We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

5,000
Open access books available

125,000
International authors and editors

140M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
1. Introduction

There are a variety of approaches to estimate above ground biomass (AGB), which can be classified according to the data source being used (Koch and Dees, 2008): field measurement, remotely sensed data or ancillary data used in GIS-based modeling. Field measurements are based on destructive sampling or direct measurement and the application of allometric equations (Madgwick, 1994). Recently, remotely sensed data, from both passive and active sensors, have become an important data source for AGB estimation. In this chapter we will focus on the use of optical multispectral data such as TM/ETM+ to estimate AGB. Generally, biomass is either estimated via a direct relationship between spectral response and biomass using multiple regression, k-nearest neighbor, neural networks, inverse canopy models or through indirect relationships, whereby attributes estimated from the remotely sensed data, such as leaf area index (LAI), structure (crown closure and height) or shadow fraction are used in equations to estimate biomass (Wulder, 1998). Here, we discuss the use of remote sensing data of moderate spatial resolution as input to estimate AGB. Research has demonstrated that it is more effective to generate relationships between field measurements and moderate spatial resolution remotely sensed data (e.g., LANDSAT), and then extrapolate these relationships over larger areas using comparable spectral properties from coarser spatial resolution imagery (e.g., MODIS) (Steininger, 2000; Lu, 2005; Phua and Saito, 2003; Foddy et al., 2003; Fazzakin et al., 1999; Roy and Ravan, 1996; Zheng et al., 2004; Mickler et al., 2002). In general terms, LANDSAT TM and ETM+ data are the most widely used data of remotely sensed imagery for forest biomass estimation, but data from other moderate spatial resolution sensors have also been used, including ASTER and HYPERION data. In this chapter we present approaches that are currently being developed in Chile. Specifically, we introduce methods for the estimation of AGB using medium spatial resolution satellite imagery and digital elevation models. The main objective is to create, calibrate and validate such methods for applications. We developed an alternative approach in the estimation of AGB using LANDSAT ETM+ imagery and SRTM digital elevation models as covariates for geostatistical modeling. From the spatial perspective, AGB data correspond to an array of points in space (x, y), while covariates correspond to a set of data that has a large number of samples in geographic space (extracted from each pixel), some of which having overlap with the available AGB
data. The method can estimate the spatial variation of AGB at a stand or sub-stand level, and measure the uncertainty attached to the estimation, depending on local conditions. These results are promising and demonstrate the feasibility of using this approach in the evaluation and monitoring of stock biomass or communal farm scale. They are applicable to the actual landscape configuration of the country, open a series of new interest in research and development, and constitute a novel way to solve the problem of assessing biomass stocks.

2. Multispectral data for forest biomass estimation

Some remote sensing studies using LANDSAT TM/ETM + imagery have focused on estimating the forest attributes through correlation or regression analysis with the spectral response obtained from the images. The estimated attributes may include basal area, AGB, canopy closure, diameter at breast height (DBH), tree height, stand density, and leaf area index (LAI) (Franklin, 1986; Horler and Ahern, 1986; Peterson et al., 1986 and 1987; Lathrop and Pierce, 1991; Ardo, 1992; Curran et al., 1992; Cohen and Spies, 1992; Gemmell, 1995; Kimes et al., 1996; Trotter et al., 1999; Turner et al., 1999; Eklundh et al., 2001; Lu et al., 2004; Hall et al., 2006). One of the methods used to estimate attributes of forests from LANDSAT imagery is the direct correlation with radiometric values where regression relationships can be calculated between spectral bands, band ratios or vegetation indices (Franklin, 1986; Roy and Ravan, 1996; Jakubauskas and Price, 1997; Foody et al., 2003; Labrecque et al., 2006). The most used way to perform this kind of method is by applying stepwise forward variable selection. Table 1 summarizes some studies using this method in several countries.

Some of the advantages of using remotely sensed data are through the provision of a synoptic view of the study zone, capturing the spatial variability in the attributes of interest such as biomass, dominant height or crown closure (Zhen et al., 2007). Remotely sensed data can also be used to fill spatial and temporal gaps in forest inventory attributional data, improving estimates of forest biomass and carbon stocks, which can done for large areas as they are not limited by the extend of forest inventories (Birdsay, 2004; Fournier et al., 2003). However, the complicated vegetation stand structures in mature forest and in advanced successional forests often make the results less accurate given similar TM reflectance even if the above ground biomass varies significantly. This disadvantage is often called data saturation. For instance, in Manaus (Brazil) the canopy reflectance of tropical secondary successional forests saturated when AGB approached about 15 kg m-2 or vegetation age reached over 15 years (Steininger, 2000). Nelson et al. (2000) analyzed secondary forest age and AGB estimation using Landsat TM data and found that AGB cannot be reliably estimated without the inclusion of secondary forest age. The main approach to overcome this problem is to include further co-variables independent to reflectance values.

