CrossMark

ORIGINAL ARTICLE

Error in the estimation of emission factors for forest degradation in central Africa

Nicolas Picard[1,5] · Matieu Henry [2] · Noël H. Fonton [1] · Josiane Kondaoulé [1] ·
Adeline Fayolle [3] · Luca Birigazzi [2] · Gaël Sola [2] · Anatoli Poultouchidou [2] ·
Carlo Trotta [4] · Hervé Maïdou[1]

Abstract The implementation of forest-based projects to mitigate greenhouse gas emissions requires the estimation of emission factors (here the difference in biomass stocks between two forest types). The estimation of these quantities using forest inventory data and allometric models implies different sources of errors that need to be prioritized to improve the precision of estimation. Using data from permanent sample plots in a tropical moist forest in central Africa and considering four allometric models with equal likelihood, the largest source of error in the estimate of the difference of biomass between intact and logged-over forest was that due to the model choice (40 % of the sum of squares). The error due to the model choice did not cancel out in the difference due to an interaction between the model's prediction and the diameter structure of the forest. The variability in biomass between plots was the second largest source of error, but was underestimated because of post-stratification. The error due to the model choice could be reduced by weighting the models' predictions.

Introduction

Following the Stern review that acknowledged the role of deforestation in global carbon emissions, forest-based projects, in particular REDD+ projects, have been put forward as an efficient way to mitigate greenhouse gas emissions (Angelsen et al. 2012). The implementation of such forest-based projects require accurate, precise, and verifiable estimates of forest carbon changes. Following decision 4/CP.15 of the UNFCCC (2010), estimates of forest carbon stocks and of carbon changes in REDD+ projects should comply with the IPCC 2006 guidelines (Eggleston et al. 2006). One general method of the IPCC to estimate forest carbon changes consists in multiplying activity data (in this context, areal extent in hectares that changes from one category of land use to another) by emission factor (in this context, difference of carbon stocks per hectare between the two categories of land use) (GOFC-GOLD 2012, §1.2.3).

For forest categories of land use, emission factors are typically estimated using forest inventory data and biomass equations. Biomass equations are models that predict tree biomass on the basis of one or more dendrometric measurements (such as diameter, height, wood density) that are non-destructive (UNFCCC 2011). Along with the biomass estimation comes an error (sensu GOFC-GOLD 2012, i.e., with a broader meaning than in statistical science) that reflects how accurate and precise this estimation is. Uncertainty assessment is a major requirement of REDD+ and has financial implications (Pelletier et al. 2015). When combining plot data and biomass equations, different

✉ Nicolas Picard
   nicolas.picard.1991@polytechnique.org

1   COMIFAC, BP 20818, Yaoundé, Cameroon

2   Food and Agriculture Organisation of the United Nations, Viale delle Terme di Caracalla, 00153 Rome, Italy

3   Laboratoire de foresterie des régions tropicales et subtropicales, Unité de gestion des ressources forestières et des milieux naturels, Gembloux Agro-Bio Tech, Université de Liège, Passage des déportés 2, 5030 Gembloux, Belgium

4   University of Tuscia, Department for Innovation in Biological, Agro-Food and Forest Systems (DIBAF), Largo dell'Università, Blocco D, 01100 Viterbo, Italy

5   CIRAD, BP 2572, Yaoundé, Cameroon

⨷ Springer

sources of error propagate to the forest-level estimation of biomass (Cunia 1987). If plots are part of a forest inventory, data contribute through the sampling error that depends on the sampling design of the inventory and on the spatial heterogeneity of the forest area being inventoried. If they are part of an experiment (e.g., permanent plots), data contribute through the variability in biomass between plots which is representative of the variability at the site level. Because biomass is not measured in sample plots, but predicted using a model, each plot-level estimate of biomass includes, in turn, a model error that consists of the error due to the model choice (different models bring different predictions), of the prediction error (i.e., the error due to the uncertainty on the model's coefficients, plus the residual model error), and of the error on the predictors of the model (i.e., measurement errors for tree diameter or height or estimation error for wood density; Chave et al. 2004). The error due to the uncertainty on the model's coefficients can further be broken into a sampling-type error due to the finiteness of the sample of trees used for model fitting and into a measurement error for the trees of this fitting data set (Mason et al. 2014).

