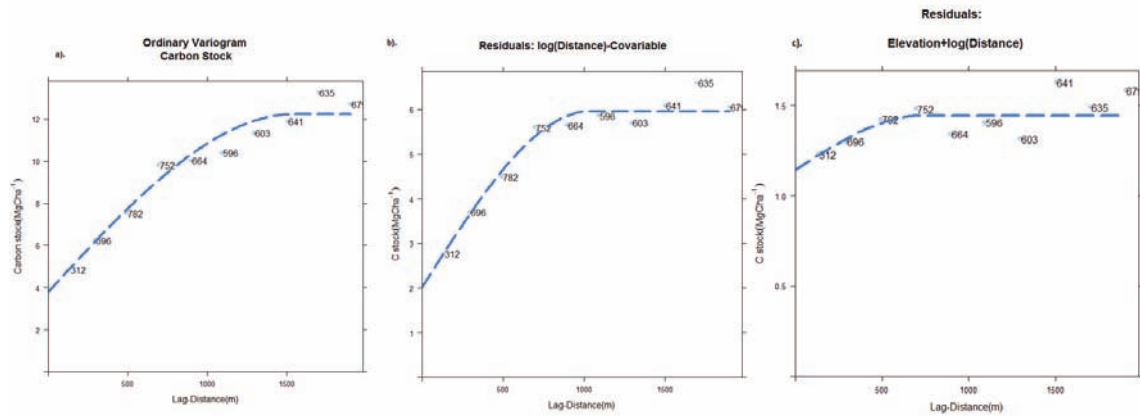


Figure 3-8: Stage wise comparison of the effects of covariates on the spatial correlation structure of carbon stock.

Covariates (elevation and distance from the nearest human settlements) accounted for most carbon stock variability. Figure 3-8 illustrates a stage wise comparison of the effects of the predictors on the spatial correlation structure of carbon stock. As illustrated in Table 3-5, log (distance) covariate accounted for 51.1% of the spatial structure in carbon stock while a linear model of elevation and log (distance) accounted for 87.3% of the spatial structure of carbon stock. Figure 2-1 shows that the best linear model of feature space (using Akaike Information Criterion, Table 2-1) accounted for most of the variability and subsequently took away most of the spatial correlation structure of carbon stock (87.3%).

The total sill was drastically reduced ($12.22 \text{ Mg C ha}^{-1}$ vs. $1.63 \text{ Mg C ha}^{-1}$), the range of spatial correlation reduced by half (1541m vs. 762m) and the nugget/total sill ratio greatly raised ($0.31 \text{ Mg C ha}^{-1}$ vs. $0.64 \text{ Mg C ha}^{-1}$).

3.7. Feature space model selection criterion

Three predictors (elevation, distance from the nearest human settlements and aspect) selected from a test of significance of the eigenvectors of the principal components (Table 3-3 and Table 3-4) were used to assess their effect on the spatial correlation structure of carbon stock.

Incorporation of the different predictor variables from the results of a principal component analysis in the modelling of the spatial correlation structure of carbon stock resulted in an additive model with elevation and distance as covariables giving the most desirable qualities for explaining the variability of the response. These qualities ranged from linear model diagnostics (meeting regression assumptions) to the amount of variability explained for carbon stock by the predictors. However, models including interaction terms had lower AIC (i.e., excessive complexity) than the most parsimonious model eventually applied for KED (Table 3-6).

Table 3-6: Summary of the selection criterion for the best candidate linear model of carbon stock prediction using kriging with an external drift.

Predictors	Model	AIC	BIC	adj. R ²
1	elevation	621.4	631.8	67.7
1	aspect	895.8	905.4	30.8
1	log (distance)	845.9	855.5	47.7
2	elevation*aspect	683.1	697.7	71.6
2	elevation + aspect	659.8	689.7	77.8
2	elevation + log (distance)	616.2	629.0	87.3
3	elevation*aspect*log (distance)	614.9	627.8	88.6
3	elevation + aspect + log (distance)	639.2	688.1	82.6

3.7.1. Linear model interpretation

A summary of the coefficients of the best linear model of feature space is presented in Table 3-7.

Table 3-7: summary of the coefficients of the best linear model of feature space

Predictors	Coefficients	P-values	Significance ($\alpha=0.05$)
Intercept	20	2×10^{-16}	Highly significant
Log (distance)	0.55	0.0095	Moderately significant
Elevation	0.02	2×10^{-16}	Highly significant

Carbon stock will decrease, on average, by 0.02 sqrt (Mg C ha⁻¹) for each 1m increase in elevation, net of the effects of changes due to distance of the community forest from the nearest human settlements. C stock will also increase by an average of 0.55 sqrt (Mg C ha⁻¹) for every 1 log(km) increase in the distance to the nearest human settlements, net of the effects of changes due to changing altitude;

(Equation: $(carbonstock)^{1/2} = 20 - 0.02Elevation + 0.55\log(distance)$, $F_{2,183}=505.3$, $P=2.13 \times 10^{-16}$, adj. $R^2=87.3\%$).

A reduction in the standard error between a design and model-based approach was calculated after deriving the standard deviations for each of the two sampling designs. The standard error for kriging predictions using ordinary block kriging with a resolution of 2 500m² was derived by averaging the prediction standard error for all the sampling locations over the entire study area. An average standard error for ordinary block kriging prediction of 2.17 Mg C ha⁻¹ compared to a standard error of 3.67 Mg C ha⁻¹ for the classical approach made a standard error reduction of 37.8%. The rest of the other error reductions for the KED and OK are shown in Table 3-8:

Table 3-8: Efficiencies of the different kriging variants in comparison to the design-based approach.

Prediction method	% standard Error Reduction
Ordinary Kriging	20.7
Ordinary Block Kriging	37.8
KED-log(distance)	49.1
KED-log(distance)+Elevation	67.8

Table 3-8 demonstrates the importance of variance reduction when explanatory information is included in the prediction of carbon stock, with KED having the highest standard error reduction compared to ordinary kriging and ordinary block kriging. This effect was also noticeable in the variogram of residuals (Figure 3-8)

3.8. Ordinary kriging

The results of ordinary kriging (Figure 3-9a) at the original support (500m²) shows predictions dominated by high carbon stock near C stock locations. The predicted carbon stock density varied from 2.37 to 10.99 sqrt (Mg C ha⁻¹) (1st and 99th percentiles) with standard error varying from 2.17 to 3.23 (Mg C ha⁻¹). The kriging prediction variances (Figure 3-9b) gives a statistical measure of uncertainty across the spatial field and shows that most of the locations near sampling locations had smaller uncertainty compared to locations remote from the known carbon stock

observations. Thus, the quality of the ordinary kriging prediction map is not better than 10.44 (Mg C ha^{-1}).

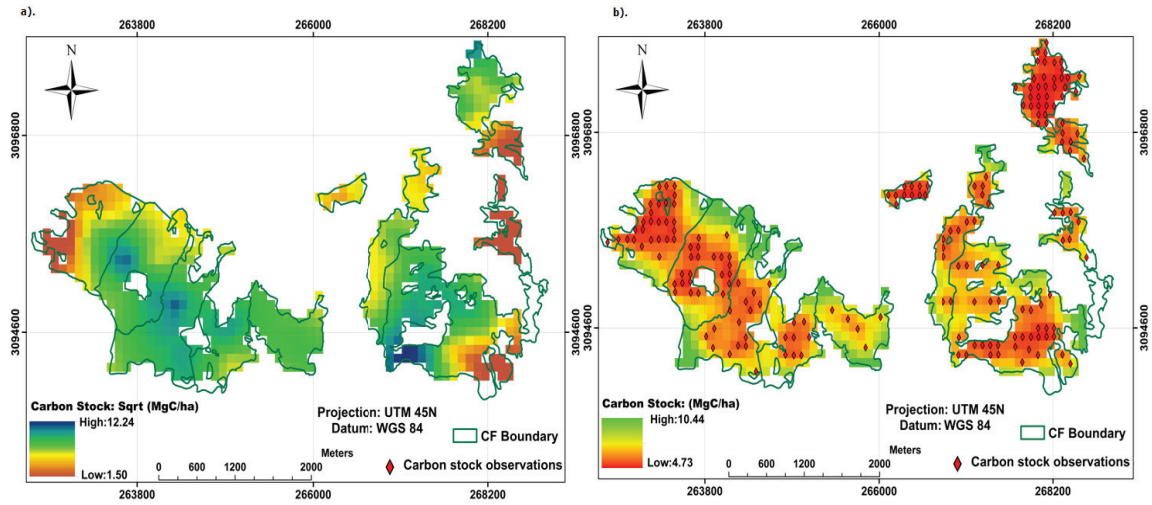


Figure 3-9: Kriging results illustrating a). Square root kriging predictions and b). Kriging variances of carbon stock (Mg C ha^{-1}).

3.9. Block Ordinary kriging

A block area with a resolution of $10\,000\text{m}^2$ (1 ha) was used to provide mean carbon stock predictions over the field of the sampled community forests since that is about the average unit of estimation for C stock reporting and assessment. The predicted carbon stock density varied from 2.37 to 10.99 sqrt (Mg C ha^{-1}) (1st and 99th percentiles) with standard error varying from 0.84 to 2.53 (Mg C ha^{-1}). The block kriging variances (Figure 3-10b) are less than point kriging variances for the same input dataset. The kriged map is smoother compared to the ordinary kriged map.

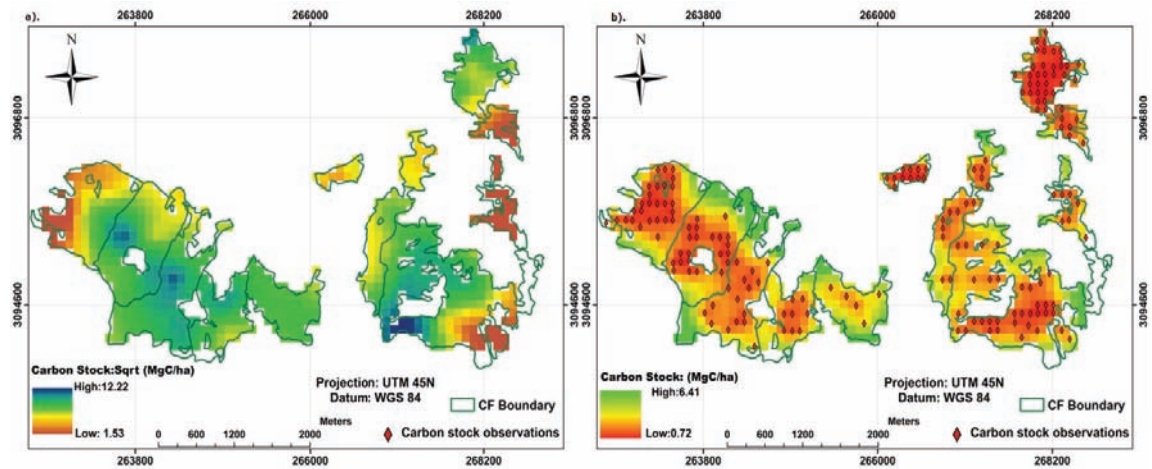


Figure 3-10: Ordinary block-kriging illustrating a). Square root kriging predictions and b). Kriging variances of C stock (Mg C ha^{-1})

3.10. Kriging with external drift (KED)

3.10.1. Log-distance covariable

Predictions for kriging with an external drift, using log (distance) as a covariable were made after modelling the ordinary variogram for carbon stock. The predicted carbon stock density varied from 2.22 to 10.71 sqrt (Mg C ha⁻¹) (1st and 99th percentiles) with standard error varying from 1.65 to 2.50 (Mg C ha⁻¹). In this case, prediction variances (Figure 3-11b) were greatly reduced at locations near the known (observed) carbon stock and showed a steady increase in locations away from them. Hence, the error tends to be higher in places where there is little or no sample data (areas approaching the periphery of the community forests).