The k-NN method has become popular for forest inventory mapping and applications of estimation over the last years (McRoberts et al., 2007; McRoberts and Tomppo, 2007; Bafetta et al., 2009; Tomppo et al., 2009; Maselli et al., 2005; Tomppo and Halme, 2004; Thessler et al., 2008). The basic idea of this method is to estimate a target attribute of an object, i.e. AGB, using its similarity to objects with known attributes. Its application has focused on estimating continuous variables such as size, age, height, basal area, mean DBH, standing volume, and leaf area (Tomppo et al., 2009; McRoberts et al., 2007; Bertini et al, 2007; Tomppo and Halme, 2004).
Sensor | Study area | Estimation method | References
--- | --- | --- | ---
TM 5 | Manaos, Brazil | Lineal and exponential regression | Steininger (2000)
TM 5 | Pará state and Rondônia, Brazil | Multiple regression | Lu (2005)
TM 5 | Sabah, Malaysia | Exponential regression | Phua and Saito (2003)
TM 4 and 5 | Manaos, Brazil; Danum Valley, Malaysia; Khun Kong, Thailand | Multiple regression and neural network analysis | Foddy et al. (2003)
TM 5 | Sweden | K nearest neighbor (k-NN) | Fazakas et al. (1999)
TM 5 | Madhav National Park, India | Multiple regression | Roy and Ravan (1996)
ETM+ | Wisconsin, USA | Multiple regression | Zheng et al. (2004)
ETM+ | Southwest USA | PnET productivity model | Mickler et al. (2002)

Table 1. Examples of methods using LANDSAT data for estimation of forest biomass.

3. Estimation of biomass in Chilean forests

Estimation studies of forest biomass in Chile began to appear in the late 80’s, primarily for plantation of *Pinus radiata* (Caldentey, 1989) and, subsequently, for some local native species (Caldentey, 1995; Garfias, 1994; Herrera and Waisberg, 2002; Schlegel, 2001; Schmidt et al., 2009). In native forest, the background data is limited. The estimation method to be applied depends on the forest composition, structure and site variability. Natural forests are highly variable in these attributes (Donoso, 1993; Gajardo, 1994; Luebert and Pliscoff, 2006) while plantations have less variation because they are monospecific and grow under intensive management regimes, designed to standardize the size and the quality of all trees (Lewis and Ferguson, 1993; Lavery, 1986; Gerding, 1991; Toro and Gessel, 1999). Secondary native forests, especially those dominated by the genus *Nothofagus*, have an intermediate degree of variation and heterogeneity (Donoso, 1981; Donoso, 1993; FIA, 2001). Traditionally, AGB estimation methods are based on sampling methods designed to assess standing timber (Husch et al., 1993; Anuchin, 1960; Bitterlich, 1984, Avery and Burkhart, 1994; Loetsch et al., 1973; Prodan et al., 1997). There is no reason for a different design because the volume/biomass ratio is relatively constant mainly depending on wood density. For the same reason, existing inventory plots can be used to estimate AGB directly. The AGB estimation method, which is usually performed for trees larger than 5 cm in diameter at breast height (DBH) and the understory is not included, should be done taking the following steps:

a. Estimate AGB at individual tree level. Given the high cost of measuring the biomass into its components (stem, bark, branches and leaves) it is preferable to use existing allometric equations for biomass by species and component. These equations depend on easy-to-measure state variables (i.e. DBH and height (H)) and allow estimating AGB in similar trees within the stand (Keith et al., 2000; Wang, 2006; Baker et al., 1984; Ares and Braeuer, 2005; Zewdie et al., 2009; Ketterings et al., 2001). The biomass components are estimated based on basic density samples (dry weight / green volume) multiplied by the total volume of the component (Keith et al., 2000; Steininger, 2000; Ishii and Tatedo, 2004; Hall et al., 2006). All these functions have the following generic form:
\[ B_i^c = f(DBH_i, H_i, S_i) + \xi_i \]

where:

- \( B_i^c \) is the biomass of component \( c \) in the \( i^{th} \) tree
- \( DBH_i \) is the diameter at breast height of the \( i^{th} \) tree
- \( H_i \) is the total height for the \( i^{th} \) tree
- \( S_i \) is an estimator of the sound/dead biomass proportion of the \( i^{th} \) tree
- \( \xi_i \) is the regression error of the \( i^{th} \) tree

**b. Estimation of AGB for areal units.** To this end, one has to use sampling plots within the target units and then their statistics as the estimates. In order for this to work properly, it is necessary to measure DBH, species and some health variable of the tree (usually subjective). In natural forests, due to the high variability for trees having the same DBH range, it is better make measurements of \( H \) and \( S \) in order to improve the precision of the results. In plantations (more homogeneous forests) the same can be made by applying a subsample for \( H \) and then to estimate the rest using a regression model for \( H \) as a function of the DBH.