Many studies have shown that the error due to the choice of the biomass equation is a major source of error in forest-level estimates of biomass stocks (Picard et al. 2015). However, it has been argued that this source of error should be less important for the estimation of carbon changes (i.e., difference of carbon stocks between two time periods) provided that the same biomass equations are used to calculate carbon stocks at both time periods (Yanai et al. 2012). If the error due to the model choice consisted of a systematic overestimation or underestimation, the prediction bias would indeed cancel out when computing the stock difference. In reality, the prediction bias of biomass equations also depends on tree diameter, so that the error due to the model choice will interact with the plot diameter structure and will not exactly cancel out when computing the stock difference. This reasoning for carbon changes also applies to any quantity that is computed as the difference between two carbon stocks, in particular emission factors. Therefore, one pending question is whether the error due to the model choice is also a major source of error in the estimation of emission factors, provided that the same biomass equations are used to calculate carbon stock for both forest categories.

In this study, we focused on the transition from an intact forest to a forest disturbed by logging and thinning (including removed trees and damage) in central Africa. Because there is no consensus on the definition of forest degradation (GOFC-GOLD 2012, pp. 1–10), and considering a disturbance intensity much higher than that commonly practised in sustainable logging in central Africa, we assimilated this disturbance to a form of degradation (the second D of REDD+). Consequently, an emission factor associated with the transition from intact to disturbed forest could be defined (Sasaki 2006). This study aimed at assessing the contributions of the different sources of errors to the total error when estimating this emission factor (not to be confused with the biomass stocks). The following questions were addressed: (1) What is the contribution of the error due to the model choice when estimating the emission factor, provided that the same biomass equations are used for intact and disturbed forests? (2) Is the error due to the model choice less important for the emission factor than for the biomass stocks? We used data from an experimental site in central Africa, where both intact forest and logged and thinned forest were monitored.

Material and methods

Study site

Forest inventory data came from the M'Baïki experimental site (3°54′N, 17°56′E) in the Central African Republic. Forest is lowland semi-evergreen moist forest dominated by *Celtis zenkeri* Engl. (Ulmaceae), *Staudtia kamerunensis* var. *gabonensis* (Warb.) Fouilloy (Myristicaceae), *Coelocaryon preussii* Warb. (Myristicaceae), *Garcinia punctata* Oliv. (Clusiaceae), *Carapa procera* DC. (Meliaceae), and *Dasylepis seretii* De Wild. (Flacourtiaceae). Average annual rainfall is 1739 mm (1981–2008 period) with a 3-month dry season (December–February). Soils are Ferralsols (World Reference Base for soil resources). Relief is flat (altitude of 500–600 m a.s.l.).

In 1982, 10 permanent sample plots, each of 4 ha, were set up and all trees with a diameter at breast height (dbh) ≥10 cm were spatially located and individually marked (Bedel et al. 1998). Species were identified. All dbh were remeasured annually (except in 1997, 1999 and 2001), and dead trees and newly recruited trees were surveyed. Seven plots were selectively logged between the 1984 and 1985 inventories (harvesting trees ≥80 cm for 16 commercial species), whereas the three other plots were left as controls. Four of the seven logged plots were additionally thinned (poison girdling of all non-timber trees ≥50 cm dbh) 2 years after logging. In 1984, an accidental fire burnt one quarter of two of the plots (one control plot and one logged and thinned plot).

The site was originally designed as a randomized block design with three blocks within a distance of 10 km, but because logging, thinning, and fire have not been homogeneous across the 4-ha plots, the planned design turned out to be irrelevant, and the 4-ha plots were no longer appropriate experimental units for the silvicultural treatments. Dividing each 4-ha plot into four 1-ha subplots and

Table 1 Biomass equations to assess the error due to model choice

References	Country	No.	a	b	c	d	σ	n
Chave et al. (2014)	Pantropical	7	-1.875^a	0.976	2.673	-0.0299	0.413	4004
Djomo et al. (2010)	Cameroon	9	-1.9644	0.3579	2.3382	0	0.325	71
Henry et al. (2010)	Ghana	4	-1.23	0	2.31	0	0.224	41
Ngomanda et al. (2014)	Gabon	5	-4.114	1.431	4.062	-0.228	0.330	101

The number (No.) is the model number in the publication. a–d are the model's coefficients, as defined in (1). σ is the residual standard error of the model. n is the number of observations used to fit the model. The bias-correcting factor $\sigma^2/2$ was removed from the a coefficient when the original publication included it

[a] The value given here includes the term depending on the climate index E, with $E = -0.013$ at M'Baïki

considering these subplots as the statistical units solved this issue, with the caveat that contiguous 1-ha subplots were no longer true independent replicates (Ouédraogo et al. 2011).

