BIOREDD+ Program

DELIVERABLE 10: MAP OF FOREST CARBON DENSITY AND UNCERTAINTY

Subcontract number: EEP-I-00-06-00013-00-GEM-00
Name: GeoEcoMap Inc.

March 2015

This document was produced for review by the United States Agency for International Development. It was prepared by GeoEcoMap Inc. for the BIOREDD+ Project, task order number AID-514-TO-11-00002.

About this Document

This document has been developed for USAID´s BIOREDD+ Program to provide a spatial information on carbon stocks in forests in BioREDD region and along the entire Pacific coastal forests of Colombia. The document summarizes the sampling methodology, the forest aboveground biomass estimation, and the spatial modeling approach for biomass estimation. The map is accompanied by a spatial uncertainty information to allow for the use of map for other REDD or regional projects for conservation and carbon emission mitigation activities.

Table of Contents

LIST OF ACRONYMS 4
EXECUTIVE SUMMARY 5
2 AIRBORNE INVENTORY APPROACH 6
2.1 Stratification 6
2.2 Lidar Sampling 8
2.3 Lidar Biomass Model 9
3. MAPPING FOREST BIOMASS CARBON STOCKS 10
3.1. Stratified Estimator 10
3.2 Spatial Estimator 11
3.3 Uncertainty Analysis 12

List of Acronyms

AFOLU	**Agriculture, Forestry and Other Land use**
AGB	**Above Ground Biomass**
ALOS	**Advanced Land Observing Satellite**
BGB	**Below Ground Biomass**
BMyF	**Bajo Mira y Frontera**
CCBA	**Carbon Communities and Biodiversity Alliance**
CELI	**Consolidation and Enhanced Livelihoods**
CONIF	**Corporación Nacional de Investigación y Fomento Forestal, National Corporation on Forestry Promotion and Research**
CSR	**Corporate Social Responsibility**
DCA	**Development Credit Authority**
DEM	**Digital elevation Model**
ESPA	**Environmental Service Purchase Agreement**
FAAN	**Fondo para la Acción Ambiental y la Niñez**
FPIC	**Free Prior Informed Consent**
GOC	**Colombian Government**
HRP	**Human Rights Program**
IDEAM	**Instituto de Hidrología Meteorología y Estudios Ambientales, Institute of Hydrology, Meteorology and Environmental Studies**
INCODER	**Instituto Colombiano de Desarrollo Rural**
IRR	**Internal Rate of Return**
LIDAR	**Laser Image Detention and Ranging**
LOI	**Letter of Intent**
MIDAS	**Más Inversión para el Desarrollo Sostenible**
MODIS	**Moderate-Resolution Imaging Spectroradiometer**
MRV	**Monitoring Registry and Verification**
NASA	**National Airspace Agency**
NPC	**Naidi Production Company**
NPV	**Net Present Value**
PALSAR	**Phased Array type L-band Synthetic Aperture Radar**
PD	**Project Document**
PES	**Payment for Environmental Services**
REDD	**Emissions Reductions from avoided Deforestation and Degradation**
SAR	**Synthetic Aperture Radar**
SMEs	**Small and medium-sized enterprises**
SPV	**Special Purpose Vehicle**
TA	**Technical Assistance**
TBD	**To be determined**
tCO2e	**Ton of carbon dioxide equivalent**
TGC	**Terra Global Capital**
USAID	**United States Agency for International Development**
US$	**United States Dollar**
VCS:	**Verified Carbon Standard**
VCU:	**Verified Carbon Unit**

Executive Summary

This document summarizes the standard operating procedures for the measurements and monitoring of carbon stocks and changes to quantify the climate benefits for the USAID BioREDD projects in the Pacific Coast of Colombia. Here, we outline the methodology for measurements to update the carbon emissions before the first verification and provide the monitoring approach, data sources, and organization for the entire crediting period. The project will monitor all required components according to the VM0006 methodology. In general, all components for calculating actual GHG benefits generated by the REDD+ project, or Net anthropogenic Emission Reductions (NERs) are included in the monitoring plan:

Monitoring of deforestation drivers, project activities and emission sources related to REDD+ project activities inside and outside of the Project Area.

Monitoring of degradation drivers, project activities and emission sources related to REDD+ project activities inside and outside of the Project Area.

Monitoring LULC class transitions in the Project Area, Leakage Area and Reference Region

A monitoring report is produced which contains all of the information above, and outlines the calculations for actual NERs generated. The tools in BioREDD MRV system will focus the measurements to update the emission factors for degraded forests and primary forests by collecting new lidar data over these areas. The monitoring plan will track down changes of forest to non-forest, non-forest to forest, forest degradation and forest recovery systematically by monitoring LULC transitions. The approach will also show how to integrate remote sensing data specific for the region (ie, Landsat, ALOS-PALSAR, Lidar, etc.) in monitoring tools along with repeated lidar flights for updates of baseline emission factors to calculate annual carbon change, within the required statistical error (10% at 95% confidence interval) over the project, reference, and leakage areas. All tools and measurements techniques and calculations are developed in accordance with VM0006 and VCS AFOLU requirements.

The BioREDD project will be using the Verified Carbon Standard VM0006 methodology for the quantification of carbon stocks and net emission reductions, as well as comprehensive analysis of satellite imagery performed by GeoEcoMap.

