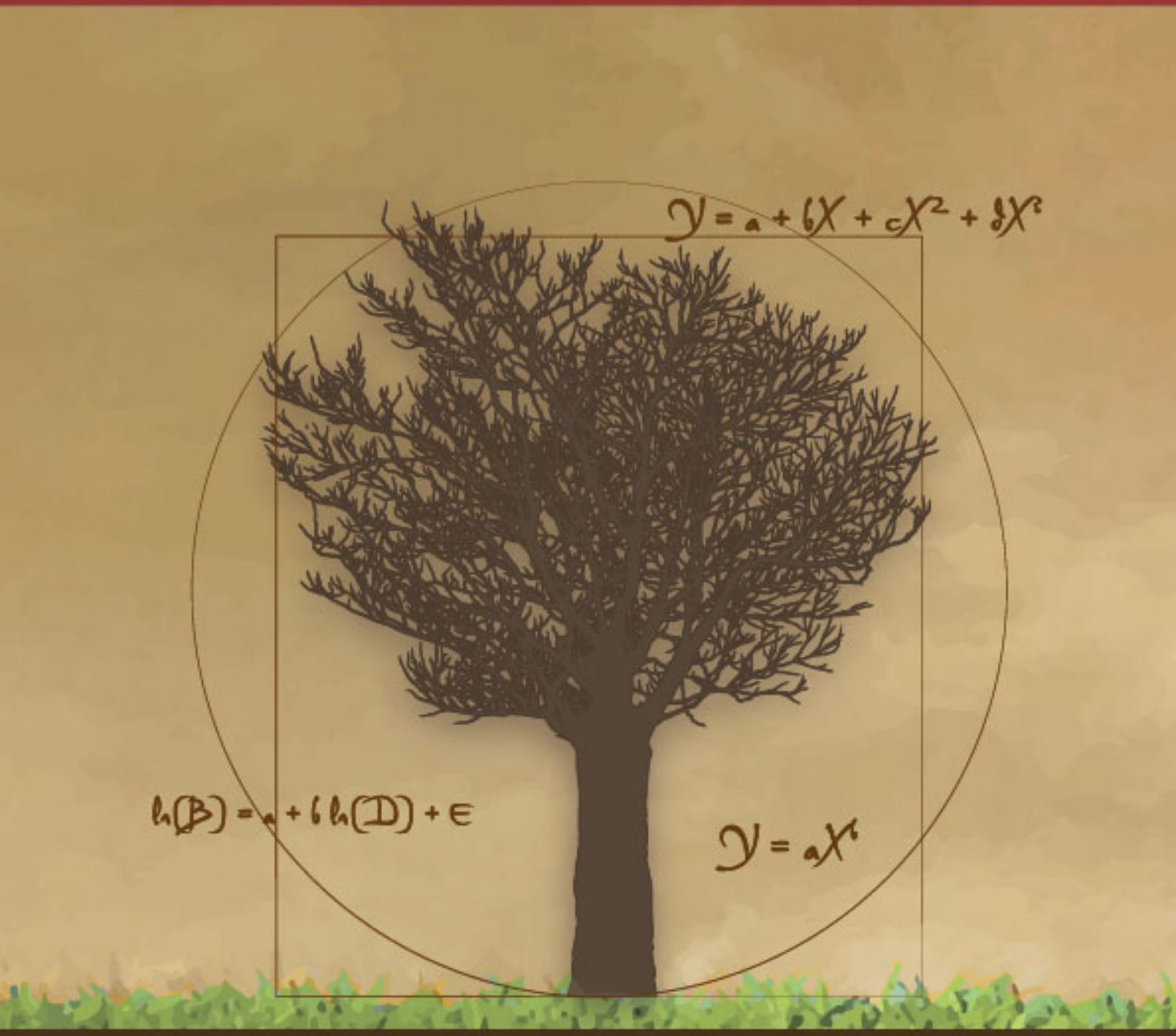
Resumen del manual de construcción de ecuaciones alométricas para estimar el volumen y la biomasa de los árboles

Del trabajo de campo a la predicción



cirad

Resumen del manual de construcción de ecuaciones alométricas para estimar el volumen y la biomasa de los árboles: del trabajo de campo a la predicción

Gael Sola,
Departamento Forestal,
Organización de las Naciones Unidas para la Alimentación y la Agricultura

Nicolas Picard,
Departamento del Medioambiente y Sociedad,
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,
Departamento Forestal,
Organización de las Naciones Unidas para la Alimentación y la Agricultura

Las denominaciones empleadas en este producto informativo y la forma en que aparecen presentados los datos que contiene no implican, de parte de la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO) y el *Centre de Coopération Internationale en Recherche Agronomique pour le Développement* (CIRAD), juicio alguno sobre la condición jurídica o nivel de desarrollo de países, territorios, ciudades o zonas, o de sus autoridades, ni respecto de la delimitación de sus fronteras o límites. La mención de empresas o productos de fabricantes en particular, estén o no patentados, no implica que la FAO y el CIRAD los aprueben o recomiende de preferencia a otros de naturaleza similar que no se mencionan.

Las opiniones expresadas en esta publicación son las opiniones de los autores y no reflejan necesariamente las opiniones de la FAO y el CIRAD.

Todos los derechos reservados. La FAO y el CIRAD fomentan la reproducción y difusión parcial del material contenido en este producto informativo. Su uso para fines no comerciales se autorizará de forma gratuita previa solicitud. La reproducción para la reventa u otros fines comerciales, incluidos fines educativos, podría estar sujeta a pago de derechos o tarifas. Las solicitudes de autorización para reproducir o difundir material de cuyos derechos de autor sea titular la FAO y al CIRAD y toda consulta relativa a derechos y licencias deberán dirigirse por correo electrónico a copyright@fao.org, o por escrito al Jefe de la Subdivisión de Políticas y Apoyo en materia de Publicaciones, Oficina de Intercambio de Conocimientos, Investigación y Extensión, FAO, Viale delle Terme di Caracalla, 00153 Roma (Italia).

Las Naciones Unidas para la Alimentación y la Agricultura (FAO)
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

Citación recomendada: Sola G., Picard N., Saint-André L., Henry M. 2012. Resumen del manual de construcción de ecuaciones alométricas para estimar el volumen y la biomasa de los árboles: del trabajo de campo a la predicción. Las Naciones Unidas para la Alimentación y la Agricultura y el Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Rome, Montpellier, 17 pp.

© CIRAD et FAO, 2012

Preámbulo

El manual de construcción de ecuaciones alométricas para la estimación de volumen y biomasa de árboles[1] ha sido diseñado para estudiantes, investigadores e ingenieros que deseen adquirir los conocimientos y la metodología necesaria para desarrollar ecuaciones alométricas para medir volumen, biomasa o mineralomasa de los árboles. No se requieren muchos requisitos o conocimientos previos gracias al empleo de numerosos ejemplos y fichas técnicas que acompañan el manual. Esta síntesis quiere ilustrar la estructura general de la obra e indicar las principales etapas para desarrollar las ecuaciones alométricas.

Un manual para mejorar las estimaciones de volumen y biomasa de los bosques

Los bosques proporcionan muchos servicios, como por ejemplo la producción de madera, leña y el almacenamiento de carbono. Ellos requieren una estimación rigurosa del volumen y de la biomasa de los árboles. En la escala de parcela, bosque o grupo de bosques, la exactitud de esta estimación sigue siendo un desafío no solo por la capacidad técnica, humana y financiera, sino también por las herramientas científicas para obtener y analizar los datos.

