Modelling for estimation and monitoring

Steen Magnussen¹ and David Reed²

THIS CHAPTER DISCUSSES THE FOLLOWING POINTS:

• The use of models and data aggregation in NFIs and NFAs.
• Estimation and modelling of biomass and carbon content in forested ecosystems.
• Aspects of model predictions, model quality, model errors and model bias.
• Issues in estimation and modelling of temporal change in NFI attributes.

Abstract

Many attributes of interest in forest inventories must be estimated using lookup tables or equations, collectively referred to as models. Stem volume, for instance, is rarely measured directly in the field, but is instead estimated from dimensional measurements such as diameter or height. Individual observations are usually aggregated for reporting purposes and may be grouped during field data collection, for example, when trees are tallied by diameter or height class. Aggregation simplifies presentation of data, but reduces the information content.

Equation-based models are used frequently to estimate variables such as volume, biomass or carbon content. However, it is sometimes difficult to judge whether a given equation is applicable to a particular situation. It is therefore important to assess the quality of all equations used in an inventory. The selection of a particular equation should be guided by this assessment, as well as the modelling objective and context.

Some inventories focus on assessing the state of the resource, while others concentrate on changes in the resource over time. Different field procedures and sampling designs are preferred for different objectives. Sampling with permanent plots provides estimates of both state and change. Designs with a smaller (and more expensive) set of permanent plots and a larger set of (less expensive) temporary plots can also provide good estimates of both state and change, but the statistical analysis of such designs is quite complex and care must be taken to ensure appropriate assessment of accuracy and precision.
1. Introduction

Many attributes of interest in forest inventory and monitoring applications are not measured directly, as to do so would be impractical or costly. The best-known example is stem volume where diameters and heights are measured on individual trees and a table or equation (collectively referred to as a model) is used to estimate the associated volume (Köhl, Magnussen and Marchetti, 2006, section 2.37). The volume of coarse woody debris (CWD) is another example (ibid., section 6.4.6; Woldendorp et al., 2004). In practice, only a small number of short segments of CWD are measured for volume, while the total amount in an area of interest is estimated through scaling of the measured pieces, the scaling itself depending on the sample design (Woldendorp et al., 2004). Advances in remote sensing favour the use of easy-to-measure and readily available variables (say X) correlated with attributes of interest (say Y). An estimate of Y for the area or population of interest is then obtained from a small sample of Y and knowledge of the relationship between X and Y (Köhl, Magnussen and Marchetti, 2006: 80).

The following sections present the types of models available for indirect estimation of quantities such as wood volume, biomass and carbon content, along with examples. They also discuss important issues to consider in the choice of model for a specific application, in addition to modelling objectives and contexts.

2. Aggregation

Aggregation is the combination of data or observations into groups. Its purpose is to simplify measurements and data processing (Stage, Crookston and Monserud, 1993) or to summarize data and observations into groups and categories of interest.

The data collected on attributes in a forest inventory are defined and prescribed according to the information needs of stakeholders (Dachang and Cossalter, 2006: 5). In a field inventory, it is important to measure and record all items in a sample plot that meet the definitions/specification of the desired attribute (for more information see chapter on Observations and measurements, p. 41). Data definitions/criteria can be based on demographic characteristics. In the case of trees, for example, this could be a minimum size limit, often expressed as a minimum diameter (e.g. 10 cm). Criteria may also limit measurements to certain species or tree characteristics of interest.

It is important to remember that limiting inventory data and observations to items (e.g. trees, shrubs, snags, etc.) that meet established definitions/criteria equally limits the inference and estimates that can be derived from these data to the parts/units of the forest population that satisfy the definitions/criteria. For example, if stem-wood volume estimates for live trees are estimated only for trees with a diameter at breast height greater than 10 cm, then it will not be possible to estimate total volume or biomass for the forest of interest. To paraphrase, the results apply only to population elements with a known positive probability of being included in the inventory sample (Thompson, 1992: 21) (see section on sampling design and estimation).

Sample unit summaries do not usually include values for all items measured. For example, trees are often aggregated into groups based on species, demographics such as diameter, height or social status (section 2.2, p. 113), hierarchy such as functional group (section 2.3, p. 114) or the total stand. Aggregation may be one-dimensional (e.g. total volume by species) or multidimensional (e.g. volume by diameter class by species) as appropriate. Aggregation of inventory items can also be based on land-use pattern (e.g. shifting cultivation) or future planned use of a resource item (e.g. fuelwood).

The consequences of any aggregation of observed/measured items on analysis and inference should always be considered carefully (Clark and Avery, 1976). Once data are aggregated, it is usually difficult – if not impossible – to recover underlying details at a later stage, should it become desirable
(Ritchie and Hann, 1997). To conserve a desired accuracy and precision of estimates, it is often better to postpone aggregation to the analytical phase that follows completion of field data collection. However, for practical and cost-saving reasons, aggregation often occurs during field observation. Instead of measuring, for example, the diameter of each tree in a field plot to the nearest 0.1 cm, counts of trees by 5 cm or even 10 cm diameter class are often practised due to expediency. In this case, the class midpoints are usually used in the analytical phase. It should be noted that an aggregation always has the potential to introduce additional error and possible bias in the resulting estimate (Ducey, 1999). Of course, aggregation greatly simplifies field procedures and may well result in greater overall precision if the savings due to efficiency are used instead to collect more field data; however, the trade-off is far from simple, and it is important to reiterate caution against uncritical aggregation.

2.1 Aggregation by species

Estimates of totals and per unit area values for a stand, forest or region, are frequently aggregated over species meeting a certain definition/criterion with respect to size or use. As stated above, it is important that all species that enter the inventory sample – and otherwise meet the definition/criterion – be duly observed and recorded. Otherwise the estimates will become biased. In cases when the field crew is unable to identify a species, it should be given a name with indications for later identification (e.g. shape or size of foliage, fruit-bodies, branching, bark, etc.). Aggregation by species during field data collection may be acceptable during surveys of biodiversity and in count-oriented surveys of disease and insect damage, but otherwise is not recommended.

It is common in temperate forests to develop estimates of stand characteristics by species. In tropical forests, this can be difficult and may prove even more difficult to interpret, due to complex stand structures or the large number of species (Higgins and Ruokolainen, 2004). In some instances, species are combined by ecological functional group (e.g. canopy dominants). This often results in a manageable number of species groups that are relatively easy to interpret in terms of forest structure (Gadow, 1999). Commercial utility and silvics are also considered for species grouping.

In national or continental scale summaries, species groups may be developed for reporting purposes due to the large numbers of potential species (Burns and Honkala, 1990). A species group labelled “pine”, for example, may include all Pinus species occurring within a geographic area of interest.

