Caractérisation des forêts de montagne par scanner laser aéroporté : estimation de paramètres de peuplement par régression SVM et apprentissage non supervisé pour la détection de sommets

Jean-Matthieu Monnet

Jean-Matthieu Monnet. Caractérisation des forêts de montagne par scanner laser aéroporté : estimation de paramètres de peuplement par régression SVM et apprentissage non supervisé pour la détection de sommets. Autre. Université de Grenoble, 2011. Français. NNT : 2011GRENT056 . tel-00652698

HAL Id: tel-00652698
https://theses.hal.science/tel-00652698v1

Submitted on 16 Dec 2011

UNIVERSITÉ DE GRENOBLE

THÈSE
POUR OBTENIR LE GRADE DE
DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE
Spécialité : Signal Image Parole et Télécom
Arrêté ministériel : 7 août 2006

Présentée par
Jean-Matthieu MONNET

Thèse dirigée par **Jocelyn CHANUSSOT** et
codirigée par **Frédéric BERGER**

préparée au sein du
Cemagref de Grenoble, UR Écosystèmes Montagnards
et du
GIPSA-Lab
dans l'**École Doctorale :**
Électronique, Électrotechnique, Automatique et Traitement du Signal

Caractérisation des forêts de montagne par scanner laser aéroporté : Estimation de paramètres de peuplement par régression SVM et apprentissage non supervisé pour la détection de sommets.

Thèse soutenue publiquement le **25 octobre 2011**,
devant le jury composé de :
Mr Paul D. GADER
Professeur, Université de Floride, Président
Mme Barbara KOCH
Professeur, Université de Freiburg, Rapporteur
Mme Florence TUPIN
Professeur, Telecom ParisTech, Rapporteur
Mr Markus HOLLAUS
Maître de conférences, Université Technologique de Vienne, Examinateur
Mr Jocelyn CHANUSSOT
Professeur, Grenoble INP, Directeur
Mr Frédéric BERGER
Ingénieur de recherche, Cemagref, Directeur

PhD Thesis
Jean-Matthieu MONNET

Thesis advisors:
Jocelyn CHANUSSOT
Frédéric BERGER

Cemagref Grenoble, UR Écosystèmes Montagnards
and
Grenoble-INP, GIPSA-Lab

Using airborne laser scanning for mountain forests mapping: Support vector regression for stand parameters estimation and unsupervised training for treetop detection.

Commitee
Mr. Paul D. GADER
University of Florida, Chair
Mrs. Barbara KOCH
University of Freiburg, Referee
Mrs. Florence TUPIN
Telecom ParisTech, Referee
Mr. Markus HOLLAUS
Technological University of Vienna, Examiner
Mr. Jocelyn CHANUSSOT
Grenoble-INP, Supervisor
Mr. Frédéric BERGER
Cemagref, Supervisor

Remerciements

Merci à Jocelyn et Fred pour leur confiance et d'avoir été disponibles pendant ces trois ans.

Merci aux membres du jury d'avoir pris le temps d'évaluer mon travail.

Merci à Éric pour la détermination de l'érable palestinien et les milliers d'arbres inventoriés.

Merci à Nicolas d'avoir partagé la bonne ambiance et les bougonnements du 024.

Merci à tous ceux qui ont participé de près ou de loin à ce travail au Cemagref et à Gipsa-Lab, sur le terrain, au bureau, à parler d'arbres, de LiDAR, de cailloux ou de plein d'autres choses passionnantes : Franck, Paul, Jérôme, Christophes, Pascal, Julie, Nathan, Sébastiens, Oliver, Catherine, Alberto, Paula, Won-Sang, Vincent, Félix, Lucia …

Merci à Laurent Descroix, Alain Munoz, Jérôme Bock et Anne Jolly de l'Office National des Forêts pour leur contribution à l'application concrète de ces travaux.

Merci à ma famille et mes amis.

Merci à Sacha et Timothé de m'avoir empêché de ramener des algorithmes à la maison.

Cette thèse a été financée par le cluster de recherche ENERGIES de la région Rhône-Alpes. Les données LiDAR utilisées pendant ces trois ans ont été financées par les projets Interreg PROVIALP, IFP et MANFRED.

"It is unfortunate that trees are tall and people are short because it is more reliable to assess diameter at some higher point."

M.D. Smith, B.C. Larson, M.J. Kelty, and P.M.S. Ashton,
"The practice of sylviculture: Applied forest ecology,"
9e éd. John Wiley & Sons, New York, 1997


Éric Mermin

Contents

Part I

Introduction

Chapter 1

Mountain forests: an unknown resource

1.1 Definitions

1.1.1 Mountain areas: a complex environment

Mountain areas are principally associated to the idea of a harsh environment due to topographic or climatic constraints. However this general idea is expressed differently in the national definition across European countries. A report about mountain areas in Europe [3] proposes a unique reference based on altitude categories combined with slope and elevation range criteria (table 1.1). This classification reflects the fact that even at lower altitudes, where climatic conditions are milder, extreme topographic features (e.g. average slope or local variations) may also qualify an area as mountain. In this document, this general definition of a mountain area is adopted.

Table 1.1: Definition of European mountain areas (after [3])

Elevation category (m)	Additional criteria
> 2500	none
1500−2499	slope greater than 2 degrees within 3 km radius
1000−1499	slope greater than 5 degrees within 3 km radius or local elevation range greater than 300 m within 7 km radius
300−999	local elevation range greater than 300 m within 7 km radius
0−299	standard deviation greater than 50 m for cardinal points

In mountain areas, rough and steep topography induces a high variability of factors affecting vegetation and more particularly trees: soil type, water resources, exposure, temperature. External stress factors such as snow and avalanches, erosion and rockfalls also have a high influence on forests. The resulting ecosystems exhibit both diversity and complexity. Forests display a great heterogeneity vertically and horizontally at local scale. At large scale, the climate variations linked with altitude gradient determine various vegetation belts characterized by different ecosystems [240] (figure 1.1). Forest are located from downslope to the subalpine areas. In this document, mountainous forests will refer not only to montane or subalpine areas, but also to forests located in the hill area.

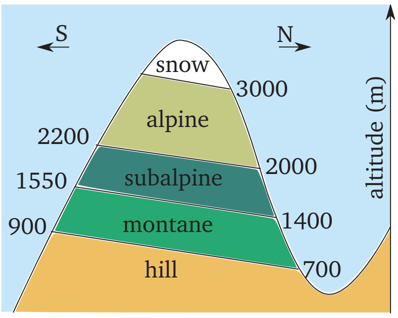
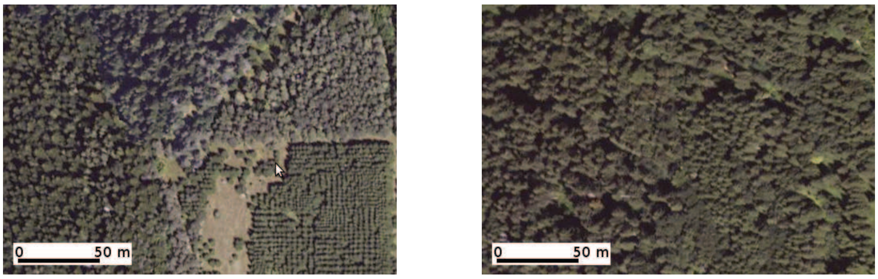


Figure 1.1: Vegetation belts in the external Alps (after [10, p. 158]).

1.1.2 Forest stands in mountainous areas

"*A stand is a spatially continuous group of trees and associated vegetation having similar structures and growing under similar soil and climatic conditions*" [239]. Due to the characteristics of mountain forests, the delineation of homogeneous areas, both in terms of forest structure and environmental conditions, is not straightforward. Moreover, in complex stands the intra-stand variability may be of similar magnitude than inter-stand variability [26, p. 88]. The objective of the selection of homogeneous areas implies scaling down to the level of small tree groups. A study conducted to determined the optimal spatial resolution for the delineation of mountainous forest stands from optical images showed that there is no specific resolution that allows a precise pixel-level delineation of all stands [58]. Figure 1.2 displays two forested areas located in the town of Villard-de-Lans, France. On figure 1.2a, five stands can be delimited, each one being rather homogeneous. The extreme example is the young plantation at the bottom right. On figure 1.2b, the forest is more heterogeneous and stand delineation is not so easy. A trade-off between intra-stand variability and stand size has to be defined, depending on both the available prior knowledge of the area and on inventory objectives.



(a) Homogeneous forest patches

(b) Heterogeneous forest

Figure 1.2: Orthophotos of two forested areas around Villard-de-Lans (BD ORTHO®).

In less complex environments, forests stands depend on growth conditions but also on the silvicultural operations that have modelled them through time. A simple example is the even-aged plantation, where all trees in a given stand have the same age. Frontiers between adjacent stands with different ages are well-delimited. In mountainous environments, silvicultural operations are made at the tree [270] or tree group level [204], so that human interventions do not imply the transformation of forests into a patchwork of homogeneous stands. Forests stands are organized in units with continuous transitions that evolve differently depending on the local conditions. Growth conditions and operational constraints thus prevent the regularization of structures but allow the persistence of "irregular", stratified stands [263, p. 95]. This heterogeneity poses a double problem. A fine mapping of forest resources is required to optimize the silvicultural operations, but the mapping operation itself is all the more problematic since forests display high variability.

1.2 Limits of current inventory methods

1.2.1 Field inventory

Exhaustive knowledge of a forest requires the full inventory of forest stands, which is all the more labor-intensive and costly as forest accessibility is limited. In mountains the access network is not dense and topographic conditions such as slope and terrain roughness make off-track moves very difficult. Except for particular cases, such as harvesting planning in stands with high economical value, full inventories are rarely implemented. Standard procedures are now sample-based inventories, where plots distributed regularly or according to a stratified scheme over the investigated area are inventoried [69]. Depending on the sampling intensity and on the variability of forests, such procedures provide reliable forest attributes estimates at the compartment or higher level, i.e. for areas larger than several hectares. Unfortunately, field data collection is still laborious and difficult to implement in remote areas, and lacks in information on the spatial distribution. Indeed, the scale of analysis is not sufficient for operational purposes, when tree groups are considered. For an efficient forest management in mountainous areas, a finer mapping of forest attributes is required, but is cost-prohibitive for large areas.

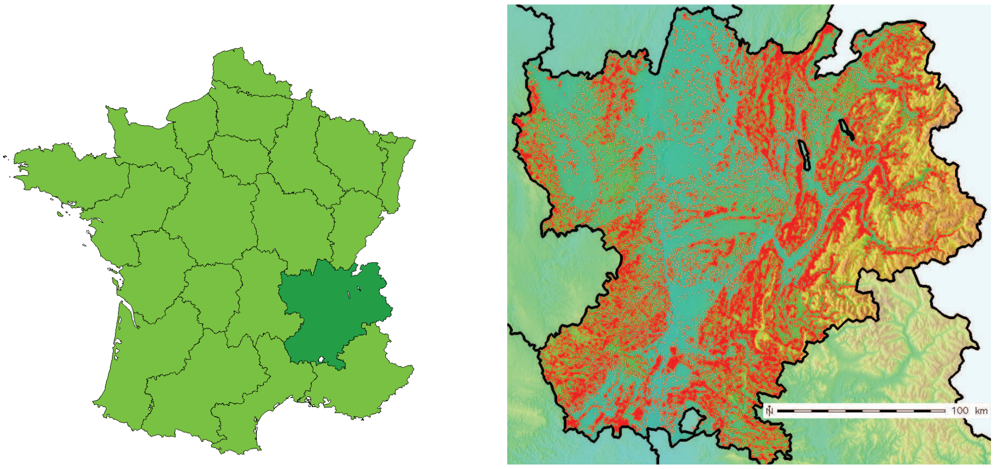
In the Rhône-Alpes region, more than 60 % of production forests (poplar plantations excluded) are located in areas with slope greater than 30 % (table 1.2). Figure 1.3 shows a map of forests in the Rhône-Alpes region. They are mainly located in mountainous areas: Alps and Jura mountains in the East, and Massif Central at the West of the Rhône valley. Therefore, areas where the lack of spatial information represents a major restraint for forest management accounts for a significant part of forest resources.

Table 1.2: Production forests (poplar plantations excluded) area and volume in Rhône-Alpes region, by slope category (data from the French Forest Inventory, available at http://www.ifn.fr/spip/?rubrique18, accessed May 9th, 2011)

Slope category (%)	Area ($\times10^3$ ha)	Volume ($\times10^6$ m^3)
0−5	109±15	16±4
5−15	138±16	22±6
15−30	279±22	53±8
30−50	448±27	86±10
> 50	462±26	90±9
unknown	38±10	7±3
total	1475±31	274±14

1.2.2 Remote sensing

With the growing number of operating sensors and commercially available data, numerous studies have tested the possibility to map forest structure variation at fine scale by combining field data with spectral imagery [123] or radar [295]. Despite the spatial resolution attained by recent sensors and the possibility to estimate attributes at the tree level, the extraction of information on the vertical and horizontal canopy distribution is not straightforward, as geometric information is only indirectly assessed, e.g. by correlation methods. Besides, the variability of forest stands in France makes the fully-automated retrieval of forest parameters from optical data a difficult



(a) Rhône-Alpes region (b) Forest areas (🟥 red)

Figure 1.3: Forest areas in Rhône-Alpes region (category 31 of Corine Land Cover database, 2006 [5]). Background is the digital terrain model (BD ALTI®).

task. Until recently, the French National Forest Inventory mainly used manual stereo photo-interpretation for land cover classification and stand delineation, and quantitative estimates were only based on field plots. At large scale (e.g. French administrative département or région), these values are trustworthy, but they are not available at finer resolution.

1.3 Increasing stakes

Despite its harshness, the mountainous environment has been modeled by human activities for centuries and forests still assume major functions regarding the local and global economy:

- timber production (figure 1.4a);
- outdoor recreation area for hiking, skiing . . . (figure 1.4b);
- protection of human stakes against natural hazards (figure 1.4c);
- watershed protection;
- reservoir of biodiversity (figure 1.4d).

(a) Timber logging (b) Leisure area (c) Block stopped by tree (d) *Glaucidium passerinum*

Figure 1.4: Some services provided by mountainous forests: economic (a), social (b), protection (c) and ecological (d).

In some rural areas, the wood sector still represents an important part of the economical activity. Since the beginning of the 1980's, mountain forest have been progressively neglected due to increasing labor costs and decreasing wood prices. As a result, many forests have been continuously increasing their growing stock and have reached a development stage where stand stability is in jeopardy. Besides, forests are now gaining consideration with the economic and social requirements for local and renewable materials and energy [89].

Beyond the wood and paper industry, forest management is also a major tool for the preservation of mountainous areas. During the second half of the nineteenth century, a broad afforestation program was undertaken by the French water and forest administration in order to mitigate erosion and natural hazards in degraded areas. 150 years later, some stands reach a stage at which their protective effect might decrease significantly. More generally, forest stands ensure protection of human assets against natural phenomenons such as avalanches, rockfalls and torrential floods. Their protective effect depends on the forest structure and varies with development stages (figure 1.5). Forest attributes have to be precisely monitored so that their protection effect is reinforced by silvicultural operations, or complemented by civil engineering if necessary [10, p. 31]

At global scale, concerns about greenhouse gases and the loss of biodiversity are growing. Mountainous forests now represent a major source of woody biomass. Besides, due to inherent ecosystems diversity and to a lower human density, they are also natural habitats to numerous plant (lady's slipper orchid *Cypripedium calceolus*...) or animal species (e..g hazel grouse *Bonasia bonasia*, Rosalia longicorn *Rosalia alpina*, Geoffroy's bat *Myotis emarginatus*...) [10, p. 260].

This fragile equilibrium is constantly threatened or altered by economic and social changes. To meet the increasing demands toward forest-related services while preserving the resilience of mountain areas, methods that allow a cost-effective monitoring of forest parameters and structures at relevant spatial resolution are needed.

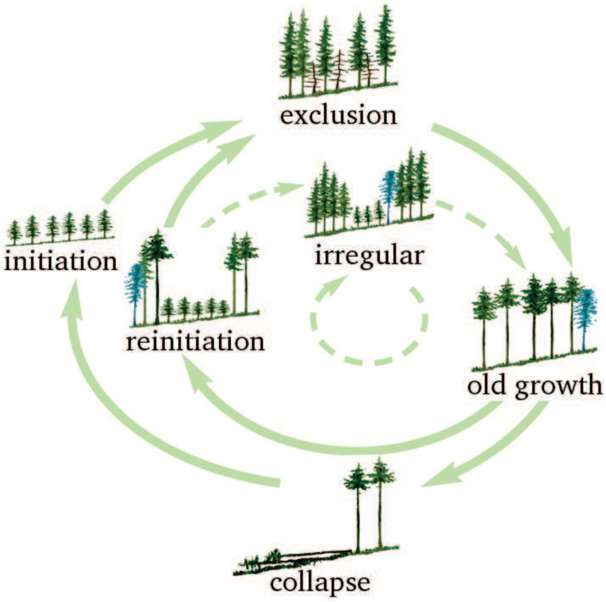


Figure 1.5: Spruce development stages in the northern Alps (adapted from [10, p. 61] and [263]).

Chapter 2

Airborne laser scanning for forest applications

2.1 A fast developing technique

2.1.1 History

Airborne laser scanning (ALS) is a remote sensing technique based on the measure of the flight time of laser pulses emitted from an aircraft and reflected by objects located on the ground. Its fast development during the two past decades has been made possible by the technological advances of global positioning system, inertial navigation, and lasers. At the beginning, LiDAR (Light Detection And Ranging) sensors boarded in aircrafts or satellites only operated on one-dimensional (1D) profiles along the platform path. Sensors are now equipped with orientating devices and are able to scan large swaths along the platform trajectory. By the end of the 1990's, the pulse repetition frequency of small footprint, commercial sensors was around 10 kHz [18], and service providers were only emerging. Now, leading sensors achieve a pulse repetition frequency of 300 kHz with the multipulse technology. Some countries have undergone a complete LiDAR coverage (Switzerland, Denmark) and wall-to-wall mapping is under way in some others (Finland, Sweden).

2.1.2 Principles

The main components boarded in the aircraft are:

- global positioning system (GPS),
- inertial measurement unit (IMU),
- laser emitter-receiver scanner,
- storage device.

A laser pulse is fired toward the Earth in a direction given by the orienting device (oscillating mirror, rotating prism). For vegetation and topographic mapping purposes, the laser wavelength is usually around 1000 nm, which assures correct reflectance values for a broad range of materials. Whenever the laser pulse is intercepted by an object, part of the energy is reflected toward the receiver and recorded. When the object is not solid or too dense (e.g. tree branches), a sufficient part of the laser beam may also continue its trajectory and be reflected by lower elements, eventually the ground surface (figure 2.1). This characteristic of laser scanning is of high importance as it allows to characterize the underlying 3D structure of the vegetation and to accurately detect the soil surface even in dense canopy areas.

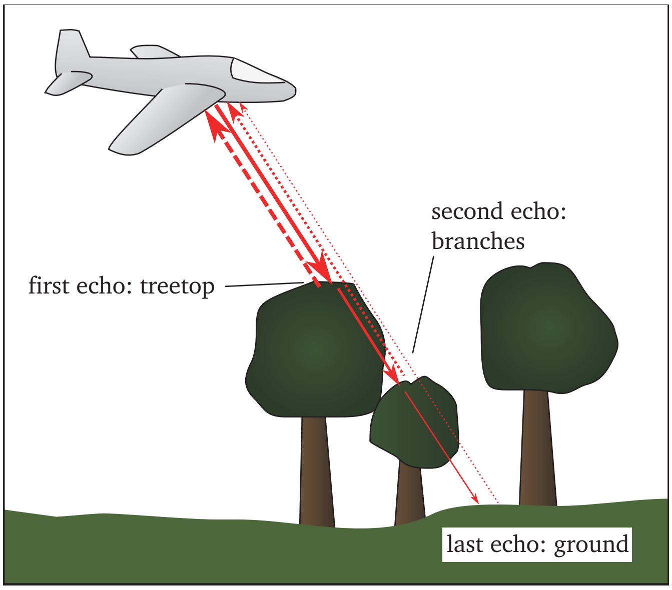


Figure 2.1: Principle of range measurement by airborne laser scanning.

The range to reflecting objects is computed as $R = c \times \frac{t}{2}$ where t is the fly-around time and c the pulse speed in the atmosphere, assumed to be equal to the speed of light. On analog scanners, t is determined in real-time by detecting the time when the leading edge of the reflected signal reaches a given threshold [17]. Such systems are able to store one to several echoes for each emitted pulse. They are called multi-echo laser scanners. In recent scanners, storage capacity has been greatly increased so that the full waveform of the reflected pulse is digitized. The post-processing of the data allows a finer detection (number and coordinates) and characterization of objects located within the footprint of the laser pulse [116, 307]. The footprint, which is the surface illuminated by the laser beam, depends on the divergence of the laser and on the range. Laser scanner with footprint size smaller than one meter are considered "small-footprint". Their range accuracy, which mainly depends on the width of the emitted pulse, is usually around 15 cm. The orienting device makes laser pulses sweep across the aircraft trajectory. Swath width is determined by the flying altitude and the scanning angle. LiDAR data are usually acquired in multiple, overlapping flying strips. A comprehensive description of relations and formulas for airborne laser scanning is available in [17].

Back in the office, objects coordinates are computed with the information from the on-board and ground-based components:

- range and pulse angle (laser emitter-receiver),
- plane trajectory (on-board GPS and ground reference stations),
- plane attitude: yaw, pitch and roll (IMU).

Further processing includes laser strip adjustment to correct for hardware calibration errors, and detection of erroneous points: high points due to atmospheric noise and low points due to multiple reflections. For a small-footprint laser scanner, accuracy of the final 3D point cloud is typically 25 cm (planimetric) and 15 cm (altimetric). Point density varies greatly depending on acquisition parameters and desired applications, from one to more than one hundred points per square meter. For small footprint lasers, the 3D point cloud contains precise geometric information related to the objects which constitute the canopy cover. For large footprint lasers, the coarser planimetric resolution is generally compensated by the full waveform digitization (e.g. LVIS [25], SLICER [100, 173] and spaceborne GLAS/ICESAT).

2.2 Airborne laser scanning in forestry

2.2.1 First steps

The ability of LiDAR sensors to detect ground surface even in forested areas was first used for topographic purposes, mainly to derive accurate digital terrain models (DTM). However, the accuracy of the geometric information about the vegetation structure quickly sparked interest among the foresters. Following the development pace of the technology, forest stand parameters were first estimated with LiDAR profiling altimeters [222], and then with laser scanners [223]. Height variables [187, 222, 223, 226] and stand volume [223, 227] were the first forest parameters investigated but laser data also proved efficient for other stand parameters, such as leaf area index, diameter and number of stems [173] or basal area [175].

At this time, large footprint LiDARs were mainly operated by NASA for research purposes, but commercial small-footprint were becoming widely available thanks to topographic acquisitions [200]. These first works showed the great potential for forestry applications, from stand parameters estimation to land cover classification, wildlife management and habitat mapping [67]. Soon, pre-operational studies for large-scale stand attributes mapping were undertaken [199, 205].

2.2.2 Fast development

In the early 2000's, research on forest applications of airborne LiDAR developed considerably. The existing methods, that had been principally tested on coniferous forests, were validated in other contexts: with SLICER data over broadleaved forest [100], temperate deciduous, temperate coniferous and boreal forest [174], with LVIS data on tropical rain forest [63, 315], with small footprint LiDAR on young forests [233] or northern hardwood forests [180].

The area-based method, which consists in the calibration of models linking quantiles of the point cloud vertical distribution to forest variables, proved very convenient for the retrieval of stand parameters such as height, mean diameter, stem density, basal area and volume [228, 231], or above ground biomass [181]. The method turned out to be efficient in various forest contexts [233] and quite robust to the laser point density [122]. It was quickly implemented at operational scale [231]. It was also successful in estimating tree properties e.g. total and crown height [237], and also more complex stand patterns such as the Weibull parameters of the diameter distribution [90].

Even tough this pragmatic approach adopted to predict forest parameters from laser-derived quantiles proved efficient, further investigation of the waveform signal and its interaction with the canopy was undertaken:

- simulation of LiDAR waveforms reflected by the canopy [285],
- decomposition of waveforms as sums of Gaussian components [116],
- modelling of waveforms with a time-dependent stochastic radiative transfer model [160].

Meanwhile, technological improvements in pulse repetition frequency soon allowed point cloud densities sufficient for single tree delineation from small footprint LiDARs [136]. Most of the studies relied on image processing techniques applied to canopy height models (CHM) [28, 137, 168, 252, 254] but some also tested point cloud processing [14, 213] (figure 2.2). The methods were rapidly improved in order to extract as much information as possible from the geometric features of the delineated trees, e.g. crown diameter [255], length or volume. Researchers soon faced the problem that suppressed trees were not easily detectable, hence the impossibility to predict height or diameter distributions with this sole method. As a result, deriving stand-level attributes from the aggregation of tree-level information remained problematic [190].

Whatever the approach, single-tree or area-based, concerns arouse about the sensitivity of the algorithms to both the LiDAR acquisition parameters and the forest structure. Indeed

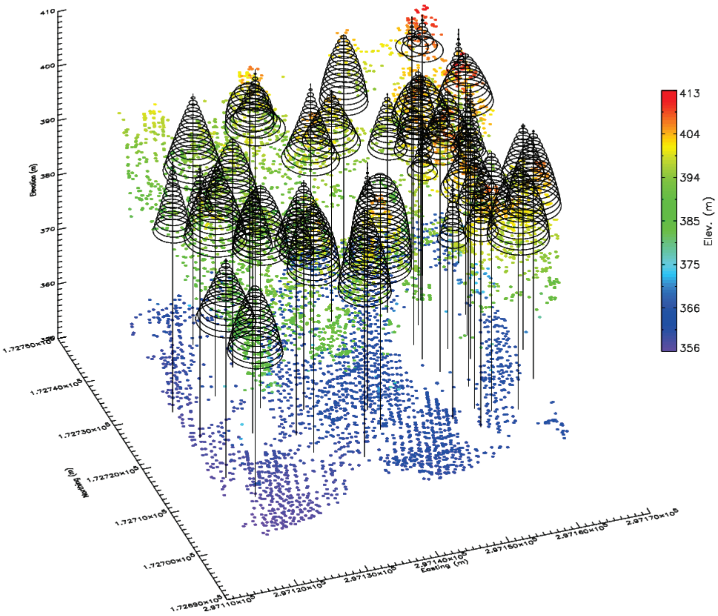


Figure 2.2: Single tree reconstruction from Bayesian object recognition in the point cloud, image from [14].

the increasing number of different scanners and the various possibilities for planning a LiDAR acquisition, altogether with the cost of LiDAR data, raised the issue of the optimization of surveys. Some studies relied on real data to assess the influence of acquisition parameters on estimated forest attributes, e.g. with LVIS data in tropical environments [64], or with small footprint LiDAR (effect of sampling on height underestimation [88], effect of flying altitude [230]). Simulations were also used as a convenient way to test several parameter combinations [120,319] which would have been unaffordable in real-size experiments.

While some studies focused on the potential of LiDAR data alone for the retrieval of various forest parameters, e.g. geometric information [121, 297] or intensity information [180, 184] for species classification, others tried to take advantage of the synergy with optical sensors. Indeed the spectral information is more relevant than ALS data for tree species identification [99, 136, 205]. Besides, the large-scale availability of optical spaceborne data is useful for wall-to-wall mapping of forest attributes such as canopy height [130,205,317]. Indeed, LiDAR can be used as a calibration or validation tool for the structural analysis of vegetated ecosystems with spaceborne optical sensors [44]. Moreover, the combination of LiDAR and spectral data proved to be superior to the use of any of them alone, e.g. for tree crown delineation [49].

The potential of LiDAR data for a broad range of ecological applications was quickly identified [177], including its use in carbon content [131, 250] or gross primary production [161] studies.

2.2.3 Diversification

At this stage a great part of exploratory analysis of LiDAR potential for forest applications was done and basic methods had been tested on numerous datasets. Then LiDAR research expanded to a wide range of applications and tackled large-scale related issues.

Investigations on waveform modelling [41, 42, 145, 247], calibration [35, 158, 307] and on the forest/signal interaction [40, 157, 286] increased as more and more full waveform small footprint scanners were commercially available. These methods allow a finer detection of echoes and the extraction of calibrated attributes. Such information proved useful for species classification [8, 306]. Reversely, pseudo-waveforms can be reconstructed from multi-echo sensors data [218].

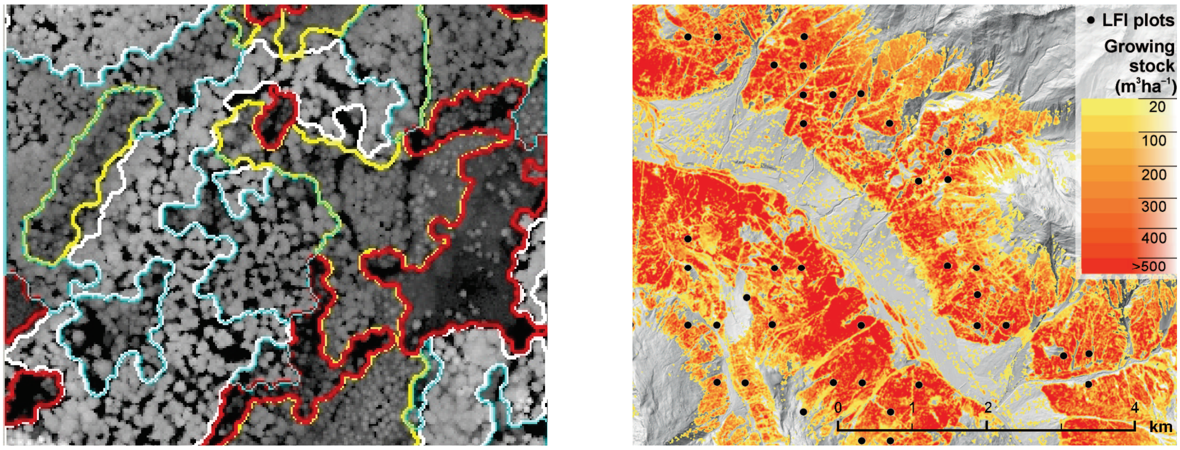
The issue of tree species classification indeed turned out to be an important pre-requisite for stand classification or species-specific attributes estimations. Some interesting results were obtained with LiDAR data alone [27, 59, 121, 159, 209, 287, 300], by taking advantage of points attributes, such as intensity [152, 241], backscatter coefficient [8, 306] or directly from the full waveform data [103, 261]. Species classification was also facilitated by the synergy with spectral data. The issue of deriving species-specific estimates received considerable attention [129], e.g. for volume [224, 243], stem number and diameter [244], or diameter distribution [253].

Synergies with other remote sensing data were also investigated. LiDAR was compared [128, 151, 271] or used with radar data [133, 186, 221, 275]. Comparison were also made with multispectral data [49, 248, 305] and combined used of the data was tested [186, 196, 256], even at the single tree level [163, 168, 283], or with the underlying idea to combine LiDAR accuracy with optical wall-to-wall cover [107, 197, 288, 308].

Indeed, the methods were increasingly used at operational scale, e.g. in Scandinavia [229,232], England [309], Alaska [11], Denmark [225]. This raised the issue of the validity of models developed at smaller scales [97, 102, 176, 188, 281, 291]. Several studies proposed or compared various prediction methods [75,189,244]. At operational scale, the problem of the aggregation of stand parameters from single tree attributes remained unsolved. The estimation of tree distribution [195] or the calibration of tree lists [182] have been tested as possible solutions, and alternative methods were proposed for the imputation of single tree attributes [299]. Some of them may be seen as a convergence between the single-tree and the area-based approaches, the considered areas being previously delineated single tree segments [30, 31].

2.2.4 Applications

The range of LiDAR applications kept broadening, applying to several scientific fields and different scale levels. Regarding land cover classification [32, 33], LiDAR proved efficient for stand delineation and forest classification [52, 57] (figure 2.3a). Some studies focused on particular ecosystems, such as riparian forests [15], or cottonwood [78–80]. LiDAR is generally a well-suited tool to qualify the structure of forest stands [47,93,98,127,147,150,167,207,214,249,326], but also of the understory [114, 141, 194] or ground vegetation, e.g. with herb-rich forests [302]. Particular or general features of the forest can be detected or assessed with laser data, e.g. canopy gaps [20, 87, 162, 323], canopy fractional cover [126, 198, 274], forest maturity [313].



(a) Automatic stand delineation in the Black Forest (Germany, image from [57])

(b) Stem volume map of Vorarlberg (Austria, image from [117])

Figure 2.3: Examples of operational applications of LiDAR remote sensing

Such information is then useful for ecological applications such as dead standing biomass assessment [153], habitat mapping [21,22,108,132,219], particularly related to avian species [46, 96,109], or ecosystem studies (light interception [170], leaf area index [38,142,267,269,278,324], forest pigment mapping [24], pest control [48,266]).

Georeferenced data is of great interest for forest management. This includes risk management, e.g. with forest fire modelling [265], and estimation of fuelwood [12,73,183,264,273] and past damages [165,316,318]. Predicted stand parameters are also useful to quantify the forest protection effect against rockfalls [210,211]. Silvicultors are of course interested in all stand parameters introduced in the previous paragraphs (height, diameter, basal area, stem number, volume...). Economical analyses showed that the investment in LiDAR-based inventory procedures are justified by better resource management [71,124,140]. At single tree level, the geometric information (e.g. crown base [54,298], 3D surface reconstruction [149], stem volume [53]) allows a better allocation by the forest practitioner, and can be used as input for virtual training environments [85]. LiDAR data is also useful for short and long-term forest change detection. Growth or harvesting operations [235,320,321], tree migration [236], pest infestation [48], disturbance dynamics [303,304] can be precisely monitored. Studies also showed the possibility of retrospective analysis of optical data [279,280,301].

Today's context of global warming and all its implications in global carbon cycle and stocks assessment have also oriented LiDAR investigations. Several studies have addressed the issue of biomass mapping [68,86,143,290,294,325].

2.2.5 Main achievements in mountainous or complex environments

After the first experiments on coniferous forests, a broader range of forest conditions was investigated. Single tree methods were tested on temperate deciduous trees, including coppice individuals [28], heterogeneous mixed stands in boreal forest [193], uneven-aged broadleaved stands [163] or mixed stands [104]. Area-based methods were also tested in various conditions, e.g. broadleaved [139], mixed [106,238], mixed multilayered forest [292]. In complex stands, such as mixed deciduous ones, the forest structure was more difficult to characterize [112]. Additional LiDAR-derived attributes were proposed to address this issue in multilayered forests [169].

Alpine environments were also investigated. The effect of footprint size and sampling density [110], topographic factors [111] on individual tree delineation were evaluated. Major works include the retrieval of canopy structure from LVIS data in montane forests [132], estimation of tree heights in sugi (*Cryptomerica japonica*) plantations [289]. Single tree [104] and area-based [106] methods were tested on forests stands located in Germany. Studies also tried to define simple indices able to predict stand parameters at operational scale [117–119,143,272] (figure 2.3b).

Chapter 3

Objectives

If LiDAR-based inventories are now commonly used for commercial mapping of forest resources in boreal forests, their operational development in areas where they would bring the most added-value is still lagging. In France, a combination of several factors explains this lateness:

- the complexity of forest stands that prevents a simple transfer of methods developed in the Scandinavian countries,
- the lack of geolocated field data required to test and calibrate the retrieval models,
- the absence of commercial or public entity willing to perform a large-scale test of LiDAR-based inventory,
- the absence of concerted effort from the public authorities at various levels toward the provision of consistent LiDAR datasets.

Meanwhile, there are high chances that this technology could solve several issues related to the management of forest stands, particularly in mountainous areas where multiple aspects are at stake.

This thesis aims at investigating the potential of LiDAR data for the retrieval of forest parameters in mountainous forests, with a specific focus on three complementary aspects.

Due to the fast pace of scanner development and to the complexity of LiDAR surveys, there is no standard procedure for the acquisition of ALS data for forestry applications, nor standard LiDAR products. Chapter 4 introduces the main issues linked with acquisition settings, and presents the workflow implemented to produce the ALS data used this work. Using airborne laser scanning for the measure of geometric features of trees implies a change of point of view for the pedestrian forester. Conventional tools used for field data collection must be questioned as the point cloud is theoretically more accurate. The issues linked with the relevance of field data protocols, emphasizing the constraints due to the mountainous environment, and the field data used in this work are presented in chapter 5.

The single tree method requires less field data for model building, and produces information at the tree level, which is of high added-value for operational use in mountain forest management. Key issues are the robustness of detection algorithms regarding the forest structures and the derivation of unbiased aggregated stand parameters. In this thesis, an algorithm based on local maxima filtering is proposed, as well as a procedure for the automated assessment of detection performance (chapter 6). With this framework, the sensitivity of the algorithm to parametrization and to forest structure is investigated (chapter 7). Finally, an unsupervised approach is proposed for the parametrization of the algorithm (chapter 8).

Unfortunately, this method is not relevant for coppice stands where each tree is constituted of multiple stems from a single stump (figure 3.1), or when the point cloud density is not sufficient to identifiy single trees. Therefore, the area-based method (chapter 9) has also been tested on coppice stands, while tempting to address the some issues related to mountain environment: variability of forest structures and lack of field data. The effect of training plot size on models accuracy is investigated in chapter 10. Chapter 11 compares the retrieval accuracy of multiple regression with a non-linear regression method based on support vector machines and addresses the problem of taking advantage of the information of the point cloud with a limited number of training data by testing dimension reduction techniques such as independent and principal component analyses.

Figure 3.1: Common whitebeam (*Sorbus aria*) coppice.

Part II

Material: airborne and field data

Chapter 4

Airborne laser scanning data

4.1 Acquisition and pre-processing

4.1.1 Acquisition parameters

LiDAR acquisition settings are more complicated than for an optical survey. Depending on the laser scanner, but also on the type of aircraft, several parameters can be tuned by the operator to obtain the desired ground sampling [17]. Along path point distance depends on scan frequency and aircraft speed. Across path point distance depends on flying altitude, scan angle, scan and pulse frequency. Footprint size depends on flying altitude and beam divergence. Final echo density also depends on swath overlap. Those settings result in differences in the interaction between laser pulses and the vegetation or ground. Several studies have tried to qualify or quantify the way in which acquisition parameters influence the final laser data and its potential for forest attributes extraction.

Due to the high cost of LiDAR data, many studies proposed to rely on simulations to optimize the acquisition parameters [94] or quantify the effect of a given parameter. Many works investigated platform altitude [93, 125, 212, 230], but also the question of footprint [13, 110, 125], scan angle [120, 212], slope [29] or point density [110, 185, 300]. The influence of acquisition parameters is enhanced in mountainous areas where topographic conditions make flight planning more complicated and pulse interaction with the vegetation more variable, mainly because of slope [110, 246, 282]. Season of acquisition also has a major effect as canopy interaction with pulses changes between leaf-on and leaf-off conditions [238, 242], even for coniferous stands.

Obviously, point density is one of the main factor regarding LiDAR potential for forest attributes extraction. It directly affects the quality of the retrieved geometric information. Regarding small footprint LiDAR, it is generally considered than area-based methods yield satisfying results with point density around one point per square meter, whereas densities superior to ten are required for single tree approaches. It is preferable to restrain scan angle, as in non-nadir views the trajectory of a laser pulse is not vertical which leads to increased occlusion effects.

4.1.2 Processing

Once the data is acquired, several processing steps are required (figure 4.1). First, aircraft trajectory and attitude during the acquisition are computed from the ground GPS reference and the on-board GPS and IMU measurements. With full waveform data, echoes have to be extracted from the raw files. Sensor coordinates of the echoes are then converted to georeferenced coordinates with the previously calculated trajectory and attitude data.

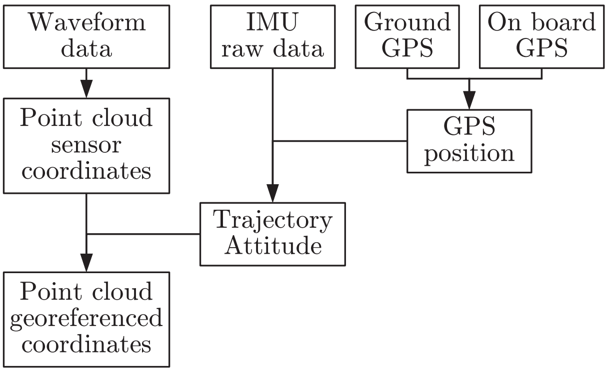


Figure 4.1: From raw acquisition data to georeferenced point cloud: processing steps.

Point clouds obtained from different flight strips are adjusted by matching overlapping areas [172] and with the control surfaces acquired by GPS on the ground. As there is no guarantee that LiDAR last echoes are reflected by the ground, terrain surface must be estimated before a final classification of echoes into ground and non-ground points is edited. This task has received considerable attention in the literature [6, 16, 19, 36, 39, 45, 72, 74, 81, 82, 138, 146, 155, 164, 202, 206, 260, 310, 311, 314]. This step is indeed crucial as altitude errors in the terrain surface will propagate in forest attributes estimates. Algorithms often require an initial parametrization and fail to handle areas where topography is highly variable [34]. A study pointed out the fact that DTM accuracy is scarcely available in the literature, and that especially in forested, rough terrain, evaluations of accuracy are likely to be rather optimistic [118]. Indeed, the low penetration rate of the laser and the high variability of the terrain surface result in a ground point density that is insufficient for correct terrain modelling.

4.1.3 Quality assessment

Quality assessment of the point cloud accuracy is generally performed by the data provider thanks to control areas positioned on flat, homogeneous surfaces such as parking lots or roofs. To evaluate the accuracy of the point cloud in forested areas with the same method, flat surfaces are installed under the vegetation cover and georeferenced [213]. However, such approaches are useful for point cloud adjustment, but do not really provide global evaluation of the DTM accuracy. Indeed, it is influenced by both the original point cloud density and the classification step. Due the limitations of GPS acquisition under forest cover, a comprehensive assessment is difficult. One alternative is to compare DTM elevation to the GPS data acquired in non forested areas, e.g. along the road network [118]. The same study reported that for gentle slopes ($< 20\%$) the root mean square error was below 20 cm, but increased for steep terrains. Similar values were obtained with GPS data acquired directly inside forest plots [105], with standard deviations around 0.3 m between reference and LiDAR data.

4.2 Quality requirement from the end-user perspective

Data quality, especially for full waveform sensors, is highly dependent on the processing steps that are generally performed by the contractor (echo extraction, geolocation, adjustment, classification and DTM computation). In mountainous areas, even adaptive filters for echo classification reach their limits due to topographic complexity, and manual editing is too time-consuming to be performed globally. Users must be aware of the precision level of the obtained data, and evaluate if it matches their requirements. Indeed, for forestry applications, an altimetric precision of 0.5 m is still an improvement compared to field height measures whose precision is of metric order.

The research presented in this document focuses on the extraction of forest parameters from the georeferenced and classified point cloud. Comprehensive evaluation of LiDAR data quality is an interesting but full-time task. Therefore, the data provider was completely in charge of data processing and quality assessment, except for a qualitative visual check of data located on the reference field plots.

4.3 Airborne laser scanning datasets

4.3.1 Location

This work benefited from numerous LiDAR datasets acquired in the framework of other research projects:

- PROVIALP, Protection de la viabilité alpine, projet n°165 Interreg III A ALCOTRA (2005-2007) [4].
- IFP, Interreg Forêt de Protection [2].
- MANFRED, Interreg Alpine Space [1].
- CARRESS, Apport de la scannérisation aéroportée pour la caractérisation et la localisation de la ressource forestière, Cluster de recherche ENERGIES de la région Rhône-Alpes.

All areas are located in the Rhône-Alpes region, France (figure 4.2). The provider was always Sintegra Photo (Meylan, France). Table 4.1 presents the LiDAR datasets.

Table 4.1: Areas covered by LiDAR acquisitions used in this work. Code for funding project. 1: PROVIALP, 2: IFP, 3: MANFRED, 4: CARRESS

Zone	Id	Acquisition date	Area (km^2)	Funding project
Murianette	Mur	2007	5.2	1
Vaujany	Vau	2007	0.7	1
Chamonix	Chm	2008	98.0	2,4
Chablais	Chb	Aug. 24-26, 2009	100.0	2,4
Saint-Paul-de-Varces	StP	Aug. 27, 2009	8.6	2,4
Valdrôme	Val	Sept. 20, 2010	6.2	3
Saint-Agnan-en-Vercors	StA	Sept. 21-22, 2010	38.7	3

4.3.2 Acquisition parameters

All acquisitions were performed using a RIEGL LMS-Q560 sensor, except for a sub area of the Chamonix acquisition, where a Leica ALS50-II was used for the high altitude areas. The LMS-Q560 was one of the first small footprint commercial full waveform scanner. The ALS50-II is one the first multipulse laser scanner. Acquisition parameters are summarized in table 4.2. Last acquisitions benefited from the experience gained during the first campaigns, in terms of both sensor calibration and processing procedures. This resulted in a better accuracy and homogeneity of the point clouds.

4.3.3 Processing procedure

The workflow implemented by the provider consists in the following steps:

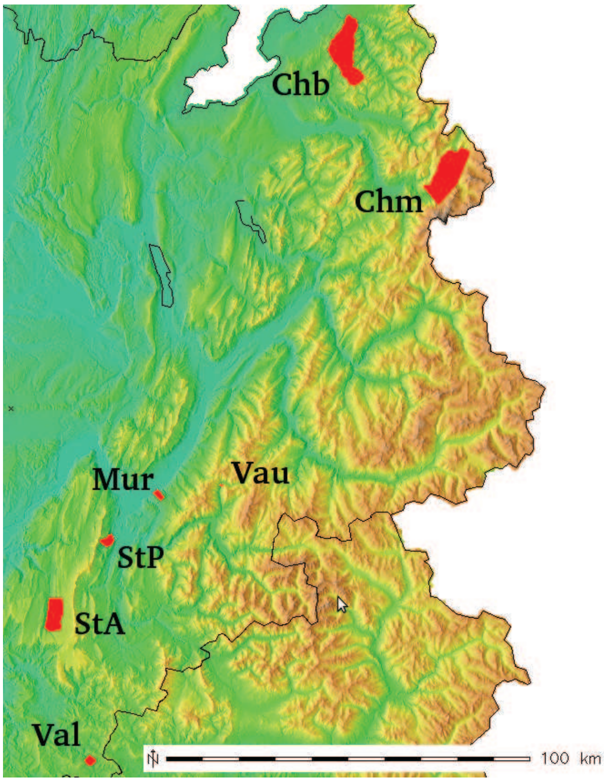


Figure 4.2: Location of airborne laser scanning surveys.

- computation of the aircraft path from the GPS and IMU data;
- echo extraction from the full waveform files with RiAnalyze;
- point cloud computation with RiWorld;
- flight strips adjustment with TerraSolid;
- point cloud automatic and manual classification with TerraSolid.

The procedure had to be adapted to the constraints of mountain environments. For example, flight paths are oriented along contour lines so that flight height remains constant. Flight height is also limited by the need to approximately follow contour lines that are quite sinuous in complex terrain. To increase swath width at lower heights, and decrease the total number of flight strips required for mapping of the whole area, scan angle is increased. However when terrain slope is of the same magnitude that scan half-angle, laser pulses shot downslope range higher distances before hitting ground, which leads to decreased position accuracy. Such points are excluded from

Table 4.2: LiDAR acquisition parameters for the covered areas. Scanner model code. 1: RIEGL LMS-Q560, 2: LEICA ALS50-II.

Area ID	Mur	Vau	Chm		StP	Chb	Val	StA
Scanner	1	1	1	2	1	1	1	1
Pulse frequency (kHz)	200	200	200	85.6	120	120	170	170
Scan frequency (Hz)	111.1	73.3	104.4	26.6	77.3	77.3	111.4	111.4
Half scan angle (°)	22.5	30	30	20	30	30	29.5	29.5
Flight height (m)	550	500	550	2800	600	600	595	595
Footprint size (m)	0.3	0.3	0.3	0.6	0.3	0.3	0.3	0.3
Point spacing (m)	0.5	0.3	0.5	1	0.6	0.6	0.5	0.5

the final datasets. When near vertical or overhanging cliffs are present, the point cloud is tilted so that the 2D classification algorithm used in TerraScan [16] performs efficiently, and then tilted back.

