Methods in Ecology and Evolution

Methods in Ecology and Evolution 2012 doi: 10.1111/j.2041-210x.2012.00266.x

Error propagation in biomass estimation in tropical forests

Quentin Molto[1,2]*, Vivien Rossi[2] and Lilian Blanc[2]

[1]*Université des Antilles et de la Guyane, UMR 'Ecologie des Forêts de Guyane' BP 709, 97 387 Kourou Cedex, France; and* [2]*CIRAD, UMR 'Ecologie des Forêts de Guyane', 97 379 Kourou Cedex, France*

Summary

1. Reliable above-ground biomass (AGB) estimates are required for studies of carbon fluxes and stocks. However, there is a huge lack of knowledge concerning the precision of AGB estimates and the sources of this uncertainty. At the tree level, the tree height is predicted using the tree diameter at breast height (DBH) and a height sub-model. The wood-specific gravity (WSG) is predicted with taxonomic information and a WSG sub-model. The tree mass is predicted using the predicted height, the predicted WSG and the biomass sub-model.

2. Our models were inferred with Bayesian methods and the uncertainty propagated with a Monte Carlo scheme. The uncertainties in the predictions of tree height, tree WSG and tree mass were neglected sequentially to quantify their contributions to the uncertainty in AGB. The study was conducted in French Guiana where long-term research on forest ecosystems provided an outstanding data collection on tree height, tree dynamics, tree mass and species WSG.

3. We found that the uncertainty in the AGB estimates was found to derive primarily from the biomass sub-model. The models used to predict the tree heights and WSG contributed negligible uncertainty to the final estimate.

4. Considering our results, a poor knowledge of WSG and the height–diameter relationship does not increase the uncertainty in AGB estimates. However, it could lead to bias. Therefore, models and databases should be used with care.

5. This study provides a methodological framework that can be broadly used by foresters and plant ecologist. It provides the accurate confidence intervals associated with forest AGB estimates made from inventory data. When estimating region-scale AGB values (through spatial interpolation, spatial modelling or satellite signal treatment), the uncertainty of the forest AGB value in the reference forest plots has to be taken in account. We believe that in the light of the Reducing Emissions from Deforestation and Degradation debate, our method is a crucial step in monitoring carbon stocks and their spatio-temporal evolution.

Key-words: Bayesian framework, modelling, REDD, uncertainty propagation

Introduction

Tropical forests are a large planetary carbon stock, with 40% of the Earth's total carbon stored in the terrestrial vegetation [from 158 to 324 Pg (Gibbs *et al.* 2007)]. Tropical forests are also a dynamic stock for carbon through land-use change (emissions $1{\cdot}3 \pm 0{\cdot}7$ Pg year^{-1}) and regrowth ($1{\cdot}6 \pm 0{\cdot}5$ Pg year^{-1}) (Pan *et al.* 2011). Preservation of this major carbon stock and the important role that forest ecosystems play in mitigating climate change are now fully recognized through the elaboration of mechanisms like Reducing Emissions from Deforestation and Degradation (REDD). An important challenge facing ecologists and foresters is to quantify as precisely as possible the carbon stocks and their fluxes at different spatial scales (Baker *et al.* 2010). However, considerable uncertainty about these figures remains.

Considerable efforts have recently been made to develop new tools to monitor tropical forest carbon stocks using an aerial approach (Goetz *et al.* 2009), together with new models of carbon estimation from tree measurements (Chave *et al.* 2005). Although uncertainty has been studied at the world scale (Pan *et al.* 2011; Saatchi *et al.* 2011), the uncertainty associated with tropical inventory data has rarely been explored (but see Chave *et al.* 2005). The above-ground biomass (AGB) of inventoried forest plots is estimated with tree-level models applied to forest inventory data. Exploring the uncertainty of AGB estimates made from tree data is essential for two reasons. First, uncertainty data are required to compare the spatial and temporal distributions of AGB. Second, the data obtained with aerial techniques must be calibrated against some reference inventory plots (Daniel *et al.* 2010; Dubois-Fernandez *et al.* 2010). Therefore, it is necessary to

*Correspondence author. E-mail: quentin.molto@ecofog.gf

have unbiased AGB estimates for these plots and to quantify the uncertainties of these estimates. The AGB values in these reference inventory plots may have broad uncertainties. It is important to take this uncertainty in account when building spatial inference models. The 2000 IPCC (Intergovernmental Panel on Climate Change) report entitled 'Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories' (IPCC 2000) points out the necessity of explicitly propagating uncertainties. In the study, we provide a method to estimate AGB from inventory data while propagating uncertainty.

Inventory data consist of the diameters at breast height (DBHs) and some taxonomic information for all the trees in a precisely known area. Sometimes, tree heights are measured in the inventoried area or in part of it. Predicting the mass of a tree requires a physical model that roughly approximates the tree to a cone, which in turn requires that its DBH, height and wood-specific gravity (WSG) are known. The gap between the inventory data and the physical model is overcome by using three sub-models. The height sub-model predicts the height of a tree from its DBH. The WSG sub-model predicts the WSG of a tree from taxonomic information. The AGB sub-model predicts the mass of each tree using its DBH, height and WSG (Fig. 1). The usual practice is to apply the three sub-models sequentially: the deterministic outputs of the height sub-model and the WSG sub-model are used as the inputs for the AGB sub-model.