The estimated AGB at plot level has the following generic form:

\[ B_j^c = \sum_{i=1}^{n_j} B_{ij}^c \times F_j \]

where:

- \( B_j^c \) is the biomass of the component \( c \) for the \( j^{th} \) plot (ton/ha)
- \( n_j \) is the number of trees in the \( j^{th} \) plot
- \( F_j \) is the hectare expansion factor for the \( j^{th} \) plot

**c. Estimation of biomass at the stand, local, regional or national levels.** The scales of interest for estimating AGB ranges from stand up to national levels according to the scale of the project. From stand level estimations the other aggregation levels can easily be archived by simply adding other stands estimation into the calculation. The biomass estimate at stand level has the following generic form:

\[ B_r^c = \sum_{k=1}^{m} B_{rk}^c \times \frac{1}{a} \]

where:

- \( B_r^c \) is the biomass of the component \( c \) for the \( r^{th} \) stand (ton)
- \( a \) is the total area of the stand (ha)
- \( m \) is the number of plots within a the stand

The use of optical satellite and topography data as auxiliary variables (covariables) allow the accuracy of AGB estimations to improve because they are based on their spatial covariation to field data by applying geostatistical interpolation. Using topographic data, the AGB variation is scaled to the actual values for that area, and then, AGB can be obtained by overlaying any available administrative division (stands, sites or districts).
According to the available data one can use different methods to estimate AGB (Figure 1). If data are traditional forest inventory plots, estimates should be made by using traditional statistical parameters for total and average quantities, which don’t give any explicit information about the spatial variability of forest attributes. If data are georeferenced, distributed in such a way that they represent the whole territory of interest, then a spatial interpolation method, such as inverse distance interpolation or kriging, can be applied. The former approach has the advantage of being an unbiased probabilistic method where the degree of confidence (accuracy) can be calculated. The basic assumption behind kriging interpolation is that the spatial dependence models described in variograms assume the behavior of a regionalized variable, which is not necessarily true in reality and should be proved. On the other hand, if it is possible to have other variables (covariates), which besides being cheaper to obtained, are spatially well correlated to the variable of interest (AGB), then they can used to somehow correct the weakness of kriging by accounting for spatial discontinuities or irregularities found in nature (Coulston, 2008).

3.1 Native forest description
Chilean forests cover an area of 15.6 million hectares (20.7% of the national territory), of which 13.4 million hectares are natural forests (85.9% of the forested area). Currently 3.6 million hectares of forest are secondary forests (CONAF et al., 1999). Mixed forest of *Nothofagus obliqua*, *Nothofagus alpina* and *Nothofagus dombeyi* are one of the most important forest types in the country and cover 1.5 million hectares (12.2% of the total native forest). The genus *Nothofagus* has ten species that have a high economic value because of the quality of their wood. These *Nothofagus* forests area concentrated between 36° 30' S and 40° 30' S and between 100 and 1,000 m a.m.s.l. in Central Chile. They are present in both mountain ranges, coastal and the Andes, where *N. obliqua* occupies the lowest areas (between 100 and 600 m, approximately), *N. alpina* intermediate ones (between 600 and 900 m) and *N. dombeyi* the highest (between 900 and 1,000 m), resulting in overlap ecotones with pure and mixed formations (CONAF et al., 1999; Donoso, 1981; Gajardo 1994). The main secondary species in these mixed *Nothofagus* second growth forest are *Aextoxicon punctatum*, *Genuine avellana*, *Laurelia sempervirens*, *Persea lingue* and *Eucryphia cordifolia* (Donoso, 1981). Today, a major part of these forests exhibit some state of degradation or have some form of human perturbation (Donoso, 1981; Gajardo, 1994). Nevertheless, they have a high productive potential and they need to be managed to improve their current condition. Usually, the quantification of these resources is done by applying volume tables or biomass functions, but these biomass functions rarely exist for native species and have only local applications.
(Caldentey, 1995; Donoso et al., 2010; Garfias, 1994; Gayoso et al., 2002; Herrera and Waisberg, 2002; Saez, 1991; Schlegel, 2001, Schmidt et al., 2009) (see Table 2). For native species, the biomass estimation methods most used in Chile are regression or allometry for the average tree (see section 3.0). In recent years, further attention has been paid in less expensive methods to assess AGB in native forests, based on remote sensing data and spatially explicit methods (Peña, 2007; Valdez et al., 2006).