LiDAR data are delivered in ASPRS LAS format. The binary files contain the following attributes for each point:

- coordinates (easting, northing, altitude or ellipsoidal height) in a given projection system, usually Lambert 2 étendu or Lambert 93,
- intensity,
- position of the echo within the returns of the pulse (single, first of many, intermediate, last of many) coded on two fields[1],
- classification code,
- point source ID,
- GPS time[1].

---

[1] When the data is processed within TerraScan binary files, the original information about the absolute position of the echo and the total number of returns for a given pulse is lost. Similarly, the GPS time is truncated which makes it impossible to distinguish between successive pulses. Using LAS files in TerraScan allows to retain the whole original information, which has been done for the Valdrôme and Saint-Agnan datasets.

Chapter 5

Field data

5.1 Introduction

5.1.1 A changing perspective

Diameter at breast height (DBH) is probably the easiest measurable dimension of a tree, and therefore is used as a common and convenient proxy for other tree metrics [55]. Basal area (planimetric area occupied by the horizontal section of one or several tree trunks at 1.30 m height), is commonly used by silviculture practitioners to quantify the space occupied by trees in a given stand, and their productivity. For a more accurate estimation of tree volume, or to estimate site quality, tree height measures are required. Due to the difficulties linked with tree height measures on the field, numerous height-diameter models have been proposed, and calibrated for various forest conditions [293].

With airborne LiDAR acquisitions, tree heights are directly measured. The point of view of the forester changes from a diameter-based to a height-based dendrometry. Some models originally developed to predict height from diameter cannot be used in reversed way. This implies the development and calibration of new models, with the LiDAR estimated height, but also other LiDAR-extracted features, as variables [268].

5.1.2 An unexpected accuracy

LiDAR points planimetric and height accuracies are around 25 and 15 cm respectively. Such values are close to or smaller than accuracy values for field measures of tree heights or global positioning system accuracy for positions acquired below forest cover [148, 220]. The availability of such detailed information makes precision forestry at the tree or group of trees levels possible. However, a point is reached where the tool to be evaluated has better precision than most of the currently available reference tools.

5.1.3 Coregistration of specific field data

Even tough the geometric information is directly measured by airborne LiDAR, calibration of models is still needed to refine height estimation or to predict other forest parameters. This requires the availability of coregistered LiDAR and ground data, at the tree or stand level. Unfortunately, field data that are routinely acquired, e.g. in the framework of French National Forest Inventory (NFI), are not always directly usable. First, their geolocation may be missing or too approximate to correctly match the LiDAR data, especially for single tree delineation purposes. Second, many inventories are based on angle-count samples instead of fixed-radius plots, which makes true correspondence of inventoried areas with LiDAR features impossible when trees are not mapped. However, some studies showed that such data could be used with satisfactorily results [117, 192]. Third, the number of plots is frequently too small within a given ALS surveyed area, which makes the modelling and validation of algorithms difficult. For example, the Chamonix dataset covers 100 km^2. 20 km^2 are forest area, which includes only six plots from the French National Forest Inventory. They are fixed-radius plots with mapped trees, unfortunately their coordinates are available with only 450 m accuracy, due to confidentiality reasons. Therefore, testing the LiDAR data potential often implies the acquisition of specific field data.

As a result, investigations on the operational use of LiDAR must take into account field data requirements and constraints, and even question current inventory procedures in order to reach a better synergy between remote-sensed and field data.

5.2 Field data collection in alpine environments

5.2.1 Geolocation

5.2.1.1 GPS acquisition in forested mountains

Precise coregistration of LiDAR and field data is a pre-requisite for successful exploitation of LiDAR data. Unfortunately, the GPS signal acquisition in mountainous areas is quite erratic due to the topographic mask. The number of visible satellites is reduced, and the probability is high that they are aligned or grouped which results in a high positional dilution of precision (PDOP). The acquisition periods with satisfactory accuracy are limited, especially in deep valleys, such as Chamonix. The problem is enhanced for plots located on the ubac, as more satellites are usually present in the southern part of the hemisphere, and for plots with dense canopy cover [148]. To avoid bad precision under dense vegetation cover, one alternative is to acquire the GPS position in a nearby cleared area and to locate the plot relatively to the GPS point [220]. This is possible only if the relative positioning is itself accurate, i.e. points are located not too far apart and positioning tools are adequate.

Another solution is to do without accurate GPS location and to position the point relatively to some man-made or natural feature visible in the LiDAR data, such as remarkable rocks, buildings, or previously installed targets [213]. However, the adjustment is then based on manual evaluation and might be time-consuming. When trees have been mapped and their height measured, it is also possible to directly match tree locations to features detected in the LiDAR data [60, 61].

For the registration of our field data, the solution of relative positioning was excluded, based on the following grounds.

- In complex environments, one cannot rely on the presence of a cleared area for better GPS acquisition.
- One cannot rely on the existence of remarkable elements, as they might not be so easily identified in the LiDAR data.
- Installation of targets is expensive and labour-intensive.

As one objective of this study is to evaluate single tree detection potential, ALS is not used to geolocate the reference data (field tree maps). GPS signal is acquired directly on the field plots by taking advantage of DTM availability to calculate of the elevation mask (see appendix B). This information is then fed to GPS software to forecast the propitious acquisition periods.

5.2.1.2 Protocol for GPS acquisition

However, the elevation mask does not account for the vegetation mask that also lowers the number of visible satellites and positioning accuracy, and one must be aware that under forest cover real position standard deviation is higher than announced by GPS post-processing softwares. The study [220] reported mean position errors of 1.86 and 3.30 m for five minutes observation periods in forest stands with basal area below 30 and above 45 $m^2.ha^{-1}$ respectively.

First trials with the available GPS equipment (Trimble Pro XRS receiver) revealed that carrier phase processing is not possible in forested areas due to the impossibility of continuous signal acquisition. Only pseudo range information is used, and the usual procedure is to acquire 300 positions on a one position per second basis, which ideally corresponds to five minutes of continuous acquisition. Due to irregular reception, total acquisition time is frequently longer. GPS data is then differentially post-processed using data from a fixed reference station located in Valence (Drôme) or Mogneneins (Ain). The position error ϵ_{post} linked with distance to the fixed station D is expected to be $\epsilon_{post} = 0.5 + \frac{D}{10^5}$ (half a meter plus 0.1 m for every additional 100 km of distance), which amounts to a maximum of 70 cm for plots that are located the further away. Given this procedure and previous findings [148, 220], final position error is expected to be around 1.5 to 5 m, depending on forest structure and local topography.

5.2.2 Mapping tools

The mapping equipment is also adapted to field work conditions. Due to the limited road access network, equipment is sometimes carried for several hundred meters of climb, including off tracks. Therefore, lightweight equipment is used. For tree mapping a clinometer and a compass mounted on a tripod are used (figure 5.1). For distance measures hypsometers are used (Vertex III and IV), as they are more convenient than laser rangefinders when understory and bushes are present. Hypsometers range is up to 40 meters in mature stands without understory, but are limited to a few meters when bushes and small trees are dense. Given the instruments specifications, relative position accuracy of trees is expected to be lower than 1 m for trees up to 40 m from the reference point.

(a) Overview (b) Clinometer and compass

Figure 5.1: Tripod with clinometer and compass used for tree mapping.

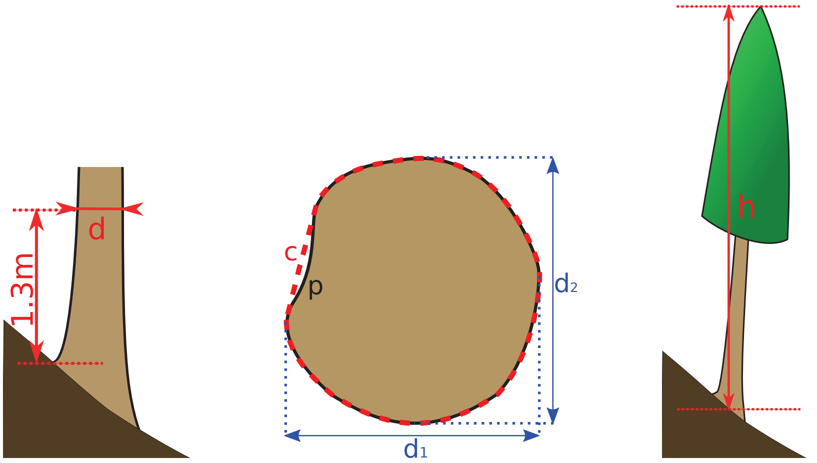
5.2.3 Forest metrics

This section introduces the definitions of the main tree and stand parameters that are considered in the document, as well as how they are estimated from field measures.

5.2.3.1 Tree parameters

Diameter at breast height DBH is the diameter of a tree at 1.3 m vertical height from the ground. In areas with slope, 1.3 m distance is measured upslope of the tree (figure 5.2a). As trees are not actually perfect circles, in practice DBH is commonly estimated in two ways depending on the used tool (figure 5.2b). With a caliper, it is estimated as the mean of two measures of tree diameter taken at orthogonal directions $\hat{d} = \frac{d_1+d_2}{2}$. With a ruban tape, it is estimated as the diameter of the circle that has the same circumference c: $\hat{d} = \frac{c}{\pi}$. Diameter at breast height d is the easiest measurable parameter for a tree, and is usually expressed in centimeters.

Total height Height is defined as the vertical distance between the ground level at tree stem and the highest bud of the tree (figure 5.2c). Contrary to DBH, measuring tree heights turns out to be tricky, principally because of the difficulty to view the terminal bud of the tree from a ground position. The literature also reports a tendency of LiDAR measures to underestimate tree height [88].



(a) Diameter at breast height

(b) Measures for diameter estimation

(c) Tree height

Figure 5.2: Measure of the diameter and height of a tree.

Basal area The basal area g of a tree is the area of the horizontal section of the tree stem at 1.3 m height. In practice it is estimated from the DBH d or directly from the circumference c. It is usually expressed in square meters (d and c in centimeters).

$$g = \frac{\pi d^2}{4.10^4} \qquad \text{or} \qquad g = \frac{c^2}{4\pi.10^4} \tag{5.1}$$

Stem volume The stem volume v of a tree depends on its diameter and height, but also on its shape. The volume of a reference tree in a given stand is usually estimated from diameter or height with specific allometric relationships. Here, the stem volume is computed according to the Alpes du Nord formulas [70] for spruce (equation (5.2)) and fir (equation (5.3)). The Algan formula [9] with a factor $f = 12$ is used for larch and with $f = 8$ for all other species (equation (5.4)). Those are formulas commonly used by forest practitioners in the area. The Alpes du Nord formulas have two inputs (height h in meters and diameter d in centimeters), whereas the Algan formula only relies on diameter. v is expressed in cubic meters.

$$v = 0.0111923 - 8.64751 \times \left(\frac{d}{100}\right)^3 + 0.476469 \times \left(\frac{d}{100}\right)^2 \times h \tag{5.2}$$

$$v = -0.0126441 - 6.98147 \times \left(\frac{d}{100}\right)^4 + 0.44908 \times \left(\frac{d}{100}\right)^2 \times h \tag{5.3}$$

$$v = \frac{8+f}{28} \times \left[0.312384 - 4.51338 \times \left(\frac{d}{100}\right) + 23.8866 \times \left(\frac{d}{100}\right)^2 - 2.89769 \times \left(\frac{d}{100}\right)^3\right] \tag{5.4}$$

It is noteworthy that stem volume estimated for a single tree is not accurate, due to the high variability of tree shape. Provided that the chosen allometric relationship is adapted to the studied forest stand, the estimation of the global volume located in the stand will be trustworthy.

5.2.3.2 Stand or forest parameters

Consider a forest area with horizontal surface S (ha). Stand parameters are estimated from the individual parameters of included trees. Those parameters are useful for forest management operations (silviculture, harvesting) at different levels. Depending on the objectives and on the scale of analysis, parameters are estimated for a stand or compartment (smallest management unit) and up to the whole forest. One important value for deriving stand attributes from tree parameters is the minimum size for a tree to be considered. It is generally based on DBH. For inventory purposes, trees with DBH above 7.5 cm are usually considered. Larger values are used for harvest planning (typically trees with an economical value, i.e. with DBH above 17.5 cm). For specific purposes, such as the inventory of stands in the initiation of reinitiation stages, smaller values may be used. In this study, a 7.5 cm threshold is used, except for the Saint-Paul inventory (5 cm). For the next definitions, it is assumed that the chosen area includes N_T trees which fulfill the DBH requirement.

Dominant height is defined as the mean height of the 100 trees with the largest diameters per hectare. For a forest area with surface S where N_T trees have been inventoried and their height and diameter $(h_i, d_i)_{i\in\{1,\dots,N\}}$ measured, if trees are sorted according to decreasing order of diameters, i.e. $\forall i \in \{1, \dots, N_T - 1\}, d_i \geq d_{i+1}$, dominant height H is estimated as:

$$H = \frac{1}{N_S}\sum_{i=1}^{N_S} h_i \qquad \text{with } N_S = max\left(1, \lfloor 100.S \rfloor\right) \tag{5.5}$$

Stem density is the average number of trees per hectare:

$$N = \frac{N_T}{S} \tag{5.6}$$

Basal area, mean diameter, volume Basal area G (m^2.ha^{-1}), mean diameter D (cm) and stem volume V (m^3.ha^{-1}) for a given forest area are calculated from their counterparts of included trees $(g_i, d_i$ and $v_i)$ as:

$$G = \frac{1}{S}\sum_{i=1}^{N_T} g_i \qquad D = \frac{1}{N_T}\sum_{i=1}^{N_T} d_i \qquad V = \frac{1}{S}\sum_{i=1}^{N_T} v_i \tag{5.7}$$

5.3 Datasets

5.3.1 Introduction

In this work two types of field data are used. Large stand plots are usually square plots larger than 0.25 ha. Plots of 50×50 m^2 have been considered suitable for the description of "elementary stands" in mountainous forests [203, p. 104]. All trees are mapped and measured (height, diameter and sometimes crown extension) so that the data can be used for simulations of stand growth [50], silviculture experiments [203], real-size rockfall experiments [62] or training purposes (marteloscopes).

Forest inventories consists in several small sample plots used as statistical descriptors of an entire forest. Trees on each plot may be mapped but usually the interest variables are the stand variables. Table 5.1 details the names and color codes used for encountered species.

5.3.2 Large stand plots

The general protocol for the full inventory of a selected stand consists in the following steps.

- Choice of plot location according to forest structure related to study objectives.
- Field delineation of plot boundaries. Square plots are usually aligned on contour lines, and slope is taken into account. Plots corners, and also intermediate points are materialized by markers. Those reference points are positioned iteratively in distance and azimuth. Field delineation is qualitatively checked when closing the perimeter by measuring the distance between the departure marker and its theoretical position estimated from the last marker.
- Tree measures and mapping. For each tree with diameter above the DBH threshold, DBH and total height are measured. Tree positions are recorded in azimuth and distance to a neighboring plot corner or intermediate marker. Additional reference points may be added when all trees are not visible from the previously materialized points. In the absence of bushes or low strata, all trees may be visible from a single point located near the center of the plot.
- Geolocation of reference points using a GPS receiver.
- Back in the office, GPS positions are differentially corrected using the data from the nearest permanent ground station (Valence or Mogneneins), and projected in the local coordinate system. Depending on the position accuracy estimated by the Pathfinder software, reference points positions are determined from the GPS data or relatively to some neighboring point with a more accurate position. Tree coordinates in the projection system are then computed relatively to their reference point with their polar coordinates and taking the grid declination (magnetic declination and convergence angle) into account.

The following information about each tree are also recorded.

- Species.
- Appearance: normal, broken stem, snag with branches, snag without branches.
- Other qualitative information, e.g. tilted or horizontal tree, presence of forks.

In non-intensively managed forest, non merchantable trees such as dead, forked or curved trees can amount to a significant part of the total woody biomass. Their presence may decrease the accuracy of estimates of commercial wood, or perturb single tree detection.

Information about large plots used in the study is displayed in table 5.2. Tree position maps, digitial terrain models (DTM), canopy height models (CHM) and height histograms of inventoried trees for each plot are gathered in appendix C.

Table 5.1: Color code, abbreviation and full name of tree species encountered in the study areas. When only the Genus could be determined on the field, trees are affected to a category called *Genus sp.*

Color	Abbrv.	Genus species	English name	French name
■	ABAL	*Abies alba*	European silver fir	Sapin blanc, sapin pectiné
■	ALVI	*Alnus viridis*	Green alder	Aulne vert
-	ACOP	*Acer opalus*	Italian maple	Érable à feuilles d'obier, érable d'Italie
■	ACPL	*Acer platanoides*	Norway maple	Érable plane
■	ACPS	*Acer pseudoplatanus*	Sycamore maple	Érable sycomore
■	ACsp	*Acer sp.*	Maple	Érable
■	BEPE	*Betula pendula*	Silver birch	Bouleau verruqueux
■	BEsp	*Betula sp.*	Birch	Bouleau
-	BUSE	*Buxus sempervirens*	Common box, European Box	Buis
-	CASA	*Castanea sativa*	Sweet chestnut	Châtaignier commun
■	COAV	*Corylus avellana*	Common hazel	Noisetier commun
■	FASY	*Fagus sylvatica*	European beech, common beech	Hêtre commun
■	FREX	*Fraxinus excelsior*	European ash, common ash	Frêne commun, frêne élevé
■	ILAQ	*Ilex aquifolium*	European holly	Houx commun
■	LADE	*Larix decidua*	European larch	Mélèze d'Europe, mélèze commun
■	PIAB	*Picea abies*	Norway spruce, European spruce	Épicéa commun
■	PRAV	*Prunus avium*	Wild cherry	Merisier
■	PINI	*Pinus nigra*	European black pine	Pin noir d'Autriche
■	QURO	*Quercus robur*	Pedunculate oak, English oak	Chêne pédonculé
-	QUPU	*Quercus pubescens*	Downy oak, pubescent oak	Chêne pubescent
■	SACA	*Salix caprea*	Goat willow	Saule marsault
■	SOAR	*Sorbus aria*	Common whitebeam	Alisier blanc
■	SOAU	*Sorbus aucuparia*	Rowan, European rowan, mountain ash	Sorbier des oiseleurs
■	TABA	*Taxus baccata*	European yew	If commun
■	TIsp	*Tilia sp.*	Lime, linden	Tilleul
■	ULGL	*Ulmus glabra*	Wych elm, Scots elm	Orme blanc, orme des montagnes
■	ULsp	*Ulmus sp.*	Elm	Orme

Table 5.2: Description of large plots situation (exposition, slope, area, altitude) and forest characteristics (dominant height H, basal area G, stem density N and species proportion). Only live trees with diameter at breast height above 7.5 cm are considered. For species proportion, only the two main species and species which represent more than 5% of stems are indicated.

Area	Id	Exp. (grad)	Slope (degree)	Altitude (m)	Area (m^2)	Inventory date	H (m)	G ($m^2.ha^{-1}$)	N (ha^{-1})	Coniferous (and species) proportions (% of stems)
Chb	chablais1		< 10	1304	50×50	07/2010	18.6	24.3	612	46 (PIAB 42, SOAU 28, FASY 23, ABAL 5)
	chablais2		< 10	795	50×50	07/2010	31.4	51.8	700	98 (PIAB 98, FREX 2)
	chablais3	311	19	1312	50×50	07/2010	23.9	23.7	408	47 (FASY 42, PIAB 25, ABAL 20)
	chablais4	362	13	1334	50×50	07/2010	35.1	44.5	328	78 (PIAB 54, ABAL 24, FASY 21)
	chablais5	336	28	1026	50×50	03/2010	31.6	61.6	580	86 (ABAL 86, FASY 14)
	chablais6		< 10	1156	50×50	03/2010	16.6	25.1	740	92 (PIAB 92, PRAV 4)
	chablais7	324	23	904	50×50	03/2010	21.3	14.9	556	17 (FASY 74, PIAB 16)
	chablais8	82	30	1052	50×50	03/2010	27.0	28.0	436	17 (FASY 61, PIAB 12, FREX 11)
Mur	combeloup	369	33	954	50×50	07/2010	27.8	30.1	460	5 (FASY 90, PIAB 5)
Chm	corruaz	169	38	1237	50×50	07/2010	27.8	41.0	744	84 (PIAB 84, FASY 4, SACA 4)
	moussoux	134	37	1244	50×50	07/2010	27.5	42.5	1124	99 (PIAB 98, LADE 1)
	orthaz3	358	30	1198	50×50	09/2010	30.1	39.4	592	99 (PIAB 92, ABAL 4)
StA	stagnan1	316	32	1241	62.5×40	07/2010	26.4	32.8	312	59 (ABAL 55, FASY 41)
	stagnan2	309	30	1232	62.5×40	07/2010	25.8	38.5	376	67 (ABAL 64, FASY 33)
	stagnan3	318	33	1269	62.5×40	07/2010	23.5	29.9	376	65 (ABAL 56, FASY 35, PIAB 5)
	stagnan4	308	31	1260	62.5×40	07/2010	23.7	28.9	372	42 (FASY 51, ABAL 38)
Val	valdrome1	284	34	793	50×50	04/2010	22.4	47.0	792	100 (PINI 100)
	valdrome2	278	33	799	50×50	04/2010	22.8	36.0	596	100 (PINI 100)
	valdrome3	268	34	811	50×50	04/2010	20.8	46.4	1060	100 (PINI 100)
	valdrome4	279	36	830	50×50	04/2010	21.2	52.3	1600	100 (PINI 100)
Vau	vaujanyba	316	34	1251	3400	05/2009	28.0	46.0	271	71 (ABAL 57, FASY 16, PIAB 14, ACPS 9)
	vaujanyha	312	35	1304	3300	05/2009	31.1	54.0	291	57 (ABAL 55, FASY 36, ACPS 7)

5.3.2.1 Murianette

The Combeloup plot is located in the Belledonne mountain, in the town of Murianette. This 50×50 m^2 area is located at 950 m above sea level, on a regular hillside oriented toward north-west (figure C.1a). Trees with DBH above 7.5 cm were mapped in 2006 (figure C.1b). Height and diameter measures were updated on Jan. 5th, 2011 using a VERTEX iiihypsometer and a ruban tape. The stand has an irregular structure and is largely dominated by European beech, with 90% of the 115 inventoried live trees (figure C.1c). Some large Norway spruce tree are also present (5% of stems). Remaining stems are other deciduous trees or small silver firs.

5.3.2.2 Chamonix

Three 50×50 m^2 plots are installed in the Chamonix valley.

Orthaz3 plot is an irregular stand dominated by Norway spruce and located on a regular hillside facing north-west, at 1200 m above sea level (figure C.2g). Its structure type according to the Northern Alps typology [263] is E1: dense irregular stand with small trees. It was originally established to test new silvicultural strategies designed to improve mountainous stands stability and resilience [203], and was first inventoried in 1994 (figure C.2h). The area was divided into squares with markers positioned every 12.5 m with a tape measure, a clinometer and a compass. Trees with DBH larger than 7.5 cm had their positions to the nearest marker recorded with the same instruments. Plot corners were georeferenced using a Trimble GPS Pro XRS receiver. On June, 2nd 2009, the plot inventory was updated. DBH and tree heights were measured with a tape and a Vertex III hypsometer. 148 live trees were inventoried, with a majority of Norway spruce (92% of stems). Some silver firs (4%) are present, as well as large European larch trees and a few broadleaved trees (figure C.2i).

Corruaz plot is also a stand dominated by Norway spruce, but it is located on a hillside facing south-east, at 1240 m above sea level (figure C.2a), and small trees represent a larger proportion than at Orthaz3. Hence its structure type is D3: coexistence of an upper stratum of old-grown trees with an abundant lower stratum, and absence of intermediate trees. It was first inventoried in the early 1990's. In July, 2010, the whole plot was inventoried anew, according to the protocol described in section 5.3.2 (figure C.2b). Norway spruce represents 84% of the 186 live trees (figure C.2c). Many deciduous species are present in the lower and medium strata (16% of stems).

Moussoux plot is located in an area similar to Corruaz. It is a very dense Norway spruce plantation (98% of stems) that now has an irregular structure. A few large European larch trees (1% of stems) dominate the stand. It was first installed in 2001 and was inventoried anew in 2010 (figures C.2d, C.2e and C.2f).

5.3.2.3 Vaujany

The Vaujany plot is a study area initially used for real-size rockfall experiments [62], located in the Belledonne moutain (45°12'07"N, 06°03'00"E, altitude 1275 m) on a 35° slope facing west (figure C.3a). It is a mixed irregular stand with two strata: large silver fir trees and a deciduous understory. In 2001 twelve markers were positioned in the hillside. They were georeferenced using a Trimble GPS Pro XRS receiver. It is noteworthy that plot boundaries were not previously delimited, which resulted in an irregular shape (figure C.3c). Trees with a DBH larger than 7.5 cm had their positions to the nearest marker recorded with a clinometer and a compass mounted on a tripod and with a laser rangefinder or a decameter. On May, 12th 2009, the plot inventory was updated. DBH and tree heights were measured with a tape and a Vertex III hypsometer.

For the purposes of this study, the plot is divided into two subplots. In the upper part of the study area (vaujanyha, 0.33 ha), 96 live trees were measured. Silver fir and European beech are dominant (respectively 55 and 35% of stems). Sycamore maples are also encountered (7%), as well as Norway spruce and European holly (figure C.3d). 92 live trees were inventoried in the lower part (vaujanyba, 0.34 ha), with a majority of Silver fir (57%). European beech (16%) and Norway spruce (14%) are also present. The remaining stems are sycamores (9%) and other deciduous trees (figure C.3b). It is noteworthy that the rockfall experiments have cleared a great part of trees with DBH smaller than 15 cm.

5.3.2.4 Chablais

In the Chablais valley, seven 0.25 ha plots were established in March and July 2010 according the protocol presented in 5.3.2. Digital terrain models, crown height models and height distributions are displayed in figure C.4 for plots 1 to 4 and in figure C.5 for plots 5 to 8.

Plot chablais1 is located at 1304 m, at the top of a hill. It is a young, mixed stand of Norway spruce, European beech and European mountain ash.

Plot chablais3 is a medium aged, irregular stand with an upper stratum of Norway spruce and silver fir. Those two coniferous species also share the lower stratum with European beech and other deciduous species. It is located on a west facing hillside with gentle slope (19°), at 1310 m above sea level.

Plot chablais4 is situated a few hundred meters away from plot chablais3. Terrain slope is gentle and the area is crossed by a small intermittent stream near its north-east corner (see DTM on figure C.4j). It is a mature, two-layered stand with an upper stratum of Norway spruce, whereas the lower stratum is mainly constituted of silver fir and European beech.

Plot chablais5 is a very dense, regular stand of medium aged silver fir trees (see on figure C.5b the closed and homogeneous canopy). A few European beech trees are also present. Due to canopy closure and its location on a north-west facing slope of 28°, GPS acquisition turned out to be particularly problematic.

Plot chablais6 is a young Norway spruce plantation located on the relatively flat upper part of a hill, at 1150 m altitude. The stand has been thinned a few years ago (see on figure C.5e that the canopy has more gaps than plot chablais5).

Plot chablais7 was once dominated by large coniferous trees that have been felled since. The remaining lower and intermediate strata constituted by European beech and a few small Norway spruce trees now represent the majority of stems. A few large coniferous trees are still present. The plot is located at the footslope, in a rockfall deposit area (altitude 900 m, slope 23°). Terrain surface is very rough due to the presence of big boulders on the plot (see DTM on figure C.5g). Canopy cover is not very dense, and bushes and shrubs are abundant. Terrain classification may have been perturbed by those features so that it is likely that LiDAR canopy height on this plot is less accurate.

Plot chablais8 is located on an east facing hillside with 30° slope. It is constituted of deciduous trees, mainly European beech, overwhelmed by a few Norway spruce trees.

Plot chablais2 is a particular case since it is a subplot of a 100×100 m^2 plot installed in 2000 in the Brevon forest. It is located in the valley bottom (800 m above sea level), in an area subject to a slow landslide. Norway spruce was planted there from 1937 on, along with a few deciduous trees (alder, ash, poplar). The inventory was updated in July, 2010. The plantation is very dense, especially in the western part where trees are also smaller. It is noteworthy that in the central part, several trees probably weakened during the 2003 drought, died during a bark beetle invasion in 2005. Several snags are now present (× symbols in figure C.4f).

5.3.2.5 Valdrôme marteloscope

Valdrôme is a one hectare plot, with rectangular shape (50×200 m^2). It is situated across a west facing hillside, under a small cliff (figure C.6a). It is an example of European black pine stands planted during the afforestation program to mitigate erosion. The plot was installed following the protocol described in section 5.3.2 except that heights were measured only on trees located along three transects (western border, median of plot and eastern border). A total of 1026 trees were inventoried, and 362 heights were measured (figure C.6b). Heights (h, in meters) for the remaining trees are estimated from their diameter (d in cm) using the regression model calibrated on the sampled trees: $h = -9.4 + 8.66 \times log(d)$. adj-R^2 for the model is 0.73. For the purpose of this study, the marteloscope is divided in four subplots of 0.25 ha. All subplots exhibit a regular structure (figures C.6c, C.6d, C.6e and C.6f). The lower part of the hillside has better site quality: trees have grown larger and the stand has been thinned. In the upper part trees are smaller and stand is more dense.

5.3.2.6 Saint-Agnan-en-Vercors marteloscope

The Saint-Agnan marteloscope is located next to the Réserve Naturelle des Hauts-Plateaux du Vercors. It was inventoried in July, 2010 for training purposes. It is a protection forest situated at 1250 m above sea level, on a hillside oriented to the west, with 32° slope (figure C.7a). A forest road crosses the middle of the plot along the north-south direction. The forest has an irregular structure, and is mainly constituted of beech and silver fir, with a few spruce trees (figure C.7b). For the purpose of this study, the marteloscope is divided in four subplots of 0.25 ha.

5.3.3 Forest inventories

5.3.3.1 Saint-Paul-de-Varces

The studied forest is located on a 4 km^2 hillside oriented to south-east (town of Saint Paul de Varces, 45°04'17"N, 05°38'25"E)(figure 5.3). It is mainly constituted of coppice stands and deciduous stands on poor quality sites, dominated by Italian maples and downy oaks. Downslope, old chestnut coppice stands are frequent. Common whitebeam is present in all the area, especially at the foot of cliffs. In thalwegs or in upper parts with better site quality, ash and beech are common. Some areas have a dense understory of holly, common hazel or box. Altitude ranges from 330 to 1270 m above sea level. High limestone cliffs overhang the area and rockfall events are frequent. No major silvicultural or harvesting operations have been performed in the area for more than fifty years.

From September to November 2009, 31 circular field plots were inventoried. Plots were distributed every 400 m along the 550, 750, 950 and 1150 m height contours, resulting in an irregular sampling scheme where horizontal distances between neighboring plots ranged from 180 to 412 m with a mean value of 302 m. Plot centers were georeferenced using a Trimble GPS Pro XRS receiver. After differential correction with the Pathfinder® software, position precision (95% confidence radius) ranged from 0.6 to 1.5 m. All live trees with diameter at breast height larger than 5 cm and located within 10 m horizontal distance from the plot center had their diameter measured with a tape. Their positions to the plot center were recorded using a Suunto KB-14 compass and Suunto PM-5 clinometer mounted on a tripod, and a Vertex III hypsometer. Maples (mainly *Acer opalus*), downy oak and common whitebeam represented nearly 60 % of stems. Ten tree heights were also measured on each plot with the hypsometer. Height sampling probability was proportional to stem basal area to ensure that dominant trees would be represented. Forest parameters statistics are presented in table 5.3.

Table 5.3: Forest stand parameters statistics (31 field plots)

Parameter	Basal area G (m^2.ha^{-1})	Stem density N (ha^{-1})	Mean diameter D (cm)	Dominant height H (m)
Mean	34.8	1735	14.5	15.6
Min	4.6	764	8.3	5.8
Max	59.7	2833	22.7	25.8
Std deviation	11.4	577	3.6	5.5

As site quality is rather poor in the area, average dominant height is only 15.6 m. Generally, stands with high values for basal area, dominant height and mean diameter are located on a few good quality sites in thalwegs with deep soil, such as the ash-dominated plot 14 with $H = 25.8$ m, $G = 59.7$ m^2.ha^{-1} and $D = 21.6$ cm. Small values are encountered on steep slopes at the bottom of cliffs with rockfall activity. For example, plot 21 is located on a scree and has $H = 10.3$ m, $G = 4.6$ m^2.ha^{-1} and $D = 8.3$ cm. In such areas stem density is highly variable.

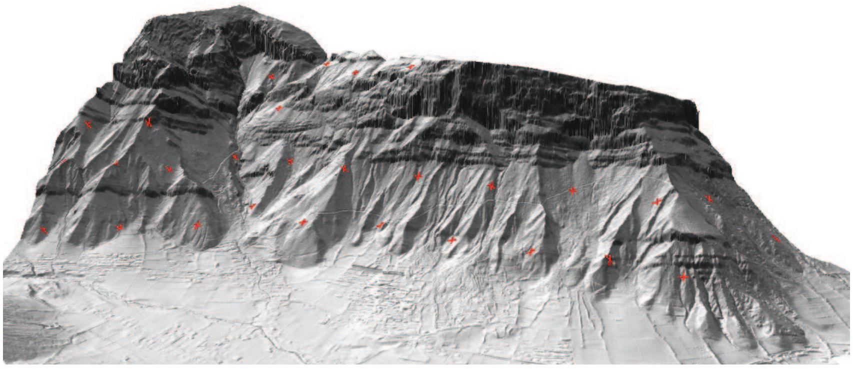
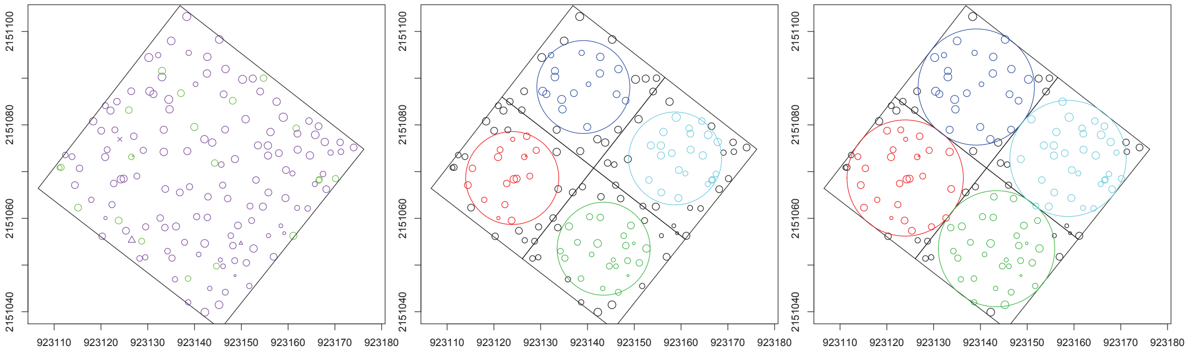


Figure 5.3: Shaded digital terrain model of the study area. +Red crosses are the locations of the sample plots.

5.3.3.2 Chablais

In the Chablais valley eight large plots (50 × 50 m^2) were installed, within an area covered by a single LiDAR flight. Even tough those plots are not representative of the forest stands of the whole area, they can be divided into several subplots that can be considered as sample plots for algorithm testing. The higher number of subplots that are extracted must not hide the fact that subplots from the same plot are likely to be similar because of spatial correlation.

The procedure for the division of plots is described hereafter. Consider a square plot of width W that have to be split in i^2 subplots. Subplot borders are created by dividing each plot borders in i segments of equal length. Corresponding nodes of the segments are connected to construct the subplot borders. If circular subplots are to be extracted, their center is positioned at the center of the quadrants. Trees belonging to the new circular plot are extracted within the chosen radius from the plot center. Figure 5.4 exemplifies the extraction of four circular plots from plot chablais5.



(a) Map of plot chablais5 (b) 10 m radius subplots (c) 12.5 m radius subplots

Figure 5.4: Extraction of circular plots from the 50 × 50 m^2 plot chablais5. (a) Plot map. (b) and (c) extracted circular subplots with radius 10 and 12.5 m respectively.

Part III

Finding the trees in the forest

Chapter 6

Introduction and methods

6.1 Single-tree approach

6.1.1 Overview

The capacity of airborne LiDAR sensors to accurately map the geometric shape of the Earth surface, and to characterize the vertical structure of non-solid objects such as trees, theoretically makes the extraction of forest information at the tree level possible. Indeed, the laser pulse is reflected by all elements constituting a tree, from leaves to the trunk, and laser footprint size (typically smaller than one meter) allows the laser beam to both reflect and penetrate inside the canopy. With a sufficiently dense LiDAR sampling, trees – with a size in the order of magnitude of a few meters – and even branches can be visually identified within a point cloud transect (figure 6.1).

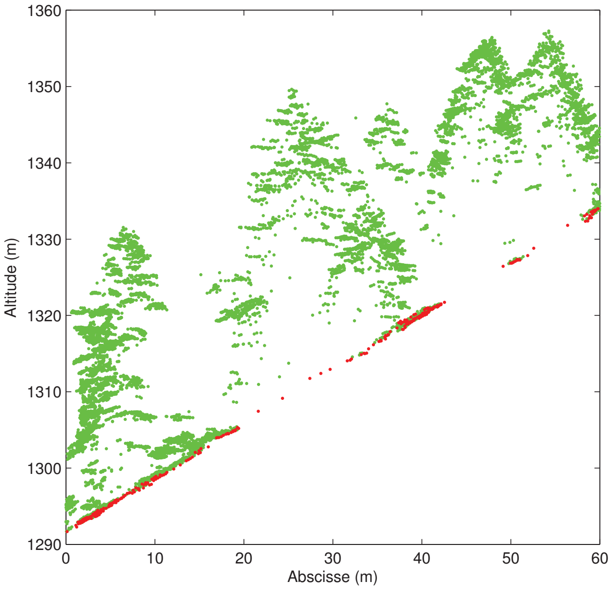


Figure 6.1: Point cloud transect along a forested hillside (Vaujany, France). Points classified as ground and vegetation are respectively in ● red and ● green. Transect width is four meters.

However, automatic tree delineation in the irregularly sampled, 3D point cloud is a complex task. Therefore, most studies on single tree delineation relied on preliminary transformation of the point cloud into raster images of the canopy height. The advantage of degrading the 3D information of the point cloud into a 2D image is that image processing techniques have been subject to intensive research with the development of digital storage and processing. They are now widely used in numerous domains such as medicine, biology, electronics, remote sensing of the Earth surface or detection of faults in manufactured goods... In forestry, they have already been used for the classification and segmentation of airborne or spaceborne optical images. With the availability of high spatial resolution data, trees could even be delineated [51,65,66,95,201,251].

Numerous approaches have been tested to identify trees in LiDAR-derived images of the canopy height. The basic idea is to detect features that are typical of trees. One of the simplest method is based on the assumption that treetops will induce local maxima in the image. Local maxima filtering is thus used for the detection of treetop candidates [137,254]. One issue is that branches may also induce local maxima, and in the case of heterogeneous stands, it is difficult to separate local maxima due to big branches from those due to small trees. This issue can be treated with field information about the relationship between the height of a tree and the size of its crown. It is then possible to adjust window size for local maximum detection to the value of the considered pixel [43,254]. Once treetops are identified, trees can be segmented with procedures derived from the watershed algorithm [43,104,115,137,156]. Indeed, limits between adjacent trees can be assimilated to watersheds in the opposite of the canopy height image. However, the algorithm makes no assumption about tree shape. To improve segmentation results, other criteria have to be taken into account, such as crown curvature or ellipsoid fit [137,156]. Assuming a round shape for tree crowns, an alternative or complementary tool to watershed segmentation is the distance transform [166].

Approaches that simultaneously handle the problems of size and shape of trees have been proposed, such as scale-space representation of the canopy height model and parabola fitting [252], spatial wavelet analysis [77] or multiscale isotropic matched filtering [245]. Marked point processes have been used on optical images [251].

The main drawback with such algorithms is that their detection performance highly depends on the forest structure [76], and parametrization is a crucial step, particularly with heterogeneous stands. To address this issue, strategies such as automatic assessment of localization and adjustment of the smoothing criterion [277], or multiple attributes decision making based on initial calibration [83] are possible.

Another limitation is that detection is performed on an image, which is a reduced version of the whole point cloud. In particular, suppressed trees are hardly detected in the canopy height model which is mainly defined by dominant or co-dominant trees. Several algorithms based on point cloud processing have thus been proposed to improve the detection rate, such as Bayesian object recognition [14], clustering [215] with several variants [171,262], voxel-based approaches [257,312] or region growing algorithm with a Delaunay triangulation [7]. However, the initialization step of point cloud algorithms is often based on local maxima detection in the canopy height image.

6.1.2 Objectives

As introduced beforehand, the advantages of using image processing techniques for single tree detection are that they are well documented, available in commercial softwares and also faster than point cloud processing algorithms, which turns out handy when millions of trees are extracted from LiDAR surveys of hundreds of square kilometers. However, their parametrization is not straightforward as it will affect detection performance differently depending on forest structure, and also on LiDAR acquisition parameters. Manual tuning is time-consuming and also site-specific, which makes it impossible for adaptive processing of large areas.

The objective in this part III is to design an automated procedure for single tree detection, based on the training of the detection algorithm on reference data. The approach based on local maxima is adopted as it is simple and fast. The idea is to build a training set of field plots for which the detection performance has been tested for several parameter combinations of the algorithm. When trees are to be detected in a new survey, the LiDAR data is compared to the training set and the algorithm parameters are set to the combination that yields the best detection performance on the training plot that is the most similar to the new plot. The whole workflow is implemented on 22 reference plots of mountain forest stands with various structures and species (section 5.3.2).

To reach this objective, several components have to be combined:

- a treetop detection algorithm,
- an automated procedure for detection performance assessment,
- a set of reference field plots,
- a similarity criterion for the matching of new LiDAR data to reference plots.

The treetop detection algorithm and the evaluation procedure are presented in this chapter. The results about the effect of algorithm parametrization on detection performance for the training plots are presented in chapter 7. Finally the automated detection procedure is detailed and discussed in chapter 8.

6.2 Local maxima extraction

6.2.1 Workflow and notation

The input data is the classified point cloud (easting, northing, altitude, classification). Each step of the sequential procedure is detailed in the following sections:

- calculation of the raster images from the point cloud,
- artifacts removal with a non-linear filter,
- lowpass filtering with Gaussian kernel,
- maxima extraction with variable window size,
- maxima selection.

The output data is a list of selected maxima with their coordinates and height. Unless specified otherwise, figures in the following steps display the example of plot chablais3, with 0.5 m spatial resolution images. The displayed area is a rectangle of 68×70 m^2, the corresponding image (or matrix) is of size 136×140 pixels.

The following notations are adopted in the whole chapter. An image is defined as a function $f(x, y)$, where x and y are the spatial coordinates and $f(x, y)$ the height or altitude of specified objects of the Earth surface (terrain, vegetation) located at this point. The whole image is composed of a finite number of pixels defined by their coordinates. Pixels are square and regularly aligned. S_{xy} is the notation for the neighborhood of the pixel located at (x, y). In this study, only square, centered neighborhoods are used. The figure axes are labelled in matrix coordinates, i.e. the matrix line index appears on the reversed vertical axis, and the row index appears on the horizontal axis. Pixels are colored according to their values, the correspondence is given by the scale bar to the right of the image. For density images, values are in points per square meter. For canopy height, terrain and surface models, the height or altitude values are in meters.

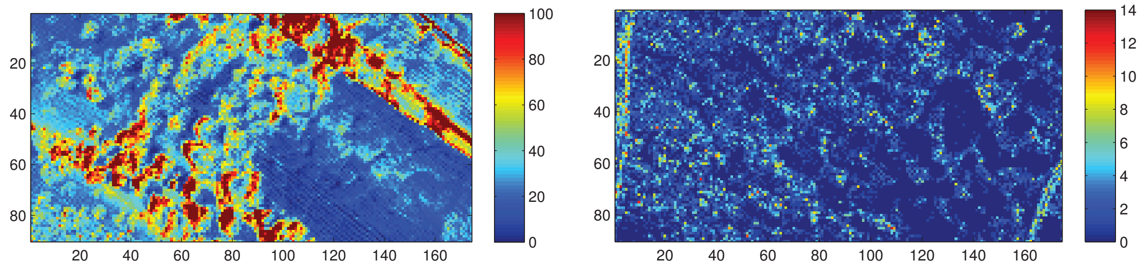
6.2.2 Image computation

6.2.2.1 Choice of image spatial resolution

The first step is to condense the information of the point cloud into regularly sampled images. The general approach is to compute one image of the terrain elevation, based on points classified as ground, and one of vegetation objects elevation with the other points. Choice of the spatial resolution is not straightforward as planimetric sampling is irregular and points are reflected at different heights in the canopy. If d_e is the mean echo density in the point cloud (points per square meter), and if we consider that each pulse will be reflected $r_p = 3$ times in average by forested surfaces (i.e. three echoes at different heights will be extracted from each returned waveform), a pixel size of approximately $res = \sqrt{\frac{d_e}{r_p}}$ seems appropriate for image spatial resolution. However, various factors affect local point density. The scanning pattern may induce more spaced pulses at the edge of the swath. Platform changes in attitude or speed also affect the distribution of points on the ground. Besides, higher densities are obtained where two acquisition strips overlap.

Figure 6.2a displays the point cloud density on the Vaujany plot. LiDAR acquisition was performed at slow speed with a helicopter. Moreover, several strips overlap in this area so that mean point density is very high (37.3 m^{-2}). This density is also highly variable, with a standard deviation of 30.9 m^{-2}. Indeed, several acquisition effects are noticeable. Only one strip covers the bottom right quarter which displays lower density despite being highly forested. On the upper left corner, densities are also small but it is partly due to the fact that trees are sparse and only one echo is recorded from one pulse sent on bare ground. Trees in this area are identifiable as patches of higher point density. The flight direction (south-west to north-east) can be guessed from the diagonal lines linked with the scanning pattern. Reduced speed of the aircraft, combined with attitude changes, may explain the presence of very dense bands, particularly one in the upper right quarter. Besides, crescent-like shapes might be due to the interaction of laser pulses with only one side of a tree when they are fired at angles far from nadir.

Concerning our final objective of detecting trees, the requirement regarding point density is that at a tree crown scale, i.e. several meters, there are enough points to characterize canopy shape. This should be possible when local point density is approximately a few points per square meter, which allows the calculation of images of submetric spatial resolution. The high differences in LiDAR acquisition hardware and forest structures, as well as the high variability of point density within small areas, are likely to make automatic choice of the resolution problematic. Therefore, the adopted strategy is to test single tree detection with a wide range of spatial resolutions and rely on spatial interpolation or on image processing techniques to tackle the issue of pixels that do not hold any LiDAR information.



(a) Whole point cloud density (m^{-2}) (b) Ground points density (m^{-2})

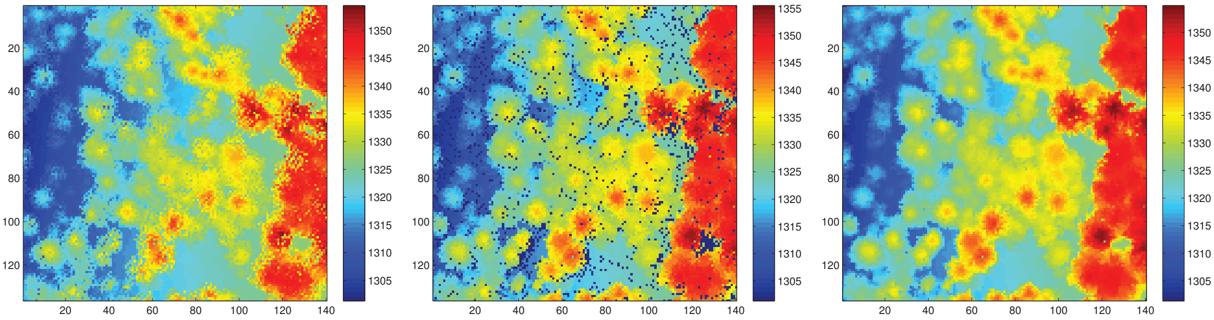
Figure 6.2: Image of point cloud density on the Vaujany plot (spatial resolution 1 m, values in points per square meter). One should read “100 and above” instead of “100” in the scale bar of subfigure (a).