We used the data of 1987 (i.e., after logging and thinning). The 40 1-ha subplots were ranked in order of increasing basal area. The 12 subplots with highest basal area (ranging 28.2–36.7 m^2 ha^{-1} with a mean of 32.5 m^2 ha^{-1} and a std. dev. of 2.8 m^2 ha^{-1}) were considered representative of undisturbed forest. With the exception of one subplot that originated from a logged plot, they all originated from control plots and excluded any burnt subplot. The 12 subplots with lowest basal area (ranging 17.1–21.3 m^2 ha^{-1}, mean: 19.7 m^2 ha^{-1}, std. dev.: 1.1 m^2 ha^{-1}) were considered representative of disturbed forest. They all originated from logged and thinned plots, including one burnt subplot.

Based on this post-stratification, logging intensity ranged between two and nine trees ha^{-1}, which is higher than commonly practised in sustainable forest management in central Africa (<2 trees ha^{-1}), and thinning removed an additional 11–41 trees ha^{-1}. The proportion of basal area lost between 1982 and 1987 for the 12 disturbed subplots was 45 % on average (min. 33 %, max. 56 %), and the proportion of number of trees lost was 24 % (min. 12 %, max. 40 %, see Ouédraogo et al. 2011 for more details). After 24 years, 63 % of the subplots had recovered their initial biomass, but only 13 % of the subplots had recovered their initial timber stock (Gourlet-Fleury et al. 2013). Correspondingly, there was a shift in species composition (Ouédraogo et al. 2011).

Biomass equations

Inventory data were converted into biomass estimates using biomass equations. Wood specific densities were extracted from Zanne et al. (2009)'s data base. When no match was found at the species level, the mean wood density across the genus was used (Slik 2006). If no match was found at the genus level, a default value of 0.6 g cm^{-3} was used (Henry et al. 2010). We selected biomass equations (1) that used dbh and wood density as predictors (but not height or any other tree mensuration that were not measured at M'Baïki), (2) that were pantropical or fitted to a data set from African moist forests, and (3) whose data set used for model fitting was publicly available. The latter criterion ensured that we would be able to get the variance-covariance matrix of the estimates of the model's coefficients (that is very rarely documented in publications). Therefore, four biomass equations were selected, all of which were particular cases of the following model (Table 1):

$$\ln(B) = a + b\ln(\rho) + c\ln(D) + d[\ln(D)]^2 \qquad (1)$$

where B is tree aboveground biomass (in kg), ρ is wood density (in g cm^{-3}), D is diameter (in cm), and a, b, c, d are model's coefficients.

Error propagation

A Monte Carlo simulation method was used to propagate errors from the tree level to the plot level, then to the forest level (McRoberts and Westfall 2014). Measurement errors on dbh were assumed to be lognormally distributed to ensure that dbh remain positive (Cohen et al. 2013), with a 2 % error (Melson et al. 2011). Therefore, at the kth Monte Carlo iterate, dbh $\hat{D}_{ijk}$ for the ith tree of plot j was drawn according to a lognormal distribution with mean $\ln(D_{ij}) - \sigma^2/2$ and standard deviation $\sigma = 1.0\,\%$, where D_{ij} was the observed dbh of the tree in the M'Baïki data set. The uncertainty on the coefficients of model m was simulated by drawing coefficients $\hat{a}_{mk}$, $\hat{b}_{mk}$, $\hat{c}_{mk}$, $\hat{d}_{mk}$ of model (1) at the kth Monte Carlo iterate according to a multinormal distribution with mean a_m, b_m, c_m, d_m (as given in Table 1) and with variance-covariance matrix Σ_m equal to that of the estimates of the model's coefficients. Matrix Σ_m was not reported in the publications and was obtained by refitting the models to their original data set. The residual error of model m was simulated by adding to the prediction of the log-transformed biomass a random error ε_{ijmk} with normal distribution with mean zero and standard deviation σ_m

(as given in Table 1). Thus, at the kth Monte Carlo iterate, the predicted biomass of the ith tree of plot j according to model m was:

$$\hat{B}_{ijmk} = \exp[\hat{a}_{mk} + \hat{b}_{mk}(\ln \rho_{ij}) + \hat{c}_{mk}(\ln \hat{D}_{ijk}) \\ + \hat{d}_{mk}(\ln \hat{D}_{ijk})^2 + \varepsilon_{ijkm}] \tag{2}$$

where ρ_{ij} was the wood density of the ith tree of plot j.

