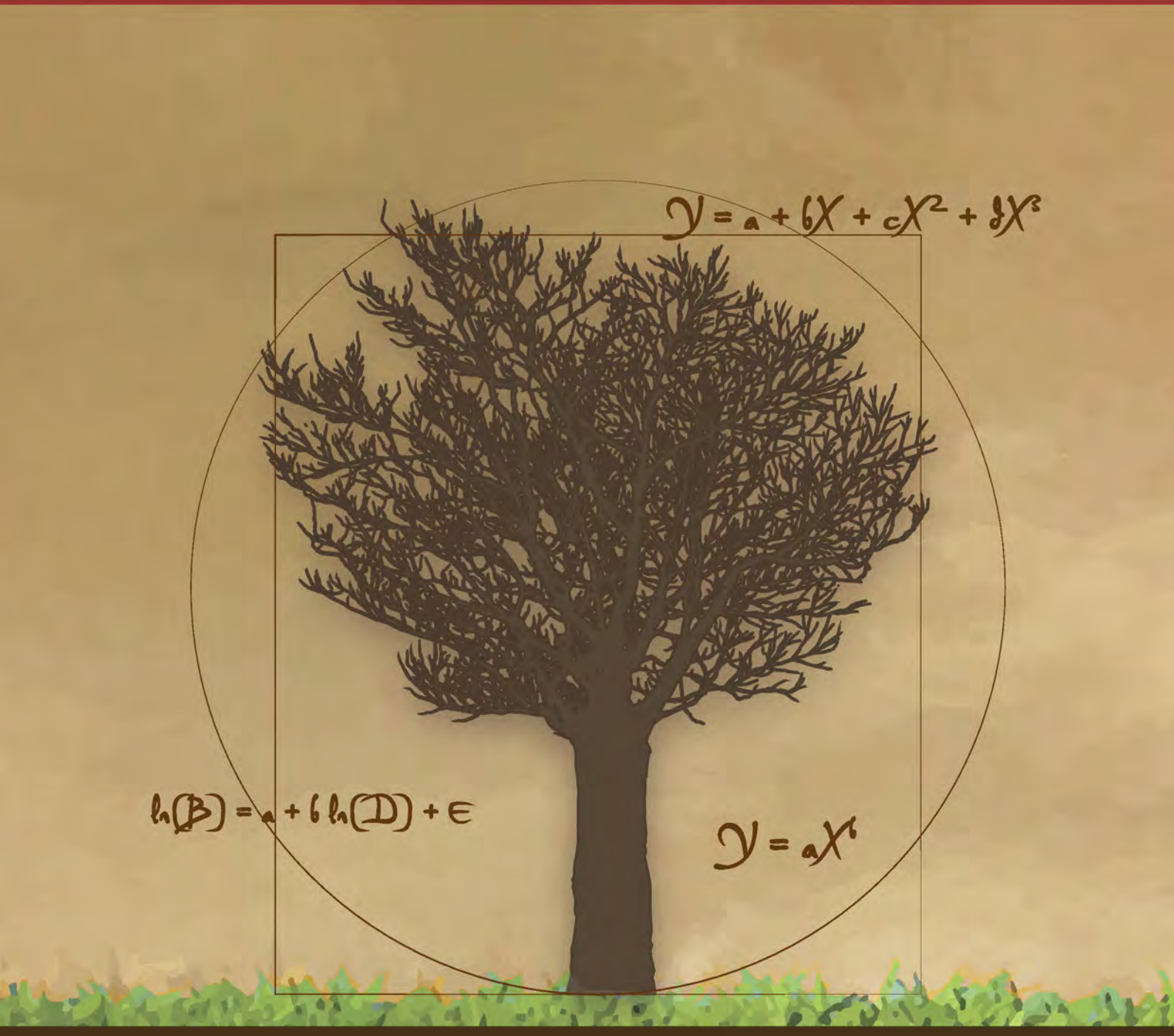
Summary of the manual for building tree volume and biomass allometric equations

From field measurement to prediction



cirad

FAO FIAT PANIS

Summary of the manual for building tree volume and biomass allometric equations: from field measurement to prediction

Gael Sola,
Forestry Department,
Food and Agriculture Organization of the United Nations

Nicolas Picard,
Environments and Societies Department,
Centre de Coopération Internationale en Recherche Agronomique pour le Développement

Laurent Saint-André,
UMR Eco&Sols,
Centre de Coopération Internationale en Recherche Agronomique pour le Développement
& UR1138 BEF, Institut National de la Recherche Agronomique

Matieu Henry,
Forestry Department,
Food and Agriculture Organization of the United Nations

Food and Agriculture Organization of the United Nations (FA0)
Viale delle Terme di Caracalla
00153 Rome, Italie

Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD)
Campus international de Baillarguet
34 398 Montpellier Cedex, France

Recommanded Citation: **Sola G., Picard N., Saint-André L., Henry M.**, 2012. Summary of the manual for building tree volume and biomass allometric equations: from field measurement to prediction. Food and Agriculture Organization of the United Nations and Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Rome, Montpellier, 20 pp.

Preamble

The manual for building tree volume and biomass allometric equations [1] is designed for students, researchers and engineers who wish to acquire the knowledge and methodology to establish allometric equations to assess the volume, biomass or mineralomass of trees. The manual provides numerous examples and detailed technical advice, so there are few prerequisites for users. The aim of this synthesis is to highlight and explain the manual's main steps.

A manual to improve estimates of forest volume and biomass and support the development of allometric equations

Forests provide many services, including the production of wood and the storage of carbon. Accounting for these services requires an accurate estimate of tree volume and biomass. At the scale of a specific plot, forest or group of forests, the accuracy of this estimate remains a challenge for technical, human and financial reasons, but also regarding the scientific tools required. Forest and tree biomass varies depending on the type of forest (Figure 1), the species concerned and which compartments of the tree are included; since it is usually impractical to measure all trees in a given area, estimates must be made on the basis of samples. Research is continuing in these fields, but by combining existing knowledge of mathematics, biology and forest science, accurate estimates can now be made.

