Comments for:

Methodology for Estimating Reductions of GHG Emissions from Frontier Deforestation
Amazonas Sustainable Foundation

Submitted by:

Kyle Holland
GHG Verification Forester
Scientific Certification Systems
145 Park Place
Point Richmond, CA 94801
Email: kholland [at] scscertified.com
Office: 510.236.9015
Fax: 510.236.8598

General Remark for Comments 1, 2 and 3:

The baseline model for deforestation is of paramount importance because it essentially determines the bulk of VCUs (carbon stocks minus projected deforestation). As stated in the methodology, the linear extrapolation and modeling approaches ("b" and "c", respectively) will grossly overestimate deforestation and therefore VCUs. Because of the financial implications of the deforestation model, it is extremely important to operate using reasonable assumptions and appropriate data. It is also extremely important to use the correct statistical methods for fitting and selecting the model, as these can have a profound effect on prediction.

Comment 1:

Section 4.1.2.1, Approach "b": Linear extrapolation.

We suggest eliminating or modifying this approach using better assumptions for the following reasons:

As presented in the methodology, linear extrapolation assumes a linear and unbounded relationship between the rate of deforestation and time,

$$A_{RR,t} = a + bt.$$

This is equivalent to writing

$$\frac{\partial A_t}{\partial t} = a + bt$$

where A_t is deforestation at time t. Its integral with respect to time is

$$= \int_{-\infty}^{+\infty} a + bt \, \partial t$$

$$= A_t = t\left(a + \frac{1}{2}bt\right)$$

which is quadratic and unbounded. However, deforestation is bounded because for any finite area of land, there can only be complete forestation or deforestation (a ratio from zero to one in-between) at the extremes. Hence, it is a poor assumption that deforestation rate and time are linear. Because this is an important assumption for reasons stated earlier, linear extrapolation as described in the methodology is inappropriate.

An alternative is to formulate the relationship in the logistic domain. The relationship between deforestation and time could be represented using a logit,

$$A_t = \frac{D}{1 + e^{-t}}$$

where D is the total potential land that could be deforested (i.e. the entire project area). This function is bounded by the interval $A_t \in [0, D]$ which is a good assumption. If one wanted to work with differentials, then

$$\frac{\partial A_t}{\partial t} = \frac{De^{-t}}{1 + 2e^{-t} + e^{-2t}}$$

$$= \frac{D}{1 + e^{-t}}\left(1 - \frac{D}{1 + e^{-t}}\right)$$

Logistic regression could be used to fit this model from a time series of sample data. In either form, the functional relationship between deforestation and time could be improved by applying some autoregressive parameters to the exponential term t or parameters such as population density, gross product, exports or agricultural prices. An example of this might be

$$A_t = \frac{D}{1 + e^{-y}}$$

in the ordinary form where y is a linear combination such that

$$y = \alpha^T\mathbf{T} + \beta^T\mathbf{X}.$$

The definition of α is a vector of autoregressive parameters of length $t - s$ lags, β is a vector of covariates, $\mathbf{T}$ is a lag matrix and $\mathbf{X}$ is a design matrix.

For additional information see *Statistical Models* by Davison (2003), *Statistical Models: Theory and Practice* by Freeman (2009) and *Time Series: Data Analysis and Theory* by Brillinger (2001).

Comment 2:

Section 4.1.2.1, Approach “c”: Modeling.

See remarks for the first comment.

Comment 3:

Section 4.1.2.1, Approach “a”: Modeling.

See remarks for the first comment.

Comment 4:

Section 4.1.2.1.

All statistical models must include statistical validation to demonstrate creditability. This point should be made overly clear in the methodology, for all approaches.

Comment 5:

Section 4.1.2.1.

Correlation analysis using a multiple correlation matrix is not always a good method for performing variable selection. This is the case because data are random variables which may result from differing distributions. The correlation matrix is misleading when variance stabilizing transformations (if they exist) are not used appropriately or variance is not correctly calculated.

Comment 6:

Section 4.1.2.1.

Mallow’s criterion prediction contains strong assumptions about the distribution of the data. It has been shown that the assumption of linearity between deforestation rate and time (or any other covariate) is definitely inappropriate. Therefore, Mallow’s criterion prediction might also be inappropriate. The assumptions of Mallow’s criterion prediction should be carefully checked before interpreting results.

Comment 7:

Section 4.1.2.1.

The methodology developer should also include Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) for variable selection. A model selection exercise using the supremum of AIC or BIC over the set of all possible models ensures that the most descriptive covariates are being selected, the model is not overly fit to the data and is robust with respect to the underlying distribution of the data.

Comment 8:

Appendix 3: Methods to Estimate Carbon Stocks.

The calculations for determining sample size are for sampling with replacement. This should be stated in the methodology.

Comment 9:

Appendix 3: Methods to Estimate Carbon Stocks.

Separate transects (to measure lying dead wood) from plots (to measure standing wood). This is will improve cost efficiency, statistical accuracy and precision. Instead, estimate the number of transects using existing transect data, if available. Systematically allocate transects to plots with a frequency equal to the proportion of the estimated number of transects to plots.

Separate soil samples (to measure soil carbon) from plots (to measure standing wood). This is will improve cost efficiency, statistical accuracy and precision. Instead, estimate the number of soil samples using existing soil sample data, if available. Systematically allocate soil samples to plots with a frequency equal to the proportion of the estimated number of soil samples to plots.

If transects are not separated from plots, then the methodology should specify that lying dead wood, soil carbon and plot-based measurements should added at the plot level, rather than the LU/LC class level. This is necessary to correctly calculate standard deviation and determine the number of plots using the specified equations.

Comment 10:

No measures of uncertainty are provided, except on page 86 (Box 2?) which is not referred to in the methodology text. Uncertainty should be measured using acceptable statistical techniques and a confidence deduction should be applied to the net carbon stocks based on the level of calculated uncertainty. Uncertainty should include the following sources of error: the estimation of carbon stocks, the validation of thematic maps and the prediction of the deforestation model.

Prediction error of the deforestation model will most likely increase over time. This is a reasonable assumption because the future is inherently uncertain, and the distant future even more so. Therefore, the prediction error of the deforestation model should be scaled with time.