At the plot level, the kth Monte Carlo iterate of the biomass for plot j according to model m was obtained by summing $\hat{B}_{ijmk}$ across i. Let $\hat{\mathcal{B}}_{jmk}$ be this plot-level estimate of biomass. The uncertainty on the biomass estimate for undisturbed forest was obtained from the empirical distribution of $\hat{\mathcal{B}}_{jmk}$, with j varying across the 12 plots of undisturbed forests, m varying across the four biomass equations, and k varying from 1 to the number K of Monte Carlo simulations (we used $K = 10{,}000$ iterations). In particular, the mean predicted biomass for undisturbed forest was: $\bar{\mathcal{B}} = \sum_{j=1}^{12} \sum_{m=1}^{4} \sum_{k=1}^{K} \hat{\mathcal{B}}_{jmk}/(4 \times 12 \times K)$, and the sum of square errors for this biomass estimate was: $\sum_{j=1}^{12} \sum_{m=1}^{4} \sum_{k=1}^{K} (\hat{\mathcal{B}}_{jmk} - \bar{\mathcal{B}})^2$. This computation gave the same weight to all biomass equations, thus implicitly assuming that they were equally likely.

The uncertainty on the biomass estimate for disturbed forest was obtained in the same way, but with j varying across the 12 plots of disturbed forest. Finally, the uncertainty on the estimate of the emission factor was obtained from the empirical distribution of $\hat{\mathcal{B}}_{jmk} - \hat{\mathcal{B}}_{j'mk'}$, with j indexing one of the 12 plots of undisturbed forest, j' indexing one of the 12 plots of disturbed forest, and k and k' being any of the Monte Carlo iterates (but the model used to predict the plot biomass in undisturbed forest must be the same as the one used for disturbed forest).

Error partition

The relative contribution of each source of error to the total error was computed by sequentially switching on/off every source of error. The error due to the choice of the model was switched off by replacing $\hat{\mathcal{B}}_{jmk}$ for all j and k with its average across models: $\sum_{m=1}^{4} \hat{\mathcal{B}}_{jmk}/4$. The between-plot variability was switched off by replacing $\hat{\mathcal{B}}_{jmk}$ for all m and k with its plot average: $\sum_{j} \hat{\mathcal{B}}_{jmk}/12$. The measurement error was switched off by replacing $\hat{D}_{ijk}$ with D_{ij}. The uncertainty on the models' coefficients was switched off by replacing $\hat{a}_{mk}, \hat{b}_{mk}, \hat{c}_{mk}, \hat{d}_{mk}$ with a_m, b_m, c_m, d_m. Finally, the residual model error was switched off by replacing ε_{ijmk} with zero. In this latter case, the predicted tree biomass had to be multiplied by the bias-correcting factor $\exp(\sigma_m^2/2)$. The proportion of the total sum of square errors attributable to a given source of error was computed as

$[\text{SQE}_{on} - \text{SQE}_{off}]/\text{SQE}_{tot}$, where SQE_{tot} was the total sum of square errors (when all sources of errors were switched on), and SQE_{on} and SQE_{off} were the sums of square errors when the subject source of error was switched on and off, respectively.

The order in which sources are switched on/off can influence the error partition between sources, depending on the interactions between sources of errors. If there is an interaction between two sources of error, the contribution of this interaction to the total error will be attributed to the source that is switched on last. For some sources of error, the order was actually indifferent as there was no interaction between them. For example, the contribution of the uncertainty on the models' coefficients and that of the residual model error can be added in any order. The question mainly arose for the error due to the model choice and for the between-plot variability.