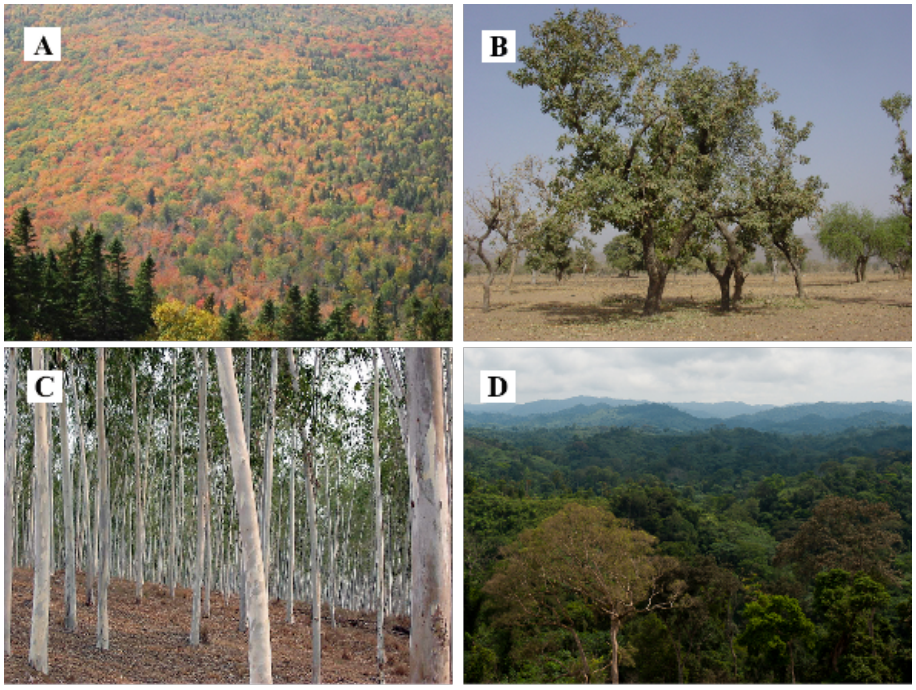


Figure 1. A: Continental temperate forest (Photo: Bruno Locatelli); B: Tropical dry forest (Photo: Régis Peltier); C: Eucalyptus plantation (Photo: Bruno Locatelli); D: Tropical rainforest (Photo: Gaël Sola).

1. **Picard N., Saint-André L.,Henry M.**, 2012. Manual for building tree volume and biomass allometric equations: from field measurement to prediction. Food and Agriculture Organization of the United Nations and Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Rome, Montpellier, 222 pp.

Theoretically, to determine the carbon stock of tree, one must weigh all its compartments (Figure 2). These measurements become difficult or impossible to obtain at the forest scale for two reasons: (1) they are destructive and often prohibited on a large scale, and (2) the cost, time and labor are unrealistic. To measure the carbon content of tree roots, field work is even more tedious and is often replaced by the use of factors directly applied to the aboveground biomass.

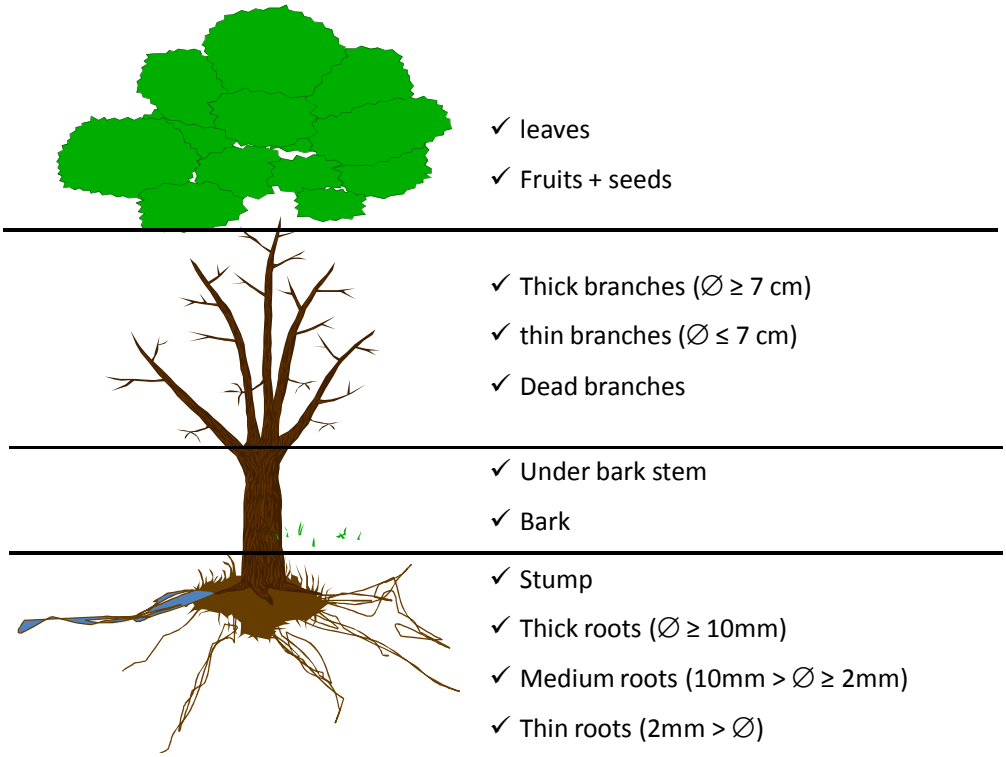


Figure 2. compartments of a tree.

Another methodology to obtain an estimate of tree biomass is based on the relation between the diameter and biomass of a given tree. This is the stage at which allometric equations are used. Allometry refers to the statistical relation between two size characteristics of individuals in a population. Therefore a statistical relation can be developed between easy to measure tree characteristics (i.e. diameter, height or density) and difficult to measure variables such as biomass or volume. Consequently, costly and destructive measurements can be limited to a sample of trees and the results can be extrapolated to all trees in a given area (box 1).

Box 1:
Parameters can be set up to obtain a power relation between tree diameter and biomass:

$$\underbrace{Biomass}_{\substack{\text{variable to}\\ \text{be explained}}} = b \times \underbrace{\text{Diameter}}_{\substack{\text{explanatory}\\ \text{variable}}}{}^{a}$$

Statistical adjustment is based on calculating parameters (a and b) which relate most accurately biomass to diameter in a target area.

With the statistical tools presented in this report, the error of prediction can be calcula-

ted. Once the needs for carbon stock assessment have been identified, the manual develops a seven step methodology (Figure 3) to provide accurate allometric equations, from the selection of explanatory variables to the best fitting allometric equation calculation and validation.