These sub-models have already been examined (Chave *et al.* 2004; Feldpausch *et al.* 2011; Flores & Coomes 2011) in attempts to determine the best ones to apply in different cases. Chave *et al.* (2005) explored different sources of uncertainty in AGB estimates. However, past studies have focused on the sub-models themselves, assuming that the AGB model could be significantly improved with the improvement of its components. This assumption has yet to be studied by quantifying the parts of the AGB uncertainty that can be attributed to each sub-model.

We stress that the height sub-model, the WSG sub-model and the AGB sub-model (sometimes called 'AGB allometry') must be integrated as the components of a unique AGB model. In the present study, we inferred the sub-model parameters with Bayesian numerical methods. The uncertainty contributed by the sub-models was propagated until the final AGB estimate was achieved.

The method is illustrated with data collected at Paracou, a long established set of permanent plots in a moist tropical forest in French Guiana. A full set of data is available (WSG, tree girth and height measurements) for this site and has already been used to estimate the carbon stocks and their fluxes in a natural forest (Rutishauser *et al.* 2010) and a logged forest (Blanc *et al.* 2009). This site was selected to provide feedback on the performances of P-band Synthetic Aperture Radar when it was used to measure the biomass and canopy height of a tropical forest with high-biomass stocks (Daniel *et al.* 2010; Dubois-Fernandez *et al.* 2010).

The objectives of this study were (1) to propose a generic method that identifies the uncertainties associated with AGB estimates based on error propagation; (2) to apply this method to the French Guiana data; (3) to make practical recommendations to guide foresters and ecologists in producing the most precise AGB estimates.

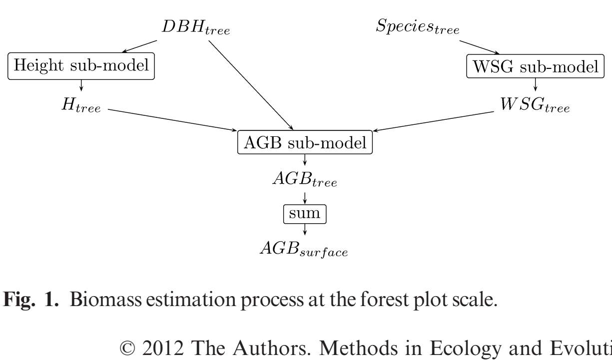

Fig. 1. Biomass estimation process at the forest plot scale.

Materials and methods

DATA

All the data were collected in French Guiana. The climate of the region is equatorial, with two main seasons: a dry season from August to mid-November and a rainy season (often interrupted by a short drier period) from December to April (Gourlet-Fleury, Guehl & Laroussinie 2004).

Census data

The census data were for a 6·25 ha plot (plot 11) at the Paracou site in French Guiana (5°11′80″N, 52°12′30″W) (Gourlet-Fleury, Guehl & Laroussinie 2004). The site was established to study the responses to different logging intensities. In each plot, the diameter of trees (DBH) was measured at breast height (1·3 m) and above the buttresses if necessary. All stems with DBH > 10 cm were mapped, tagged and measured biannually. In the present study, the biomass estimation method was applied to the unlogged inventory plots. We knew the diameter and some taxonomic information for each of the 3992 trees. The DBHs ranged from 10 to 105 cm. Taxonomic information was available to the species level for 76% of the trees.

WSG data

The data were collected during the Bridge Project (Baraloto *et al.* 2010; Sarmiento *et al.* 2011). WSG is defined here as the mass (in grams) of an oven-dried sample divided by its green volume (in cm^3) divided by the density of water (in grams per cm^3) (Chave *et al.* 2006; Williamson & Wiemann 2010). The sample cores were 6 mm long, and therefore contained sapwood only. We do not assume these WSG measures to represent the WSG of the whole tree volume; we use them as statistical predictors of the tree AGB. The specific gravity of the sapwood had been measured for 2504 trees in French Guiana, representing 466 species in 201 genera in 56 families. The number of measurements made for each species ranged from 1 (107 species) to 50 (*Lecythis persistens*); 100 species have eight or more measures.

Height data

We used data collected at the Paracou site. Height and DBH were measured in 1603 trees. DBH ranged from 10 to 172 cm and height ranged from 4 to 47 m.

Mass data

The data were collected in 1972 during the ECEREX Project (Lescure *et al.* 1983). Three hundred and sixty-one trees were cut down on a 1 ha plot in French Guyana. The trees had been selected to include a range of diameters and heights. The mass, DBH and height of each tree were measured along with some taxonomic information. The DBHs ranged from 5 to 118 cm, the heights from 2·7 to 47 m and the masses from 3·6 kg to 25 500 kg. The biggest trees were not actually weighed but their masses were extrapolated from trunk, branches and crown samples, and precise volumes (Lescure *et al.* 1983). The taxonomic information was known up to the species level (260 trees).

SUB-MODEL DEFINITIONS

Height sub-model

The role of the height sub-model is to predict the height of the trees from their diameters (Fig. 1). The height model is only used for biomass predictions. Measuring the heights of harvested trees presents no problem, although measuring standing trees does. Height–diameter relationships have long been studied for temperate forest species (Huang, Titus & Wiens 1992). Because of the huge species richness, a species-specific height model would require huge, unavailable data sets to be inferred. Very few global height models have been published (Brown, Gillepsie & Lugo 1989; Feldpausch *et al.* 2011). After considering a Weibull model, Feldspausch *et al.* recommended the log–log model. After comparing four different model shapes (see Supporting Information), we chose the Michaelis–Menten one: $\log(H_i) = \log(\alpha \cdot \mathrm{DBH}_i/(\beta + \mathrm{DBH}_i)) + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$,where ε represents the error of the model, assumed to be normally distributed (Eqn. 1).