It is important to assess the interaction between the forest type (disturbed/undisturbed) and the model choice because, if this interaction was null, the estimate of the emission factor would be independent of the model choice. Therefore, a two-way analysis of variance was also conducted to specifically address the forest type-model interaction:

$$\mathcal{B}_{jmt} = \mu + \alpha_{jt} + \beta_m + \gamma_{mt} + \varepsilon_{jmt} \tag{3}$$

where $\mathcal{B}_{jmt}$ was the estimated biomass of the jth plot ($j = 1, ..., 12$) in forest type t (disturbed or undisturbed) according to model m (with all other sources of errors switched off), μ was the overall mean, α_{jt} was the plot effect, β_m was the model effect, γ_{mt} was the forest type-model interaction, and ε_{jmt} was the residual error. We considered a plot effect α_{jt} rather than a forest type effect α_t because the different model predictions were computed for the same plots (and not for independent plots).

Results

On average across all sources of errors, aboveground biomass at M'Baïki was 312 tonne ha^{-1} in undisturbed forest and 165 tonne ha^{-1} in disturbed forest. Therefore, the emission factor due to forest degradation was 147 tonne ha^{-1} on average. However, there was strong disagreement among models for the estimate of the emission factor, with a 2.3-fold variation between the lowest estimate of the emission factor (Djomo et al. 2010's model: 87 tonne ha^{-1}) and the highest one (Henry et al. 2010's model: 197 tonne ha^{-1}).

A strong disagreement among models for the estimate of the biomass stocks was also found (Fig. 1), but because the emission factor is the difference between two biomass
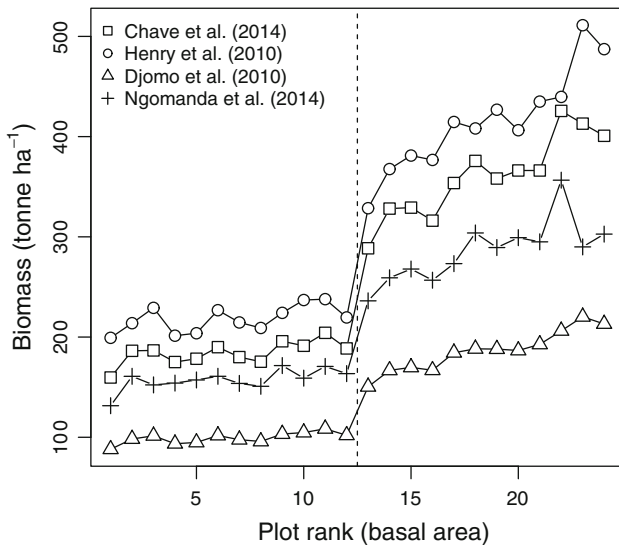
Fig. 1 Aboveground tree biomass in 24 plots at M'Baïki, Central African Republic, according to four biomass equations: *square symbol* Chave et al. (2014)'s model; *circles* Henry et al. (2010)'s model; *triangles* Djomo et al. (2010)'s model; *crosses* Ngomanda et al. (2014)'s model. Plots are ranked in the order of increasing basal area, with the 12 first plots (*on the left of the vertical dotted line*) being representative of disturbed forest, and the 12 last ones (*on the right*) being representative of undisturbed forest

Table 2 Analysis of variance table for the model that predicts plot biomass depending on plot, model choice, and the interaction between the model and the forest type (disturbed/undisturbed), for 24 1-ha plots and according to four biomass models at M'Baïki

Effect	D.f.	Sum sq.	Mean sq.	F value	Pr (>F)
Plot	23	573,266	24,925	129.3	<0.001
Model	3	401,363	133,788	693.9	<0.001
Forest type-model	3	43,785	14,595	75.7	<0.001
Residuals	66	12,725	193		

stocks estimated with the same model, this disagreement alone could not explain the observed differences in emission factors across models. What mattered was the interaction between the forest type and the model choice, which was significant at 5 % level (Table 2). Therefore, the choice of the model not only strongly influenced the level of biomass estimates, but its interaction with the plot dbh structure was also strong enough to influence the estimate of biomass stock differences.

The model choice was the largest source of error for the estimation of the emission factor, representing approximately 40 % of the sum of square errors (Fig. 2c). The between-plot variability (that also included the plot-model interaction) was the second largest source of error. As expected, the error due to the uncertainty on the model's

coefficients was the greatest for those models that were fitted to the smallest data sets (Djomo et al. 2010's and Henry et al. 2010's models). Measurement error was negligible with respect to the other sources of error.