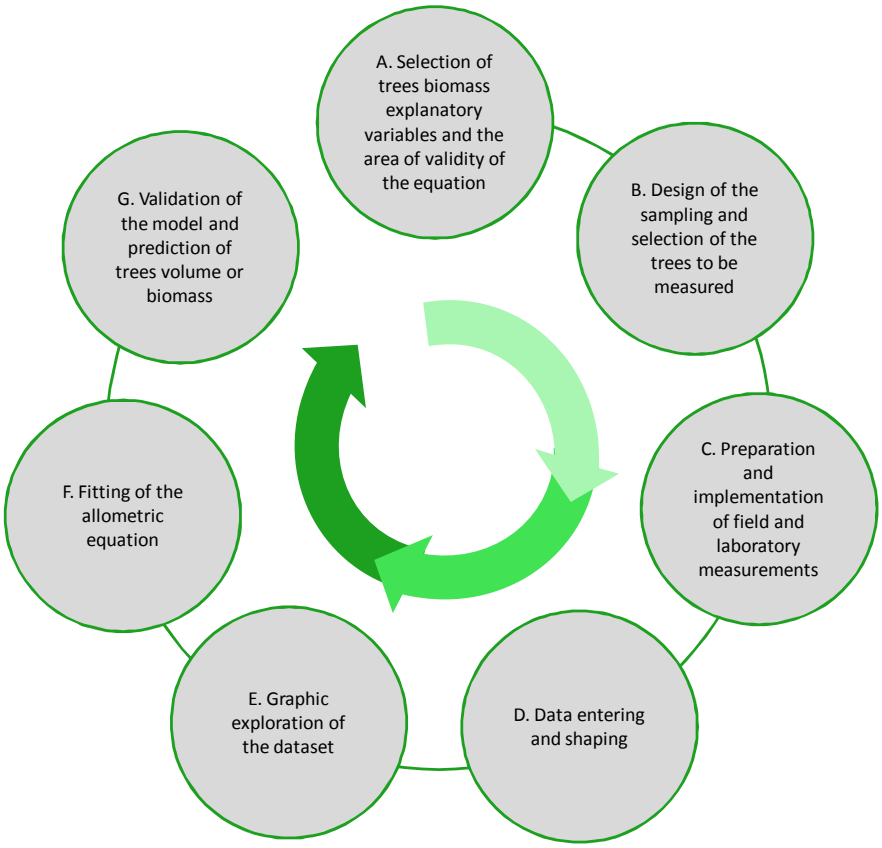


Figure 3. The seven steps of the methodology presented in the manual.

Choice of the factors influencing tree biomass and the area of validity of the equation

The relevance of allometric equations is based on the proportional relation between relative growth of individual measurements. Thus, the biomass of a tree is related to its diameter (see Box 2). However, finding a statistical relation between biomass and explanatory factors is meaningless if this relation does not have a biological reality and its use can lead to error. The explanatory variables should be sought among those that influence tree growth.

Box 2:
The mathematical expression of the proportional relation between biomass (B) and diameter (D) is:

$$\frac{dB}{B} = a \times \frac{dD}{D}$$

which has the integral:

$$B = bD^a$$

where a and b are the parameters of the model.

Tree growth is the result of two processes: growth in length, resulting from the activity of buds, and growth in thickness, due to cambium activity. These processes depend on tree genetics, the stage of development (tissue age) and the environment (soil, atmosphere, competition from other trees, and human influence on this competition and on the tree itself). These factors can be translated into mathematical variables, such as:

- *station fertility*: determined through the age of the tree population and its dominant height H_0 (Eichhorn rule extended, even-aged and monospecific stand);
- *stand density*: (Hart-Becking spacing factor, Reinecke Density Index RDI, slenderness ratio H/D, robustness coefficient $D^{1/2}/H$) to reflect competition between trees;
- *social status of trees*: (H/H_0 or D/D_0, where D_0 is the dominant diameter of the population).

The introduction of fertility allows the extension of an allometric equation area of validity, and all of these factors can be used to determine the most accurate equation. Once the factors have been selected and the area of application determined (plot, forest, group of forests, range of species, etc.), the next step is choosing the most appropriate sampling method.

Trees sampling and stratification to improve the cost-accuracy compromise of the final model

Two sources of error can be calculated in the statistical modelling of biomass or volume: the ***sampling error*** due to the fact that only few trees are measured; and ***the prediction error***, which is related to the fact that the final model will always give an approximation of the volume or biomass. Measurement and data-entering errors cannot be determined statistically and the manual provides advice on how to avoid them. The sampling error can be reduced through the design of an appropriate sampling plan, i.e. the selection of the trees to be measured on the field. The prediction error is closely related to the construction of the model.

Sampling plan to determine the biomass of a tree

Sample size

The sample size must be adapted to the tree diversity of the stand. The more variability there is within a stand, the bigger the number of trees sampled should be. Thus, a monospecific even-aged stands (e.g. an industrial plantation) requires a smaller sample size than a tropical rainforest to achieve the same sampling error. In a monospecific stand, the trees will tend to have a similar carbon stock: they have the same age, are from the same species or clone and have similar social status (Figure 1, picture C). In a tropical rainforest (Figure 1, picture D) species diversity will be higher and social status and many other factors may be wide ranging.

Tree selection

The selection of the trees to be measured also has an influence on the sampling error. The theory shows that to build a linear regression between biomass and tree variables, the accuracy increases with the size of the sample, and it is preferable to sample trees across the whole range of values of the input variable (see Box 3). For example, when the input

variable is tree diameter, it is preferable to select a constant number of trees per basal area class, to better represent the biggest trees.

Box 3:
For the sampling step, the confidence interval of the predictions is an inverse function of the standard deviation of a selected explanatory variable X. The half amplitude of the confidence interval at a given threshold is illustrated by the following formula:

$$t_{n-2}\frac{\hat{\sigma}}{S_X\sqrt{n}} \text{ where } S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ and } \bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i$$

where t_{n-2} is the $1 - \alpha/2$ of a Student t-distribution with n-2 degrees of liberty, $\hat{\sigma}$ is the empirical standard deviation of the model's residuals, n is the size of the sample and S_X the standard deviation of the empirical explanatory variable X in the sample.

This formula shows that the bigger the sampling size (i.e $\sqrt{n}$ increases) the smaller the confidence interval (i.e. the accuracy is high). The standard deviation S_X follows the same trend, so sampling trees among the entire selected explanatory variable range leads to higher accuracy.

Stratification of the sample

In some cases it may be useful to add exogenous criteria to assist in the more efficient sampling of a population. All explanatory variables can be used: stand age (mainly plantations), fertility, station, silviculture, species, altitude, depth of groundwater, etc. According to these criteria, the inventory area can be divided into different zones (for example: rich soil or poor soil zones). This stratification is useful when the variability of biomass is higher on some strata than others. Increasing sampling effort in these strata will lead to greater accuracy for a given sample size (see Box 4).

Box 4:
Theoretically, the sampling size should be proportional to the standard deviation in each strata. In practice, the sampling size is often determined by technical, financial and human resources and by on field constraints.

Sampling to estimate the biomass of a stand

To estimate the biomass of a stand, individual tree selection becomes impractical. In stands, therefore, biomass estimates should be obtained by measuring all the trees in a given area called a plot and determining the number of plots to reach the desired sample size. As for forest inventories, the error is a function of the number of plots (see Box 5).

The evenness of the distribution of biomass in a stand should also be considered in the choice of plot size. For the same sample size, the error will be reduced in a stand in which the biomass distribution is uneven if a large number of smaller plots is used, whereas in the case of a regular spatial distribution of biomass, fewer, larger plots would be more efficient (Figure 4).

Box 5:
The number of plots (n), the sampling error (E) and the biomass coefficient of variation within the stand (CV_B) are linked by a mathematical formula. After simplifications, the formula gives the number of plots:

$$n \approx (\frac{2CV_B}{E})^2 + 1 \text{ where } CV_X = \frac{S_B}{\bar{B}}$$

$\bar{B}$ and S_B are respectively the biomass mean and standard deviation within the stand.

The coefficient of variation within a plot is a key element in building the sampling plan. It is linked to the plot area (A) by a power relation, as illustrated by the following formula:

$$CV_B = kA^{-c}$$

In this formula, c represents the biomass aggregation within the stand. If $c < 0.5$ the stand biomass is aggregated, or else its spatial distribution is regular.