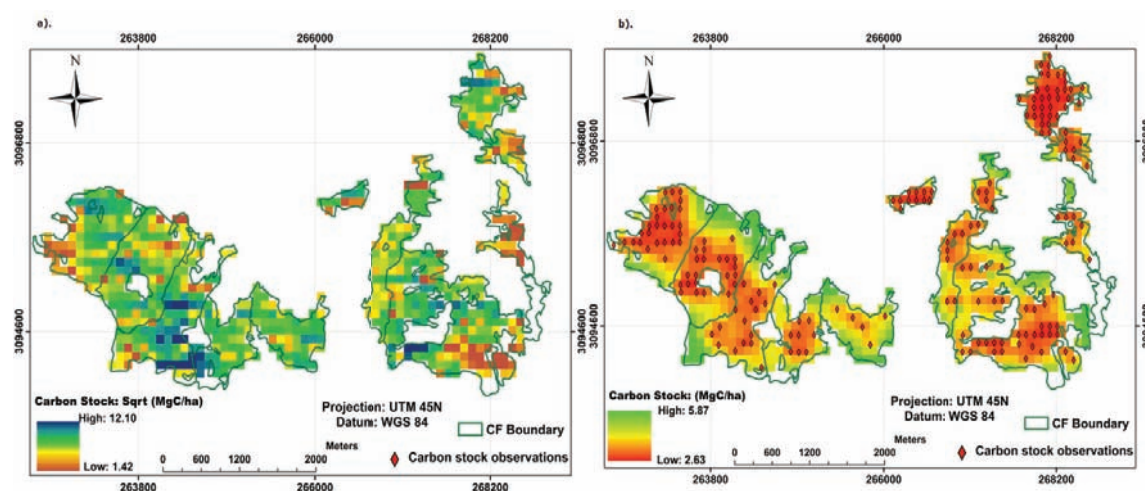


Figure 3-11: Kriging with external drift (KED) results with log (distance) as a covariable, illustrating, **a**). Square root kriging predictions and **b**). Kriging variances of Carbon stock (Mg C ha⁻¹)

3.10.2. Elevation and log (distance) covariables.

The inclusion of elevation and distance as predictor variables resulted in a decrease of the total sill of the variogram (Figure 3-8c) and a shortening of the range of influence. The best feature space model predicted carbon stock density with a range of (1st and 99th percentiles) 3.82 to 13.32 sqrt (Mg C ha⁻¹) with a standard error varying from 1.13 to 1.35 (Mg C ha⁻¹). The shortening of the range of influence and the decrease in the total sill (Figure 3-8c) demonstrates the predictive power with which a linear additive model of elevation and distance has on the spatial correlation structure of carbon stock than when distance was regressed on its own. Hence, the quality of prediction was greatly improved as a result of using covariates in the prediction of C stock with the highest predictions occurring in the south west part of the sampled community forests Taksatari and Birechok CFs) (Figure 3-13a).

Consistent with theory, the KED estimated error variance seemed to be dependent on the observed data configuration, where the uncertainty of estimation decreased towards the sampling locations (Figure 3-13b).

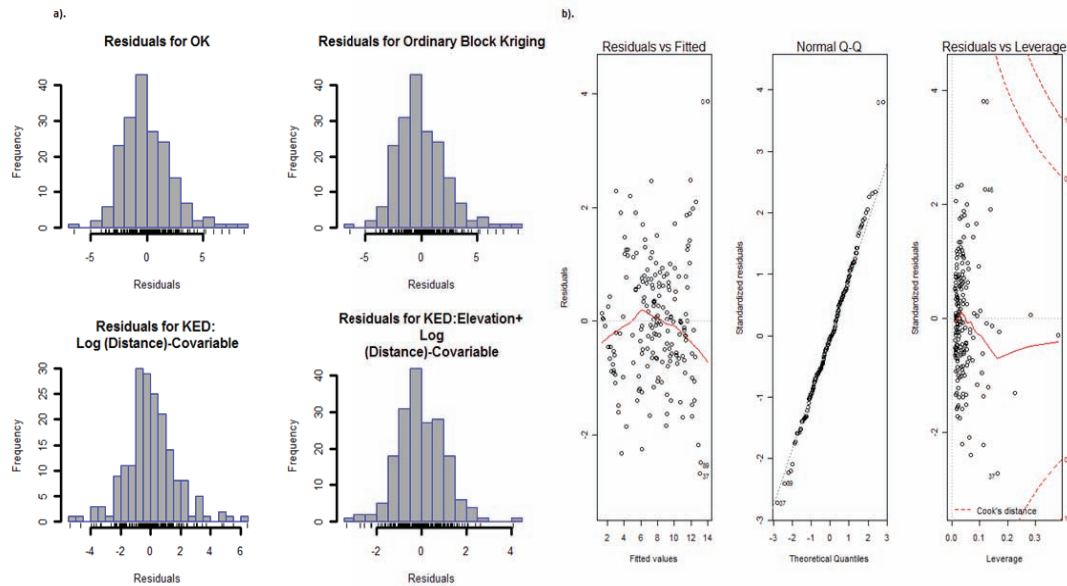


Figure 3-12: **a).** Diagnostic plots of the residuals of the kriging prediction variants and **b).** Model diagnostics of the best linear model (Elevation + log (distance)) for carbon stock.

3.11. Cross-validation Statistics and model diagnostics

Variogram models for the aforementioned three kriging variants were cross-validated to assess the validity of the fitted models and to compare estimates from the variogram model with actual values (Utset, et al., 2000). Table 3-9 gives a summarized overview of these statistics for each prediction variant. Summary statistics of the leave one out cross validation procedure (Isaaks & Srivastava, 1989) shows that the mean prediction errors approach zero, with external drift kriging giving a far superior error distribution than ordinary kriging and therefore not biased.

Table 3-9: Summary of cross-validation statistics.

PREDICTION METHOD	ME	MSE	RMSE	PRESS
Ordinary Kriging	0.00086	0.00015	2.02	903.65
Ordinary Block Kriging	0.00086	0.00015	2.02	903.65
KED: log (distance)	-0.0027	0.0013	1.70	539.19
KED: Elevation + log (dist)	-0.00057	5.4×10^{-6}	1.05	202.96

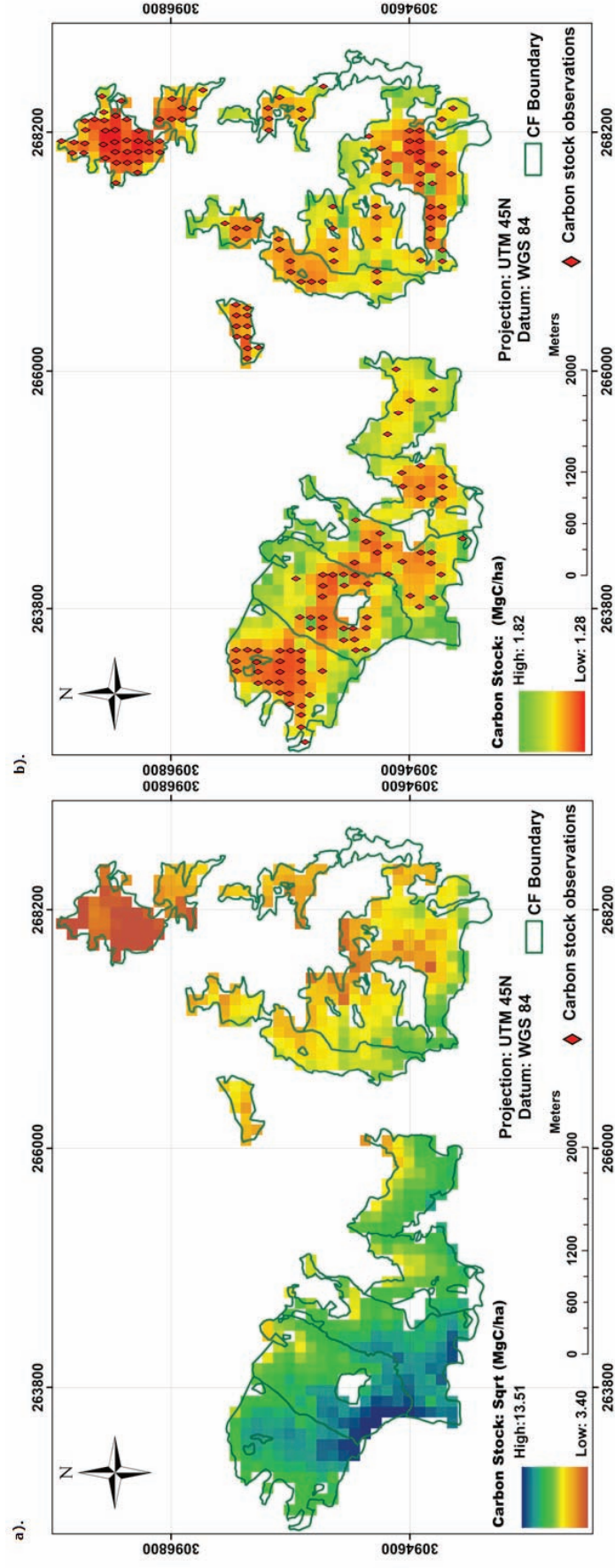


Figure 3-13: Kriging with external drift (KED) results with (elevation and log (distance)) as covariables, illustrating, a). Square root kriging predictions and b). Kriging variances of Carbon stock (Mg C ha⁻¹)

Figure 3-12a presents a comparative overview of the model diagnostics whereas Figure 3-12b presents the diagnostics for the best carbon stock predictive model chosen using Akaike Information criterion. The ideal situation is to have the diagnostics in Figure 3-12a have a more symmetrical error distribution. Kriging with an external drift using elevation and distance as covariates, ordinary kriging and ordinary block kriging gave more symmetrical error distributions (histograms) whereas the histogram for KED using log (distance) as covariate showed a symmetrical distribution but with some peakedness towards the tails.

3.11.1. Carbon stock estimates: the model-based approach

The mean standard error derived from the KED predictions of the best linear model was used to derive total C stock estimates for the study area map as illustrated in Table 3-10. It is evident from the table that higher C stock estimates are attained when secondary information is included, which in this case was supplied as the distance of the community forests from the nearest human habitations and topographic elevation. In both cases, a block of 10 000m² grid resolution was used to calculate total C stock estimates for the study area since it is a unit of area better suited for forestry management (Burges & Webster, 1980).