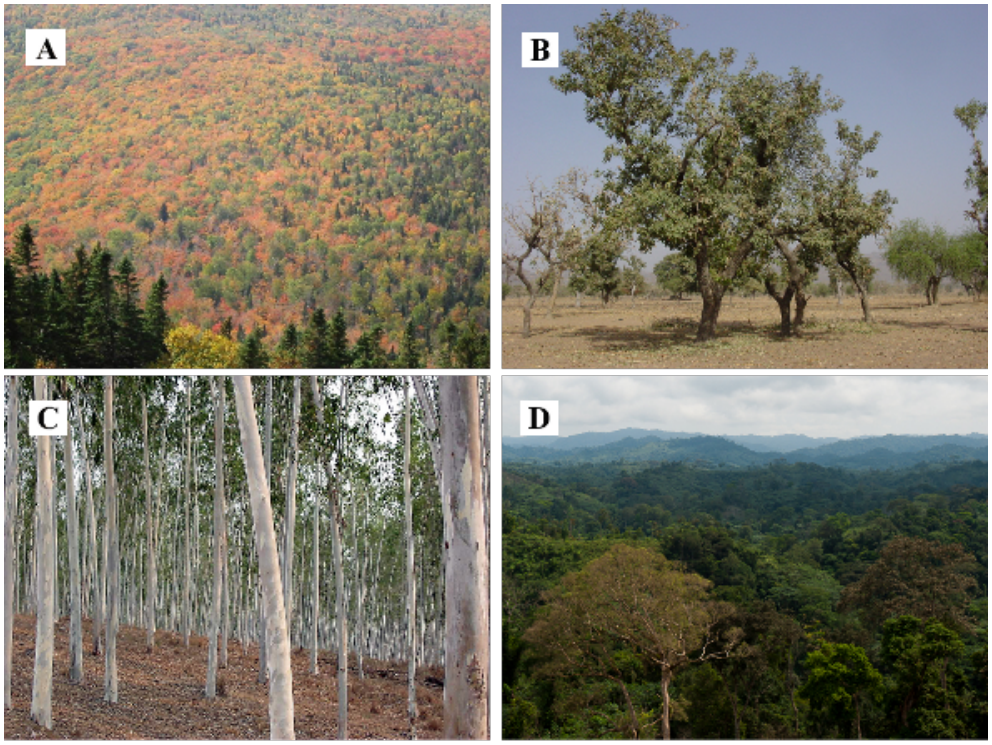


Figura 1. A: bosque templado continental (foto: Bruno Locatelli); B: bosque seco Tropical (foto: Régis Peltier); C: plantación de Eucalyptus (foto: Bruno Locatelli); D: selva Tropical (foto: Gael Sola).

La demanda y la conciencia de estos servicios suscitan sin embargo un fuerte interés en

[1] **Picard N., Saint-André L., Henry M.**, 2012. Manual de construcción de ecuaciones alométricas para estimar el volumen y la biomasa de los árboles: del trabajo de campo a la predicción. Las Naciones Unidas para la Alimentación y la Agricultura, y el Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Rome, Montpellier, 218 pp.

el campo de la investigación forestal y estimulan la mejora de las herramientas técnicas. La variabilidad de la biomasa en función del tipo de bosque (Figura 1), de las especies y del compartimiento de árbol y la imposibilidad de medir el conjunto completo de arboles pone un problema al cual las matemáticas, la biología, y las ciencias forestales pueden dar respuestas precisas.

Las reservas de carbono se estiman a partir de la biomasa del árbol, es decir, a partir de la masa seca de materia orgánica. Para obtener el valor de biomasa es necesario por tanto pesar todos los elementos constitutivos de un árbol (Figura 2). Estas medidas se hacen difíciles o imposibles a escala de bosque por dos razones: (1) se trata de mediciones destructivas y, por lo tanto, es indeseable o está prohibido efectuarlas a gran escala, (2) el costo y el tiempo necesario para el trabajo son excesivos. Con respecto a la biomasa subterránea, además, el trabajo de medición es aun más desafiante y, muchas veces, en lugar de medirla directamente, se calcula a través de factores de multiplicación específicos a partir de la biomasa aérea.

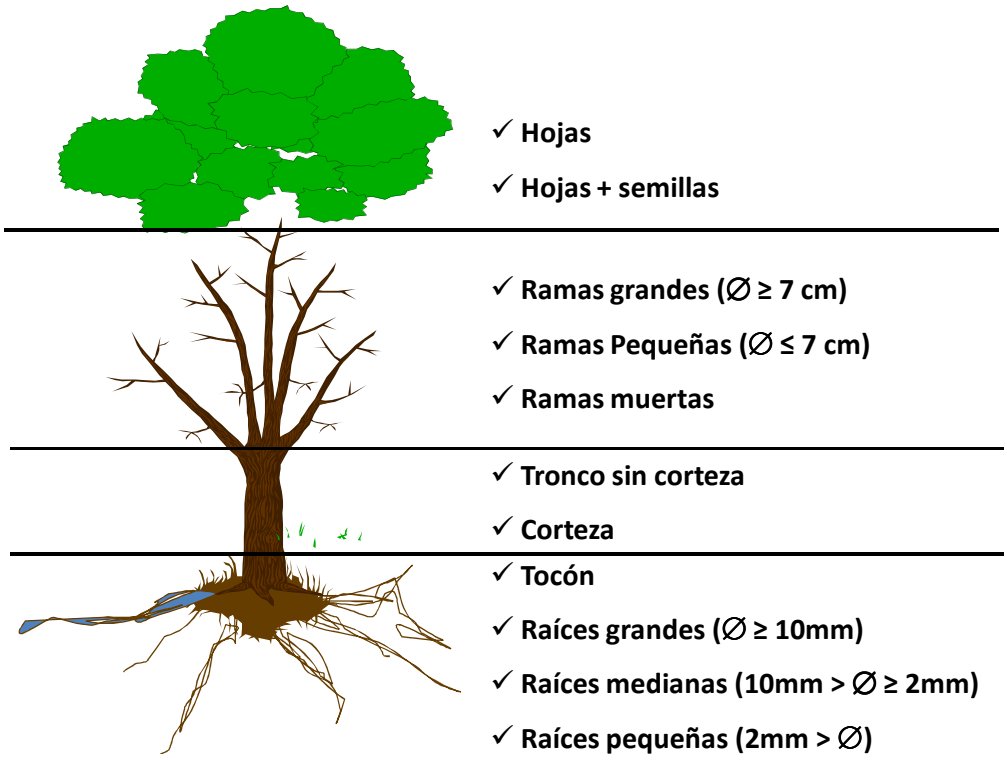


Figura 2. Los diferentes compartimentos de un árbol.

Otro método más accesible para obtener una estimación de la biomasa de un árbol consiste en utilizar la relación existente entre el diámetro del árbol y de la biomasa. Es en este nivel las ecuaciones entran en juego. La alometría estudia las relaciones existentes entre las características dimensionales de los individuos en una población. De hecho existe una relación estadística entre cantidades fáciles de medir a gran escala (diámetro, altura, densidad de madera) y variables difíciles de medir, como la biomasa o el volumen. Por lo tanto las mediciones costosas y destructivas podrán ser limitadas a una muestra de árboles y tendrán como objetivo identificar estadísticamente los parámetros alométricos para todo el conjunto de arboles presente en el área de investigación (ver el recuadro 1).

Ricuadro 1:
Los parámetros pueden ser ajustado para conseguir una relación de potencia entre la biomasa y el diámetro del árbol:

$$\underbrace{Biomasa}_{\substack{\text{variable a}\\ \text{explicar}}} = b \times \underbrace{\text{Diámetro}}_{\substack{\text{variable}\\ \text{explicativa}}}{}^{a}$$

El ajuste consiste en determinar estadísticamente los parámetros a y b que dan la mejor relación entre la biomasa y el diámetro.