WSG sub-model (Fig. 2)

The role of the WSG model is to provide the WSG distribution for each tree, based on its species (Fig. 1). The WSG prediction is usually a deterministic call in a database. This is not appropriate in our study because our goal was to account for the variance in AGB. Our WSG sub-model explicitly accounts for the WSG variance within species. We assumed that for each species, the WSG has a truncated normal distribution. The distribution is truncated for physical realism so the wood density cannot be lower than 0·15 or higher than 1·3. The precision of the normal distribution follows a gamma distribution $G(r, s)$. At this level, the species were weighted by their numbers of measures. Species with a higher number of measures, and therefore a more reliable variance estimate, were given greater weight. This hierarchical structure based on the precision parameter allows meaningful estimates for species with very few measures.

In our predictions, the species of some trees was unknown or the species was not recorded in the WSG data set. Such trees were given a WSG distribution compiled from a mixture of the WSG distributions of all the known trees of the forest plot.

Biomass sub-model

The response variable was the fresh mass of the tree (kg). Different biomass models have been published and compared (Brown 1997; Araujo, Higuchi & Junior 1999; Chave *et al.* 2005). These models are generally built upon the mechanistic mass of a cylinder with the height, diameter and WSG of the tree (Fig. 1), and some statistical corrections. The equation is log-transformed to achieve linearity. Some authors have

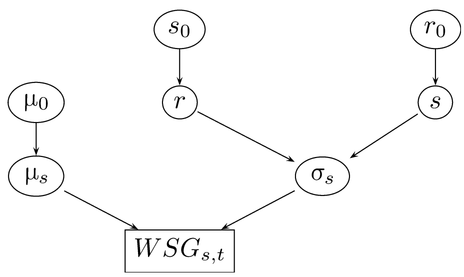

Fig. 2. Directed acyclic graph of the wood-specific gravity sub-model. $\mathrm{WSG}_{s,t}$ is the wood-specific gravity of tree t of species s; μ_s is the mean parameter of species s; σ_s is the variance parameter of species s. At a higher level, r and s are the parameters of the gamma distribution of all the σ_s values. μ_0, s_0, and r_0 are low-informative priors of their respective parameters.

chosen not to use these three predictive variables because tree height and WSG are difficult to obtain at the inventory scale (Ketterings *et al.* 2001; Chave *et al.* 2005; Pilli, Anfodillo & Carrer 2006). We preferred to include them all and deal explicitly with missing data. We have given each variable its own coefficient to allow for any deviation from the mechanistic model: $\log(\mathrm{AGB}_i) = \beta_0 + \beta_1\log(\mathrm{DBH}_i) + \beta_2\log(H_i) + \beta_3\log(\mathrm{WSG}_i) + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$,where ε is the error term of the model, assumed to be normally distributed (Eqn. 2).

When inferring parameters β_0, β_1 , β_2, β_3, and σ^2, WSG_i is predicted from taxonomic information and is therefore uncertain. DBH_i, H_i, and the tree species are directly obtained from the harvested trees.

When predicting the biomass of standing trees, both WSG and height were predicted with their corresponding sub-models. DBH was measured and the species identified on the living trees.

SUB-MODELS INFERENCE

Bayesian methods

In the Bayesian paradigm, the parameter inference consists of updating the prior knowledge on the parameter from the data, generating the posterior distribution of the parameter. The parameter posterior distribution of the three models was inferred with numerical Bayesian methods (Monte Carlo Markov chain algorithms) (Gelman *et al.* 2003). These methods produce a sequence of values for each parameter, called a 'chain', which converges to the parameter posterior distribution. Once algorithm convergence is achieved, it is also said that the chain has reached its 'stationary state', and the values of the chain constitute a sample of the parameter posterior distribution.

The first iterations are discarded to ensure that the chain has reached the stationary state. Because the values of the chain could be correlated, we used a thinning procedure to produce a 1000 quasi-independent samples from the posterior distribution of each parameter.

Height sub-model

The parameters were estimated with Metropolis and Gibbs' algorithm (Gelman *et al.* 2003).

WSG sub-model

Two parameters were estimated for each species. The mean parameter μ_s has an improper prior. The variance parameter σ_s^2 follows a $\mathrm{IG}(r,s)$

distribution, and r and s have low-informative priors $N_{[0\ +\ \infty]}(0, 10^6)$. The posterior distributions of the parameters were estimated with the Metropolis–Hasting algorithm. The truncature introduced for physical reasons on the WSG distribution $N_{[0\cdot15,1\cdot3]}(\mu_s, \sigma_s^2)$ also helps the algorithm convergence by forcing him to explore physically realist values. Data were introduced sequentially, starting with the species with the highest weights (more than 10 measures). This helped the chains to reach their stationary states by avoiding some local minima. After 5000 iterations, the remaining data were introduced. Finally, for each parameter, we retained one value each for 70 of the last 70 000 iterations to reduce the autocorrelation between them. After this thinning, we obtained 1000 quasi-independent samples of the posterior distributions of the parameters (Fig. 3). The algorithm details are given in Appendix S1.