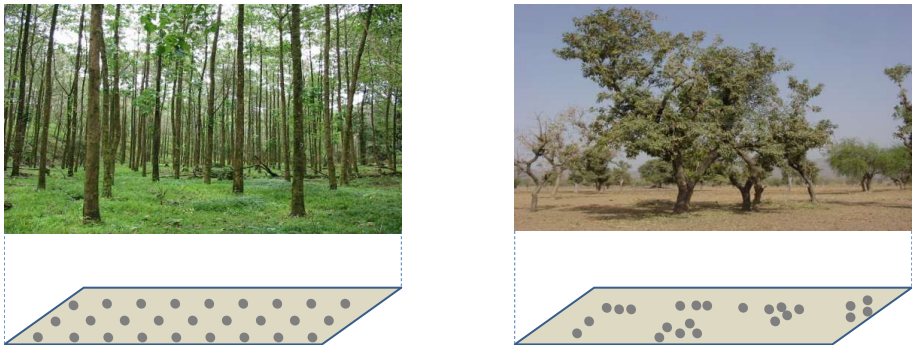


Figure 4. On the left the figure shows a regular spatial distribution of a tree population. On the right, the population is aggregated.

Finally, it is important to remember that apart from these tools to improve the accuracy of sampling, the construction of an allometric equation depends on many other financial, human, technical and environmental factors. Each will impose its own limits and the art of achieving the best equation lies in finding the optimal compromise.

Advice for field and laboratory data collection

Field errors are costly and cannot be corrected. To minimize them, three key principles should be followed:

→ it is better to weigh all tree compartments in the field;

→ for each sample, it is always better to weigh the total mass of the sample and of the sample at the same time to monitor moisture loss from the plant material;

→ biomass campaigns consume time and money, other measurements can be done to avoid going back on the ground (for example stem profile or mineralomass sampling).

Field measurements are usually destructive because of the need to weigh all the tree compartments on the site. However, it is not always possible (e.g. trees are too heavy, or there is a logging ban) or desirable (dry forests, for example) to fell all trees in a given area. Therefore, instructions will be given for direct measurement in the field, followed by advice on overcoming a lack of capacity for destructive field measurement.

Box 6: advice before going to the field.

→ In most cases, field constraints will lead to slight changes in the initial sampling plan.

→ Tree compartments may have different wood density and moisture content. Thus, it is better to measure each compartment individually.

→ To avoid bias when selecting the samples, only one person should select all of them.

→ If the samples are not measured in the field, they should be conserved in an ice chest.

→ The preparation of the equipment, data sheets and bags for the aliquots should be made before going to the field to save time during the field work.

Organization of the work in the case of direct measurement in the field

To avoid time loss while doing fieldwork, an organizational structure with seven operations (Figure 5) has been developed by France's *Institut National de la Recherche Agronomique* (INRA). The number of people per operation and the chronology of the various activities involved are detailed in the manual.

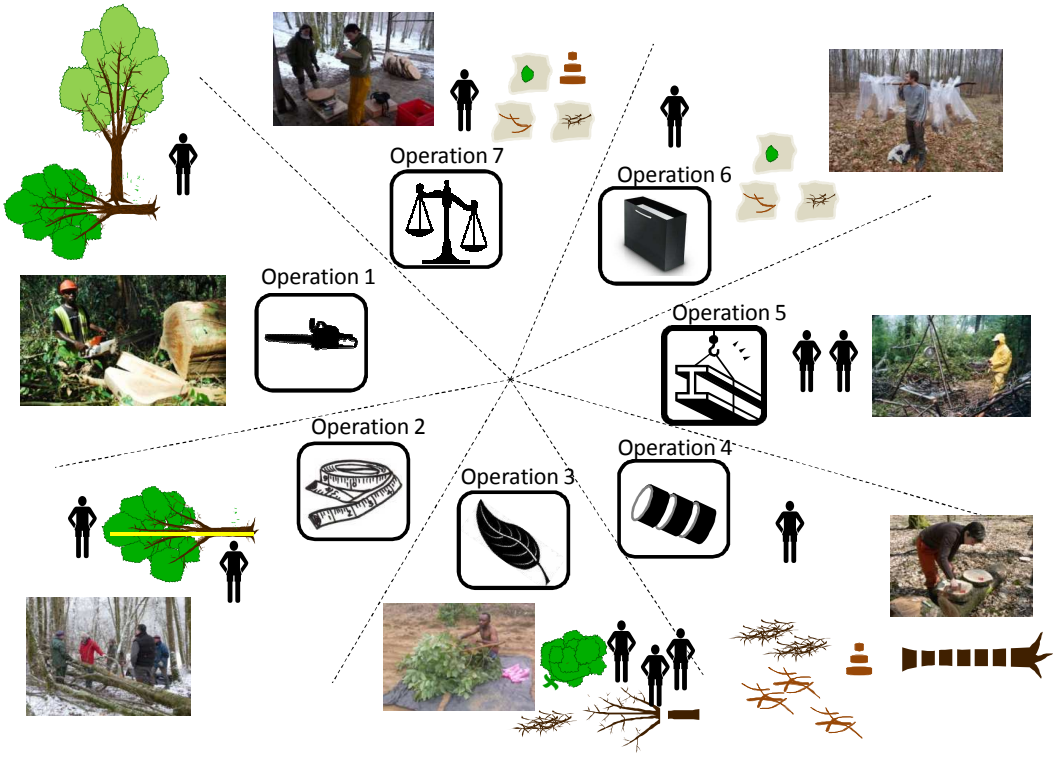


Figure 5. Operation 1: preparation of the ground and tree felling (photo: M. Henry). Operation 2: measurement of felled trees (stem profile, position of the short-logs) (photo: M. Rivoire). Operation 3: defoliation and delimbing (photo: R. D'Annunzio). Operation 4: crosscut and labelling of short-logs (photo: C. Nys). Operation 5: weighing of short-logs and bunch of branches (photo: B. Locatelli). Operation 6: Branches sampling (photo: M. Rivoire). Operation 7: weighing of the samples (photo: M. Rivoire).

When not all measurements can be destructive

Several methods exist to avoid the destruction of whole trees. Two cases are presented here - for dry forests and for trees too large to be fully weighed in the field.

In the case of ***dry forests***, trees are pruned regularly but are too few to be felled. Thus, the pruned branches are used to calculate the biomass of the leaves (in practice at least three leaves from three different branches are required to form the aliquot) as well as biomass and fresh density of branches. On the other parts of the tree, the diameter and length are measured (Figure 6). When this information is combined with the density of pruned elements and the volume of all compartments, the total biomass of the tree can be assessed.

In the case of very ***large trees***, a complete weighing on the ground may be impractical and only the biomass of the branches with a diameter under 10 cm (this value can be modified to better reflect the reality of the studied forest) is measured. The branches with a diameter greater than 10 cm have their biomass deduced from their volume (diameter and length measured in the field, see Figure 7) and the tree average density.