2.2 Aggregation by size

Trees are often aggregated by demographics with the result that trees of similar sizes or social status are combined into groups for data summaries or field data collection. Aggregation by size for reporting and analysis purposes is generally straightforward, except when the inclusion probabilities (viz. expansion factors) associated with the inventory data are linked to the size of trees (Köhl, Magnussen and Marchetti, 2006: 155). Aggregation of trees by diameter (e.g. 5-cm diameter class) is probably the most common method of aggregation in the field. The number of trees in each diameter class is tallied and diameter class midpoints are used in the subsequent analyses to estimate variables such as volume or biomass. The minimum diameter of trees to be measured is usually specified in the field measurement protocols. Thresholds for inclusion/measurement are defined according to the purpose of the survey. Usually, the minimal commercial diameter is used, or one to two diameter classes below that threshold. Stratification of sampling efforts by size is an inevitable practical constraint when sampling odd-shaped, highly variable and scattered objects. An example is sampling for coarse woody debris (Roth et al., 2003; Williams, Ducey and Gove, 2005).

If trees are selected with a probability
proportional to their size, the use of diameter classes can be problematic (Ducey, 2000). Since forest inventory data are often used repeatedly for many diverse purposes, it is preferable to keep data in the form they were captured. For example, a user wanting to model the diameter distribution (Cao, 2004; Gove and Patil, 1998) would be better served by non-aggregated data.

Trees may also be aggregated by height class, which may have a bearing on the resulting wood products that may be derived (Köhl, Magnussen and Marchetti, 2006: 36). Height class may also be used to aggregate trees for analysis of grazing forage availability or wildlife habitat structure (Spetich and Parker, 1998). Height aggregation is probably more common in ecological or grazing surveys than in surveys that focus on assessment of commercial fibre utilization opportunities.

Finally, trees may be grouped by their position in the forest canopy, such as canopy dominants (Gargaglione, Peri and Rubio, 2010; Nigh and Love, 2004). This type of classification is more useful in ecological surveys than in surveys attempting to inventory commercial material. Structural classification may have a great deal of importance in evaluating forage potential or other non-fibre commercial potential.

### 2.3 Aggregation by hierarchy

Trees may be aggregated in various hierarchical systems for analysis and reporting (Mairota, Florenzano and Piussi, 2002). Trees may be aggregated in a biological hierarchy (individual trees → species → ecological functional group → stand) or a utilization hierarchy (individual trees → diameter class → product class → species → stand). Reporting may include summaries for any or all levels of the hierarchy.

Aggregation also occurs at the spatial scale (e.g. Wolf, 2005). The spatial unit can be a stand or a smaller unit (e.g. a pixel in a remotely sensed imaged). Aggregation is then undertaken across all units that meet a certain requirement relating to their forest attributes or data values. Spatial aggregation may be used for reporting and analysis or to improve sampling efficiency by stratification (Köhl, Magnussen and Marchetti, 2006: 105). During analysis of spatially aggregated data, it is imperative to recover all pertinent information concerning the genesis of the data (observed, sample-based, model-based, predicted, imputed, interpolated, etc.), as well as any available estimates of accuracy and bias. Spatial covariance among aggregated forest resource data types can greatly complicate the statistical analysis of spatially aggregated data (Mairota, Florenzano and Piussi, 2002; Rossi et al., 2009; Schwab and Maness, 2010; Waser and Schwarz, 2006).

In spatial aggregations the spatial units may be aggregated by forest type (e.g. moist tropical) or geographic region for analyses at regional or national levels. Geographic region may be based on political delineations (e.g. state or provincial boundaries) or ecological zones, for example, Holdridge life zones (Ni, Ouyang and Wang, 2005). Aggregation at the stand level may occur after estimation (prediction) of the variables of interest for each sampling unit in a stand. The appropriate methodology for the aggregation is determined by the sampling design and survey objectives. Aggregation may be built into the sampling design through the use of stratified sampling (De Vries, 1986: 31).

### 2.4 Aggregated class estimation

\[
Y_{ij} = A^{-1} \sum_{k}^{n_{ij}} X_{ijk}
\]

For trees measured on fixed area plots, estimates of per hectare values are obtained by dividing the respective individual tree characteristics by the size of the area sampled: \(Y_{ij} = \text{estimated per hectare quantity of the}\)
measured variable for the \( i \)th sampling unit and the \( j \)th aggregation class with \( n_{ij} \) observations of \( X \);

\( X_{ijk} = \) value of the measured variable for the \( k \)th tree in the \( i \)th sampling unit and \( j \)th aggregation class;

\( A = \) size in hectares of the individual sampling unit.

For trees measured on variable radius sampling units (see chapter on Observations and measurements, p. 41) estimates of per hectare values are obtained by dividing the respective individual tree characteristics by the basal area of the measured tree, and multiplying by the basal area expansion factor:

\[
Y_{ij} = BAF \sum_{k=1}^{n_{ij}} \frac{Y_{ijk}}{B_{ijk}}
\]

\( Y_{ij} = \) estimated per hectare quantity of the measured variable for the \( i \)th sampling unit and the \( j \)th aggregation class with \( n_{ij} \) observations of \( X \);

\( BAF = \) basal area factor, equivalent to the basal area per hectare represented by each measured tree;

\( X_{ijk} = \) the value of the measured variable for the \( k \)th tree in the \( i \)th sampling unit and \( j \)th aggregation class;

\( B_{ijk} = \) basal area in m\(^2\) of the \( k \)th measured tree in the \( i \)th sampling unit and \( j \)th aggregation class.

\( \hat{Y} = \sum_{i \in S} Y_i \pi_i^{-1}, \hat{Y} = \hat{Y} \times N^{-1} \)

An estimate of a total (\( \hat{Y} \)) or a mean (\( \hat{Y} \)) for a population parameter – obtained by a probability sampling design – follows the basic principles behind the Horwitz-Thomson estimator (Overton and Stehman, 1995):

where summation is over the units \( (i) \) in the sample \( (s) \) and \( p_i \) is the sample-inclusion probability of the \( i \)th sampled unit. For a population with \( N \) units, the estimation is sometimes done with a mixture of \( n \) sample-based unit-level observations, and \( N-n \) estimates derived from models or otherwise imputed (McRoberts, 2006; McRoberts, Nelson and Wendt, 2002). In this case, the estimator for the total is simply the sum of all \( N \) unit-level values (observations and estimates). The associated estimator of accuracy may follow directly from statistical theory (Overton and Stehman, 1995). In more complicated cases application of the delta technique is required (Davison, 2003: 33–35).

Estimation of population parameters from units of observation is performed on a routine basis in forest inventories. There are two basic sampling units commonly used in forest inventory applications: (i) fixed area plots and (ii) variable radius plots (e.g. Corona et al., 2010). Transect sampling are special cases of fixed area plots and can be treated similarly in many applications (Hedley and Buckland, 2004). In most cases, estimates of the variables of interest are obtained for each sample unit, and then combined to obtain estimates for the larger area (aggregate) of interest. The method of combining estimates from individual sampling units, and the methods of estimating associated precision of the estimates, depends on the sampling design.