Para mejorar el conocimiento sobre las ecuaciones alométricas y facilitar su desarrollo, los autores del manual proponen una metodología estructurada en siete etapas (Figura 3), a partir de la selección de las variables explicativas hasta la determinación de la mejor ecuación alométrica, pasando por las mediciones de campo. Gracias a las herramientas estadísticas que se presentarán, será posible calcular también el error en la predicción.

Figura 3. Los diferentes pasos del método descrito en el manual.

Selección de los factores que influyen en la biomasa de los árboles y zona de validez de la ecuación

La validez de las ecuaciones alométricas está basada en el hecho de que existe una relación proporcional entre los crecimientos relativos de los componentes de un individuo. La biomasa de un árbol está relacionada con su diámetro (ver Recuadro 2). Pero encontrar

una relación estadística entre la biomasa y los factores explicativos tendría poco significado si esta relación no tuviera una realidad biológica, por lo que su uso podría conducir a errores. Las variables explicativas entonces tienen que ser buscadas entre aquellos parámetros que efectivamente tienen una influencia sobre el crecimiento de los arboles.

Recuadro 2:
La relación de proporcionalidad entre la biomasa (B) y el diámetro (D) se expresa como:

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

Se intégra en:

$$B = bD^a$$

donde a y b son parámetros del modelo.

El crecimiento de los árboles es el resultado de dos procesos: el crecimiento en longitud, debido a la actividad de los brotes, y el crecimiento diametral, causado por la actividad del cámbium. Este crecimiento depende del patrimonio genético de los árboles, del estadio de desarrollo (envejecimiento de los tejidos) y de los factores ambientales (suelo, atmósfera, la competencia de otros árboles, la influencia humana). Es posible traducir estos factores en variables matemáticas, como por ejemplo:

- La fertilidad de la estación a través de la relación entre la edad de su población y su altura dominante H_0 (ley Eichhorn extendida, válida para poblaciones de la misma edad y mono específicas),
- La densidad de la población (factor de espaciamiento de Hart-Becking, Índice de Densidad de Reinecke, RDI, coeficiente de esbeltez, H/D, coeficiente de solidez $D^{1/2}/H$) para tener en cuenta la competencia entre árboles,
- El estado social de los árboles (H/H_0 o D/D_0, donde D_0 es el diámetro dominante de la población).

La introducción de la fertilidad permite ampliar la zona de validez de una ecuación y todos estos factores se pueden utilizar para determinar la ecuación más exacta. Una vez seleccionados los factores y el área de aplicación de la tarifa determinada (parcela, bosque, conjunto de bosques etc.), el siguiente paso es definir la estrategia de muestreo más adecuada.

Toma de muestras y estratificación de los árboles para mejorar la relación entre costo y precisión del modelo final

Dos fuentes de error pueden ser enmarcadas estadísticamente en el proceso de modelización de biomasa o volumen. El primero error es asociado con el hecho de que sólo algunos árboles se miden, es decir el ***error de muestreo***. El segundo error es asociado con el hecho de que el modelo final dará siempre una aproximación del volumen o de la biomasa, que es el ***error de predicción***. Los errores de medición y en la entrada de los datos no pueden ser determinados o enmarcados estadísticamente y, por eso, el manual ofrece muchos consejos para evitarlos. Relativamente al error de muestreo esto puede reducirse durante el diseño del plan de muestreo, es decir durante la selección de los árboles que serán parte de la muestra. El error de predicción está estrechamente relacionada con la construcción del modelo.

Toma de muestras para determinar la biomasa de un árbol

Tamaño de la muestra

Cuanto más variable es la población, mas grande tiene que ser el número de árboles a muestrear. Por lo tanto, una población mono-específica y coetánea (una plantación industrial, por ejemplo) requiere un tamaño de muestreo más pequeño que un bosque tropical para el mismo error. En el primer caso los árboles son más similares entre ellos a nivel de stock de carbono: tienen la misma edad, son de la misma especie o el mismo clon y tienen casi todos el mismo estado social (ver la Figura 1 foto C). En la selva tropical (ver la Figura 1 foto D) el número de especies diferentes es muy alto, hay mucha diferenciación social, etc. El tamaño de la muestra debe ser escogida en función de la variabilidad de los árboles en el rodal.

Selección de los árboles

La selección de los árboles en la muestra también tiene una influencia sobre el error de muestreo. La teoría muestra que en el caso de una regresión lineal (1) la precisión aumenta con el tamaño de la muestra y (2) es preferible muestrear árboles en todo el rango de valores de la variable de entrada (ver el Recuadro 3). Por ejemplo, cuando la variable de entrada seleccionada es el diámetro, es preferible seleccionar un número constante de árboles por clase de diámetro, para representar mejor los árboles de gran diámetro.

Recuadro 3:
Durante el muestreo, la amplitud del intervalo de confianza de la predicción es una función inversa de la desviación estándar de la variable de entrada seleccionada. La mitad de la amplitud al umbral α del intervalo de confianza de esta regresión se define por la fórmula:

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

donde t_{n-2} es el cuantil $1-\alpha/2$ de la ley de distribución de Student con $n-2$ grados de libertad, $\hat{\sigma}$ es la desviación estándar empírica de los residuos del modelo, n es el tamaño de la muestra, y S_X la desviación estándar empírica de la variable de entrada X en la muestra.

En esta fórmula es posible constatar que cuanto más grande es el tamaño de la muestra ($\sqrt{n}$ aumenta) más pequeño es el intervalo de confianza de la predicción (es decir, más alta es la precisión). Lo mismo se verifica con la desviación estándar S_X, es decir que es preferible muestrear los árboles en todo el rango de la variable de entrada seleccionada.

La estratificación de la zona a inventariar

En algunos casos, puede ser útil añadir criterios exógenos para soportar un muestreo más eficiente de la población. Todas las variables explicativas se pueden utilizar: edad de la población (principalmente para las plantaciones), fertilidad, estación, silvicultura, variedad o especie, altitud, profundidad de las aguas subterráneas, etc. De acuerdo con estos criterios, una división del área del inventario en diferentes estratos podrá ser efectuada (suelo rico, pobre, por ejemplo). Esta estratificación es útil cuando la variabilidad de la biomasa es mayor en algunos estratos que en otros. Aumentando el esfuerzo de muestreo en estos estratos, habrá una mejor precisión de la estimación para un tamaño determinado de la muestra (ver el Recuadro 4).

Recuadro 4:
En teoría, el esfuerzo de muestreo debe ser proporcional a la desviación estándar dentro de cada estrato. En la práctica, depende de los condicionantes técnicos, financieros y humanos.

El muestreo para estimar la biomasa de una población

Para estimar la biomasa a la escala de la población, la selección individual de los árboles resulta impracticable. Es preferible medir todos los árboles en un área determinada llamada parcela y repetir esta unidad para alcanzar el tamaño deseado de la muestra. Como para los inventarios forestales, el error está determinado por el número de parcelas (ver el Recuadro 5).

Recuadro 5:
Existe una relación entre el número de parcelas, el error de muestreo y el coeficiente de variación de la población estudiada. En términos simplificados, esta fórmula se utiliza para determinar el número de parcelas que tienen que ser seleccionadas:

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

donde n es el número de parcelas, CV_B es el coeficiente de variación de la biomasa ($\bar{B}$ es la biomasa media de las parcelas y S_B su desviación estándar) y E es el error debido al muestreo.