Biomass sub-model

The parameters of the AGB model were estimated with Gibbs' algorithm, using noninformative priors. For each tree in the mass data set, we generated 1000 samples from the WSG posterior distributions of the WSG model parameters. At each step of Gibbs' algorithm, a sample from the WSG distribution for each tree was randomly chosen. See Appendix S3 for details of the algorithm.

BIOMASS PREDICTION

The biomass of a tree was predicted with a Monte Carlo scheme from the posterior distribution of the model parameters. For each tree, we generated:

1000 samples from its WSG distribution, using the WSG sub-model with parameters associated with the tree species. If the tree species is not known, the parameters are sampled from a mixture of the WSG distributions of all the trees in the forest plot. The truncature on [0·15, 1·3] of the WSG distribution ensure that the sampled values are realistic.

Thousand samples from its height distribution, using the Height sub-model and its DBH.

Thousand samples from its AGB distribution, using the AGB sub-model its DBH and each sample from its WSG distribution, and its height distribution.

This protocol was applied to the 3992 trees in the forest plot. Samples from the plot biomass distribution were obtained by summing the mass of each tree. This mass was converted into tonnes per hectare, the usual units. We can calculate any information from the predictive distribution of the plot biomass, such as the mean, quantiles or credibility interval (Table 3).

QUANTIFICATION OF THE ERROR SOURCES

The prediction of the AGB of the forest plot was based on three sub-models: the height sub-model, the WSG sub-model and the AGB sub-model. These three sub-models were sources of uncertainty in the estimation of the plot AGB. To quantify the uncertainty each sub-model contributed to the AGB estimates, each was considered as deterministic, one at a time.

To consider a model deterministic, its parameters were replaced by their expected values computed from their chains as the mean of 1000 values. We then examined the changes in the biomass prediction distributions, for both the trees and the forest plot (Table 3, Fig. 4).

When a model is considered deterministic, the uncertainty it brings to the biomass estimate disappears. This causes the AGB uncertainty to decrease. We used this decrease to quantify the part of the AGB uncertainty that could be ascribed to that model.

Results

MODEL PARAMETER VALUES

Height sub-model

The height (Table 1) model was calibrated on a local data set and thus cannot be used for other tropical sites. Because many authors have found site and environmental effects in tree

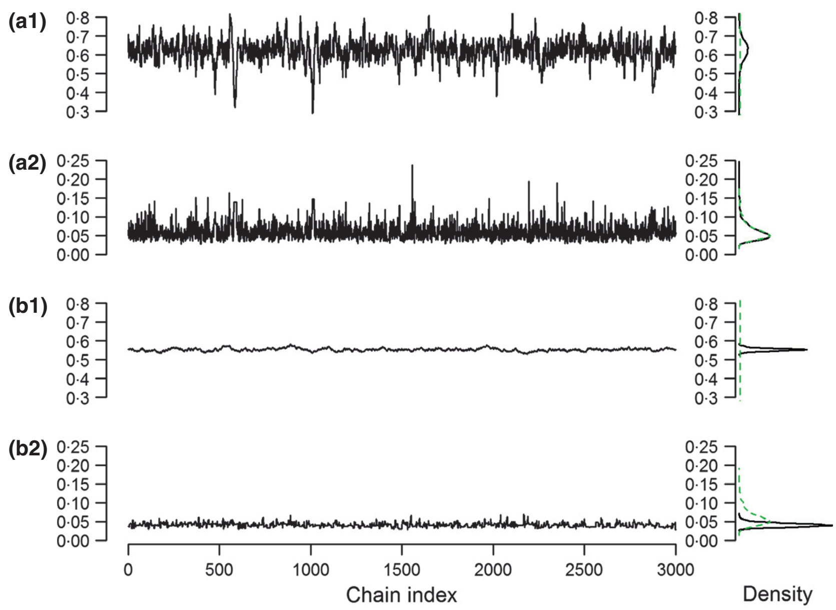

Fig. 3. Posterior distributions of the WSG model parameters for two species, plus the WSG distribution simulated from these parameters. a: *Andira inermis* (one measure only); b: *Carapa procera* (27 measures): 1, mean parameter ; 2, variance parameter . Dashed line: prior distribution.

height modelling (Feldpausch *et al.* 2011; Lines *et al.* 2012), we believe that it is always better to use local models.

WSG sub-model

The posterior distributions for μ_s (mean parameter) and σ_s (variance parameter) were sampled for the 463 species of the data set (Fig. 3). If the number of observations is low (e.g. *Andira inermis*, Fig. 3), the variance parameter distribution is the same as the hierarchical prior distribution. If the number of observations is high (e.g. *Andira inermis*, Fig. 3), the model allows deviations from the prior. As expected, the predicted WSG of each species was centred on the mean of the observed values.