### 2.5 Implications of aggregation in estimation and modelling

Aggregation during the field data collection phase simplifies field data collection and may improve the relative accuracy. Some of the statistical issues in connection with data aggregation have already been outlined. A summary is provided next.

The potential downside of data aggregation during the data collection phase is the possible introduction of bias and an almost certain reduction in both accuracy and precision of resulting estimates due to the introduction of error (Clark and Avery, 1976). Combining all trees within a species class, for example, results in loss of information on individual tree sizes. Since forest inventory data are typically used for multiple purposes and in multiple combinations, it is in general advisable to limit any aggregation to the
reporting/analysis phase of an inventory. Decisions about aggregation during field data collection also affect the future utility of the collected data. New and emerging issues important to forestry may require details that were lost due to economic pressures of expediency. It is not possible, for example, to explore numerous aspects of biodiversity if species have been aggregated during field data collection. The increased use of remote sensing techniques in forest inventories (Tomppo et al., 2008) can be viewed as an aggregation process (in extremis).

Disaggregation of aggregated data is only possible if one is willing to make assumptions about the frequency distribution of possible data values that have been aggregated into a single value (Papalia, 2010). Only rarely can such assumptions be justified.

3. Volume estimation

Volume is the most widely used measure of wood quantity. It is usually estimated as part of the assessment of economic value or commercial utilization potential. The wood volume may refer to a specific portion or part of a tree or the whole tree. The total wood volume of a tree includes the volumes of stem(s), branches, stump and roots. For standing trees, above ground volume production is generally based on stem wood volume for conifers, but may include branch volume for broad-leaved tree species.

Depending on the measurement objective and local traditions, measurements or predictions of wood cubic volume may refer to total stem volume, total tree volume (stem and branches) or the volume of portions of a tree intended for a specific use (Köhl, Magnussen and Marchetti, 2006: 47). Volume estimates may include or exclude bark and, for above ground estimates, include or exclude the stump. Volume is always a cubic measure and is usually expressed in cubic metres. Merchantable volume, however, is sometimes expressed in other units relating to commercial use (Skovsgaard, 2004).

In the field, the volume of standing trees is typically estimated from such measurements as diameter, or diameter plus some height of interest (e.g. merchantable height, total height or height to a usage specified diameter limit). Subsequent application of suitable volume equations, taper equations or a log-rule will then produce the desired volume estimate (Lynch, 1988, 1995; Tesfaye, 2005; Tomé et al., 2007; Yamamoto, 1994a; 1994b).

Volume may be measured directly on felled trees or logs, but is often estimated from dimensions such as minimum diameter or piece length (Husch, Miller and Beers, 1972). Direct measurement of volume is usually performed by sectioning a tree into smaller pieces assumed to be cylinders (Köhl, Magnussen and Marchetti, 2006: 50). Volume may be estimated for stacks of logs or processed products by measuring their dimensions. Local knowledge is needed to make the appropriate transformation to an estimate of the solid wood volume.

Advances in remote sensing technology, especially Lidar (see chapter on Remote sensing, p. 77), now allow field-based estimates of volume for a spatial unit (plot) to be combined with a suite of remotely sensed ancillary variables, in order to obtain either model-based predictions of per-unit area volume or per-tree estimates of volume for trees large and distinct enough to be identified with a high degree of confidence (Maltamo et al., 2004a; Maltamo et al., 2004b; Parker and Evans, 2004; Popescu, Wynne and Nelson, 2003).

3.1 Volume equation forms

Stem volume (V) is usually expressed quantitatively as a function of diameter (D), diameter and height (H) or merchantable length. Occasionally, other variables such as clear bole length are used to estimate volume (Husch, Miller and Beers, 1972). An important consideration is that any variable needed to predict volume should be observed during field data collection. The following two “classic” models are often used (Köhl, Magnussen and
Marchetti, 2006: 50): \[ V = \alpha + \beta \times D^2 \times H \]
alternatively \[ V = \omega \times D^2 \times H^p \] are coefficients to be determined from specific (small) samples in which tree volumes are carefully determined, or known from previous studies (viz. subject knowledge). When tree height can be expressed as a function of diameter (Begin and Raulier, 1995; Huang and Titus, 1992; Jayaraman and Lappi, 2001; Moore, Zhang and Stuck, 1996; Nanos et al., 2004; Zhou and McTague, 1996) the relationship can be built into a volume prediction based on diameter alone. Tabulated lookup tables of stem volume for a given species, location and stem diameter are called volume-tariffs (Fonweban and Houllier, 1997; Magnussen, 1998; Paine and McCadden, 1988).

The choice of model may depend on the modelling objective and data (Skovsgaard, 2004). The listed equations implicitly assume a single-stemmed tree form and may require modification or replacement for species with a more complex form. At times a volume equation is easier to fit to data after a logarithmic transformation because the transformation brings the model into a linear form. However, a negative bias is introduced when the predicted logarithm of \( V \) is converted back to arithmetic units (Baskerville, 1972; Bi et al., 2001; Lee, 1982; Wiant and Harner, 1979). This bias is approximately the order of magnitude of one-half of the residual variance of the equation, at least when it can be justified to assume a normal distribution of model residuals.

In the absence of a trusted local volume equation(s), it is possible to utilize geometric relationships to approximate volume. The volume of a cylinder is simply the area of the base times the height, and the volume of a cone is one-third of the volume of a cylinder with the same area of the base and height. Trees are neither cones nor cylinders, however empirical analyses often indicate that the volume of a single-stemmed tree is between that of a cone and a cylinder, with tree volume often lying between 0.40 and 0.45 times that of an equivalent cylinder. A value of 0.42, for example, gives \[ V \approx 0.42 \times B \times H \] where \( B \) is tree basal area at breast height and \( H \) is tree merchantable height. This equation will often overestimate the volume of open-grown trees with more conic form and underestimate the volume of trees with more cylindrical form, and may need to be modified for species with more complex forms. Nevertheless, it does provide a first approximation that can be modified subsequently following local experience.

Volume equations derived from remotely sensed predictors are typically linear (possibly after a logarithmic transformation) with a model form that depends on the sensor type, data resolution and scale (Biggs, 1991; Magnnusson, Fransson and Holmgren, 2007; McRoberts et al., 2007; Straub, Weinacker and Koch, 2010; Yu et al., 2008).