Table 3-10: summary of total C stock estimates for KED

Prediction method	Mean sqrt (Mg C ha ⁻¹)	Total C stock Sqrt (Mg C)
KED-log (distance)	8.14 ±0.27	5114.8
KED-(Elevation+log(dist))	9.89 ±0.17	6803.3

The best linear model of feature space using kriging with external drift gave predictions with a narrower margin of error (9.89 ± 0.17) Sqrt (MgCha⁻¹) compared to the C stock confidence intervals obtained using the design-based approach (9.91 ± 0.8) Sqrt (MgCha⁻¹). However, the design-based approach gave slightly higher total C stock estimates in comparison to the geo-statistical approach.

3.12. The design-based approach

3.12.1. Exploratory data analysis

An exploratory analysis to assess the adequacy of modelling assumptions on the design-based C stock data for the population in question (community forests) gave the results shown in the graphical plots of Figure 3-14.

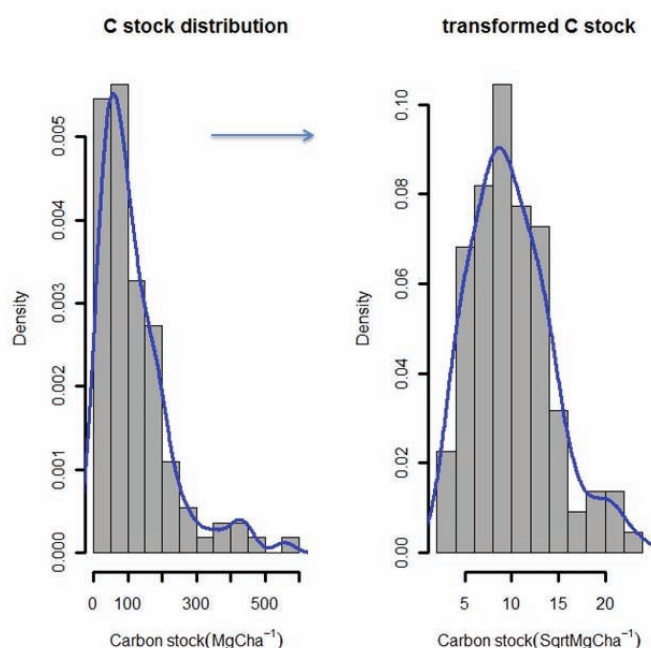


Figure 3-14: C stock distribution for the design-based approach

In general, the unprocessed raw data from the field showed a heavy positively tailed distribution; an indication of a population (community forests) dominated by young individuals. In order to dampen the influence of outliers, a square root transformation as shown in Figure 3-14 resulted in a more symmetrically shaped distribution of C stock. The same kind of a trend (positively skewed C stock distribution) was observed for the model-based approach as illustrated in Figure 3-1.

3.12.2. Test of homogeneity of variance

Transformation of carbon stock for the design-based approach was largely motivated by the need to meet the assumption of the linear analysis of variance (ANOVA (Gregoire, 1998)). Although ANOVA models are relatively robust to violations of non-normality, the normalizing transformation can result in more similar variances between groups (Longford, 2008a).

The assumption of the homogeneity of variance of the community forests was therefore validated by a significance test using Bartlett's statistic as shown in Table 3-11. The test rejected the null hypotheses of the homogeneity of variances before a transformation of the response variable was applied to the independent variable. The same test failed to reject the null hypotheses after a square root transformation of C stock. The stabilization in the variances of the community forests strata is therefore evident from the graphical plots (Figure 3-17) showing constancy of the residual variance. The plots show about the same extension of scatter of the residuals around zero for each factor level (Kutner et al., 2005) .

Table 3-11: Homogeneity of variance test for the community forests strata.

	Bartlett's k-squared	Degrees of freedom (d.f)	P-value ($\alpha = 0.05$)
Before transformation	35.53	6	3.4×10^{-6}
After transformation	10.64	6	0.069

A graphical box and whisker plot showing the variability of mean carbon stock in the different community forests shown in Figure 3-15, shows that Taksatari CF had the most density of carbon stock, 14.91 sqrt (MgCha⁻¹) and Chisapani had the least carbon stock, 8.57 sqrt (MgCha⁻¹). A single factor ANOVA model was fitted to the data and analyzed for significant differences of mean carbon stock amongst the sampled community forests.

There was a significant difference of mean carbon stock density amongst the sampled community forests under the management of different community forest user groups (CFUGs) ($F_{6, 103} = 5.21$, $P = 0.00017$). The Turkey-Kramer (Kramer, 1956) multiple comparison method was used to establish significant differences amongst the community forests. This method has the advantage of taking into account the risk of accumulating false positives as a result of conducting multiple tests simultaneously. Figure 3-16 shows confidence intervals which do not intersect the vertical reference line at zero. These comparisons were statistically significant and hence corresponded to pairs of mean carbon stock (CFs) which can be declared different by Turkey's honestly significant difference (HSD) method.

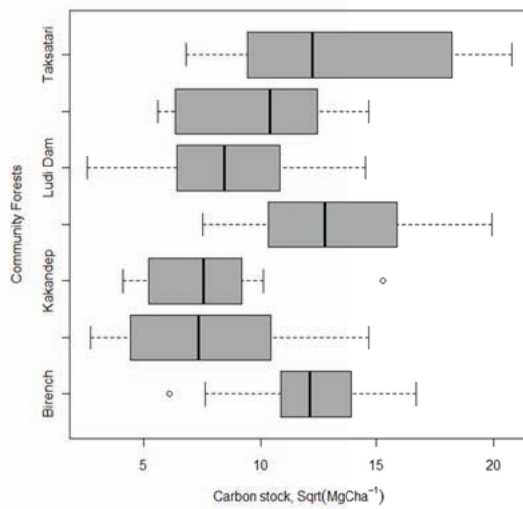


Figure 3-15: Box and whisker plot of the mean carbon stock density within each of the seven sampled community forests.

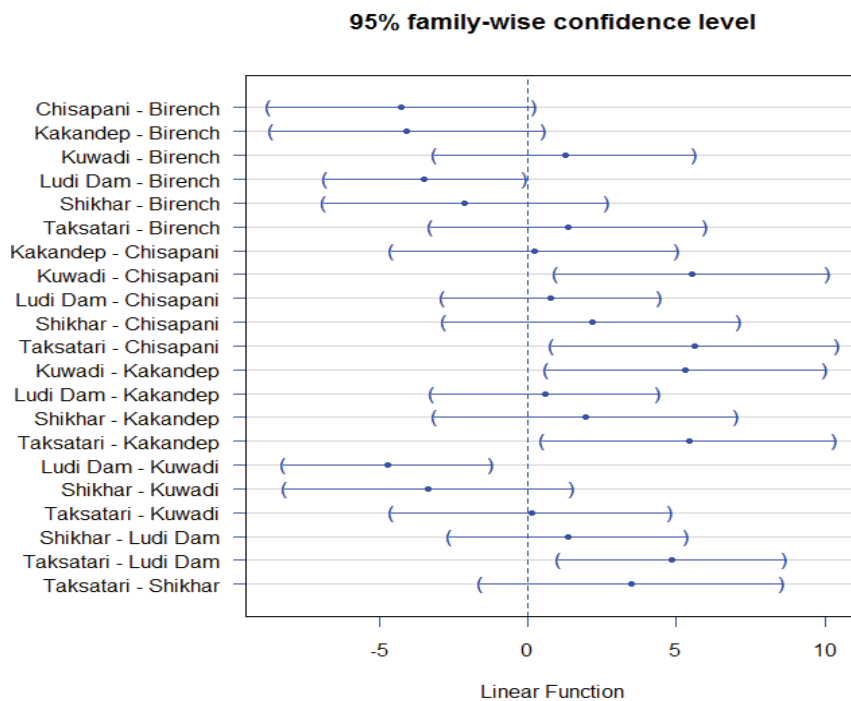


Figure 3-16: Simultaneous 95% confidence limits, Turkey method with carbon stock density as the response variable.

3.13. Carbon stock estimates: the design-based approach

A total of 15 different tree species were recorded, with *Shorea robusta* and *Schima wallichii* being predominant over the other species in the seven community forest sampled. Estimates for the

mean and total carbon stock obtained using each of the two methods is presented in the paragraphs below. C stock estimates based on management regimes in the different community forests is presented in Table 3-12. Estimates of the sampling error or the standard error of the mean C stock showed that the community forests (strata) have different C stock density. For example, there is a big difference in the standard error of Birechok and Chisapani CFs

Table 3-12: Summary of mean and total C stock for the community forests

CF	n	S.E.	Variance	Mean Sqrt(MgCha ⁻¹)	Total Sqrt (MgC)
Birechok	12	0.8	8.7	10.9 ±1.7	914.31
Chisapani	10	1.2	14.1	6.7 ± 2.3	335.6
Kakandepakha	9	1.2	11.9	6.9 ± 2.3	351.6
Kuwadi	11	1.4	14.3	11.2 ±2.2	1125
Ludi Damgade	47	0.48	11.4	7.7 ± 0.9	2007.9
Shikar	8	1.2	11.6	8.8 ± 2.1	447
Tarksatari	13	1.2	18.3	12.4 ±2.3	1107.3
Population	115	0.42	19.18	9.91±0.8	6817

Mean and total C stock estimation for the different community forests showed the highest mean C stock in Tarksatari CF and the lowest in Chisapani CF with a mean C stock of 6.7 ± 2.3 sqrt (Mg C ha⁻¹). In spite of the bigger area of Ludi Damgade (207 ha), community forests of smaller area like Tarksatari (table 2-5) gave higher mean C stock estimates.

3.14. Jackknifing

The ANOVA model was validated by using a technique called jackknifing. Two measures of accuracy of the estimator, $\hat{\theta}$, calculated from the data with the i^{th} observation removed, namely the standard error and the bias of the estimator were calculated and gave values of 0.16 MgCha⁻¹ and 0 sqrt (MgCha⁻¹) respectively with a 95% confidence interval of $\{9.74 \leq \hat{\theta} \leq 9.77\}$ sqrt (MgCha⁻¹). A zero bias is indicative of unbiasedness of the estimates since the overestimates and the underestimates are cancelling each other. A plot of the residuals against the predicted values of the ANOVA model as shown in Figure 3-17 shows that the model was relatively robust to the violation of the homogeneity of variance assumption as the residuals show equal spread across groups.