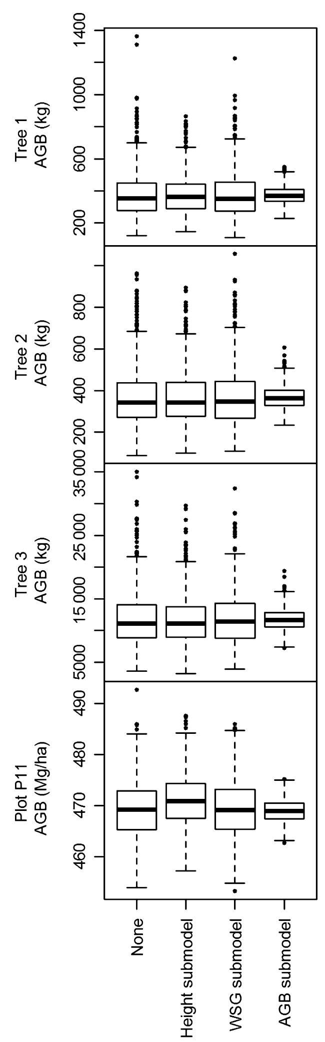

Fig. 4. Boxplot of the AGB posterior distribution for three trees (described in Table 3) and for the 6·25 ha Paracou forest plot P11.

Table 1. Posterior distributions of the height model parameters $\log(H_i) = \log(\alpha \cdot \mathrm{DBH}_i/\beta + \mathrm{DBH}_i) + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$.

Parameter	Median	CI
α	40·3	38·8, 42·1
β	9·43	9·77, 9·68
$\tau = 1/\sigma^2$	27·6	24·6, 30·6

AGB sub-model

The posterior values of the AGB model coefficients are given in Table 2. Under our conditions, the WSG signal was weak (coefficient value: 0·198). Previous studies have found WSG to be a key predictor of AGB at the tree level or have chosen to fix the coefficient value to 1 for physical reasons (Chave *et al.* 2005). This difference is certainly attributable to the fact that WSG uncertainty was taken into account in inferring the AGB sub-model parameters.

The parameter posterior values (Table 2) were very close to Chave's values (Chave *et al.* 2005), especially the error parameter (Chave *et al.* 2005: RSE = 0.302; present model: $1/\sqrt{\tau} = 0.323$).

AGB value

Using the previous models, we predicted the AGB for individual trees and for the forest plot (Table 3). The final biomass estimate for the forest plot was 451 Mg ha^{-1}, with a credibility interval of [441, 461]. This estimate is slightly higher than previous estimates [421 Mg ha^{-1} in 2007; (Chave *et al.* 2005; Goetz *et al.* 2009)]. We believe our estimate to be more accurate because we used local models only, rather than pan-tropical equations.

It should be noted that the AGB of a surface is the sum of the AGBs of the trees on that surface. Therefore, the variance in the AGB of the surface is the sum of the variances of the trees. As the surface increases, the variance increases as an absolute value in Mg, but decreases as a relative value in Mg ha^{-1}.

Identification of error sources

The AGB of the P11 plot at Paracou, with all error sources taken in account, was 451 Mg ha^{-1}. The Bayesian 95% credibility interval around this estimate had a range of 20 Mg ha^{-1} (±2·2%). This range was compared with the range of the credibility intervals of the AGB estimates when one sub-model was considered deterministic.

When the height model was considered deterministic, the credibility interval range was similar: 19 Mg ha^{-1}. Therefore,

Table 2. Posterior distributions of the biomass model parameters $\log(\mathrm{AGB}_i) = \beta_0 + \beta_1\log(\mathrm{DBH}_i) + \beta_2\log(H_i) + \beta_3\log(\mathrm{WSG}_i) + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$.

Parameter	Median	CI
β_0	−2·91	−3·12, −2·69
β_1	2·19	2·07, 2·31
β_2	0·756	0·604, 0·92
β_3	0·187	−0·0149, 0·381
$\tau = 1/\sigma^2$	9·79	8·31, 11·3

Table 3. Estimates of AGB and descriptors for three trees (AGB in kg) and for the entire plot P11 (AGB in Mg ha^{-1}). Each estimator is presented with its median and its 95% Bayesian credibility interval in brackets.

	Species	DBH	Height	WSG	AGB (kg for single trees, Mg ha^{-1} for the plot)			
					All error sources	Height model deterministic	WSG model deterministic	AGB model deterministic
Tree 1	Eperua falcata	20·2	18·9 [11·3, 26·2]	0·637 [0·545, 0·728]	334 [160, 660]	337 [182, 624]	339 [156, 662]	359 [253, 450]
Tree 2	unknown	20·0	18·9 [11·2, 26·1]	0·691 [0·429, 0·888]	329 [160, 653]	333 [175, 637]	334 [161, 686]	360 [249, 455]
Tree 3	Eperua grandiflora	82·1	32·5 [24·8, 40·4]	0·679 [0·567, 0·788]	11 100 [5590, 21 100]	11 300 [6040, 21 400]	11 100 [5970, 21 100]	12 000 [10 000, 14 000]
Plot P11					451 [441, 461]	452 [443, 462]	452 [442, 462]	451 [448, 454]

the height sub-model contributed no uncertainty to the AGB estimate.

When the WSG model was considered deterministic, the credibility interval range was also similar: 20 Mg ha^{-1}. Therefore, the WSG model contributed no uncertainty to the AGB estimate.

When the AGB sub-model was considered deterministic, the credibility interval range was narrower: 6 Mg ha^{-1}. Therefore, the AGB sub-model is the model that contributed the largest part of the uncertainty in the AGB estimate.

Discussion

ERROR PROPAGATION

Using a statistical model for prediction makes that prediction uncertain. This uncertainty has different sources:

Measurement error

When collecting data, measurements are more or less repeatable. The distribution of repeated measures around a 'true value' is the measurement error. When a measured variable is used to infer a model, this error is part of the error term of the model. We did not quantify the part of the error model that derived from the measurement error, but a previous study found it to be negligible (Chave *et al.* 2004).