Application of any model means that the ensuing estimates are not exact. Estimates derived from models may be biased due to limitations of the model, and in all cases are predictions of the expected value given the value of the predictors (Kangas, 1996). For example, if a stem volume is predicted from \( D \) and \( H \) using a local volume equation, then the estimated value of \( V \) should be interpreted as the average value of all trees in the population with the exact same \( D \) and \( H \) values. The actual tree in question may have a volume that is either greater or less than the expected value (Gregoire and Williams, 1992). If a model produces estimates that have an unacceptably high level of bias, it may become necessary to either develop an improved model or calibrate an existing model (Erdle and MacLean, 1999; Kangas and Maltamo, 2000; Lappi, 1991).

### 4. Biomass estimation

Biomass is defined as the total mass of living plant organic matter expressed as oven-dry tonnes or oven-dry tonnes per unit area. Estimates of biomass may be restricted to the above ground portion of the vegetation, trees or tree components (e.g. foliage, wood, etc.) (Gschwantner et al., 2009).
Biomass of a forest stand (compartment) is often proportional to the volume and basal area of the stand. Conversely, the biomass of a single tree is typically proportional to its diameter and height (Teobaldelli et al., 2009). Allocation of biomass to various functional components is related to species, growing conditions and the water, nutrient and energy requirements of individual plants and stands (Gargaglione, Peri and Rubio, 2004). The carbon content of vegetation is directly related to biomass, as discussed in the following section.

Direct estimation of forest biomass is a labour-intensive and costly proposition. Stratified sampling for a field-based estimation of biomass per unit area is the only realistic approach (Loaiza Usuga et al., 2010). Strata are typically defined based on clearly identifiable components of the living vegetation (e.g. fungi, mosses, herbs, grasses, shrubs, seedlings, saplings, trees and epiphytes). The biomass in each sampling unit is determined by weighing after drying following standard protocols (Gabriëls and Berg, 1993). Sample items (e.g. trees) too large for practical and cost-effective handling are sectioned into smaller pieces, and a sample of the smaller pieces is then taken for biomass estimation. It is vital that the dissection and sample plan ensure an unbiased estimator of the biomass of large items (Ahmed, Bonham and Laycock, 1983; Cancino and Saborowski, 2005; Good et al., 2001; Gregoire, Valentine and Furnival, 1995).

4.1 Biomass components

Biomass may be estimated in total for stands or portions of stands as noted, but information on biomass distribution by plant component is often needed. Biomass components may be divided as necessary for a given application, but often include categories such as stem wood, branch wood, foliage, bark, roots, and so on, with more or fewer subdivisions as needed. A common constraint is that the sum of the component biomass estimates must equal the total biomass for the stands or portions of stands of interest.

In many applications, only above ground biomass estimates are used. There are obviously below-ground components to biomass (e.g. coarse roots, fine roots, etc.), but studies quantifying these values are available for only a small number of species and ecosystems, and are difficult to conduct and typically produce low precision data (Lukac and Godbold, 2010; Macinnis-Ng et al., 2010; Niiyama et al., 2010; Pramod and Mohapatra, 2010; Zhang et al., 2010).

At the stand level, biomass may be estimated for the overstory, shrubs, herbs, lichens, moss and so on. In forested situations, the overstory biomass usually dominates. There are cases where tree cover is low and overstory or tree biomass is smaller compared with that of other ecosystem components. The decision as to which biomass components to consider depends on the choice of ecosystems to be surveyed and the intended use of the information.

Cannell (1982) presents a compendium of worldwide biomass data from a cross-section of ecosystems. The compendium includes ratios for various biomass components for many forest types. As of 2010 this compendium remains the single most authoritative compilation of benchmark biomass figures. A smaller set of biomass estimates can be found in a recent re-evaluation of forest biomass and carbon storage (Keith, Mackey and Lindenmayer, 2009).

4.2 Biomass equations

Biomass equations are used to predict biomass from readily available ancillary variables (X). The equations may predict the biomass of a single tree or the tree biomass on a unit of forest land. Tree-level equations express biomass as a function of tree dimensions (diameter and height). Equations for unit-area predictions of biomass vary according to the ancillary variable(s) (X). Equations driven by field-related X-variables generally apply stand-level attributes such as basal area, mean tree size (height/diameter) or similar aggregates of tree-level attributes. Equations
driven by $X$-variables obtained via remotely sensed data (Gallaun et al., 2010; Wijaya et al., 2010; Zhao, Popescu and Nelson, 2009) vary according to the sensor type and resolution behind $X$. In many cases, the biomass used as the dependent variable ($Y$) in these equations is rarely a direct estimate of biomass, but rather an estimate obtained by another set of models that “expands” available inventory estimates of tree and stand attributes to the desired biomass component(s) (Albaugh et al., 2009; Gallaun et al., 2010; Jalkanen et al., 2005; Lehtonen et al., 2004; Levy, Hale and Nicoll, 2004; Schroeder et al., 1997; Somogyi et al., 2008; Teobaldelli et al., 2009; Wijaya et al., 2010; Zhao, Popescu and Nelson, 2009).

Equations applied to forest inventory data are usually developed for particular species or species groups, and may be developed with data collected from narrow geographic ranges. Some examples described below show more widely applicable equations developed through a synthesis of published studies. Cannell (1984) presented equations to estimate stand-level woody biomass from total stand basal area and average tree height for a wide range of temperate and tropical stand types; most of the equations are for temperate coniferous forest types. These equations are simple to apply since they use variables commonly obtained during field data collection. Their application is, as a rule, used for stand-level (plot) estimation rather than for tree-level estimation.

At the individual tree level, Jenkins et al. (2003) give composite equations applicable for temperate species across North America. Teobaldelli et al. (2009) provide a similar set of generalized equations for five species groups in Europe. These equations could be applied, with appropriate qualification, to other temperate forest types.

An example is presented here of a generalized equation used by Jenkins et al. (2003), specifically, the Schumacher equation where total above ground biomass is estimated for individual trees based on an allometric relationship with diameter at breast height: $B = e^{b_0 + b_1 \ln(D)}$ where $B$ is total above ground biomass (kg) for trees 2.5 cm and larger in diameter at breast height ($D$). Coefficients are given for both deciduous and coniferous species groups throughout all regions of the United States. Broad species groupings are utilized (e.g. pine and spruce with a total of five coniferous and four deciduous species groups). Teobaldelli et al. provide equations for the expansion factor (BEF) needed to convert an estimate of growing stock ($X$) to an estimate of biomass ($B$). A widely used expansion equation has the form $BEF = b_0 + b_1 X^{b_2}$, whereby $X$ is a measure of the growing stock.

Brown (1997) presents equations for individual trees in tropical forests. For broadleaved species, two equations are presented for tropical dry forests, two for tropical moist forests and one for tropical wet forests. In addition, one equation is presented for palms and another for tropical conifer forests. All of these equations express individual tree biomass as a function of diameter and height, although different specific equation forms are used in different applications.