In comparison, for biomass estimation, the model choice was also the largest source of error, but its contribution to the sum of square errors (80 %) was about twice that for emission factor estimation (Fig. 2a, b). For each model, the partition of error for the estimation of biomass in disturbed plots was similar to that obtained for the estimation of the emission factor. The variability of biomass estimates was less marked in disturbed plots than in undisturbed plots.

Discussion

As already pointed out by several studies (Kenzo et al. 2009; Rutishauser et al. 2013), the choice of a biomass equation to convert inventory data into biomass estimates is critical for tropical rain forests. A new result shown here is that this choice is also critical for the estimation of emission factors, i.e., differences between two biomass stocks estimated with the same model. Thus, although prediction biases cancelled out when computing differences in biomass stocks (Holdaway et al. 2014), the interaction between the dbh structure of the plot and the model (that does not cancel out when computing differences) was still strong enough to influence the estimate of the difference. Accordingly, the model choice contributed much less to the total error for the estimate of the emission factor at M'Baïki than for the estimate of biomass stocks, but it was still the main source of error of the estimate of the emission factor.

As compared to other forms of forest degradation and in line with the gradient of degradation defined by Thompson et al. (2013), the one considered here would correspond to low degradation. Both the estimates of the emission factor and of the biomass stocks given here were consistent with those given by Gourlet-Fleury et al. (2013). For instance, using Chave et al. (2005)'s pantropical equation based on dbh and considering between-plot variability as the only source of error, they estimated the biomass stock to be 374.5 ± 58.2 (std. err.) tonne ha^{-1} in undisturbed forest in 1982 at M'Baïki.

In current pilot REDD+ projects that operate at a local scale, a single allometric equation is often chosen based on the UNFCCC (2011) guidelines, which is equivalent with options C, D, H, or N of Fig. 2. However, the error due to the model choice should be included in the uncertainty assessment for carbon emissions reporting. This error can be reduced by assigning different weights to the models'
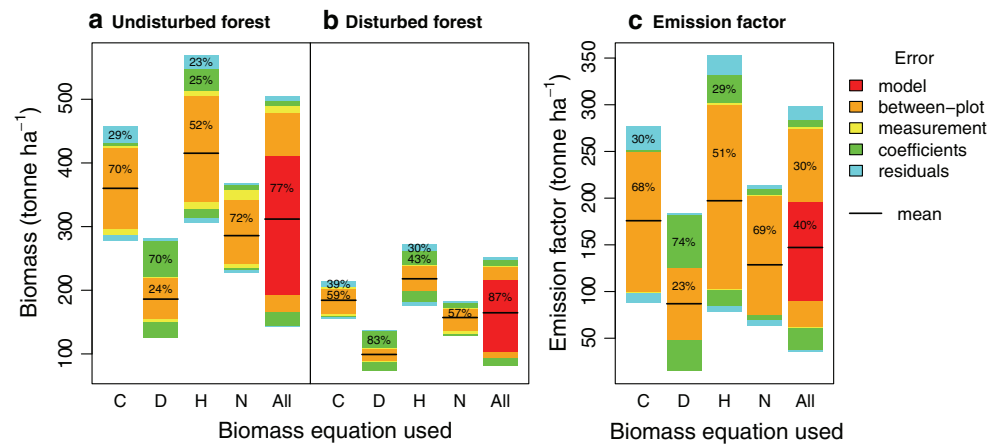
Fig. 2 Estimate of aboveground biomass in **a** undisturbed forest and **b** disturbed forest at M'Baïki, Central African Republic, and **c** estimate of the emission factor due to forest degradation, for different biomass models (*C* Chave et al. 2014's model, *D* Djomo et al. 2010's model, *H* Henry et al. 2010's model, *N* Ngomanda et al. 2014's model) and when combining all models with equal weights (*All*). *Boxes* give the 95 % prediction intervals: the upper and lower limits of the box give the upper and lower bounds of the 95 % prediction interval, respectively. The *black line* gives the mean. The prediction intervals are non-symmetrical, thus giving a different size of the box above and below the mean. The *different colors* correspond to the different sources of errors: *red* error due to the model choice; *orange* between-plot variability; *yellow* measurement error; *green* uncertainty on the models' coefficients; and *blue* residual model error. *Numbers* give the proportions of the sum of square error associated with each source of errors (one single value for each source of error; proportions less than 20 % in **a** and **c** or 30 % in **b** were not typed) (color figure online)

predictions, reflecting the probability of each model to be the "best" among the ensemble of models. In the ideal case where one model would outperform all the other ones (i.e., one model with a weight of one and all other models with a weight of zero), the error due to the model choice would vanish. A posteriori weights can be computed using Bayesian model averaging (Picard et al. 2015) but this technique requires a training data set of tree biomass with all the variables used in the models. Few tree biomass data sets are currently availably in central Africa (with a total of 819 measured trees), and none close to M'Baïki (Fig. 3).