6.2.2.2 Digital surface model

The digital surface model (DSM) is the image of the elevation of objects located at the surface of Earth (buildings, vegetation...). Computation of the DSM by bilinear interpolation of all vegetation points over a regular grid leads to a noisy effect in vegetated surfaces (figure 6.3a). Indeed, vegetation points located in the canopy are also considered, and act as attractors whereas the object of interest is the global shape or envelope of trees. A widely use method to avoid such artifacts is to consider only the laser point with maximum altitude within each pixel. Then two mains possibilities exist for the calculation of the DSM:

- affect to each pixel the altitude of the highest laser point (figure 6.3b), or
- compute the altitude at the center of each pixel by interpolation of laser points that are maxima in their pixels only (figure 6.3c).

With the first possibility, pixels which do not contain any laser points have no value, but can be filled by considering their neighborhood. The problem with filling empty pixels by interpolated values is that the added information can not be separated from the information that really comes from the point cloud. By interpolating maximum points, canopy appears somewhat shrunk compared to the image obtained by only retaining the maximum value. In particular, the height of treetops might be underestimated. However, comparison of the DSMs shows that mean difference between the non-interpolated and interpolated images is only 0.12 m, with 1.17 m standard deviation. Large differences are mainly located in areas with large height variation, such as tree edges.
Besides, considering the height of the maximum point without taking its position in the pixel into account might lead to over or underestimation of tree heights when the terrain altitude (estimated at the center of the pixel) is subtracted in areas with significant slope. In the case of a 30% slope, 0.5 m spatial resolution, and assuming plane ground surface, pixel height additive error will be at most 0.11 m with a mean value of zero.
Such values are of the same order of magnitude as LiDAR altimetric accuracy, and lower than height measures performed on the field. Therefore, the method based on retaining the maximum value for each pixel, without interpolation, will be used from now on. The issue of void pixels will be addressed later on.



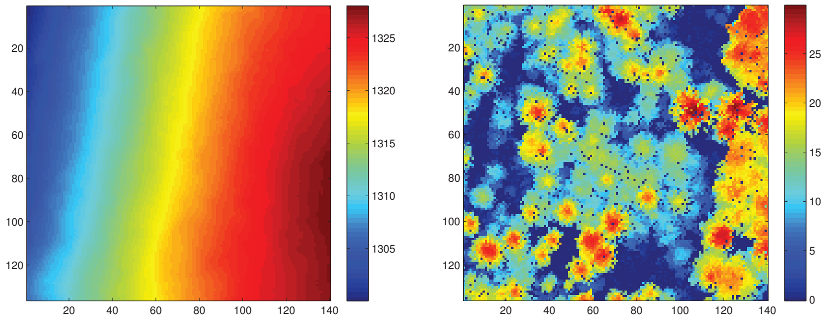
(a) Interpolation of all points (b) Maximum value in pixel (c) Interpolation of maximum values

Figure 6.3: Digital surface models of plot chablais3 (0.5 m spatial resolution).

6.2.2.3 Digital terrain model

The digital terrain model (DTM) is the image of ground elevation. Ground points account for a very small proportion of the whole point cloud in forested areas. For example, on the Vaujany plot, ground point density is only 1.3 m^{-2} (whole point cloud: 37.3), and the standard deviation 1.8 m^{-2}. In particular, laser pulses are unable to reach the forest floor below large trees with dense crowns, even in areas with high point density. Such areas are identifiable as the dark blue patches in figure 6.2b. Ground point density is higher in areas where trees are more sparse (bottom left quarter) or in unforested areas such as roads (lines along the left border and in the bottom right corner). The low density of ground points has two origins. Either canopy is very

dense and few pulses reach the ground, or pulses are actually reflected by the ground, but terrain surface is so rough or so crowded by low vegetation that a very conservative strategy is adopted for point classification. In both cases, the number of points classified as ground points is low and terrain surface is only coarsely modeled.
Although the resulting ground point density is theoretically not sufficient to derived a DTM with the same spatial resolution as the DSM, assumption is made that terrain aspect is locally plane. The missing values will be interpolated so that terrain height is available to estimate tree height by subtracting it to the DSM at treetop locations. The values of the DTM are calculated by bilinear interpolation of ground points at the coordinates of pixel centers. The canopy height model (CHM) is calculated as the difference between the DSM and the DTM (figure 6.4). One must bare in mind the fact that in rough terrain, errors in the DTM will propagate into canopy height estimates.



(a) DTM (b) CHM

Figure 6.4: Digital terrain model and crown height model of plot chablais3 (0.5 m spatial resolution).

6.2.2.4 Choice of DSM or CHM for maxima detection

At this stage, it is tempting to use the CHM for treetop detection. Indeed, the vegetation height is directly accessible. However, the subtraction of the DTM to the CHM results in a distortion of tree crowns in areas with steep slope. In particular, treetop position is shifted toward downslope for trees with relatively flat or rounded tops (figure 6.5). The problem exists for broadleaved trees, but also for some conifers, such as silver firs, whose tops tend to flatten with age for dominant, mature trees. The order of magnitude of position shift is metric, so that is is not crucial for forest inventory purposes, but it represents a additional error source for automatic matching of field trees to detected treetops. In such cases the treetop height is also over-estimated. To avoid this shifting effect, the DSM is used for local maxima detection. The CHM is displayed for the convenience of visual analysis in the following figures and used for a posteriori retrieval of treetop height. Anyway, preliminary tests showed that detection performance was only marginally affected when the CHM is used for local maxima detection instead of DSM.

6.2.3 Non-linear filters

The scanning pattern results in several artifacts in the DSM:

- void pixels where no laser points are present (insufficient sampling density, shading effect of trees at non nadir angles),
- low pixels in tree crowns (the laser pulse enters inside the tree crown through gaps and is reflected by low branches or leaves).

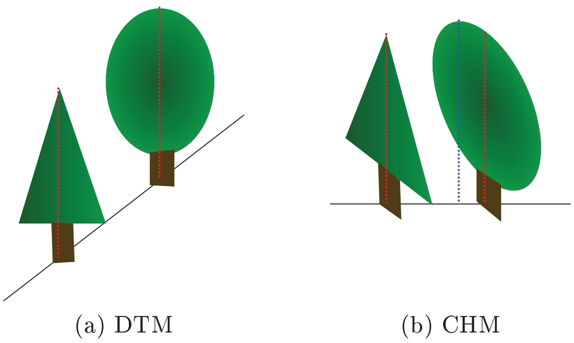
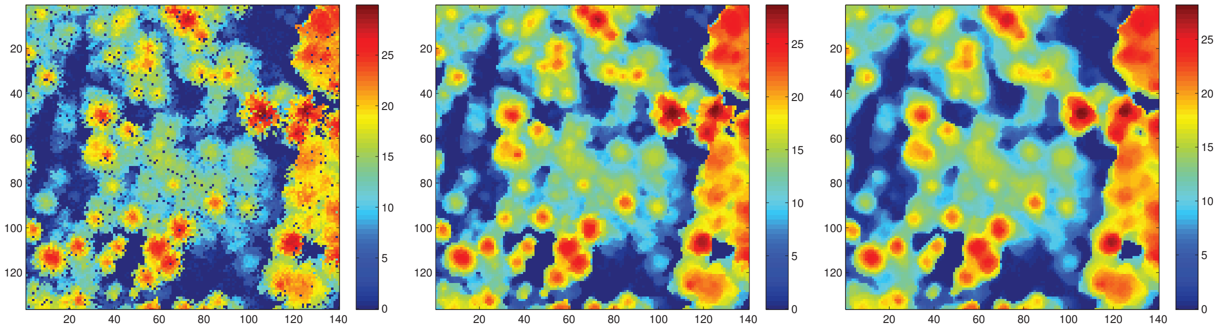

(a) DTM (b) CHM

Figure 6.5: Deformation of tree crown due to the correction of vegetation height with terrain height. For the round tree, height is overestimated and treetop position shifted (blue line).

As built-in image processing algorithms have no specific procedure to handle void pixels, the values of the corresponding pixels in the DTM are affected to the DSM, which is equivalent to setting them to zero in the CHM.

6.2.3.1 Median filter

The presence of low pixels in the DSM, or more precisely of zero values in the CHM can be considered as impulse or "salt and pepper" noise. In this case, the median filter, which is an order-statistic (non linear) filter, is particularly effective [92, p. 326]. The median filter replaces the value $f(x, y)$ by the median of the values of pixels located in a given neighborhood S_{xy}. Effect of the median filter for different sizes of the neighborhood (defined by n where $2n+1$ is the width of the square window) is depicted in figure 6.6. In this case, the noise is removed even with an application of the median on a small neighborhood ($n = 1$). Tree contours are also smoothed. Trees that appear as stars in the CHM (characteristic of spruce branches, such as dominant trees in plot chablais3, for example the tree at matrix coordinates 50,105) appear more rounded. One disadvantage is that local maxima (possible treetops) also have their values replaced by the median, which might lead to an under-estimation of height. Moreover, the "pepper-like" noise in not uniformly distributed, as patches (shading effect of trees, as exemplified in the canopy gap in the bottom-right of figure 6.3b) or lines (lower pulse density across platform trajectory) might appear. The neighborhood must be expanded to satisfactorily filter such cases, but this will result in a loss of information in other areas.



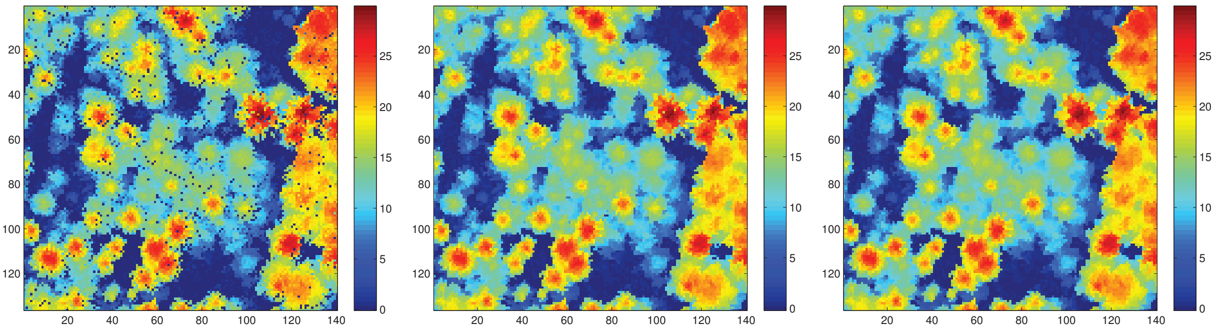
(a) Original image (b) Filtered $n = 1$ (c) Filtered $n = 2$

Figure 6.6: Effect of median filter with different neighborhood size (square window of $2n+1$ pixels width).

A solution for both aspects is the adaptive median filter. This filter acts in two steps. Consider a pixel (x, y) and its neighborhood S_{xy}. First the algorithm compares the median value of S_{xy} to the minimum and maximum in this neighborhood. If the median is equal to one of those value, it is likely that a majority of pixels in the neighborhood are actually noise pixels, and therefore the neighborhood is increased and the analysis re-iterated. Otherwise the algorithm proceeds to the second step. It is now the value at the considered pixel $f(x, y)$ that is compared to the minimum and maximum. If it is equal to one of them, it is likely that the pixel is indeed noisy and it is replaced by the median value. Otherwise, its value remains unchanged.

The adaptive median filter has two advantages over the basic median filter: it adapts the neighborhood until the supposed proportion of noisy pixels is low enough, and it replaces only the noisy pixels. In our case, where only the "pepper-like" noise is present, the algorithm is modified so that maximum values are not taken into account in the first step, nor replaced in the second one. To take advantage of the adaptive neighborhood, void pixels in the DSM are not set to the values in the DTM but to zero, contrary to other tested non-linear filters. Indeed, a patch of void pixels in the DSM will be replaced by several different values from the DTM, as terrain is generally not flat. The patch will not appear as a uniform, noisy area and the algorithm will not increase the neighborhood.

Besides, this problem still exists for low values within the vegetation. For example within a crown of height above 30 m, two adjacent pixels with values 21 and 22 m can be found. The adaptive median algorithm will only correct the lower. It must be applied twice if both values are to be corrected. Figure 6.7 displays the effect of one or two successive applications of the adaptive median filter. With one iteration, most of low pixels are corrected. With two iterations (figure 6.7c), remaining low pixels within the star-shaped spruce crown have disappeared.



(a) Original image (b) One application (c) Two applications

Figure 6.7: Effect of adaptive median filter.

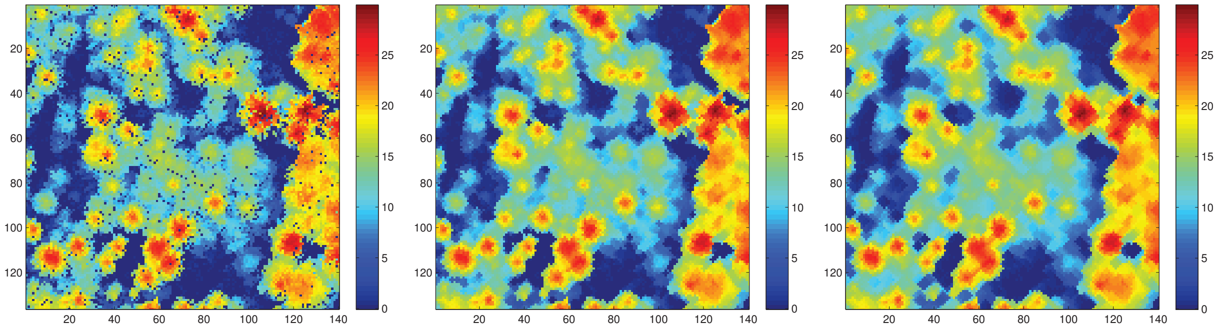
6.2.3.2 Morphological filters

Alternative tools for the removal of low pixels are provided by mathematical morphology. This powerful theory has numerous applications for image processing problems. Operations in mathematical morphology are based on structuring elements (SE) i. In this study, only centered, binary disks are used. This choice is driven by the fact that trees, which are the elements to be highlighted in the image, are generally of circular shape. The two basic operations of mathematical morphology are erosion and dilation. In a gray-scale image, the erosion ϵ^i (resp. dilation δ^i) of an image consists in replacing every pixel by the minimum (resp. maximum) value of its neighborhood defined by the structuring element i. With these two basic operators, more complex operations are defined. Closing ψ of an image f consists in a dilation followed by erosion with the same SE i.

$$\psi^i(f) = \epsilon^i(\delta^i(f)) \tag{6.1}$$

Its effects are to remove small holes and fill narrow gaps (figure 6.8). Even with a small structuring element, holes are filled and maximum heights are not lowered (figure 6.8b). The discrete approximation of a circle of five pixels width is actually diamond-shaped, which tends to produce diagonal artifacts in the filtered image. Besides, some bridges appear where trees were previously separated by narrow gaps, and bigger holes, possibly true canopy gaps, are reduced. The discrete approximation SE_n^{disk} of the disk structuring elements of size n used in the figures are:

$$SE_1^{disk} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \qquad SE_2^{disk} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{6.2}$$



(a) Original image (b) Closing with SE_1^{disk} (c) Closing with SE_2^{disk}

Figure 6.8: Effect of closing filter with disks of different sizes as structuring elements.

To avoid this tendency to over-fill the gaps, and more generally the degradation of original shapes by the structuring element, the closing by reconstruction operator ψ_R^i can be introduced as [276]:

$$\psi_R^i(f) = R_f^\epsilon(\delta^i(f)) \qquad \text{with } R_f^\epsilon \text{ the geodesic reconstruction by erosion} \tag{6.3}$$

The advantage of reconstruction operators is that features smaller than the structuring element are removed, but without altering the shape. Closing by reconstruction removes the unconnected low features (figure 6.9). The low pixels are satisfactorily filled while the original shape is preserved. However, some low pixels remain in the tree crowns.

6.2.4 Lowpass filter

The previous step of non-linear filtering mainly aims at removing the artifacts in the vegetation model due to the acquisition parameters. The next step is to keep only the information related to trees. Indeed, the final objective is to detect the treetops with a local maxima filtering. A simplified frequential interpretation of a canopy-height or digital-surface model image is to consider that:

- very low frequencies are linked to terrain variations, provided that slope is constant a scale smaller that the width of a tree (approx. 10 m period),
- low frequencies are linked to tree envelopes (2 m),
- medium frequencies to branches (0.5 m),
- and high frequencies to small branches, leaves and LiDAR noise (0.1 m).

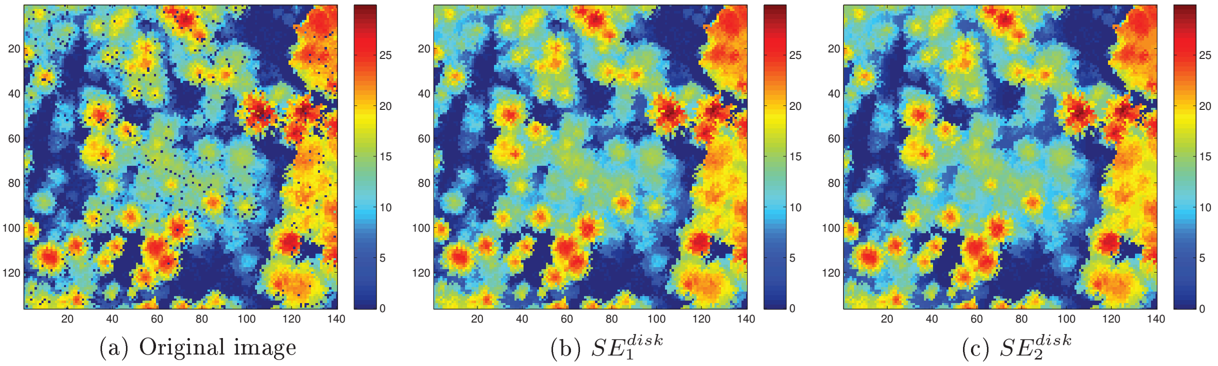

(a) Original image (b) SE_1^{disk} (c) SE_2^{disk}

Figure 6.9: Effect of closing by reconstruction filter with disks of different sizes as structuring elements.

The high frequency content is likely to induce several local maxima that will not correspond to actual treetops. Under the strong hypothesis that branches, leaves and LiDAR noise can be considered as a random noises with zero mean, one efficient way to remove them is linear smoothing. Here discrete approximations of Gaussian lowpass filters are used. The discrete approximation G_σ of a Gaussian kernel of parameter σ is convolved with the non-linear filtered vegetation height image f to obtain the smoothed image $\hat{f}$.

$$\hat{f} = f \star G_\sigma \tag{6.4}$$

Examples of Gaussian smoothing applied to the digital surface model previously filtered by the adaptive median filter are displayed in figure 6.10. With $\sigma = 0.25$ m (figure 6.10a), the effect is barely noticeable. In particular, branches of the largest spruce trees are still remarkable, and low pixels in the crown remain visible (matrix coordinates 50, 105). With $\sigma = 1$ m, small adjacent trees that were identifiable as small yellow patches in the center of the image filtered by adaptive median image (figure 6.7b) are merged into a single, banana-shaped blob (matrix coordinates 70, 60 in figure 6.10c). With $\sigma = 0.5$ m, trees appear considerably less branchy but small trees are preserved.

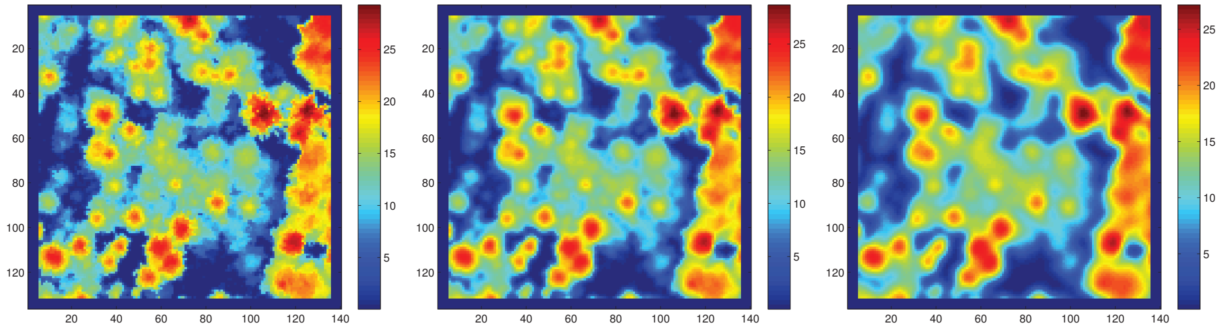
It is noteworthy that Gaussian smoothing does not handle the same type of noise as the previous non-linear filters. Those two types of filter are complementary. For example, if a very low value is present within a crown, the low-pass filter will only partially increase it, and will also lower the nearby values. On the contrary, non-linear filters do not remove all the high-frequency noise, even tough the median filter (and to a lesser extent, closing filter) has a smoothing-like effect when employed on large windows.

6.2.5 Local maxima filter

Local maxima are detected by checking for each pixel (x, y) if its value is equal to the maximum value in its neighborhood S_{xy}. If $f(x, y)$ is indeed a global maximum in S_{xy}, the operation is reiterated with a larger neighborhood. Here, considered neighborhoods are centered, square windows of width $2n + 1$ pixels, S_{xy}^n. During the sliding process of the variable size window, a maxima image m is computed.

$$m(x,y) = \begin{cases} 0 & \text{if } f(x,y) < max_f(S_{xy}^1) \\ \frac{2n+1}{2} & \text{if } f(x,y) = max_f(S_{xy}^n) \text{ and } f(x,y) < max_f(S_{xy}^{n+1}) \end{cases} \tag{6.5}$$

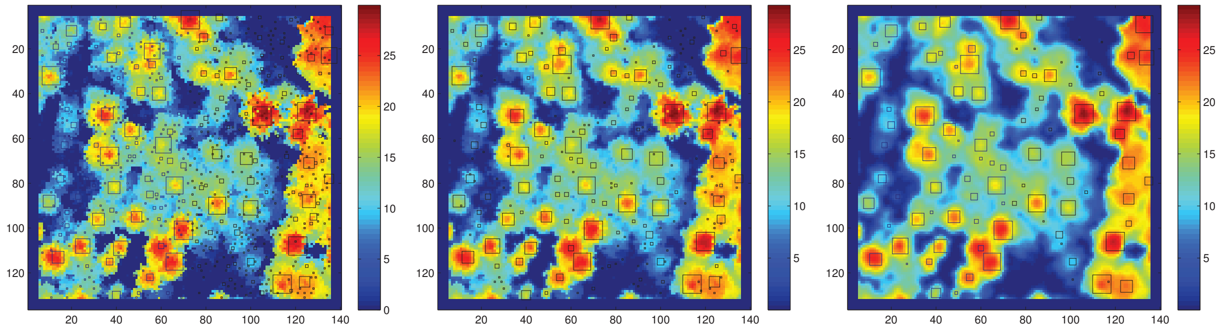
where $max_f(S_{xy}^i)$ is the maximum value of f in the neighborhood S_{xy}^i. The value of a pixel in m is zero if this pixel is not a local maximum, and is otherwise equal to half the width of the largest



(a) $\sigma = 0.25$ (b) $\sigma = 0.5$ (c) $\sigma = 1$

Figure 6.10: Effect of Gaussian smoothing of the image filtered by adaptive median (figure 6.7b), with different values for σ (m). Spatial resolution is 0.5 m. Dark-blue frames in the filtered images mask the border effects.

square window where it is global maximum. In this process, 8-adjacent maxima with equal height may be extracted. In such cases, only one maxima among the maxima that are 8-connected is retained. Results of maxima extraction on differently filtered images are displayed in figure 6.11. Without lowpass filter (figure 6.11a), a high number of maxima is detected. Their number seems more reasonable when a narrow Gaussian kernel is use for smoothing (figure 6.11b). However, some false detections remain, mainly in non forested gaps (dark blue areas), but also in the form of nearby maxima with a small value in m (compact groups of small squares). With $\sigma = 0.5$ m, small maxima groups have disappeared. Unfortunately, many local maxima with small values in m that where close to larger ones have also been removed, for example in the spruce tree at matrix coordinates (50, 105), or near the silver fir at (70, 35).



(a) Adaptive median only (b) Adaptive median + $\sigma = 0.25$ (c) Adaptive median + $\sigma = 0.5$

Figure 6.11: Local maxima detected on images obtained by a combination of adaptive median filtering and Gaussian smoothing. Background is the filtered image used for local maxima detection. Black squares are the detected maxima. Symbol size is proportional to the value of pixels in $m(x, y)$.

6.2.6 Maxima selection

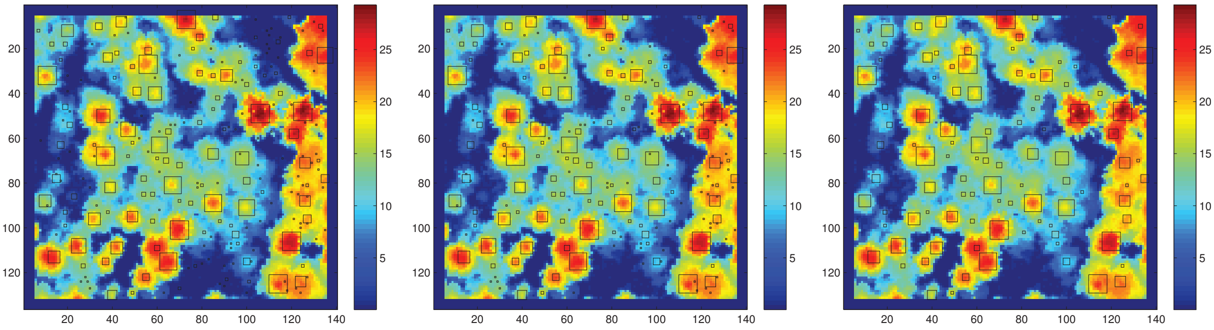
To address the problem of the over detection of treetops, an additional procedure for maxima selection is performed. It relies on two main assumptions:

- trees have a minimum height;
- the higher a tree, the bigger its distance to its neighbor of equal height.

The maxima selection procedure is based on the mathematical transcription of those hypotheses into two inequalities involving a minimum height and the relationship between a maxima height $h_l(x,y)$ and its distance $m(x,y)$ to the nearest higher pixel. h_l is the canopy height model calculated as the difference between the non-linear filtered DSM and the DTM. Even tough the local maxima are detected on the smoothed DSM, the previous non-linear filtered DSM is preferred for the extraction of maxima height as smoothed heights are generally decreased for maxima pixels. In the end, a local maximum, with coordinates (x,y), which does not fulfil the following inequalities, is discarded.

$$\begin{cases} m(x,y) \geq m_{min} + m_{prop} \times h_l(x,y) \\ h_l(x,y) \geq h_{min} \end{cases} \tag{6.6}$$

$(m_{min}, m_{prop}, h_{min})$ are the selection parameters. The final image of selected maxima $m_s(x,y)$ is obtained by setting to zero the values in $m(x,y)$ that do not comply with inequalities 6.6. Figure 6.12 displays the result of maxima selection. The height threshold removes local maxima in the low vegetation (dark blue areas in figure 6.12b). The threshold applied to m removes maxima on the smallest neighborhoods (figure 6.12c). Some of them might actually be treetops.

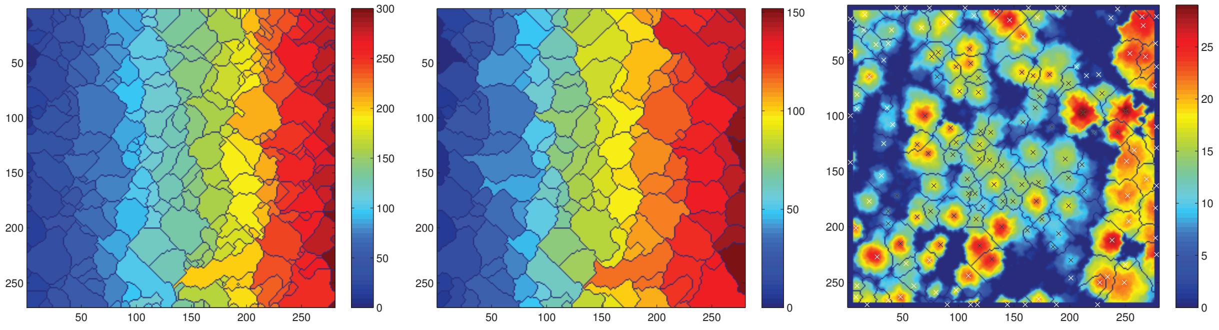


(a) No selection (figure 6.11b) (b) $h_{min} = 5$ m (c) $m_{prop} = 0$, $m_{min} = 1$ m

Figure 6.12: Effect of local maxima selection. Symbols are the same as in figure 6.11.

6.2.7 Extracted parameters

For each selected maxima – pixel (x,y) with $m_s(x,y) > 0$ – two values are available. $m_s(x,y)$ is the distance to the nearest pixel that has higher value, and is of little interest once the selection has been performed. $h_l(x,y)$ is the estimated height. An additional information about the alleged tree is obtained by segmentation of the DSM to delineate tree boundaries. The limits between adjacent trees can be determined using the watershed segmentation algorithm [92, p. 769]. This morphological operation consists in determining areas where all pixels are connected to a single minimum with a monotonically decreasing path. These areas are called the watersheds. Pixels that divide two or more adjacent watersheds are the watershed lines. Each watershed in the flipped-over DSM can be considered as the crown of the tree associated to the local maximum in the DSM. Figure 6.13a shows the segments obtained by watershed segmentation of the smoothed DSM. As each maximum induces a new watershed, the image is oversegmented. To obtain a segmentation that is consistent with the selected maxima, the watershed algorithm is constrained by previously determined markers: a local minima can not give birth to a new watershed if it is not located on a marker. Here, markers are the selected maxima. Figure 6.13a displays the segments obtained by marker-constrained watershed, and figure 6.13c displays the watershed lines on the CHM, as well as the selected maxima. After segmentation, an additional LiDAR parameter is associated to each selected maxima: the area of its watershed a_l, given by the number of pixels within its segment multiplied by the pixel size in square meters.



(a) Segments obtained without . . . (b) . . . and with markers (c) Watershed lines and markers

Figure 6.13: Tree delineation with watershed segmentation: (a) resulting segments, (b) result of segmentation constrained by the selected maxima as markers and (c) watershed lines superimposed on the CHM along with selected maxima (×, white crosses are the selected maxima located outside the field plot mask). In (a) and (b) colors refer to segment numbers.

6.3 Performance assessment

The objective is to test the sensitivity of the detection algorithm to parametrization and to forest structure. As numerous cases will be tested, it is not possible to manually check all detection results. An automatic assessment procedure will not be able to handle all particular cases, however it is much faster and will treat all cases consistently. The procedure used in this work includes two aspects.

- An automated matching procedure to assess the detection rate. It is based on the 3D distance between selected maxima and the positions of field treetops.
- A “score” value for the detection performance that takes both true and false positives into account.

Those two aspects are presented in the following paragraphs.

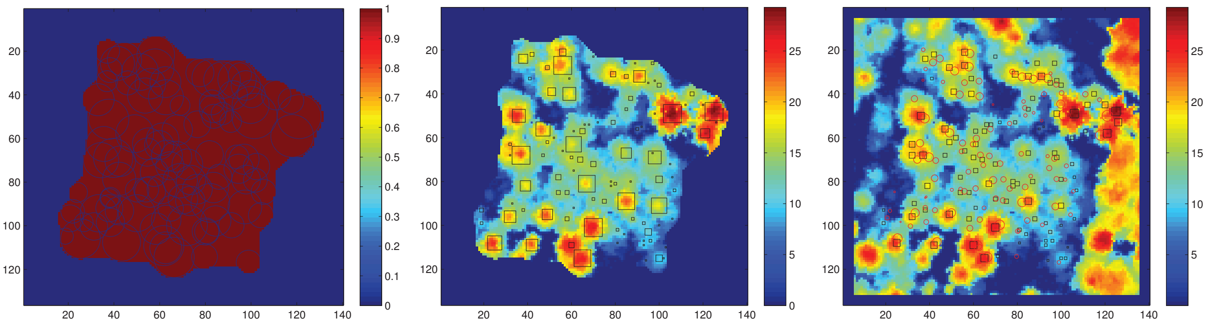
6.3.1 3D matching procedure

To assess the accuracy of treetop detection, an automated matching procedure is performed. It is based on the comparison of distances between selected maxima and the positions of treetops alleged from the field inventory, and estimates the proportion of correctly detected trees. It is assumed that GPS positioning planimetric error ϵ_{gps} is inferior to 1.5 m and that the slope of a tilted tree s_{tree} is inferior to 14% (value determined by measures performed on the 26 highest trees in Vaujany). Height measures accuracy is estimated to be $\epsilon_h = 15\%$. Under these asumptions, the limit matching distance between a selected maxima and a theoretical treetop located vertically at h_t meters above the GPS position of tree T is:

$$d_{max}(h_t) = \epsilon_{gps} \times \sqrt{1 + s_{terrain}^2} + s_{tree} \times (\epsilon_h + 1) \times h_t \qquad (6.7)$$

with $s_{terrain}$ the terrain slope, set to 30%.

To avoid border effects, the first step is to remove selected maxima that are not located on the field plot. One option is to geolocate the field boundaries with a GPS and to exclude maxima that do not fall inside the theoretical plot. Due to topographic constraints, real field plot may be quite different from the plot defined by the GPS coordinates of its four corners. Besides, the top of an inventoried tree may be located outside plot boundary because of tilt. Therefore, the plot area is defined as the union of buffer disks created around each tree with a radius proportional to tree height, and given by equation 6.7. The field plot mask is thus a binary image (figure 6.14a). Possible holes in the buffer are filled by a closing operation with six meter radius circular structuring element. Previous border is reconstructed by morphological reconstruction. Final matrix of selected maxima on the field plot is obtained as the elementwise product of the selected maxima image with plot mask image. Only those maxima are used during the matching procedure.



(a) Plot mask (b) Remaining maxima (c) Remaining maxima and field trees

Figure 6.14: Matching procedure. (a) Construction of plot mask (background image) from the regularization of tree buffers (circles). (b) Selected maxima located inside the buffer (black squares, size proportional to pixel value in m) and (c) comparison with field trees (red circles, symbol size proportional to height).

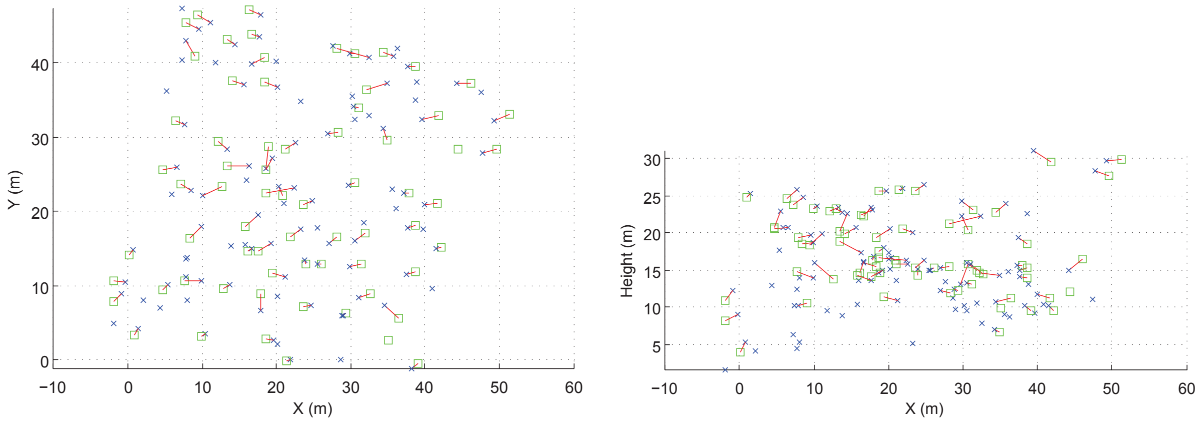
For each pair constituted of a tree T and a maximum M of respective heights h_t and h_m, a matching index $I_{T,M}$ is computed as the ratio of the distance $d_{T,C}$ between the assumed treetop and the detected maximum, and $d_{max}(h_t)$.

$$I_{T,M} = \frac{d_{T,M}}{d_{max}(h_t)} \tag{6.8}$$

The potential pair with the lowest matching index is validated and the lists of remaining maxima and trees are updated before reiterating the procedure. Linking ends when all remaining values of $I_{T,M}$ are greater than 1. From this procedure, three cases can be distinguished:

- true positive (or correctly detected tree), when a selected maximum is successfully linked to a field tree,
- false positive (or false alarm, commission error), when a selected maximum can not be matched to any field tree,
- false negative (omission error), when a field tree can not be matched to any selected maximum.

Let N_{TP}, N_{FP}, N_{FN} be the numbers of true positives, false positives and false negatives respectively. We have $N_{TP} + N_{FN} = N_T$ and $N_{TP} + N_{FP} = N_M$ where N_T is the number of field trees and N_M the number of selected maxima. An example of result of automated matching of selected maxima with field trees is displayed on figure 6.15. The fact that detected trees are mainly dominant ones appears clearly on figure 6.15b.



(a) Top view (b) Lateral view

Figure 6.15: Result of automated matching (– red lines) between □ LiDAR selected maxima and × field treetops (scales are in meter).

6.3.2 Optimality criterion

Effect of an algorithm parameter on detection performance can be displayed on a receiver operating characteristic (ROC) curve, which is the plot of the true positives ratio versus the false positives ratio. Here true and false positive ratios are defined as $R_{TP} = \frac{N_{TP}}{N_T}$ and $R_{FP} = \frac{N_{FP}}{N_T}$.

Besides, for accuracy assessment a trade-off between the true and false positives ratios has to be found. It is assumed that for every five additional true positives, one false positive is tolerated. The detection score S is thus computed as:

$$S = (w \times R_{FP})^2 + (1 - R_{TP})^2 \qquad \text{with } w = 5 \tag{6.9}$$

which corresponds to the squared Euclidean distance to the perfect match ($N_{FP} = 0$ and $N_{TP} = N_T$) with the false positives ratio weighted by a factor 5.

Chapter 7

Exploratory sensitivity analysis

7.1 Material and methods

The global workflow for treetop detection on a plot is displayed in figure 7.1. Each step of the algorithm is presented in the previous chapter.

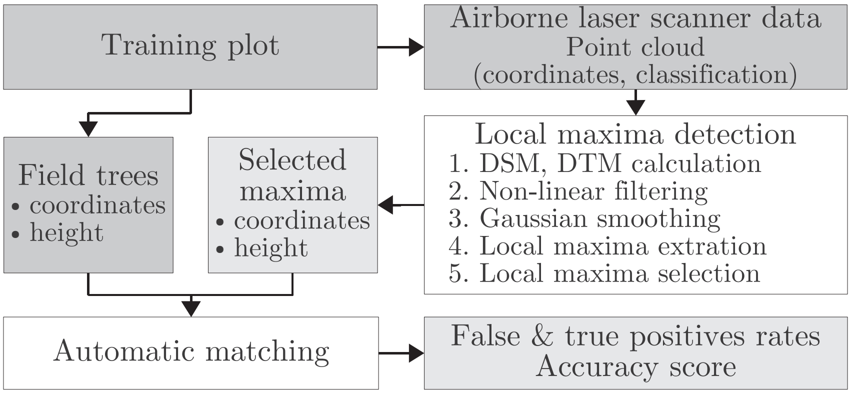


Figure 7.1: Workflow for treetop detection.

Field data The algorithm is tested on 22 plots from the following LiDAR acquisitions:

- Chamonix (three plots: Moussoux, Orthaz and Corruaz),
- Chablais (eight plots numbered from 1 to 8),
- Murianette (Combeloup plot),
- Vercors (Saint-Agnan marteloscope divided in four plots numbered from 1 to 4),
- Valdrôme (marteloscope divided in four plots numbered from 1 to 4).
- Vaujany (rockfall experiment area divided in two plots: lower and upper part).

All plots are square plots of 0.25 ha, except at Saint-Agnan (rectangular shape: 40×62.5 m^2) and Vaujany (no particular shape, approximately 0.33 ha). The size of these training plots is necessary to limit border effects during the matching process. Indeed, errors may occur at plot borders due to poor GPS precision and stems being tilted. With bigger plots, the area to perimeter ratio increases, which decreases the influence of border effects on detection score.

LiDAR data Acquisition parameters of LiDAR surveys are presented in table 4.2. Resulting echo densities for the field plots are displayed in table 7.1.

Table 7.1: LiDAR echo densities (m^{-2}) of points classified as ground (G) and vegetation (V) for the 22 field plots.

Plot id	G	V	Plot id	G	V	Plot id	G	V
chablais1	0.8	19.0	combeloup	3.0	3.6	stagnan4	0.9	8.4
chablais2	1.1	10.7	corruaz	0.4	6.0	valdrome1	1.9	8.5
chablais3	1.1	12.4	moussoux	0.3	8.0	valdrome2	1.7	7.5
chablais4	0.9	13.1	orthaz3	0.5	8.7	valdrome3	1.5	12.8
chablais5	0.5	17.6	stagnan1	0.9	9.9	valdrome4	1.6	14.0
chablais6	0.8	12.3	stagnan2	1.1	11.9	vaujanyba	1.4	37.8
chablais7	0.7	8.0	stagnan3	0.7	8.7	vaujanyha	0.8	20.7
chablais8	0.8	26.0						

Tested parameters To assess the sensitivity of the detection algorithm to parametrization and forest structure, for each training plot numerous combinations of parameters are tested.
The tested values for the spatial resolution of the DSM and DTM images are $res \in \{\frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{2}{3}, 1, \frac{3}{2}\}$ in meters. Preliminary tests with pixel width greater than 2 m showed that detection potential is drastically limited for coarse resolutions.
The preliminary tests also showed that the interpolation method used for DSM calculation has little effect on detection potential. Similarly, extracting the maxima on the DSM or on the CHM yielded similar results. Based on those findings, the DSM is chosen as the basis image for maxima extraction and it is calculated by retaining in each pixel the maximum height of contained LiDAR points.
Non-linear filters are tested separately. Adaptive median is applied successively $n_{adapt} \in \{1, 2, 3, 4\}$ times. Median and morphological filters are applied with filter size $r_m \in \{0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4\}$ in meters. Depending on the resolution res, discrete filter size is approximated to the nearest integer number of pixels. r_m corresponds to the window half-width for the median filter and to structuring element size (disk radius) for morphological filters.
Tested values for the Gaussian smoothing parameter are $\sigma \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1, 1.25, 1.5\}$ in meters. Discrete approximations of the Gaussian distribution are used.
Criteria for the height threshold of detected maxima are $h_{min} \in \{0, 2.5, 5, 7.5, 10, 12.5, 15\}$ in meters.
Tested criteria for the relationship between a maxima height and its distance to the nearest higher pixel are $(m_{min}, m_{prop}) \in \{(0.5, 0), (0.75, 0), (1, 0), (1.25, 0), (1.5, 0), (1.75, 0), (2, 0), (1.5, \frac{1}{80}), (1, \frac{1}{40}), (\frac{2}{3}, \frac{1}{30}), (\frac{2}{5}, \frac{1}{25}), (0, \frac{1}{20})\}$

For each plot, a total of 188160 combinations are tested. All computations are performed using Matlab® and its Image Processing Toolbox.

7.2 Results

7.2.1 Detection performance

Due to some redundant discrete approximations of parameters, the final number of obtained combinations is only 111720 for each plot. Table 7.2 presents the best detection results obtained on each plot and the corresponding settings. True positives ratio ranges from 28.7 to 80.5 %. False positives ratio is lower than 6.5 %. Scores are in the interval [0.06, 0.54]. The best performance is for plot chablais6, which is a recently thinned young spruce plantation on a flat area ($R_{TP} =$ 80.5% with five false positives). Detection ratio is only 28.7% in Combleloup, with four false positives.

Pixel sizes for these best settings are always below 1/3 m, except for plot 2 in Saint-Agnan with 2/5 m. All non-linear filters are represented, save closing by reconstruction (adaptive median: 7, median: 6, closing: 7). On one plot, the best result is obtained without non-linear filtering (orthaz3). Regarding other parameters (non-linear and Gaussian filters sizes, selection criteria), no trend is remarkable.

The height distributions of detected and missed trees, as well as false positives is available in appendix C.3.1 (figures C.8 and C.9). Dominant trees are generally well detected. The detection rate decreases for smaller trees.

Table 7.2: Best detection results and corresponding algorithm parameters for the 22 tested plots. Non-linear parameter n corresponds to the number of iterations of the adaptive median filter, or to the width $2n+1$ in meters of the window (median filter) or of the structuring element (closing filters).

Plot id	res	Non-linear filter		σ	h_{min}	m_{min}	m_{prop}	R_{TP}	R_{FP}	S	Field
	(m)	Type	n	(m)	(m)	(m)		(%)	(%)		trees
chablais1	1/3	ad. med.	4	0.3	5	0	0.05	51.6	3.9	0.27	153
chablais2	1/5	closing	2	0.3	15	0.75	0	46.6	2.9	0.31	174
chablais3	1/4	closing	0.5	0.2	2.5	1	0	64.7	2.0	0.13	102
chablais4	1/3	closing	2	0.1	5	0.4	0.04	47.6	3.7	0.31	82
chablais5	1/4	closing	0.75	0.5	15	0.75	0	55.9	2.1	0.21	145
chablais6	1/3	ad. med.	3	0.3	7.5	0.5	0	80.5	2.7	0.06	185
chablais7	1/5	closing	0.4	0.75	5	0	0.05	36.7	5.8	0.48	139
chablais8	1/4	median	1	0.1	10	0.5	0	45.9	3.7	0.33	109
combeloup	1/5	closing	1.4	0.3	10	2	0	28.7	3.5	0.54	115
corruaz	1/5	closing	0.8	0.1	10	0	0.05	31.7	2.7	0.48	186
moussoux	1/5	ad. med.	2	0.5	12.5	1.25	0	44.5	6.0	0.40	281
orthaz3	1/2	none	0	0.3	7.5	0.67	0.03	45.3	4.1	0.34	148
stagnan1	1/5	closing	0.8	0.4	10	0.5	0	55.1	3.8	0.24	78
stagnan2	2/5	ad. med.	1	0.5	15	1.5	0	47.9	0.0	0.27	94
stagnan3	1/3	median	0.67	0.75	7.5	0.5	0	40.4	1.1	0.36	94
stagnan4	1/3	median	0.67	0.5	10	1.25	0	43.0	4.3	0.37	93
valdrome1	1/3	median	0.67	0.2	10	0.4	0.04	67.7	3.0	0.13	198
valdrome2	1/4	ad. med.	4	0.4	12.5	1.5	0	67.1	4.0	0.15	149
valdrome3	1/5	ad. med.	4	0.4	12.5	0.4	0.04	58.9	5.3	0.24	265
valdrome4	1/5	median	0.2	0.2	10	0.4	0.04	58.2	6.5	0.28	400
vaujanyba	1/5	ad. med.	4	1	7.5	0.5	0	38.0	0.0	0.38	92
vaujanyha	1/3	median	1	0.5	7.5	1.25	0	38.5	2.1	0.39	96

In case of tied scores for a given plot, the following criteria are successfully considered until only one result remains:

- lower false positives ratio,
- larger pixel width,
- median filter is used,
- closing filter is used,
- closing by reconstruction filter is used,
- window width for non-linear filtering is the biggest, or number of iterations for adaptive median is the smallest,
- height threshold h_{min} is the largest,
- value for the distance threshold m_{min} is the largest,
- value for coefficient m_{prop} is the largest.

7.2.2 Estimation of tree parameters

7.2.2.1 Introduction

After the matching procedure of trees with maxima selected with the best setting, two lists are obtained for each plot: the list of correctly detected trees with their parameters measured on the field (height h and diameter d) and the list of corresponding selected maxima with their LiDAR parameters (height h_l and crown area a_l). For trees on a given plot, linear regression models are trained to fit relationships between the field parameters h and d and the extracted LiDAR parameters h_l and a_l.

For the Valdrôme plots, all heights have not been measured on the field. Only correctly detected trees with a field value are used in the analysis.

7.2.2.2 Height estimates

Statistics of the difference between LiDAR height and the field measured height of the correctly detected trees are shown in table 7.3. Mean difference is either negative or positive (-1.0 to 1.3 m). The distribution of $h - h_l$ is not symmetric about 0 ($p < 0.05$, two-sided Wilcoxon signed-rank test) for thirteen plots. The highest biases are for plots chablais6 and orthaz3. Standard deviation of differences ranges from 0.77 m to 2.2 m, respectively for plots chablais6 and combeloup. The root mean square of difference $h - h_l$ ranges from 1.0 to 2.2 m. Low values are obtained in plots chablais3, chablais5 and chablais7.