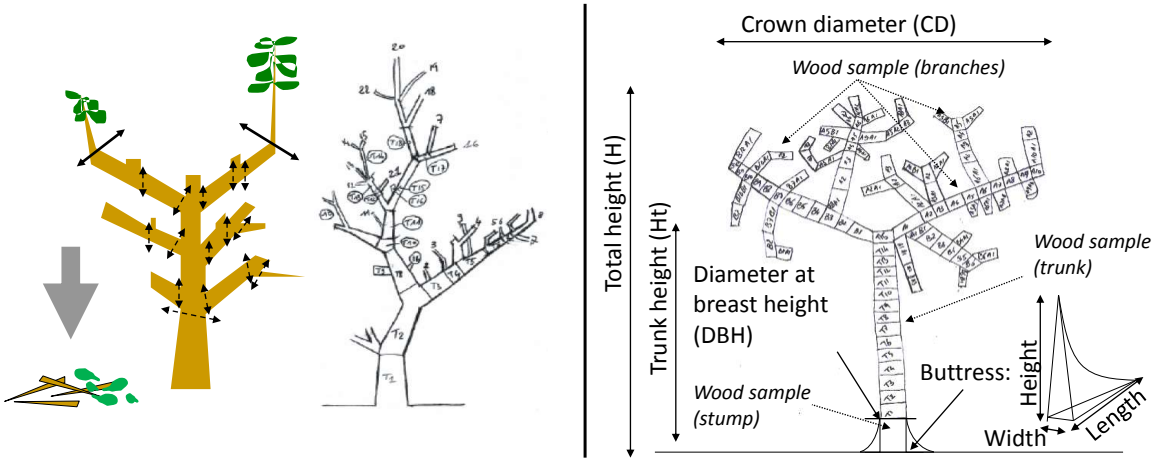


Figure 6. On the left, tree pruning and measurement is shown and non-pruned tree parts are numbered. On the right, non-cut tree compartments are numbered for their length and basal diameter measurement.

Laboratory work

The laboratory work involves measuring the volume, fresh and dry mass of samples. The aliquots are dried in an oven (Figure 7). the leaves and fruits are dried at 70 °C (65 °C if chemical analyzes are included) and, for wood biomass only, the samples can be dried at 105 °C.

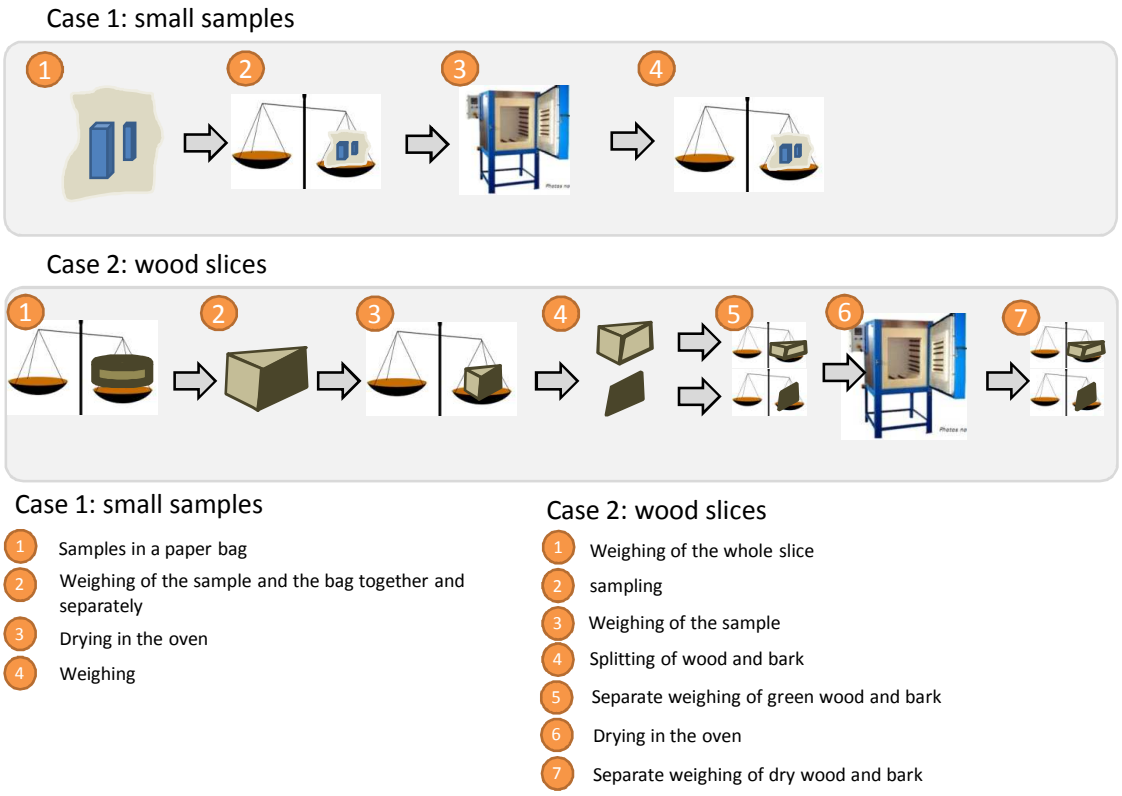


Figure 7. The laboratory steps of aliquots weighing and drying.

Data entering and formatting

Once fieldwork is completed, the preparation and formatting of the data is a relatively simple but crucial step to limit errors during the model-fitting.

Data tables should be nested when several levels of data are considered, and each table should include one row per data stratum: for example a table with one row per tree (if we consider a set of trees) can be connected to a table with raw per-plot data at the forest level. This table can itself be nested in a table for a region, where each row represents a forest.

To ensure a good understanding of the data, it can be useful to add information about it (i.e. meta-information), such as the nature of the variables, measurement date, and the name of the person who entered the data. When entering data, the qualitative variables must be clearly distinguished from the quantitative variables. It is preferable to use codes with few characters to avoid typing errors.

Figure 8. Operators entering data (Photo: S. Giaccio).

An expensive but reliable method is to have two operators enter the same data (Figure 8). In all cases, clearance procedures can be used to identify entry errors, such as unrealistic values, unwanted spaces, double spaces, and capitalization errors (the latter being the same modality input with and without capital letters). Automatic clearance is also possible through the creation of dedicated scripts. The graphical representation of variables in pairs can also help in the detection of aberrations.

Graphical exploration of the data

When the dataset is ready, the graphical exploration of the data is used in the search for relevant explanatory variables to model biomass. The results of a model are composed of two terms: the ***mean*** and the ***error*** (or ***residual***). The graphical exploration shows both the form of the mean and the residual without determining the model parameters (this is the purpose of the model-fitting section, below). The objective of this section is therefore not to find the model that best explains the variability of the dataset but to retain the three or four associations between variables that appear most relevant. The method consists of representing the variables in pairs and observing if the scatter plots follow a particular shape, linear or not, and if the variance of the residuals is constant. Figure 9 illustrates the four possible form types.