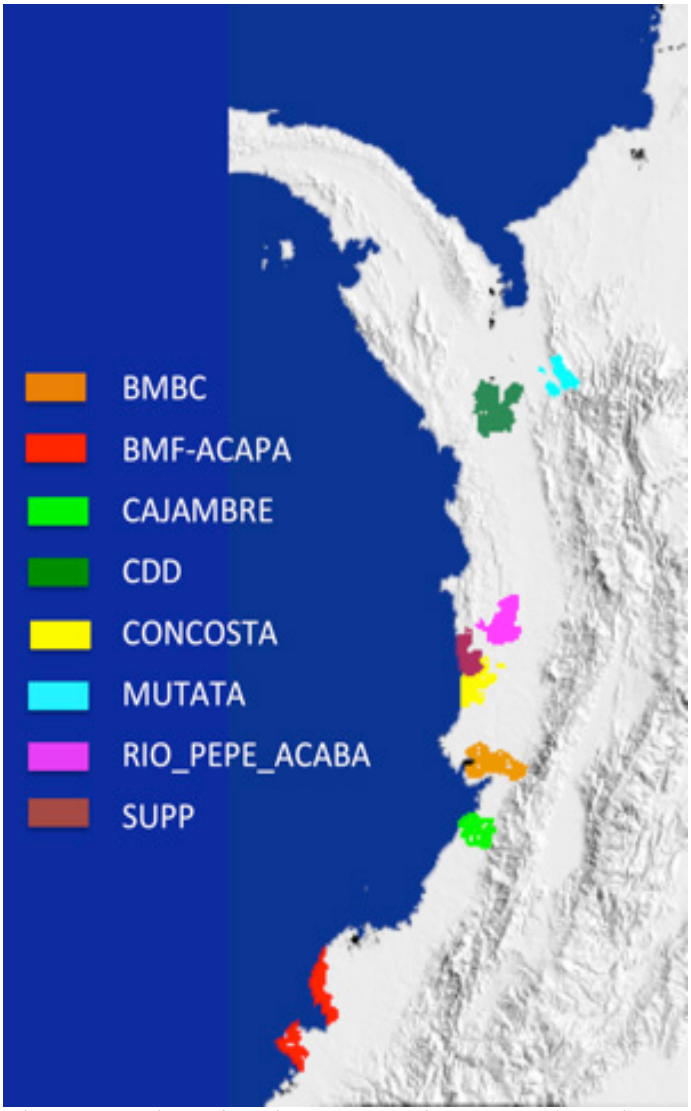

Fig. 1. Location of 8 BioREDD project areas along the Pacific Coast of Colombia.

1 Introduction & Context

The BioREDD project started in 2012 as the flagship environmental program of the United States Agency for International Development (USAID) in Colombia. This US$ 27.9 million program is designed to strengthen Colombian capacity to mitigate and adapt to climate change, protect biodiversity and support the development of remote, impoverished communities. The development of REDD+ projects is a key element of the Program, which seeks to promote sustainable livelihoods compatible with forest conservation. BIOREDD+ is developing a portfolio of 8 REDD+ projects in the Colombian Pacific to be validated under the Climate, Community and Biodiversity Alliance (CCBA) and the Verified Carbon Standard (VCS). The projects are located in four geographic nodes covering over 700,000 hectares.

Number of BioREDD Projects

The BioREDD project selected eight areas along the Pacific coastal region of Colombia for developing REDD+ projects (Figure 1). These areas are distributed in different municipalities along the pacific coastal forests of Colombia, covering a variety of ecosystems, transitioning from coastal mangroves and wetlands to paramos and high mountain forests. This area forms part of the Chocó Biogeographic corridor, one of the 10 world´s mega-diverse hotspots that accounts for more than 40% of the total vertebrate population of Colombia. The REDD+ projects will be established in the territories of afrocolombian and indigenous communities along the coast where forests have been degraded and are under the threat of further degradation and deforestation for a variety of reasons. The main drivers of forest cover change have been identified as degradation from timber extraction for fuel and development needs, illegal logging, gold mining, and conversion of forests to agriculture and livestock. The projects are being developed jointly with local communities who have clear legal title to their land and have themselves agreed to REDD+ project development. They will be implementing REDD+ project activities with the expectation of generating revenues in exchange for their conservation efforts.

2 Airborne Inventory Approach

The monitoring plan will include collection of airborne lidar data before the first verification event to update the emission factors for degraded forests and primary forests. These are the only measurements to update the baseline emission factors before the first verification to reduce the uncertainty discounts required by VM0006. Following the VM0006 requirements (9.3.2 & 9.3.9), the emission factors will be updated before each baseline updates every 10 years. All calculations of emissions and removals during the crediting period will be based on monitoring LULC transitions and using the emission factors fixed between the two baseline updates except for the intact forests that may have undergone through selective and small scale degradation that are not detected by remote sensing LULC monitoring techniques. In this case, additional measures of conservativeness will be applied for updating the emission factors for intact or primary forests before the next lidar flights. A summary of measurement methods and standard operating procedures for lidar measurements for baseline updates are provided below and described extensively in references cited here.

2.1 Stratification

The stratification of landscape has been performed using a series of spatial data sets. The goal of the stratified map was to develop a set of forest strata that have similar environmental variables and most probably homogeneous or unique forest structure and biomass. We used the following data sets to guide our stratification:

1. Climate data: A series of climate metrics were chosen to stratify the landscape based on the total amount of the rainfall or the rainfall of dry season. The climate surfaces were created from a number of databases by R. Hijmans et al. (2004) and are available from WorldClim website (http://biogeo.berkeley.edu/). These climate metrics, known also as the bioclimatic variables, We used only the rainfall of the driest quarter to include in the landscape

stratification. This rainfall metric is known to be the most important indicator of forest structure and dynamics in tropical regions (Malhi et al., 2006). Although rainfall in the Choco region reaches to above 8000 mm in some regions but they are large variability along the latitudinal and elevational gradients. We used three categories to separate the rainfall of driest quarter by dividing them into areas < 300 mm, 300-600 mm, and >600 mm rainfall.

2. Elevation Data: We also included the SRTM digital elevation data, at 100-meter resolution. The SRTM data were classified into three categories to separate the coastal regions from various landscapes with elevational gradients in inland areas. The segmented classes are 1-100 m, 101-600 m, and >600 m elevation. These general categories separate areas along the coast from the hilly areas of inlands and areas along the elevational gradients on the foothills of Andes.

3.Soil Data: We developed a soil class/landform map for the Pacific coastal region of Colombia to help with the stratification of the forested landscape. The soil class data for the entire study area have been derived from the Soil and Terrain Database for Latin America and Caribbean (SOTERLAC, version 2) released in 2005 at 1:5 million scale (Dijkshoorn, et al., 2005); the assignment of the soil class was based on matching the descriptions of the map units and comparing with the landforms and geographical description provided by Sombroek (2000). The categories were reduced to only 8 classes to focus on the most important categories for forest types and dynamics. The classes included: Heavily leached white sand soils, less infertile lowland soils (ultisols and entisols, More fertile lowland soils, alluvial deposits from the Holocene, contemporary alluvial soils including acrisols with plinthic and gleyic content, gleysols, luvisols, histosols, young, submontane soils, and other soil types with less information (probably histosols).