The biomass of various biomass components is commonly estimated from models of the distribution (allocation) of above ground forest tree biomass to specific components (stem, bark, stump, branches, foliage, fruit/seed). Continuing the example from Jenkins et al. (2003), the proportion of the total biomass in the $i^{th}$ biomass component of a tree can be estimated from the tree’s diameter at breast height, as in $r = e^{b_i \ln(D)}$.

It should be noted that the “conversion” from one or more readily available inventory attributes of growing stock to biomass, and then to biomass components via a set of generalized equations, is only simple in principle. Many factors and circumstances can cast doubt on an estimate of biomass obtained with generalized equations (Albaugh et al., 2009; Jalkanen et al., 2005; Lehtonen et al., 2004; Retzlaff et al., 2001). It is therefore
incumbent upon the analyst to exercise great diligence with respect to choice of model and intended application of a chosen model. The limitation and error structure of many generalized models and allometric biomass-allocation formulae are often not well documented.

5. Carbon content estimation

Regional and national estimates of ecosystem carbon content, and change in ecosystem carbon content over time, are important components of any assessment of global carbon cycling and its impact on atmospheric greenhouse gases and climate (Birdsey, 2006; Cairns and Lasserre, 2006; Waterworth and Richards, 2008; Watson, 2009). Several international agreements now require improvements in the ability to assess forest carbon stocks and their change (Dutschke and Pistorius, 2008; Kägi and Schmidtke, 2005; Zhang, Zhu and Hou, 2009).

In this context, it has become increasingly important to quantify the carbon content that resides in forests and forested ecosystems, and its contribution to the carbon cycle. Forest inventories make significant contributions to estimates of carbon in forested ecosystems because the carbon content is relatively easy to assess for the components of the vegetation captured by an inventory (Dupouey et al., 2010; Nabuurs, 2010; Rodeghiero et al., 2010; Tupek et al., 2010). In many cases, vegetative carbon is used as a surrogate for total ecosystem carbon, since it is relatively easy to derive from existing information or ongoing inventory efforts. Total ecosystem carbon, which includes inorganic ecosystem components such as soil, is more difficult to assess, especially if the precision of the estimates must be quantified (Baritz et al., 2010; Loaiza Usuga et al., 2010). Expensive estimates of carbon are typically derived from a few intensively studied plots, each considered as representative of a very large area with similar soils, vegetation and climate.

Today, many unit-area estimates of carbon content in forest vegetation are generated from a suite of explanatory variables (regressors) delivered from various satellite or airborne sensors (Maselli et al., 2010; Sánchez-Azofeifa et al., 2009; Tagesson et al., 2009). Invariably these estimates build on a modelled relationship between field-based estimates of biomass (carbon) and one or more sensor-based ancillary variables.

5.1 Carbon content of vegetation

The carbon content of vegetation is surprisingly constant across a wide variety of tissue types and species (Baritz et al., 2010; Mäkelä, Valentine and Helmsaari, 2008; Munishi and Shear, 2004; Nogueira, Fearnside and Nelson, 2008; Rana et al., 2010; Wauters et al., 2008). Schlesinger (1991) noted that the \( C \)-content of biomass is almost always found to be between 45 percent and 50 percent (by oven-dry mass).

In many applications, the carbon content (\( C \)) of vegetation may be estimated by simply taking a fraction of the estimate of oven-dry biomass (\( B \)), as in \( \hat{C} = 0.475 \times \hat{B} \). The accuracy of an estimate of this nature is typically not great due to errors in \( \hat{B} \), and one should also expect it to be biased.

For dead material, carbon content is a function of the state of decomposition (Boulanger and Sirois, 2006; Garrett et al., 2010; Mukhortova and Trefilova, 2009; Vävrová, Penttilä and Laiho, 2009; Yang et al., 2010). For material that can still be identified, such as fresh litter or standing dead trees, the above equation may be used to estimate the \( C \)-content if the mass of the material can be estimated (see section 5.2 below 5.2, p. 121). For severely decomposed material, it may be necessary to determine the \( C \)-content in subsamples taken from the material collected at a site, and then combine this with an estimate of the total (bio) mass of that class of material before the \( C \)-content for that vegetative component can be estimated. Even
small errors due to sampling, measurement and handling of the material can have a serious impact on the accuracy of an estimate for a vegetation component that is orders of magnitude larger than the taken sample (Woodall, Heath and Smith, 2008).

The total carbon content of vegetation goes beyond trees. It includes all parts and components of the plant community, such as herbs, shrubs and mosses. Field-based estimation of carbon typically begins with an estimation of biomass and then a conversion along the lines detailed above. To accomplish this task it becomes necessary to stratify the community and sample from each stratum. The necessary strata must be defined based on the composition, structure and extent of the community in question (Clark et al., 2008; Friedel, 1977; Kenow et al., 2007). In some cases, it may make sense to obtain C-content estimates for lifeforms such as epiphytes, while in other cases this is irrelevant. The approach follows the above for all classes of vegetation: first, the biomass is estimated in each stratum (component) using appropriate sampling methods and then the ratio is applied to estimate the C-content.

### 5.2 Ecosystem carbon content

In addition to the carbon content of vegetation, it may be necessary to estimate total ecosystem carbon content (Jia and Akiyama, 2005; Wang and Sun, 2008; Wise et al., 2009). This includes biotic as well as abiotic carbon pools. Avian (Pautasso and Gaston, 2005) and mammalian (Desbiez, Bodmer and Tomas, 2010; Plumptre and Harris, 1995) biomass and carbon content is often ignored, since it usually constitutes a small fraction of total ecosystem carbon. At times it may be required to estimate arthropod biomass and carbon content in order to obtain a good estimate of total ecosystem C (Fisk, Fahey and Groffman, 2010; Tovar-Sanchez, 2009). Colonizing insects may comprise a significant portion of the total biomass of some systems (Vasconcellos, 2010; Yamada et al., 2003), and abiotic materials incorporated into nests and colonies may also be a significant portion of total C.

Soil organic matter is a major abiotic carbon pool (Chang et al., 2010; Rovira, Jorba and Romanyà, 2010; Tipping et al., 2010) of particular importance at high latitudes or high altitudes. This may in some cases be greater than the vegetative carbon. Dead plant material at the soil surface and in the upper soil horizons may also have a significant C-content that should be considered in any estimate of ecosystem C-content (Fisk, Fahey and Groffman, 2010; Gasparini et al., 2010). McKenzie et al. (2000) provide a compendium of methods for field data collection for carbon estimation in soil, litter and coarse woody debris. Quantitative data on forest litter may be sparse. However, several countries with an elevated risk of forest fires may have extensive information, drawn from surveys conducted of elements on the forest floor that significantly increase fire hazards during periods of drought (Fernandes, 2009; Kessell et al., 1978).