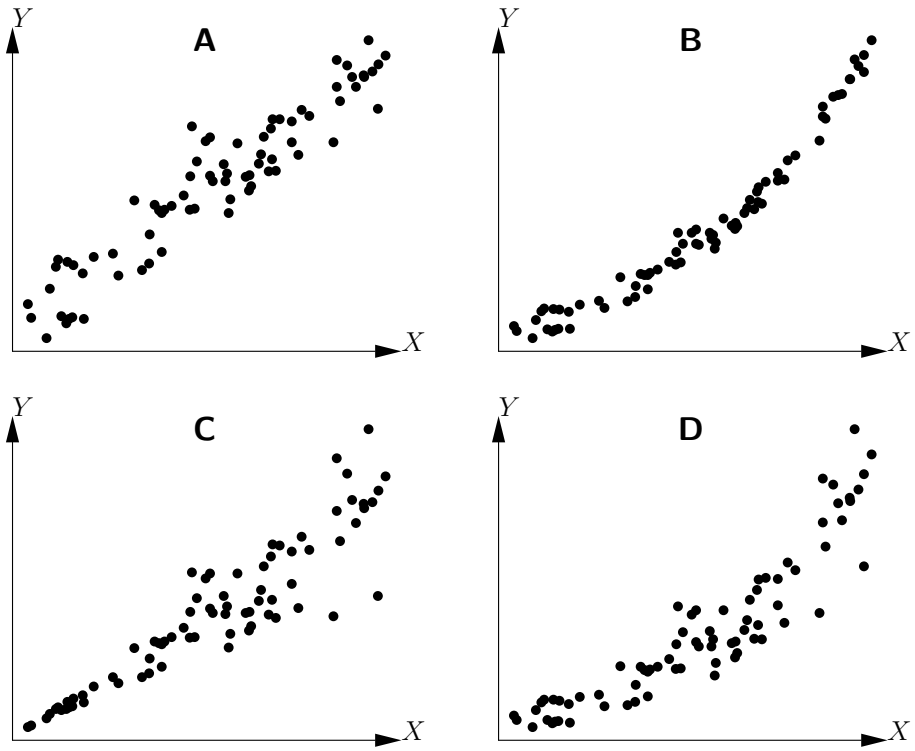


Figure 9. A: linear relation, residuals' variance constant. B: non-linear relation, residuals' variance constant. C: linear relation, residuals' variance not constant. D: non-linear relation, residuals' variance not constant.

The variable to be explained can also be plotted with:

→ coupled variables from the basic variables (for example D^2H);

→ transformed variables, the most common of which is the logarithmic transformation

to obtain a linear relation from a power relation (it is easier to visually distinguish whether a relation is linear rather than to identify the type of relation from curves);

→ variables created from several variables, interdependent or not (e.g. sum, multiplication sum and multiplications).

The graphical exploration step is also useful for detecting (1) modelling artefacts, i.e. some non-linear scatter plot forms but with a high coefficient of linear correlation (R^2); and (2) large errors in data entery.

Fitting the allometric equation

After identifying the variables that could generate a relevant model, statistical tests are used to confirm or modify the choice of these variables and to determine the parameters of the model. All these tools are based on the behaviour of the model residuals. The first part of this section is devoted to them, followed by a more complete presentation of the various possibilities of model-fitting.

Definition and hypothesis on residuals

The model results are called predictions. In this case, the dependent variable is biomass, so there will be as much biomass measured in the field as biomass predicted by the model using explanatory (also called independent) variables (diameter, height, density, species, etc.). The difference between each observation and its prediction is called residual, so there is as much residuals as observations (see Box 7).

Box 7:
For each observation (B_i) of the dependent variable there is a prediction ($\hat{B}_i$) by the model and a residual (ε_i) defined by:

$$\varepsilon_i = B_i - \hat{B}_i$$

The statistical tools used to determine the parameters of a model are based on the behavior of these residuals and the verification a posteriori of three hypotheses:

1. ***Independence of residuals:*** this assumption is generally not verified by calculation; it should be verified when sampling trees. These trees should be far enough apart to ensure that the choice of one tree does not influence the choice of the next tree. In case of doubt, the Durbin-Watson test can be used to check the independence of residuals.

2. ***Normal distribution of residuals:*** this hypothesis can be verified by the graphic representation of the empirical quantiles with the theoretical quantiles. In this graphical representation, the points must be aligned approximately along a straight line (Figure 10). Statistical tests can also be used to verify the normality of residuals.

3. ***Constant variance of residuals:*** this is also known as homoscedasticity and heteroscedasticity when the residuals' variance is not constant. Homoscedasticity of residuals' variance is controlled by the graph of residuals against predicted values, in which the scatter plot does not show any particular trend or structure (Figure 10)

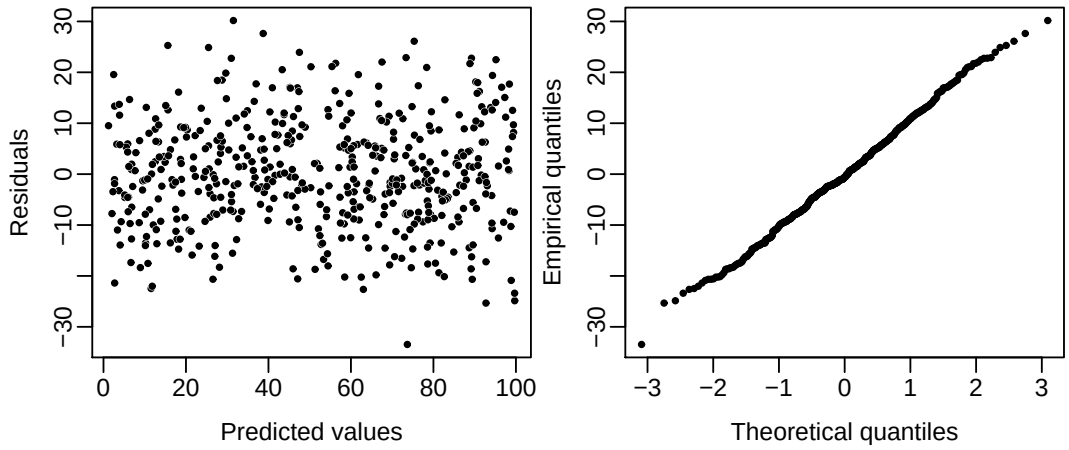


Figure 10. On the left, the residuals' variance is constant (homoscedasticity). On the right, the residuals follow a normal distribution.

Because of the ***robustness*** of the model types presented below, the second and third hypotheses can be verified by simple graphical visualization. The quality of these models is sufficiently high, even if the hypotheses are not completely verified.

Several methods for model-fitting

Several types of model-fitting are presented in this section, within two broad categories: (1) simple, multiple or weighted linear regressions and those with a modelling of the variance, and also non-linear regressions with a known or to be estimated exponent; and (2) the selection between pre-existing nested models to determine the same or a different dependent variable. Table 1 shows the statistical method adapted to each type of fitting.