Linear regression models with field height as dependent variable and LiDAR height as independent variable are fitted by ordinary least squares with the data of correctly detected trees on each plot:

$$h = \alpha + \beta h_l \qquad \text{with } (\alpha, \beta) \text{ the model parameters} \tag{7.1}$$

Table 7.3 displays regression model parameters. Data points are very close to the 1:1 line (figure C.10), and regression parameter β is close to 1, except for vaujanyba and some plots in Saint-Agnan and in Valdrôme where it is lower than 0.9. Values for adj-R^2 are above 0.8, except in valdrome2 and valdrome3. Due to the fairly good correspondence between LiDAR height and measured height, the root mean of squared residuals of regression models is only slightly reduced compared to the root mean squared differences (respectively $rmsr$ and RMS in table 7.3). Improvement is mainly due to bias correction in plots with small variance of $h - h_l$ (e.g. plots chablais6, orthaz3 and chablais1).

Plots of h as function of h_l are displayed in the appendix C.3.2 (figures C.10 and C.11), along with the obtained regression models. In certain plots, the heights of the tallest trees seem overestimated (figure C.11, Saint-Agnan plots, valdrome2 and vaujanyba). However, Spearman correlation coefficients ρ between h and $h - h_l$ indicate that significant relationships are found in Chablais plots, combeloup, moussoux and three of the Valdrôme plots, and are in fact positive correlations. In these plots, tree heights are all the more underestimated as trees are tall.

7.2.2.3 Basal area estimates

Two types of regression models are tested for basal area retrieval from two LiDAR derived parameters: height h_l and segment area a_l. The additive model is an ordinary least squares multivariate linear regression, the multiplicative model is obtained by log-transformation of all variables prior to applying the additive model.

$$g = \alpha + \beta h_l + \gamma a_l \tag{7.2}$$

$$g = \alpha h_l^{\beta} a_l^{\gamma} \tag{7.3}$$

Table 7.3: Correctly detected trees with field-measured height: statistics of the difference between ground truth and LiDAR height (mean, standard deviation, root mean square, and Spearman correlation coefficient ρ with measured height), and regression model between field height h and LiDAR height h_l: $h = \alpha + \beta h_l$ (model parameters, adj-R^2 and root mean square residuals). Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $^{\bullet}$ $p < 0.1$. Significance levels in the "Mean" column correspond to the Wilcoxon signed-rank test applied to $h - h_l$.

Plot id	Number	Statistics of $h - h_l$				Regression model parameters			
	of trees	Mean	Std dev.	RMS	ρ	α	β	adj-R^2	rmsr
chablais1	79	0.97^{***}	1.02	1.40	0.48^{***}	0.35	1.05	0.93	1.00
chablais2	81	0.76^{***}	1.62	1.78	0.41^{***}	0.12	1.02	0.86	1.61
chablais3	66	0.46^{***}	0.90	1.00	0.29^{*}	0.17	1.02	0.97	0.88
chablais4	39	0.87^{***}	1.34	1.59	$0.28^{\bullet}$	0.88	1.00	0.98	1.33
chablais5	81	-0.05	1.11	1.10	0.30^{**}	1.13	0.96	0.80	1.10
chablais6	149	1.27^{***}	0.77	1.49	0.22^{**}	1.82	0.96	0.85	0.77
chablais7	51	0.36^{*}	1.19	1.23	0.30^{*}	-0.31	1.04	0.96	1.16
chablais8	50	-0.04	1.66	1.65	0.42^{**}	-1.39	1.06	0.95	1.59
combeloup	33	-0.28	2.24	2.23	0.38^{*}	1.03	0.95	0.79	2.20
corruaz	59	0.57^{**}	1.50	1.59	0.00	1.50	0.95	0.95	1.45
moussoux	125	0.62^{***}	1.76	1.86	0.39^{***}	0.56	1.00	0.83	1.75
orthaz3	67	1.31^{***}	1.18	1.75	0.11	1.06	1.01	0.95	1.17
stagnan1	42	-0.98^{***}	1.45	1.74	0.24	2.27	0.87	0.86	1.35
stagnan2	45	-1.03^{***}	1.89	2.13	0.07	5.26	0.75	0.72	1.64
stagnan3	38	-0.52	1.91	1.95	0.03	3.14	0.84	0.82	1.73
stagnan4	40	-0.16	1.69	1.67	-0.04	2.17	0.89	0.87	1.59
valdrome1	45	0.91^{***}	1.09	1.41	$0.28^{\bullet}$	2.43	0.92	0.81	1.06
valdrome2	33	0.07	1.91	1.89	$0.32^{\bullet}$	6.16	0.71	0.52	1.73
valdrome3	52	0.15	1.49	1.49	0.11	4.76	0.74	0.70	1.30
valdrome4	96	0.39^{**}	1.42	1.47	0.29^{**}	2.00	0.89	0.77	1.38
vaujanyba	35	-0.91^{*}	2.03	2.20	-0.10	2.20	0.89	0.89	1.88
vaujanyha	37	-0.50	1.91	1.95	0.10	1.12	0.94	0.92	1.85

The purpose here is not to develop new LiDAR-based allometric relationships for tree diameter or basal area prediction. It is just to show that, whereas height prediction from LiDAR is quite straightforward, derivation of other tree parameters is a far more difficult task in complex, heterogeneous forests.

Table 7.4 displays the adj-R^2 coefficient for both types of models, when only LiDAR height is used as variable and when crown area is also included as variable. adj-R^2 ranges from 0.06 to 0.80 when only h_l is employed as variable in the additive model. On some plots, including a_l increases the degree of fit. In the first three plots of Valdrôme, additive regression models with only h_l are poorly fitted, but the inclusion of a_l greatly increases the adj-R^2. Improvement is also noticeable in plots chablais1 and stagnan1. With two LiDAR variables, best results for the additive model are obtained for plots chablais1, chablais8, corruaz and chablais3, with adj-R$^2 > 0.8$.

With the multiplicative model, adj-R^2 values are generally better but similar trends are noticeable. Minimum and maximum values are $(0.06, 0.93)$ and $(0.36, 0.93)$ for models without and with a_l respectively. The inclusion of a_l yields a great improvement in all Valdrôme plots, and in chablais8. Obtained best fits are slightly better with adj-R^2 for chablais4, chablais8, Vaujany plots, chablais3 and valdrome2 above 0.8.

It is noteworthy that Spearman correlation coefficient between the adj-R^2 value of the multiplicative model with two LiDAR variables and the stem density has a negative value (-0.62), significant at the $p < 0.01$ level. Correlation with other stand parameters (stand basal area and dominant height) are not significant ($p > 0.1$).

Table 7.4: adj-R^2 of multiple regression models for the prediction of basal area with LiDAR variables (height h_l and crown area a_l for correctly detected trees with field-measured heights).

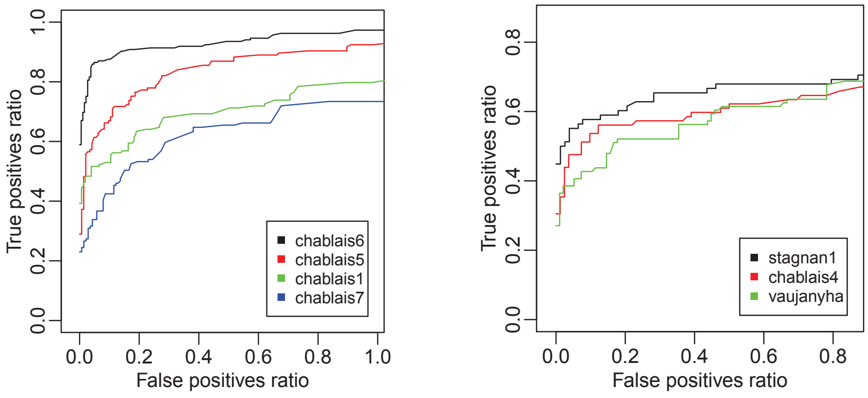
Plot id	Number of trees	adj-R^2 of multiple regression model $g = \alpha + \beta h_l + \gamma a_l$	$g = \alpha + \beta h_l$	$g = \alpha h_l^\beta a_l^\gamma$	$g = \alpha h_l^\beta$
chablais1	79	0.82	0.68	0.66	0.62
chablais2	81	0.32	0.32	0.40	0.40
chablais3	66	0.81	0.76	0.83	0.77
chablais4	39	0.64	0.58	0.93	0.93
chablais5	81	0.47	0.41	0.45	0.43
chablais6	149	0.37	0.35	0.66	0.59
chablais7	51	0.67	0.67	0.77	0.77
chablais8	50	0.82	0.80	0.86	0.77
combeloup	33	0.42	0.39	0.64	0.65
corruaz	59	0.81	0.80	0.72	0.73
moussoux	125	0.38	0.38	0.57	0.57
orthaz3	67	0.71	0.64	0.79	0.77
stagnan1	42	0.62	0.49	0.56	0.47
stagnan2	45	0.62	0.56	0.61	0.52
stagnan3	38	0.55	0.56	0.76	0.76
stagnan4	40	0.65	0.65	0.70	0.71
valdrome1	45	0.66	0.37	0.45	0.44
valdrome2	33	0.77	0.45	0.81	0.44
valdrome3	55	0.49	0.06	0.43	0.06
valdrome4	96	0.37	0.38	0.39	0.36
vaujanyba	35	0.69	0.70	0.84	0.85
vaujanyha	37	0.77	0.74	0.84	0.83

7.2.3 Sensitivity to parametrization

Due to the high number of combinations, only a graphical overview of the results is presented. On the following figures, the monotonically increasing border of the point groups envelopes (efficiency frontier) are represented in the (R_{FP}, R_{TP}) plane, with $R_{FP} \leq 0.2$ and $R_{TP} \geq 0.2$. Each point represents a parameter combination, groups are constituted according to values of a given parameter. The upper border of the different point groups indicates whether a given parameter has limiting effect on detection performance. Obviously, this rough approach cannot address the combined effects of parameters.

Regarding the trade-off between true and false detections, some general trends are noticeable. With homogeneous stands, true positives ratio increases greatly when a few false positives are tolerated. Then true positives ratio further increases slowly even if more false positives are tolerated. Example for this case is given by plot chablais6, which is a young spruce plantation. For stands that are more heterogeneous, such as plot chablais5, chablais1 and chablais7 (figure 7.2a), the observed ROC curve is more round. In heterogeneous stands, the ROC curve also seems more flat for stands dominated by a few very large trees, such as plots vaujanyha, stagnan1 or chablais4 (figure 7.2b).

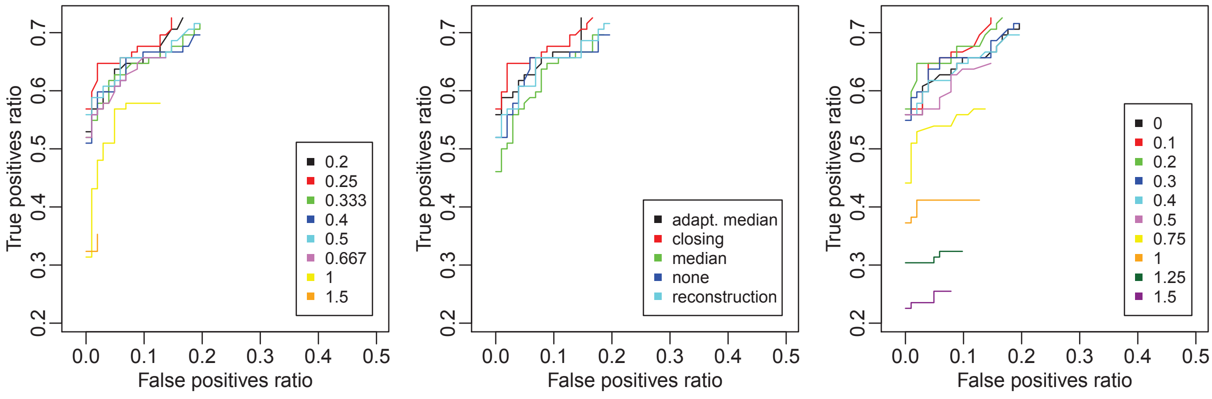
With pixel size smaller than approximately 0.5 meter, achieved detection performance is sim-



(a) (b)

Figure 7.2: Efficiency frontier for (a) plots with various heterogeneity or (b) heterogeneous plots dominated by a few large trees.

ilar for each plot. For coarser spatial resolution, detection ratio appears all the more limited as pixel size is big (figure 7.3a).
Different non-linear filters yield very similar results. However, closing by reconstruction is frequently less efficient than closing or median filters. (figure 7.3b).
The size of the Gaussian kernel for lowpass filtering greatly affects the trade-off between false and true positives ratios. For σ above 0.75 meter true positives ratio is bounded, but so does the false positives ratio. Indeed, with $\sigma = 1$ m, R_{FP} generally remains below 10%. For values below 0.5 m, no general pattern is noticeable (figure 7.3c).

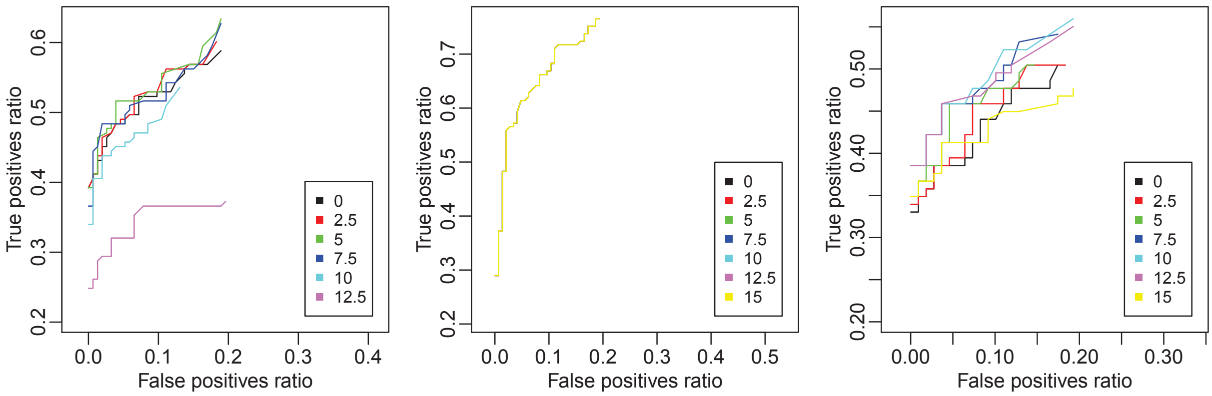


(a) *res* (m) (b) Non-linear filter (c) σ (m)

Figure 7.3: Effect of parameters on the efficiency frontier of the detection algorithm for plot chablais3.

Effect of the height threshold really depends on the forest structure. On some plots, high thresholds bound the true positives ratio. On plot chablais1, which is a rather sparse, young irregular stand, high thresholds remove small detected trees but other values have no noticeable effect (figure 7.4a). On plots with a very high and homogeneous canopy cover, thresholding the height values has non effect (figure 7.4b). In other cases, intermediate values for the threshold help removing detected shrubs and while keeping small trees (figure 7.4c).

Effect of the distance threshold displays no general trend, showing that it might be only a secondary aspect. On most plots however, selection based of values of m_{min} greater than 1.5 m result in lower detection performance.



(a) Plot chablais1 (b) Plot chablais5 (c) Plot chablais8

Figure 7.4: Effect of different values for the height threshold parameter h_{min} (m) on the efficiency frontier.

7.2.4 Sensitivity to forest structure

7.2.4.1 General considerations

Despite the low number of field plots, some trends appear when correlations between detection performance and stand parameters are investigated. Table 7.5 present the Spearman rank correlation coefficient ρ between stand parameters and the true positives ratio obtained with the best setting for each plot. True positives ratio is positively correlated to coniferous tree proportion ($p < 0.05$) but is negatively correlated with dominant height ($p < 0.05$). Correlation is significant at the $p < 0.1$ level for stem density ($\rho = 0.37$). It is noteworthy that the null hypothesis that dominant height and coniferous proportion are not correlated cannot be rejected ($p = 0.7$). Plot chablais6 appears as an outlier and, when removed from the dataset, correlations are less significant for dominant height and stem density.

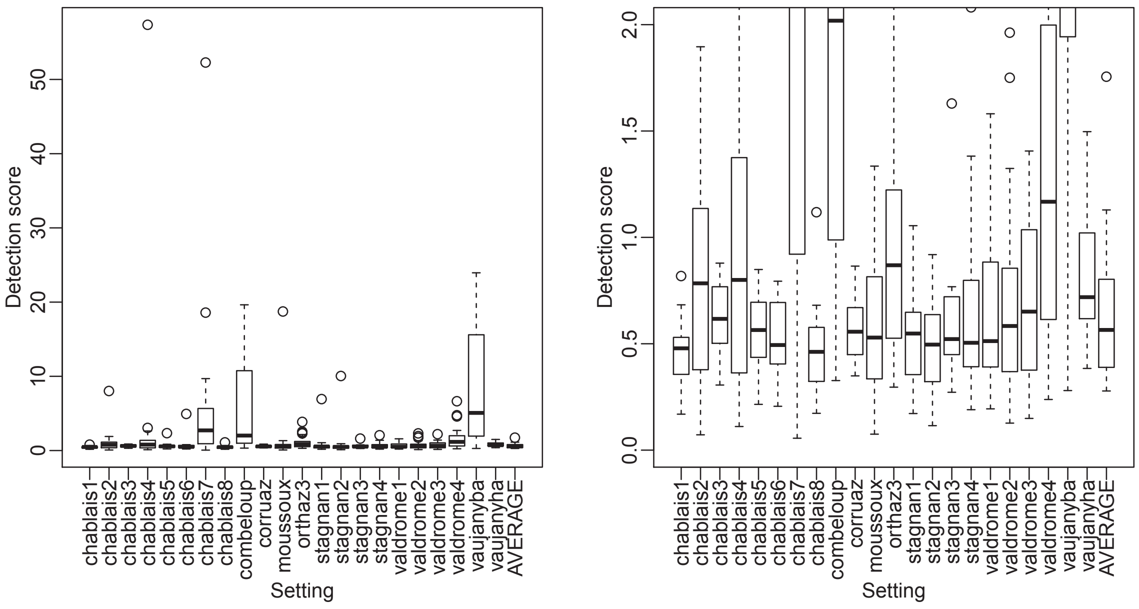
Table 7.5: Spearman correlation coefficient ρ between the true positives ratio and forest stand parameters. First row of the table displays results when all 22 plots are considered. Second row presents results for all plots except the outlier chablais6. Significance level: $^{\bullet}$ $p < 0.1$, * $p < 0.05$.

Subset	Basal area	Stem density	Dominant height	Coniferous proportion
(all)	0.07	$0.37^{\bullet}$	-0.45^{*}	0.53^{*}
(- chablais6)	0.19	0.33	$-0.37^{\bullet}$	0.53^{*}

7.2.4.2 Cross-validation of optimal settings

Figure 7.5 displays boxplots of scores obtained on all plots for the settings optimized on each plot individually, and with a global "average" setting. This setting is the parameters combination that yields the lowest mean score on all plots, i.e. $s_{min} = argmin(\sum_i S_i^j)$ with S_i^j the score obtained on plot P_i with the setting s_j. The obtained s_{min} consists in the following parameters: computation of images of spatial resolution $res = 0.4$ m, filtering with one iteration of the adaptive median filter, Gaussian smoothing with $\sigma = 0.75$ m, removal of maxima with height h under $h_{min} = 5$ m, and of maxima whose values in the maxima image do not satisfy $m \geq 0.05 \times h$ ($m_{min} = 0$ and $m_{prop} = 0.05$).

For several settings, outlier scores are present (dots far outside the whiskers in figure 7.5a). These values are mainly obtained on plot combeloup. Three settings turn out to yield very variable results: these are the optimal settings determined on plots valdrome4, chablais8 and chablais6 (large boxes in figure 7.5a). The mean score with the average setting is 0.47 with a standard deviation of 0.15. The minimum and maximal scores are obtained with the plots that also yielded the minimum and maximum scores with the optimal setting, respectively plot chablais6 and combeloup. The mean score with optimal settings is 0.30, with 0.12 standard deviation.



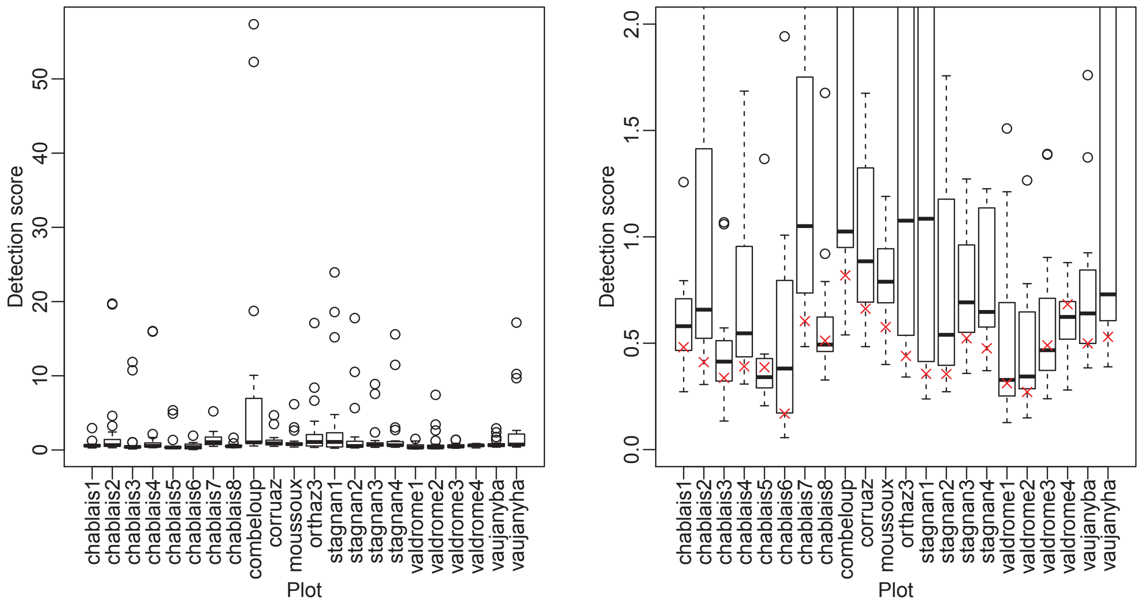
(a) Overview

(b) Zoom on scores in [0, 2]

Figure 7.5: Boxplot of detection scores obtained on all 22 plots for the settings determined on each plot individually, and for the average setting.

Figure 7.6 displays boxplots of scores obtained with all optimal settings for each plot. Detection score values are the same as in figure 7.5, except that they are grouped by tested plot instead of tested setting. Many outliers are noticeable (figure 7.6a), which means that for all plots some inappropriate settings exist. Such points corresponds to settings optimized for plots chablais6, chablais8 and valdrome4, that have shown poor performance in the previous graph. The combeloup plot gives the most variable detection results. The average setting detection score is under the first quartile of detection scores for 15 plots (figure 7.6b). It is located above the median for only four plots (chablais5, chablais8, valdrome3 and valdrome4).

Mean scores can be calculated plot-wise or setting-wise. Mean score for plot P_i is $\bar{S}_i = \frac{1}{N}\sum_j S_i^j$ and mean score for setting s_j is $\bar{S}^j = \frac{1}{N}\sum_i S_i^j$. Figure 7.7 displays the scatterplot of $(\bar{S}^i, \bar{S}_i)$, i.e. the mean detection score obtained with all settings on plot i as a function of the mean score obtained on all plots with the setting determined on plot i. The mean value of detection scores for plot combeloup, and to a lesser extent, plot stagnan1 are high (7.9 and 3.7 respectively). However, the settings determined on these plots are quite robust as they yield mean detection scores on all plots of 0.57 and 0.93 respectively. On the contrary, settings determined on plots valdrome4, chablais8 and chablais6 give poor results with mean detection scores of 8.1, 6.0 and 5.9, but scores obtained on these plots with other settings are satisfactorily with mean detection scores of 0.62, 0.59 and 0.52. Plot chablais3 lies in a intermediate position compared to the two aforementioned groups. The remaining points are gathered in an area bounded by $\bar{S}_i < 3$ and $\bar{S}^i < 2$. Spearman rank correlation ρ between $(\bar{S}_i)$ and $(\bar{S}^i)$ is -0.63 ($p < 0.01$) if all points are considered, and -0.56 ($p < 0.05$) if the six points mentioned previously are removed. This



(a) Overview (b) Zoom on scores in $[0, 2]$

Figure 7.6: Boxplot of detection scores obtained with all settings for each plot. Scores obtained with the average setting are displayed as × red crosses.

trend shows that plots where trees are not easily detected yield the most robust parametrization when used for algorithm training.

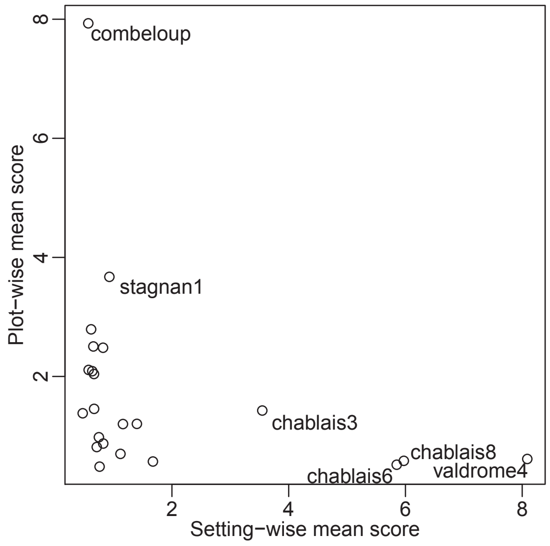


Figure 7.7: Scatterplot of $(\bar{S}^i, \bar{S}_i)$.

7.3 Discussion

7.3.1 Algorithm

7.3.1.1 Local maxima extraction

The local maxima extraction algorithm consists of numerous steps. For each one, several parameters have been tested, leading to a important number of tested combinations. However, results indicate that the whole process could be somewhat simplified. Besides, some slight modifications can be performed in order to make comparisons between parameter sets more meaningful.

Regarding resolution, it appears that even when LiDAR data is theoretically not dense enough to compute images with small pixel size, good detection performance are obtained. This might be linked with the fact that non-linear filters are successful in filling empty pixels. Then the position of the treetop is estimated with better precision for images with finer spatial resolution. This leads to a marginally better detection performance with the matching procedure based on distances between the assumed treetop and the position of the local maximum. The information originally contained in the point cloud is indeed better retained in finer grids. However, the computational burden for processing such images is higher. If similar performance is obtained with slightly coarser resolution, e.g. 0.4 or 0.5 m, working with such pixel size might be more efficient in operational applications. Moreover, if delineation of trees – which demands even more CPU – is performed after maxima selection, an alternative for better estimation of tree position and height would be to select the position of the highest LiDAR echo in the LiDAR point cloud contained in the segment.

Regarding non-linear filters, it appears that filters that are the less conservative with shapes (median and closing) yield better results that closing by reconstruction. This shows that in the process of detecting treetops, fine details of tree crowns disturb the detection of the global shape of trees. The case of the adaptive median filter is quite different, as it is quite conservative with crown details but also adjusts the window size to the estimated noise level. This might explain why it produces good results. However, for a more truthful comparison of non-linear filters, similar window shapes should be used. Indeed, square windows are used for the filters based on median, whereas disk structuring elements are used for closing filters. Given the shape of trees, circular windows might be a better choice. This remark also stands for the window used for local maxima detection. However this is likely to give only minor differences [258].

The additional selection based on minimum height proved useful, mainly in sparse stands. In fact, detected maxima might really correspond to small trees or bushes. But as these plants are not considered in the field inventory, or are not interesting to the forest manager, it is better to remove them in order not to bias the detection performance or stand parameters estimates (e.g. number of stems above a certain DBH).

Selection based on the relationship between the value in the maxima image m, which corresponds to the size of the window where the pixel is local maximum, and pixel height is more difficult to interpret. Using criteria based on relationships between tree height and crown width [43, 258] seems more meaningful but requires prior knowledge of allometric relationships and might not be able to handle the cases of trees with overlapping crowns. A selection based only on the value in the maxima image (which occurs in fourteen of the best settings) might be sufficient. Besides, this selection step might be considered redundant with the preceding filters (non-linear and low-pass). To address the issue of selection of maxima depending on their height, without using a complementary selection, height-dependent filters could be used. Indeed, it is possible to adapt the size of the structuring element or of the smoothing linear kernel to the value of pixel in the height image [156]. However it is likely that for optimal performance, the stand structure will have to be taken into account to adjust the height-dependent window size.

Finally it turns out that a better understanding of both tree crowns interaction in complex forests and of LiDAR crown models is required to design improved local maxima extraction procedures. Procedures could be simplified for better robustness, or adapted to various forest structures for better handling of particular cases.

One example is that the criterion based on the value in the maxima image is not adapted to tree crowns that are not symmetric. For example, in Valdrôme, most of the crowns are flag-shaped (figure 7.8). As trees have the same age, an upslope tree has a competitive advantage for light with respect to its neighboring downslope tree. As a result, trees tend to bend over trees located downslope, resulting in an asymmetric tree crown. For a given local maximum detected in the DSM, the size of the window where it is maximum is very small and do not reflect its crown size. An alternative it to use a relationship between the size of the segmented tree crown and maximum height as selection criterion, but this requires an iterative approach.

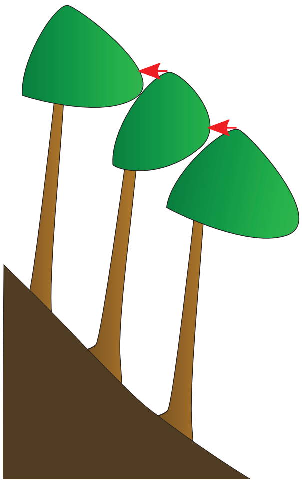


Figure 7.8: Distance of treetop to the nearest higher crown. In this case of asymmetric, flag-shaped crowns, this distance (red arrows) is not a good indicator of crown width.

7.3.1.2 Matching procedure

The final outcome of the workflow implemented previously heavily depends on the automated evaluation of detection performance. However, it is obvious that numerous aspects might lead to errors in the process. The first one is ground truth. Indeed, there are several types of error during field inventories [69]:

- measure errors due to instruments,
- omission or double inventory of trees,
- measure and reporting errors due to operators.

Errors linked with instruments have already been presented in section 5.2. They are smaller than the errors linked with trees tilt and geolocation, and are accounted for in the matching buffer. Trees are rarely inventoried twice as they are usually numbered with chalk on their stem. Some errors linked with erroneous reading or reporting of instruments measures might be detected with careful checking of field data, e.g. values range, map of tree positions, height to diameter relationship. Outliers might be due to false measures and should be checked by going back to the field, which is not always possible. Besides some errors might remain unnoticed, and lead to a wrong tree position or height. In such cases, the treetop with wrong position is likely to be considered as a false negative. If the treetop is actually detected in its right position, a false positive is also considered, instead of a true positive. Similarly, omitted trees will lead to no result or to a false negatives. However, experience with plots that have been inventoried twice show that such errors are rare, and mostly restrained to small trees that are generally not detected.

Second type of error is linked with the unknown position of the treetop. Indeed, it is the tree stem position that is recorded, with precision directly linked to GPS data. However, when a tree is tilted, the treetop projection on the ground might be located several meters away from the stem. That is why the buffer approach is adopted, which allows to look for a selected maximum inside a sphere centered on the assumed treetop position. The choice of the sphere size (proportional to tree height) is designed to reach a trade-off between the possibility to refuse a true positive (tree that is very tilted, with its treetop outside the buffer), and the possibility to create an erroneous true positive by linking another nearby selected maximum to this tilted tree. The setting that is adopted has been chosen with data from Vaujany, where trees are very tilted toward downslope. Thus one can expect that very few true positives will actually be rejected, but that matching with nearby maxima might occur. In such cases, either another true positive was linked with the wrong treetop, or a false positive is considered as a true positive. A comparison with results from external, manual matching for a subset of plots showed that such cases are also rare.

The third type of error is due to border effects. To be consistent with the buffer approach procedure, the plot boundaries are not calculated from the ground delineations, but from a 2D buffer created around each field tree stem. When a tree has its stem outside the plot, but its treetop inside the buffer (e.g. when a nearby big tree it tilted over a small tree located inside the plot), it is likely to be detected as a false positive, whereas it should be a true positive. The two trees detected as the black squares near matrix coordinates (35, 95) in figure 6.14c are examples of how border effects lead to wrongly considered false positives. There might be between zero and two false positives due to border effects in the tested plots. To mitigate this problem, the construction of a 3D (grayscale) buffer might help removing false positives.

The fourth type of error is linked with the greedy nature of the matching algorithm. Indeed, pairs are successively created based on the lowest matching index, and never changed, whereas breaking one pair might make the creation of two other pairs possible (figure 7.9). To address this issue, an optimization procedure based on simulated annealing [154] has been tested. However, the choice of the energy function turned out to be complicated. Indeed, taking only the created pairs as criterion leads to the matching of false treetops with nearby undetected field trees, given the permissive size of the matching buffer. Taking the quality of matching into account with the integration of a parameter linked to the distances between maxima and treetops in constituted pairs was also tricky. In the end, the greedy matching algorithm is preferred as it is more simple and will tend to underestimate detection performance, whereas the optimized procedure tended to overestimate it. Besides, cases where the breaking of one pair would allow the creation of two other pairs are also rare.

In order to be able to design automated matching procedures for algorithm testing, one has to rely on simple assumptions. It is obvious that the adopted procedure could be refined, but taking into account all particular cases occurring in complex forests would be an endless task, and might involve challenging the field protocol. For example, snags (dead standing trees) are not considered in the analysis, meanwhile trees that recently died and still have their small branches will induce in LiDAR data a response similar to live trees. Besides, in the field inventory, all individual stems at 1.3 m height are considered as trees. When coniferous trees have forks at higher levels, this might lead to two separate crowns for one single tree. If two treetops are detected, one will be considered false positive according to ground truth. Despite its limitations, this matching procedure meets the requirement of comparing detection results quickly and in a

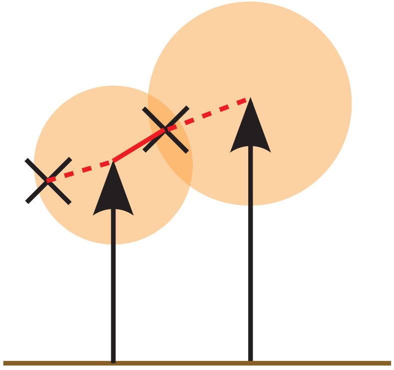


Figure 7.9: Example of bad matching by the greedy algorithm. One of the field treetops (black arrows) is linked (– red solid line) to the nearest detected maximum (× black cross). Untying them will make matching of both trees with detected maxima possible (··· red dotted lines). Orange shades represent the matching buffer limits.

stable way, which makes further analysis of detection performance possible.

For final detection performance comparison, a unified criterion is required. Here a trade-off between the true positives and false positives ratios is proposed. It has been arbitrarily chosen in order to limit false positives while allowing satisfactorily true positives ratio. For real applications it is obvious that the criterion has to be adapted. The weight factor should be adjusted to take the relative cost of false and true positives into account. Moreover, the number of trees might also be weighted by a factor that reflects their value, such as tree height, in the absence of reliable estimate of the basal area or volume. Such criteria depend on the users of remote-sensed data and require further cost plus loss analyses, such as the study for evaluating the economical opportunity of LiDAR inventories [124].

7.3.2 Detection performance

7.3.2.1 Detected trees

Comparison with detection results from the literature must be taken with care as there are as many performance assessment procedures as published results. Main methods match detected maxima based on the inclusion within field delineated crowns, or reversely, based on field trees included in segmented tree crowns. In most cases, it is a 2D matching process, even tough a height difference threshold is sometimes considered. The procedure based on the 3D distance takes advantage of the fact that all tree heights have been measured. It is more conservative in this regard, but it is more flexible regarding the position of the treetop. Moreover, forest structures are difficult to compare and LiDAR data is acquired with different parameters or sensors. In sugi (*Cryptomeria japonica*) plantations in mountainous terrain, detection results (R_{TP}, R_{FP}) of $(0.74, 0.03)$, $(0.86, 0.1)$ and $(0.92, 0.08)$ have been reported for steep slope, gentle slope and gentle slope with rough terrain respectively [113]. These results are better than those obtained in plots chablais5 and chablais6. Besides, sugi plantations were more dense than those two plots.
In richly structured forests in Bavaria [104], the global detection results for 29 plots were $(R_{TP}, R_{FP}) = (45.4, 5.5)$ for all trees $(n = 2584)$. Of our 3378 trees in the 22 plots, 51.5% are detected with settings optimized for each plot, but only 34.7% if the average setting is used. Corresponding false positives ratios are respectively 3.8 and 3.6%. The Bavarian study obtained similar results regarding the facts that detected trees are mainly from the upper stratum and that detection performance is lower for stands with high stem density or deciduous proportion.

7.3.2.2 Parameter estimation

Height It is noteworthy that, compared to the 2D based matching procedure, the 3D approach used here has the indirect effect of preventing large relative differences between the measured and LiDAR heights of a correctly detected tree. Possible outliers are de facto considered as false positives.

Underestimation of tree heights in coniferous stands has been commonly reported [88]. It is explained by the small probability for a laser pulse to actually hit the treetop, or at least to induce sufficient backward signal to accurately measure its position. When treetops are flat-shaped, such as for broadleaved species or for ageing sliver fir trees, the underestimation is expected to be less important. Here, a large value is obtained for plot chablais6. The terminal shoot and the last whorl of young spruce trees are indeed very thin which might explain why their height is not correctly reported by laser returns. All plots dominated by spruce display an underestimation of tree height. Height underestimation in the Chamonix plots might also be caused by the gap between the LiDAR acquisition (2008) and the measures (2009 and 2010). Plots dominated by silver fir or beech show no or small overestimation. In the Bavarian study a mean difference of 0.54 m, with a standard deviation of 1.44 m is reported for the difference between measured and LiDAR height. These values are also similar to the ones presented in table 7.3. In some plots (Saint-Agnan 1 and 2, vaujanyba), an overestimation is recorded. All these plots are protection forests located in areas with altitude greater than 1200 m and slope greater than 30°. Environmental factors lead to a tendency of trees to be tilted toward downslope. This results in tree height overestimation by LiDAR, as reported in [113], and also in an underestimation of field height (figure 7.10).

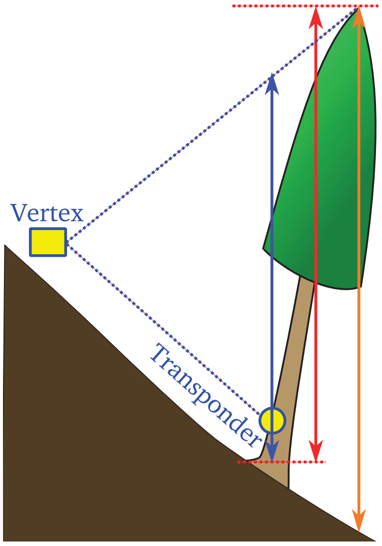


Figure 7.10: In slope areas, it is usually more convenient to measure tree height (red arrow) from an upslope position. Unfortunately, if the tree is tilted downslope, height is underestimated (blue arrow), whereas LiDAR data overestimates it (orange arrow).

Basal area Other studies mainly focused on the retrieval of diameter at breast height, which is proportional to the square root of basal area. The Bavarian study [104] fitted regression models with LiDAR height, crown radius and their squared values as independent variables. adj-R^2 of the models were 0.79 for deciduous trees and above 0.89 for coniferous trees. These results are clearly better, which could be explained by different factors. First the investigated relationship (i.e. diameter versus height, crown radius and their squared counterparts with an additive regression model) has a better fit. Second there might be important differences between broadleaved and coniferous species which makes the use of two separate models more relevant. Third, the watershed algorithm used for crown segmentation is not constrained by segment shape or pixel height. Forest gaps will be integrated in the nearby trees, which erroneously increases their crown area. Besides, in multi-layered forest, small trees are not detected but have crowns that partially join crowns of larger, detected trees, which also increases the segment size.

7.3.3 Algorithm sensitivity

7.3.3.1 General considerations

Results show that for homogeneous, single-layer stands, optimal parametrization leads to better performance than for complex stands. Indeed, trees have similar height, width and shape, so that with an adapted setting treetops can be efficiently detected in every part of the plot while limiting the amount of false positives. On the contrary, in complex plots, the variability of shapes and tree arrangements makes optimization of detection parameters at the plot scale more complicated. The trade-off between false and true positives leads to a setting that is more robust but necessarily less efficient from the score criterion. In such stands, true positives ratio can be increased if more false positives are tolerated. In homogeneous stands, gaining more true positives is more difficult as deviations from the optimal parameters set lead to a lot of false positives with relatively few additional true positives.

Two complementary aspects may explain why the true positives ratio is positively correlated with coniferous proportion. First it has been reported in the literature that broadleaved trees are less easily detected due to the irregular shape of their crowns. A single tree may have two or more clearly separated local maxima, whereas in coniferous trees the apical dominance is highly pronounced, and shape more regular, which makes algorithm parametrization easier. Second, there might be a link between forest structures and coniferous proportion in our studied area. Homogeneous stands are frequently coniferous plantations (e.g. chablais5, chablais6, Valdrôme plots) whereas deciduous species in mountainous forests have often been considered less valuable and restrained to the lower strata of multi-layered, complex stands.

The negative correlation of true positives ratio with dominant height might also be linked with forest structure. Indeed, studied plots with high values of dominant height are mainly multi-layered stands with a dominant stratum of large mature trees. Thus, dominant height value is not related to an homogeneous cover of large trees but to an upper stratum that overwhelms smaller trees, making them hard to detect on the DSM.

7.3.3.2 A few typical examples

Homogeneous stands chablais6 is the typical example of an homogeneous stand. In this young spruce plantation, recently thinned, trees have the same height and crown size (figure C.5e). Crowns are also well delimited. The best true positives ratio is obtained on this plot with 80.5%. False negatives are mainly due to border effects. The ROC curve (figure 7.2a), is square-shaped, which means that once the majority of trees are correctly detected, additional true positives require a important trade-off with false positives. The optimal parameter setting is very specific to this forest structure, and yields poor results on other plots. chablais5 is also a very homogeneous coniferous stand, but with higher canopy closure, so that achieved true positives ratio is only 55.9%.

Dense stands valdrome3, valdrome4 and moussoux plots have stem densities higher than 1000 ha^{-1}. In the Valdrôme plots, which are one-layered stands, true positives ratio of 58% are obtained, whereas in the irregular stand of moussoux, the value is only 44.5%. In both cases, the high stem density leads to a pretty high number of false positives (> 5.3%). In the moussoux plot, the bad detection score might also be explained by a lower LiDAR echo density. The optimal setting obtained on plot valdrome4, which has $N = 1600$ ha^{-1}, performs poorly on plots with lower stem density.

Broadleaved stands Plots combeloup, chablais7 and chablais8 have more than 80% broadleaved trees proportion. Bad detection score in combeloup is mainly due to the poor quality of the LiDAR data. Acquisition was realised in leaf-off conditions, which results in a low proportion of vegetation echoes. The resulting DSM is of poor quality (figure C.1b), it is likely that echo extraction from the full waveform files was too restrictive. This shows that the choice of LiDAR acquisition parameters and the quality of pre-processing is crucial regarding the potential of final ALS data for forest parameters retrieval. In chablais7, the stand is complex as several broadleaved trees are actually coppiced, and terrain surface is very rough. This may result in additional errors in both LiDAR heights and field data. In plot chablais8, stand structure is more homogeneous and a better detection performance is obtained, with 45.9% detection ratio.

7.4 Conclusion and perspectives

Whereas the estimation of tree height from LiDAR data is straightforward, retrieval of other parameters is more complicated. Fitting correct models requires additional steps, such as: forest structure stratification or species identification in order to apply specific models, extraction of other relevant LiDAR metrics and testing new models.

With complex forest stands, parametrization has a major effect on detection performance. Satisfactorily results can be obtained in any forest stand with an average setting. However, this will greatly limit the detection performance for homogeneous stands where trees are theoretically easily detected with an appropriate setting. When comparing the performance of settings optimized for a given plot when used on the other reference plots, three main points are remarkable:

- homogeneous plots require a specific tuning of the algorithm,
- detection on complex stands requires an important trade-off between true and false positives,
- forest structure and species have a major influence on detection performance.

In operational applications, field data is usually rare, and manual delineation and parametrization of the algorithm for small areas is not possible. An unsupervised alternative would be to take advantage of available reference data, i.e. field plots with LiDAR data where the algorithm parametrization are known, to automatically adjust detection parameters based on the similarity between the new LiDAR data and reference plots. Indeed, for each plot, there is a setting optimized for another plot that performs better than the average setting. This approach is presented in the following chapter.

Chapter 8

Unsupervised algorithm parametrization

8.1 Method

8.1.1 Introduction

The simple local maxima detection algorithm presented in the previous chapter yields good detection performance provided it is appropriately parametrized. Results obtained with an average set of parameters are generally satisfactorily but are bonded for homogeneous cases where trees are theoretically easier to detect. To deal with this sensitivity of algorithms to forest structure, various approaches can be adopted. Parameters are generally based on prior knowledge about forest structures, e.g. with crown diameter – height relationships [43, 258]. Some algorithms automatically adjust the parameters in the course of the detection procedure, e.g. with multi-scale approaches [245, 252]. From the results of the previous chapter, it is noticeable that even tough the optimal parameters set obtained with a given plot generally performs poorly when applied on the others, there is frequently one or several of them that still performs better than the average parametrization. When no information is available on a forest area covered by a new LiDAR survey, it is not possible to rely on previous field knowledge of allometric relationships. The proposed idea is to use the existing coregistered field and LiDAR data as training data in a three step procedure. The training step, i.e. the determination of the best parameters setting among all combinations for each training plot is first performed, as has been presented in the previous chapter. Then the new data is compared to existing LiDAR datasets. Finally, the optimal parameters set of the most similar reference dataset is used for treetop detection in the new data.

To implement this workflow, a matching procedure based on LiDAR data is required. Due to the complexity of comparing two 3D point clouds, two simple metrics are investigated. The first one is the height distribution of points within the plot. It reflects the vertical structure of the forest stand, which is alleged to be one important criterion regarding the potential for image processing techniques to detect trees. The second one is an integrated spectral-energy description of the canopy heigth model. It characterizes the spectral content of the surface model which is the raw material for image processing techniques used for local maxima extraction.

8.1.2 Point cloud metrics computation

8.1.2.1 Point cloud extraction

In the previous chapter, the single-tree based procedure did not require that field plots have similar shape. It was possible to finely adjust the plot mask to inventoried trees in order to minimize border effects. Here, the objective is to compare the point cloud in a given area to the point clouds located in reference field plots. In order to be able to extract stand metrics based on Fourier-transform techniques, square-shaped plots are preferred. Unfortunately, the real field plot shape is never a perfect square, and its corner coordinates are not known with accurate precision. Therefore, the proposed procedure is to automatically adjust a 50 m wide square to the inventoried trees positions and to extract the included point cloud for further processing.

We consider a square plot were N trees have been inventoried. Trees planimetric coordinates in the current projected system are $(x_i, y_i)_{i\in\{1,\cdots,N\}}$. A rotation angle that approximately aligns the plot with coordinate grids is estimated by finding the argument α_R that minimizes the range of the rotated coordinates $(X_i, Y_i)_{i\in\{1,\cdots,N\}}$:

$$\begin{pmatrix} X_1 & Y_1 \\ \vdots & \vdots \\ X_N & Y_N \end{pmatrix} = \begin{pmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_N & y_N \end{pmatrix} \times \begin{pmatrix} cos(\alpha) & sin(\alpha) \\ -sin(\alpha) & cos(\alpha) \end{pmatrix} \tag{8.1}$$

$$\alpha_R = \underset{\alpha}{argmin}(min(max(X) - min(X), max(Y) - min(Y))) \tag{8.2}$$

α_R is estimated by applying a built-in optimization function of the R software based on simulated annealing [154]. The rotation is applied to the LiDAR point cloud, as well as a translation of vector $\frac{1}{2}(max(X) - min(X), max(Y) - min(Y)$. Considered LiDAR points are those located in the quadrant of width 50 meters and centered on the new origin.

8.1.2.2 Height distribution

The vertical distribution of LiDAR points is related to the vertical structure of the forest. The underlying hypothesis is that detection performance in complex stands is linked to the presence of multiple layers in the canopy cover. The height distribution of LiDAR points is expected to reflect this vertical organization.