El coeficiente de variación de una parcela de superficie A es por lo tanto esencial para construir el plan de muestreo. La relación entre los dos es una función de potencia:

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

En la relación entre CV_B et A, c representa la tipología de agregación de la biomasa en la población. Si $c < 0,5$ la población es agregada, si $c > 0,5$ su distribución espacial es regular.

En estos cálculos, para elegir el tamaño de las parcelas hay que tomar en cuenta la agregación de la población. Si la población tiene una biomasa agregada, para el mismo esfuerzo de muestreo el error se reducirá a través de la selección de un gran número de pequeñas parcelas, mientras en el caso de una distribución espacial regular de la biomasa, es mejor seleccionar un menor número de parcelas pero con una superficie mayor (Figura 4).

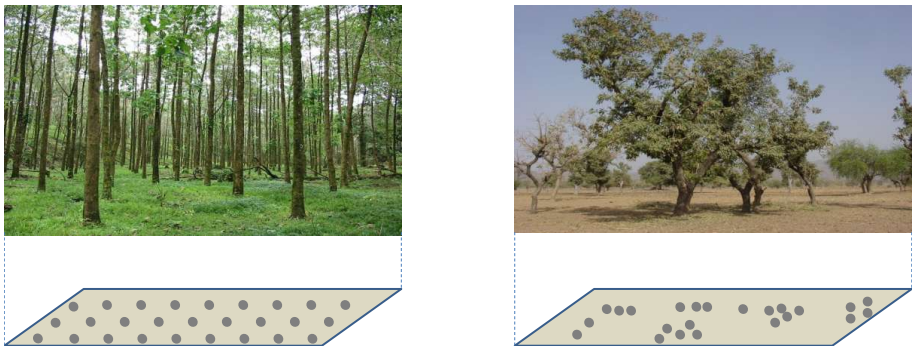


Figura 4. A la izquierda, el diagrama muestra una estructura regular y a la derecha una estructura agregada.

Es igualmente importante recordar que, aparte de estas herramientas para mejorar la precisión de la toma de muestras, la realización de una ecuación alométrica depende de muchos otros factores, como los recursos financieros, humanos, técnicos y medioambiental . Cada factor impone ciertos límites y el arte de lograr la mejor ecuación consiste en encontrar el mejor compromiso.

Consejos para la recogida de datos en el campo y en laboratorio

Los errores de campo son costosos y no pueden ser corregidos. Por lo tanto, hay que tener en cuenta tres principios claves para reducirlos al mínimo:

→ La mejor solución es pesar todos los compartimentos de los árboles en el campo,

→ Cada vez que hay que muestrear una alícuota, es siempre preferible pesar sistemáticamente la masa total de la muestra y después la de la alícuota para controlar la pérdida de humedad del material vegetal,

→ Siendo las campañas de medición de biomasa costosas en términos de recursos y de tiempo, es posible hacer paralelamente otras medidas para evitar regresar otra vez al campo (por ejemplo el perfil del tallo, muestreo para mineralomasa).

Las mediciones de campo son preferentemente destructivas dado que hay que pesar todos los compartimentos vegetales in situ. Sin embargo, no es siempre posible (prohibición de la tala o árboles demasiado grandes) o deseable (por ejemplo en los bosques secos) cortar todos los árboles. En este documento se indicarán las acciones a realizar directamente sobre el terreno. Además serán presentados algunos consejos para superar la imposibilidad de efectuar una medición completa de los arboles.

Recuadro 6: unos consejos antes de salir.

→ No es raro que a causa de unos impedimentos o obstáculos (acceso, ...) haya que modificar el plan original de muestreo.

→ Los diferentes compartimentos del árbol no tienen todos la misma densidad ni la misma humedad. Por tanto, es preferible medir compartimento por compartimento para mantener una cierta homogeneidad.

→ Para evitar sesgos en la selección de las muestras, es importante que las muestras sean recogidas siempre por la misma persona.

→ Para evitar la pérdida de humedad, se recomienda utilizar una nevera en el caso no sea posible medir las muestras en el campo.

→ Una preparación racional de los materiales antes de la salida al campo (hojas de entrada, bolsas para las muestras) permitirá ahorrar tiempo de trabajo en el campo.

Medición directa sobre el terreno

Para evitar los tiempos de inactividad durante el trabajo de los diferentes equipos, un plan de organización del trabajo de 7 etapas (Figura 5) fue desarrollado por el Instituto Nacional de Investigación Agronómica (INRA). El número de personas por turno y la cronología de los diferentes actores involucrados están detallados en el manual.

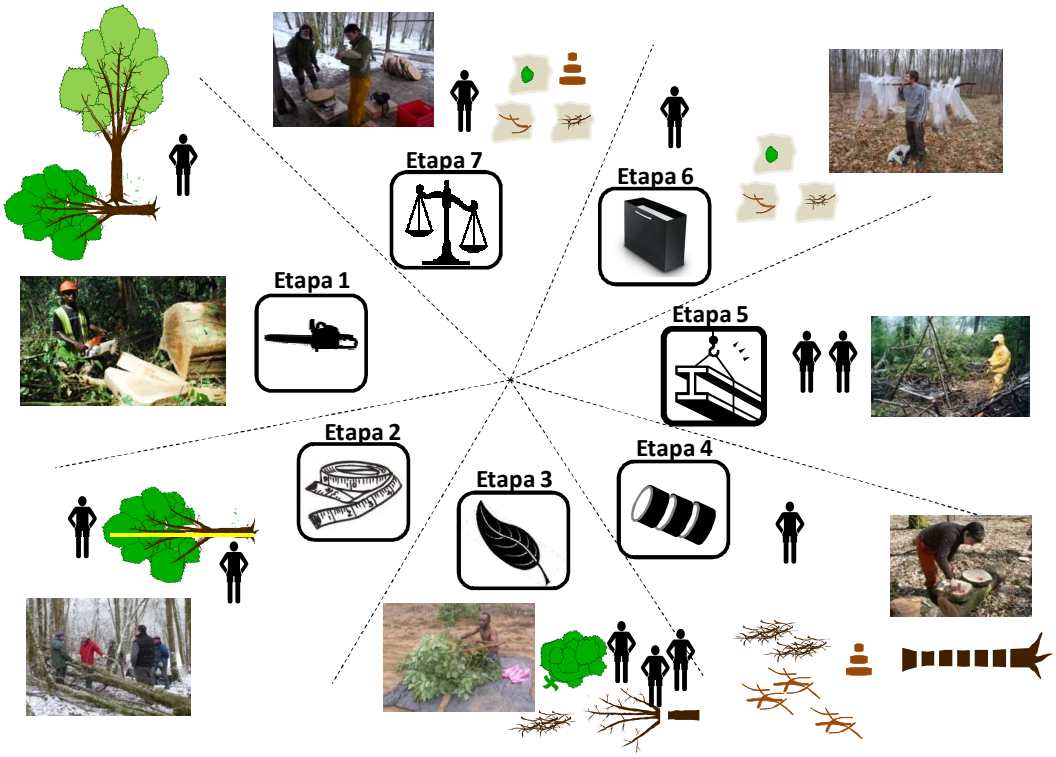


Figura 5. Etapa 1: preparación y tala de los arboles (foto: M. Henry); Etapa 2: Medición de los árboles talados (foto: M. Rivoire); Etapa 3: recogida de las hojas y de las ramas (foto R. D'Annunzio); Etapa 4: partición del tronco y etiquetamiento de las arandelas (Foto: C. Nys); Etapa 5: peso de las porciones de tronco y de las ramas (foto: B. Locatelli); Etapa 6: recogida de muestras de ramas (foto: Mr. Rivoire); Etapa 7: pesar las muestras (foto: M. Rivoire).