3.2 Plantation description
In Chile, forest plantations are the main supply source of industrial raw material, with a total of near 40 million cubic meters. *Pinus radiata* D. Don is the main commercial species behind these massive commercial activities, with 80% of participation, followed by eucalyptus with 20% participation (CORMA, 2011). The plantations are concentrated in the central coastal area of Chile, between latitude 32 ° S and 42 ° S (Figure 2), covering a wide range of soils types (Schlatter, 1977; Schlatter and Gerding, 1984), environmental conditions (Fuenzalida, 1965; Madgwick, 1994) and silvicultural management regimes (Lewis and Ferguson, 1993; Lavery, 1986; Gerding, 1991). The land ownership structure is highly concentrated in two large forestry companies, which together owned 53% of the total planted area (CNE, GTZ / INFOR, 2008; Leyton, 2009), and the problem of quantifying wood supply has been primarily addressed by the private sector. Therefore, little public information exists on the area, location, age, species and management regime of plantations (CORMA, 2011). In contrast, the private sector has a large network of forest inventory information and has built growth simulators for the main species for different areas (i.e. *Radiata* and *Eucasisim* models), types of site and management schemes and bucking (Fundación Chile, 2005; Morales et al., 1979). Some studies on the availability of wood for the forest industry (CORMA, 2011), carbon sequestration by plantations (Gilabert et al., 2010), AGB stocks for energy projects (INFOR, 2010, CNE / GTZ / INFOR, 2008) have been made by combining regional forest inventory data and these simulation models.

<table>
<thead>
<tr>
<th>Chile Region</th>
<th>Forest Type</th>
<th>AGB (Mg/ha)</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIII</td>
<td><em>N. alpina</em> second growth forest</td>
<td>104</td>
<td>Garfias, 1994</td>
</tr>
<tr>
<td>X</td>
<td>Evergreen</td>
<td>194 to 663</td>
<td>Schlegel, 2001</td>
</tr>
<tr>
<td>X III</td>
<td>Mixed <em>N. obliqua</em>, <em>N. alpina</em> and <em>N. dombeyi</em></td>
<td>285</td>
<td>Herrera and Waisberg, 2002</td>
</tr>
<tr>
<td>XII</td>
<td><em>Nothofagus pumilio</em></td>
<td>380 to 447</td>
<td>Caldentey, 1995</td>
</tr>
</tbody>
</table>

Table 2. Above ground biomass of different types of forest and locality.

4. Remote sensing based biomass estimation

4.1 Description of the study areas and field data collection
We consider four study areas, two predominantly plantations (pine and eucalyptus) and two covered with mostly second growth Chilean oaks: *Nothofagus obliqua*, *Nothofagus alpina* and *Nothofagus dombeyi*. In both cases, one of the areas has a farm spatial scale and the other a municipality level (Table 3). Figure 2 shows the geographic location of the four areas of study.
Table 3. Description of the study areas

<table>
<thead>
<tr>
<th>Forest type</th>
<th>Study area</th>
<th>Ha</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plantations</td>
<td>Pantanillos</td>
<td>400</td>
<td>Mainly <em>Pinus radiata</em> and few <em>Eucalyptus globulus</em> stands, both at different ages with intensive management.</td>
</tr>
<tr>
<td></td>
<td>Quivolgo</td>
<td>20,000</td>
<td></td>
</tr>
<tr>
<td>Native, secondary forest</td>
<td>Monte Oscuro</td>
<td>1,600</td>
<td>Mainly <em>N. obliqua</em> with several secondary tolerant species and informal management.</td>
</tr>
<tr>
<td></td>
<td>Curacautín</td>
<td>20,000</td>
<td>Mixed of <em>N. obliqua</em>, <em>N. Alpina</em>, <em>N. dombeyi</em> at different ages and without management.</td>
</tr>
</tbody>
</table>

We used a random sampling design and concentric circular sampling units (8m radius). All the data were collected in summer 2010 and the total size of the sampling was 348 plots, with an average sampling intensity of one sampling units every 63.2 hectares. Each selected tree was measured for DBH and one out of three for total height (H). For young plantations where understory foliage did not allow seeing trees at DBH we used circular plots with a radius of 5.65 m. In plantations stands where individuals grew in clusters - coppice type - we counted the number of individuals per unit and DBH and H was measured for the smaller and bigger trees. Biomass calculations were performed using the methods presented in Section 3.

4.2 Methods

4.2.1 ETM+ and SRTM data acquisition and preprocessing

LANDSAT ETM + images were obtained from the Earth Explorer Web site of the United States Geological Survey (USGS). Additionally, the SRTM digital elevation model (90 m) was freely downloaded from the site Earth Explorer. Bands ETM+ 1, 2, 3, 4, 5 and 7 were grouped into a single file and then projected to WGS84 UTM 18 South or 19 South, according to the area to which they correspond. Subsequently, there were a series of preprocessing steps as detailed below:

a. SLC-off Correction: The images acquired after 2003 have missing data due to malfunction of an instrument called Scan Line Corrector (Figure 3). This so-called SLC-off error was corrected using ENVI software application, which fits the flaw by using two images with the error in different areas, i.e. non-overlapping. Interpolation is performed considering the local histogram of the two images.

b. Geometric Correction: The images were rectified using Chilean regular cartography 1:50,000 with at least 30 control points per image and a root mean square error less than 30 m was achieved.

c. Radiometric corrections: Standard radiometric corrections were applied on all images to reduce the atmospheric effect following the method proposed by Chavez (1996), and for the topographic the one proposed by Riaño et al. (2003).