The normalized heights of selected LiDAR points are computed by subtracting the terrain altitude at their orthometric coordinates. Terrain surface is estimated by bilinear interpolation of points classified as ground points. The height distribution is computed as the frequency of points that fall in height bins of 0.5 m width, starting from two meters height. Lower points are not considered in order to remove the influence of ground, shrubs and bushes. For each plot, the height distribution H is a vector of 83 values that correspond to the frequency of points in 83 height bins of 0.5 m width from $[2, 2.5[$ to $[43, 43.5[$.

8.1.2.3 Module of the Fourier Transform of image

The raw data for the treetop detection is the the DSM. In section 6.2.4, the terrain and objects in the DSM are interpreted in terms of frequency components. It is expected that the optimal detection parameters depend on the arrangement and shape of trees, and on LiDAR acquisition parameters. Forest structure and LiDAR parameters are related to the relative importance of frequency contents of the DSM, which can be regarded as global texture patterns. The Fourier spectrum is a widely-used tool for describing the frequency contents of images [92, p. 838]. Considering an image f of size (M, N), the 2D discrete Fourier transform (DFT) F and the corresponding spectrum S are given by:

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)e^{-2j\pi(ux/M+vy/N)} \qquad \text{with } j^2=-1 \tag{8.3}$$

$$S(u,v) = |F(u,v)| \tag{8.4}$$

In order to avoid border effects, the DSM surface is calculated for the point cloud extracted according to the procedure detailed in section 8.1.2.1. The DTM model of the area is subtracted to the DSM in order to obtain the CHM, where the contribution of the terrain variations to the frequency content is removed. The DFT is performed on the CHM. Simplifying the interpretation of the obtained 2D spectrum is possible by converting the spectrum $S(u,v)$ in polar coordinates $S(r,\theta)$. Indeed, in forest areas the texture patterns are assumed to be isotropic. Besides, the directionality of patterns is of little importance for local maxima extraction. The function of spectrum versus radius is obtained by integrating over the angle values in the range $[0,\pi]$, as the spectrum if symmetric around the origin for images with values in $\mathbf{R}$:

$$S(r) = \sum_{\theta=0}^{\pi} S_\theta(r) \tag{8.5}$$

In practice, (u,v) values are regularly spaced, but the values in the corresponding radii vector (r) are not. S values at regular intervals (r_i) are estimated by averaging the values of $S(r)$ with a Gaussian kernel:

$$S(r_i) = \frac{\sum a_i S(r)}{\sum a_i} \qquad \text{with} \qquad a_i = exp\left(-\frac{(r-r_i)^2}{2\sigma^2}\right) \tag{8.6}$$

The spatial resolution for the CHM is 0.5 meter. Its dimensions in pixels are 100 × 100. The frequency resolution of the DFT is $\frac{1}{50}$ m-1. S is estimated for values $r_i = \frac{i}{50}$ with $r_i \in \left[0, \sqrt{2}\right[$ (figure 8.1c). The upper boundary corresponds to the maximum value of the radius (frequency) along the diagonal of the spectrum image. σ is set to $\frac{1}{50}$ m-1.

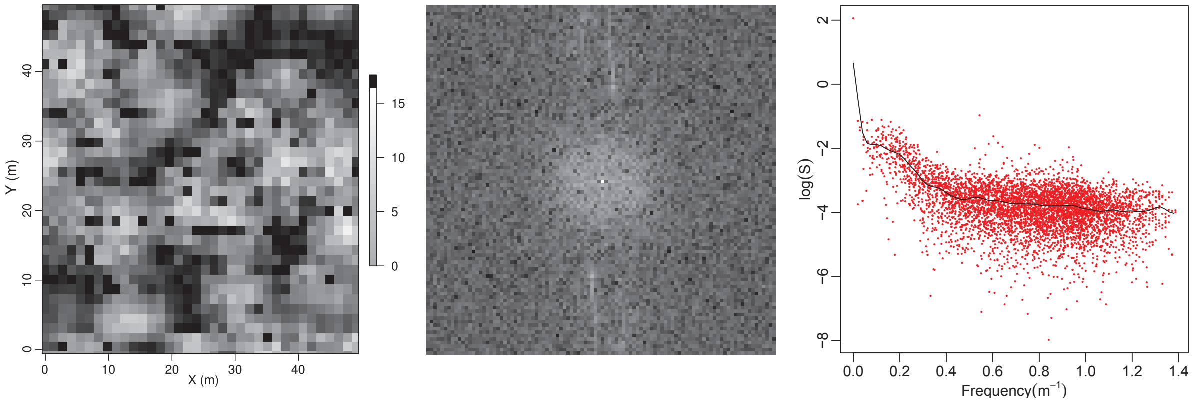
The steps for the estimation of $S(r_i)$ are exemplified with plot chablais6 in figure 8.1. The CHM at 0.5 m spatial resolution displays horizontal black lines. The aircraft trajectory was probably oriented from the bottom to the top of the image. Flight speed was slightly too fast which results in a pulse spacing wider along than across the platform track. With a spatial resolution of 0.5 m lines of empty pixels appear on the CHM. These horizontal, almost periodic patterns induce the vertical lines of energy in the spectrum (figure 8.1b). Otherwise, the energy is mainly concentrated in the low frequencies. In figure 8.1c, values of $S(r)$ appear as the red dots. They are highly scattered, but the main features from the spectrum are present in the averaged curve. High values are present from $f = 0$ to 0.3 m-1. To avoid the major influence of the mean value, the first coefficient, which corresponds to $f = 0$ is not considered in the following steps. For each plot, the final spectrum S has 69 values that correspond to frequencies from 0.02 to 1.38 m-1 with 0.02 increment.

8.1.3 Nearest neighbor selection

For nearest neighbor selection, the Euclidean metric is used to compute the distance between two plots P^a and P^b, whose coordinates are $(p_i)_{i\in\{1,\cdots,n\}}$:

$$d(P^a,P^b) = \sqrt{\sum_{i=1}^{n}(p_i^a - p_i^b)^2} \tag{8.7}$$

Here, (p_i) in a given plot corresponds either to the height distribution H or to the spectrum S.



(a) Canopy height model $f(x, y)$

(b) 2D spectrum $S(u, v)$ (log scale)

(c) 1D spectrum $S(r)$ and $S(r_i)$

Figure 8.1: Computation of the spectrum for plot chablais6. (a) CHM at 0.5 m resolution. (b) 2D spectrum obtained by DFT. (c) 1D spectrum: red dots are all the points $S(r)$, black line is the $S(r_i)$ average curve.

8.1.4 Validation procedure

The validation procedure aims at comparing the detection performance of various possibilities for algorithm parametrization. We consider a validation plot P^v and a set T of n training plots $(P^i)_{i\in\{1,\cdots,n\}}$.
The first possibility to parametrize the algorithm is to chose the setting s^T_{min} that yields the lowest mean score on all training plots. This average setting is determined as $s^T_{min} = argmin(\sum_i S^j_i, j \in A)$ with S^j_i the score obtained on plot with the setting s_j. The set A of 111720 combinations tested in the previous chapter is considered for $(s_j)_{i\in A}$. The detection performance score obtained by using s^T_{min} on the validation plot P^v is numbered $S_{average}$.
The second possibility is to apply the optimized setting obtained on a training plot that best resembles the validation plot. The best-match training plot is the one with the minimum Euclidean distance to the validation plot for the height distribution or the spectrum. The score obtained by detecting trees on P^v with the optimal setting of the plot with lowest distance for the height distribution (resp. spectrum) criterion is numbered S_{H-opt} (resp. S_{S-opt}).
For comparison purposes, the lowest score obtained on P^v with the optimal setting (T_{opt}) of the training plots is also recorded. The setting is chosen as $s^T_{opt} = argmin(\sum_i S^j_i, j \in T_{opt})$, and the resulting score on P^v is S_{opt}. The score obtained with a randomly selected setting in T_{opt} is also recorded as S_{random}.

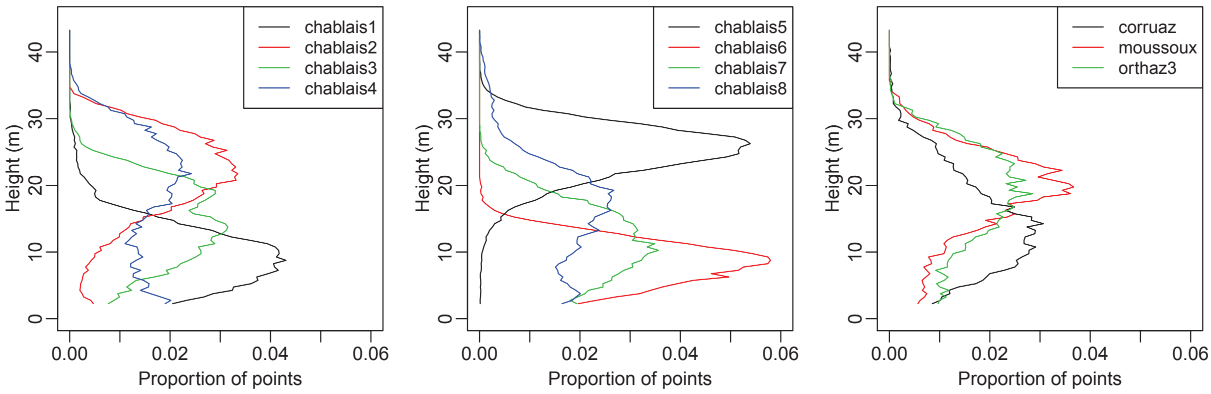
To assess the influence of the number of training plots on the detection performance of each of these parametrization alternatives, the following procedure is implemented. A validation plot is sampled from the 22 reference plots. n training plots are sampled from the 21 remaining plots. Values of $S_{average}$, S_{H-opt}, S_{S-opt} and S_{opt} obtained with the proposed workflow are recorded. The computation is repeated 10^4 times for each value of n in $\{5, 6, \cdots, 21\}$.

8.2 Results

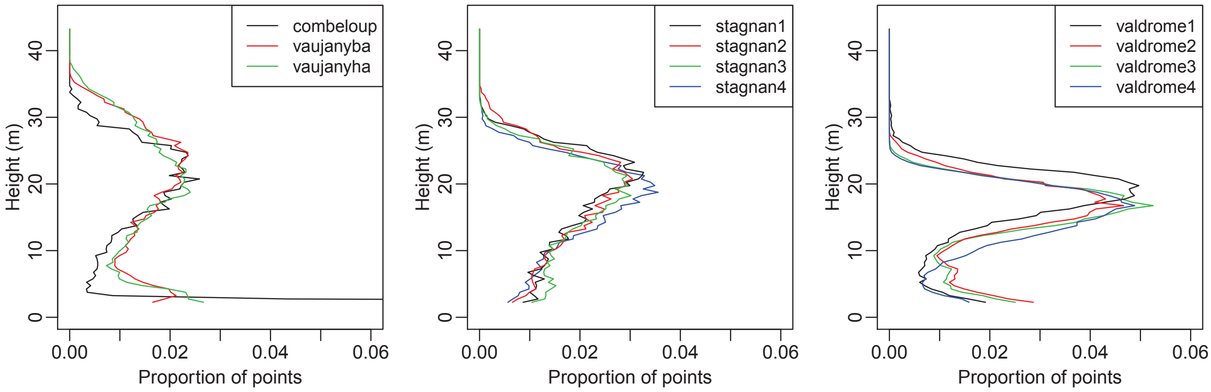
8.2.1 Metrics

8.2.1.1 Height distribution

Figure 8.2 shows the height distribution of LiDAR points of the reference plots. It is noteworthy that height distribution of points is quite similar to height distribution of trees in corresponding plots (appendix C). For example, the homogeneous canopy cover in the single-layered plots chablais5 and chablais6 appears as sharp-peaked distributions in both cases (figure 8.2b). The existence of two layers in plot chablais4 could explain the presence of two modes in the points height distribution (figure 8.2a). Plot corruaz has more small trees than the other Chamonix plots, which is reflected by the higher proportion of low LiDAR points (figure 8.2c). The differences in the height distributions of the two plots of Vaujany and of the four plots of Saint-Agnan are hardly noticeable (figures 8.2d and 8.2e). The distributions in the Valdrôme plots reflect the fact that all plots exhibit a single-layer, homogeneous structure, but with slightly different dominant heights. They are all peak-shaped, but shifted along the height axis (figure 8.2f).
For low heights, the correspondence is not so straightforward, as small trees and bushes are not inventoried. The very closed canopy in plot chablais5 results in the absence of points near the ground, whereas when canopy is less dense or presents gaps, low points amount to a larger proportion. The combeloup plot is a extreme example, as points between 2 and 2.5 meters high represent almost one fourth of the total. This might be explained by the fact that the LiDAR survey was performed in leaf-off conditions, or by a bad pre-processing of the point cloud.



(a) Chablais plots 1 to 4 (b) Chablais plots 5 to 8 (c) Chamonix plots

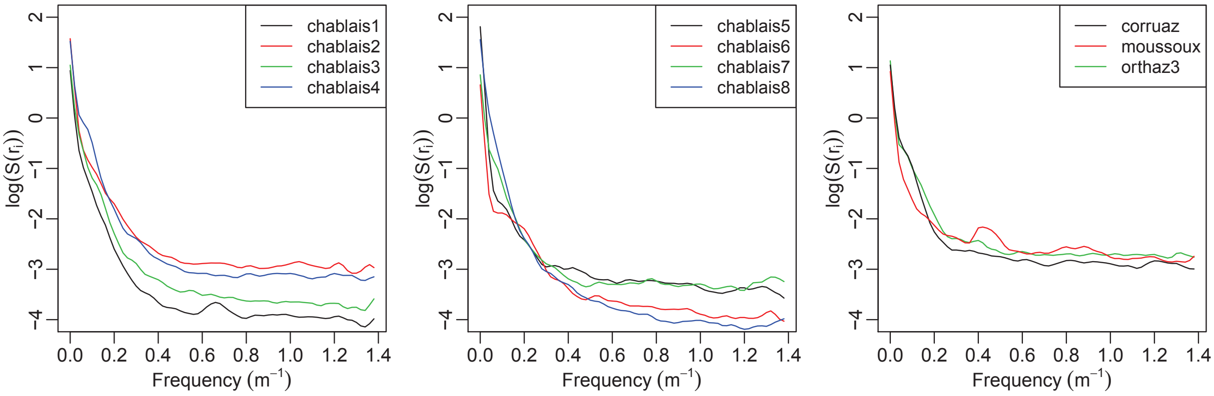


(d) Combeloup and Vaujany plots (e) Saint-Agnan plots (f) Valdrôme plots

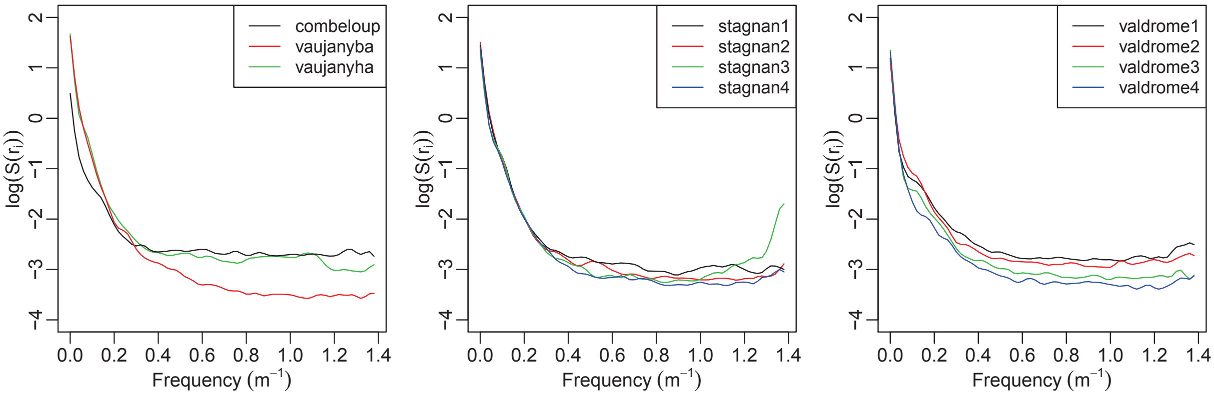
Figure 8.2: Distribution of point cloud heights in the reference plots.

8.2.1.2 Canopy height model spectra

Figure 8.3 displays the 1D spectra $S(r_i)$ of the reference plots. Frequencies that correspond to tree envelopes are approximately in the range $[0.05, 0.5]$ m^{-1} (period in $[2, 20]$ meters). Plot chablais4 has an homogeneous upper layer of large trees which might explain why its spectrum has the largest values for frequencies between 0.06 and 0.16 m^{-1} (figure 8.3a). Plot chablais6 has an homogeneous canopy of smaller trees, and the spectrum displays a table-like area between frequencies 0.06 and 0.2 m^{-1}, which might correspond to plantation spacing (figure 8.3b). Some remarkable peaks occur in the spectra of moussoux and chablais1 plots at frequencies 0.42 and 0.66 respectively. This indicates periodic features that might be linked with aligned strips of void pixels in the CHM caused by the scanning pattern. These lines are visible in the 0.5 m CHM of plot moussoux (figure C.2e), but are barely noticeable for the CHM of chablais1 as their value in the spectrum is more than ten times smaller (figure C.4b). The void pixels might also be responsible for the high energy values in the very high frequencies of the spectrum of stagnan4. In many areas, the 0.5 m CHM resembles a chess board at pixel level. Field plot are usually aligned on contour lines in order to ease fieldwork. For the same reason, ALS survey flight paths are aligned on contour lines. When the CHM is rotated to align the plot with image grid, the scanning pattern is also aligned with the image grid, which enhances the resulting noise in the CHM, as whole horizontal lines remain without information.



(a) Chablais plots 1 to 4 (b) Chablais plots 5 to 8 (c) Chamonix plots



(d) Combeloup and Vaujany plots (e) Saint-Agnan plots (f) Valdrôme plots

Figure 8.3: 1D-spectra $S(r_i)$ the reference plots. Spectrum coefficients on the y-axis are log-scaled. x-axis is labelled in absolute frequencies (m^{-1}).

8.2.2 Unsupervised parametrization

8.2.2.1 Influence of training set size

Figure 8.4 displays the mean and standard deviation of score values obtained with the 10^4 repetitions. Curves show that random selection of one of the settings from the training plots results in an average score of 1.8, with a standard deviation of around 4.6. The theoretical score obtained by the selection of the best setting in the available training plots yields mean values lower that 0.46 with 0.20 standard deviation. These values decrease slightly when training set size is increased and reach 0.38 ± 0.16 for $n = 21$. The mean score obtained with the average setting determined on the training plots starts from 0.73 and seems to decrease until $n > 10$ where it remains quite stable around 0.60. The standard deviation of $S_{average}$ is more than halved when the number of training plots increases from five to fifteen and then decreases more slowly to 0.5 for $n = 21$. Mean scores obtained with setting selection based on spectrum or height distribution distances have mean values higher than those of the average setting, but with a faster decreasing trend with the number of training plots. The improvement of the mean score with training set size is better for the spectrum than for the height-based selection, with values decreasing from 1.51 ± 4.21 to 0.58 ± 0.38 compared to 1.16 ± 2.13 to 0.64 ± 0.50. Spectrum-based selection gives better results than height selection for larger training set sizes ($n > 11$), but still produces worse detection performance than the average setting. Both methods reach similar performance than the average setting for the highest tested size of training set ($n = 21$). Standard deviation for S_{S-opt} is even smaller than the one of $S_{average}$.

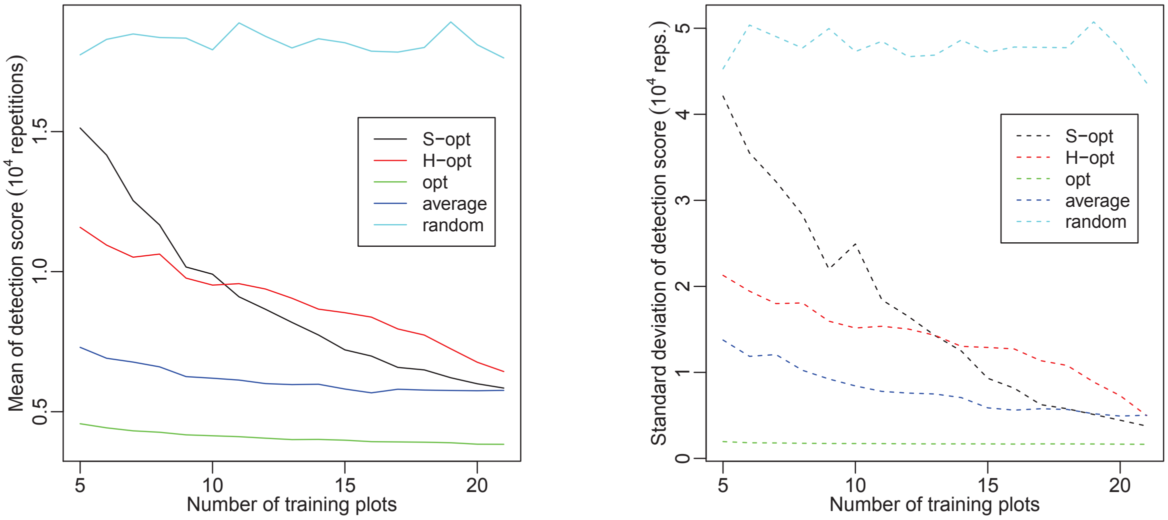
For each training set size, a Wilcoxon signed-rank test is performed to test the null hypothesis that the distribution of the difference of scores obtained with two parametrization criteria is symmetric (two-sided test). Distributions of S_{opt} and S_{random} are significantly different from the other distributions ($p < 0.001$). Distribution of S_{S-opt} is significantly different from $S_{average}$ for $n \in \{5, \cdots, 20\}$ ($p < 0.001$) and $n = 21$ ($p < 0.05$). Distribution of S_{H-opt} is also significantly different from $S_{average}$ for small training sets ($n \in \{5, \cdots, 19\}$, $p < 0.001$), but not for $n = 20$ ($p = 0.76$) and $n = 21$ ($p = 0.08$).

Table 8.1: Mean and standard deviation values of detection performance scores (10^4 repetitions) for different numbers of training plots n, and for the different parametrization criteria.

n	5	10	15	21
S_{S-opt}	1.51±4.21	0.99±2.49	0.72±0.93	0.58±0.38
S_{H-opt}	1.16±2.13	0.95±1.52	0.85±1.29	0.64±0.50
S_{opt}	0.46±0.20	0.41±0.17	0.40±0.17	0.38±0.16
$S_{average}$	0.73±1.38	0.62±0.84	0.58±0.59	0.58±0.50
S_{random}	1.77±4.53	1.79±4.73	1.82±4.72	1.76±4.36

8.2.2.2 Zoom on results for training set size $n = 21$

With 22 reference plots, results obtained with $n = 21$ correspond to a leave-one-out cross validation. Table 8.2 details the scores obtained on each plot when the algorithm is parametrized with the $n - 1$ remaining ones. The selection of the best available setting is done only for plots chablais3 and valdrome4 (both selection criteria), and chablais6 (height-based selection). The average setting yields only one score above 1, for the combeloup plot. Height-based selection gives four scores above 1 (for plots chablais2, chablais4, chablais7 and valdrome2), whereas it is only the case for two plots (chablais7 and valdrome3) for the spectrum based selection.



(a) Mean value

(b) Standard deviation

Figure 8.4: Mean and standard deviation values of detection performance scores (10^4 repetitions) as a function of the number of training plots. Line colors correspond to the different parametrization criteria: – black for spectrum distance S_{S-opt}, – red for height distribution distance S_{H-opt}, – blue for average setting $S_{average}$, – green for theoretical optimal setting from training plots S_{opt} and – cyan for random selection of setting from training plots S_{random}.

8.3 Discussion

8.3.1 Extracted area

The extraction of metrics is based on the point cloud located in a 50 × 50 m^2 area centered and aligned on the inventoried trees. In the case of plots with different shape (62.5 × 40 in Saint-Agnan, irregular in Vaujany), the extracted area does not exactly match the field plot. For example, in Saint-Agnan only 2000 of the 2500 m^2 (80%) are actually located inside the field plot. As plots are adjacent, 250 m^2 (10%) are located in a neighboring plot, and the remaining 10% are outside any inventoried area. Extracted point cloud for plots stagnan1 and stagnan4 share 20% in common area, and so do plots stagnan2 and stagnan3. The approximation may have two opposite effects. As the extracted metrics do not exactly correspond to the validation plot, the selection of a similar stand in the training plots should be less efficient. On the contrary, as the metrics are computed from a common area with an adjacent plot that may be included in the training set, the chances that this plot is selected are enhanced. Due to the spatial correlation in forests, the adjacent stands are likely to exhibit the same structure and will behave similarly with respect to algorithm parameters. This second effect would be in favor of a better selection of the parametrization. However, as only two pairs are concerned, the impact on the whole set of simulations is likely to be only minor.

The approximation can be corrected for rectangular plots provided that an appropriate range is then chosen for the values r_i where the spectrum is estimated. For plots with irregular shape, a rectangular shape is nevertheless required to perform the Fourier transform. A trade-off has to be found between the amount of extracted area that lies within and without the plot limits. For the height distribution no particular shape is required so that plot area could even be defined in a similar manner as the plot mask (section 6.3.1).

Table 8.2: Comparison of the detection score and corresponding setting for different parametrization criteria. For each tested plot, the training is constituted of the 21 others.

	Average	Height selection		Spectrum selection		Best in training set	
Plot	$S_{average}$	S_{H-opt}	setting	S_{F-opt}	setting	S_{opt}	setting
chablais1	0.54	0.44	chablais6	0.48	chablais7	0.33	corruaz
chablais2	0.45	2.42	moussoux	0.60	stagnan4	0.36	chablais4
chablais3	0.34	0.21	corruaz	0.21	corruaz	0.21	corruaz
chablais4	0.39	1.21	vaujanyba	0.57	vaujanyha	0.38	chablais2
chablais5	0.39	0.39	chablais2	0.29	valdrome3	0.24	corruaz
chablais6	0.17	0.07	chablais1	0.39	combeloup	0.07	chablais1
chablais7	0.60	1.33	corruaz	1.90	chablais1	0.65	orthaz3
chablais8	0.51	0.49	stagnan3	0.49	stagnan2	0.40	valdrome3
combeloup	2.93	0.96	vaujanyha	0.88	moussoux	0.84	chablais2
corruaz	0.66	0.69	chablais3	0.61	orthaz3	0.54	stagnan1
moussoux	0.58	0.61	stagnan2	0.75	combeloup	0.51	chablais5
orthaz3	0.64	0.45	stagnan2	0.45	corruaz	0.38	stagnan1
stagnan1	0.36	0.39	stagnan2	0.39	stagnan2	0.27	chablais7
stagnan2	0.35	0.37	stagnan1	0.37	stagnan1	0.32	chablais7
stagnan3	0.52	0.55	stagnan2	0.41	stagnan4	0.38	vaujanyha
stagnan4	0.48	0.48	stagnan3	0.61	chablais2	0.46	chablais7
valdrome1	0.41	0.19	stagnan4	0.23	valdrome2	0.19	stagnan3
valdrome2	0.37	1.27	valdrome3	0.33	corruaz	0.22	stagnan4
valdrome3	0.49	0.36	valdrome2	1.39	valdrome4	0.30	moussoux
valdrome4	0.68	0.40	valdrome3	0.40	valdrome3	0.40	valdrome3
vaujanyba	0.50	0.53	chablais4	0.53	stagnan2	0.46	chablais7
vaujanyha	0.58	0.62	vaujanyba	0.57	stagnan2	0.56	chablais5

8.3.2 Metrics

LiDAR metrics The proposed metrics reflect the vertical structure of the forest stand (height distribution) and the texture of the canopy cover (spectrum of the CHM). Other metrics should be considered in order to improve the description of the forest structure. Similarly to area-based approaches adopted for stand parameter estimation, other height metrics could be considered. It is possible to separate echoes according to their order of arrival within a given pulse (single, first of many, last of many. . .) and compute corresponding height distributions. Simplified metrics related to the height distribution (quantiles, mean and higher order statistics) might be sufficient. Regarding the horizontal structure, other texture variables can be used to characterize the CHM, such as co-occurrence matrices. Unfortunately, those 2D tools are applied to the CHM which encompasses only a limited part of the information of the whole point cloud. The development of 3D descriptors of irregular point clouds remains a challenging task.

LiDAR acquisition parameters Even tough the LiDAR acquisition parameters induce particular features in the CHM, which in turn are visible in the spectrum, it is likely that their influence in the algorithm selection is limited. Meanwhile, algorithm parametrization would probably benefit from taking them into account as LiDAR settings affect the interactions between pulses and the canopy, and hence the filter operations required for optimal detection of treetops. To investigate this effect, reference data with a wider range of LiDAR settings is required, as well as data from other sensors. Indeed, in this study only data from the RIEGL LMS-Q560 scanner are used and only one plot has a point cloud density lower than 6 m^{-2}.

Species information The other important aspect is the information about species, or at least coniferous–broadleaved distinction. Indeed, previous results showed that detection performance is linked to conifers proportion. Such information is not directly available in geometrical features, but could be extracted from other attributes related to LiDAR points. Indeed intensity or calibrated backscatter coefficient values can be used for species classification [180,184]. Another possibility is the fusion with optical images as spectral data remains more efficient than LiDAR data for species classification.

8.3.3 Nearest neighbor selection

For selection of a setting from the training plots, only the Euclidean distance was used. Other distances could also be tested, such as the Bhattacharyya distance. Besides, the potential of the two metrics was investigated separately. Combining both metrics, and maybe additional ones might be more efficient. When combining different metrics, two main approaches are possible. The first possibility is to apply different distances to each metric, and perform the final selection based on weighted or fuzzy criteria. The other alternative is to feed simultaneously all metrics to a most similar neighbor algorithm, which is a commonly used tool for forest inventories [208].

8.3.4 Reference data

The reference data only include 22 plots. Besides, all plots are not independent as ten of them are actually extracted from bigger plots (Saint-Agnan, Valdrôme and Vaujany). Subplots from a single area are not independent due to spatial correlation of forest structures and LiDAR acquisition characteristics. This may result in two opposite effects regarding the tested approach. On the one hand, unsupervised selection of a good setting is easier for subplots when other subplots from the same area are included in the training set. Indeed, table 8.2 shows that for a training set size of 21, i.e. other subplots are systematically included in the training set, the settings of Saint-Agnan and Valdrôme plots are frequently chosen among other subplots from the same area. On the other hand, the number of really independent plots in the training set is lower which decreases the matching possibilities.

Even tough the trend suggests that for large training sets the setting selection based on height or spectrum distributions would become more efficient than using an average setting, results obtained for this limited reference should be handled with care. Indeed, the training sets are not always and completely independent from the tested plot, and the mean score for the average setting is greatly influenced by the bad result on plot combeloup.

8.4 Conclusion and perspectives

Detection performance, and height estimates obtained with the proposed algorithm are similar to those from previous studies in mountainous environments. However, the results obtained with the unsupervised parametrization of the algorithm are promising. This framework for algorithm training represents an alternative to manual parametrization or to the development of complex detection algorithms. Indeed, any algorithm that requires parametrization can be used within the proposed workflow, provided that:

- parameters can be externally tested during the training step,
- the algorithm outputs the coordinates and height of the detected treetops to allow the automated matching and assessment of detection performance.

The nearest neighbor selection procedure still requires improvement, as other metrics and other selection procedures should be tested. Besides, metrics to be considered might be adapted to a given detection algorithm.

For operational applications, the framework is also quite flexible. The score criterion can be adapted to reflect the end-user requirements regarding the trade-off between true and false positives. The training set of reference plots can be build progressively from various datasets.

However, the issue of the aggregation of tree parameters into stand parameters estimates remains challenging. More investigation is required to check if the information on the plot selected from the training set proves useful to automatically assess the detection performance for the tested plot. The problem of the optimal size of field plots should also be addressed, but it is likely to be a complex task as it impacts different aspects of the approach: stand structure assessment is scale-dependent and border effects in the matching process are enhanced for small plots.

Part IV

Estimating stand parameters

Chapter 9

Introduction

9.1 Area-based methods

9.1.1 Overview

Even though the modelling of the interaction between the laser pulses and the elements constituting the forest cover remains a challenging task, it is clear that the laser reflections exhibit a strong relationship with forest structure. The pragmatic approach based on the calibration of empirical models between the statistical distribution of the laser echoes and forest parameters quickly proved successful for the estimation of mean tree height [187, 226], stand volume [227], leaf area index [187], basal area [199] or stem density [233]. This approach was then formalized into a two-step method designed to be used at operational scale by forest practitioners [228]. The idea is first to calibrate empirical models with stand parameters as dependent variables and selected laser metrics as independent variables, using a set of field plots as training data. The models are then used to estimate stand parameters for all the area covered by LiDAR acquisition. For such application, a LiDAR sampling rate around one pulse per square meter is sufficient [90, 228, 231].

The method was soon tested and proved successful at operational scale [229, 232]. Various other forest cases were investigated with similar methods, mainly in boreal conditions [90, 122, 191], including cases such as plantations [237], young forest [233], mixed forest [238], boreal reserve [237]. It was then tested in other contexts, such as tolerant northern hardwood forests [180, 181] or in complex alpine environments [106].

As the interaction between laser pulses and the vegetation heavily depends on the stand structure and species composition, some studies relied on prior stratification of forest types before models are calibrated for each determined type, even within a small LiDAR acquisition. Stratification can be based on prior knowledge of the area or on photo-interpretation [228]. Potential stratification criteria are based on factors affecting forest structure more or less directly, e.g. stand age [228], management stage [180], coniferous–broadleaved proportion [106, 238], site quality [228, 232], altitude [106]. In a study conducted in the Bavarian Forest National Park, stratum-specific models generally had better prediction accuracy than general ones [106]. In boreal forests, conifer models used on birch dominated stands yielded accurate height estimates but poor results for stem number and volume [231].

As usual with LiDAR data, one of the other concerns is the sensitivity of the method to acquisition parameters. The method was found to yield stable results, even for low point densities or perturbation of acquisition conditions. In [122], prediction error remained almost constant when density was reduced from 4.3 to 0.1 points per square meter. In [238], leaf-on and leaf-off acquisition turned out to affect laser metrics in mixed stands. Besides, changing the sensor configuration had an effect on first pulse data. However, models trained with each dataset separately, although different, had similar prediction accuracy. The method was also shown to be robust against flying altitude variations [230].

If the procedure seems to be able to handle a broad range of forest structures and LiDAR acquisition parameters, one key point is the field data used to train the models. Differences in training datasets can lead to lower accuracies [122]. [234] emphasizes the fact that field data should be a representative sample of encountered forest types in order to avoid systematic errors. However, field data from different districts could be combined provided that they are affected to the correct strata. Another study combining training data from two districts, one dominated by pine and the other by spruce, nevertheless obtained a common regression model with satisfactory prediction accuracy [232].

Most of the studies relied on ordinary least squares multiple regression to establish relationships between laser metrics and forest parameters. A comparison of seemingly unrelated, partial least squares and ordinary least squares regressions showed only minor discrepancies [234]. However, parametric methods reach their limits when dealing with a small number of field observations combined with high dimensional data. The non parametric k-most similar neighbor method has been successfully tested for species-specific stand attributes estimation from laser data [244]. In a comparison with k-most similar neighbor and artificial neural networks models such as self organizing map and multilayer perceptron, support vector machines regression (SVR) turned out to be one of the best suited method for prediction purposes [224].

9.1.2 Objectives

In French mountainous areas, many coppice stands are encountered in the hill and lower montane belts. In such stands, each tree stump has several stems that are very close one to another (figure 3.1). Such structures can not be easily characterized by single tree methods. Besides, information at the tree level is not of utmost importance for the management of coppice stands. Stand parameters such as stem density, dominant height and mean diameter are more interesting to the forest practitioner. Therefore, the area-based method seems more appropriate for this particular case. As no study of estimation of parameters of coppice stands from airborne laser scanner data was available in the literature, the first objective of this chapter is to test the two-step procedure [228] on mountainous coppice stands and to compare results to those obtained for a mixed forest.

It was exposed in 5.2 that topographic constraints limit the productivity of both the inventory and GPS acquisition. As collection of data on large plots is costly, the effect of sample plot size on the prediction accuracy of LiDAR models must be quantified so that efficient procedures can be designed. That is why the second objective is to assess the effect of sample plot size on the training data and on the accuracy of resulting LiDAR models. Those two aspects will be treated in chapter 10.

If the total number of field plots is limited by field constraints, the number of metrics that can be extracted from the laser data is very important, and it is likely that several of them are correlated. Dimension reduction methods are an alternative way to extract meaningful information from original variables into a limited number of components. The third objective is thus to evaluate the effect of the choice of laser metrics on models accuracy, and consists in two complementary aspects. One is to assess the influence of different numbers and types of laser metrics in the models, including an entropy metric designed to input information about the horizontal dispersion of the point cloud. The other is the relevance of dimension reduction of the raw laser metrics prior to model training. Principal component analysis and independent component analysis are tested.

Relationships between LiDAR metrics and forest variables might be non linear. As an alternative to ordinary multiple least squares regression, the fourth objective was to test a non linear regression method based on support vector machines. Those aspects will be handled in chapter 11.

For all these investigations, two areas are studied: the coppice dominated forest of Saint-Paul-de-Varces, and the Chablais. Estimated forest parameters are basal area, stem density, mean diameter, dominant height and stem volume (Chablais data only). General workflow and methods are presented in this chapter.

9.2 General workflow

Figure 9.1 displays the global workflow implemented to estimate forest stand parameters from airborne laser scanning data. From now on ordinary multiple least square regression and support vector regression will be referred to as *ols*-MR and SVR respectively. All computations are performed with the R software [259].

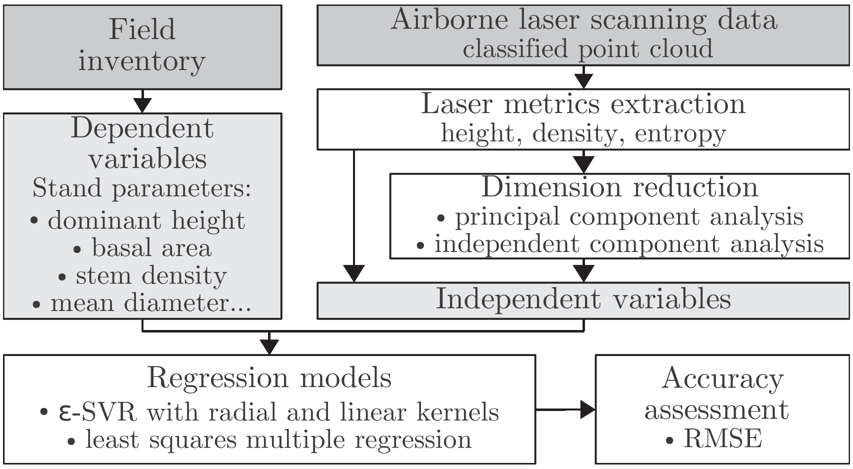
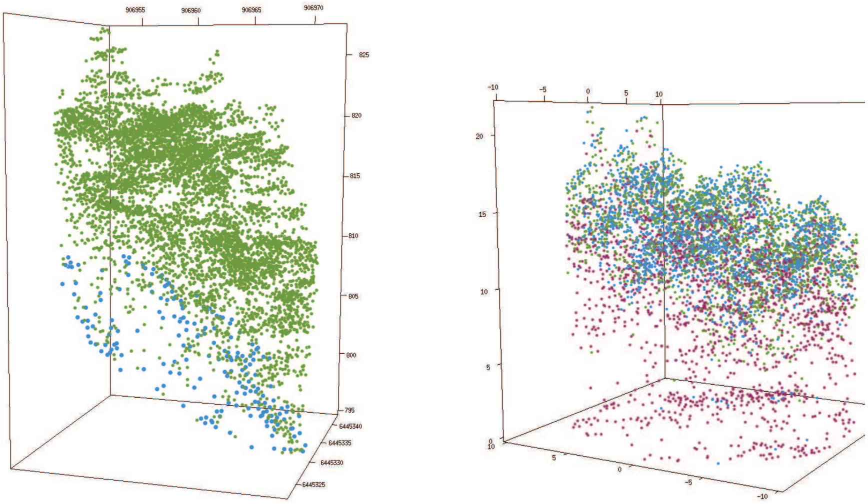


Figure 9.1: General workflow for estimating forest stand parameters with an area-based method.

9.2.1 Point cloud extraction

We consider a set of N_F field plots $(F_j)_{j\in\{1,\dots,N_F\}}$. Plots are of circular shape with radius r, and the coordinates of their centers $(x_j, y_j)_{j\in\{1,\dots,N_F\}}$ are known. We also have the coregistered laser point cloud with following information about each point: 3D coordinates, return position category, ground or non-ground classification. Return position is factor variable which levels are: s for single echoes (only one echo for a given pulse), f for first echoes and l for last echoes. The classification of laser points in such categories is directly known from the echo extraction process. Classification into ground or non ground echoes results from the procedure implemented by the LiDAR contractor (see 4.3.3).

For each plot j, laser points within r meters horizontal distance form the plot center are extracted (figure 9.2a). Their normalized heights are computed by subtracting the terrain altitude at their orthometric coordinates. Terrain surface is estimated by bilinear interpolation of points classified as ground points. Points with height lower than 2 m are excluded to avoid influence of dense shrubs understory, as suggested by [228]. Three point groups are then constituted according to return positions: single, first and last echoes (figure 9.2b). Many studies implementing this methodology only relied on two points groups as their data was acquired with multi-echo scanners with only first and last echo classes. Here single echoes are considered as a separated group. Intermediate points, i.e. echoes occurring between the first and the last echoes, represent less than a few percent of the total number and are not included in the analysis.



(a) Original point cloud (b) Normalized point cloud

Figure 9.2: 3D view of the point cloud in plot 17 of the Saint-Paul-de-Varces inventory. In (a) points are colored according to the classification: ● ground or ● vegetation. In (b) points are colored according to their return position: ● single, ● first or ● last.

9.2.2 Laser metrics

Several types of laser metrics can be extracted from the laser point cloud. Usual metrics mainly consists in two categories [106, 228]. Height metrics are the height quantiles, mean, relative median, standard deviation and coefficient of variation. Density metrics are computed as the proportion of laser points lying under given height thresholds. In more recent works, other types of metrics have been added, such as higher order statistics or intensity-related metrics [129]. Here only basic height and density quantiles are considered. Indeed, the number of observations in our field data is limited, and considering too many independent variables may not be parsimonious. For each group, n_h height metrics and n_d density metrics are calculated. The height metrics include the minimum, maximum, mean and q-quantiles [134, definition 7] of the height distribution (total: $n_h = q + 2$). The density metrics are computed as the proportions of points located below fractions $(\frac{i}{n_d+1})_{i \in \{1,\dots,n_d\}}$ of the maximum height of the point group in each plot.

Usual metrics only reflect information about the vertical distribution of laser points. However, information about the horizontal distribution of laser points may also be extracted from the point cloud and could be particularly interesting in the case of heterogeneous stands. Entropy, which is a statistical measure of randomness, is introduced at this point. A histogram estimator of the entropy of the 2D distribution of laser points located in one or several height bins is used (figure 9.3). The whole point cloud is horizontally divided in 2 m wide square pixels (s_i) and vertically divided in n_e height bins of equal width $(h_j)_{j \in \{1,\dots,n_e\}}$. Only pixels which have more than 90% of their area inside the plot circle are considered. For each height bin h_j, the entropy metric is computed as:

$$e_j = \sum_i p_{i,j} log(p_{i,j}) \qquad \text{with } p_{i,j} = \frac{card(s_i \cap h_j)}{card(h_j)} \tag{9.1}$$

A set of independent variables $(v_i)_{i \in \{1,\dots,n_v\}}$ is thus composed of $n_v = 3 \times (n_h + n_d) + n_e$ laser metrics. For example, the variables set $(n_h, n_d, n_e) = (6, 3, 3)$ has 30 laser metrics. For each of the three point groups and each plot, six height metrics are calculated (minimum, first quartile, median, third quartile, maximum and mean), and three density metrics (proportion of echoes

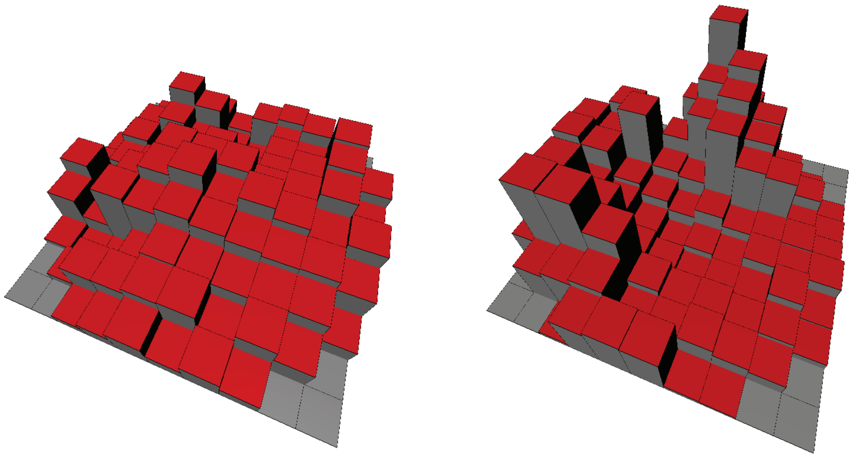

(a) (b)

Figure 9.3: Example of the extraction of $n_e = 2$ entropy metrics on plot 17 of Saint-Paul. The histograms are the 2D horizontal distributions of points located (a) above and (b) below $h = \frac{h_{max}}{2}$, used to calculated the entropy metrics. The plot has 10 m radius, pixel size is 2×2 m^2.

below 25%, 50% and 75% of the maximum height of the group). Entropy metrics for the whole point cloud are calculated for each plot as the entropies of the horizontal distribution of echoes located in three height bins: $[0, \frac{h}{3}[$, $[\frac{h}{3}, \frac{2\times h}{3}[$ and $[\frac{2\times h}{3}, h]$ with h the maximum height. The matrix of extracted metrics $\underset{N_F \times n_v}{M}$ consists of N_F observations of $n_v = 3 \times 9 + 3 = 30$ laser metrics.

9.2.3 Dimension reduction

9.2.3.1 Objectives

Concerning prediction models using laser data, one key issue to avoid over-fitted or complex models is the selection of relevant laser metrics from the laser variables set. As mentioned beforehand, the number of features that can be extracted from the point cloud is very important, whereas the number of observations is often limited by labor-intensive field data collection. The extraction of a high number of possibly correlated metrics puts up several issues. With *ols*-MR, models with too many variables will tend to be over-fitted, and models will often fail to fulfil the linear model assumptions. For non linear regression methods, accuracy may decrease as a result of the curse of dimensionality. Therefore, it might be interesting to further process the numerous laser metrics in order to obtain a reduced number of features while retaining as much information as possible.

This process of dimension reduction (DR) can be implemented by selecting a subset of the original features, or by mapping the original data in the high dimensional space into a new feature space with lower dimension. For multiple regression most of the studies relied on stepwise variable selection [106,228] or exhaustive comparison of combinations [122]. Principal component analysis was also used both for laser metrics selection and for the extraction of a limited number of components [178].

When regression models such as SVR have to be tuned by evaluating for several hyperparameters combinations a prediction error, testing all possible combinations of laser metrics represent a huge computational burden. Multi objective genetic algorithms (MOGA) can be used to perform global exploration while limiting the processing time compared to greedy alternatives [224]. However, in this last reference which compared different non linear regression models for the prediction of stand variables, MOGA was used only with one reference method for feature selection.

As a result, regression methods were compared on their ability to process a single set of features that has been chosen optimal for one given reference method. This eludes the fact that different methods may be able to take different advantage of the information included in the raw data.

In this work, a “brute-force” strategy is adopted since for *ols*-MR all possible combinations of a maximum of three variables (laser metrics or extracted components) are tested. This is computationally possible thanks to the low number of observations and the fact that the number of extracted laser metrics is also limited. For SVR, all extracted laser metrics are systematically entered as input to the algorithm, but when DR is performed, different numbers of features are used to fit the models.

9.2.3.2 Methods

Principal component analysis (PCA) is considered one of the most applicable DR method [56]. It reduces dimensionality by extracting from the original data set uncorrelated components which encompass the highest variance [144]. As our data include different variables (heights, densities, entropies) the original data set is centered and standardized beforehand. The PCA is then equivalent to an eigenanalysis of the correlation matrix.