Methods for linear and non-linear regressions

The ***least squares*** method consists of the calculation of the sum of the squares of the residuals, and then the determining of the set of parameters that minimizes this sum. This method is relatively easy but is only suitable for the simplest models: i.e. single or multiple linear regressions.

In the case of the ***weighted least squares*** method, a positive weight is associated with each observation, defined by a proportional relation with the inverse of the residuals' variance. This relation can be simplified for biological data such as biomass or volume, leading to the determination of an additional exponent compared to the previous method.

The ***maximum likelihood*** method consists of the calculation of the argument that maximizes the likelihood of the observations. The likelihood of an observation is the probability density to obtain this observation under the specified model. In this case, the dependent variable is assumed to follow a normal distribution and its standard deviation depends on three parameters (θ,k,c). The method consists of finding the values of these three parameters that maximizes the likelihood.

In the case of non-linear models, either with the weighted least squares or the maximum likelihood method, a numerical optimization is necessary to verify that the minimum (for weighted least squares) or maximum(for maximum likelihood) is achieved. The manual describes several optimization processes.

Finally, during the fitting step, several linear and non-linear models may correspond to observations. To assist with the choice of the final model, Table 2 presents several of the

advantages and disadvantages of multiple linear and nonlinear models.

Fitting methods	**Method**	**Equation form**	**Residuals distribution**
Linear regression:			
Simple	Last squares	$Y = a + bX + \varepsilon$	$\varepsilon \sim \mathcal{N}(0,\, \sigma)$
Multiple	Last squares	$Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_pX_p + \varepsilon$	$\varepsilon \sim \mathcal{N}(0,\, \sigma)$
Weighted	Weighted last squares	$Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_pX_p + \varepsilon$	$\varepsilon_i \sim \mathcal{N}(0,\, \sigma_i)$
With a variance modeling	Maximum likelihood	$Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_pX_p + \varepsilon$	$\varepsilon \sim \mathcal{N}(0,\, kX_1^c)$
Non linear regression:			
With a known c exponent	Weighted last squares	$Y = f(X_1, X_2, \ldots, X_p; \theta) + \varepsilon$	$\varepsilon_i \sim \mathcal{N}(0,\, \mathrm{Var}(\varepsilon))$ $\mathrm{Var}(\varepsilon) = g(X_1, \ldots, X_p; \vartheta)$
With c exponent to be determined	Maximum likelihood	$Y = f(X_1, X_2, \ldots, X_p; \theta) + \varepsilon$	$\varepsilon_i \sim \mathcal{N}(0,\, \mathrm{Var}(\varepsilon))$ $\mathrm{Var}(\varepsilon) = g(X_1, \ldots, X_p; \vartheta)$
Selection between preexisting models:			
Nested (linear)	Fisher distribution	Tests if the parameters added by the full model to the nested one are different from 0	
Nested (non-linear)	χ^2 distribution	Tests if the parameters added by the full model to the nested one are different from 0	
Same dependent variables	AIC	The best model will have the minimal AIC value	
Different dependent variables	Furnival index (F)	The best model will have the minimal F value	

Table 1. Various statistical tests used to fit tree biomass or volume models. All symbols and formulae are explained in the manual.

Selection between pre-existing models

In the case of ***nested models***, the selection consists of testing whether the additional parameters of the most complete model are significant. The null hypothesis is rejected if the p-value of the test (Fischer for linear models or χ^2 for non-linear models) is less than the significance level (the full model is best in this case).

In the case of ***models with the same dependent variable***, the best model is that which minimizes the Akaike information criterion (AIC). This criterion is based on the likelihood of the model used.

Finally, to compare ***models with different dependent variables*** (one being a transformed variable of the other), the Furnival index can be used. This index is defined for a model whose residuals have a constant variance without taking into account the transformation of variables. The model with the smallest value of F will be considered the best.

	Advantages	Disadvantages
Multiple regression	Explicit expression of the coefficients of the model.	Restrictions on the form of the residuals, small flexibility in the form of the model.
Non-linear regression	No restrictions on the form of the model-fitting, the mean or the variance.	No explicit expression of the model's coefficients, leading to risks of a wrong estimation of the parameters.

Table 2. Advantages and disadvantages of multiple linear and non-linear models.

Stratification and aggregation of observations

When the dataset is stratified, two methods can be used for the fitting: (1) each stratum can be fitted separately and the methods described above can be used; or (2) the full dataset can be analysed globally. In the second case, the analysis used is called ***analysis of covariance***. It is assumed that all residuals have the same variance within and between the different strata. A Fisher test is then used to test whether the parameters of the model are significantly different from 0. When the number of strata is too large, such as when the stratification is by tree species in a diverse stand, the manual presents other solutions. Finally, the manual presents solutions for integrating models for each tree compartment into a single model, such as through the use of ***multivariate models***.

The tools presented in this section can be used not only to test different models from a dataset and determine the most accurate equation (Figure 11), but also to compare several existing models to choose that one that best matches the study area.

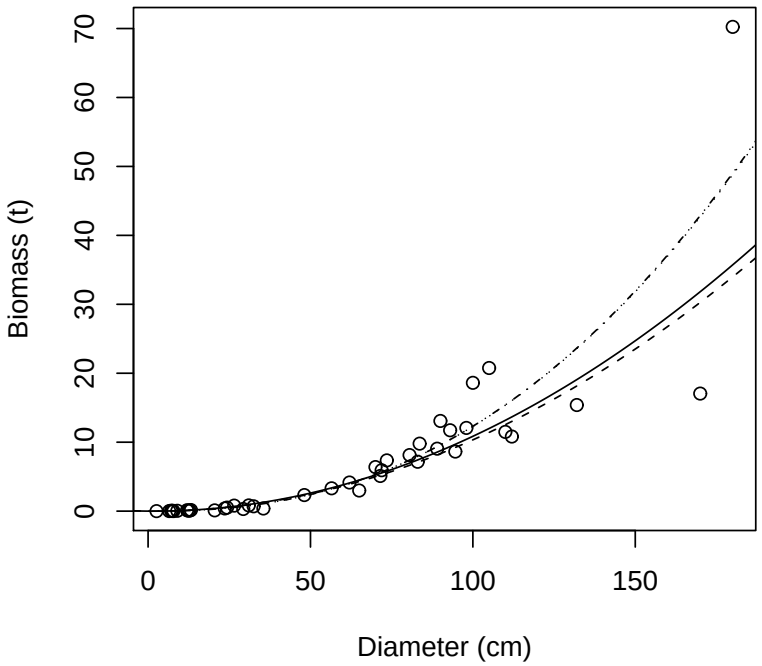


Figure 11. Four biomass models have been fitted from a 42 trees dataset measured in Ghana (Henry et al., 2010)[2].

2. **Henry, M., Besnard, A., Asante, W.A., Eshun, J., Adu-Bredu, S., Valentini, R., Bernoux, M. et Saint-André, L.**, 2010. Wood density, phytomass variations within and among trees, and allometric equations in a tropical rainforest of Africa. Forest Ecology and Management, 260(8): 1375-1388.