Once ETM+ and SRTM were co-registered, several variables were then derived:

a. Normalized Difference Vegetation Index or NDVI (Tucker, 1979).

b. Tasseled Cap bands: Brightness (TC1), Greenness (TC2) and Wetness (TC3). The other three components (TC4, TC5 and TC6) do not have a direct biophysical interpretation but were also calculated to take into account the complete data variation.

c. Slope, Aspect and Altitude were directly calculated from SRTM data.
Fig. 2. Forest resources in Chile and geographical location of the study areas.
4.2.2 Data integration
The integration of all layers of information was carried out in GIS environment (ArcGIS 9.3RM) and grouped into four classes:

a. Field data, which contain the observed AGB (target variable) and its coordinates.

b. ETM spectral bands, excluding the thermal band, plus their derivatives: NDVI and Tasseled Cap components (covariates).

c. Topographic information derived from SRTM data: elevation, slope, and orientation (covariates).

d. Vector files with property boundaries, stands, and base cartography support, which are useful in the stratification of the results and administrative aggregation.

For each of the four study areas a database was generated, which was used subsequently in the geostatistical analysis. Data were treated as point processes or count variables in the bi-dimensional space. Each set of field data is associated with a specific geographical area, for which all values to a resolution of 16 or 30 meters were considered in a comprehensive manner. Figure 4 shows the general methodological approach and the generic structure of the databases for each area.

From here and for the remaining sections, we considered biomass and all covariates mentioned above as regionalized variables to be able to follow the formality of geostatistical analysis. First, we performed a variogram analysis with the aim of studying the spatial autocorrelation of biomass and its spatial dependence with covariates. The result is a set of spatial models called variograms for the variable of interest (biomass) and the covariates. In all cases, we analyzed the anisotropy of the models, incorporating it explicitly in subsequent interpolations whenever possible. The variograms were used in the spatial estimation of biomass via cokriging. The exploratory analysis and modeling process were performed using IsatisRM geostatistical software and MatlabRM environment.
4.2.3 Variogram analysis

The values of a regionalized variable such as biomass are not independent, in the sense that the value observed in a site provides information about the values of neighboring sites (i.e. one is more likely to find a high value of biomass near other high values). In the geostatistical interpretation of the regionalized variable this intuitive notion of dependence is described by the formulation of random functions, which model the way the values observed in different sites relate to each other by a multivariate probability distribution.

The moment of order 1 of a random function (expected value or mean) involves a single spatial position and does not actually deliver information on spatial dependence. In contrast, the moments of order 2 (especially, the variogram) are defined with the help of two spatial positions, i.e. the smallest set that can be considered to describe the spatial "interaction" between values. These are the moments that give a basic description and operations of the spatial continuity of the regionalized variable.

Variogram analysis consists in calculating an experimental variogram from the available data, and subsequently fitting a theoretical variogram model. The experimental variogram measures the mean squared deviation between pairs of data values, as a function of the separation vector (distance and orientation) between the two data. The characteristics of this variogram are related to the characteristics of the regionalized variable (Isaaks and Srivastava, 1989; Chilès and Delfiner, 1999), in particular:

a. The behavior of the variogram near the origin reflects the short-scale variability of the variable.

b. The increase of the variogram reflects the loss of spatial correlation with the separation vector. It may depend on the direction of this vector, indicating the existence of directions with greater spatial continuity than others (anisotropy).

c. The range of correlation, for which the variogram reaches a sill (plateau), indicates the distance at which two data do no longer have any correlation.

In practice, the anisotropy can be identified by comparing experimental variograms calculated along various directions of space, for example oriented 0°, 45°, 90° and 135° with respect to the x-axis. Often, this test is completed by drawing a “variogram map”. When there is isotropy, the experimental variograms in different directions overlap and concentric circles are drawn in the variogram map. Otherwise, one may distinguish geometric anisotropies, in which the variogram map draws concentric ellipses; in such a case, the modeling rests upon the experimental variograms along the main anisotropy axes, which correspond to the axes of the ellipses.
Once calculated, the experimental variogram is fitted with a theoretical model, which usually consists of a set of basic nested structures with different shapes, sills and/or ranges (Gringarten and Deutsch, 2001). Examples of basic structures include the nugget effect (discontinuous component), as well as the spherical, exponential, cubic, Gaussian and power models (Chilès and Delfiner, 1999). In the multivariate case, one has to calculate the experimental variogram of each variable (direct variogram), as well as the cross-variograms between each pair of variables, which measure the spatial correlation structure of the set of coregionalized variables. The fitting of a nested structure model is subject to mathematical constraints, reason for which one usually resorts to automatic or semi-automatic fitting algorithms (Wackernagel, 2003).