The carbon content of litter depends on the stage of decomposition, and can usually be determined from field samples designed for this specific purpose. Application of a ratio approach such as that described for vegetation can be used (see section 5.1), but will often underestimate the C-content of the litter layer due to the escape of carbonic gases during the process of decomposition (Fioretto et al., 2007; Hosseini and Azizi, 2007; MacDicken, 1997).

Estimates of soil C may be obtained from field sampling. This is the most precise and appropriate method to estimate site-specific carbon content. Field data collection should be used whenever precise estimates of soil C are needed, but it is important to consider temporal variation throughout a growing season in large studies that may require an extended sampling period. If a soil classification map for the area of concern exists, there may be information on carbon content for different soil types in the area.
A given soil type may yet have different mean carbon content depending on the dominant vegetative cover and land use. Soil under an agricultural field may have a very different C-content than a similar soil under a mature forest. Estimates of soil C-content may or may not be available for all conditions in an area of concern. Batjes (2009) provides access to an extensive database of global soil physical and chemical properties, including information that may be used to approximate soil C-content in the absence of site-specific information. These estimates will be less precise than those obtained from field samples, but may be cost-efficient when high precision, site-specific estimates are not required. In many applications, it may be more cost effective, and ultimately result in higher precision in final estimates, to use a greater number of less precise estimates of C-content for individual sampling units, than to measure C-content of a subset of sampling units with high precision (MacDicken, 1997). The trade-offs are a function of sampling design and cost, and must be evaluated in that context. It should be noted, however, that if expedient less expensive C-estimates are biased, the opportunities for an attractive trade-off between a small sample with expensive observations and a larger sample with less expensive observations can be severely curtailed (Köhl, Magnussen and Marchetti, 2006: 79).

### 6. Judging model quality

A model summarizes a conceptual relationship between one or more dependent variables (Y) and one or more predictors (X). The model can be stated as a single equation (e.g. Fehrmann et al., 2008), a system of related models (e.g. Gertner, Fang and Skovsgaard, 2002) or a hierarchical (multilevel) model (e.g. Pedersen, 1998). Models are mostly used for predicting new value(s) of an unobserved entity from available predictors. The volume, biomass and carbon equations given above provide examples of the most basic types of models. A model may be formulated through subject knowledge (Curtin, 1970), adopted from other studies or suggested by apparent trends in observed data. The following references provide access to a broad selection of forest models (Amaro, Reed and Soares, 2003; Dykstra and Monserud, 2009: 276; Schwab and Maness, 2010; van Laar and Akça, 2007).

In forest inventory and biological sciences, data exhibit a large amount of natural variation and models are limited to predicting the expected value of the dependent variable given the input data. The quality of any model is judged by its ability to provide unbiased (accurate) estimates of these expectations and the precision of model predictions. Models with deterministic (fixed, invariable) model parameters generate a single prediction (the expected value) given a set of predictor values. Stochastic models contain one or several parameters that are random (Biging and Gill, 1997; Rennolls, 1995). Hence, they can generate both conditional predictions for a random unit (e.g. a tree, a plot or a forest stand) and population averaged predictions (Schabenberger and Gregoire, 1996).

When fitting a model to data, a comparison of values predicted by a model and the actual values of the dependent variable provide an initial assessment of model quality. It is generally desirable for models to be unbiased, meaning departures from model predictions (residuals) to average to zero for any input, and precise, meaning residuals are distributed tightly around the predicted values.

The quality of a model for prediction purposes is assessed by comparing a prediction of a new observation not used in model development to the actual value of the new observation. Common criteria for assessing model quality include a t-test of the hypothesis of a zero mean model prediction error, the variance of model errors, the magnitude of the median absolute deviation (Venables and Ripley, 1994), and the sign test for testing equal medians of the observed and predicted values.
The Wald-Wolfowitz runs test can be used to test the hypothesis that the elements of a sequence of model errors along a gradient of predictor values are independent (Conover, 1980: 136). Additional assessments are often geared towards testing the assumption of a normal distribution of model residuals (Brown and Hettmansperger, 1996), an analysis of errors in “curve” models (Ducharme and Fontez, 2004; Huang, 1997), and the homogeneity of error variances across a range of inputs (McKeown and Johnson, 1996; O’Brien, 1992; Shoemaker, 2003: 432). Reynolds (1984) provides a basic approach to model quality assessment. Vanclay and Skovsgaard (1997) provide a brief overview and an operational framework for judging model quality. Some common assessments of model quality exclude a portion of the data from the model-fitting phase or use a leave-one-out cross-validation approach whereby repeated model estimations leave out one observation and then compare the actual and predicted value for the withheld datum (Efron, 2004). The latter approach is preferred, since a large portion of the data can rarely be withheld without affecting the properties of the assessed model. In summary, fewer observations for modelling may result in the chosen model being suboptimal. When sample sizes are small, model-building based on robust techniques is recommended (Choi, Li and Zhu, 2010; Lange et al., 1989; Wang and Leng, 2007).

When applying a general model, such as the volume and biomass equations shown earlier, or a model developed for a given species in a different geographic area, it is important to attempt to assess model quality prior to application. This may require the collection of new field data, although it may be possible to utilize existing data for this purpose. Failure to assess model quality forces the user to make an untested, implicit assumption that the model used is appropriate for the species and geographic area to which it is applied, which may or may not be true. Users of models should always keep in mind that a model may generate unusual predictions. Extrapolations (application of models with one or more of the predictor values falling outside the range of the data used during model fitting) should be avoided whenever possible because bias and precision may quickly become unattractive for otherwise well-founded models (Schreuder and Reich, 1998).

With the advance in models that rely on input from remotely sensed data, it is increasingly important to verify that the predictors are actually the same as the data used during model fitting (i.e. with identical information content, collected at identical spatial scales and with identical measurement error-structures). If this is not the case, the impact of errors-in-variables must also be considered (Carroll, Ruppert and Stefanski, 1995; Fuller, 1987).

Users of existing models are rarely in a position to conduct a full-fledged model check or even a validation. Key information about the statistical properties and data behind a model is often missing or difficult to retrieve. Instead of relying entirely on model predictions a better strategy may be to take a small probability sample of the variable(s) of interest and then combine them with predictions from a model. This model-assisted type of estimation (Särndal, Swensson and Wretman, 1992) has become popular. In the statistical literature the approach goes under the name of “Small Area Estimation” (Pfeffermann, 2002; Tomppo, 2006). Users concerned with the quality of a model may also adopt a Bayesian paradigm, whereby user-defined prior distributions on model parameters capture model uncertainty and possibly bias, and integrate this uncertainty into their predictions (Gertner, Fang and Skovsgaard, 2002; Green et al., 1999; Green and Strawderman, 1996; Green and Valentine, 1998).

Validation of complex models for large-scale applications (e.g. ecosystem predictions of carbon content) is rarely possible. Validation of individual components of the model may not guarantee that all interactions of the
model components are adequately captured. It is always the user’s responsibility to check model assumptions and model predictions.