Prediction error

When a model is used for prediction, the predicted values are uncertain. This uncertainty comes from (1) the model itself, through its error term and the uncertainty of its parameters and (2) the uncertainty of the other variables used in the model. These prediction errors are our great concern because they are often disregarded. Whereas it is quite natural to propagate these uncertainties in a Monte Carlo environment, the process usually makes the computation more complex (but not impossible; see for example (Gourlet-Fleury *et al.* 2011). The IPCC recommend the use of the Monte Carlo scheme when required (IPCC 2000).

Here, we identified the AGB sub-model as the main source of error in the prediction of AGB, but we did not explore the uncertainty sources inside it: measurement errors or the natural diversity among the trees. However, because measuring harvested trees is not difficult, we can hypothesize that the measurement error is only a tiny part of the AGB sub-model error.

CONSEQUENCES AND RECOMMENDATIONS

On variance

When focusing on the precision of AGB estimation, it is unnecessary to improve the quality of the WSG or height sub-models. The actual basic models are precise enough compared with the AGB sub-model.

The only way to improve the precision of the AGB predictions is to improve the precision of the AGB sub-model. Adding more trees to the calibration data may be a way to improve the precision of the model. However, there will always be diversity in AGB among trees of the same height, DBH and WSG. Because our error term in the AGB model is the same as those found previously for large data sets (Chave *et al.* 2005: RSE = 0·302; present model: $1/\sqrt{\tau} = 0 \cdot 323$), we strongly believe that we had already reached the limit.

Another way to improve the precision of the AGB predictions could be to change the AGB sub-model, adding new variables to explain the residual variance. Variables that extend our knowledge of the tree volume (such as trunk taper, diameters at various heights, crown size, etc.) or tree density (WSG measured along the trunk, at all depths, etc.) would increase the precision of the model. However, the AGB model is used on large inventory data sets. If it requires too much data to be measured in the field, it will not be useful.

ON BIAS

Bias introduced by the sampling strategy

First, the trees used to infer the model parameters must be representative of the trees whose AGB, height and WSG are to be predicted. For this reason, we used only local data sets. Our models work best around the site at Paracou, where the height and weight data were collected. This point has already been addressed (Chave *et al.* 2005; Goetz *et al.* 2009), but has not been identified as a source of bias.

Second, models should never be used outside their domains of definition. This happens when, for example, we predict the height of a tree with a higher DBH than any tree in the height calibration data set. Because the behaviour of the model outside its range of definition has not been calibrated against the data, predictions outside this range may be biased. This can be corrected by extending the range of the calibration data sets. Measuring the heights and weights of the largest trees should make us more confident of the mass predictions for very large trees.

Bias from predictive variables

Bias is a deviation of the expected value of a variable from its definition. This definition is not a physical or biological definition, although it should be close. The definitions of the variables are chosen with the calibration data set. Once the models are inferred, the definitions of the variables must remain absolutely immutable.

If, during the prediction, a variable does not have the same expected value that it would have had in the calibration data set, it is biased. A biased variable at this stage creates bias in the AGB estimate. The bias $p_{\mathrm{AGB}}\%$ in the AGB estimate induced by a bias of $p_x\%$ in the predictive variable X associated with a coefficient βx can be calculated with the formula: $p_{\mathrm{AGB}} = \left((1 + \frac{px}{100})^{\beta x} - 1\right) * 100$ We calculated p_{AGB} for the three predictive variables (WSG, DBH and H) and for six values of p_x (−10%, −5%, −2%, +2%, +5%, +10%). The results are given in Table 4.

Diameter

The DBH variable is the most standard variable in forest inventory data. In the AGB calibration data set, it is necessary to measure the diameter with respect to the inventory procedure. In this way, no bias can occur. DBH is the predictive variable that is most sensitive to bias (Table 4) because it has the highest coefficient (Table 2).

Height

In the AGB data set, height is easily measured after the tree is harvested. Therefore, this height definition is used in the biomass model, and this height must be the objective when height is measured in the field. Bias occurs if the height measured in the field differs from this definition. In Table 4, we see that a bias of 2%–5% in height creates a bias that can force the AGB estimate outside its confidence interval. A comparison of height measurements with different methods and on harvested trees should be enlightening.

Wood-specific gravity

In Table 4, we see that our model is not very sensitive to WSG bias. This derives from the low value of the coefficient (Table 2). Models with different coefficient values would have different sensitivities to WSG bias. For example, some authors fixed the WSG coefficient to one for physical reasons (Chave *et al.* 2005). In this case, the bias induced in the AGB prediction was the same as the bias in WSG.

The definition of WSG and the measurement methods used have been widely discussed in the literature. A consensus seems to have emerged for WSG [oven-dried mass (72 h at 104 °C) divided by green volume]. Other measurements can be converted to WSG with various formulae (Muller-Landau 2004; Chave *et al.* 2009).

The definition of 'wood sample' is also contentious. It seems difficult to argue that one definition is better than another, but because there is evidence for large within-individual variations (Parolin 2002; Woodcock & Shier 2002; Williamson & Wiemann 2010), it seems a bad idea to mix different sampling methods.

Again, the most important point is consistency. The definition of a variable must remain the same in every data set, from the calibration of the model to its use in prediction. If necessary, variables can be corrected to avoid bias. Note that these corrections are additional models that can also contribute uncertainty.