Independent component analysis (ICA) extracts components that are also called latent variables. Their calculation is based on maximization of the statistical independence of the estimated components. An implementation of the FastICA algorithm [135] is used (R package *FastICA*). ICA has already proved quite efficient for the reduction of laser metrics [129].

When the number of observations N_F is greater than the number of variables n_v, PCA and ICA are performed for dimension reduction. Subsets of dimension $n_v \in \{1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24\}$ of the obtained principal and independent components are also used as sets of independent variables, i.e. the matrix of extracted metrics is replaced by the matrix of selected components during the model training step.

9.2.4 Regression models

9.2.4.1 Ordinary least squares multiple regression

Given a dependent variable (stand parameter) y and a independent variables set $v = (v_i)_{i\in\{1,\ldots,n_v\}}$, the field observations $(y_j)_{j\in\{1,\ldots,N_F\}}$ and the matrix of extracted metrics $\underset{N_F\times n_v}{M}$ are used as training data to fit a multiple regression model by ordinary least squares:

$$y = b + \sum_{i=1}^{n_v} a_i \times v_i \qquad \text{with } ((a_i)_{i\in\{1,\ldots,n_v\}}, b) \text{ the model parameters} \tag{9.2}$$

Models including a maximum of three variables are tested by exhaustive search among all the possible combinations (package *leaps*). Models which do not fulfill the linear model assumptions (package *gvlma*, which tests a linear model for skewness, kurtosis, link function and heteroscedasticity) or including a variable with a partial p-value greater than 0.05 are discarded. For each variables set the model with highest adjusted coefficient of determination (adj-R^2) was selected.

9.2.4.2 Support vector regression

Support vector machines are a training approach based on the framework of statistical learning theory. They have proved their robustness to dimensionality and generalization abilities [101]. Moreover non-linear relationships can be accounted for thanks to the kernel trick. Support vector machines are now widely used in the field of remote sensing, mainly for the purpose of hyperspectral image processing [179], but also for continuous variables estimation [23, 37, 216, 217, 224, 284, 322].

Datasets are also used to train an ϵ-SVR. It is a common implementation of SVR which aims at approximating a function $f\ :\ y = f(x)$ that has at most ϵ deviation from the actual targets y_j. The solution has the form:

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i) + \beta \qquad \text{with} \begin{cases} ((\alpha_i)_{i \in \{1,\ldots,n\}}, \beta) \text{ parameters from the training process} \\ k \text{ a kernel function} \\ x_i \text{ samples from the training set} \end{cases} \tag{9.3}$$

The R package *e1071* is used for fitting the ϵ-SVR models.

9.2.5 Validation procedure

Due to the absence of external validation data, prediction accuracy is evaluated in cross-validation by computing the root mean square error ($RMSE$) and its coefficient of variation CV_{RMSE}.

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{N_F}\sum_{j=1}^{N_F}(y_j - \hat{y}_j)^2} && \text{with} \begin{cases} y_j \text{ the observed values} \\ \hat{y}_j \text{ the predicted values} \end{cases} \\ CV_{RMSE} &= \frac{RMSE}{\bar{y}} && \text{with } \bar{y} = \frac{1}{N_F}\sum_{j=1}^{N_F} y_j \end{aligned} \tag{9.4}$$

Two procedures are used for computing the cross-validated $RMSE$. Leave-one-out (LOO) cross-validation consists in successively removing one observation from the training data, fitting the model with the $N_F - 1$ remaining ones and then predicting the value for the left-over plot with the obtained model. When several subplots are extracted from a single plot, a “leave-one-plot-out” cross-validation is performed, i.e. all subplots from a given plot are excluded of the training data. This procedure aims at avoiding that the predicted value for a given subplot is estimated with a model fitted to similar data, as it is the case when subplots are extracted from one single homogeneous plot (e.g. plot 6 in Chablais).

200 repetitions of ten-fold cross-validation are also performed to estimate the standard deviation of the $RMSE$. Ten-fold cross-validation relies on the same principle as LOO cross-validation except that observations are first grouped in ten groups, and all observations from a given group are removed from the training data before the model is fitted and then used to calculate their predicted values. LOO cross-validation is equivalent to a N_F-fold cross-validation but with a ten-fold one the computational burden is lower and more combinations are possible for the sampling of observations in the folds. When a ten-fold cross-validation is used, the fact that some subplots may be extracted from the same plot is not taken into account.

During cross-validation procedure, the following rules are applied for successive model fitting. For *ols*-MR, the variables selection procedure is not performed again. The coefficients of the three (at most) previously determined variables in the model that yielded the best coefficient of determination are fitted by ordinary least squares. For ϵ-SVR, the model parameters C (Cost) and γ for radial kernel are not tuned again. Their values are set to the previously determined scalars.

Chapter 10

Multiple regression models

The objective is to test the basic two-step method based on ordinary least squares multiple regression for the prediction of stand attributes in a coppice forest and in a mountainous, mixed forest, and to assess its sensitivity to training plot size.

10.1 Methods

The workflow described in figure 9.1 is implemented in a simplified procedure: no dimension reduction is performed and only multiple regression models are fitted. It is applied to data from Saint-Paul-de-Varces and Chablais. Field data ($N_F = 31$ circular plots of 10-m radius and $N_F = 32$ circular plots of 12.5 m radius respectively) are described in section 5.3.3.1. Only live trees are considered and minimum diameter is 5 cm in Saint-Paul-de-Varces and 7.5 cm in Chablais, unless specified otherwise. The LiDAR data are presented in section 5.3. To evaluate the effect of training plot size on prediction accuracy, plot radius is reduced by excluding trees situated further than various distance thresholds before reiterating the procedure. Tested radii are $r \in \{5, 5.5, ..., 10\}$ for the Saint-Paul-de-Varces data and $r \in \{5, 5.5, ..., 12.5\}$ for Chablais. It is noteworthy that differences in field protocols may have an effect on results:

- trees above 5 cm DBH inventoried in Saint-Paul, 7.5 cm in Chablais;
- inventory performed systematically from plot center in Saint-Paul, not in Chablais;
- all tree heights measured in Chablais, only for a sample in Saint-Paul.

Stand parameters The following stand parameters are computed for each plot $j \in \{1, 2, ..., N_F\}$ and each radius r: basal area (G_j^r: surface occupied by the horizontal section of tree stems at 1.30 m height), stem density (N_j^r), mean diameter at breast height (D_j^r), dominant height (H_j^r: mean height of the 100 trees with largest diameters per hectare) and stem volume (V_j^r).

For the Saint-Paul-de-Varces data, dominant height is calculated for $r = 10$ m only, as the field sampling scheme does not ensure that enough measured trees are included within each radius. As all tree heights are not measured, the following procedure is applied for dominant height calculation. If at least one of the three largest trees on the plot has its height sampled, dominant height is calculated as the mean value of the available heights, or else dominant height is estimated as the height of the largest sampled tree. Stem volume is not calculated for Saint-Paul-de-Varces.

Wilcoxon signed-rank tests are performed to compare forest parameters observations obtained with different plot sizes. Correlation between field observations statistics and plot radius is evaluated by computing Spearman ρ.

Table 10.1: Basal area (G), stem density (N), mean diameter (D), dominant height (H) and stem volume (V) field observations statistics (mean, minimum, maximum and standard deviation) for plot radius $r \in \{5, 7.5, 10, 12.5\}$ of Saint-Paul-de-Varces and Chablais inventories ($N_F = 31$ and 32 respectively).

Parameter	r	Saint-Paul-de-Varces				Chablais			
		Mean	Min	Max	Sd	Mean	Min	Max	Sd
G (m^2.ha^{-1})	5	32.3	4.9	66.8	16.7	34.3	2.1	96.0	24.6
	7.5	34.6	4.7	90.1	15.9	30.8	5.1	85.1	17.9
	10	34.8	4.6	59.7	11.4	32.3	4.8	79.2	18.5
	12.5	Data not available				34.1	8.5	72.5	17.2
N (ha^{-1})	5	1668	382	3310	724	525	127	891	209
	7.5	1714	736	3112	560	502	226	1019	199
	10	1735	764	2833	577	529	191	923	196
	12.5	Data not available				538	244	917	169
D (cm)	5	14.5	8.9	30	4.6	26.4	8.3	64.0	12.0
	7.5	14.4	8.3	23.3	3.9	25.4	12.8	47.2	8.3
	10	14.5	8.3	22.7	3.6	25.3	12.4	48.3	8.2
	12.5	Data not available				25.9	13.3	42.3	8.3
H (m)	5	Data not available				23.5	8.3	38.7	8.9
	7.5					25.0	10.1	38.7	7.8
	10	15.6	5.8	25.8	5.5	24.6	12.2	39.2	7.3
	12.5	Data not available				25.4	14.3	38.3	7.2
V (m^3.ha^{-1})	5	Data not available				412	22	1483	396
	7.5					358	37	1300	300
	10					374	34	1207	307
	12.5					398	56	1094	299

Laser metrics The laser metrics are calculated for each of the three point groups g: s single echoes (only one echo for a given pulse), f first echoes and l last echoes. They include only $n_h = 6$ height metrics: minimum ($h_{g,0}$), first quartile ($h_{g,0.25}$), median ($h_{g,0.5}$), third quartile ($h_{g,0.75}$) and maximum ($h_{g,1}$) values, plus mean height ($h_{g,mean}$), and $n_d = 3$ density metrics: proportion of echoes recorded below height thresholds corresponding respectively to 0.75, 0.5 and 0.25 of the maximum echo height recorded on the plot for this group ($d_{g,0.75}$, $d_{g,0.5}$ and $d_{g,0.25}$ respectively). For each radius r, the variables set $(v_i^r)_{i\in\{1,\ldots,27\}}$ consists of N_F observations of $n_v = 3 \times 9 = 27$ laser metrics.

10.2 Results

10.2.1 Influence of plot radius on field observations

Forest plots statistics for 5, 7.5, 10 and 12.5 m radii are displayed in table 10.1. Figure 10.1 presents boxplots of field observations of basal area, stem density and mean diameter depending on plot radius. Graphical analysis shows that only basal area, mean diameter and volume have data points outside the whiskers (except one value for stem density calculated on 7 m plot in Saint-Paul). Besides, such values are mostly large values. Similarities between basal area and mean diameter can be explained by the fact that for a plot with N_T trees, the parameters basal area G, mean diameter D and the sum of squared residuals to mean diameter are linked by the following relationship:

$$G = \frac{\pi}{4S}\sum_{i=1}^{N_T} d_i^2 = \frac{\pi}{4S}\left(N_T \times D^2 + \sum_{i=1}^{N_T}\epsilon_i^2\right) \qquad \text{with } (\epsilon_i = d_i - D)_{i\in\{1,\cdots,N_T\}} \tag{10.1}$$

Volume is also quadratically linked with diameter [see equations (5.2), (5.3), (5.4)].

Otherwise, no remarkable trend can be distinguished. Wilcoxon signed-rank tests indicate that the distributions of the observations differences ($y_{r1} - y_{r2}$ with $r1 \neq r2$ and $y_r \in \{N^r, D^r\}$) are not significantly asymmetric about 0 ($p > 0.05$). For basal area, there were six and one cases with a significant value, out of a total of 120 and 55, for the Chablais and Saint-Paul data respectively. For volume (resp. dominant height), 4 (resp. 40) out of 120 tests had a significant value for the Chablais data. Dominant height values indeed greatly depend on the inclusion of a few very large trees (only one for small radius plots). This effect is reduced for volume and basal area where large trees have an important contribution to the value, but with all other trees taken into account. For mean diameter, the importance of large trees is further reduced (no quadratic effect) and that may explain why Wilcoxon tests are not significant.

Despite the low number of significant results, numerous statistics exhibit a correlation with plot radius. Spearman correlation tests between stand variables statistics and plot radius are presented in table 10.2.

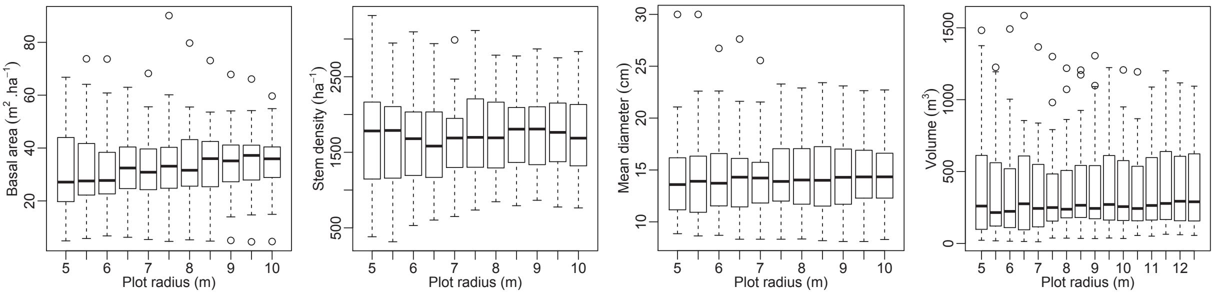
Except for stem density in Saint-Paul, the standard deviation of stand parameters is negatively correlated with plot radius. Indeed with larger plots, more trees are taken into account and local spatial effects are averaged, hence a lower variability of estimates. In the coppice stands of Saint-Paul, stem density is very high and stems were inventoried from 5 cm DBH. Even in small plots, local effects are less pronounced as many trees are inventoried.

Besides, the interquartile range (IQR) is significantly correlated only for basal area (Saint-Paul), and dominant height (Chablais). As exposed beforehand, dominant height values rely on a limited number of trees, and so does basal area, to a lesser extent. In Saint-Paul, bigger trees from seedling origin are usually found inside the multiple coppice stems. Such particular trees have an important effect on total basal area. IQR is a non parametric measure of the variability of a sample that is less sensible than standard deviation to extreme values. The fact that regarding standard deviation almost all variables are correlated to plot radius, whereas IQR is not, shows that the effect is mainly linked with outliers.

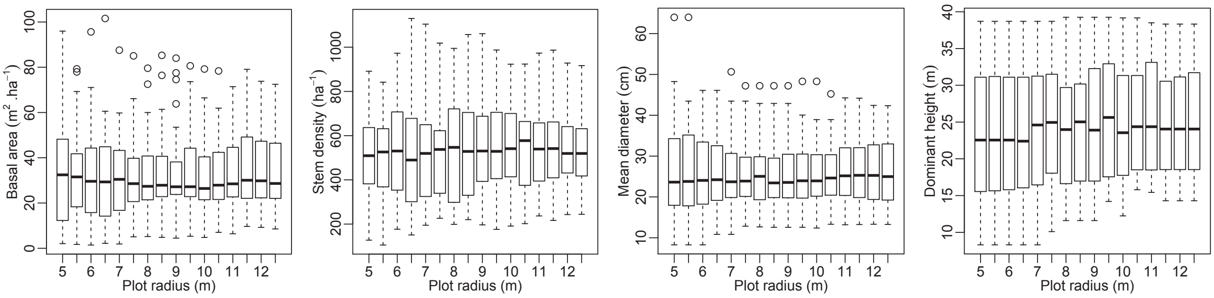
This remark about extreme values is confirmed by the fact that the minimum value is correlated to plot radius for all variables. As could be expected, the correlation is positive, expect for basal area and mean diameter in Saint-Paul. ρ for maximum values are all negative, but correlation is significant only for stem density (Saint-Paul), basal area and volume (Chablais) and mean diameter. The first quartile of all variables has a significant, positive correlation with plot radius (except for mean diameter in Chablais), but the third quartile does not exhibit any particular trend. This shows that lower values (minimum and first quartile) are more sensitive to plot radius than large ones.

10.2.2 Influence of plot radius on multiple regression models

Selected multiple regression models for $r \in \{5, 7.5, 10, 12.5\}$ are detailed in table 10.3. With 10 m radius plots in Saint-Paul-de-Varces, prediction accuracies (coefficient of variation of the $RMSE$ obtained by leave-one-out cross-validation) are respectively 19.4, 25.5, 14.3 and 19.4 % for basal area, stem density, mean diameter and dominant height. For $r = 7.5$ m results slightly improve for mean diameter and stem density (respectively 13.6 and 23.0 %), but are worse for basal area (36.4 %). The lowest accuracies are achieved with 5 m radius: 32.9 % for stem density and 36.7 % for basal area.



(a) G - StP (b) N - StP (c) D - StP (d) V - Chb



(e) G - Chb (f) N - Chb (g) D - Chb (h) H - Chb

Figure 10.1: Boxplots of field parameters observations: basal area G ($m^2.ha^{-1}$), stem density N (ha^{-1}), mean diameter D (cm), dominant height H (m) and volume V ($m^3.ha^{-1}$) depending on plot radius. Subfigures (a)-(c) correspond to Saint-Paul-de-Varces data (StP), subfigures (d)-(h) to Chablais data (Chb).

Table 10.2: Spearman correlation coefficient ρ between field observations statistics and plot radius for basal area (G), stem density (N) and mean diameter (D). Statistics are minimum (Min), first quartile (Q1), median, third quartile (Q3), maximum (Max), mean, standard deviation (Sd) and interquartile range (IQR). Significance levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ (two-sided test).

Area		Min	Q1	Median	Q3	Max	Mean	Sd	IQR
Saint-Paul-de-Varces	G	-0.68^{*}	0.98^{***}	0.91^{***}	-0.02	-0.31	0.85^{**}	-0.89^{***}	-0.76^{***}
	N	0.84^{**}	0.90^{***}	0.12	0.11	-0.76^{**}	0.87^{***}	-0.54	-0.60
	D	-0.87^{***}	0.77^{**}	0.75^{*}	0.57	-0.93^{***}	-0.52	-0.98^{***}	-0.14
Chablais	H	0.93^{***}	0.96^{***}	0.43	0.15	-0.30	0.86^{***}	-0.91^{***}	-0.79^{***}
	G	0.89^{***}	0.68^{**}	-0.36	0.14	-0.84^{***}	0.50^{*}	-0.85^{***}	-0.25
	N	0.77^{***}	0.66^{**}	0.31	-0.04	-0.12	0.84^{***}	-0.75^{**}	-0.44
	D	0.88^{***}	0.49	0.61^{*}	-0.21	-0.73^{**}	-0.37	-0.85^{***}	-0.27
	V	0.86^{***}	0.54^{*}	0.59^{*}	0.19	-0.87^{***}	0.36	-0.75^{**}	0.02

With Chablais data, prediction accuracies for 12.5 m plots are respectively 18, 23, 12.6, 11.7 and 30.4 % for basal area, stem density, mean diameter, dominant height and stem volume. For basal area, stem density and volume, accuracy decreases considerably for smaller plots. For mean diameter and dominant height, the effect is less pronounced.
Wilcoxon signed-rank tests indicate that the distribution of the differences between predicted and observed values is not significantly asymmetric about 0 ($p > 0.05$) for any forest parameter. It is noteworthy that laser metrics included in the models depend both on the forest parameter and on training plot radius.

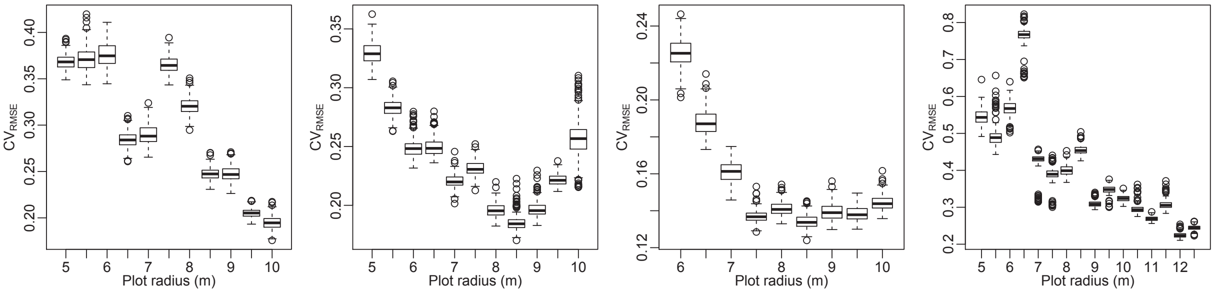
Table 10.3: Selected multiple regression models for radius $r \in \{5, 7.5, 10\}$, and corresponding coefficient of determination adj-R^2 and coefficient of variation of the $RMSE$ computed by leave-one-out (LOO) cross-validation and 200 repetitions of ten-fold cross-validation (mean ± standard deviation).

Site	Parameter	Radius	Laser metrics in the model	*adj*-R^2 (%)	CV_{RMSE} LOO (%)	CV_{RMSE} 10-fold (%)
Saint-Paul-de-Varces	G	5	$h_{s,0} + h_{s,1} + h_{f,mean}$	55.5	36.7	34.3±0.8
		7.5	$h_{f,0.75}$	46.0	36.4	36.2±1.1
		10	$h_{f,0.25} + h_{f,0.5} + d_{s,0.5}$	70.8	19.4	19.5±0.7
	N	5	$h_{l,mean} + d_{f,0.75} + d_{l,0.25}$	51.0	32.9	31.8±0.9
		7.5	$h_{f,0.75} + h_{l,1} + d_{l,0.25}$	59.1	23.0	22.9±0.6
		10	$h_{l,0.5} + d_{f,0.25}$	58.0	25.5	25.7±1.5
	D	5	linear model assumptions not satisfied			
		7.5	$h_{f,0} + h_{l,0.25} + d_{l,0.75}$	76.6	13.6	13.6±0.4
		10	$h_{s,0.75} + h_{f,0.75} + d_{l,0.25}$	71.4	14.3	14.4±0.4
	H	5, 7.5	no relevant due to sampling scheme			
		10	$h_{s,0.5} + h_{l,0.5} + h_{l,0.75}$	73.1	19.4	19.5±0.6
Chablais	G	5	$h_{f,0.75} + h_{l,0} + d_{s,0.5}$	67.9	44.0	44.5±1.4
		7.5	$h_{s,0.75} + d_{s,0.25} + d_{l,0.75}$	79.5	29.4	30.4±1.4
		10	$h_{f,0.5} + h_{means,} + d_{s,0.25}$	85.9	24.1	23.4±0.7
		12.5	$h_{f,0.75} + h_{l,0.75} + d_{s,0.25}$	90.6	18.0	16.3±0.4
	N	5	$h_{s,0.75} + h_{l,0.75} + d_{s,0.5}$	44.2	34.2	32.2±1.0
		7.5	$h_{l,0.5} + h_{s,means} + d_{f,0.25}$	67.6	25.3	24.5±0.7
		10	$h_{f,0.75} + d_{f,0.25}$	62.1	23.9	23.8±0.5
		12.5	$h_{s,0.5} + d_{f,0.25}$	53.1	23.0	22.2±0.5
	D	5	$h_{s,0.75} + h_{f,0.25} + h_{f,0.5}$	74.4	26.4	24.3±0.5
		7.5	$h_{s,0.75} + h_{f,0.5} + d_{l,0.75}$	72.2	22.9	18.8±0.6
		10	$h_{l,1}$	50.3	27.3	23.9±0.6
		12.5	$h_{f,0.25} + h_{l,0.5} + d_{s,0.75}$	85.2	15.6	13.3±0.4
	H	5	$h_{s,0.75} + h_{f,0.5} + d_{f,0.25}$	92.9	11.0	10.9±0.3
		7.5	$h_{s,0.25} + h_{s,0.5} + h_{s,0.75}$	84.6	14.3	13.2±0.4
		10	$h_{l,0.75} + d_{s,0.25} + d_{l,0.5}$	90.5	10.8	9.8±0.3
		12.5	$h_{l,0.75} + d_{f,0.5} + d_{l,0.25}$	91.0	11.7	9.5±0.4
	V	5	$h_{l,0} + h_{l,0.75} + d_{f,0.25}$	73.5	54.6	54.4±2.2
		7.5	$h_{s,mean} + d_{s,0.25} + d_{l,0.75}$	87.1	34.9	38.5±2.5
		10	$h_{s,0.5} + h_{s,1} + h_{s,mean}$	86.8	31.1	32.5±1.1
		12.5	$h_{s,0} + h_{s,1} + h_{l,0.75}$	91.0	30.4	24.4±0.8

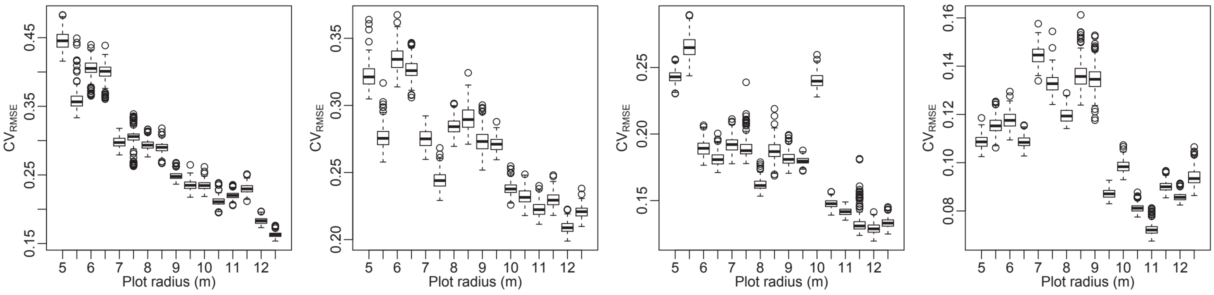
Figure 10.2 displays boxplots of prediction accuracy of multiple regression models as a function of plot radius. At Saint-Paul-de-Varces, Spearman correlation coefficient ρ between prediction accuracy (coefficient of variation of the $RMSE$ calculated by LOO cross-validation) and

plot radius is significantly different from zero for basal area only ($\rho = -0.87$, $p < 0.001$, two-sided test). Stem density displays a similar decreasing trend but prediction accuracy increases again when $r > 9$. For mean diameter the coefficient of variation of the $RMSE$ decreases when plot radius increases from 6 to 7.5 m, and then remains stable around 14%.

With Chablais data, all correlations are significant if all radii are taken into account (G, N and V: $p < 0.001$, D: $p < 0.01$, H: $p < 0.05$). However, when only radii between 5 and 10 m are included in the analysis, the correlation remains significant only for basal area. This suggests that the overall decreasing trend is somewhat driven by values obtained for larger plots.



(a) Basal area - StP (b) Stem density - StP (c) Mean diameter - StP (d) Volume - Chb



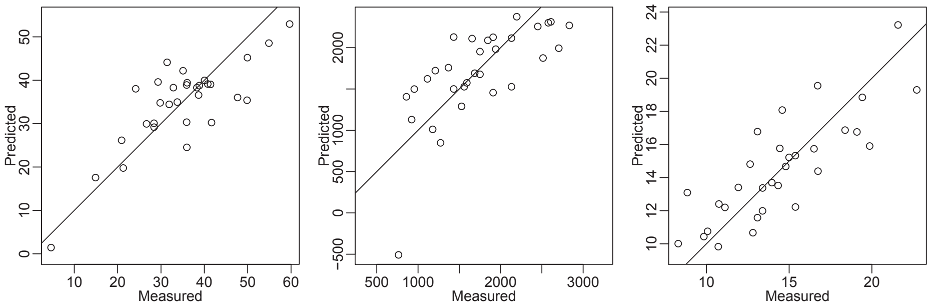
(e) Basal area - Chb (f) Stem density - Chb (g) Mean diameter - Chb (h) Dominant height - Chb

Figure 10.2: Boxplots of prediction accuracy (coefficient of variation of the $RMSE$ calculated by 200 ten-fold cross-validations) for basal area G, stem density N, mean diameter D, dominant height H and volume V depending on plot radius. Subfigures (a)-(c) correspond to Saint-Paul-de-Varces data (StP), subfigures (d)-(h) to Chablais data (Chb).

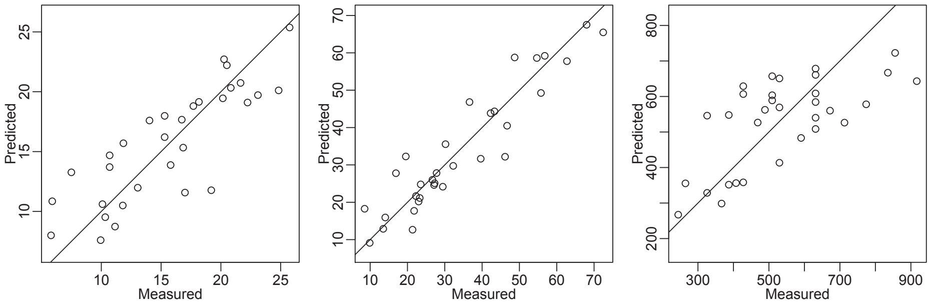
10.2.3 Prediction errors analysis for selected regression models

For the regression models obtained with the maximum radius of training plots in Saint-Paul ($r = 10$) and Chablais ($r = 12.5$), a short analysis of prediction errors is performed. Regarding the difference between values predicted by leave-one-out cross-validation and observed values, Wilcoxon signed-rank tests indicate that the null hypothesis (the distribution of the differences between predicted and observed values is symmetric about 0) can not be rejected.

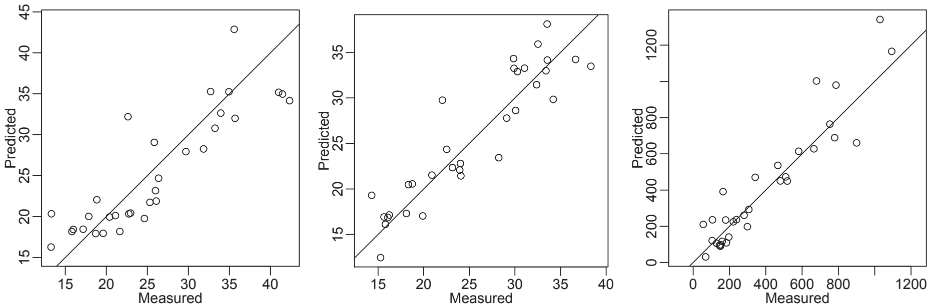
Figure 10.3 plots the predicted values against the observed values for basal area, stem density, mean diameter, dominant height and volume (Chablais only). For stem density in Saint-Paul, one negative value is predicted. This outlier is obtained for plot 21, which is located on a scree and displays large canopy gaps. For this plot, other stand parameters are better predicted (G: 1.5 for 4.6 $m^2.ha^{-1}$, D: 10.0 for 8.3 cm, H: 9.5 for 10.3 m). Otherwise, no clear trends are remarkable, expect that stem density values in Chablais seems overestimated for intermediate values, but underestimated for large ones. A similar pattern is hardly noticeable with stem density estimates in Saint-Paul.



(a) G - StP (b) N - StP (c) D - StP



(d) H - StP (e) G - Chb (f) N - Chb



(g) D - Chb (h) H - Chb (i) V - Chb

Figure 10.3: Predicted values plotted against field values for basal area G (m^2.ha^{-1}), stem density N (ha^{-1}), mean diameter D (cm), dominant height H (m) and volume V (m^3.ha^{-1}). Subfigures (a)-(d) correspond to Saint-Paul-de-Varces data (StP), subfigures (e)-(i) to Chablais data (Chb). Predicted values are estimated by leave-one-out cross-validation.

Table 10.4 presents prediction errors statistics. Bias is lower than 2.2 % in absolute value, expect for stem volume with -5.5 %. The widest range of relative errors is also obtained for volume with -81.3 to 60.2 %. The smallest is obtained for dominant height in Chablais with -30.4 to 19.1 %. In Saint-Paul, the smallest range for relative error is obtained for mean diameter. In coppice stands, where stems growing on stumps usually have the same age, diameter distribution is less variable than in irregular stands.

Table 10.4: Prediction errors statistics expressed in percentage of mean field value for basal area (G), stem density (N), mean diameter (D), dominant height (H) and stem volume (V). Prediction errors are estimated by leave-one-out cross-validation for the models described in table 10.3. Statistics are mean (Bias), minimum (Min), maximum (Max) and standard deviation (Sd).

Area	Radius	Variable	Bias	Min	Max	Sd
Saint-Paul-de-Varces	10	G	0.5	−39.8	41.5	19.7
		N	2.1	−40.2	73.2	25.9
		D	−0.4	−29.4	27.3	14.5
		H	−0.2	−37.0	47.5	19.8
Chablais	12.5	G	0.8	−37.2	40.9	18.2
		N	1.3	−41.0	50.8	23.4
		D	2.2	−36.9	31.6	15.7
		H	−1.3	−30.4	19.1	11.8
		V	−5.5	−81.3	60.2	30.4

10.3 Discussion

10.3.1 Accuracy of multiple regression models for stand parameter estimation

The ALS method produces accurate estimates of forest stand parameters, showing that it is suitable for mountainous forests such as coppice stands and complex, mixed forests. Results obtained with 10 m radius training plots in Saint-Paul are consistent with those obtained with 34 deciduous plots located in the Bavarian Forest National Park (Germany) [106]. In this study a multiple regression was also performed with forest parameters as dependent variables and laser metrics as independent variables. Plot size was 20×20 m^2 which is close to the area of the circular plots (314 m^2). Accuracy was similar for basal area (20.6%) and mean diameter weighted by basal area (13.2%). Stem density also yielded the greatest error with 29.8%, whereas a better result was achieved for dominant height with 8.1%.

In the same study, prediction accuracy was 44.0% for volume, 37.4% for basal area, 51.1% for stem density, 6.3% for dominant height and 16.8% for mean diameter weighted by basal area when all plots (coniferous, deciduous and mixed forest, $N_F = 108$, size 20×20 m^2) were considered. Results obtained in Chablais with 12.5 m radius plots (491 m^2) are better except for dominant height, with respectively 30.4, 18.0, 23.0, 11.7 and 15.6. Results obtained with 11 m radius plots (380 m^2) are respectively 29.6, 21.4, 23.9, 9.1 and 16.9%, which suggests that accuracy differences are probably linked to the wider range of forest structures in the Bavarian study, rather than training plot size. Indeed, stands from valley bottom, montane slopes and subalpine areas were altogether considered, whereas here a narrower range of altitude is considered (800-1330 m).

Another factor that may explain difference in prediction accuracy is the difference in metrics extraction and selection. More metrics were extracted in the Bavarian study, but for only two groups (first and last). Here the single echoes group is also considered, and metrics related to this group are often included in the best models. Moreover, variable selection was based on a stepwise selection. Some relevant metrics combinations might have been missed, whereas in this study all combinations are tested. Besides, the better result obtained in the Bavarian study for dominant height might be explained by the fact that height percentiles such as 90th or 95th have been reported to better model dominant height than the maximum height. Indeed, such metrics were often included in dominant height models whereas in our study maximum height metrics are seldom used.

10.3.2 Effect of plot radius on field observations

Comparison of field parameters values obtained with different radii show that the distributions of differences are not significantly asymmetric about 0. However the mean and standard deviation, as well as several distribution quantiles, are significantly correlated with plot radius for some of the forest parameters.

In Saint-Paul the median and mean of basal area are positively correlated with plot radius. Even tough special attention is paid to ensure that field inventories are not biased, operational constraints and stand characteristics in that case may explain such patterns, e.g.:

- need for a small stable platform to install and operate the tripod;
- minimum distance to large tree trunks to ensure visibility from the tripod and acceptable GPS signal;
- minimum distance to small trees and understory to unfold the 4 m GPS antenna;
- presence of compact groups of several stems in coppice stands.

The effect of the inclusion of a cleared area at the center of the plot to operate the mapping instruments may result in the underestimation of the total basal area for small radii, because the void area amounts to a significant portion of the plot. In Chablais where plot centers are defined afterwards when large plots are divided in smaller circular plots, this effect should not be observed. Indeed, the mean value of basal area is only slightly correlated ($\rho = 0.5$) and the median is not ($\rho = -0.36$).

The extreme values are affected by plot radius. Indeed with small plots, the radius is of the same order of magnitude as the elements of the forest stand: trees. Therefore, at such a small resolution, stand parameters are highly variable. To avoid this nugget effect, [69] suggests that plot size is chosen so that plots contain at least twelve trees. Minimum and first quartile values for parameters other than stem density are highly correlated with plot radius since their are mostly obtained in plots that have smaller trees and higher stem density. The local heterogeneity is less important so that stand parameters estimates are less influenced by plot size. Maximum values are obtained in plots with less but larger trees, so that stand parameters will vary greatly depending on the inclusion of any additional tree when plot radius is increased.

10.3.3 Consequences on models accuracy

It is noteworthy that prediction accuracy also varies greatly with plot calibration radius (figure 10.2). As the distribution differences between observed and predicted values are not significantly asymmetric about 0 for any of the plot radii, plot size may influence residuals variance or some higher order statistics. This is consistent with the previous observation that the presence of outliers in field observations is all the more likely since small plots exhibit wider range of forest parameters. Such data points may affect the coefficient of determination and prediction accuracy of multiple regression models.

These results are consistent with the findings of a study on the effect of plot size on stand properties estimated by LiDAR conducted in coniferous boreal stands [91]. In mature forests, $RMSE$ for volume and basal area models decreased from around 25% to 14% when plot size increased from 200 m^2 to 400 m^2. With the Chablais data, $RMSE$ for volume (resp. basal area) decreases from 39.3 to 29.4% (resp. 28.5 to 22.8%) when plot radius increases from 8 to 11.5 m (plot area increases from 201 to 415 m^2). Better overall accuracy in the boreal study may be explained by the fact that models were calibrated after stratification, i.e. with more homogeneous data. However both results show an important accuracy improvement when plot size doubles from 200 to 400 m^2. Such findings with real data are confirmed by comparisons of models accuracy obtained for simulated plots of size ranging from 314 to 1963 m^2 [84]. $RMSE$

for above ground biomass estimates of 60 simulated validation plots decreased from 47.1 to 38.5% when plot radius increased from 10 to 15 m (plot area increased from 314 to 707 m^2). As pointed out in [91], it is noteworthy that for basal area and volume, plot size has more influence than for dominant height.

Those two studies [84, 91] also emphasize the fact that plot size effect is enhanced by GPS positioning errors. For small radii, there are higher chances that the extracted point cloud is only partially located within the actual field plot, resulting in erroneously fitted linear models. However, the study on boreal forests reported that for plot position errors lower than 5 m, the median of the mean prediction error for basal area, volume and Lorey's height were similar. Given the estimated precision of our GPS geolocation, coregistration error is likely to have a small effect on models accuracy compared to plot size.

10.4 Conclusion and perspectives

With an area-based method, forest stand parameters such as dominant height, basal area, stem density and mean diameter can be precisely estimated from airborne laser scanning data for coppice stands and complex, mixed stands. Comparison of regression models obtained with plot radii ranging from five to twelve meters show that prediction accuracy depends on calibration plot size. Indeed, the spatial heterogeneity of mountainous stands results in a high variability of stand parameter estimates obtained from small field plots. This variability affects linear regression models prediction accuracy, mostly because of the effect of outlier data points. This effect is enhanced by biases in field data collection and approximate GPS coregistration.

Comparison with results obtained in previous studies suggests that model accuracy depends on the extracted and selected metrics. However, laser scanner hardware improvements make additional information regularly available, so that the extraction of relevant features for prediction models training might remain a perpetual task.

Accurate (sub-meter) positioning of plots would be problematic for Chablais data, and even more for Saint-Paul. Thus it will be impossible to compare the effect of co-registration approximation on models accuracy with respect to an ideally registered situation. However it would be possible to simulate additional registration noise in order to quantify the available margin regarding position accuracy and its possible dependence on plot size. Such information will be valuable when designing operational protocols for airborne laser scanner based inventories.

Chapter 11

Support vector regression models

The objective of this chapter is triple:

- compare prediction models based on ordinary least square multiple regression (*ols*-MR) and ϵ-support vector regression (ϵ-SVR);
- assess the effect of the use of different numbers and types of laser metrics on prediction accuracy;
- assess the relevance of dimension reduction of laser metrics prior to model training.

11.1 Methods

The general workflow described in section 9.2 is tested on the data from Saint-Paul-de-Varces and Chablais. The LiDAR data are presented in section 5.3. Field data are described in section 5.3.3.1. Only live trees are considered and minimum diameter is 5 cm in Saint-Paul-de-Varces and 7.5 cm in Chablais. Contrary to the chapter 10, all computations are done for 10 m radius plots. Besides, models are not calibrated for volume and dominant height, but only for basal area (G), stem density (N), mean diameter (D) and top height ($\mathcal{H}$, calculated as the height of the tallest sampled tree on the plot).

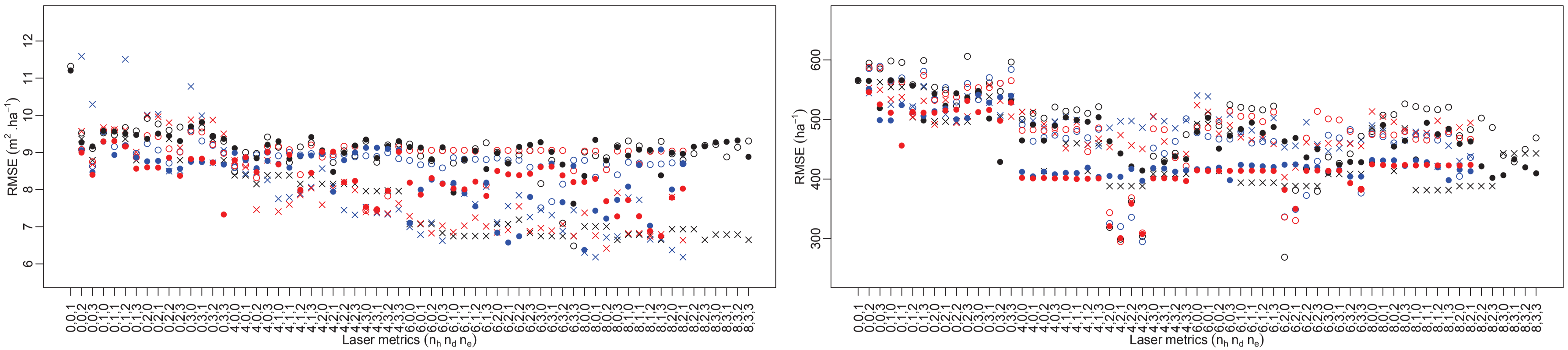
ols-MR and ϵ-SVR models are fitted. Regarding ϵ-SVR, linear $k(x, z) = \langle x, z \rangle$ and radial-basis $k(x, z) = \exp(-\gamma \|x - z\|^2)$ kernels are tested. Hyper parameters C and γ are selected by tuning over a range of pre-defined values and selecting the combination which yields the lowest root mean square error in leave-one-out cross-validation. Cost parameter C is a positive constant that defines the trade-off between training error and model flatness. Tested values are $C \in (10^i)_{i=-5,-4,\ldots,3}$. The same range is investigated for radial kernel γ parameter.

To evaluate the effect of the number and type of laser metrics on prediction accuracy, variables sets obtained by combination of $(n_h, n_d, n_e) \in (\{0, 4, 6, 8\} \times \{0, 1, 2, 3\} \times \{0, 1, 2, 3\})$ are tested. When the number of observations N_F is greater than the number of variables n_v, PCA and ICA are performed for dimension reduction (DR). Subsets of dimension $n_v \in \{1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24\}$ of the obtained principal and independent components are also tested as sets of variables.

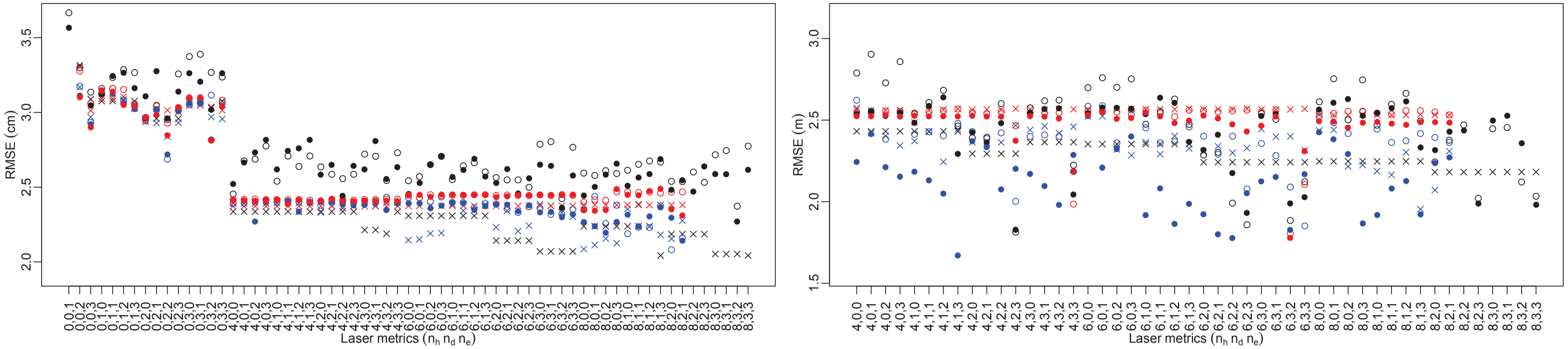
11.2 Results

11.2.1 Overview

Figure 11.1 and 11.2 provide an overview of prediction accuracy, depending on regression type, on dimension reduction method and on the number and type of laser metrics, for the Saint-Paul and Chablais data respectively.

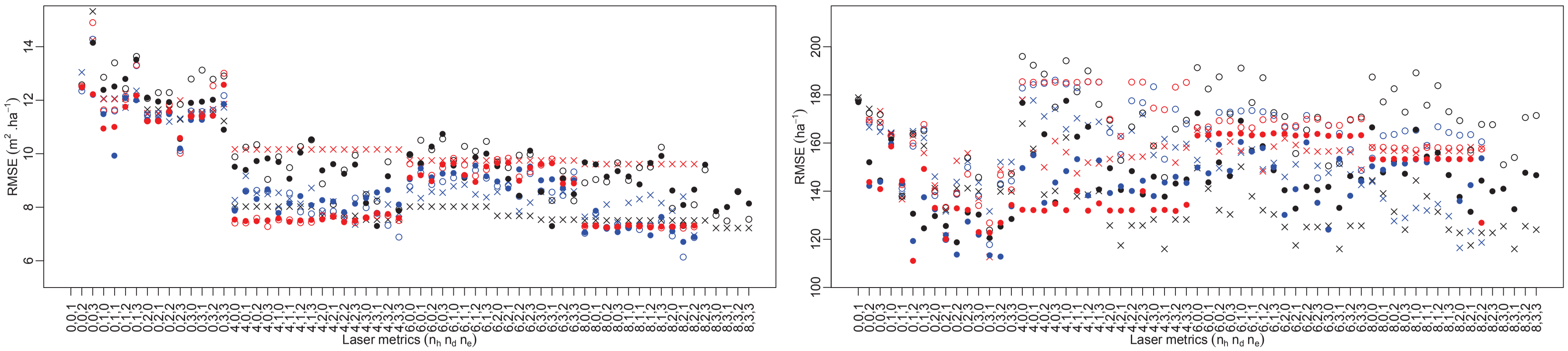


(a) Basal area (m^2.ha^{-1}) - StP (b) Stem density (ha^{-1}) - StP



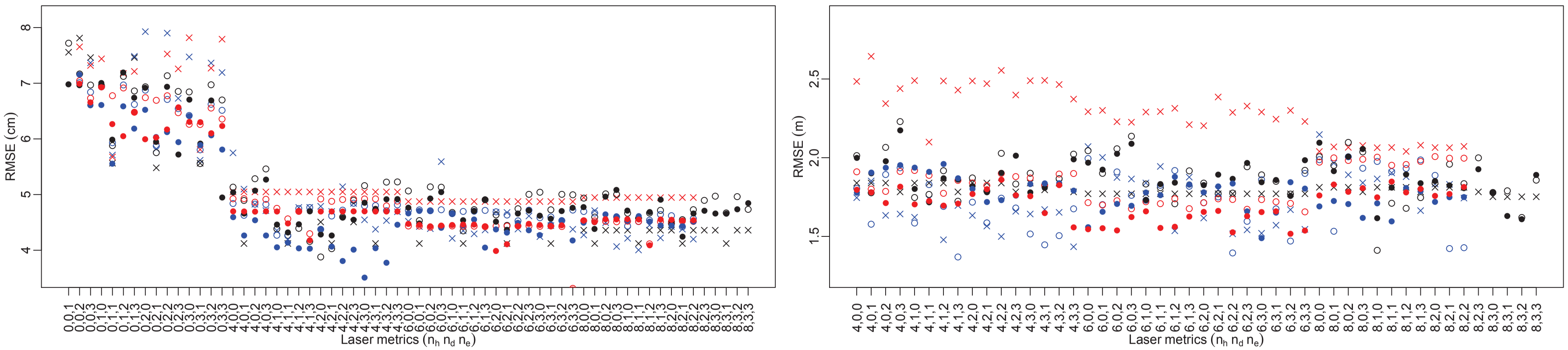
(c) Mean diameter (cm) - StP (d) Top height (m) - StP

Figure 11.1: Influence of the number and type of laser metrics on prediction accuracy ($RMSE$ obtained by leave-one-out cross-validation) of *ols*-MR (×) and ϵ-SVR with linear (○) and radial (●) kernels. Triplets on the x-axis refer to the number of height, density and entropy metrics (n_h, n_d, n_e) used to construct the variables sets. Symbol colors refer to the dimension reduction method: PCA (blue), ICA (red) or none (**black**). Data are from the Saint-Paul inventory. For top height, results with $n_h = 0$ are not presented as $RMSE$ is much higher than when height metrics are included.