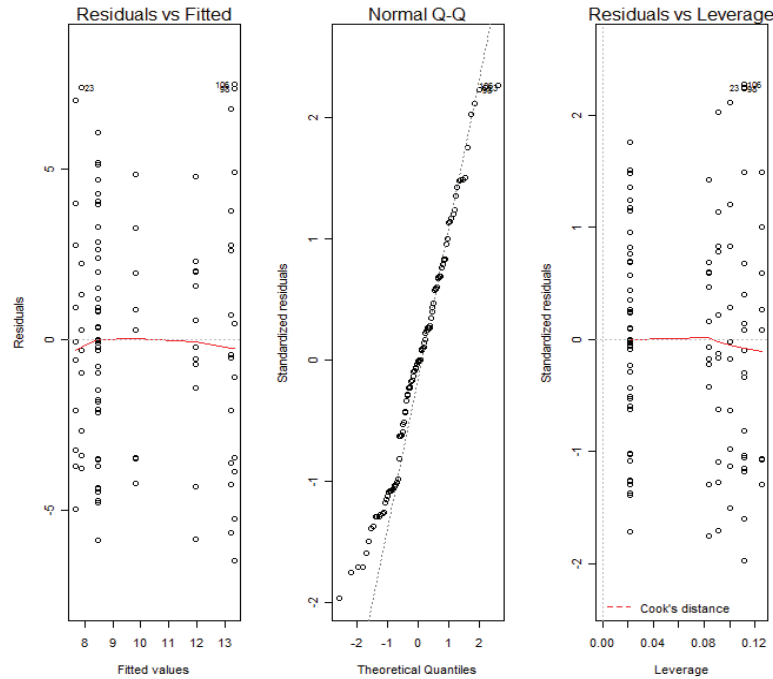


Figure 3-17: Diagnostic plots for the ANOVA model.

The difference between the estimated mean carbon stock and the true mean (bias) for the design-based approach was shown to be zero whereas a comparable statistical measure of bias (ME) for the model-based approach for ordinary kriging, ordinary block kriging, KED with log (distance) as covariable and KED with elevation and log (distance) as covariables (Table 3-9) were 0.00086, 0.00086, -0.0027, and -0.00057 respectively. This shows that when the design-based approach is compared to the geo-statistical approach using ordinary kriging and block kriging methods, the design-based approach is unbiased.

However, the power of including covariates in C stock estimation is apparent when KED was used for carbon stock predictions. There were marginal differences in terms of accuracy of kriging with external drift (MSE= 0.0013 and 5.4×10^{-6} for KED-log (distance) and KED elevation + log (distance) respectively compared to the design-based framework. Hence, covariates can significantly improve the estimation of C stock.

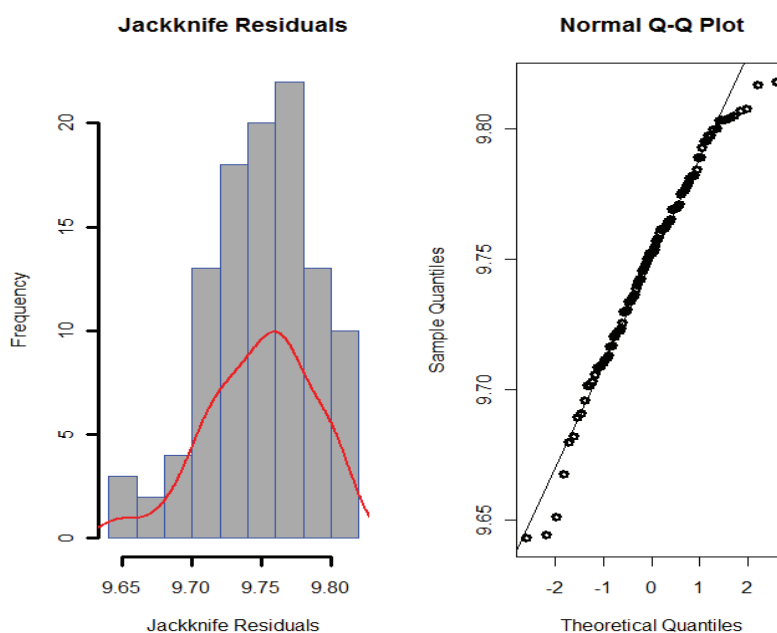


Figure 3-18: Jackknife diagnostics for the design-based approach.

In addition to basing judgments of accuracy on the calculated ME and MSE statistics (Table 3-9), graphical plots of residual diagnostics of the model provide a superior way of assessing model quality (Kohl et al., 2011). Statistical theory and regression assumptions therefore favour models with a more symmetric error distribution. Judging on this criterion alone, the model-based framework using the various kriging variants provide a far superior error distribution of residuals (Figure 3-12a. and Figure 3-12b.) compared to the design-based approach (figure 16). Furthermore, the KED with distance and elevation were most precise (Table 2-1) with RMSE of 1.05 (Mg C ha^{-1}).

3.15. Comparison of mean confidence intervals

A comparison of the confidence intervals of the estimates from the two approaches to assert whether the model-based approach (using the best feature model) is significantly better than the design-based approach was done by assessing the confidence 95% confidence bounds of the two estimators. The results of Table 3-13 shows that there is insufficient evidence for the rejection of the null hypothesis that there is no significant difference between the mean C stock estimated by the model based approach and the mean C stock estimated using the design-based approach.

Table 3-13: Comparison of confidence intervals (CI)

Estimation Method	Mean (Sqrt Mg C ha ⁻¹)	95% CI	
		Lower limit	Upper limit
Model-based approach	9.89	9.72	10.06
Design-based approach	9.91	9.11	10.71

The model-based approach gave a narrower confidence interval with a smaller margin of error than its design-based counterpart. The design-based CI contains the mean of the model-based approach and there is a 0.003 probability that the two confidence intervals will not overlap. Essentially, the confidence intervals of the two mean C stock estimators will overlap 99.7% of the time. This result is also supported by the rule of thumb that we will fail to reject the null hypothesis that there is no significance difference between the mean C stock estimated by the model-based approach and the mean C stock estimated by the design-based approach when either confidence interval includes the mean of the other (Payton, Greenstone, & Schenker, 2003). The confidence intervals for the two estimators, as shown in

Figure 3-19 are $\{9.72 \leq \bar{X}_1 \leq 10.1\}$ Sqrt (Mg C ha⁻¹) for the model-based approach and $\{9.11 \leq \bar{X}_2 \leq 10.72\}$ for the design-based approach.

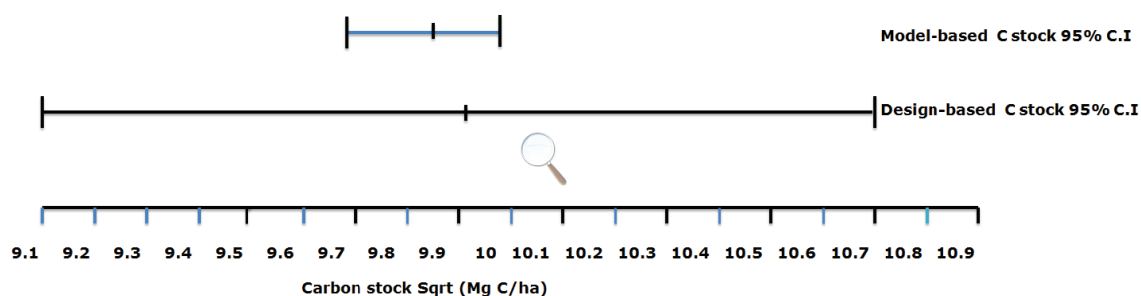


Figure 3-19: Overlapping of confidence intervals

4. DISCUSSION

4.1. The geo-statistical approach

4.1.1. What spacing distance/ interval is required to model the spatial correlation structure of C stock?

The establishment of the appropriate scales of spatial variability of the sampled resource can lead to accurate predictions. The results of the present study using the preliminary variogram (Figure 3-5) gave a range of 1541m, implying a spacing distance of around 350-400m. Subsequent modelling of C stock using the fitted variogram (spherical model) resulted in unbiased and accurate C stock predictions using the best model of feature space (Table 3-9 and Figure 3-13). However, if the scales of spatial variability are not properly captured and modelled, the chances that the predictions from the modelled variable will be accurate become remote. In the event that the initial determination of such scales underestimates the actual scales on the ground like the situation of the present study, there is no detriment since the most critical small scale variability will still be captured. Such an underestimation can however result in inefficient and costly surveys since time will be wasted in sampling a smaller unit area compared to a larger area had the right scales been used. On the other hand, an overestimate of the scales will usually prove detrimental since it will be difficult to harness the scales of spatial variability occurring at smaller distances. This can result in less accurate and imprecise predictions.

It is the first time the geo-statistical approach has found its application in this study area. With no previous knowledge of the spatial structure of C stock, the variogram model estimated from the samples of this study displayed an acceptable moderately spatially dependent C stock, with the limit of spatial dependence ranging from 762-1541m, much longer than previously estimated from the reconnaissance variogram (Figure 3-5). This was caused by the underestimation of the scales of spatial variability that were determined by the reconnaissance variogram. The sample size of the present study was considerably larger ($n=189$) than the 60 samples previously used for the reconnaissance variogram, thus resulted in a better estimate of the scale of spatial variation. 60 samples are quite small to give a reliable variogram (Webster & Oliver, 2001). On the other hand, though the variable (elevation) used for the reconnaissance variogram is strongly correlated with the resource of interest, there are other variables like distance that accounted for the variability of C stock density. Apparently, the true scale of spatial variation is coarser than expected. The sampling grid distance established by the present study is therefore approximately 350metres.

This means that future sampling protocols making use of geo-statistics would have to be done with a sampling distance around 350m. Hence, more area will be covered and sampled in a cost effective manner.

4.1.2. How effective is the model-based approach at generating C stock estimates

Isaaks and Srivastava (1989) assert that small nugget to sill ratios ($C_0/(C_0 + C_1)$), coupled with large spatial correlation ranges in the variogram are indicative of higher accuracy in the mapping of the variable. The class of spatial dependence established by carbon stock (Figure 3-8) means that the values of C stock can be predicted in the entire study area using the known sampled values. In regard of Cambardella et al., (1994) classification of spatial dependence, the successful modeling of C stock in this study using geo-statistics represents an important milestone towards the improvement of carbon accounting methods for the attainment of REDD by offering a different and most promising paradigm for the accurate estimation of C stock in the long run. This is important as it unlock opportunities and avenues that must be used for an alternative, cost effective and sustainable C stock accounting exercise.

Modelling of C stock using the spherical model (Figure 3-8) is reflective of abrupt changes of C stock density in space at irregular distances (Jongman, Ter Braak, & Van Tongeren, 1995). This fact is vindicated by the unbalanced demographic trends of the sampled plant communities (CFs) evident from the distribution of C stock in Figure 3-1 and Figure 3-14. Actually, there is a disproportionately bigger fraction of young individuals as compared to the older individuals and little of average aged individuals. This implies that sampling programmes should consider stratifying the forests ecosystems on variables like the age of the CF and level of forest disturbance to improve the accuracy and reliability of C stock accounting.