4 Land Cover: We used ALOS PALSAR HH and HV data acquired in 2007-2010 at 100 m spatial resolution to derive three categories of land cover type for the region. This PALSAR data is L-band radar data and is sensitive to forest type and structure, and has the additional advantage of being unaffected by cloud cover. We created 3 classes of forest, nonforest, and swamps. These three classes helped to do the first sampling design for lidar and plot inventory sampling. Note that degradation as an important forest type in the region was not classified at the time of lidar sampling. We beleived this could not introduce any bias in our sampling design as we use a random sample with the forest class that should include both intact and degraded forests in the forest class.

Stratification Approach: We performed the final stratification by combining four data layers of climate, soil, topography, and the land cover map. A program in IDL was written to perform the analysis automatically. First, we made sure all data layers have the same size and spatial resolution. Climate and soil data were resampled to 100 m

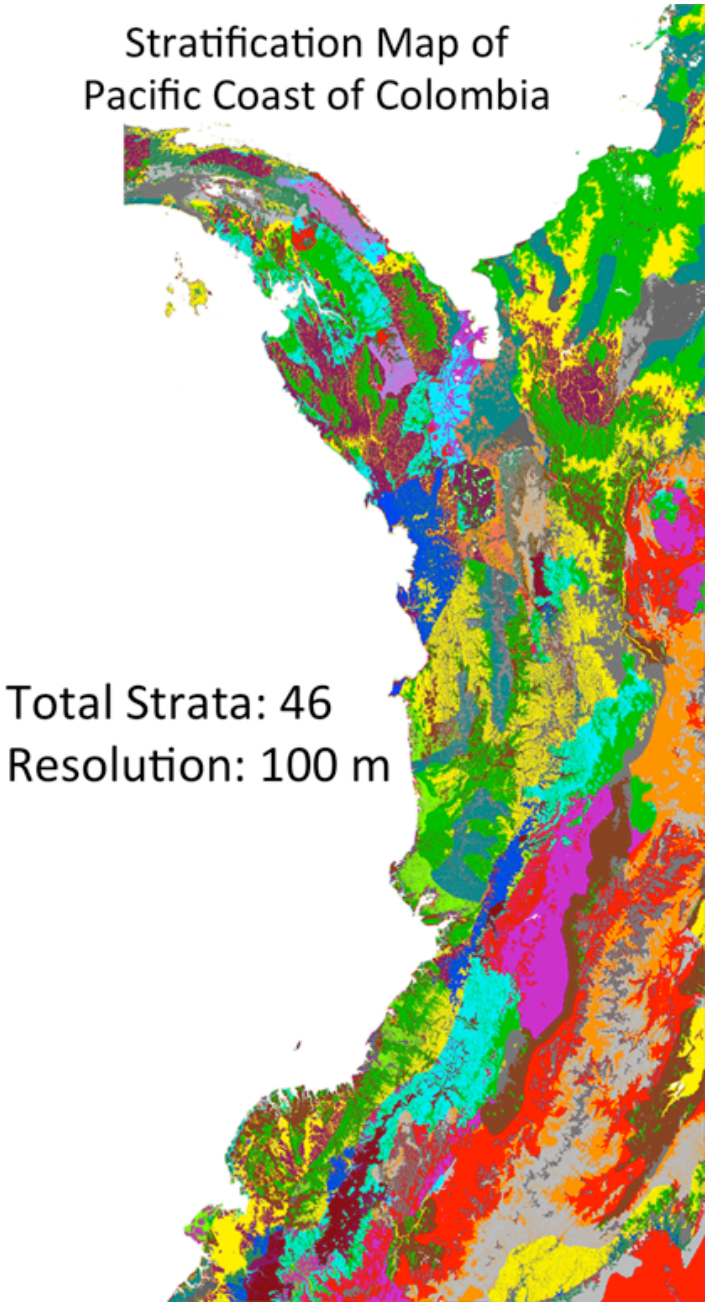

Fig. 2. Final landscape stratification map of the Pacific coastal region of Colombia.

resolution to match the SRTM and ALOS products. The statification was then performed by multiplying all data layers to create 3 x 3 x 3 x11=297 classes where among them there were only 103 distinct strata in the coastal region with only 46 classes covering more than 90% of the entire image area (Table 1). We colored the stratified image into 30 distinct colors to demonstrate the potential variations of the landscape features that may influence the forest structure, carbon stock and dyamics. The stratified map will be used to define areas for forest inventory and airborne Lidar samples. Out of the 46 strata in the region, there were 36 that fell into the BioREDD regions and were used later on to capture the biomass variations by lidar sampling.

Table 1. Data layers and partitions used in developed the stratification map of the region.

Data Layer	Properties and Class Type
SRTM	1: < 100 m, 2: 100-600 m, 3: > 600 m
Rain	1: < 300 mm, 2: 300-600 mm, 3: >600 mm
ALOS/PALSAR	1: Forest, 2: Swamp, 3: Non-forest
Soil	1:AL+CL, 2: CM, 3: AR+RG, 4: FL+GL, 5: FR, 6: PT, 7: AC + LX, : 8: NT, 9:PZ, 10: LP; 11: Water

2.2 Lidar Sampling

The variations of stratified map over areas selected for the BioREDD project along the Pacific coastal region of Colombia was used to design the potential flight lines and total flight hours to achieve the systematic sampling of the landscape. The stratified layers will be used in conjunction with other GIS layers such as roads, logistics, and access to design the final location of field plots and airborne Lidar flights over the region. This process will be achieved in person with the field crew and Lidar company selected to perform the lidar acquisitions and processing over the BioREDD project regions. After several steps of negotiations with the BioREDD team in Colombia, we designed Lidar acquisitions for the 30 strata using 49 flight lines with each flight line having 1000- 2000 ha areas for each for a total of 83,000 ha of randomly sampling areas. The process of the Lidar flight design followed the VT0005 methodology tool:

Estimation of areas of stratified landscapes within the BioREDD project nodes.

Estimation of percent area of Lidar coverage based on the number of strata and area of each stratum within each project nodes.

Developing Lidar coverage using the long transect design to increase the coverage over few strata with one flight line and optimize the flight time and acquisitions.

Flight lines were selected randomly in each strata within the project nodes using the GIS tool the Reversed Randomized Quadrant-Recursive Raster (RRQRR) algorithm that is based on the implementation of the Generalized Random Tessellation Stratified (GRTS) algorithm. The RRQRR toolbox allows for probability-based spatiality balanced sample designs to be implemented within a Geographic Information System (GIS).