Mediciones semidestructivas

Existen varios métodos para evitar el corte total del árbol. Dos de ellos se discuten aquí, el caso de un bosque seco y el caso de un árbol demasiado grande para ser completamente pesado en el campo.

En el caso de los ***bosques secos***, los árboles pueden ser podados regularmente, pero la tala no es una opción debido a la rareza de las especies. Las ramas podadas se utilizan para calcular la biomasa de las hojas (normalmente tres muestras de hojas de tres ramas diferentes son necesarios para formar la alícuota necesaria) y la biomasa de las ramas, así como la densidad de las ramas frescas. Además, se miden el diámetro y la longitud del árbol (Figura 6). Con la densidad de los elementos podados y el volumen de todos los compartimentos, se puede estimar la biomasa total del árbol.

En el caso de los ***árboles demasiado grandes*** para un pesaje completo en el campo, las ramas con un diámetro mayor de 10 cm (valor fijado arbitrariamente, puede ser modificado para reflejar mejor la realidad del bosque en cuestión) se tratan diferentemente que las más pequeñas. La biomasa de los elementos de diámetro superior a 10 cm se deduce a partir de la medición del volumen (diámetro y longitud medida en el campo, ver Figura 6) y del cálculo de la densidad promedio del árbol. La biomasa de los elementos de diámetro inferior a 10 cm se calcula a partir de una regresión entre su diámetro basal y su biomasa.

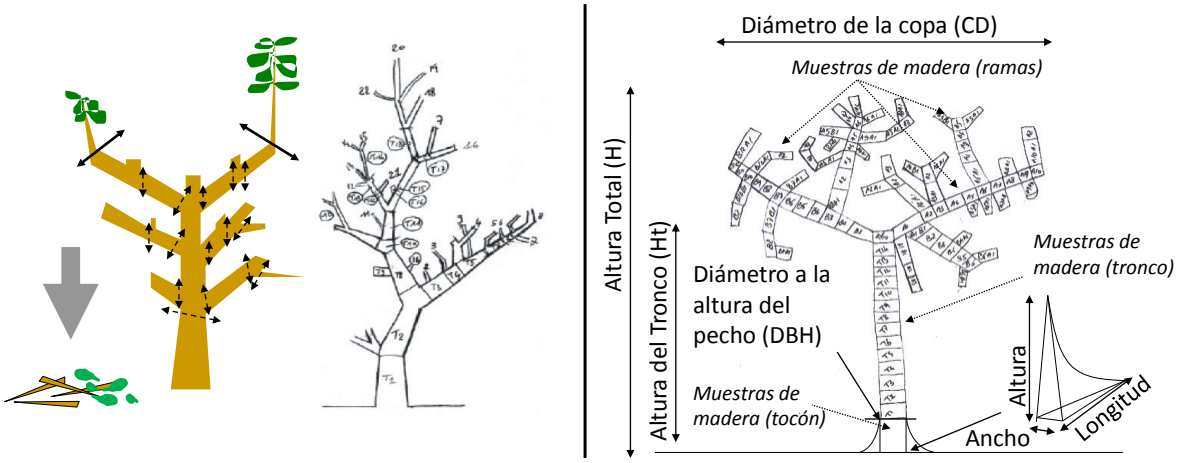


Figura 6. A la izquierda, separación de la biomasa podada y medición de las secciones del árbol no podadas. Las secciones son numeradas en el momento de la medición. A la derecha, numeración de los compartimentos que no han sido cortados y medición de su longitud y diámetro.

Mediciones en laboratorio

El trabajo de laboratorio consiste en medir el volumen de las alícuotas, su masa fresca y su masa seca, obtenida por calentamiento en un horno (Figura 7). En el horno las hojas y los frutos se secan a 70°C (65°C si también se planea de hacer análisis químicos), para las mediciones de biomasa y solamente para la madera , las muestras se pueden secar a 105°C.

Figura 7. Diferentes etapas de secado y pesaje de muestras en el laboratorio.

Entrada de los datos

Una vez que se haya completado el trabajo de campo, la preparación y la entrada de los datos es un paso relativamente simple pero crucial para limitar los errores en la fase de ajustamiento de modelo.

Las tablas tienen que ser anidadas si se consideran varios niveles de datos. La tabla con los datos de los arboles se puede anidar dentro de la tabla de las parcellas, que a su vez se puede anidar dentro de una tabla a nivel de bosque.

Para asegurar una buena comprensión de los datos, puede ser útil incluir información sobre ellos (meta-información): el tipo de variables, fecha, nombre de la persona que ha introducido los datos, etc. Al introducir los datos, las variables cualitativas deben distinguirse claramente de las variables cuantitativas. Es preferible utilizar códigos con pocos caracteres para evitar errores de escritura.

Figura 8. Operadores durante la entrada de los datos (foto: S. Giaccio).

Un método costoso pero que asegura una buena fiabilidad de los datos consiste en realizar una doble entrada por dos operadores diferentes (Figura 8). En todos los casos los procedimientos de verificación utilizados permiten corregir ciertos errores, especialmente la entrada de valores aberrantes o irreales, las entradas de espacios no deseados o espacios dobles, errores de mayúsculas/minúsculas, etc. Es también posible crear scripts para la verificación automática. Los gráficos que representan las variables en pareja son también una buena manera de detectar valores atípicos.

Exploración gráfica de los datos

Cuando el conjunto de datos está listo, la fase de exploración gráfica es un paso clave en la búsqueda de variables explicativas relevantes para la modelización de la biomasa. Los resultados de un modelo se componen de dos términos: ***la media*** y ***el error*** estándar (***o residual***). La exploración gráfica permite definir la forma de la media y de los residuos sin determinar los parámetros del modelo (esto será parte del ajuste). Por lo tanto no se trata de encontrar el modelo que mejor explica la variabilidad del conjunto de datos, sino de retener los tres o cuatro que aparecen más pertinentes. El método consiste en la representación de los valores de la variable a explicar en función de cada variable explicativa y observar si la nube de puntos obtenidos sigue una forma particular o no, y si la varianza de los residuos es constante o no. Los gráficos siguientes (Figura 9) ilustran estos cuatro tipos de relación.

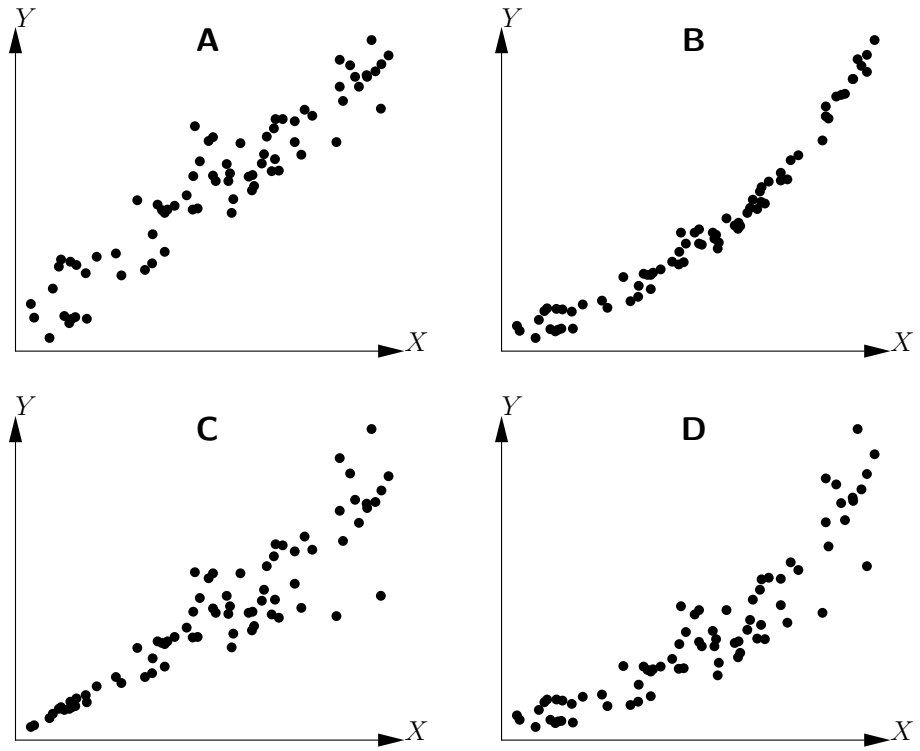


Figura 9. A: relación lineal, varianza constante de los residuos; B: relación no lineal, varianza constante de los residuos; C: relación lineal, varianza de los residuos no constante; D: relación no lineal, varianza de los residuos no constante.