### 4.2.4 Interpolation by kriging

Kriging aims at estimating the value of the regionalized variable (here, AGB) at a position with no data, as accurately as possible, using the data available in other positions. It is structured according to the following steps:

a. **Linearity constraint:** The estimator must be a linear combination (i.e., a weighted average) of the data.

b. **Unbiasedness constraint:** This step expresses that the expected error is zero: if one considers a large number of identical kriging configurations, the average estimation error approaches zero. The absence of bias does not guarantee that errors are small, only that they cancel out on average.

c. **Optimality constraint:** according to the above steps, the estimator is subject to one or more constraints but is not fully specified. The last step consists in minimizing the estimation error variance, which measures the dispersion of the error around its zero mean, therefore the uncertainty in the true unknown value. Intuitively, this restriction means that if a large number of identical kriging configurations are considered, the statistical variance of the estimation errors is as small as possible. In addition to the estimated value, one can also calculate this minimal estimation error variance, known as the “kriging variance”.

The weighting of the data depends on the spatial continuity of the regionalized variable, as modeled by the variogram, and on the spatial configuration of the data and location to estimate. In general, closer data receive larger weights than data located far from the target location, but this may be counterbalanced by other effects. For instance, data that are clustered in space tend to convey redundant information and may therefore receive small weights; data located along directions of little spatial continuity (with a small variogram range) may also be underweighted with respect to data located along directions of greater continuity.

There exist many variants of kriging, depending on the assumptions about the first-order moment or mean value (in particular, whether it is known or not, constant in space or not) and on the areal support of the value to estimate (“point” support identical to that of the data, or “block” support larger than that of the data) (Chilès and Delfiner, 1999). The multivariate extension of kriging is known as cokriging, which makes use not only of the data of the variable to estimate (primary variable), but also of covariates that are cross-correlated with this primary variable (Wackernagel et al., 2002). The determination of cokriging weights and of the error variance relies on the set of modeled direct and cross variograms (Wackernagel, 2003).
4.2.5 Kriging and cokriging neighborhood

Quite often, the amount of available data is large and using all of the available data is impractical and a selection of the most relevant data is required. One therefore defines a local window or “kriging neighborhood” in which to search for nearby data and to perform local estimation. The shape of the neighborhood should explicitly consider the detected anisotropies, by making the search radius dependent on the variogram range: the larger the range along a direction, the larger the search radius. Typically, the neighborhood is defined by placing an ellipse around the target location, and selecting a given number of data within this ellipse. The justification for such a practice is the so-called screening effect, according to which the closest data screen out the influence of farthest data (David, 1976; Stein, 2002). Also, to avoid the selection of clustered data that convey redundant information, it is good practice to divide the neighborhood into several angular sectors (e.g., quadrants or octants) and to look for data within each sector (Isaaks and Srivastava, 1989).

Such a definition of the kriging neighborhood is, however, mainly heuristic and one is usually not aware of which data are really worth being included in the neighborhood, and which data can be discarded without deteriorating the estimation. The optimal neighborhood actually depends on the variogram of the variable of interest, as the screening effect does not occur with every variogram model. Some generic guidelines have been provided by Rivoirard (1987) to validate the choice of a given neighborhood in the univariate context.

In the multivariate case, the design of the neighborhood is more complex and critical. For instance, the data of a covariate may be screened out by the collocated data of the primary variable or, on the contrary, they may supplement the primary data and provide useful information to improve local estimation. As suggested by recent publications (Rivoirard, 2004; Subramanyam and Pandalai, 2008), the decision to include or not the covariate data should consider the correlation structure of the coregionalized variables and the sampling scheme (in particular, whether or not all the variables are measured at all the sampling locations). Some options include:

a. Single search: the data points are selected according to their proximity to the target location, irrespective of which variable(s) is (are) informed at these data points.

b. Two-part search: the first one is a selection of the nearby data of the primary variable, ignoring the information of covariates; the second one is a selection of the nearby data of the covariates, ignoring the information of the primary variable. The advantage of this strategy is to decouple the search of data according to the nature of the variable (primary or covariate). The disadvantage is that the search and the resolution of the cokriging equations must be carried out as many times as there are variables to estimate.