After the locations were selected, the flight lines were designed in three different sizes, all with 2 km width and with 5 km, 10 km, or 20 km length to allow the required coverage in each stratum. The design will allow the airborne flights to be acquired in an optimum configuration with transects covering large areas and using less time

for turns and realigning for flight headings. The orientation or heading of the flight lines were also selected randomly at 45o intervals.

2.3 Lidar Biomass Model

In the BioREDD project, the plots are only used for calibration and validation of lidar to AGB model following the VT0005 methodology. The minimum number of plots required for developing and validating the remote sensing to AGB predictive model was estimated to be 45, with 30 plots for calibration and 15 plots for validation. This requirements were based on a similar methodology as in CDM A/R tools for Calculation of the number of sample plots for measurements within A/R CDM project activities, and based on minimum number for developing remote sensing and AGB predictive model (Asner and Mascaro, 2014). In general one predictive model is enough to estimate forest biomass from lidar in BioREDD regions if the ground allometric model used in estimating AGB at the plot level is also one. If the number of ground allometry changes in the project area because of the presence of various forest types, as in Chave et al., 2005 for wet, moist, and dry forests, then the number of remote sensing predictive models will also change.
In the BioREDD project we increased the number of calibration and validation plots in order to make sure that the numbers are enough for developing more than one lidar to AGB predictive model if necessary. A total 15 cluster plots were used in 15 lidar transects covering different strata in the BioREDD regions. Each cluster plot met the following requirements:

1. Each cluster plot had one permanent plot of 1-ha in size (100 m x 100 m) and 8 satellite or auxiliary plots at 0.25 ha in size (50 m x 50 m).
2. The center location of permanent plot was selected randomly in the lidar transect in the forest area to allow sampling all forest conditions as in fragmented, degraded or secondary forests. The satellite plots were located at 250 m and 500 m away from the center permanent plots in four cardinal directions (Fig. 6.1).
3. We also used 45 0.25 ha systematically sampled in a lidar flight line and was used to develop the model and tested over other study sites to estimate the uncertainty associated with the model and potential improvement of the model.
4. The measurements in the permanent plots included all trees with DBH ≥ 10 cm in diameter that were measured in each plot, with trees being tagged and recorded, x,y coordinates were recorded for permanent plots but not the satellite plots, and trees were identified in all plots by species for wood density values.
5. A minimum number of 50 trees in different DBH classes were selected for height measurements and development of DBH-H allometry for the plots or forest type. The number of trees with height measurement represented between 7 and 16% of the total number of trees registered in each site. All trees in San Pablo already had height measurements.
6. A tree-diameter height (H-D) relationship or predictive model was developed for each cluster plot and also combined for the region. We used a Weibull model to estimate the height of trees that were not measured directly.
7. The AGB of plot data were estimated using the regional allometric equation developed for the project. The estimated AGB values were

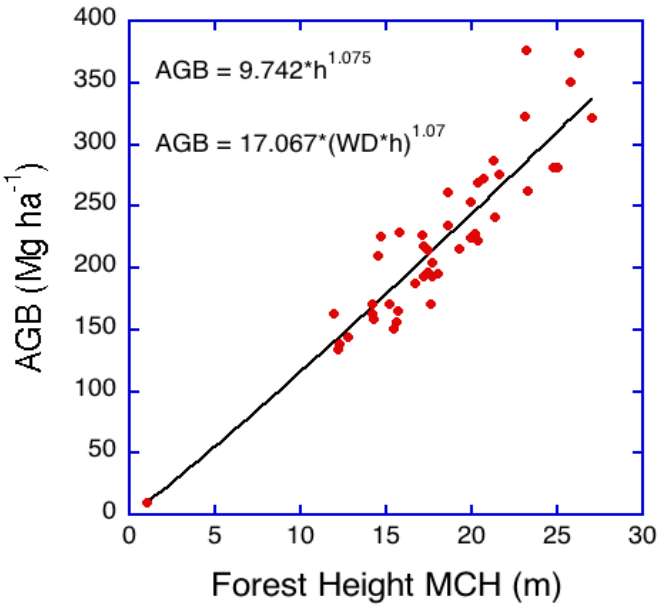

Fig. 3. Relation between lidar mean top canopy height and forest aboveground biomass derived from lidar measurements and 45 systematically designed ground plots in a degraded forest scene used for baseline

related to the lidar data extraction over the same plots.

8. The errors in plot estimates of AGB have been accounted in the overall uncertainty anaylsis and included errors from the validation process, QA/QC data, and an statistical error propagation model.
9. The project will use the same lidar biomass model developed during the baseline lidar data analysis to estimate the biomass of all forest types and revise the baseline emission factors. This model was developed using the lidar data and ground plots as discussed in GeoEcoMap report #8&9 and the uncertainty report # 16.The model has the general form of:

$$AGB = a(WD \times h)^b + \epsilon \qquad (1)$$

where WD represents the plot mean value of wood density in units of g cm^{-3}, h represents the mean top canopy height MCH in units of m, and $\epsilon \sim N(0, \sigma^2)$ represents the uncertainty in measurements or when observations deviate from a power-law by accounting for the heteroskedasticity. The model used in the baseline estimates are shown here for reference. This model is derived from the 45 sample plots at 0.25 ha with corrected geolocation errors. To implement the model for biomass estimation from lidar data use the following steps:

1. Create forest height map at the native resolution of lidar (e.g. 1-m) by creating CHM (canopy height model) from the difference between DSM (Digital Surface Model: the first return) and DTM (Digital Terrain Model: last return).
2. Use the CHM which represent the top canopy height at the lidar native resolution at 1-m to develop average mean top canopy height MCH at 50 m (0.25-ha) or 100 m (1-ha) to apply the lidar biomass models developed for the same resolution from relating the lidar metric and the ground data in the region.
3. Develop a raster forest biomass map at 0.25 ha or 1-ha spatial resolution from each lidar image and produce a spreadsheet with the location of each lidar 1-ha pixel with geolocation error and the biomass estimates.
4. In the case of mixed intact and degraded forest in the lidar coverage, separate the two by delineating areas of degraded forests using a combination of canopy cover estimate from lidar and visual interpretation of digital airborne color imagery.