También es posible representar la variable a explicar en función de :

→ variables acopladas a partir de las variables básicas (por ejemplo, D^2H),

→ variables transformadas, lo más común es la transformación logarítmica para obtener una relación lineal (es más fácil de distinguir visualmente si una relación es lineal o no más que identificar el tipo de relación a partir de curvas),

→ variables creadas a partir de las relaciones entre varias variables interdependientes o no (sumas, multiplicaciones, sumas + multiplicaciones).

La fase de exploración permite detectar artefactos de modelización, es decir que algunas formas de las nubes de puntos no muestran una relación lineal entre las dos variables representadas mientras la modelización estadística da un coeficiente de correlación (R^2) alto. Este paso también permite encontrar errores importantes de entrada de datos.

Ajuste de la ecuación a partir de los datos

Después de identificar las variables que generan un modelo relevante, se tendrá que utilizar métodos y pruebas estadísticas para confirmar o modificar la selección de las variables y determinar los parámetros del modelo. Todas estas herramientas se basan en el comportamiento de los residuos del modelo, la primera parte está dedicada a ellos antes de una presentación más completa de las diferentes posibilidades de ajuste.

Definición y hipótesis sobre los residuos

Los resultados del modelo son las predicciones. Siendo nuestra variable dependiente (también llamada variable de respuesta del modelo) la biomasa, vamos a tener valores de biomasa medidos en el campo y valores de biomasa predichos por el modelo a partir de variables explicativas (diámetro, altura, densidad, especies, etc.). La diferencia entre cada observación y su modelo de predicción se llama residuo y así hay tantos residuos como observaciones (ver el Recuadro 7).

Recuadro 7:
A la observación i-ésima de la variable de respuesta (B_i) le corresponde una predicción del modelo ($\hat{B}_i$) y un residuo (ε_i) definido por :

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

Las herramientas estadísticas para determinar los parámetros del modelo se basan en el comportamiento de estos residuos y sobre la verificación a posteriori de tres hipótesis:

1. ***Independencia de los residuos:*** esta hipótesis no suele verificarse mediante cálculo y se debe averiguar en el momento del muestreo de los árboles. Los arboles deben estar suficientemente separados para que la elección de un árbol no vaya a influir en el árbol siguiente. En caso de dudas, la prueba de Durbin-Watson permite comprobar la independencia de los residuos.

2. ***Distribución normal de los residuos:*** esta hipótesis se verifica gráficamente mediante una representación grafica de los cuantiles empíricos en función de los cuantiles teóricos. En este gráfico, los puntos deben estar alineados aproximadamente a lo largo de una línea recta (Figura 10). También hay pruebas estadísticas para probar la normalidad de los residuos (Kolmogorov-Smirnov, D'Agostino et al, etc.).

3. ***Varianza constante de los residuos:*** esa hipótesis es también conocida como homocedasticidad. Se habla de heterocedasticidad cuando este supuesto no se ha verificado. La homocedasticidad de varianza de los residuos se verifica mediante el gráfico de los residuos en función de los valores previstos. Este gráfico no debe presentar ninguna estructura o tendencia particular de la nube de puntos (Figura 10).

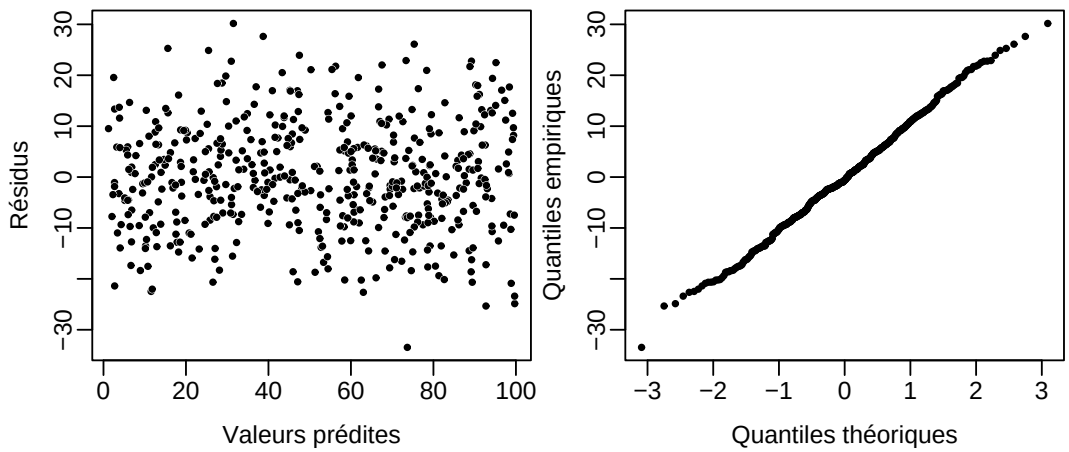


Figure 10. A la izquierda la varianza de los residuos es homogénea, a la derecha la distribución de los residuos es normal.

Las hipótesis dos y tres pueden ser verificadas mediante una simple visualización gráfica dado que los tipos de modelos aquí presentados son ***robustos***, es decir, que son de calidad satisfactoria incluso si las hipótesis no estén completamente verificadas.

Los diferentes tipos de ajuste

Los diferentes tipos de ajuste presentados en esta sección son: (1) las regresiones lineales simples, múltiples, ponderadas o con modelo sobre la varianza y regresiones no lineales con exponente conocido o a estimar, y (2) la selección entre modelos anidados preexistentes, con la misma variable de respuesta o con diferentes variables de respuesta. El método estadístico adaptado para cada tipo de ajuste se muestra en la siguiente tabla (Tabla 1).