(a) Basal area (m^2.ha^{-1}) - Chb

(b) Stem density (ha^{-1}) - Chb



(c) Mean diameter (cm) - Chb

(d) Top height (m) - Chb

Figure 11.2: Influence of the number and type of laser metrics on prediction accuracy ($RMSE$ obtained by leave-one-out cross-validation) of *ols*-MR (×) and ϵ-SVR with linear (○) and radial (●) kernels. Triplets on the x-axis refer to the number of height, density and entropy metrics (n_h, n_d, n_e) used to construct the variables sets. Symbol colors refer to the dimension reduction method: PCA (blue), ICA (red) or none (**black**). Data are from the Chablais inventory. For top height, results with $n_h = 0$ are not presented as $RMSE$ is much higher than when height metrics are included.

General remarks There is no clear common tendency across the different stand parameters. Moreover, for a given parameter, results obtained for Saint-Paul and Chablais data are also quite different. This suggests that optimized models are parameter- but also site-specific. For mean diameter and top height, results obtained without height quantiles are far poorer than with four or more height quantiles. This tendency is less pronounced for basal area, and stem density in Saint-Paul. Stem density in Chablais is the only case where results without height quantiles are similar or even better. With the Chablais data, accuracy obtained for radial kernel ϵ-SVR used on ICA variables mainly depends on the number of height metrics.

Basal area With Chablais data, poor results are obtained with less than four height metrics in the variable set. Regarding *ols*-MR, dimension reduction of the variables set reduces accuracy, particularly for ICA. Results obtained with PCA are intermediate and quite erratic. When no dimension reduction is performed, accuracy is better. Accuracy values are similar whether four or six height metrics are extracted. Results are improved when two or three density metrics are added to the variables set. When eight height metrics are used, at least three density metrics are needed to provide a further improvement of prediction accuracy.
For ϵ-SVR, better and more stable results are achieved with prior dimension reduction of the metrics set. Meanwhile, some good accuracy is obtained with a radial kernel for variables sets $(n_h, n_d, n_e) \in (\{4,6\} \times \{3\} \times \{1\})$. On the whole, linear and radial kernels display similar accuracy, especially when ICA is employed. The main factor influencing accuracy is the number of height metrics in the variables set. With $n_h = 4$, ICA obtains slightly better results than PCA. Accuracy decreases for both when $n_h = 6$, but is improved when $n_h = 8$, especially for PCA. With variables sets reduced by ICA, accuracy seems less sensitive to the number of other metrics.

With Saint-Paul data, trends are somewhat different. Results are similar when $n_h = 0$ but when more metrics are included, *ols*-MR models accuracy generally increases more than for ϵ-SVR. When no dimension reduction is used, *ols*-MR accuracy improves when the number of height metrics increases from four to six but then depends mainly on the number of additional density metrics. Best results are obtained for $(n_h, n_d) \in (\{6,8\} \times \{1,3\})$. For $(n_h, n_d) \in (\{6,8\} \times \{0,2\})$ an improvement is noticed when three entropy metrics are included. With $n_h = 4$, dimension reduction benefits to *ols*-MR models accuracy. The overall best results are obtained with PCA for $n_h = 8$ and $n_d = \in \{0,1\}$.
Overall, ϵ-SVR accuracy is lower than with *ols*-MR, save for $n_h = 0$. Except for a few cases, ϵ-SVR without dimension reduction and linear kernel ϵ-SVR accuracy remains stable. Radial kernel gives better accuracy. Results obtained with ICA are slightly poorer but seems less dependent on the number of laser metrics.

Stem density With Chablais data, it is the only case where best accuracy is obtained without any height quantiles. On the whole, linear kernel ϵ-SVR yields poor results, whatever the dimension reduction. Radial kernel ϵ-SVR gives better results, especially when no height quantiles are extracted. Radial kernel used with ICA gives good and stable results when $n_h = 4$, but accuracy decreases when more height quantiles are extracted. With PCA or without dimension reduction, radial kernel yields very variable results. Accuracy seems better when at least two density metrics are present.
For *ols*-MR, accuracy obtained with ICA is poor, except for the case $(n_h, n_d, n_e) \in \{0\} \times \{2,3\} \times \{1\}$. Best results are obtained without dimension reduction, especially when one entropy metric is combined with at least two density metrics. Some good accuracies are obtained with PCA when $n_h \geq 8$.

With Saint-Paul data, linear kernel also gives poor results, especially with no dimension

reduction of variables sets with $n_d = 1$, and except for combinations of four or six height quantiles with two density metrics. ϵ-SVR models accuracy is poorer when no dimension reduction is performed, even tough results are slighlty better when at least two density metrics are included. With radial kernel, results obtained with ICA and PCA are similar and hardly dependent on the number of laser metrics, save the above mentioned combinations. Besides, accuracy decreases when the total number of metrics increases.
For *ols*-MR, dimension reduction does not allow any accuracy improvement. Good results are obtained without dimension reduction when two density metrics are added to height quantiles. Extracting three density metrics yields lower accuracy, whereas with only one density metric combined to eight height metrics, accuracy is better, suggesting the existence of combined effects of laser metrics of different types.

Mean diameter With both datasets, results are improved when height metrics are included in the variables set. With Chablais data, accuracy obtained with ICA components displays similar trends for ϵ-SVR and *ols*-MR. Indeed, prediction errors are close and depend mainly on the number of height metrics. Slightly better results are obtained with $n_h = 6$. Without dimension reduction or with PCA, SVR accuracy is close to ICA results but more variable, except when $n_h = 4$. Indeed, for variables sets based on four height metrics, PCA gives the better results although the effect of the other metrics is not easily explainable. Regarding *ols*-MR, best results without dimension reduction are obtained when one entropy metric is used. PCA yields similar prediction errors.

With Saint-Paul data, ϵ-SVR without dimension reduction gives poor results. Prediction error obtained with ICA hardly depends on the included laser metrics and on the regression method. Regarding ϵ-SVR, linear and radial kernels yield similar accuracy. Results obtained with ICA and PCA are fairly similar when $n_h = 4$, but accuracy tends to decrease (resp. increase) for ICA (resp. PCA) when more metrics are included in the variable sets. However accuracy is better for radial ϵ-SVR with ICA for $(n_h, n_d) \in \{6, 8\} \times \{0, 2\}$.
For *ols*-MR, accuracy mainly depends on the number of height and density metrics. Results with a higher number of height metrics are better, and are further improved when more than two density metrics are added. However, accuracy for $(n_h, n_d) = (6, 2)$ is better than for $(8, 2)$. Entropy metrics have no effect, save for the $(n_h, n_d, n_e) = (8, 1, 3)$ and $(4, 3, 3)$ cases, where accuracy is respectively increased and decreased.

Top height With Chablais data, *ols*-MR models obtained without dimension reduction yield almost the same prediction error (1.75 to 1.84 m). With ICA, *ols*-MR models are less accurate, even tough $RMSE$ decreases down to 2 m when $n_h = 8$. With PCA, results are more variable but better results are achieved ($RMSE = 1.43$ m). With ϵ-SVR, accuracy achieved without dimension reduction is poor and very variable, especially when no density metrics are included. With PCA, results are better but also more variable, especially for linear kernel. With ICA, radial and linear kernels display similar tendencies, the former being slightly more accurate. Prediction error is globally linked to the number of height metrics, and is worse for $n_h = 8$.

With Saint-Paul data, ICA performed poorly for all methods, except for a few metrics combinations. For *ols*-MR without dimension reduction, accuracy is increased when at least two density metrics are added to the height metrics. However when $n_h = 4$, increasing n_d from two to three results in higher error. When PCA is used some better results are achieved but they are also more variable. For ϵ-SVR, PCA combined with radial kernel yields good but highly variable accuracy. When no dimension reduction is performed, error is higher, except for some particular cases which include combinations of $(n_d, n_e) \in \{2, 3\} \times \{3\}$.

11.2.2 Best settings of each regression method

Comparison of regression methods Table 11.1 details the settings that allow the lowest prediction error for each stand parameter and type of regression. For each case, the prediction error ($RMSE$) obtained by leave-one-out cross-validation is printed, as well as the mean value obtained with 200 ten-fold cross-validations. Values predicted by LOO cross-validation are plotted against the field values in figure 11.3. When only the leave-one-out results are considered, obtained accuracies for most cases are similar. ϵ-SVR seems even more accurate for mean diameter in Chablais and stem density in Saint-Paul. Indeed, error differences are lower than 1.6%, except for mean diameter in Chablais (2.6% difference between radial ϵ-SVR and *ols*-MR) and stem density in Saint-Paul (6.5% difference between linear ϵ-SVR and *ols*-MR). Some similarities in the extracted metrics that yield the best accuracy for a given stand parameter are noticeable with the Chablais data.

However, when looking at mean prediction error and its standard deviation estimated by 200 ten-fold cross-validations, *ols*-MR appears to be more robust. Indeed, standard deviation for *ols*-MR is below 0.7% except for stem density in Chablais where ICA has been performed for dimension reduction. For ϵ-SVR, standard deviation value reaches 6.5 and even 10.9% for basal area in Chablais (for this latter case, computation was particularly slow which might indicate troubles during model fitting).
The location shift of the distributions of the $RMSE$ obtained in ten-fold cross-validation for *ols*-MR and linear ϵ-SVR is significantly different from 0 (Mann-Whitney test, $p < 0.001$), except for stem density in Saint-Paul. The test is also significant for *ols*-MR compared to radial ϵ-SVR, except for dominant height in Saint-Paul. When linear ϵ-SVR and radial ϵ-SVR are compared, location shifts are significantly different from 0 only for mean diameter, and for basal area in Saint-Paul ($p < 0.001$). In all other cases, the null hypothesis that the location shift is equal to zero cannot be rejected ($p > 0.1$). Standard deviation values are also higher for the prediction error obtained by ϵ-SVR, compared to those obtained by *ols*-MR, except for stem density predicted on the Chablais data.

Comparison with the variables set $(n_h, n_d, n_e) = (6, 3, 0)$ With the variables set $(6, 3, 0)$ without entropy metrics or dimension reduction, the obtained prediction errors for basal area, stem density, mean diameter and top height are 23.3, 23.7, 27.3 and 6.4% for Chablais data, and 19.4, 25.5, 14.3 and 12.6% for Saint-Paul [1]. By using different laser metrics and dimension reduction methods, *ols*-MR models are improved by 0.2 to 3.5 points of percentage with respect to the basic approach relying only on height and density metrics and multiple regression. The higher improvement for mean diameter in Chablais is due to the fact that the original model was particularly bad, probably because of the low number of tested models complying with the linear model assumptions.

---

[1] Values are different from those presented in section 10.2 because top height is considered instead of dominant height, and with the Chablais data $RMSE$ is assessed by leave-one-out cross-validation whereas a "leave-one-plot-out" cross-validation (see details in section 9.2.5) was used previously. $RMSE$ is lower when the second procedure is employed with the Chablais data as in leave-one-out cross-validation, models are fitted with subplots that are spatially correlated to the tested remaining subplot. As our objective is not to obtain absolute values but to compare different methods, this effect is secondary. Meanwhile, support vector regression is less sensible to outliers, so that it should theoretically be less affected than multiple regression by a leave-one-plot cross-validation. It is also noteworthy that prediction error for top height is much lower than for dominant height.

Table 11.1: Best prediction accuracy obtained with *ols*-MR and ϵ-SVR (linear and radial kernels) and corresponding variables sets, for basal area (G), stem density (N), mean diameter (D) and top height ($\mathcal{H}$). Values are the coefficients of variation of the $RMSE$ obtained by leave-one-out (LOO) and 200 repetitions of ten-fold (10-F) cross-validations (mean ±standard deviation).

Site	Parameter	Regression method	CV_{RMSE} LOO (%)	CV_{RMSE} 10-F (%)	Laser metrics (n_h, n_d, n_e)	Metrics or components used
Saint-Paul	G	*ols*-MR	17.8	17.9±0.5	8,2,1	PCA 10,17,20
		linear ϵ-SVR	18.6	21.0±4.3	6,3,3	none
		radial ϵ-SVR	18.3	20.4±1.9	8,0,0	PCA 1-12
	N	*ols*-MR	22.0	22.2±0.6	8,1,·	$h_{f,0.67} + h_{l,0.83} + d_{s,0.5}$
		linear ϵ-SVR	15.5	22.8±4.3	6,2,0	none
		radial ϵ-SVR	17.3	24.9±4.4	4,2,1	ICA 1-19/19
	D	*ols*-MR	14.1	14.2±0.4	8,{1,3},3	$h_{l,0.33} + d_{l,0.5} + e_{0.66-1}$
		linear ϵ-SVR	14.4	17.6±2.0	8,2,0	PCA 1-24
		radial ϵ-SVR	14.8	17.3±1.9	8,2,1	PCA 1-20
	$\mathcal{H}$	*ols*-MR	11.0	11.0±0.3	8,1,3	PCA 3,26
		linear ϵ-SVR	10.1	20.1±6.5	6,3,2	ICA 1-24/24
		radial ϵ-SVR	9.4	11.6±2.2	4,1,3	PCA 1-10
Chablais	G	*ols*-MR	21.5	21.6±0.7	8,2,2	PCA 1,8,9
		linear ϵ-SVR	19.0	23.4±10.9	8,2,1	PCA 1-10
		radial ϵ-SVR	20.7	22.8±2.2	8,2,1	PCA 1-10
	N	*ols*-MR	21.3	34.0±6.8	0,3,1	ICA 5,6,8/8
		linear ϵ-SVR	22.3	22.6±0.6	0,3,1	PCA 1-6
		radial ϵ-SVR	21.0	22.0±1.4	0,1,2	ICA 1-2/2
	D	*ols*-MR	15.8	16.0±0.4	8,1,1	PCA 7,20,22
		linear ϵ-SVR	13.1	25.4±4.5	6,3,3	ICA 1-28/28
		radial ϵ-SVR	13.0	26.1±5.5	6,3,3	ICA 1-28/28
	$\mathcal{H}$	*ols*-MR	5.2	5.2±0.1	4,3,3	PCA 7,10,12
		linear ϵ-SVR	4.9	5.9±0.8	4,1,3	PCA 1-14
		radial ϵ-SVR	5.4	6.1±0.7	6,3,0	PCA 1-16

11.3 Discussion

11.3.1 Influence of the number and type of laser metrics

Numerous studies have shown that the upper height quantiles are a good indicator of tree height. This is exemplified by the fact that prediction accuracy for top height models is greatly improved when height metrics are included in the variables set. Previous studies [122, 228] showed that dominant height and, to a lesser extent, mean diameter are correlated with the upper height quantiles. This may explain why those stand parameters display similar tendencies regarding prediction models: stability of results when enough height metrics are included in the variables, and inefficiency of DR for *ols*-MR. Relationships between basal area or stem density and forest structure are more complex, resulting in patterns that are more difficult to interpret. The case of stem density in Chablais is an exception, as good accuracies are achieved without any height metrics. Indeed, a great part of training field plots are irregular stands, where some high, mature trees co-exist with a more or less dense lower stratum. With such structures, top height exhibits lower correlation with stem density. On the contrary, in Saint-Paul, most plots are constituted by even-aged coppice stands. In the absence of thinning or logging, height growth goes along with a decrease in stem density as a consequence of stem exclusion, which may explain why height metrics are successful in predicting stem density.

Entropy metrics have little impact with Saint-Paul data. Indeed, coppice stands are very dense and quite homogeneous at local scale, so that there is only little information to be extracted from the horizontal distribution of points. On the contrary, in Chablais numerous stands are irregular so that they display a high horizontal variability within a given plot. That may explain why adding entropy metrics helps increase prediction accuracy for stem density or mean diameter. The number of height bins for entropy computation also seems to be a major criterion as in Chablais prediction accuracy is improved when entropy is calculated with the whole range of height. In Saint-Paul, for top height the extraction of three entropy values benefits to ϵ-SVR models. This suggests that depending on the forest structure and on the forest parameter, the relevant entropy information may be located within the totality or only parts or the point cloud. In Saint-Paul, some broadleaved or coniferous trees originated from seedlings emerge from the homogeneous coppice canopy in some plots. That may explain why entropy metrics are relevant only when computed for selected height bins, whereas in Chablais all layers are heterogeneous.

Density metrics effect greatly depends on the forest parameter. For top height they have nearly no influence. For the other forest parameters, they improve prediction accuracy, especially for mean diameter in Saint-Paul. It is also noteworthy that depending on the site and on the forest parameters, the number of extracted metrics also has an effect on accuracy. For example, with basal area in Saint-Paul, *ols*-MR models without dimension reduction have better accuracy with $n_d = 1$ and $n_d = 3$ than with $n_d = 2$, suggesting that the relevant information is in the metric $d_{g,0.5}$. On the contrary, with mean diameter in Saint-Paul, *ols*-MR models without dimension reduction have better accuracy when more density metrics are included in the models, which may be explained by the fact that more valuable information is extracted when more height bins are considered.

Some combined effects are also noticeable. For example, *ols*-MR models without dimension reduction display better accuracy when $n_h = 6$ and n_d increases from one to two. However, when $n_h = 8$, accuracy decreases when the number of density metrics comes from one to two. This could be linked to the fact that metrics of different type might encompass some redundant information.

11.3.2 Influence of regression method

On the whole, *ols*-MR performs better with raw laser metrics, whereas ϵ-SVR yields better accuracies with reduced sets of features. This might be linked with the fact that regarding *ols*-MR a selection of the best model with at most three variables is performed, whereas the whole variables set is used in the case of ϵ-SVR. The training with a lower number of components (from ICA or PCA) might make models more accurate than when the whole set of raw metrics is used. Indeed, several SVR models that yield the best accuracy (table 11.1) are obtained with sets with a lower number of components than originally included in the raw metrics.

Radial and linear kernels sometimes yield very similar results (mean diameter in Saint-Paul), but ϵ-SVR with radial kernel seems more robust to the dimension reduction method and to the laser metrics. This could be expected as the radial kernel can account for non-linear relationship. However, the small difference in achieved accuracy suggests either that relationships between laser metrics (or extracted components) can be considered linear, or that radial kernel models might actually be over-fitted.

This hypothesis of over-fitting might also explain why ϵ-SVR models have performances similar to *ols*-MR when only leave-one-out cross-validation results are considered, but are worse both in terms of mean and standard deviation of $RMSE$ when ten-fold cross-validations are repeated. During the ten-fold cross-validation, hyperparameters are not tuned again, which suggests that the model determined in LOO cross validation might actually be over-fitted. A better parametrization of the ϵ-SVR algorithm might be attained by tuning the ϵ size of the insensitive tube. Previous studies suggest that this value should be set according to the noise level, which is unfortunately difficult to estimate.

11.3.3 Influence of dimension reduction

The fact that *ols*-MR performs better with raw laser metrics also suggests that most of the valuable information lies within a limited number of metrics, and that this information is not transferred to or usable within the components extracted with PCA or ICA. Indeed, PCA aims at maximizing the amount of variance encompassed in the orthogonal components, but there is no reason why the information related to the forest parameter should be linked to these features. Indeed, the results obtained with PCA in table 11.1 seldom use the first components. For example, upper height quantiles are correlated to some forest parameters, and the valuable information that is originally present in one or two laser metrics might be diluted into several components during dimension reduction. Concerning ϵ-SVR, it has already been mentioned that the results are better with dimension reduction. Moreover, using ICA is on the whole less accurate but somewhat more robust than PCA.

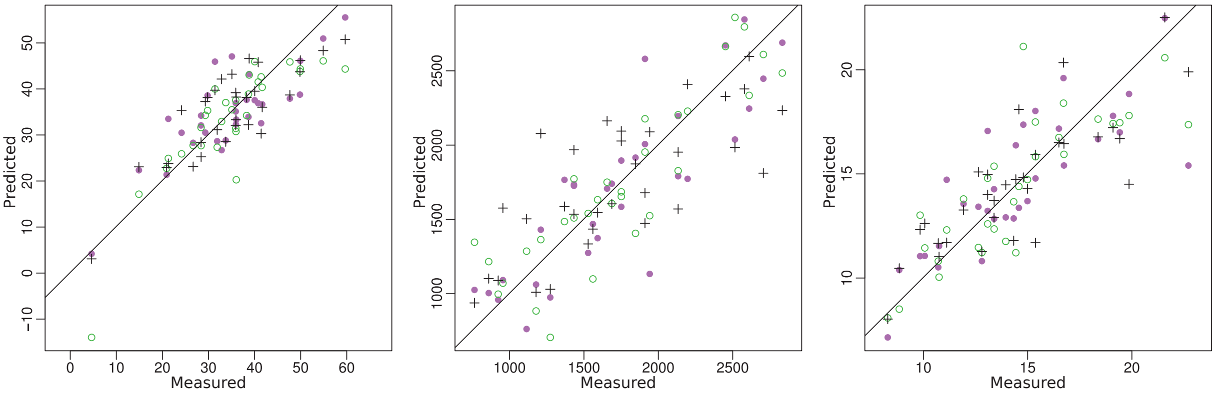
Besides, using dimension reduction methods might turn out tricky in operational applications. Laser metrics computed in a given location (e.g. plot) are independent from the rest of the data. Meanwhile, PCA and ICA components for the laser metrics in a given plot will depend on the number and location of other plots. If a trained model is to be applied for prediction purposes on the whole LiDAR dataset (which is the objective of operational applications), laser metrics and components have to be previously calculated for the whole area. The resulting computational task might be problematic if several metrics combinations or number of components are tested. Besides, it is all the more important for ICA since components are not ordered. As a result, even if the transformation than extracts components from the raw data is similar whether the whole data is considered or not, components selected for multiple regression may not come out in the same place in two successive independent components analyses. This is probably why the $RMSE$ estimated with ten-fold cross-validation in table 11.1 differs greatly from the LOO value obtained with *ols*-MR for stem density in Chablais. The order of ICA components should not matter here for ϵ-SVR as all extracted component are systematically used in the fitting process.

11.4 Conclusion and perspectives

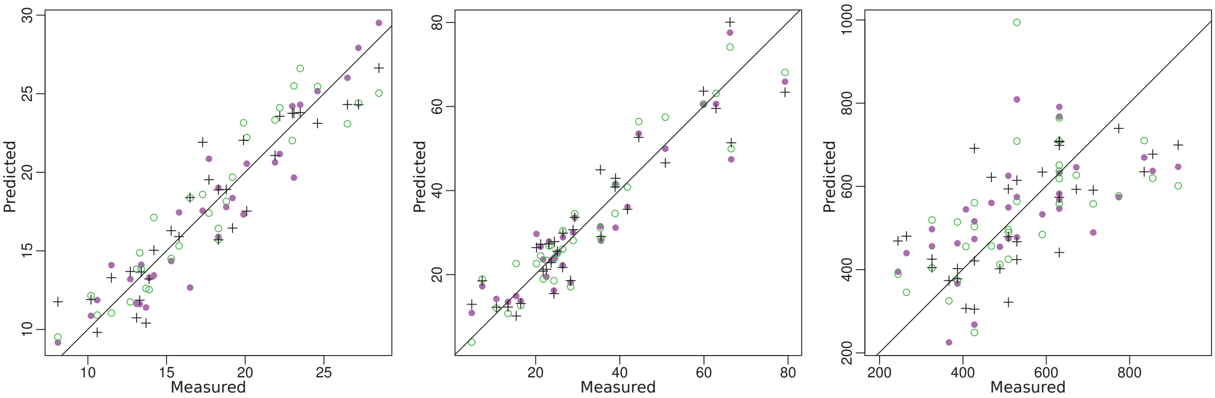
Upper height quantiles are generally the most important features, but additional relevant metrics depend on both the forest structure and the forest parameter, which suggest that prior stratification of the forest structure is a important requirement before accurate models are trained.
Finally the basic ordinary least squares multiple regression turns out to be the most accurate and robust for forest stand parameters estimation in this study. ϵ-support vector regression obtained almost similar performance but results are less robust. Multiple regression seems to take advantage of the selection of a few relevant metrics, such as those upper height quantiles. Support vector regression has trouble handling the whole raw metrics sets, and performs better on a reduced number of components, mainly from PCA but with less success. Despite its theoretical advantages, the radial kernel is only slightly superior to linear kernel.

Further research should focus on factors that may improve support vector regression. The first possibility is to improve SVR model parametrization, for example by using a different formulation of the algorithm, such as (ν-SVR). In that case, the ϵ parameter is replaced by the parameter ν, which affects the number of support vectors of the model, and might be easier to tune. The second possibility is to adapt the dimension reduction technique. Indeed, feature selection has not been tested here. In a study where the same selected laser metrics were used in multiple and support vector regression models, the latter was more accurate [224]. This suggests that feature selection would be more adapted for model training than using the whole laser metrics or than mapping them into a lower dimensional feature space.

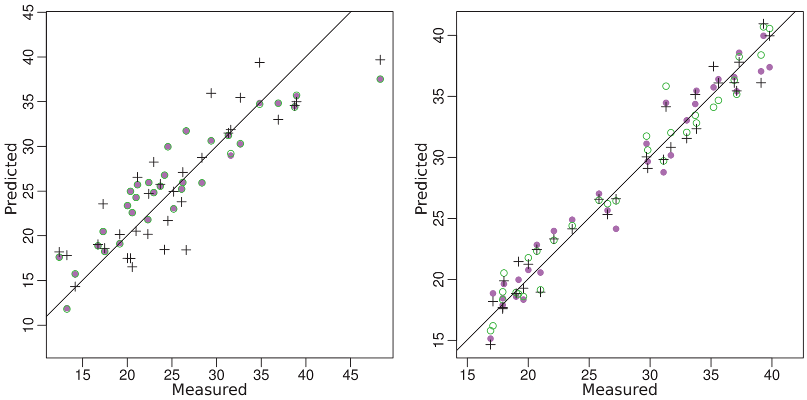
Besides when more observations are available, linear model assumptions might be less frequently fulfilled, so that non linear regression method such as radial kernel SVR will really be at their advantage.



(a) G (m^2.ha^{-1}) - StP (b) N (ha^{-1}) - StP (c) D (cm) - StP



(d) $\mathcal{H}$ (m) - StP (e) G (m^2.ha^{-1}) - Chb (f) N (ha^{-1}) - Chb



(g) D (cm) - Chb (h) $\mathcal{H}$ (m) - Chb

Figure 11.3: Predicted values plotted against field values for basal area G, stem density N, mean diameter D and top height $\mathcal{H}$ for each model described in table 11.1. Subfigures (a-d) correspond to Saint-Paul-de-Varces data (StP), subfigures (e-h) to Chablais data (Chb). Symbols correspond to the regression method: + *ols*-MR, ○ linear and ● radial ϵ-SVR. Predicted values are estimated by leave-one-out cross-validation.

Conclusion

This thesis dealt with the use of airborne laser scanning (ALS) data to characterize forest mountains. The aim was to improve methods used to derive forest information from small footprint airborne laser scanning data and test them on stands located in the french Alps.

A prerequisite for these investigations was to address the issue of the field data required to build and validate the models. Indeed, the accuracy of ALS data calls the protocols used for reference data collection into question. Due to the constraints of fieldwork in mountainous environments, a trade-off between accuracy and labor-intensity was adopted for the collection of the field data. Finally, a total of approximately 5.7 ha of forest plots were fully calipered within large plots (more than 0.25 ha). Plots represent a wide range of forest structures, from monospecific coniferous plantations to mixed, multilayered stands. Moreover, 31 small plots representing a total of 1 ha were inventoried in a coppice-dominated forest.

Single-tree approaches are designed to extract the information about the position and size of individual trees. In mountain forests, the complexity of the horizontal and vertical structure of stands makes the parametrization of detection algorithms difficult. Here, a method based on unsupervised training was implemented.

1. A framework for the automatic assessment of detection performance was proposed. It is necessary for the evaluation of different parametrization options, both at the training and validation steps. It is based on the comparison of the position of the treetops detected in ALS data and those measured on the field. Uncertainty in field and ALS data are taken into account. A unique criterion that balances true detections with false positives is used, which allows for a fast and consistent comparison of results.

2. The treetop detection algorithm itself consists in successive image processing filters (morphological and low-pass) designed to remove undesired features from the digital surface model derived from the point cloud. Treetop positions are extracted by local maxima filtering and a final selection is performed, based on the relationship between treetop height and distance to its neighbor. The sensitivity analysis showed that specific settings are required to achieve optimal detection performance on homogeneous plots whereas on heterogeneous plots detection requires an important trade-off between true and false positives. Globally, detection performance turned out to be very sensitive to parametrization and forest structures.

3. The unsupervised training approach aims at taking advantage of the existence of reference data where the performances of the algorithm are known. In the case of new ALS data, it is compared to the ALS data of the reference plots in order to decide which algorithm setting is most relevant. For matching of the test data to the training data, two indices have been tested separately: the vertical distribution of laser points, which reflects the vertical structure of the forest, and the spectrum of the Fourier transform of the canopy height model, which can been seen as a texture descriptor of the image used for treetop detection. Cross-validations showed that when the training set size is large enough, detection performance with unsupervised training is similar to the one obtained with an average setting of the algorithm.

This framework for unsupervised training is promising and can be applied to other detection algorithms. However, larger training sets have to be tested in order to evaluate its limits. This approach is a good example of the growing convergence between single-tree and area-based methods. Indeed, it can be seen as a most similar neighbor approach for the selection of (detection) parameters according to area-based descriptors. Other descriptors (e.g. quantiles) or imputation techniques already used for area-based methods could be tested. Moreover, the size of the plots makes the use of texture indices possible. Another perspective is to use the information from the training data to improve the stand estimates derived from the attributes of detected trees.

The area-based method based on the calibration of regression models was successfully tested on two datasets: a coppice forest and a mixed, complex forest.

1. Investigations on the effect of training plot size on ordinary least squares multiple regression models showed that plot size has a major effect on model accuracy. Moreover, for small plots, fieldwork constraints might be responsible for systematic biases.
2. Comparisons of ϵ-support vector regression (SVR) and ordinary least squares multiple regression showed that, even tough SVR models are sometimes more accurate, multiple regression is more robust.
3. Depending on the forest parameter, but also on the study area, the ALS metrics selected in the models varied greatly. Using principal component analysis and independent component analysis for previous dimension reduction of the variables sets did not benefit to multiple regression models accuracy, suggesting that the relevant information is not available in the extracted features. Besides, those transformations are not space- and time-independent, which makes them more difficult to use for operational applications.

Meanwhile, these findings have to be confirmed by testing the workflow with larger datasets, as support vector regression might handle cases with non-linearity in a better way. Besides, SVR may also be be more efficient when performed after a laser metrics selection.

As a final conclusion, the perspectives toward operational applications should be highlighted. The results obtained with both approaches, single-tree and area-based, show that models are highly dependent on the forest structures. Developing stratum-specific models requires all the more field data. An alternative would be to include environmental factors in the models, such as altitude and landform patterns, which can be computed from the ALS data as well. Efficient use of ALS in forestry practices involves taking advantage of any available information by integrating it into the whole workflow of forest inventory.

Appendices

Appendix A

Thesis publications

This appendix lists the publications co-authored during the thesis.

Journal papers

- J.-M. Monnet, J. Chanussot, F. Berger, "Support vector machines regression for the estimation of forest stand parameters using airborne laser scanning," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 580–584, May 2011.

- N. Clouet, F. Berger, J.-M. Monnet, L. Descroix, "CARTUVI : un modèle sous SIG pour la cartographie des surfaces débardables en zone de montagne," *Revue Forestière Française*, vol. 62, no. 2, pp. 155–170, 2010.

Conference papers

- J.-M. Monnet, É. Mermin, J. Chanussot, F. Berger, "Tree top detection using local maxima filtering: a parameter sensitivity analysis," *Proceedings of Silvilaser 2010, the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems, September 14-17, Freiburg, Germany*, 9 p., 2010.

- J.-M. Monnet, É. Mermin, J. Chanussot, F. Berger, "Estimation of forestry parameters in mountainous coppice stands using airborne laser scanning," *Proceedings of Silvilaser 2010, the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems, September 14-17, Freiburg, Germany*, 8 p., 2010.

- J.-M. Monnet, J. Chanussot, F. Berger, "Support vector machines regression for estimation of forest parameters from airborne laser scanning data," *Proceedings of IGARSS 2010, IEEE Geoscience and Remote Sensing Symposium, July 25-30, Honolulu, Hawaii, USA*, pp. 2711–2714, 2010.

- J.-M. Monnet, É. Mermin, J. Chanussot, F. Berger, "Using airborne laser scanning to assess forest protection function against rockfall," *Proceedings of the Interpraevent International Symposium in Pacific Rim, Taipei, Taiwan*, pp. 586–594, 2010.

- J.-M. Monnet, N. Clouet, F. Bourrier, F. Berger, "Using geomatics and airborne laser scanning for rockfall risk zoning: a case study in the French Alps". *Proceedings of the 2010 Canadian Geomatics Conference and Symposium of Commission I, ISPRS, Calgary, Alberta, Canada, June 15-18*, 5 p., 2010.

Technical papers (non peer-reviewed)

- N. Clouet, J.-M. Monnet, F. Berger, “SIG et LiDAR pour la réalisation de plans de prévention du risque de chute de blocs : application dans les Alpes françaises,” *Géomatique expert*, no. 82, pp. 52–57, 2011.
- J.-M. Monnet, N. Clouet, É. Mermin, F. Berger, “Le scanner laser aéroporté en zone de montagne : localiser la ressource forestière et évaluer ses conditions de mobilisation,” *Rendez-Vous Techniques de l’ONF*, no. 31, pp. 49–54, 2011.
- N. Clouet, J.-M. Monnet, F. Berger, “Estimation du volume de bois exploitable en montagne par scanner laser aéroporté (LiDAR),” *Géomatique expert*, no. 70, pp. 32–37, 2010.

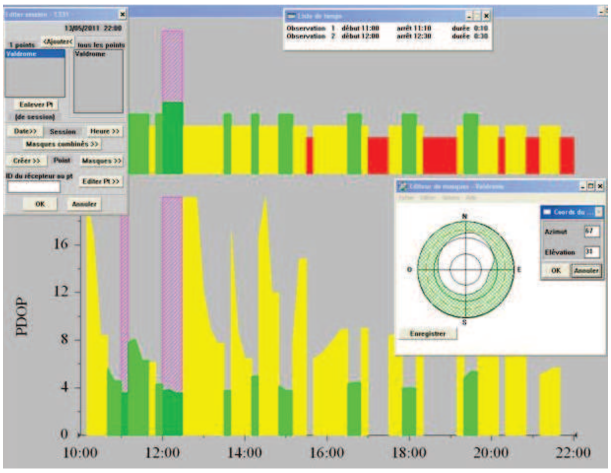
Appendix B

GPS elevation mask computation with digital terrain models

B.1 Motivations

In mountainous areas, periods with satisfactorily GPS signal quality are rare. Indeed, the topographic mask considerably limits the number of visible satellites and the signal quality (positional dilution of precision, PDOP). Besides, moving in such environments is quite slow so that is it time-consuming to both get back to a plot when previous GPS acquisition failed and to stand waiting for a sufficient number of available satellites after other field tasks have been completed.

However, one can take advantage of prior knowledge of the plot location, even approximate, and of the digital terrain model (DTM) to improve GPS acquisition productivity. Indeed, GPS softwares can output forecasts of signal quality at any given location, based on the scheduled satellite map [296, p. 190]. Such forecast can be refined by taking into account the elevation mask effect on satellite availability (figure B.1). Such masks are determined on the field with a compass and a clinometer, but in this appendix a quick way to roughly estimate it from DTMs is proposed.



Observation 1 début 11:00 arrêt 11:10 durée 0:10
Observation 2 début 12:00 arrêt 12:30 durée 0:30

Azimut 67
Elévation 31

Figure B.1: Screenshot of the utility that displays the theoretical PDOP and number of available satellites given the location and elevation mask. In that case, the total period of acquisition with six satellites and a PDOP lower than four is only 40 non consecutive minutes for the whole daytime.

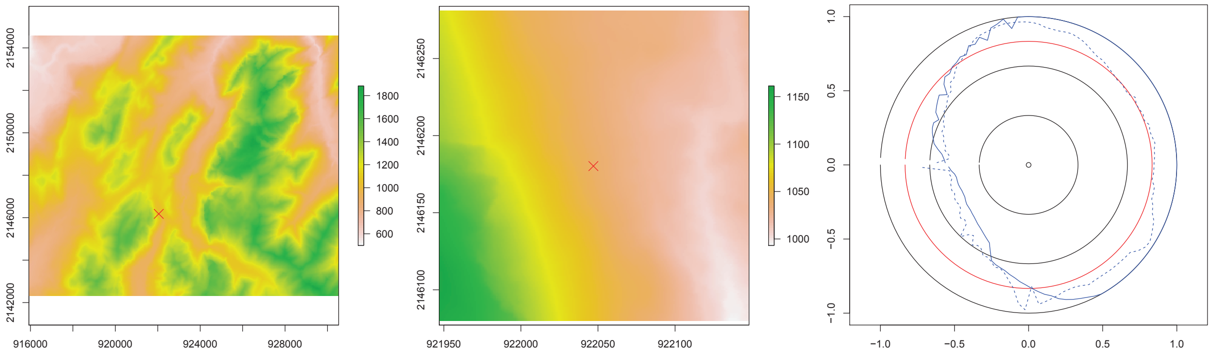
B.2 Method

When a given field plot location L with coordinates (x_L, y_L) is already known, either because a coarse GPS acquisition was made during a previous campaign, or because the sample plot coordinates are chosen in the office, the 50 m DTM of BD Alti® (figure B.2a) is used to compute a long-distance elevation mask of the topography. First z_L, altitude of L, is estimated by bilinear interpolation of the neighboring points of the DTM. Then the coordinates of the DTM pixel centers (P_i) are converted from the matrix Cartesian system (x_i, y_i, z_i) to the spherical coordinate system centered on L: $(r_i, \theta_i, \varphi_i)$. The pixel value is used as altitude for the corresponding pixel center. A rough elevation mask $M_r = (\theta_p)_{p \in \{1 \ldots n\}}$ is computed by retaining for each azimuth step the maximum value of the elevation angle of the pixel centers situated in this section S_p:

$$\theta_p = max(\theta_j)_{P_j \in S_p} \qquad \text{with } S_p = \left\{ P_j \backslash \frac{2\pi p}{n} \leq \varphi_j < \frac{2\pi(p+1)}{n} \right\} \tag{B.1}$$

This rough mask does not account for the short-distance mask induce by high frequency variations of the topography such as small cliffs. When a LiDAR DTM is already available (figure B.2b), the same method is implemented to compute a short-distance elevation mask M_s with a 1 m DTM extracted at 100 m around L. The final mask M is computed as the elementwise maximum of M_r and M_s. On the example of plot 8 in the Chablais (figure B.2c), the complementarity of the two masks appears. The short distance mask does not account for the mask due to the hillside situated across the valley (East-North-East direction). Long distance mask underestimates the elevation occlusion due to upslope local topography.

With the above mentioned procedure, the estimated quality of GPS acquisition can be taken into account when planning field campaigns, so that the productivity of geolocation is improved.



(a) 50 m DTM (b) 1 m DTM (c) Elevation masks

Figure B.2: DTM and corresponding planimetric elevation masks for plot 8 in Chablais. Blue dotted (resp. solid) line is the elevation mask computed with the 50 m (resp. 1 m) DTM. Black circles correspond to 0, 30, 60 and 90 degrees elevation. Red circle is the 15 degrees elevation, which is a common default setting for the lower elevation limit to take a satellite into account.

Appendix C

Description and detection figures for large plots

C.1 Introduction

This appendix gathers the figures corresponding to large plots described in section 5.3.2. For each plot, three figures are displayed for plot description:

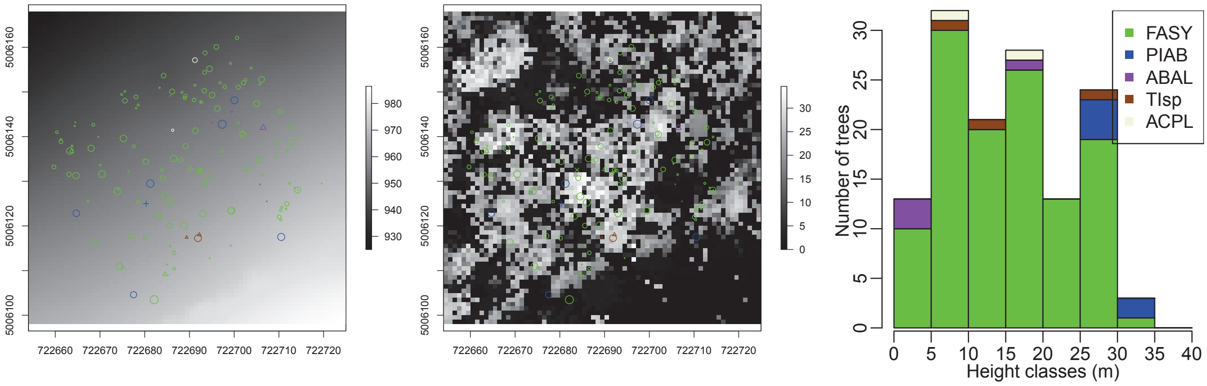
- the digital terrain model (DTM) with tree positions,
- the canopy height model (CHM) with tree positions,
- the histogram of height distribution.

Spatial resolution of the DTM and CHM is one meter. Pixel color refers to the terrain altitude (DTM) or vegetation height (CHM). Scale bar units are in meters. Trees positions are marked as symbols whose size is proportional to tree height. Symbol colors in the images and bar color in the histogram refer to tree species (see table 5.1). Symbol types refer to the appearance of a tree: ○ normal, △ broken stem, + snag with branches, × snag without branches. Only live trees (normal and broken stem categories) are considered in the height distribution.

Two types of figures corresponding to detection results obtained with the best setting determined for each plot are also presented:

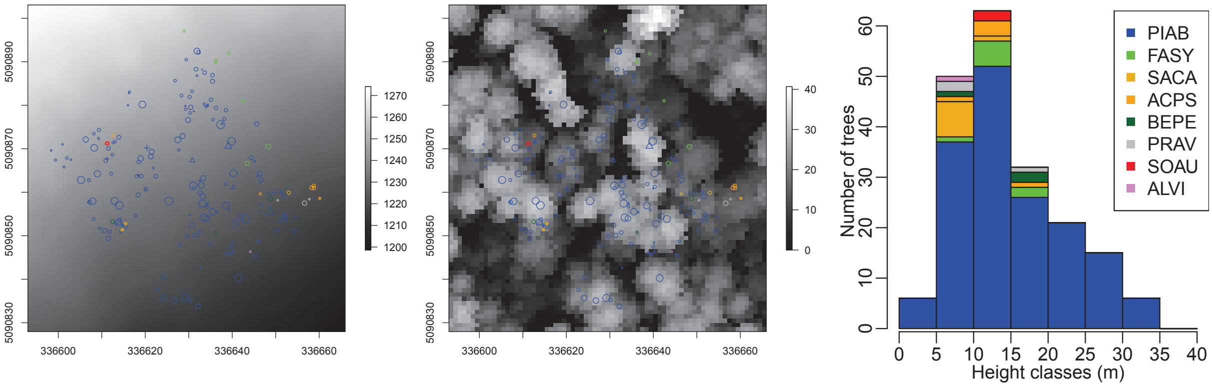
- the height histogram of true positives, false negatives and false positives,
- the scatterplot of field height as a function of LiDAR height, for true positives.

C.2 Description

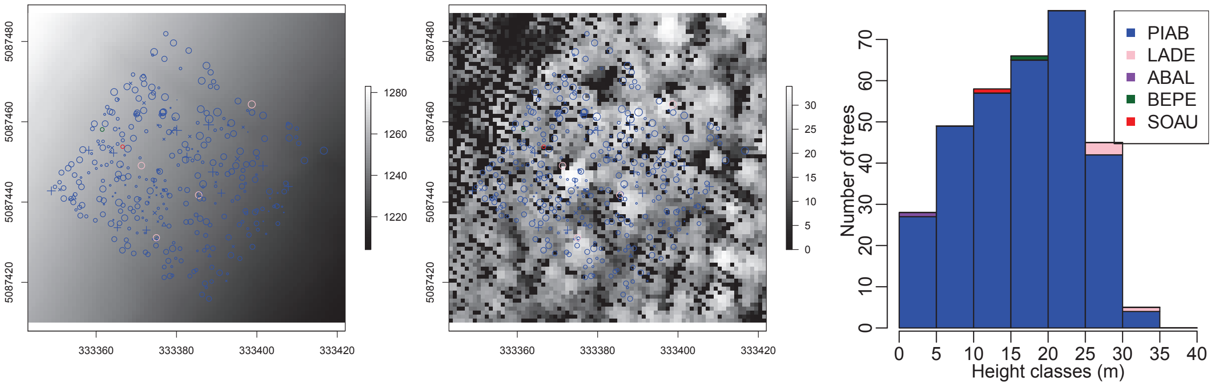


(a) 1 m DTM (b) 1 m CHM (c) Height distribution

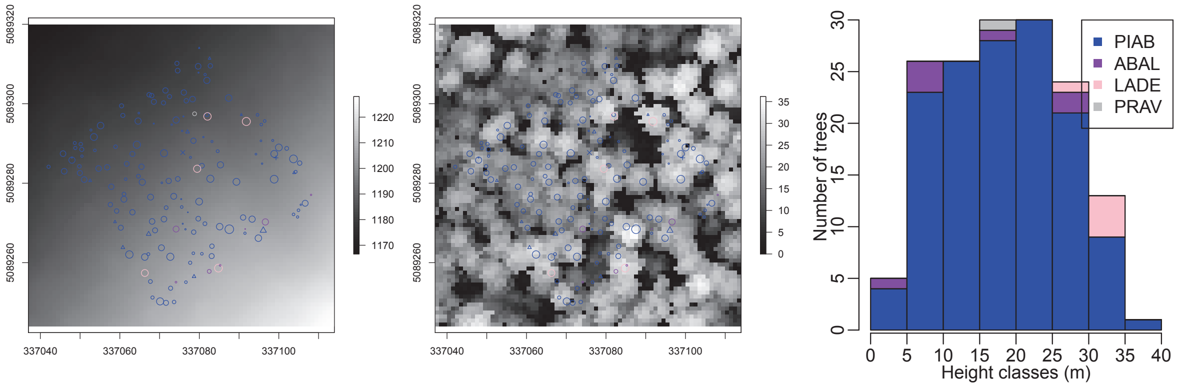
Figure C.1: Murianette plot.