The variability in biomass between the M'Baïki plots (coefficient of variation of 28 % for undisturbed plots, 26 % for disturbed plots) was of the same order as that reported for 1-ha plots in other undisturbed tropical rain forests (28 % in Brazilian Amazonia, Keller et al. 2001; 15 % in Panama, Chave et al. 2004). Yet it was presumably underestimated for disturbed plots at M'Baïki because of the post-stratification. Assessing this variability at the forest rather than at the site level would require a forest inventory, thus allowing to estimate the sampling error. Sampling error is the source of error that has the most often been addressed in studies devoted to the estimation of forest biomass (Réjou-Méchain et al. 2014), but considering it as the only source of error (e.g., Kearsley et al. 2013) would result in underestimated prediction intervals. In the case of emission factors for forest degradation, using forest inventories further raises the question of controlling for the logging degradation that is generally not spatially homogeneous across the forest.
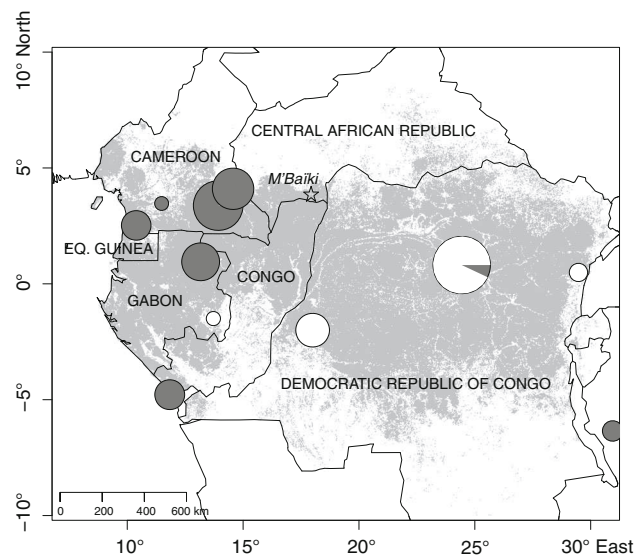

Fig. 3 Location of sites with destructive tree biomass measurements in central Africa (*circles*). The size of the *circle* is proportional to the number of trees measures (from 6 to 160), and its *color* indicates if the data were published (*dark grey* published, *white* unpublished). The *star* shows the location of the M'Baïki experimental site. The *grey* area cumulates the five following types of Verhegghen et al. (2012)'s classification of Congo basin forests: dense moist forest, submontane forest, mountain forest, edaphic forest, and mangrove

Measurement error in our study was negligible with respect to the other sources of errors, but was restricted to dbh measurement. This result is consistent with other studies where measurement errors were restricted to dbh, height, or wood density (Molto et al. 2013). In contrast,

when measurement errors are more comprehensively accounted (including, e.g., species identification, measurement of the plot area, missed or double counted trees), their contribution to the total error increases (Holdaway et al. 2014).

The prediction error is composed of the uncertainty on the model's coefficients and of the residual error (Cunia 1987). Focusing on the former and failing to account for the latter may strongly underestimate the model error (Holdaway et al. 2014). The uncertainty on the coefficients decreases as the number n of observations in the data set used for model fitting increases (asymptotically, proportionally to $1/\sqrt{n}$). Contrary to Ståhl et al. (2014), who found a tenfold decrease in the error due to the uncertainty on the coefficients when estimating biomass changes rather than biomass stocks, the contribution of this uncertainty was about the same from biomass stocks and for emissions factors in our study. The way this error was propagated may explain these different results. Whereas Ståhl et al. (2014) used the same set of model's coefficients for the two biomass estimates that defined the change, we used independent Monte Carlo outcomes for the two terms of the difference.