Ajuste	**Método**	**Forma de la ecuación**	**Ley de los residuos**
Regresión lineal:			
Simple	mínimos cuadrados	$Y = a + bX + \varepsilon$	$\varepsilon \sim \mathcal{N}(0,\, \sigma)$
Múltiple	mínimos cuadrados	$Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_pX_p + \varepsilon$	$\varepsilon \sim \mathcal{N}(0,\, \sigma)$
Ponderada	mínimos cuadrados ponderados	$Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_pX_p + \varepsilon$	$\varepsilon_i \sim \mathcal{N}(0,\, \sigma_i)$
Con modelos sobre la varianza	Máximo de verosimilitud	$Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_pX_p + \varepsilon$	$\varepsilon \sim \mathcal{N}(0,\, kX_1^c)$
Regresión no lineal:			
Si el exponente c es conocido	mínimos cuadrados ponderados	$Y = f(X_1, X_2, \ldots,\, X_p; \theta) + \varepsilon$	$\varepsilon_i \sim \mathcal{N}(0,\, \mathrm{Var}(\varepsilon))$ $\mathrm{Var}(\varepsilon) = g(X_1,\, \ldots,\, X_p; \vartheta)$
Si hay que estimar el exponente c	Máximo de verosimilitud	$Y = f(X_1, X_2, \ldots,\, X_p; \theta) + \varepsilon$	$\varepsilon_i \sim \mathcal{N}(0,\, \mathrm{Var}(\varepsilon))$ $\mathrm{Var}(\varepsilon) = g(X_1,\, \ldots,\, X_p; \vartheta)$
Selección entre modelos preexistentes:			
Anidados (lineal)	Ley de Fisher	prueba si los parámetros agregados por el modelo completo al modelo anidados no son diferentes de 0	
Anidados (no lineal)	Ley del χ^2	prueba si los parámetros agregados por el modelo completo al modelo anidados no son diferentes de 0	
Variables respuestas idénticas	AIC	El mejor modelo es aquello que minimiza esto criterio	
Variables respuestas diferentes	Índice de Furnival (F)	El modelo con el valor de F más pequeño será considerado el mejor	

Tabla 1. Tabla que muestra varias pruebas estadísticas utilizadas para el ajuste del modelo. El conjunto de símbolos y los cálculos es explicado en el manual.

Les méthodes de régression linéaire ou non

El método de ***los mínimos cuadrados*** consiste en el calcular la suma de los cuadrados de los residuos (SCE), y luego en el encontrar el conjunto de parámetros que minimiza esta suma. Este cálculo es relativamente fácil, pero es adecuado sólo para los modelos más simples: regresión lineal simple o múltiple.

En el caso del método de ***los mínimos cuadrados ponderados***, un peso positivo se asocia a cada observación. Este peso es definido por una relación proporcional con el inverso de la varianza residual. Esta relación se puede simplificar para los datos biológicos tales como la biomasa o el volumen y luego esto se traduce en un exponente adicional a identificar en comparación con el método anterior.

El método de ***máxima verosimilitud*** consiste en calcular los argumentos que maximizan la verosimilitud de las observaciones. La verosimilitud de que una observación es la densidad de probabilidad de obtener esta observación bajo el modelo especificado. La variable de respuesta se distribuye en este caso con una ley de distribución normal por la cual la esperanza y la desviación estándar dependen de tres parámetros (θ,k,c). Este método consiste en encontrar el triplete de parámetros que maximizan esta función.

En el caso de los modelos no lineales, sea con el método de mínimos cuadrados ponderados o con el de máxima verosimilitud, una optimización numérica es necesaria para verificar que se obtenga el mínimo (o el máximo, respectivamente). El manual describe varios procesos de optimización. Por fin, durante la fase de prueba, varios modelos lineal y no lineal pueden corresponder a las observaciones. Para seleccionar el modelo final, la Tabla 2 presenta algunas ventajas y desventajas de modelos lineales múltiples y modelos no lineales.

	Ventajas	**Desventajas**
Regresión lineal:	Expresión explícita de los coeficientes del modelo.	Restricciones en la forma de residuos, poca flexibilidad en la forma del modelo.
Regresión no lineal:	No hay restricción en la forma del modelo por la media o la varianza.	No hay expresión explícita de los coeficientes del modelo por lo tanto hay el riesgo de hacer estimaciones sesgadas de los parámetros.

Tabla 2. Esta tabla presenta algunas ventajas y desventajas los modelos lineales múltiples y no lineales.

Selección entre los modelos ya existentes

En el caso de ***modelos anidados***, la selección vuelve a probar si los parámetros adicionales del modelo más completo son significativas o no. La hipótesis nula se rechaza si el p-value de la prueba (Fischer para los modelos lineales o χ^2 para los modelos no lineales) es menor que el nivel de significación (el modelo completo es el mejor en este caso).

En el caso de los ***modelos con la misma variable de respuesta***, el mejor modelo es el que minimiza el criterio de información de Akaike (AIC). Este criterio se basa en la verosimilitud del modelo utilizado.

Si se trata de comparar ***modelos con diferentes variables de respuesta*** (siendo la primera la variable transformada de la otra), se puede utilizar el índice de Furnival. Este índice es definido por un modelo cuyos residuos tienen una varianza constante, independientemente de la forma de la transformación de la variable. El modelo con el menor valor de F es considerado el mejor.

Estratificación y agregación de las observaciones

Cuando los datos son estratificados, hay dos métodos para ajustar el conjunto de datos: (1) cada estrato se ajusta separadamente utilizando los métodos descritos anteriormente o (2) se analiza el conjunto de datos en su totalidad. En el segundo caso, el análisis utilizado se llama ***análisis de covarianza***. Se supone que todos los residuos tienen la misma varianza dentro y entre los diferentes estratos. Se utiliza un test de Fisher para probar si los parámetros del modelo son significativamente diferentes de 0. Cuando el número de estratos es demasiado grande, el manual presenta otras soluciones. Este caso puede ocurrir en la estratificación por especies de árboles. Por último, el manual también presenta soluciones para integrar los modelos para cada compartimento del árbol en un único modelo, por ejemplo mediante el uso de los ***modelos multivariados***.

Las herramientas presentadas en esta sección permiten probar diferentes modelos a partir de un conjunto de datos y determinar la ecuación más precisa (Figura 11), pero también permiten comparar varios modelos existentes para elegir uno que coincida mejor con el área de estudio.

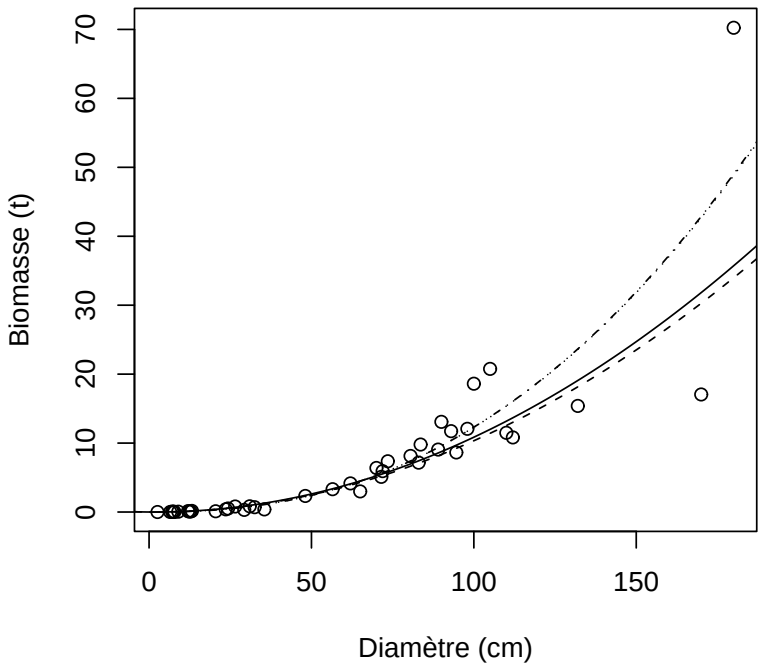


Figura 11. Predicción de la biomasa en cuatro tipos diferentes (sólidos, trazos, puntos, punto-de trazos) ajustado a los datos de 42 árboles (puntos) medido en Ghana por Henry et al., (2010)[2].