4.1.3. What is the gain in efficiency between a model-based and design-based scheme

De Gruijter (2006) prescribed three preconditions that can make the model-based approach to have an edge over the design-based approach which are herein listed:

- There must be a sizeable number of samples
- There must be evidence of spatial correlation in the modelled variable
- A greater proportion of the sample must be taken at spacing distances much smaller than the range of spatial dependence

A definition for the efficiency of alternative A relative to alternative B can be expressed as the variance of A divided by the variance of alternative B (Kohl, et al., 2011; McCune, 2010). This definition therefore entails the design effect of a sampling plan relative to another. Calculation of the design effect of the model-based approach using the different kriging prediction methods relative to the design-based approach showed some significant reduction in the standard error due to the model-based design effect. The error reduction ranged from 20% to 67% for ordinary kriging and KED respectively (Table 3-8). The reason for this tremendous error reduction can be related to the factors that influence the margin of error. A sample with a larger size (model-based approach $n=186$) will result in decreased standard error as compared to the sampling plan with a smaller size (design-based $n=115$). Thus, increasing sample size increases statistical power.

However, this comparison can still pose challenges in that the same methods that are used to determine sample sizes in the design-based framework cannot be used in the geo-statistical approach (Webster & Oliver, 2001). Instead, Webster and Oliver (2001) recommend sample sizes ranging from 100 to 200 for the model-based approach. The model-based approach is driven by spatial autocorrelations in the modelled variable and therefore the classical sample size determination approach does not hold true. The most pragmatic way to handle this dilemma will be to consider the estimation methods as equally valid, considering the theoretical bases from which they emanate from.

In a slightly different study, estimation of soil properties and the comparison of the two types of designs under discussion, McBratney and Webster (1983), established a considerably smaller kriging estimation variance than the classical estimation variance. However, a follow up study by Papritz and Webster (1994) on the comparison of the efficiency of the two approaches on the basis of McBratney and Webster (1983) found out that the geo-statistical estimate had a smaller precision than the paired stratified and paired systematic estimates.

Results of the studies revealed in the forgoing paragraphs present a paradox that can make recommendation of one of the approaches difficult. This is only a manifestation of the difficulties and complexities that can be encountered in the modelling and characterization of natural phenomena. Environmental variables are subject to temporal and structural changes which are also complicated by the lack of independence in the information they contain from one location to another. As a result, the design-based approach using CFs as blocks can earn itself an advantage over the model-based approach as it does not depend on spatial correlations.

4.1.4. Are there covariables that can be used to improve the spatial correlation structure of carbon stock? What proportion of the apparent spatial correlation structure (ordinary variogram) is removed by covariables?

Despite the popular and widespread application of NDVI and other vegetation indices as predictors of vegetation activity, many studies, including Danson and Curran (1993) and Puhr and Donoshe (2000) have reported poor performance of NDVI in predicting forest above ground biomass. Irregularities in vegetation densities and distribution as shown in Figure 2-1 can cause poor reactivity of the Near-infrared (NIR) bandwidth forest structure parameters. The rejection of NDVI as a potential predictor of C stock by a principal components analysis of the present study is therefore justified by the ongoing discourse. Variability in species composition and differences in the ages of the background vegetation, typically in successional forests (Figure 4-1), can lead to enhanced spectral variability of background vegetation in the NIR band of satellite sensors than other wavelengths (Gemmell, 1995). Notwithstanding the successes reported by Gonzalez-Alonso et al., (2006) and Zheng et al., (2004) on the performance of NDVI in mapping forest structure parameters, the results of the present study concurs with the findings of Foody et al., (2001) and Foody, Boyd and Cutler (2003) who found poor correlations of NDVI with forest above ground biomass.

Likewise, slope could not be found in the short list of predictor candidates for the modelling of C stock since there is usually a strong relationship between slope and terrain aspect in natural ecosystems (Liu, Shao, & Wang, 2011). Hence, the information carried by terrain aspect is not independent of the slope at any given location and so the variable carrying the most information at a particular time and location is likely to reflect C stock variability. On the other hand, the disturbances in the ecosystems, especially due to human and animal encroachment can alter the smooth spatial variation of these factors in space.

Three covariates, namely elevation, distance to the nearest human settlements and aspect were used in the modeling of carbon stock using kriging with external drift. A criterion –based model selection procedure resulted in an additive model of elevation and the distance to the nearest human settlements as the best model explaining the most variability of C stock (Table 3-6). As illustrated in table 3-5, log (distance) covariate accounted for 51.1% of the spatial structure in carbon stock while a linear model of elevation and log (distance) accounted for 87.3% of the spatial structure of carbon stock.

The apparent increased explanatory ability with which the additive model of elevation and distance have on the variability of C stock is illustrative of the dominance of topographic gradients which have a greater sway on biomass density and distribution (Gotway & Hartford, 1996). The auxiliary predictors explained 87.3% of the variation in the original data hence, the mapping precision is considerably higher for KED than for OK and OBK. Thus, the incorporation of more predictors resulted in much of the variability of C stock being taken out in the feature space, leaving less for spatial structure.

There is strong evidence of the interplay of climatic (temperature and precipitation) gradients with topographic gradients (elevation). Altitudinal gradients have varying levels of nutrient cycling and rainfall amounts, with faster rates of biogeochemical cycling and higher rainfall at lower elevations (high temperatures) and slower rates of nutrient cycling and lower rainfall amounts at higher altitudes, thus resulting in different primary productivities (Shijo et al., 2008). Such communities are therefore capable of supporting a wide range of soil fauna and flora due to the favourable conditions existing in the soil. On the other hand, more complex and recalcitrant compounds take several years to break and decompose in colder and lower rainfall environments.

The distance of the community forest to the nearest human settlements exert its effect by either making easily accessible forests prone to attack or forests that are difficult to access less prone to exploitation by the communities. In areas with topographic complexity like the Ludikhola watershed, forests that are easily accessible are in danger of attack by the nearest communities ((Table 2-6) and therefore such ecosystems suffer heavily from human and animal encroachments. This can result in loss of biodiversity and the decline in forests vitality. Even illegal loggers would not log in areas where it is difficult to reach and where there is no easy access to the transport network system since it can complicate the transportation of the forest resources to the nearest markets.

The question of which covariables can best explain the variability of the regionalized variable at hand cannot be answered decisively at this moment. Unless and until all the possible covariates have been measured and tested over a period of time can we establish the most influential factors behind the variability of C stock density. This means that the findings of the present study is just but a stepping stone towards the establishment of a concrete set of explanatory information for the variability of C stock in the study area.

4.1.5. What strata should be used for C stock estimation

It cannot be explicitly established that the stratification variable used (management) provided the best C stock variability and density. As such, other stratification variables like the degree of forests encroachment by the community could be used and assessed over a period of time in order to establish the most influential factors for C stock density and variability. The design-based approach analysis relied on samples collected by fellow colleagues in the same study area during the sampling programme.

An analysis of variance for the differences in mean C stock among the sampled community forests rejected the null hypothesis that there is no significant difference in mean carbon stock amongst the CFs. The rejection of this test therefore suggests that stratification due to management units has an effect in the density of C stock. Table 2-6 shows that the community forests do not have the same status in terms of management style, disturbance and trees densities. The low density in C stock in Chisapani and Ludi Damgade CF can therefore be explained by the fact that the ecosystems have suffered from degradation and are predominated by secondary forests. A survey made by ANSAB (2010) showed that C stock in Ludikhola watershed increases with decreasing elevation, with the most biomass found in low lying areas where there are heavy downpours. On the other hand, high altitude areas tend to experience drizzle precipitation, leading to slow growth patterns in the forest. Despite the interplay in the physical factors influencing C stock density, the most important factors that will determine the actual C stock at any given time is dependent on the limiting factors in the environment. Visualisation of the study design map shows that Ludi Damgade (low biomass) lies in high altitude areas (figure 2-3) which is confirmatory to the findings of this study and those by ANSAB (2010). Furthermore, as illustrated in table 2-6, the low biomass of this CFs is justified by the fact that it is predominated by successional and degraded forests.

Figure 3-6 showed 8 confidence interval comparisons which were declared statistically significant by Turkey's HSD method (Kramer, 1956). For carbon stock accounting and assessment, these CF can be used as reference benchmarks on which any improvements and forestry management initiatives can be based on.

4.1.6. What are the accuracies of the different methods in estimating C stock?

The apparent differences in the precision of predicting C stock without auxiliary variables and with explanatory information (Figure 3-9, Figure 3-11 and Figure 3-13) showed that consideration of auxiliary predictors in the modelling of a natural resource variable is worth the effort. The results of KED appear justified in terms of the known physical and presumed anthropogenic relationships which are imposed on the local mean by the elevation covariable and to a lesser extent, by the distance to the nearest human settlements. It is therefore the partitioning of the data into a deterministic trend component and a residual ‘noise’ component that vindicates non-stationary geo-statistics (Hengl, 2007). The precision and the uncertainty of predicting C stock using ordinary kriging or ordinary block kriging could not match KED. This emanates from the assumptions that govern the ordinary kriging algorithm compared to KED. A test of directional dependence for carbon stock as elucidated in Figure 3-4 showed the variable to be highly anisotropic, with different nuggets and total sills in the various directions. On the other hand, KED modelled this trend using explanatory variables in the form of elevation and log (distance) thereby smoothening the variance in the predictions (Berterretche et al., 2005).

It’s not only the incorporation of secondary information that can make KED better to other kriging variants, but it also depends on the quality of the secondary information used (Ahmed & De Marsily, 1987). A brief look at the correlations of the selected predictors with carbon stock in Figure 3-2 shows that there are plausible correlations between carbon stock and the selected covariates. With the increasing availability of topographic information at finer resolutions than before, it is easy to access information like elevation, slope and aspect from the Shuttle Radar Topography Mission (SRTM) 90m elevation online databases. In that case the application of KED outweighs the pure based approach of ordinary kriging. There were marginal differences in the total C stock predicted by block kriging compared to ordinary kriging, but the differences in the reduction of the location-specific uncertainty of the sampled C stock using block ordinary kriging is evident. Both the KED and the OK models yielded acceptable RMSE and ME values, with residuals almost meeting linear regression assumptions requirements. However, both OK and OBK performed poorly as shown by the variance maps in Figure 3-9 and Figure 3-10.

In comparison to the design-based approach, KED gave a narrower margin of error for the mean (9.89 ± 0.17) Sqrt (MgCha⁻¹) stock estimates than the margin of error reported by the design-based approach (9.91 ± 0.8) Sqrt (MgCha⁻¹). This is partly because of the lower variances and a

relatively larger sample size used for the KED algorithm. However, the design-based approach predicts slightly higher total C stock estimates than the kriging variants (Table 3-10 and Table 3-12). The design-based technique is known to overestimate since it assigns the same weighting to all the predictions and to all the residuals. On the other hand, KED has the ability of incorporating auxiliary information to further improve estimates of a primary variable (C stock) (Isaaks & Srivastava, 1989). The kriging variants, especially the KED, greatly reduced the uncertainty associated with predicting the variable in question by using predictors as its aides, which is very important in forests management. The findings of the present study are in conformity with the results of a comparative study by Guibal (1973).