3. Mapping Forest Biomass Carbon Stocks

To provide a forest biomass carbon density estimate for the BioREDD regions or the entire coastal forests of Colombia, we used two techniques to extent the lidar estimates of the biomass regionall:

3.1. Stratified Estimator

For the stratified estimator methodology we use a class of conventional design-based estimators used in national forest inventory when sampling methods and intensities differ by strata (McRoberts et al., 2002). In this approach inventory data located in each strata will be use to estimate the mean carbon density of the stratum and the total carbon by multiplying the density to the forest area. In our approach the inventory data have been replaced by biomass estimates derived from lidar samples. Using the function in equation (2) as the model to estimate forest biomass from GLAS lidar shots, the mean biomass density of stratum j is given by (1):

$$\mu_j = \frac{\frac{1}{n}\sum_{i=1}^{n} F_{ij}(\alpha_j)}{\frac{1}{n}\sum_{i=1}^{n} \eta_{ij}} \qquad (2)$$

where μ_j is the mean carbon density of stratum j, F_{ij} is the carbon density derived from individual lidar pixels located in stratum j along the lidar transect i, n is the number of lidar transects intersecting stratum j, α_j represents the vector of parameters used in the biomas model (e.g. a and b in lidar model in equation (2)), and η_{ij} is the number of lidar pixels along each lidar flight transect intersecting stratum j. The estimate of the variance of the mean carbon density for each stratum is given by:

$$V(\mu_j) = \frac{1}{\bar{\eta}_j^2} \frac{\sum_{i=1}^{n} \left(F_{ij}(\alpha_j) - \mu_j \eta_{ij} \right)^2}{n(n-1)} + \frac{1}{\bar{\eta}_j^2} \sum_{k_1}^{p_j} \sum_{k_2}^{p_j} Cov(\alpha_{k_1 j}, \alpha_{k2j}) \bar{F}'_{k_1 j} \bar{F}'_{k2j} \tag{3}$$

where

$$\bar{F}'_{k_1 j} = \sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\partial f(x_{it}, \alpha_{k_1})}{\partial \alpha_{k_1}} \tag{4}$$

representing the first derivative of the lidar biomass model as given in equation (1) with respect to the

parameters for each strata. In equation (3), $cov(\alpha_{k_1 j}, \alpha_{k_2 j})$ is the covariance of k_1 and k_2 coefficients of the lidar biomass predictive model for strata j. The first term in equation (4) represents the sampling error and the second term describes the model error for each stratum.

In this study we use a total of 36 strata with the lidar coverage and estimate the mean and variance of the aboveground biomass in each stratum using the above relations (Table 1).

3.2 Spatial Estimator

To map the forest biomass at high spatial resolution (100 m) over the BioREDD region or the entire coastal forests of Colombia, we make use of a non-parametric machine learning approach based on maximum entropy estimator (Saatchi et al., 2011). The Maximum Entropy (MaxEnt) estimator has been used for national and continental scale biomass mapping (Saatchi et al., 2011) and provide relatively similar results as other machine learning approaches with some additional advatages such as developement of an unceratinty map based on an embedded Baysian algorithm, and a spatially unbiased estimation. To implement the approach, we first divide more than 83000 ha of biomass estimates of lidar pixels (1-ha) into ranges of biomass (i.e. 0-25 Mg/ha, 25-50, 50-75, 75-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, and > 400 Mg/ha) and then we run the MaxEnt model for each given range to create the probability of predicting the biomass range for each pixel. Within the MaxEnt model, the spatial probability density functions (pdf) for each biomass range is optimally estimated using the Bayesian algorithm. We also adopt a similar Bayesian statistical approach to combine the biomass pdf values over the entire domain of the study. Following Baysian statistics, we can interpret the output from the MaxEnt model for a range A for a specific pixel (i,j) over the study domain (i.e. entire Gabon) as $Pr\left(AGBmin < AGB_{i,j} < AGBmax \middle| A\right)$: the probability of the $AGB_{i,j}$ at pixel (i,j) being inside range A (where $AGBmin$ is the lower bound of range A, and $AGBmax$ is the upper bound of range A) given condition A (here meaning that we are in the domain of the estimation of each pixel of the studying area being either inside or outside of range A, i.e. the Maximum Entropy model run for range A).

For an area of study where we divide the AGB into n ranges, we obtain a set of probability distributions $Pr(AGBmin_k < AGB_{i,j} < AGBmax_k | A_k)$ for k=1 to n. If $Pr(A_k)$ is the prior probability of having condition A_k, then the expectation value of a pixel can be calculated as

$$\boldsymbol{AGB_{i,j}} = \frac{\sum_{k=1}^{n} \Pr\left(AGBmin_k < AGB_{i,j} < AGBmax_k \middle| A_k\right)^m \Pr(A_k) AGBmean_k}{\sum_{k=1}^{n} \Pr\left(AGBmin_k < AGB_{i,j} < AGBmax_k \middle| A_k\right)^m \Pr(A_k)} \quad (5)$$

where $AGBmean_k$ is the mean AGB for range A_k , and m=3 similar to the optimum value used in Saatchi et al (2011). This creates the correct AGB distribution in the final product while keeping the prior distribution from being over-powering.

The prior probabilities Pr (A_k) are calculated from the number of lidar derived AGB values that fall into each range A_k. Ideally, a random sample of lidar AGB would give a good estimation of $\Pr(A_k)$. However, no truly random sample exists at the global scale. The lidar based AGB distribution approaches the true distribution as the area of interest increases and the number of orbits increase and become more random. We use the lidar derived AGB at the BioREDD regions or strata to find the prior probabilities Pr (A_k).

The result of the spatial modeling is provided in Figure 4 , showing detailed information about the variations of forest biomass density over Gabon at 100 m (1 ha) spatial resolution. The map shows the concentration of high forest biomass density comparable with the field inventory and lidar data in central to northern Choco region and along higher eelavations away from the coastal degradation. The southern coastal region with the highest degradation shows lower biomass values than other regions. Most degraded and low density forests are concentrated in the coastal region in Guandal forests and in areas with < 10% slopes.