Validación del modelo y control de las predicciones

Una vez que el modelo final está ajustado, es posible ***validarlo*** si hay a disposición un conjunto de datos independientes de aquellos utilizados para el ajuste. Hay varios criterios para comparar las predicciones del modelo con las observaciones: el sesgo, la suma de los

[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.

cuadrados de las diferencias residuales, la varianza residual, el error residual ajustado, R^2 o el Criterio de Información de Akaike (AIC). Por último, no hay que olvidar que las predicciones del modelo seleccionado tienen una variabilidad intrínseca. Una vez completado el modelo, es aconsejable que se le asocie una medida de incertidumbre (Figura 12): ***el intervalo de confianza al 95 %***. El método de cálculo del intervalo de confianza es diferente si la predicción se basa en un árbol promedio o en un árbol escogido al azar en la población, y también según la forma del modelo, lineal o no lineal. Todos estos métodos están descritos detalladamente en el manual.

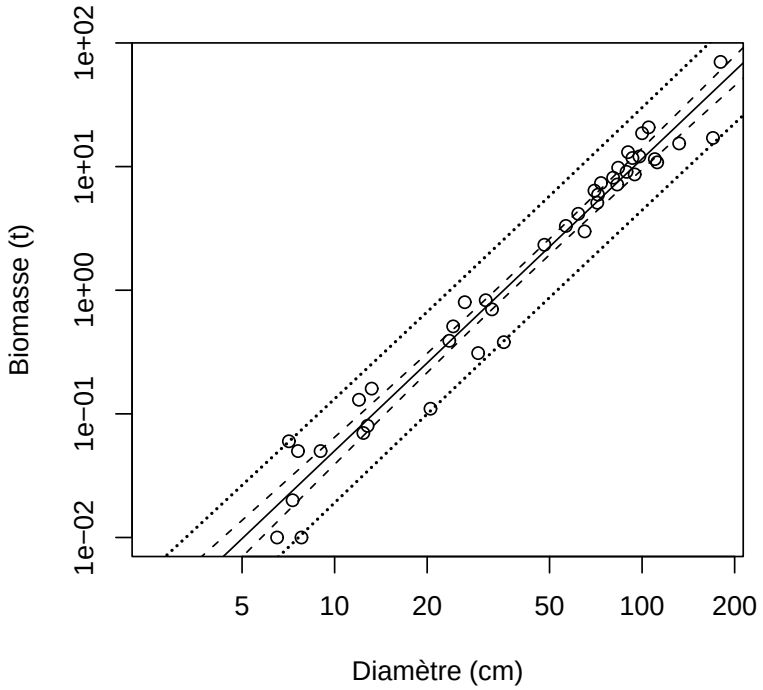
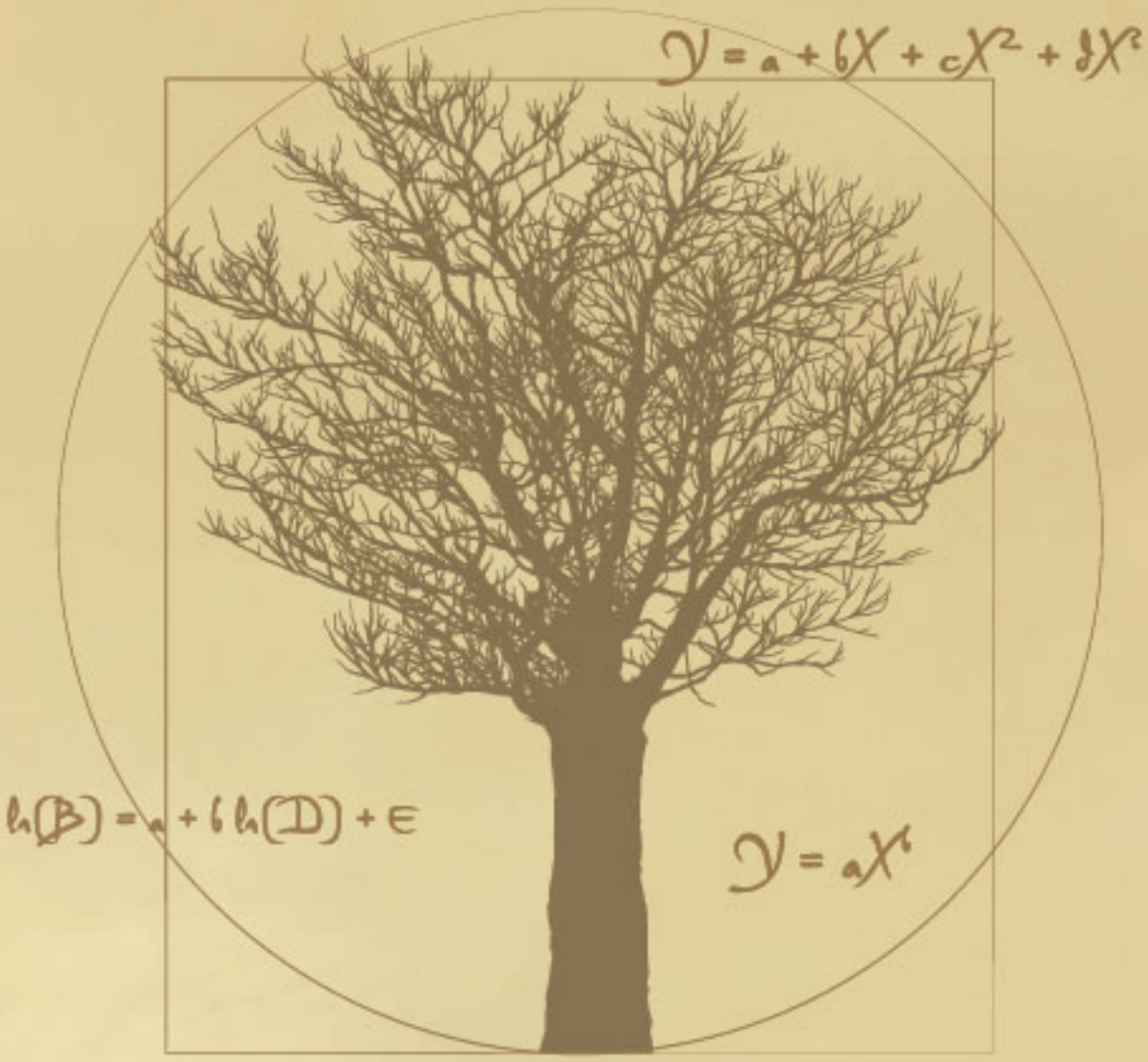


Figura 12. Datos de 42 árboles (puntos) medidos en Ghana, con la predicción (línea negra) y los intervalos de confianza para un árbol tomado al azar (línea de puntos) y para el árbol promedio (línea de rayas)

Conclusiones

Las herramientas estadísticas y los diferentes métodos presentados en el manual de construcción de ecuaciones alométricas tienen la originalidad de ofrecer un enfoque holístico. El objetivo final de este manual es proporcionar herramientas para aumentar la precisión de las estimaciones de las reservas forestales de carbono. Para lograr esto, es importante conocer con exactitud el proceso global que conduce a las ecuaciones, desde el muestreo hasta los análisis estadísticos. No están representados todos los posibles casos dado que algunas situaciones aún constituyen un reto técnico y científico. A pesar de que ya se ha hecho mucho en este campo, los conocimientos estadísticos sobre este sujeto están aún evolucionando. Cuando las herramientas para construir ecuaciones alométricas sólidas y precisas estén bastante generalizadas, nuevas prospectivas podrán ser explorada para mejorar la estimación del carbono. En particular, la creación y el mantenimiento de bases de datos de todas las ecuaciones existentes dará la posibilidad a los usuarios de encontrar rápidamente la ecuación más adecuada a su contexto específico.


$Y = a + bX + cX^2 + dX^3$
$\ln(B) = a + b\ln(D) + \epsilon$
$Y = aX^b$