From the assessment of the extent of overlap between the C stock mean estimators of the two sampling protocols, it is clear that the mean C stock estimates for both methods fall within the 95% confidence bounds of each other, figure 3-19. The estimators will overlap 99.7% of the time, implying insufficient evidence to suggest that the mean C stock estimates are significantly different from each other. The reason for this substantial extent in overlap emanates from the fact that the larger mean C stock estimate (design-based) is lower than the upper confidence limit of the lower mean (the model-based approach) (Moore & McCabe, 2002). Hence, the model-based approach is not significantly better than its design-based counterpart as previously hypothesised. However, the model-based approach looks much superior to the design-based approach as it can be noted from figure 3-19 that the estimated C stock has a narrower confidence interval and margin of error. This is very important for judging and evaluating an estimator as it shows the amount of uncertainty that we have in estimating the population parameter. In other words, there is a very small distance between the sample statistic and the population parameter for the model-based approach than for the design-based approach, a position that favours the model-based approach as an option for C stock estimation.

4.1.7. Sources of error, allometric equations and potential improvements

Estimates of total C stock derived by both methods could be improved if species specific allometric equations were used in the derivation of plot C stocks. Even though Ravindranath and Ostwald (2007) recommended the use of general allometric equations for the estimation of biomass, the dynamics of forests growth and distribution of tree species is subject to earth processes, which entails different carbon sequestration by different species at different locations and at different times. As a result, aggregating species can limit the ability to unlock and realize potential sites for optimal forestation programmes of important and appropriate tree species.

Measurement error was possible in the field, especially considering the difficulty encountered in defining DBH in plots lying in steep valleys and slopes. Instead of exact locations, some plots could be taken to the nearest point since some plots fell in dangerous valleys. Possible sources of error emanating from the selection of plots predefined on a regular grid map as a result of subjectivity in the location of some plots could all contribute to bias. Of particular concern was the bias due to tape placement, stem fluting, bryophytes and lianes.

Table 2-6 gives a summarised description of the community forests as recorded from the field observations during the sampling exercise. It is clear that the density and level of disturbance of the different CFs is not the same. Botkin and Simpson (1990) demonstrated how a lack of categorization of forests into successional stages with bare or below low biomass led to overestimation of carbon stored in the boreal forests of North America. In an attempt to minimize biases that could come from sampling easily accessible areas only, the present study accounted for the forests heterogeneity by including low to zero carbon biomass plots, thereby eliminating selective location of high biomass and tree density (Figure 4-1).

The density of C stock is affected by the stage of development and hence, it could be appropriate to apply a factor to biomass estimates for different growth stages within the CFs. Although square root transform of the response variable is not very common in forestry and carbon estimation, it proved to be an elegant way of meeting linear regression assumptions and validation of the KED linear models as adduced by the plot of the best feature space KED model residuals shown in Figure 3-12b. However, the predictions could not be back-transformed before kriging as there is no known bias free method of back transforming square root transformed response in the gstat library. For this reason, C stock estimates and their uncertainty have been given in the units of the transformed variable rather than the original units.



Figure 4-1: Successional forests and low biomass plots.

4.2. Limitations of the study

The study covered only 7 community forests out of a possible total of 31 community forests in the study area (Ludikhola watershed, Gorkha district Nepal) and thus, the results detailed in this study may not apply directly to all the CFs in the watershed.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

The major objective of this study was to find a sampling approach that gives the most accurate estimates of C stock. Additional information can be utilised to make significant improvements on the quality of interpolation. There is incontrovertible evidence that a linear additive model of elevation and the approximate distance of the community forests from human habitations carry useful information for C stock parameters.

The design-based approach generalises uncertainty of estimation for the entire study area and is therefore less useful in carbon stock monitoring and accounting since forests growth and development responds to changing earth processes. The model-based approach can aid the discovery of patching, illegal logging and forests disturbance, which is useful as an early warning signal towards the adoption of proactive monitoring systems.

In comparison to ordinary kriging, KED using the best linear model of feature space had the most reliable estimates with respect to the estimated error. Kriging with external drift had the best performance in terms of prediction in comparison to ordinary kriging and ordinary block kriging. In spite of this considerably better precision, the design-based approach still predicted higher total C stock than the model-based approach. However, the model-based approach had a much lower margin of error than the design-based approach and therefore appears attractive and more accurate than the later. There is a significant difference in the mean C stock among the sampled CFs. Out of the five covariates tested, elevation, distance to the nearest human settlements and aspect are statistically significant predictors for C stock density and variability.

There is insufficient evidence to suggest that the model-based approach is significantly better than the design-based approach as elucidated by the overlapping confidence interval bounds of the two estimates. There is a marginally small probability (0.003) that the two mean C stock estimators will overlap. However, the model-based is still superior and attractive in comparison to the design-based approach since it takes far less time and effort to collect the same number of plots that can be obtained when using the design-based approach. Furthermore, the model-based approach gave a narrower margin of error (9.89 ± 0.17) than the design-based approach (9.91 ± 0.8), implying that the former is more accurate than the later. This implies that we are closer to the estimated population parameter (total C stock for the study area) with the model-based

approach than with the design-based approach. The time and effort spent in the field for sampling using a model-based approach is far less as compared to the design-based technique since much time is spent in trying to locate the random sampling plots. For that reason, the model-based is cost effective and more practical especially considering the complexity in the topography of the present study area (Ludikhola watershed). The successful modelling of C stock in this study area using geo-statistics offers a new paradigm in the carbon estimation methods by opening avenues for better and more accurate carbon stock accounting methods.

The paradigm presented in the foregoing discourse suggests that both methods could equally be valid if applied appropriately and objectively. However, the balance of probabilities favour the model-based approach as it can essentially answer many questions, including questions that the design-based approach can answer. On the other hand, the design-based approach cannot answer questions that can be answered by the model-based approach, for example, the where and how much question as presented in figure 5-1. Carbon stock accounting allows Carbon-dense areas to be given priority in regional and local land use planning. The only way to tell the exact concentrations of stock in a spatial field is through the application of geo-statistical modelling and the model-based approach is therefore more appropriate for local interventions like in the case of detecting patchiness and as an early warning system for forest and habitat fragmentation. On the other hand, the design-based approach offers an advantage in global C stock estimation that are usually required for REDD reporting.

5.2. Recommendations

The recommendations that can be made by an assessment of the findings of this study are therefore outlined in the paragraphs below:

- For the optimization of C stock estimation, the design-based technique is appropriate since the optimization criterion is defined at the global level. On the contrary, optimization of C stock estimation using the model-based approach is primarily relevant for local estimation. For accurate estimation of the spatial correlation of carbon stock and small scale distances, clustering of the samples for a fraction of the study area is recommended.
- There is need to model the same resource of interest taking non-forests areas out of the calculation and estimation of C stock confidence intervals through an land cover classification of the study area map.

- The use of other variables like tree height derived from high resolution LIDAR technologies in the calculation of above ground biomass and the subsequent derivation of C stock.
- There is need to use other stratification variables like vegetation species, age of vegetation, soil type, terrain elevation and proximity to settlements/roads as this can go a long way in lessening the heterogeneity that is integral in field data samples. The same approach also applies for the model-based approach in that; more predictors need to be tried and tested on their power to improve C stock estimates.
- Instead of using a variable related to the management units of the community forests in as a covariable in the modelling of C stock using KED, it is worth trying the use of management blocks as a direct covariate for use with kriging with external drift.

Moreover, a thorough discussion and important recommendation approach that might be reached by this study is based on an approach adopted by Brus and de Gruijter (1997). An informed decision on the way forward regarding the selection of the appropriate statistical approach to use is therefore dependent on the nature and type of request that might be queried by the different stakeholders.

REFERENCES

- Ahmed, S., & De Marsily, G. (1987). Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. *Water Resour. Res.*, 23(9), 1717-1737.
- Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, 30, 9-14.
- ANSAB. (2010). *Forest carbon stock of community forests in three watersheds (Ludikhola, Kayarkhola & Charnawati)*. Kathmandu: ICIMOD, ANSAB, FECOFUN.
- Basuki, T. M., van Laake, P. E., Skidmore, A. K., & Hussin, Y. A. (2009). Allometric equations for estimating the above-ground biomass in tropical lowland Dipterocarp forests. *Forest Ecology and Management*, 257(8), 1684-1694.
- Berterretche, M., Hudak, A. T., Cohen, W. B., Maieresperger, T. K., Gower, S. T., & Dungan, J. (2005). Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest. *Remote Sensing of Environment*, 96(1), 49-61.
- Binkley, C. S., Brand, D., Harkin, Z., Bull, G., Ravindranath, N. H., Obersteiner, M., et al. (2002). Carbon sink by the forest sector--options and needs for implementation. *Forest Policy and Economics*, 4(1), 65-77.
- Botkin, D. B., & Simpson, L. G. (1990). BIOMASS OF THE NORTH-AMERICAN BOREAL FOREST - A STEP TOWARD ACCURATE GLOBAL MEASURES. [Article]. *Biogeochemistry*, 9(2), 161-174.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, 26, 211-252.
- Brown, S. (2002). Measuring carbon in forests: current status and future challenges. *Environmental Pollution*, 116(3), 363-372.
- Brus, D. J., & de Gruijter, J. J. (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*, 80(1-2), 1-44.
- Bryan, J., Shearman, P., Ash, J., Kirkpatrick, J. B., Hwoor, G. H., Hoodra, R., et al. (2010). Estimating rainforest biomass stocks and carbon loss from deforestation and degradation in Papua New Guinea 1972-2002: Best estimates, uncertainties and research needs. *Journal of Environmental Management*, 91(GEOBASE), 995-1001.
- Burges, T. M., & Webster, R. (1980). Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. *Soil science* 31, 333-341.
- Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., et al. (1994). Field-scale variability of soil properties in central Iowa soils. *Soil Science Society of America*, 58(1), 1501- 1511.
- Clark, D. B., & Clark, D. A. (2000). Landscape-scale variation in forest structure and biomass in a tropical rain forest. *Forest Ecology and Management*, 137(1-3), 185-198.
- Cochran, W. G. (1977). Sampling techniques (Third edition ed). New York. John Wiley and sons. *Sampling techniques* (pp. 219-222). New York: John Wiley and sons.
- Conerly, M. D., & Mansfield, E. R. (1988). An Approximate Test for Comparing Heteroscedastic Regression Models. *Journal of the American Statistical Association*, 83(403), 811-817.
- Dahal, N., & Banskota, K. (2009). Cultivating REDD in Nepal's community forestry: A discourse for capitalizing on potential? . *Journal of Forest and Livelihoods.*, 8(1), 40-51.
- Danson, F. M., & Curran, P. J. (1993). Factors affecting the remotely sensed response of coniferous forest plantations. *Remote Sensing of Environment*, 43, 55-65.
- de Gruijter, J., Brus, D. J., Bierkens, M. F. P., Knotters, M., Hardq, H., Jafroc, K., et al. (2006). *Sampling for Natural Resource Monitoring*. Springer.
- Deutsch, C., & Journel, A. (1998). *GSLIB: Geostatistical Software Library and user's guide (2nd ed.)*. New York: Oxford University Press.
- Dhakal, S., & Raut, A. K. (2010). Potential and bottlenecks of the carbon market: The case of a developing country, Nepal. *Energy Policy*, 38(7), 3781-3789.