4.3 Results

4.3.1 Selection of variables

To estimate AGB we used ordinary cokriging (i.e., cokriging with unknown means that are assumed constant at the scale of the neighborhood) considering all the available covariates. Cokriging only works with linearly independent variables, for which there is no colinearity or “redundancy” of information. For this reason, the spectral bands TM1 to TM7 were excluded from the analysis since these variables can be deduced from that of the other variables such as Tasseled Cap components and NDVI. Also, the orientation variable
corresponds to an angle between 0 and 360 degrees, so the value 0 represents the same information as the value 360. In order to prevent very different values to represent equal or nearly equal angles, we replaced the angles by their cosines and sines. Thus, in addition to the primary variable (AGB), we used the following 11 covariates in all the analyses: altitude (ALT), orientation cosine and sine (ASPECT), slope (SLOPE), normalized difference vegetation index (NDVI), and the six Tasseled Cap components (TC1, TC2, TC3, TC4, TC5 and TC6).

4.3.2 Variogram analysis
While there is much information from the covariates (tens of thousands to millions of records) from which experimental variograms can be calculated in a very detailed way, information is scarcer with the primary variable (AGB) that has only a few hundreds of positions with field data. Because of the limited data available for the country, the inference of the variograms of the primary variable and the cross variograms between this variable and all covariates is difficult. To determine the spatial correlation structure, we chose one of the following alternatives, depending on the case under study:

a. Use of traditional experimental directional variograms, calculated along the identified main directions of anisotropy (Pantanillos case study).

b. Use of omnidirectional variograms, when anisotropy was not detectable for the primary variable (AGB) due to the scarcity of data (Quivolgo and Curacautín case studies).

c. Use of centered covariance as a substitute to the variogram (Chiles and Delfiner, 1999) in case the experimental variogram is too erratic (Monte Oscuro case study).

As an illustration, Figure 5 shows the experimental and modeled variograms for AGB, Brightness (TC1), Greenness (TC2) and Wetness (TC3), along the two identified main directions of anisotropy (north-south N0ºE and east-west N90ºE), for the Pantanillos area. For TC1, TC2 and TC3, the spatial variability appears to be greater along the north-south direction than along the east-west. These direct variograms, together with that of the other covariates and with the cross-variograms (a total of 78 variograms), have been jointly fitted thanks to the algorithm proposed by Goulard and Voltz (1992), using a nugget effect and nested exponential and power basic structures.

4.3.3 Cokriging neighborhood definition and application
To run cokriging, it is also necessary to define a neighborhood containing the data relevant to the local estimation. Given the very different number of data between the variables (AGB with very few data in comparison with covariates), we decided to use a two-part search:

a. For AGB we considered the 15 closest available data.

b. For each covariate, the 50 closest data were considered.

Cokriging was performed with an ad hoc MATLAB routine, since no known commercial software is able to perform cokriging with the above specifications and 11 covariates. The results are estimated values and error variances for AGB. The estimates were made for all the study areas, at the nodes of a grid with cells of 16m × 16m or 30m × 30m, depending on the case study, assuming unknown mean values for all the variables (ordinary cokriging).

Since not all the land in each area is covered by vegetation, we subsequently multiplied the estimates and error variances by the fraction of the cells located inside the identified stands, using vector digital layers of their boundaries. Figures 6 and 7 present the field data and identified stands, as well as the cokriging results for the Pantanillos area.
Fig. 5. Experimental (dots and dashed lines) and modeled (solid lines) variograms along the directions N0°E (black) and N90°E (blue) for AGB (top left), TC1 (top right), TC2 (bottom left) and TC3 (bottom right) in Pantanillos study area.

www.intechopen.com
4.4 Conceptualizing and generalizing the method

The estimation approach we developed is a real alternative for assessing AGB at communal and regional scales. It has a relatively low implementation costs when compared to traditional methods (i.e. inventories of land) and can be used to assess and monitor the stock of biomass over time on a country-wide scale. Although it requires field data, it is possible to estimate other areas using the already gained knowledge (in particular, the spatial correlation structure, via the direct and cross variogram models). The idea is based on two hypotheses:

a. The amount of biomass available at any point in the geographical space depends, first, on the productivity of the site and, second, on the status of the vegetation that grows on it.

b. Using satellite data (reflectance in several bands of optical range of the electromagnetic spectrum) and topography attributes, we can estimate both the productivity of the site and the vegetation status in an integrated manner. Then we can estimate the spatial variation of biomass concentrations. Using terrain data, the variation is scaled to the actual values for specific areas and subsequently aggregated at any available administrative division (stands, sites or districts).
In order to apply this approach in the estimation of AGB in any area the following steps should be done:

a. Selection of the geographic area: It is first necessary to define the limits of the study area and the stands of interest. One may not be able to divide into stands if one does not have enough previous information and the results can be aggregated in the proper spatial division later (see step f).