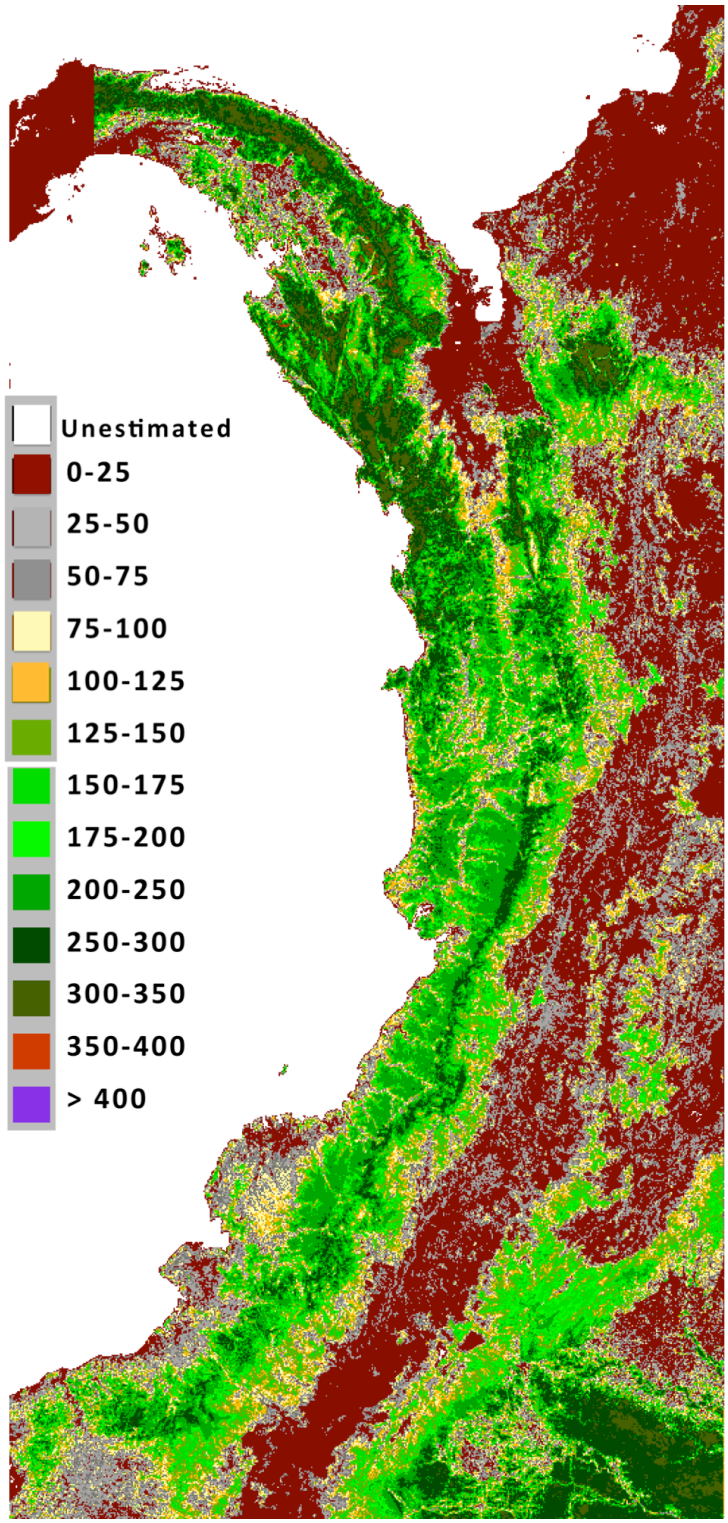

Fig. 4. Distribution of above ground forest biomass at 1-ha spatial resolution in units of Mg/ha.

3.3 Uncertainty Analysis

In estimating forest above ground biomass distribution everywhere in the BioREDD domain, each step in the entire process is evaluated for possible sources of error, and associated uncertainties are quantified. The sources of error on AGB value are, in the order of our model procedures, measurement error associated with estimation of Lidar height (very small error),

sampling error associated with representativeness of lidar height samples as the true height distribution of the strata, as well as heterogeneity of forest biomass in the 100 m pixels ($\varepsilon_{sampling}$), prediction errors from the Maximum Entropy model ($\varepsilon_{prediction}$), and allometric error when converting lidar height metrics to AGB ($\varepsilon_{allometry}$). RMSE for Lidar measurement of height is < 1 m at 1 m pixels and less than few centimeter at 1-ha where the AGB is evaluated. The lidar measurement error is much smaller (< 1%) than all other errors and for all practical purpose can be ignored in the error propagation approach. The main source of error in estimating biomass is in ground measurements of DBH, height, wood density, and location of the plots that have been quantified in our uncertainty analysis document (GeoEcoMap, Report # 14). Using the methodology developed in report # 14, we will assume the uncertainty from ground estimation of biomass can be approximately 20% (10-30%) at 1-ha scale.

Allometric errors for height to biomass equation can be estimated from the relationships in converting lidar measurements to ground estimated biomass. We have a good handle of this allometry as shown in the above section. The errors associated with ground allometry is approximately 1-5% at the 1-ha scale. We assume 5% for the sake of conservativeness.

We calculate $\varepsilon_{prediction}$ using 20% of the samples that were set aside and not used in the MaxEnt model. We