Conclusion

In this study, we have proposed a method to quantify uncertainty and identify its sources in biomass estimation. The most

Table 4. Percentage of bias in AGB for three possible bias in the predictive variables DBH, H and WSG. 95% confidence intervals are indicated in brackets

% of biais	−10	−5	−2	+2	+5	+10
DBH	−20·6 [−19·7, −21·6]	−10·6 [−10·1, −11·2]	−4·33 [−4·12, −4·57]	4·44 [4·21, 4·69]	11·3 [10·7, 12]	23·2 [21·9, 24·7]
Height	−7·63 [−5·97, −9·14]	−3·79 [−2·95, −4·56]	−1·51 [−1·17, −1·82]	1·5 [1·16, 1·82]	3·74 [2·89, 4·54]	7·44 [5·73, 9·06]
WSG	−1·94 [0·313, −4·05]	−0·949 [0·152, −1·99]	−0·375 [0·0599, −0·79]	0·369 [−0·0587, 0·78]	0·911 [−0·145, 1·93]	1·79 [−0·282, 3·81]

important source is the biomass sub-model, and the other sources are negligible. New height models or new WSG modelling methods, even if better than previous ones, will not necessarily improve the quality of carbon stock assessments. This study provides a methodological framework that can be broadly used by foresters and plant ecologist to provide accurate confidence intervals for tree biomass estimates from inventory data. This method can also be applied at the regional scale with error propagation. We believe that in the light of the REDD debate, such a method is a crucial step in accurately monitoring carbon stocks and their spatio-temporal evolution.

Acknowledgements

This study is part of the GUYAFOR and GUYASIM projects (30075 and 31032, programme operational FEDER 2007–2013), with financial support provided by European structural funds. Since 1984, forest inventories at the Paracou experimental site have also been funded by the National Forestry Fund (FFN), Ecofor, Silvolab, Fonds National de la Science and CIRAD, the institution in charge of Paracou. This work has benefited from an 'Investissement d'Avenir' grant managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-0025), and a BGF grant from the French Ministry of Agriculture (MAAP).

References

Araujo, T.M., Higuchi, N. & Junior, J.A.D. (1999) Comparison of formulae for biomass content determination in a tropical rain forest site in the state of Para, Brazil. *Forest Ecology and Management*, **117**, 43–52.

Baker, D.J., Richards, G., Grainger, A., Gonzalez, P., Brown, S., DeFries, R., Held, A., Kellndorfer, J., Ndunda, P., Ojima, D., Skrovseth, P.E., Souza, C. & Stolle, F. (2010) Achieving forest carbon information with higher certainty: a five-part plan. *Environmental Science & Policy*, **13**, 249–260.

Baraloto, C., Paine, C.E.T., Patino, S., Bonal, D., Herault, B. & Chave, J. (2010) Functional trait variation and sampling strategies in species-rich plant communities. *Functional Ecology*, **24**, 208–216.

Blanc, L., Echard, M., Herault, B., Bonal, D., Marcon, E., Chave, J. & Baraloto, C. (2009) Dynamics of aboveground carbon stocks in a selectively logged tropical forest. *Ecological Applications*, **19**, 1397–1404.

Brown, S. (1997) Estimating biomass and biomass change of tropical forests: A Primer. *FAO Forestry Paper 134*, p. 55. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.

Brown, S., Gillepsie, A.J.R. & Lugo, A.E. (1989) Biomass estimation methods for tropical forests with applications to forestry inventory data. *Forest Science*, **35**, 881–902.

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S. & Perez, R. (2004) Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 409–420.

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Folster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.P., Nelson, B.W., Ogawa, H., Puig, H., Riera, B. & Yamakura, T. (2005) Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, **145**, 87–99.

Chave, J., Muller-Landau, H.C., Baker, T.R., Easdale, T.A., Ter Steege, H. & Webb, C.O. (2006) Regional and phylogenetic variation of wood density across 2456 neotropical tree species. *Ecological Applications*, **16**, 2356–2367.

Chave, J., Coomes, D., Jansen, S., Lewis, S.L., Swenson, N.G. & Zanne, A.E. (2009) Towards a worldwide wood economics spectrum. *Ecology Letters*, **12**, 351–366.

Daniel, S., Dubois-Fernandez, P., Thuy Le, T., Chave, J., Blanc, L. & Davidson, M. (2010) P-band SAR study of tropical forest in French Guiana. *Remote Sensing for Agriculture, Ecosystems, and Hydrology Xii* (ed. C.M.U.M.A. Neale), SPIE Digital Library, Toulouse, France.

Dubois-Fernandez, P., Oriot, H., Coulombeix, C., Cantalloube, H., du Plessis, O. R., Thuy Le, T., Daniel, S., Chave, J., Blanc, L., Davidson, M., Petit, M. (2010) Tropisar: exploring the temporal behavior of P-band SAR data. *Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium Publication*, pp. 1319–1322.