7. Model error contribution to total error

Methods to ascertain the precision of inventory estimates are dependent on the sampling design used to collect the data. These methods, however, generally assume that the individual observations are measured without error. For model-based estimates such as volume, biomass and carbon, however, there are model errors to consider. Consequently, there are three main sources of error: measurement error, model error and sampling error. Sample-based precision estimates should, therefore, be considered as underestimating the variance, or conversely as implying confidence intervals that are too narrow for derived variables such as volume, biomass or carbon content. Similarly, methods of estimating the sample requirements to achieve a desired level of precision will indicate fewer samples than really needed, unless consideration of model error is taken into account in addition to sampling error.

Inventory models are never perfect. The discrepancy between the actual (unknown) value ($Y_A$) and the predicted value from a model ($Y_P$) is called the model error ($e_p$). In equation form, this becomes: $Y_A = Y_P + e_p$. This simple (linear) equation also implies that the variance of a series of predicted values is less than or equal to the variance of the actual values. Equality holds only for perfect models with no error variance. For example, if we predict the volume of trees in a plot from a suitable volume equation then the calculated variance of the volume predictions will be less than the actual variance of the volume of the trees in the plot. Consequently, the standard error of a predicted mean volume for a plot will be biased downwards. The variance of prediction errors must be included to obtain an unbiased estimate of the total error.

In many applications it should also be considered that the parameters in models used in an estimation procedure are themselves estimates with associated errors. One may choose to include also this extra source of uncertainty in the estimation of the total errors. For sample surveys with large sample sizes this type of model error would usually constitute a large portion of the total error (variance).

The variance of prediction errors may be substantially larger than the residual variance obtained during model fitting, especially when the mean and covariance of the input variables vary from those of the data used for model fitting. Application of the model outside the recommended application domain raises the spectre of serious additional underreporting of error.

8. Monitoring over time

Monitoring over time allows estimation of change and trends in forest attributes (Köhl, Magnussen and Marchetti, 2006: 143). The changes and trends can be estimated from a set of permanent sample plots (see section 8.1) or temporary plots (see section 8.2) or a combination of both. Temporary plots can be used to obtain estimates of the current state of the forest, while permanent plots or a mixture of permanent and temporary plots are prerequisites for obtaining estimations of change over time (Picard et al., 2010). Estimation of change is a complex challenge. There are three major types of temporal changes in forestry: (i) change conforming to the expected progression of living and dead material in a forest during the period of interest (e.g. volume increment of living trees); (ii) unexpected biotic or abiotic disturbances (e.g. mortality due to insect, snow, wind, fire, etc.); and (iii) forest management activities (thinning, harvesting, planting, seeding, etc.). Each category operates at different temporal
and spatial scales. Given the multivariate nature of forest resources, and the wide range of rates and modes of change, it follows that the efficiency of most sampling designs for estimation of change can be highly efficient for one attribute of change (e.g. net volume increment), yet inefficient for capturing other types of change (e.g. rates of deforestation, volume destroyed by fire, insect mortality). Few practical designs are efficient for capturing change in sensitive but small subpopulations (e.g. number of specimen of a rare or possibly endangered species) (Christman, 2000; Magnussen et al., 2005). To adequately capture changes related to abiotic and biotic disturbances and to forest management practices, it is common to conduct a census of correlated ancillary variables via remote sensing at the start and the end of the period covered by the change estimate(s) (Coppin et al., 2004; Stehman, 2009; Tomppo et al., 2008).

Change estimates are frequently evaluated against expectations or a set of targets, and estimates of the precision of the change estimate are important in this situation. When the change is estimated from a combination of field observations and remotely sensed ancillary variables, estimators of change and their precision can become very complex, and the actual estimation may require the assistance of a statistician (Stehman, 2009). Unless it can be argued on statistical grounds that an estimate of change is unbiased, bias should be accepted as a potential issue.

The simplest change equation is for a trait $Y$ observed at time $t$ and then again at some future time $t + \Delta_t$. Here $Y_{t+\Delta_t} = Y_t + \Delta Y_{\Delta t}$, where $Y_t$ is the initial measurement at time $t$, $Y_{t+\Delta_t}$ is the future measurement at time $t + \Delta t$, and $\Delta Y_{\Delta t}$ is the change in $Y$ from time $t$ to time $t + \Delta t$. The variance of the estimate of $\Delta Y_{\Delta t}$ depends on the type of plots (or mixture of plots) used for the data collection. Any correlation between measurements at two points in time must be accounted for when estimating the variance of a change.

Continuing this simple example, the estimate of change is $\Delta Y_{\Delta t} = Y_{t+\Delta t} - Y_t$. In this case, the variance of the change estimate is equal to:

$$\text{var}(\Delta Y_{\Delta t}) = \text{var}(Y_t) + \text{var}(Y_{t+\Delta t}) - 2\rho(Y_t, Y_{t+\Delta t}) \sqrt{\text{var}(Y_t) \text{var}(Y_{t+\Delta t})}$$

where $\text{var}$ denotes a variance and $\text{cov}$ a covariance, and $\rho$ is a correlation coefficient (between the original and future measurements). A strong positive correlation reduces the variance of a change measurement. When a sample selection has been derived with an unequal probability sampling design, the analyst must take into account that these probabilities may have changed over time (Roesch, Green and Scott, 1993).

As discussed in section 7 (p. 124), if the above change estimation involves the use of quantities that are predictions of expected values from one or more models, then it will again be necessary to account for the “hidden” errors in $Y_t$ and $Y_{t+\Delta t}$. This will commonly be the case in forestry. Compounding the issue is the fact that the errors in $Y_t$ and $Y_{t+\Delta t}$ often tend to be correlated. Additional complications arise when the method (protocol/process) for obtaining $Y_t$ differs from that of $Y_{t+\Delta t}$. The assistance of a professional statistician may be called for.

### 8.1 Estimating change using remeasured permanent plots

Permanent plots refer to forest sampling locations that are monumented or otherwise uniquely identified and remeasured at different points in time (Köhl, Magnussen and Marchetti, 2006: 144).

From a statistical and data analysis perspective, the major advantage to permanent plots is improved precision of estimates of change due to a typically strong correlation of sampling errors (see the above expression for the variance of a change). Higher data quality
may also materialize from additional attention and quality control. Finally, permanent plots permit inference regarding cause and effect (Augustin et al., 2009). For undisturbed and carefully measured permanent plots the correlation between subsequent measurements tends to be both positive and relatively strong, which, as outlined above, lowers the variance of an estimate of change. Yet the correlation between successive measurements can deteriorate quickly with the length of a measurement interval and disturbances (e.g. fire, wind, snow, drought, forest management interventions). High data quality facilitates error checking in current data and scanning for anomalies in past data.