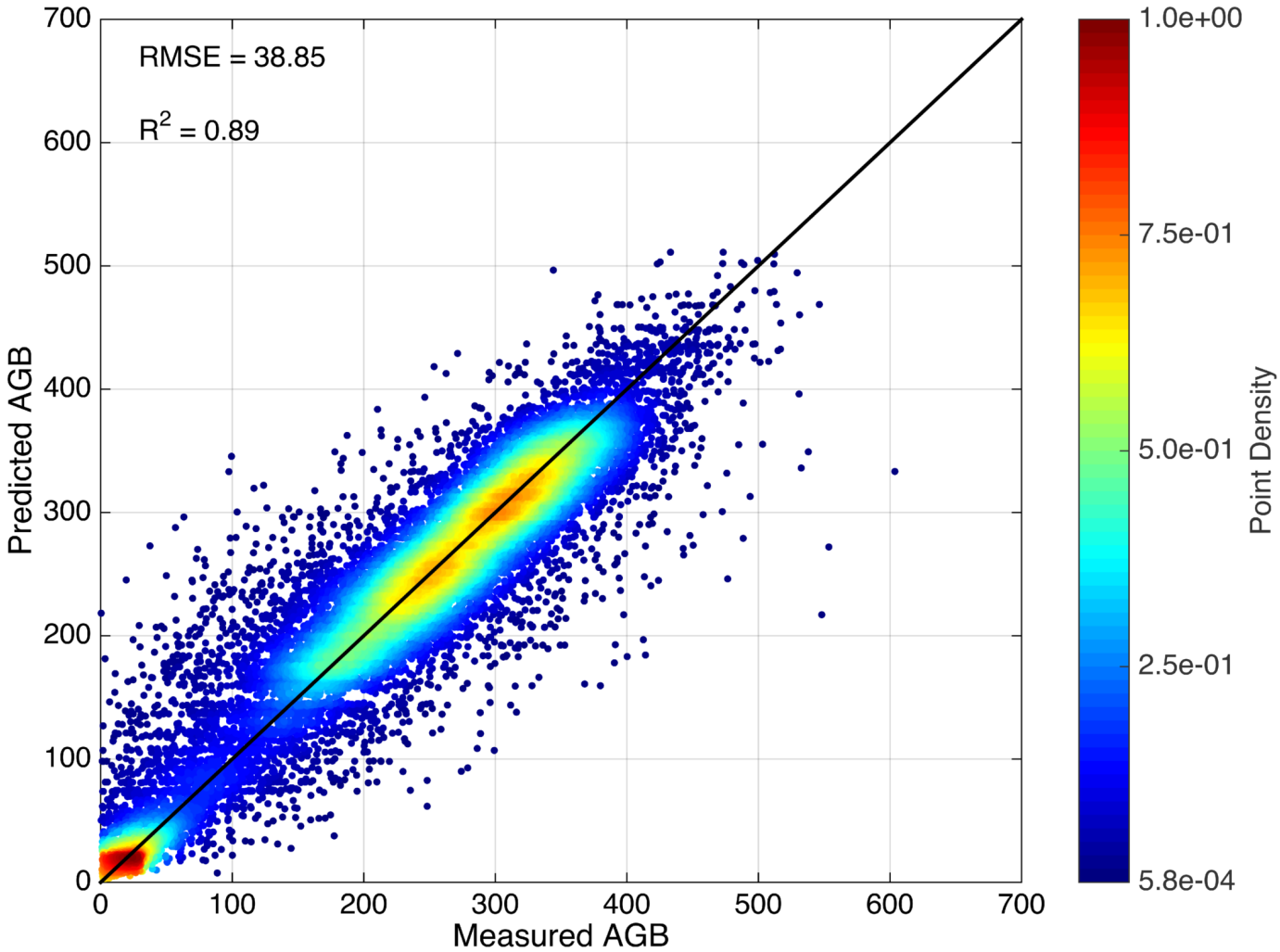

Fig. 5. Uncertainty associated with the pixel level prediction of the above ground biomass from the spatial modeling approach.

estimate the average uncertainty to be 30.2% from model prediction of AGB.

We also estimate spatial uncertainty at the pixel-level by using the predicted probabilities of the MaxEnt model in

$$\sigma_{\hat{B}} = \sqrt{\frac{\sum_{k=1}^{N}\left(B_k - \hat{B}\right)^2 P_k P(A_k)}{\sum_{k=1}^{N} P_k P(A_k)}} \tag{6}$$

where B_k is the mean biomass of the kth range, $\hat{B}$ is the predicted biomass value, P_k is the MaxEnt generated probability for biomass range k, and $P(A_k)$ is the prior probability of any pixel being in biomass range k as used in SI Equation 2. The relative uncertain for each pixel is then $\varepsilon_{prediction} = \frac{\sigma_{\hat{B}}}{\hat{B}} \times 100$. We can then calculate the total uncertainty in estimating AGB, assuming all errors were independent and random, by using

$$\varepsilon_{AGB} = \sqrt{\varepsilon^2_{measure} + \varepsilon^2_{allometry} + \varepsilon^2_{sampling} + \varepsilon^2_{pr}} \tag{7}$$

where each of the terms are the relative errors at that pixel.

The error distribution shows large uncertainty in areas where ground and lidar data were not collected in the north of Choco region bordering the Darien forests. However, in areas along the central coastal plane and away from the high elevation areas where most of the ground and lidar data collection were concentrated the map appears to have less 40 Mg/ha uncertainty on the average for each 1-ha pixel. Using the uncertainty from the map and the spatial correlation of errors, we arrive at very small error at the level of strata or LULC for the BioREDD regions.

To further examine the results of the spatial modeling, we will compare the mean biomass values for each stratum for the entire region from both the stratification estimation and the spatial estimation. In theory, both approaches must provide the same mean values. The results of comparison for the entire stratum region for the mean over all LULC and only the forest sector are provided in Fig. 7.

The comparison between the two methodology will ensure that the sampling strategy and spatial