Validation of the model and control of the predictions

Once the final model has been fitted, it can be ***validated*** with another dataset independent of the one used for the model-fitting. Several indicators can be used to compare model predictions with observations: the sum of the squares of residual differences, the residual variance, the adjusted residual error, the R^2 and the AIC. Finally, it is important to remember that the predictions of any model have an intrinsic variability. To express this variability, it can be associated with an uncertainty indicator: ***confidence interval at 95 %*** (Figure 12). The methodology for calculating this interval depends on the shape of the model (linear or nonlinear) and whether the prediction is based on an average tree in the stand or a random tree. The various possible methodologies are presented in the manual.

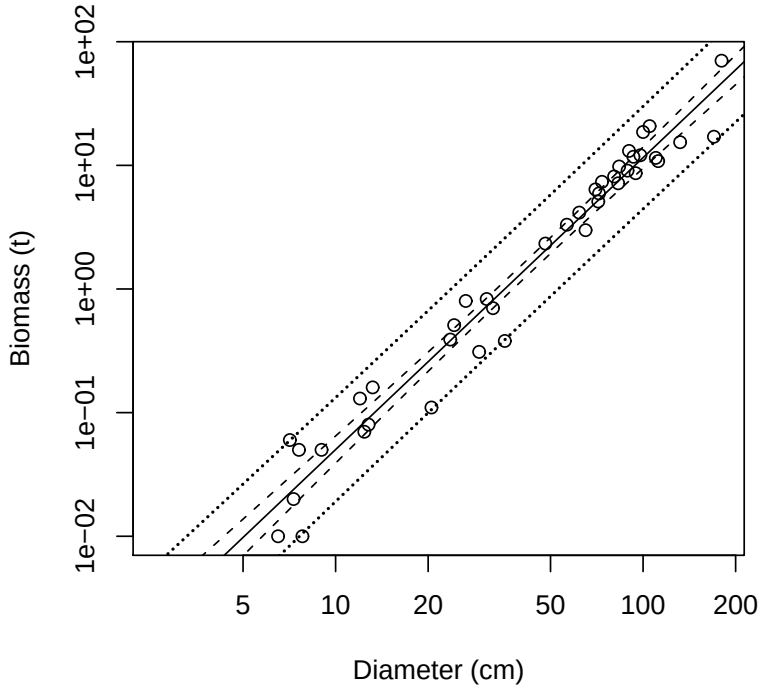
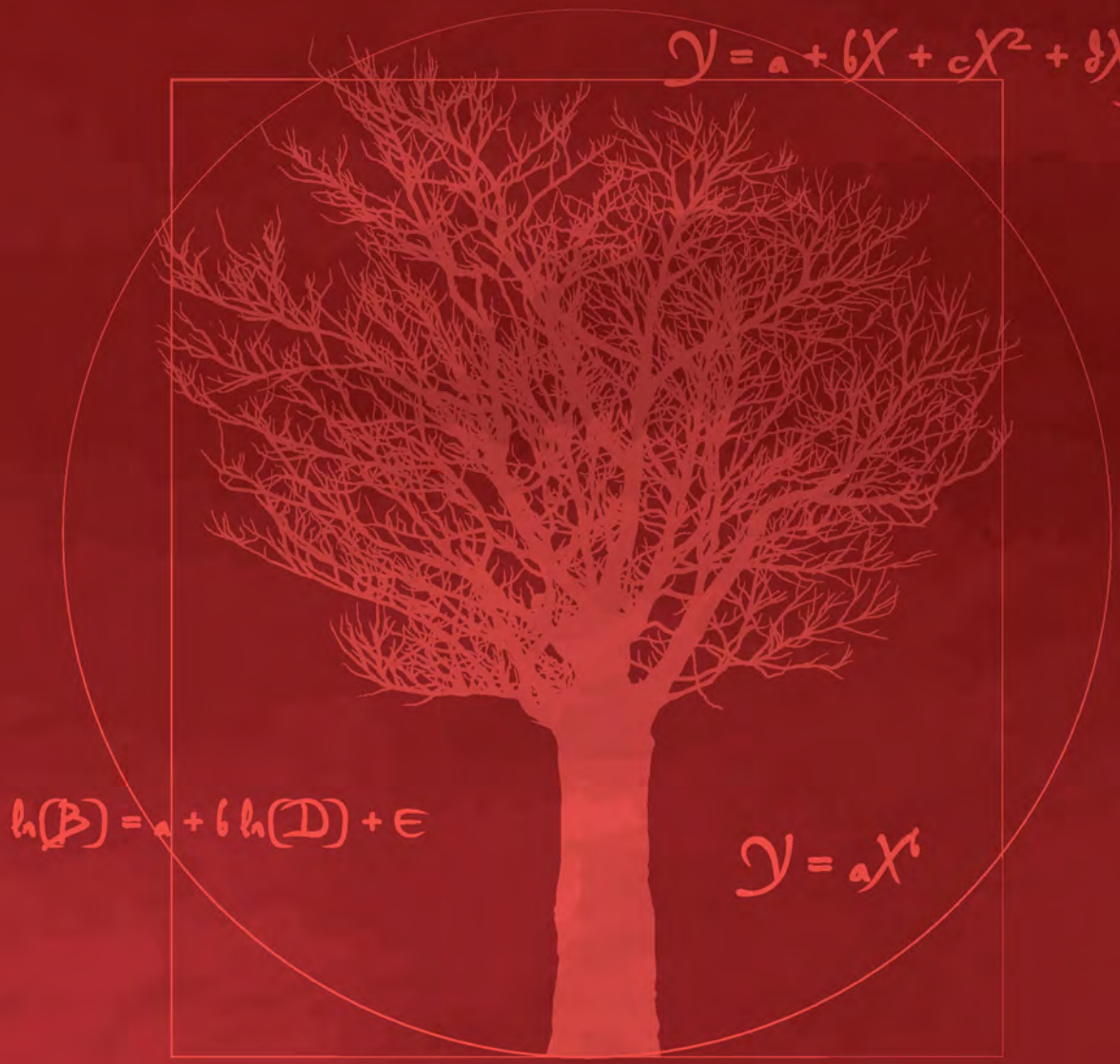


Figure 12. Biomass prediction (line) with its confidence interval, whether the prediction is based on an average tree in the stand (dot) or a random tree (dash), fitted from a 42 trees (points) measured in Ghana.

Conclusion

The ultimate goal of the manual is to provide tools to increase the accuracy of forest carbon stock estimates. To achieve this, it is important to ensure accuracy at each stage of the overall process used to develop the equations: sampling design, fieldwork, measurement and statistical analyses. The statistical tools and methodologies presented in this manual are for the construction of allometric equations to provide a holistic approach to forest carbon stock estimates. Not all possible cases are presented because some situations face specific technical and scientific challenges. However, statistics on this subject are evolving, and much progress has been made in this area. As accurate allometric equations become more widely used, new opportunities will be able to be explored to improve carbon accounting. For example, the creation and maintenance of databases of all existing equations could allow users to quickly find a specific equation adapted to their context.



E-ISBN 978-92-5-107347-6