8.2 Estimating change using temporary plots

Temporary plots offer a maximum of flexibility: independent surveys can be established at different times, with plots only measured at one time. Since surveys at different times are taken on different plots, the advantage of the above-discussed positive correlation of plot-specific observations at time \( t \) and \( t+\Delta t \) no longer exists (there is no natural pairing of the two sets of observations).

Individual trees on temporary plots are usually measured more quickly and with less precision than those on permanent plots, reducing the precision of the estimates and the resulting estimate of change. Less precise observations will also make it more difficult to spot outliers and anomalies in the data (Cerioli, 2010). A lower data precision can, to a degree, be offset by the use of a greater number of temporary plots, but the final trade-off depends, in a complicated way, on where and how errors enter the observations and on the assumed model behind the observations (Carroll, Ruppert and Stefanski, 1995).

### Self-study exercises

#### Abbreviations:

- \( B \) = basal area (of a tree or trees in a plot)
- \( \text{BAF} \) = basal area factor (in Bitterlich sampling)
- \( \text{BEF} \) = biomass expansion factor
- \( \text{CWD} \) = coarse woody debris
- \( D \) = stem diameter of a tree at a given reference height (typically 1.3 m)
- \( H \) = tree height
- \( \text{Lidar} \) = light detection and ranging (aka laser scanning)

#### Exercise 1.

Your task is to estimate total stem volume for a fixed area plot in an inventory. The example assumes that there is only one species in the plot, and that a single volume equation is adequate.

The area of the plot is 600 m\(^2\).

You have measured the diameter (\( D \)) at breast height for all trees with a height greater than 1.3 m.

You have also measured the height of eight trees for the purpose of constructing of a height-diameter model and the prediction of height for trees with no measured height.

Trees selected for a height measurement are highlighted (bold) in the data given below.

There are 42 trees in the plot. The diameter (\( D \)) at breast height was measured to the nearest 0.5 cm, but the data were rounded to the nearest 2-cm class. The recorded values of \( D \) are as follows:

\[
D = \{18, 18, 20, 22, 22, 22, 22, 22, 22, 22, 24, 26, 26, 28, 28, 28, 28, 28, 28, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 32, 32, 32, 32, 32, 32, 32, 34, 34, 36, 36\}
\]

Q1.0 Compute the basal area (\( B \)) in square metres per hectare from the given data of \( D \)
and plot size (give the result to the nearest 0.1 m²).

Q1.1 Has the rounding of D introduced a bias in the estimate of B?

Q1.2 What is the magnitude of the relative bias (i.e. bias in percent of the estimate of B) introduced by rounding measurements of D in B? Pick the most appropriate answer from the following list of relative bias: −5%, −2%, 0%, 2%, 5%.

Heights of the eight measured trees were measured to the nearest 10 cm, but only recorded to the nearest 0.5 m. The height data were:

\[ H = \{26., 27.5, 29.5, 30., 30., 31., 31.5, 31.5\} \]

To predict the expected height (eH) for trees with no height measurement you opt for the following height-diameter model:¹

\[ eH = b_0 + b_1 \cdot \log(e[D]) \]

Estimate via ordinary least squares the regression parameters \(b_0\) and \(b_1\).

Q1.3 Do you think the rounding rule applied to D has an effect on your estimates of \(b_0\) and \(b_1\)?

With the estimates of \(b_0\) and \(b_1\), compute the expected height of all plot trees (call them eH; notice: the expected height of a measured tree is the measured height). Round all expected heights to the nearest 0.5 m.

Compute the expected total stem volume of the 42 trees (eV) using the following volume-equation (D in cm, H in m):²

\[ eV = 3.03 \times 10^{-5} \cdot D^{1.7} \cdot eH^{1.3} \]

Round each tree's volume estimate to the nearest 0.02 m³.

Q1.4 In your opinion, does the above rounding rules for D and H introduce a bias in the estimate of the mean tree height (i.e. mean of eH)?

Q1.5 Does the rounding of D and H introduce a bias into the expected total volume (eV)?

**Exercise 2.**

You are tasked with estimating the total above ground biomass for a species AA in a newly inventoried forest. The inventory provides you with the diameter at breast height for all trees of species AA with a height larger than or equal to 1.3 m. You have identified the following four above ground biomass (AGBM in kg oven-dry mass) equations as suitable candidates.³

\[ D \text{ is diameter at breast height (in cm).} \]

EQ1: AGBM = 0.0983 \( \times D^{2.3773} \)

EQ2: AGBM = 0.0617 \( \times D^{2.5328} \)

EQ3: AGBM = 0.0842 \( \times D^{2.5715} \)

EQ4: AGBM = 0.0629 \( \times D^{2.6606} \)

To help you decide which equation is most suitable, you have estimated AGBM for ten trees sampled at random from the inventory sample.

The diameter (D cm) and above ground oven-dry biomass (AGBM kg) of your ten trees are as follows:

<table>
<thead>
<tr>
<th>Tree no.</th>
<th>D cm</th>
<th>AGBM kg</th>
</tr>
</thead>
</table>

₁ Taken from Loetsch, Zöhrer and Haller, 1971, *Forest inventory*, Vol. 2. BLV, p. 130.


Q2.1 Based on this information, which of the above four equations appears to be most suitable for use? Explain your choice of equation.

**Exercise 3.**
Your task is to estimate total tree stem volume of trees in the NFI plots of your country. You have made observations of diameter at breast height \((D)\) and height \((H)\) of all trees in all measured NFI plots. For species AA, BB and CC you are using volume equation models that can estimate the total stem volume from \(H\) and \(D\). However, for species DD you do not have a model for volume estimation. For this exercise, assume that you have only observed these four species in your NFI plots. How would you go about estimating volume for species DD in this situation? Please explain your suggested approach. What factors might prevent you from reaching a decision? In the event that no decision is possible, what additional information might be required to reach a decision and estimate the volume for trees of species DD?

**Exercise 4.**
You are part of an NFI/NFA team responsible for taking field measurements in NFI plots. Stakeholders of the NFI are keenly interested in estimating ecosystem carbon content. A funded request is put forward to select data in the NFI plots to estimate carbon content in forest floor litter. How would you respond to this request? If a decision is taken to meet this request what field procedure would you recommend?

**Exercise 5.**
An NFA client/stakeholder has asked you to make predictions of total above ground biomass of live trees within a 1000 km\(^2\) region with approximately 60 percent forest cover. There are only 25 NFI plots in this region. The biomass estimates from these 25 plots are highly variable. A direct estimate of biomass for the region based on the 25 samples would have a very low precision. You have access to national archives of remotely sensed data. What course of action would you embark on in this situation?

**References**


Oxies, J. 1976. *Permanent plots in forestry growth and yield research*. Royal College of Forestry, Department of Forest Yield Research, Research Notes, 43.