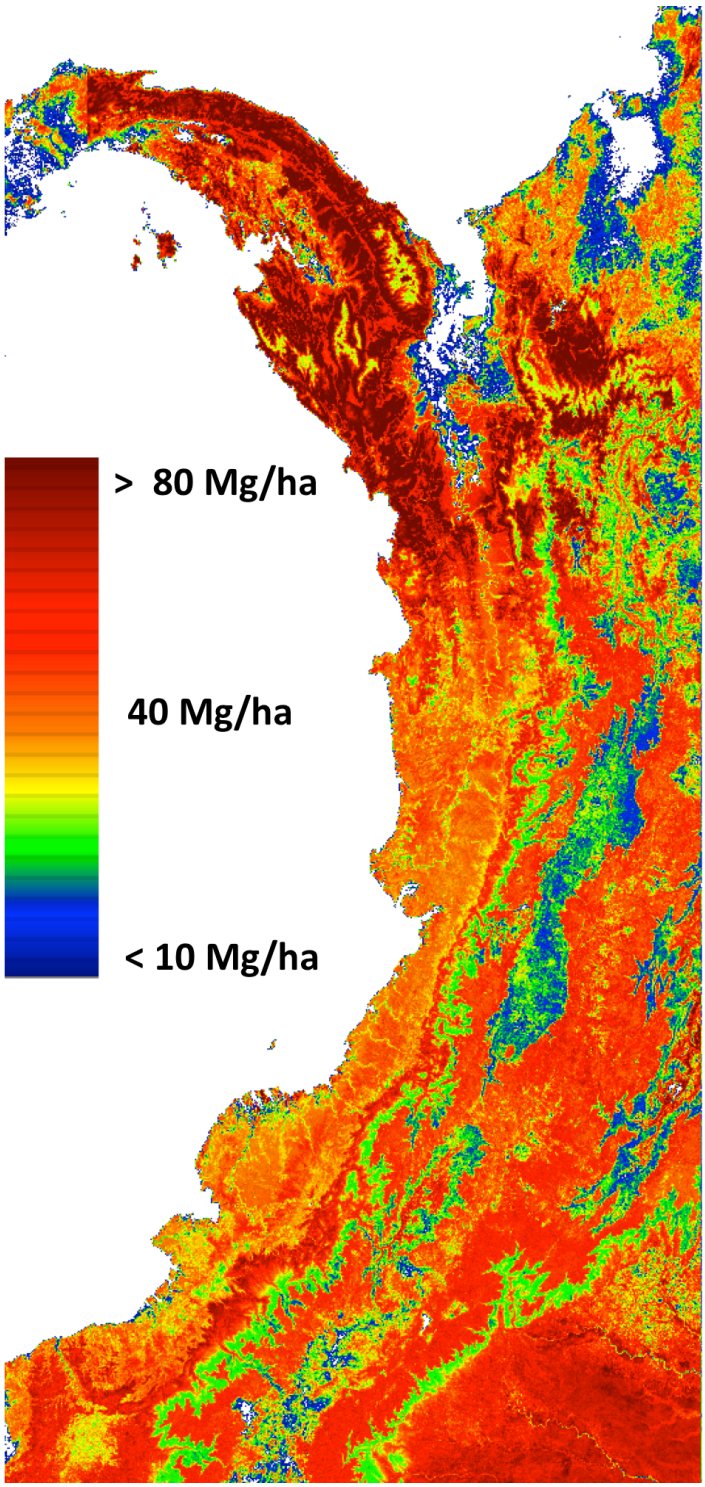

Fig. 6. Distribution of forest biomass uncertainty in Mg/ha estimated at 1-ha spatial resolution.

modeling approaches are consistent. Results shows that if the mean AGB is calculated for each strata regardless of the LULC classes, there is a strong bias particulaly in areas where there are some level of degradation and fragmentation in the stratum. This is mainly due to the fact that our original stratification did not have a degraded class types, we were able to separate this class and all other fragmentations after performing the LULC classes.

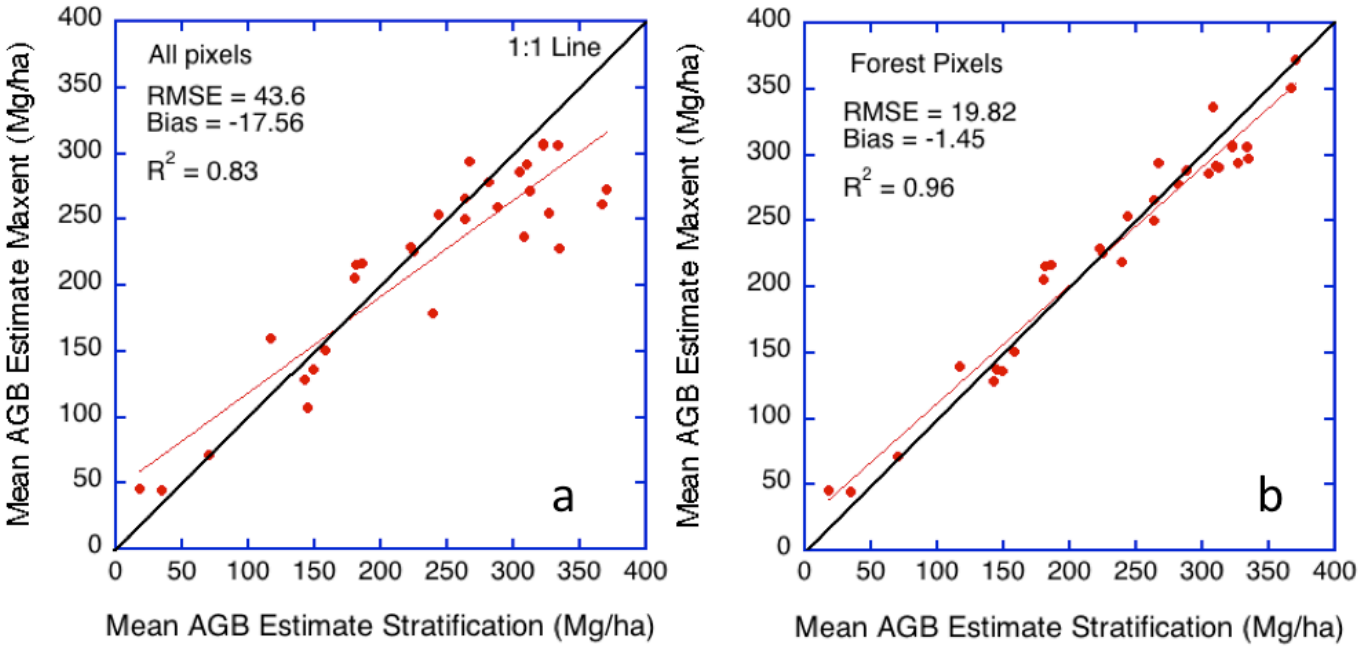

Fig. 7. Comparison of spatial modeling results and lidar sampling stratification approach in estimating the mean forest biomass for each strata within the BioREDD domain with a) showing the comparison when all pixels within the stratum were used in calculating the mean from the maxent spatial map, and b) showing when the average were performed only on the forest pixels identified by the LULC map.

Therefore, the lidar samples, although random only covered the forest classes in each stratum. Our original sampling strategy also included much large area of coverage that were reduced because of limited flight permissions particularly over areas with less security. The lidar sampling also concentrated mainly in the BioREDD project regions and strata with no REDD projects were excluded. To demonstrate that the bias can be minimum and the uncertainty in mean biomass estimation will be reduced, we only calculated the mean for the forest class only by using the LULC map as a mask. The results shown in Fig. 7, clearly demonstrates that first the two methodology converge in estimation of mean over the strata and the bias reduces signfiicantly, suggesting that the maxent mapping approach can provide unbiased estimate of the mean biomass over each forest strata.

Reference

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., ... & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, *108*(24), 9899-9904.