Feldpausch, T.R., Banin, L., Phillips, O.L., Baker, T.R., Lewis, S.L., Quesada, C. A., Affum-Baffoe, K., Arets, E.J.M.M., Berry, N.J., Bird, M., Brondizio, E.S., de Camargo, P., Chave, J., Djagbletey, G., Domingues, T.F., Drescher, M., Fearnside, P.M., Franca, M.B., Fyllas, N.M., Lopez-Gonzalez, G., Hladik, A., Higuchi, N., Hunter, M.O., Iida, Y., Salim, K.A., Kassim, A.R., Keller, M., Kemp, J., King, D.A., Lovett, J.C., Marimon, B.S., Marimon, B.H., Lenza, E., Marshall, A.R., Metcalfe, D.J., Mitchard, E.T.A., Moran, E.F., Nelson, B.W., Nilus, R., Nogueira, E.M., Palace, M., Patino, S., Peh, K.S.H., Raventos, M.T., Reitsma, J.M., Saiz, G., Schrodt, F., Sonke, B., Taedoumg, H.E., Tan, S., White, L., Woll, H. & Lloyd, J. (2011) Height-diameter allometry of tropical forest trees. *Biogeosciences*, **8**, 1081–1106.

Flores, O. & Coomes, D.A. (2011) Estimating the wood density of species for carbon stock assessments. *Methods in Ecology and Evolution*, **2**, 214–220.

Gelman, A., Carlin, J., Stern, H. & Rubin, D.. (2003) *Bayesian Data Analysis*, 2nd edn (Chapman & Hall/CRC Texts in Statistical Science). Chapman and Hall/CRC, Boca Raton, Florida.

Gibbs, H.K., Brown, S., Niles, J.O. & Foley, J.A.. (2007) Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, **2**, 1–13.

Goetz, S.J., Baccini, A., Laporte, N.T., Johns, T., Walker, W., Kellndorfer, J., Houghton, R.A. & Sun, M. (2009) Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. *Carbon Balance and Management*, **4**, 2.

Gourlet-Fleury, S., Guehl, J.M.L. & Laroussinie, O.. (2004) *Ecology and Management of a Neotropical Rainforest - Lessons Drawn From Paracou, a Long-Term Experimental Research Site in French Guiana*. Elsevier, Paris, France.

Gourlet-Fleury, S., Rossi, V., Rejou-Mechain, M., Freycon, V., Fayolle, A., Saint-Andre, L., Cornu, G., Gerard, J., Sarrailh, J.-M., Flores, O., Baya, F., Billand, A., Fauvet, N., Gally, M., Henry, M., Hubert, D., Pasquier, A. & Picard, N. (2011) Environmental filtering of dense-wooded species controls above-ground biomass stored in African moist forests. *Journal of Ecology*, **99**, 981–990.

Huang, S.M., Titus, S.J. & Wiens, D.P. (1992) Comparison of Nonlinear Height Diameter Functions for Major Alberta Tree Species. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **22**, 1297–1304.

IPCC (2000) *IPCC Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories*. Hayama, Kanagawa, Japan.

Ketterings, Q.M., Coe, R., van Noordwijk, M., Ambagau, Y. & Palm, C.A. (2001) Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests. *Forest Ecology and Management*, **146**, 199–209.

Lescure, J.P., Puig, H., Riera, B., Leclerc, D., Beekman, A. & Beneteau, A.. (1983) La phytomasse épigée d'une forêt dense en Guyane française. *Acta Oecologica*, **4**, 237–251.

Lines, E.R., Zavala, M.A., Purves, D.W. & Coomes, D.A.. (2012) Predictable changes in aboveground allometry of trees along gradients of temperature, aridity and competition. *Global Ecology and Biogeography*, **10**, 1017–1028.

Muller-Landau, H.C. (2004) Interspecific and inter-site variation in wood specific gravity of tropical trees. *Biotropica*, **36**, 20–32.

Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S., McGuire, A.D., Piao, S., Rautiainen, A., Sitch, S. & Hayes, D.. (2011) A Large and Persistent Carbon Sink in the World's Forests. *Science*, **333**, 988–993.

Parolin, P. (2002) Radial gradients in wood specific gravity in trees of Central Amazonian floodplains. *Iawa Journal*, **23**, 449–457.

Pilli, R., Anfodillo, T. & Carrer, M. (2006) Towards a functional and simplified allometry for estimating forest biomass. *Forest Ecology and Management*, **237**, 583–593.

Rutishauser, E., Wagner, F., Herault, B., Nicolini, E.A. & Blanc, L. (2010) Contrasting above-ground biomass balance in a Neotropical rain forest. *Journal of Vegetation Science*, **21**, 672–682.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M. & Morel, A. (2011) Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 9899–9904.

Sarmiento, C., Patino, S., Paine, C.E.T., Beauchene, J., Thibaut, A. & Baraloto, C. (2011) Within-individual variation of trunk and branch xylem density in tropical trees. *American Journal of Botany*, **98**, 140–149.

Williamson, G.B. & Wiemann, M.C. (2010) Measuring wood specific gravity … correctly. *American Journal of Botany*, **97**, 519–524.

Woodcock, D.W. & Shier, A.D. (2002) Wood specific gravity and its radial variations: the many ways to make a tree. *Trees-Structure and Function*, **16**, 437–443.

Received 18 June 2012; accepted 16 September 2012
Handling Editor: Robert Freckleton

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Height sub-model parameters inference.

Appendix S2. WSG sub-model parameters inference.

Appendix S3. AGB sub-model parameters inference.

Appendix S4. Diameter-Height model shape selection.

Table S1. Maximum likelihood for four height models on nine sites in French Guiana.

Figure S1. Height-diameter relationship. Each black point is a tree from the Height dataset.