- Dhital, N. (2009). Reducing emissions from deforestation and forest degradation (REDD) in Nepal: Exploring the possibilities. *Journal of Forest and Livelihoods*, 8(1), 55-62.
- Diggle, P. J., & Ribeiro, P. J. (2007). An overview of model-based geostatistics *Model-based Geostatistics* (pp. 27-45): Springer New York.
- Du, H., Zhou, G., Fan, W., Ge, H., Xu, X., Shi, Y., et al. (2010). Spatial heterogeneity and carbon contribution of aboveground biomass of moso bamboo by using geostatistical theory. *Plant Ecology*, 207(1), 131-139.
- Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., et al. (2000). The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System. *Science*, 290(5490), 291-296.
- Fang, S., Xue, J., & Tang, L. (2007). Biomass production and carbon sequestration potential in poplar plantations with different management patterns. *Journal of Environmental Management*, 85(3), 672-679.
- Faraway, J. J. (2002). *Practical Regression and Anova using R*. London, United Kingdom Published by CRC Press.
- Fletcher, S. R. (2011). Kyoto Protocol: Climate Change.
- Foody, G. M., Boyd, D. S., & Cutler, M. E. (2003). Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment*, 85(463-474).
- Foody, G. M., Cutler, M. E., Mcmorrow, J., Pelz, D., Tangki, H., Boyd, D. S., et al. (2001). Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Global Ecology & Bio-geography*, 10, 379-387.
- Fox, J. C., Yosi, C. K., Nimiago, P., Oavika, F., Pokana, J. N., Lavong, K., et al. (2010). Assessment of Aboveground Carbon in Primary and Selectively Harvested Tropical Forest in Papua New Guinea. *Biotropica*, 42(4), 410-419.
- Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., et al. (2006). Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison. *Journal of Climate*, 19(14), 3337-3353.
- Fuchs, H., Magdon, P., Kleinn, C., Flessa, H., Chebd, G., & Vickedy, T. (2009). Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. *Remote Sensing of Environment*, 113(3), 518-531.
- Gautam, K. H. (2006). Forestry, politicians and power--perspectives from Nepal's forest policy. *Forest Policy and Economics*, 8(2), 175-182.
- Gammel, F. M. (1995). Effects of forest cover, terrain, and scale on timber volume estimation with thematic mapper data in a Rocky Mountain site. *Remote Sensing of Environment*, 51, 291-305.
- Gibbs, K. H., Brown, S., & Niles, O. J. (2007). Monitoring and estimating tropical forest carbon stocks: Making REDD a Reality. *Environmental Research Letters*, 2(4).
- Goetz, S., Baccini, A., Laporte, N., Johns, T., Walker, W., Kelldorfer, J., et al. (2009). Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. *Carbon Balance and Management*, 4(1), 1-7.
- Goldstein, H., & Healy, M. J. R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society: Series A*, 158, 175-177.
- Gonzalez-Alonso, F., Merido-De-Miguel, S., Roldan-Zamarron, A., Garsia-Gigorro, S., Cuevas, J. M., Gadow, K., et al. (2006). Forest biomass estimation through NDVI composites. The role of remotely sensed data to assess Spanish forests as carbon sinks. *International Journal of Remote Sensing*, 27, 5409-5415.
- Gotway, C. A., & Hartford, A. H. (1996). Geostatistical Methods for Incorporating Auxiliary Information in the Prediction of Spatial Variables. *Journal of Agricultural, Biological, and Environmental Statistics*, 1(1), 17-39.
- Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28(10), 1429-1447.
- Guibal, D. (1973). L' estimation des oukoumés du Gabon *centre de Morphologie Mathématique*, 333.
- Hengl, T. (2007). A Practical Guide to Geostatistical Mapping of Environmental Variables *JRC Technical and Scientific Reports*, 120-130.
- Heo, J., & FitzHugh, T. W. (2000). A standardized radiometric normalization method for change detection using remotely sensed imagery. *Photogrammetric Engineering & Remote Sensing*, 66, 173-181.

- Houghton, R. A. (2005). Aboveground Forest Biomass and the Global Carbon Balance. *Global Change Biology*, 11(6), 945-958.
- Isaaks, E. H., & Srivastava, R. M. (1989). An introduction to applied geostatistics (pp. 561): Oxford University Press New York.
- Jaenicke, J., Englhart, S., & Siegert, F. (2011). Monitoring the effect of restoration measures in Indonesian peatlands by radar satellite imagery. *Journal of Environmental Management*, 92(3), 630-638.
- James, F. C., & McCulloch, C. E. (1990). Multivariate analysis in ecology and systematics: Panacea or Pandora's box? *Annual Review of Ecology & Systematics*, 21, 129-166.
- Jongman, R. H. G., Ter Braak, C. J. F., & Van Tongeren, O. F. R. (1995). *Data analysis in community and landscape ecology*. London: Cambridge University Press.
- Kaipainen, T., Liski, J., Pussinen, A., Karjalainen, T., Jerrinm, t., & Anderws, Y. (2004). Managing carbon sinks by changing rotation length in European forests. *Environmental Science & Policy*, 7(3), 205-219.
- Kangas, A. (2006). Model-Based Inference. In A. Kangas & M. Maltamo (Eds.), *Forest Inventory* (Vol. 10, pp. 39-52): Springer Netherlands.
- Keeling, H. C., & Phillips, O. L. (2007). The global relationship between forest productivity and biomass. *Global Ecology and Biogeography*, 16(5), 618-631.
- Keller, M., Palace, M., & Hurtt, G. (2001). Biomass estimation in the Tapajos National Forest, Brazil: Examination of sampling and allometric uncertainties. *Forest Ecology and Management*, 154(3), 371-382.
- Killeen, T. J., Schroth, G., Turner, W., Harvey, C. A., Steininger, M. K., Dragisic, C., et al. (2010). Stabilizing the agricultural frontier: Leveraging REDD with biofuels for sustainable development. *Biomass and Bioenergy*, In Press, Corrected Proof.
- Kiyono, Y., Furuya, N., Sum, T., Umemiya, C., Itoh, E., Araki, M., et al. (2010). Carbon Stock Estimation by Forest Measurement Contributing to Sustainable Forest Management in Cambodia. *Jarq-Japan Agricultural Research Quarterly*, 44(1), 81-92.
- Kohl, M., Lister, A., Scott, C., Baldauf, T., Plugge, D., & Nertyg, G. (2011). Implications of sampling design and sample size for national carbon accounting systems. *Carbon Balance and Management*, 6(1), 10.
- Kramer, C. Y. (1956). Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 12(1), 309-310.
- Kutner, M., Nachtsheim, C., Neter, J., Li, W., Hert, G. T., & Kanddegy, H. N. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill/Irwin, USA.
- Lecomte, J. (1990). Statistics for the Life Sciences: Myra L. Samuels. Dellen Publishing Co., San Francisco; Collier MacMillan Publishers, London 1989. 591 pp.; 14 chapters, numerous appendices and statistical tables, index. ISBN 0-02-405501-8. *Biochemical Systematics and Ecology*, 18(2-3), 206.
- Leemans, R., van Amstel, A., Battjes, C., Kreileman, E., Toet, S., Kwwort, R., et al. (1996). The land cover and carbon cycle consequences of large-scale utilizations of biomass as an energy source. *Global Environmental Change*, 6(4), 335-357.
- Lehtonen, A., Mäkipää, R., Heikkinen, J., Sievänen, R., Liski, J., Kevnro, H., et al. (2004). Biomass expansion factors (BEFs) for Scots pine, Norway spruce and birch according to stand age for boreal forests. *Forest Ecology and Management*, 188(1-3), 211-224.
- Liu, Z., Shao, M. a., & Wang, Y. (2011). Effect of environmental factors on regional soil organic carbon stocks across the Loess Plateau region, China. *Agriculture, Ecosystems & Environment*, In Press, Corrected Proof, 5,10.
- Longford, N. T. (2008a). ANOVA and Ordinary Regression *Studying Human Populations* (pp. 1-35): Springer New York.
- Longford, N. T. (2008b). Sampling Methods *Studying Human Populations* (pp. 67-101): Springer New York.
- Lopez-Granados, F., Jurado-Exposito, M., Atenciano, S., Garcia-Ferrer, A., De la Orden, M. S., & Garcia-Torres, L. (2002). Spatial variability of agricultural soil parameters in southern Spain. *Plant and Soil*, 246, 97-105.
- Mac Nally, R. C. (2000). Regression and model-building in conservation biology, biogeography & ecology: The distinction & reconciliation of, predictive & explanatory models. *Biodiversity & Conservation*, 9, 655-671.
- Malhi, Y. (2010). The carbon balance of tropical forest regions, 1990-2005. *Current Opinion in Environmental Sustainability*, 2(4), 237-244.

- Mansfield, E. R., & Helms, B. P. (1982). Detecting Multicollinearity. *The American Statistician*, 36(3), 158-160.
- Matheron, G. (1971). The theory of regionalized variables and its applications. *Less Cahiers du Centre de Morphologie mathématique*, 5(19), 333-347.
- McBratney, A. B., & Webster, J. T. (1983). How many observations are needed for regional estimation of soil properties? *Soil Science*, 135, 177-185.
- McCune, S. K. (2010). *Practice Makes Perfect Statistics*. US: McGraw-Hill USA.
- Miettinen, J., & Liew, S. (2009). Estimation of biomass distribution in Peninsular Malaysia and in the islands of Sumatra, Java and Borneo based on multi-resolution remote sensing land cover analysis. *Mitigation and Adaptation Strategies for Global Change*, 14(4), 357-373.
- Miranda, H., & Souto de Miranda, M. (2011). Combining Robustness with Efficiency in the Estimation of the Variogram. *Mathematical Geosciences*, 43(3), 363-377.
- Moore, D., & McCabe, G. (2002). *Introduction to the Practice of Statistics* (4th edition ed.). New York: Freeman, New York.
- Mugo, J. M., Njunge, J. T., Malimbwi, R. E., Kigomo, B. N., Mwasi, B. N., & Muchiri, M. N. (2011). Models for Predicting stem diameter from crown-diameter of open grown trees in Sondu-Nyando river catchment, Kenya. *Asial Journal of Agricultural Sciences*, 2011, 3(2), 119-126.
- Nelson, B. W., Mesquita, R., Pereira, J. L. G., de Souza, S. G. A., Batista, G. T., & Couta, L. B. (1999). Allometric regressions for improved estimate of secondary forest biomass in the Central Amazon. *Forest Ecology and Management*, 117, 149-167.
- Neyman, J. (1978). About the theory of probability and mathematical statistics. *Historia Mathematica*, 5(4), 485-486.
- Oliver, M. A., & Webster, R. (2008). *Geostatistics for environmental scientists*. GB: John Wiley & Sons Ltd.
- Panta, M., Kim, K., Joshi, C., Okocha, H., Chen, H., & Zako, M. (2008). Temporal mapping of deforestation and forest degradation in Nepal: Applications to forest conservation. *Forest Ecology and Management*, 256(9), 1587-1595.
- Papritz, A., & Webster, R. (1994). Estimating temporal change in soil monitoring, II. Sampling from simulated fields. *European Journal of Soil Science*, 46, 13-27.
- Payton, M. E., Greenstone, M. H., & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science*, 3:34, 6.
- Phachomphon, K., Dlamini, P., & Chaplot, V. (2010). Estimating carbon stocks at a regional level using soil information and easily accessible auxiliary variables. *Geoderma*, 155(3-4), 372-380.
- Puhr, C. B., & Donoghue, D. N. (2000). Remote sensing of upland conifer plantations using Landsat TM data: a case study from Galloway, south-west Scotland. *International Journal of Remote*, 21, 633-646.
- Rachina, S. (2011). *Comparison of individual tree delineation methods for carbon stock estimation using very high resolution satellite images*. University of Twente (ITC), Enschede.
- Rahman, M., & Pearson, L. M. (2009). *Estimation of the Multivariate Box-Cox Transformation Parameters* (Vol. 4).
- Ravindranath, N. H., & Ostwald, M. (2007). Methods for Estimating Above-Ground Biomass Carbon Inventory Methods Handbook for Greenhouse Gas Inventory, Carbon Mitigation and Roundwood Production Projects (Vol. 29, pp. 113-147): Springer Netherlands.
- Ribeiro, J., & Diggle, P. J. (2001). geoR: A package for geostatistical analysis: R News 1(2), 15-18.
- Richard, C. D., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random forests for classification in ecology. 88, 2783-2792. Retrieved from <http://esapubs.org/archive/ecol/E088/173/appendix-B.htm>
- Rossi, J., Govaerts, A., De Vos, B., Verbist, B., Vervoort, A., Poesen, J., et al. (2009). Spatial structures of soil organic carbon in tropical forests—A case study of Southeastern Tanzania. *CATENA*, 77(1), 19-27.
- Rossiter, D. (2004). An Introduction to Applied Geostatistics: a lecture note. International Institute for Geo-Information Science and Earth Observation (ITC)