The residual error is often very high for a single tree prediction: a residual error of σ on log-transformed data corresponds to a margin of error (i.e., one-half the width of the prediction interval at level α) of approximately $\sinh(q_{1-\alpha/2}\sigma)$, where q_α is the α quantile of the centered and scaled Gaussian distribution. For the models listed in Table 1, this tree-level margin of error ranges between 90 % (for Chave et al. 2014's model) and 45 % (for Henry et al. 2010's model) at level $\alpha = 5\%$. At plot level, the residual error decreases as $A^{-0.5}$ as plot size A increases when trees randomly accumulate, but may not completely level off when there is a plot effect on biomass.

We conclude that the choice of the biomass equation was critical, both for biomass estimation and for the estimation of emission factors at M'Baïki. Even if some sources of error (in particular the between-plot variability and the measurement error) have certainly been underestimated in our study, the error due to the model choice will remain a large source of error. Furthermore, when two different equations are applied at two different time periods to measure the carbon stock change from one land use to another (e.g., natural forest to plantation), the contribution of the model choice to the total error will be even greater than reported here, with the risk that inaccurate carbon gains may be reported.

References

Angelsen A, Brockhaus M, Sunderlin WD, Verchot LV (eds) (2012) Analysing REDD+: challenges and choices. CIFOR, Bogor

Bedel F, Durrieu de Madron L, Dupuy B, Favrichon V, Maître HF, Bar-Hen A, Narboni P (1998) Dynamique de croissance dans les peuplements exploités et éclaircis de forêt dense africaine: dispositif de M'Baïki en République Centrafricaine (1982–1995). No. 1 in Forafri, CIRAD-Forêt, Montpellier, France

Chave J, Condit R, Aguilar S, Hernandez A, Lao S, Perez R (2004) Error propagation and scaling for tropical forest biomass estimates. Philos Trans R Soc Lond B 359(1443):409–420

Chave J, Andalo C, Brown S, Cairns MA, Chambers JQ, Eamus D, Fölster H, Fromard F, Higuchi N, Kira T, Lescure JP, Nelson BW, Ogawa H, Puig H, Riéra B, Yamakura T (2005) Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145:87–99

Chave J, Réjou-Méchain M, Búrquez A, Chidumayo E, Colgan MS, Delitti WBC, Duque A, Eid T, Fearnside PM, Goodman RC, Henry M, Martínez-Yrízar A, Mugasha WA, Muller-Landau HC, Mencuccini M, Nelson BW, Ngomanda A, Nogueira EM, Ortiz-Malavassi E, Pélissier R, Ploton P, Ryan CM, Saldarriaga JG, Vieilledent G (2014) Improved allometric models to estimate the aboveground biomass of tropical trees. Glob Chang Biol 20:3177–3190

Cohen R, Kaino J, Okello JA, Bosire JO, Kairo JG, Huxham M, Mencuccini M (2013) Propagating uncertainty to estimates of above-ground biomass for Kenyan mangroves: a scaling procedure from tree to landscape level. For Ecol Manag 310:968–982

Cunia T (1987) Error of forest inventory estimates: its main components. In: Whraton EH, Cunia T (eds) Estimating tree biomass regressions and their error. Proceedings of the workshop on tree biomass regression functions and their contribution to the error of forest inventory estimates, May 26–30, 1986, Syracuse, N.Y.—Part E, USDA Forest Service, Northeastern Forest Experiment Station, Broomall, PA, USA, no. NE-117 in General Technical Report, pp 1–14

Djomo AN, Ibrahima A, Saborowski J, Gravenhorst G (2010) Allometric equations for biomass estimations in Cameroon and pan moist tropical equations including biomass data from Africa. For Ecol Manag 260:1873–1885

Eggleston S, Buendia L, Miwa K, Ngara T, Tanabe K (eds) (2006) 2006 IPCC guidelines for national greehouse gas inventories. Agriculture, forestry and other land use, vol 4, Institute for Global Environmental Strategies (IGES) on behalf of the Intergovernmental Panel on Climate Change (IPCC), Hayama, Japan

GOFC-GOLD (2012) A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals associated with deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation. No. version COP18-1 in GOFC-GOLD Report, GOFC-GOLD Land Cover Project Office, Wageningen University, The Netherlands

Gourlet-Fleury S, Mortier F, Fayolle A, Baya F, Ouédraogo D, Benedet F, Picard N (2013) Tropical forest recovery from logging: a 24 year silvicultural experiment from Central Africa. Philos Trans R Soc Lond B 368(1625):20120302