b. Gathering and processing satellite images: The images used should be subject to the following treatments:
i. Eliminate the presence of clouds and systematic errors. In the case of Landsat ETM + error must be corrected SLC-off explicitly.
ii. Geometric correction for integration into GIS environment.
iii. Radiometric correction to reduce topographic and atmospheric errors.
iv. Calculation of vegetation indexes and Tasseled Cap components.
c. Collection of field data: Field data can be collected especially for AGB estimation or can be available from previous inventories. Data should be expressed in biomass per unit area (i.e. hectare) and associated with a point in geographical formal space (i.e., UTM coordinates). Whatever the case, the field estimations of biomass can be made in any of the following cases:
i. If data are available at the individual tree, apply allometric equations by species.
ii. If data are aggregated for each plot, apply biomass functions.
d. Data checking and cleaning: To perform the geostatistical analysis, it is necessary to create integrated data tables that incorporate both the biomass data, obtained from field data, plus all available covariates (satellite and topographical). The data should be reviewed, validated, and one should eliminate those displaying an aberrant behavior.
e. Geostatistical modeling: Using the data structure obtained in step d, perform geostatistical analysis including the following sub-steps:
i. Analysis of multivariate spatial correlation (variogram analysis) to obtain experimental direct and cross variograms and to fit variogram models required for interpolation.
ii. Obtain local estimations of biomass and levels of uncertainty (estimation error variances) via cokriging throughout the study area.
iii. Masking for estimating administrative units or stands of interest. To get accurate estimates, it is important to rule out areas with no vegetation.
f. Report results and prepare associated maps.

In the previous paragraphs, one of the most important steps is the collection of field data. These data allow modeling the autocorrelation of biomass in space and its cross-correlation with satellite and topographic variables. When these data are available, the procedure for the estimation should be the one already described. However, sometimes no data are available or they cannot be collected on time or the costs are not reasonable. In this case, one may use known variogram models to make a "blind" estimation using a slightly alternative method: simple cokriging, which assumes that the average biomass is known. We propose that this average can be obtained using the forest growth simulators that are available in many countries, e.g., Radiata and Eucasis growing models and software in Chile.

5. Conclusions
The use of covariates extracted from SRTM elevation model and LANDSAT images allows for estimates with a greater level of detail than those obtained by using only data field. This can be corroborated by comparing the estimates by cokriging with those that would be obtained with univariate kriging. We hypothesize that the set of covariates accounts, through their spatial dependence, for two fundamental aspects to explain existing biomass:
a. The quality of the forest site. Topographic variables (altitude, slope and orientation) inform in a synergistic way about the quality of the specific potential growing conditions (forest site).
b. Vegetation health and vigor are jointly captured by the vegetation indices and Tasseled Cap components.
The approach to this scale of work, using medium spatial resolution sensors, is to estimate AGB independently of species or plant community types and then overlap stands boundaries, or any other administrative division, to obtain the associated total AGB stocks. Coregionalization models (direct and cross variograms) are simplifications of reality; in particular, they may not detect anisotropies when having a too small amount of field data, allowing only for an omnidirectional inference. The number of field data should be such that the inference of AGB variogram is feasible, which is usually achieved with more than 100 data in each stratum. In order for this number to be reduced, we recommend the sampling of the area (site) to be as regular as possible, i.e. avoiding samples clustered in space.

In the future, we suggest identifying other covariates correlated with AGB, for example, an indicator of the amount of sunlight available or a local indicator of the ground geometry (concave, convex or plane) that is related to the availability of water. These indicators could be calculated from a digital elevation model, already available at several spatial resolutions.

6. Acknowledgment

We would like to express our thanks to Biocomsa Consortium, and therefore INNOVA CORFO CHILE, for funding most of the field work, data acquisition and processing. Also, we thank Miss Paz Acuna and Miss Lissette Cortes from GEP Lab. at Universidad de Chile, for their support in reviewing the text and for providing the reference cartography. Also we thank Helen Grover for proof reading the text.

7. References


www.intechopen.com


www.intechopen.com


The accurate measurement of ecosystem biomass is of great importance in scientific, resource management and energy sectors. In particular, biomass is a direct measurement of carbon storage within an ecosystem and of great importance for carbon cycle science and carbon emission mitigation. Remote Sensing is the most accurate tool for global biomass measurements because of the ability to measure large areas. Current biomass estimates are derived primarily from ground-based samples, as compiled and reported in inventories and ecosystem samples. By using remote sensing technologies, we are able to scale up the sample values and supply wall to wall mapping of biomass. Three separate remote sensing technologies are available today to measure ecosystem biomass: passive optical, radar, and lidar. There are many measurement methodologies that range from the application driven to the most technologically cutting-edge. The goal of this book is to address the newest developments in biomass measurements, sensor development, field measurements and modeling. The chapters in this book are separated into five main sections.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:


InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
Chinese International Commerce College, 200040, China
Phone: +86-21-62489820
Fax: +86-21-62489821

Remote Sensing of Biomass - Principles and Applications
Edited by Dr. Lola Fatoyinbo
Hard cover, 322 pages
Publisher InTech
Published online 28, March, 2012
Published in print edition March, 2012
© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.