- Sales, M. H. S. J., Carlos M. Kyriakidis, Phaedon C. Roberts, Dar A. Vidal, Edson. (2007). Improving spatial distribution estimation of forest biomass with geostatistics: A case study for Rondônia, Brazil. *Ecological Modelling*, 205(1-2), 221-230.
- Shijo, J., Reddy, C. S., Thomas, A. P., Srivastava, S. K., Srivastava, V. K., Ganh, H. T., et al. (2010). 1-Making Effective Use of Geostatistics. from <http://www.esri.com/news/arcuser/0807/geostatistics.html>
- Shijo, J., Sudhakar, R. C., Chiranjibi, P., Sudhakar, S., Drakre, H., & Kaipander, T. (2008). Distribution of plant communities along climatic and topographic gradients in Mudumalai Wildlife Sanctuary (southern India) *Journal of Tropical Forest Science*, 45, 29-41.
- Simpson, L. G., Botkin, D. B., & Nisbet, R. A. (1993). The potential aboveground carbon storage of north American forests. *Water, Air, & Soil Pollution*, 70(1), 197-205.
- UNFCCC. (1998). Kyoto Protocol to the United Nations Framework Convention on Climate Change.
- Utset, A., Lopez, T., & Diaz, M. (2000). A comparison of soil maps, kriging and a combined method for spatially prediction bulk density and field capacity of Ferralsols in the Havana-Matanaz Plain *Geoderma*, 96(1), 199-213.
- Webster, R., & Oliver, M. A. (1992). SAMPLE ADEQUATELY TO ESTIMATE VARIOGRAMS OF SOIL PROPERTIES. *Journal of Soil Science*, 43(1), 177-192.
- Webster, R., & Oliver, M. A. (2001). *Geostatistics for Environmental Scientists*. Chichester: John Wiley & Sons, Ltd.
- Wells, N. A. (1994). Statistical analysis of circular data: N.I. Fisher, 1993. Cambridge University Press, Cambridge, U.K., 277 pp. Price: \$35.00 hardcover. ISBN 0521-35018-2. *Earth-Science Reviews*, 36(3-4), 249-250.
- Wulder, M. A., Han, T., White, J. C., Sweda, T., Tsuzuki, H., Tahui, H., et al. (2007). Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization. *Remote Sensing of Environment*, 110(1), 123-137.
- Wysowski, B. (2010). *Mapping and estimation of carbon stock of roadside woody vegetation along roadways in eastern Overijssel, the Netherlands*. University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede.
- Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., et al. (2004). Estimating aboveground biomass using Landsat ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment*, 93, 402-411.

6. APPENDICES

Appendix A: Field data collection form

DATA COLLECTION FORM FOR LUDIKHOLA WATERSHED, GORKHA NEPAL											
Name of recorder.....			Date.....			Plot radius size.....12.62m.....					
Sampling Plot No.	Coordinates				Elevation	Slope (%)	Aspect				
	X Y										
Management type:											
Strata Name:											
Forest type:						Crown cover (%)					
Tree No.	Species	DBH (cm)	Height (m)	Crown diam. (m)	Remark	Tree No.	Species	DBH (cm)	Height (m)	Crown diam. (m)	Remark
1						6					
2						7					
3						8					
4						9					
5						10					

Appendix B: Slope correction table

Slope correction table
Plot size 500m²

Slope%	Radius(m)	Slope%	Radius(m)	Slope%	Radius(m)
0	12.62			71	13.97
1	12.62	36	13.01	72	14.00
2	12.62	37	13.03	73	14.04
3	12.62	38	13.05	74	14.07
4	12.62	39	13.07	75	14.10
5	12.62	40	13.09	76	14.14
6	12.63	41	13.12	77	14.17
7	12.63	42	13.14	78	14.21
8	12.64	43	13.16	79	14.24
9	12.64	44	13.19	80	14.28
10	12.65	45	13.21	81	14.31
11	12.65	46	13.24	82	14.35
12	12.66	47	13.26	83	14.38
13	12.67	48	13.29	84	14.42
14	12.68	49	13.31	85	14.45
15	12.69	50	13.34	86	14.49
16	12.70	51	13.37	87	14.52
17	12.71	52	13.39	88	14.56
18	12.72	53	13.42	89	14.60
19	12.73	54	13.45	90	14.63
20	12.74	55	13.48	91	14.67
21	12.75	56	13.51	92	14.71
22	12.77	57	13.53	93	14.74
23	12.78	58	13.56	94	14.78
24	12.79	59	13.59	95	14.82
25	12.81	60	13.62	96	14.85
26	12.82	61	13.65	97	14.89
27	12.84	62	13.68	98	14.93
28	12.86	63	13.72		

Appendix C: R Code

###: LOADING THE DATASET

```
join=readShapePoints("Export_Output1")
crds <- coordinates(join)
drift=readShapePoints("tsikid")
names(drift)
crds=coordinates(drift)
gridded(drift)=TRUE
```

##: BOX-COX TRANSFORMATION

```
g <- lm(join$carbon~elev+aspect+distance,join)
boxcox(g,plotit=T)
boxcox(g,plotit=T,lambda=seq(-2,1.5,by=0.1))
```

##: DIRECTIONAL DEPENDANCE

```
explore2=read.geodata("spatial.txt",header=T,sep="\t")
plot(explore2)
par(mfrow=c(1,1),mar=c(4,4,2,1))
vario.4 <- variog4(explore2,data=explore2$carbon,
max.dist = 3000,estimator.type = c("classical", "modulus"))
plot(vario.4, lwd=3,type="b",xlab="Distance(m)",ylab=
expression(paste("Carbon stock",plain(MgCha^-1))))
vgm1.carbon.ev <- variogram(explore$carbon.rt~1, data=explore,
alpha=c(0,45,90,135))
plot(vgm1.carbon.ev,type="b",lwd=3,col=c("blue","green","pink","red"))
```

####: MODELLING OF SPATIAL DEPENDENCE OF CARBON STOCK

```
va=variogram(join$carbon.rt~1,join,cutoff=2000,width=200)
vma=vgm(12,"Sph",1500,5)
plot(va,plot=T,model=vma)
(vmaf=fit.variogram(va,vma))
plot(va,model=vmaf,plot=T,main="Ordinary Variogram\n Carbon Stock",
lwd=4,lty=5,xlab="Lag-Distance(m)",ylab=
expression(paste("Carbon stock",plain(MgCha^-1))))
```

####: SPATIAL STRUCTURE OF THE RESIDUALS: BEST LINEAR MODEL

```
vr2=variogram(join$carbon.rt~elev+d.l,Cressie=T,locations=join,cutoff=2000,width=200)
print(plot(vr2, plot.numbers = T, main = "Residuals: Elevation+log(Distance)"))
(vrm2f <- fit.variogram(vr2, vgm(psill = 0.04, model = "Sph",
range = 1500, nugget = 0.028)))
print(plot(vr2,plot.numbers=T,model=vrm2f,main="Residuals:
\n Elevation+log(Distance)",lty=5,lwd=4,
xlab="Lag-Distance(m)",ylab=expression(paste("C stock",plain(MgCha^-1))))))
```

####: BEST MODEL: KRIGING WITH EXTERNAL DRIFT

```
kd360<- krige(join$carbon.rt~ elev+d.l, locations = join,
newdata = drift, model = vrm2f)
proj4string(kd360) <- CRS(proj4string(drift))
summary(kd360)
```

CROSS-VALIDATION STATISTICS

```
kcv.kd360 <- krige.cv(carbon.rt ~ e.l*d.l*a.l, locations=join,
model = vrm2f)
kcv.kd360res=as.data.frame(kcv.kd360)$residual
ED=hist(kcv.kd360res,breaks=20,col="darkgray",border="blue",
lwd=3,xlab="Residuals",main="Residuals for KED:Elevation+\nLog
(Distance)-Covariables")
```

##: DESIGN-BASED APPROACH (19/12/2011)

```
comstockII <- read.table("D:/NEPAL_M_14_15/Variogram_Plots/comstockII.txt",
header=TRUE, sep="\t", na.strings="NA", dec=".", strip.white=TRUE)
library(multcomp, pos=4)
library(abind, pos=4)
AnovaModel.1 <- aov(carbon ~ comff, data=comstockII)
boxplot(carbon.qrt~comff, ylab="Community Forests", xlab=
expression(paste("Carbon stock, Sqrt", (plain(MgCha^-1))))),
col="darkgray",lwd=1,horizontal=T,data=mmm)
library(bootstrap)
y=mmm$carbon.qrt
ThetaHat=function(y)
{
return(mean(y))
}
results=jackknife(y,ThetaHat)
results$jack.values
results$jack.bias
(results$jack.se)^2
print(results)
jackknife
jackreps <- c()
for (i in 1:n){
jackreps[i] <- mean(y[-i])
}
jackreps
results$jack.values
(n-1)^2*var(jackreps)/n
(results$jack.se)^2
mean(results$jack.values)
var(results$jack.values)
names(results)
hist(results$jack.bias)
```