(a) 1 m DTM (b) 1 m CHM (c) Height distribution

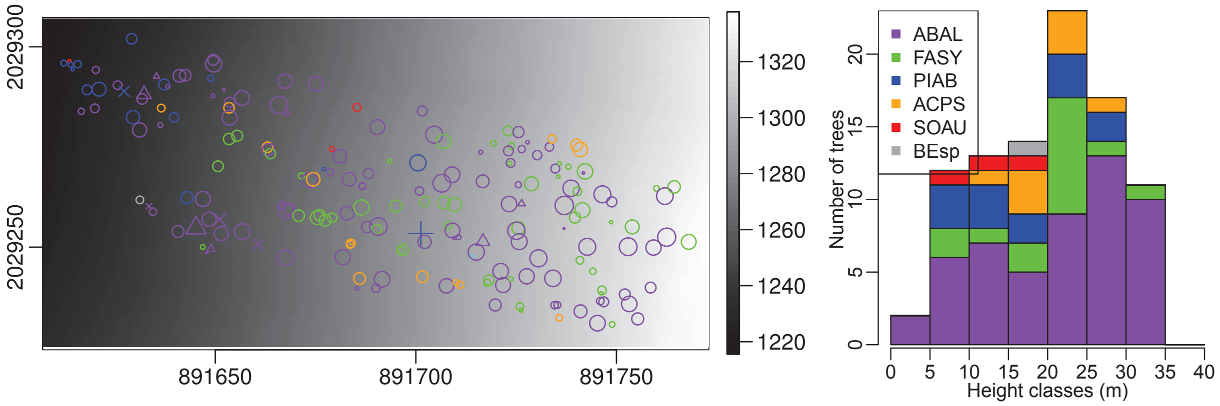


(d) 1 m DTM (e) 1 m CHM (f) Height distribution

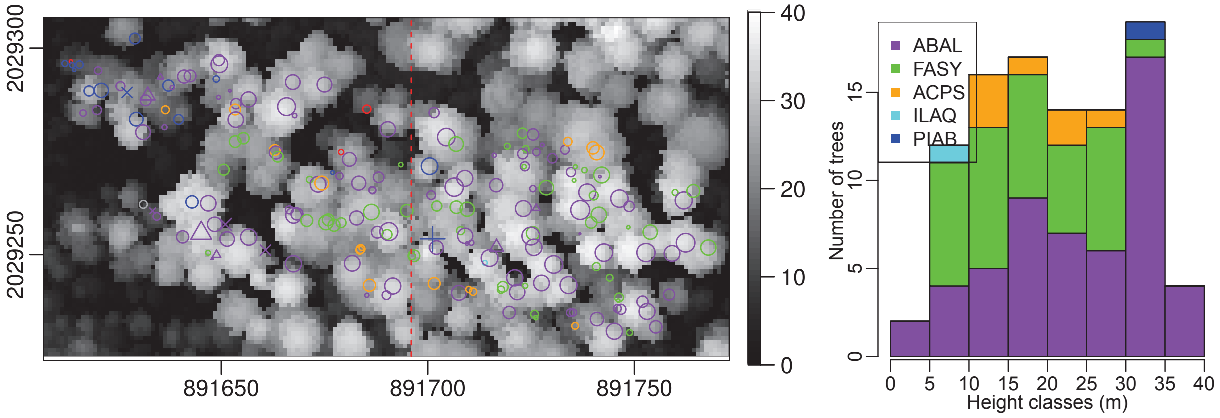


(g) 1 m DTM (h) 1 m CHM (i) Height distribution

Figure C.2: Overview of the Chamonix plots: Corruaz (top row), Moussoux (middle row) and Orthaz3 (bottom row).

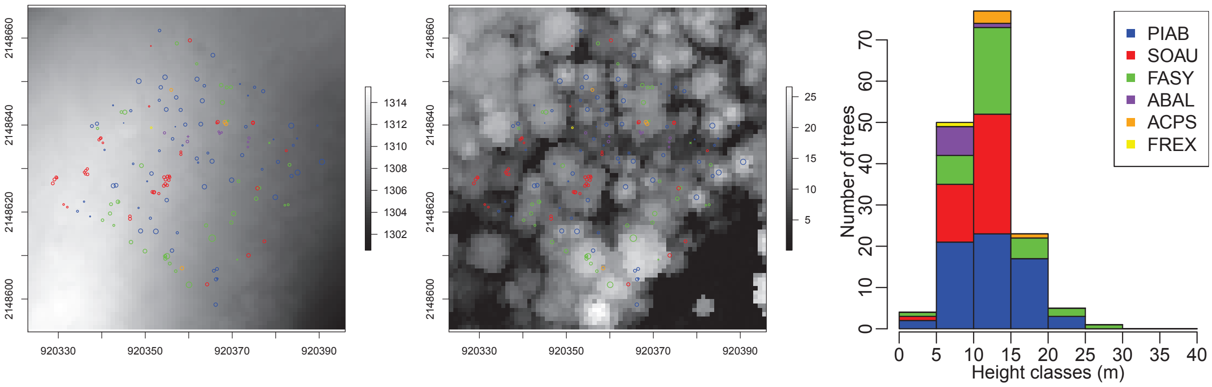


(a) 1 m DTM (b) Height distribution

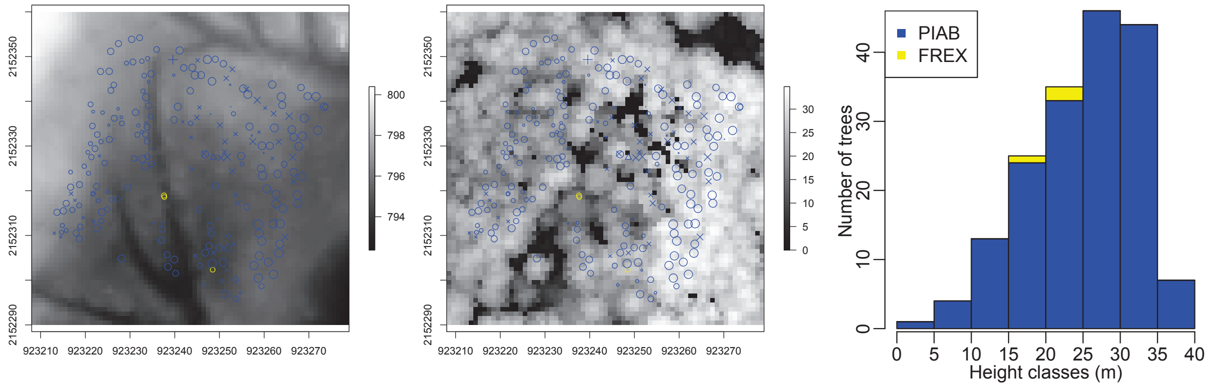


(c) 1 m CHM (d) Height distribution

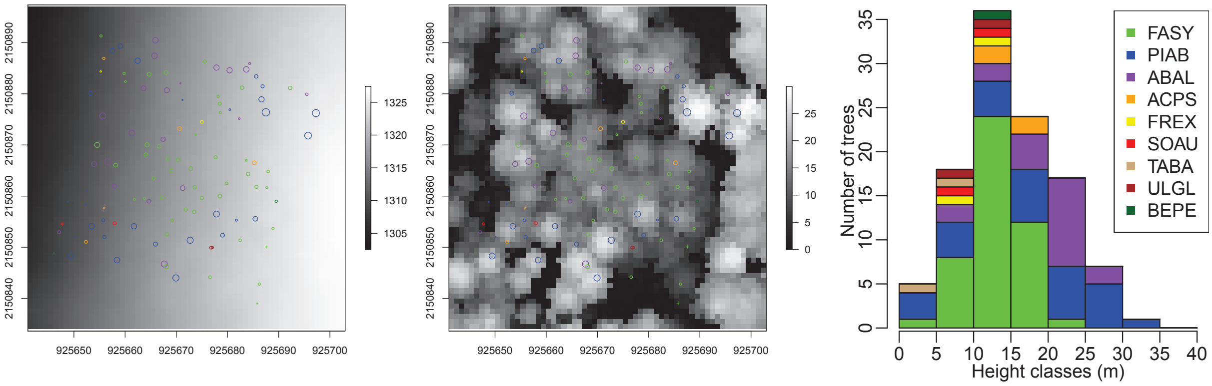
Figure C.3: Overview of the Vaujany study area. Height distribution in (b) refers to the lower part of the plot (vaujanyba), which correspond to trees at the left of the red dotted line in (c). Height distribution in (d) refers to the upper part of the plot (vaujanyha).



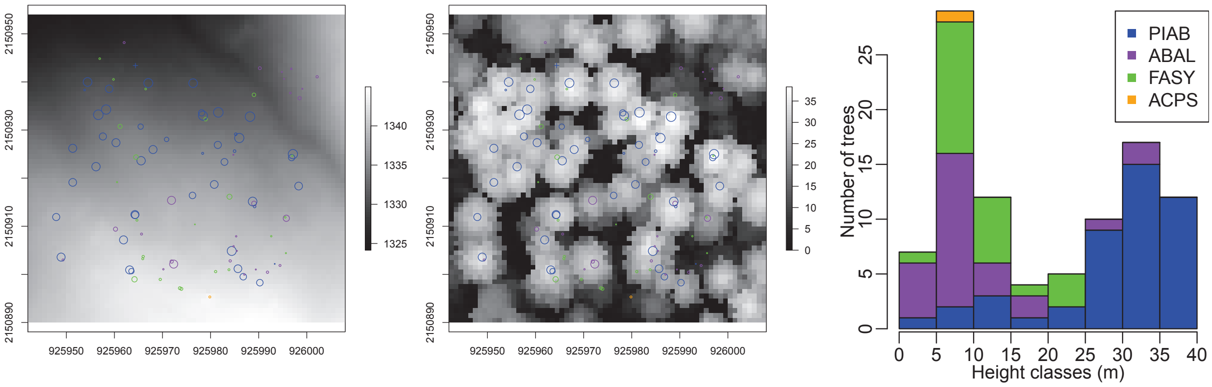
(a) 1 m DTM (b) 1 m CHM (c) Height distribution



(d) 1 m DTM (e) 1 m CHM (f) Height distribution

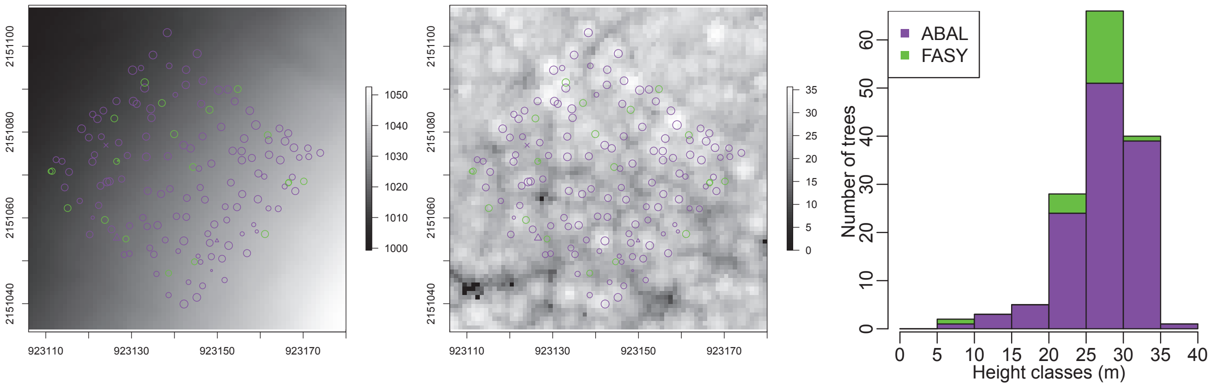


(g) 1 m DTM (h) 1 m CHM (i) Height distribution

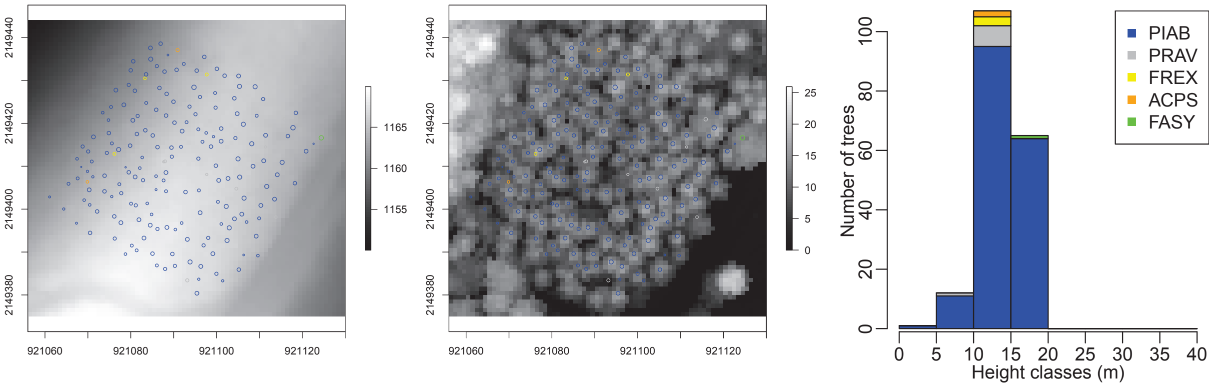


(j) 1 m DTM (k) 1 m CHM (l) Height distribution

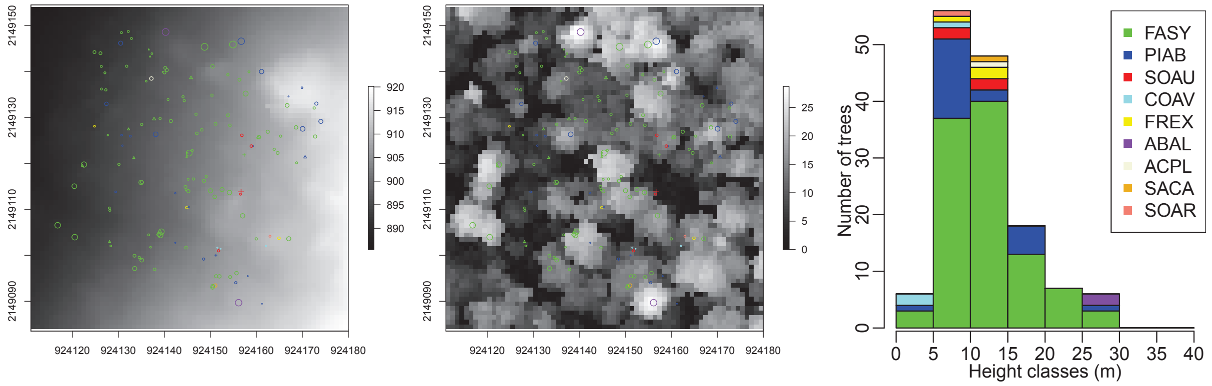
Figure C.4: Overview of the Chablais plots: chablais1 (first row), chablais2 (second row), chablais3 (third row) and chablais4 (fourth row).



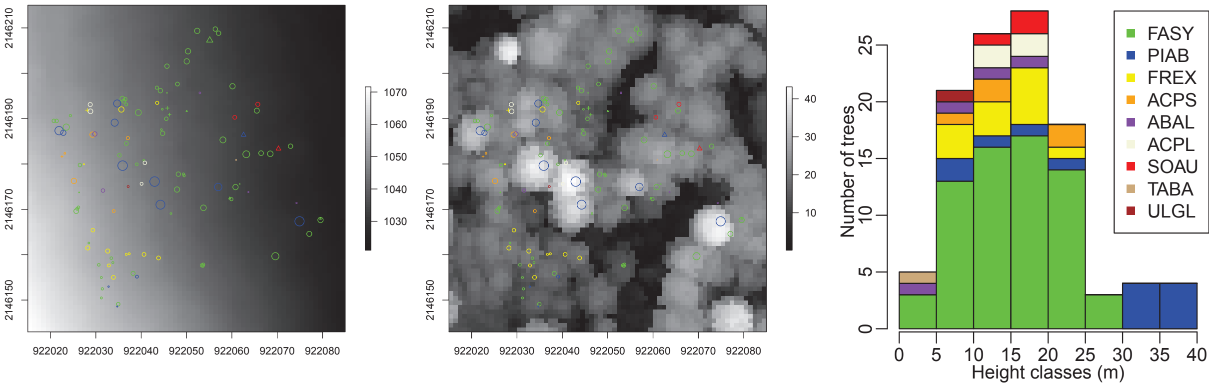
(a) 1 m DTM (b) 1 m CHM (c) Height distribution



(d) 1 m DTM (e) 1 m CHM (f) Height distribution

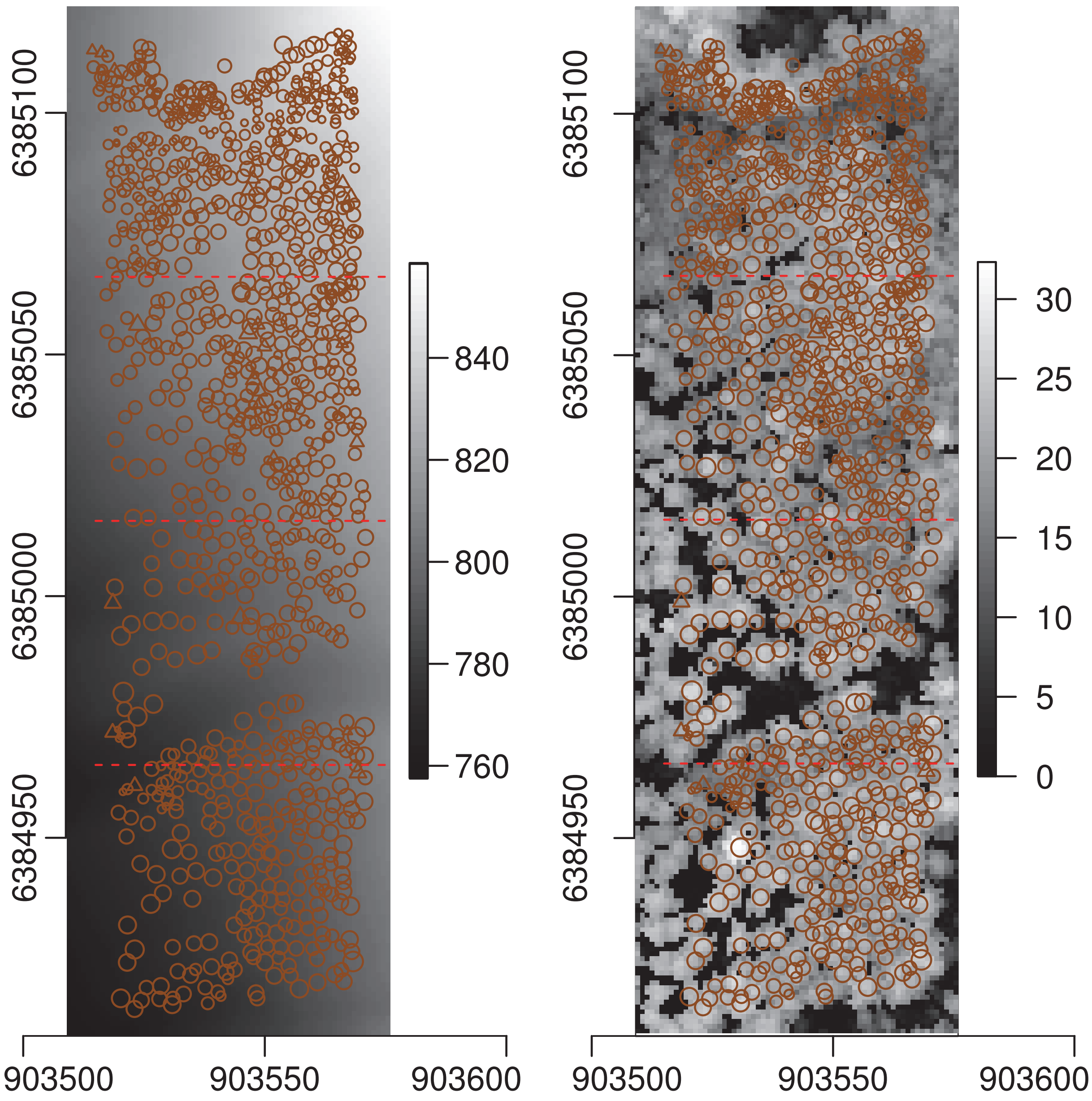


(g) 1 m DTM (h) 1 m CHM (i) Height distribution



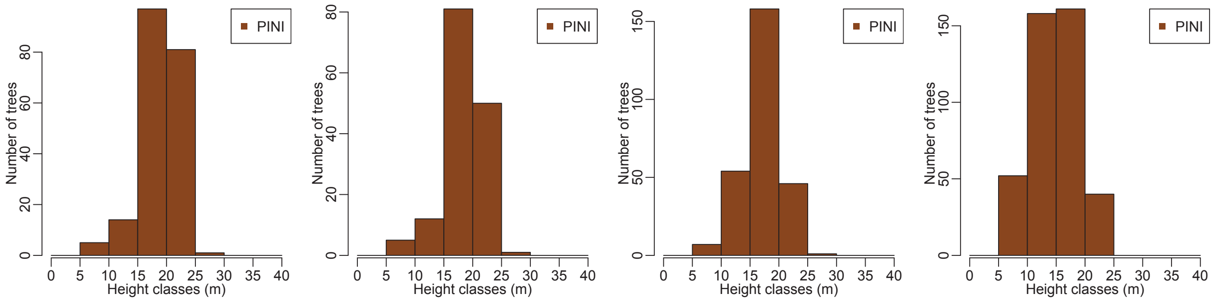
(j) 1 m DTM (k) 1 m CHM (l) Height distribution

Figure C.5: Overview of the Chablais plots (continued): chablais5 (first row), chablais6 (second row), chablais7 (third row) and chablais8 (fourth row).



(a) 1 m DTM

(b) 1 m CHM



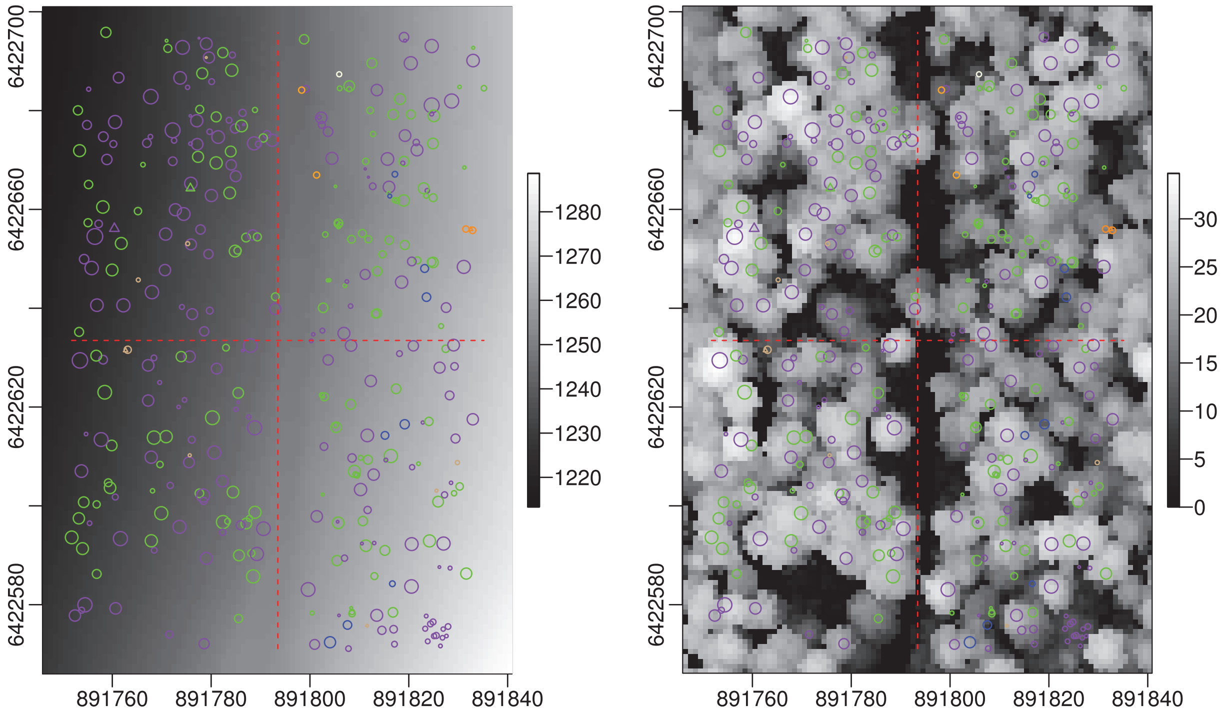
(c) Valdrome1

(d) Valdrome2

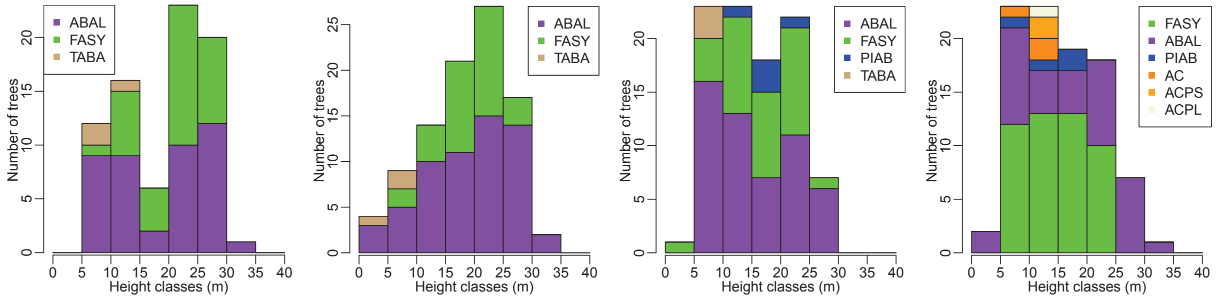
(e) Valdrome3

(f) Valdrome4

Figure C.6: Overview of the Valdrôme marteloscope. Height distribution in (c, d, e, and f) refers to subplots delimited in (b) by the red dotted lines. Subplots are indiced from 1 to 4 from south to north (bottom to top on the figure).



(a) 1 m DTM (b) 1 m CHM

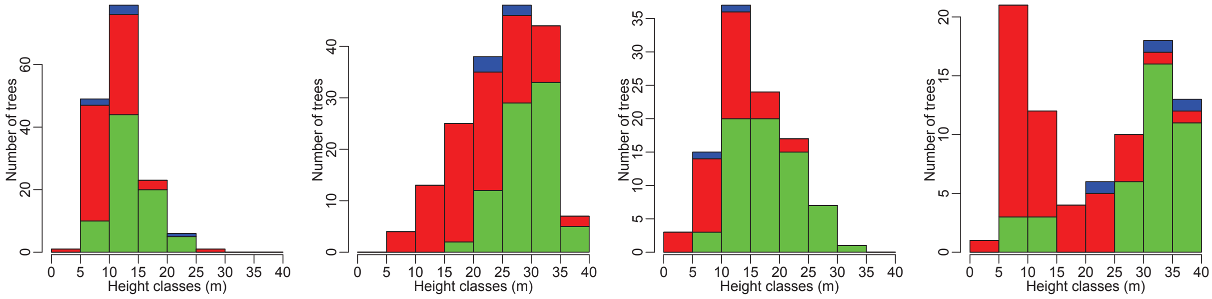


(c) stagnan1 (d) stagnan2 (e) stagnan3 (f) stagnan4

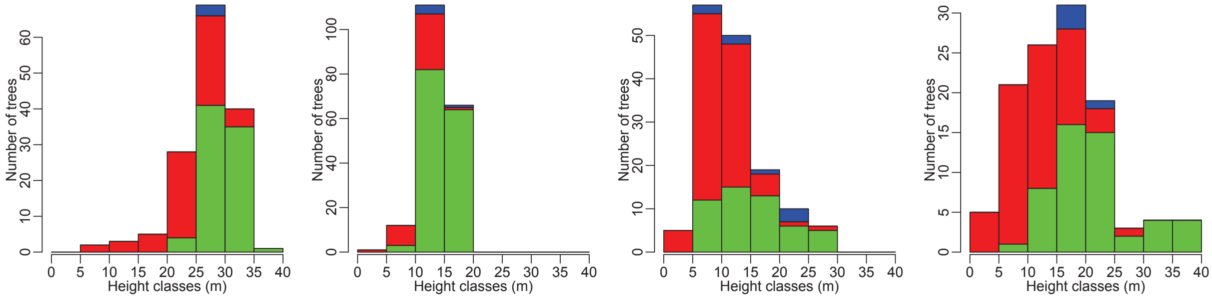
Figure C.7: Overview of the Saint-Agnan marteloscope. Height distribution in (c, d, e, and f) refers to subplots delimited in (b) by the red dotted lines. Subplots are numbered from 1 to 4 with stagnan1 at the bottom left and the following ones numbered in clockwise order.

C.3 Detection results

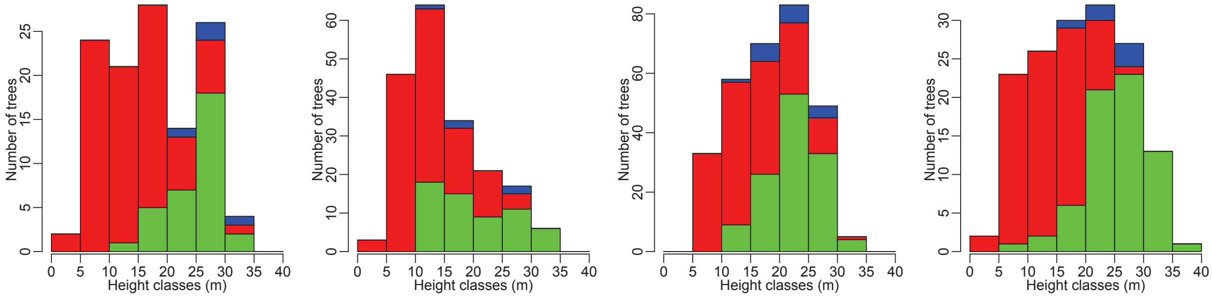
C.3.1 Height histogram



(a) chablais1 (b) chablais2 (c) chablais3 (d) chablais4

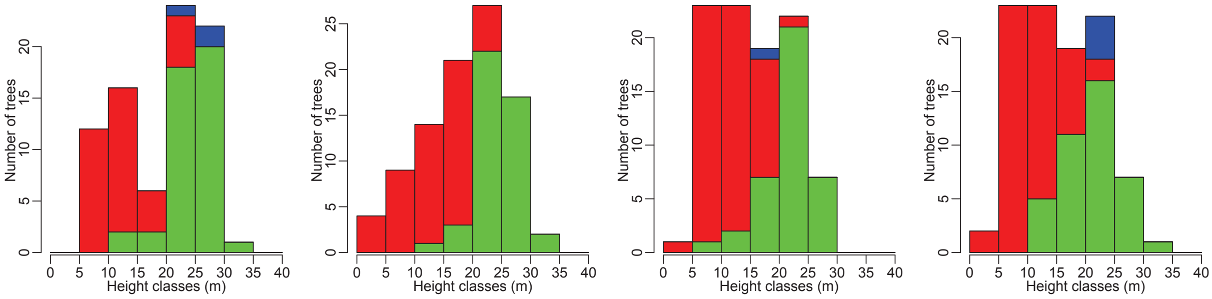


(e) chablais5 (f) chablais6 (g) chablais7 (h) chablais8

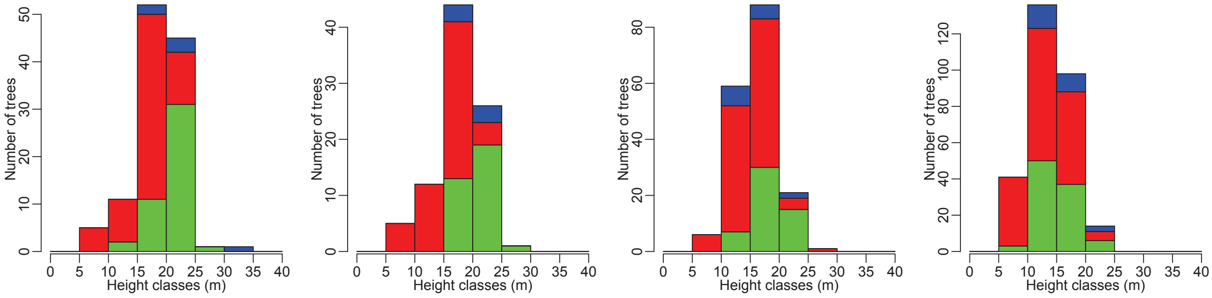


(i) combeloup (j) corruaz (k) moussoux (l) orthaz3

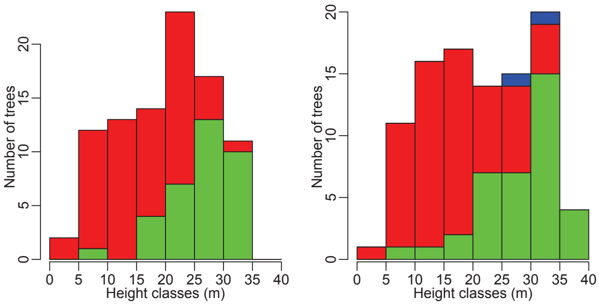
Figure C.8: Field height histogram for detection results: true positives (🟩 green), false negatives (🟥 red) and false positives (🟦 blue). For false positives, LiDAR height is used.



(a) stagnan1 (b) stagnan2 (c) stagnan3 (d) stagnan4



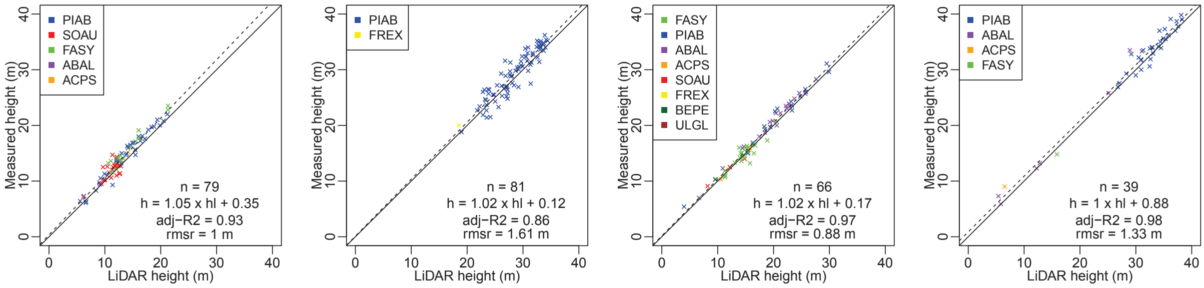
(e) valdrome1 (f) valdrome2 (g) valdrome3 (h) valdrome4



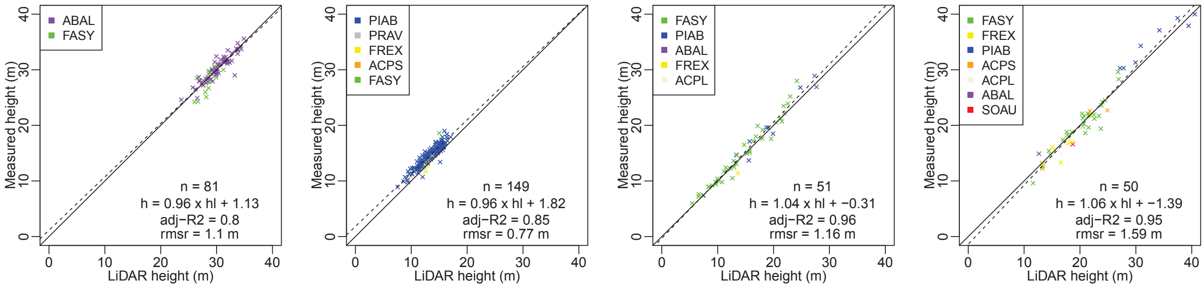
(i) vaujanyba (j) vaujanyha

Figure C.9: Field height histogram for detection results (continued from figure C.8).

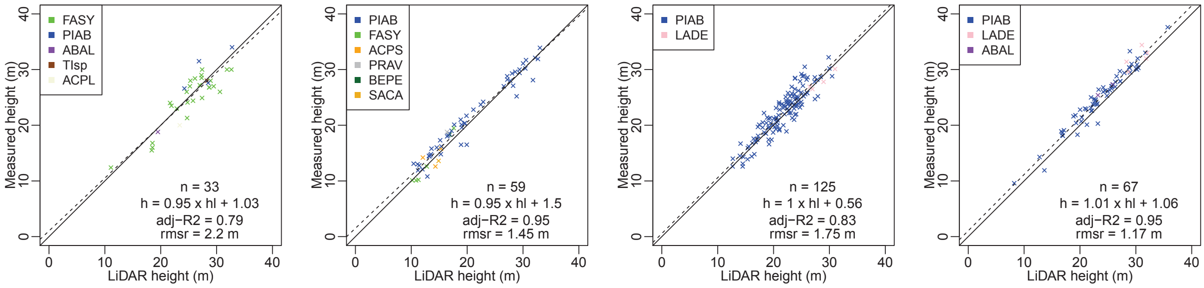
C.3.2 Field-LiDAR height scatterplots



(a) chablais1 (b) chablais2 (c) chablais3 (d) chablais4

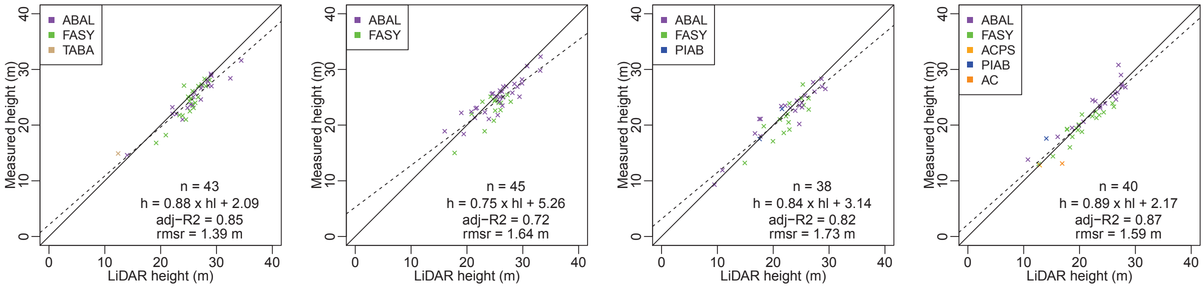


(e) chablais5 (f) chablais6 (g) chablais7 (h) chablais8

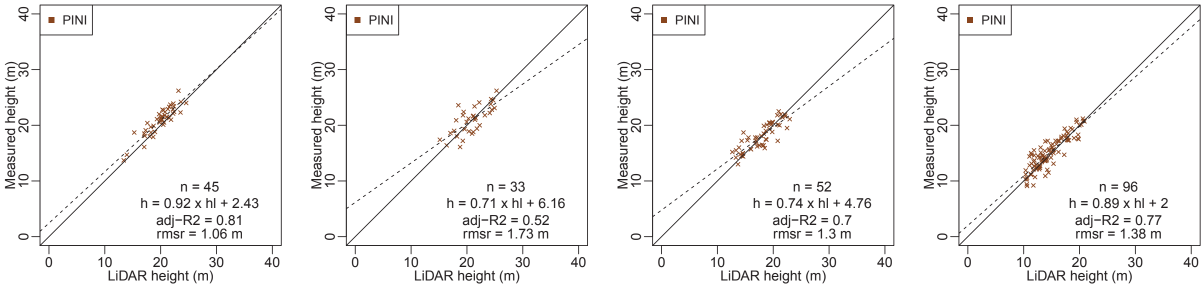


(i) combeloup (j) corruaz (k) moussoux (l) orthaz3

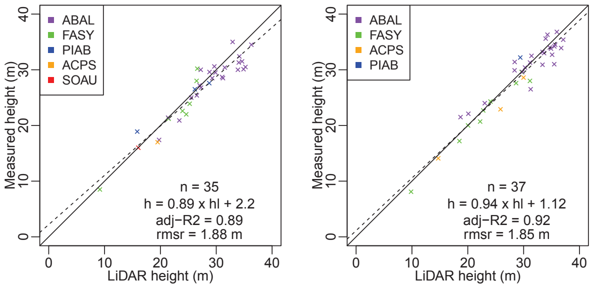
Figure C.10: Field measured height of correctly detected trees plotted against LiDAR height. Regression model is indicated on the figures with the corresponding adj-R^2, root mean square residuals (rmsr) and the number of detected trees. For Valdrôme plots, only detected trees with field measures are considered. Symbol colors refer to tree species (see table 5.1).



(a) stagnan1 (b) stagnan2 (c) stagnan3 (d) stagnan4



(e) valdrome1 (f) valdrome2 (g) valdrome3 (h) valdrome4



(i) vaujanyba (j) vaujanyha

Figure C.11: Field measured height of correctly detected trees plotted against LiDAR height (continued from figure C.10).

Appendix D

Detection improvement by ellipse modelling of tree crowns

D.1 Introduction

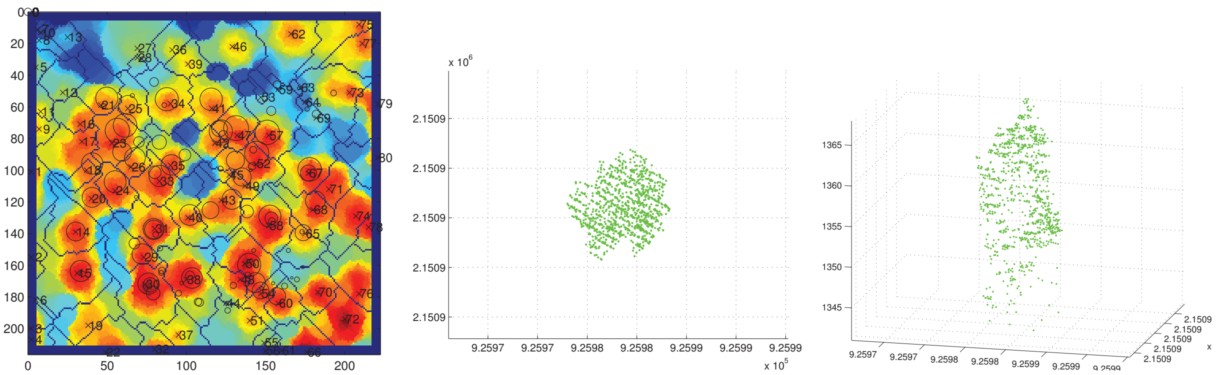
Motivations Single tree detection methods based on image processing techniques do not take advantage of the ability of LiDAR remote sensing to retrieve information from the inside of the canopy cover. Indeed, they are generally applied to the crown height model (CHM), which is the upper envelope of the point cloud. However, in complex or multilayered stands, some trees have mixed crowns or are completely over-topped by taller trees. In the CHM, such trees are merged with other trees, or do not appear at all. Besides, the shape of the visible trees is altered by those suppressed trees. In this appendix, a procedure for improving the delineation of trees obtained by the image-based algorithm detailed in chapter 6 is presented. It relies on the the construction of ellipses on the point cloud within previous segments. This work was done in collaboration with Paula Mursa from the Military Technical Academy of Bucharest.

Material The method is tested on plot chablais4 (field plot description in section 5.3.2). In this square plot of 0.25 ha, positions and heights of trees with diameter at breast height larger than 7.5 cm are known. Single tree detection is performed with the algorithm presented in chapter 6. The adopted setting leads to the correct detection of 35 trees out of 82, with 2 false positives. A marker-constrained watershed segmentation is then performed to delineate tree crowns associated to each detected treetop (section 6.2.7). The input data for the algorithm are the segments obtained by the watershed algorithm, and the LiDAR point cloud with point coordinates and classification (ground, non ground). LiDAR points with relative height lower than two meters are removed. Remaining points are affected to each corresponding segment (figure D.1).

D.2 Methods and results

Assumptions The proposed algorithm is designed to improve tree detection of previous local maxima filtering and watershed segmentation, based on the following statements:

- local maxima filtering only takes the distance to the nearest pixel with higher value into account, which is in fact the distance to the nearest higher crown, and is not a good indicator of the distance to the nearest higher tree,
- watershed segmentation only relies on the fact that the tree crown is monotonically decreasing from the treetop to ground, without any other constraint on shape.



(a) Segmented trees (b) XY view of segment 43 (c) Perspective view of 43

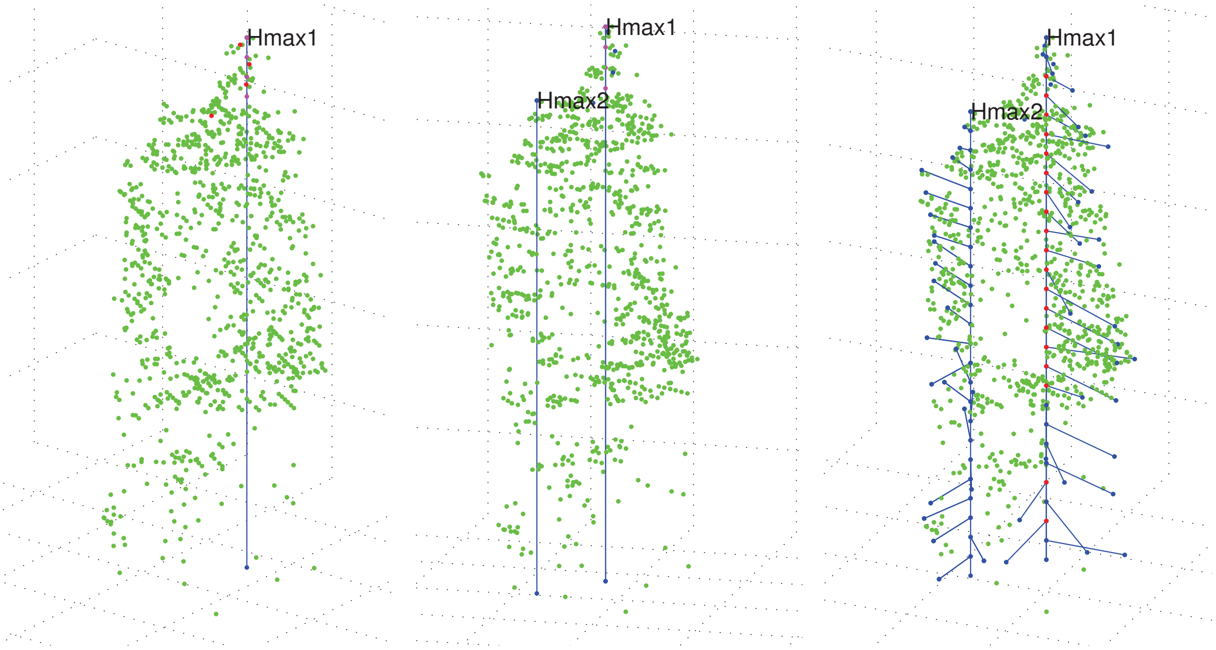
Figure D.1: (a) Result of detection and segmentation of trees on plot chablais4. Numbers correspond to segments id, black circles are the positions of field trees. Symbol size is proportional to tree height. background image is the canopy height model of 0.5 m resolution, axes are labelled in pixels. (b) and (c) view of the point cloud located inside segment 43. Only points higher than 2 m above ground are considered.

The proposed algorithm is based on the assumption that tree crown slices at various heights are concentric, successive ellipses.

Processing of main tree The highest point of the point cloud is identified. From now on, its coordinates are considered as the fixed center $H_{max1}(x_1, y_1)$ of the tree (figure D.2a). The point cloud $(x_i, y_i, z_i)_{i\in\{1,\cdots,N\}}$ is split into n vertical bins (crown slices) of one meter width.
For each level l, starting from the top, an ellipse whose axes are aligned on the grid axes is roughly fitted to the point cloud. The center of the crown slice l is considered to be located at $(x_l, y_l) = \frac{1}{2}(x_{min,l} + x_{max,l}, y_{min,l} + y_{max,l})$, with $x_{min} = min(x_i, i \in l)$. Corresponding ellipse axes lengths are $\sqrt{x_{min,l} + x_{max,l}}$ an $\sqrt{y_{min,l} + y_{max,l}}$.
After the position of the slice center is computed, it is compared to the position of the original center H_{max1}. If they are separated by a distance larger than a threshold value d, the center shift indicates that it is likely that another tree is present in this crown slice. In this case, the iterations are interrupted and the algorithm follows a second procedure.

Processing in case of second tree The top of the second tree is considered to be the point in l with the largest distance to the fixed center H_{max1} of the main tree, and is now recorded as H_{max2} (figure D.2b).
On each level, centers of the ellipses for both trees will now be considered to be located at H_{max1} and H_{max2}. We still consider ellipses that are aligned on the coordinates axes. The preceding procedure for ellipse axes estimation is used except that for the second tree, only points of the half plane where it is located are considered. For the main tree, the points of the other half plane are used (figure D.2c).

Results The algorithm allows the detection of five additional trees (table D.1) on the plot, with two additional false positives. One additional tree is detected in the upper layer, and four in the lower layer. These results are slightly less accurate than those obtained in a study that investigated the improvement of watershed delineated segments by normalized cut segmentation of the point cloud [262]. With a true 3D approach based on voxels, the results from watershed segmentation (48% of true positives with 4% false positives) were improved to 60% true positives with 9% false positives.



(a) Positions of mobile center (b) Position of second tree (c) Maximum crown extension of trees

Figure D.2: Steps of crown splitting. (a) Ellipse center (red dot) on the fourth level is far apart from the original fixed center (blue line). (b) Determined second treetop. (c) Estimation of the maximum extension of trees in their respective half-planes, for each level.

Table D.1: Detected trees by height class for the local maxima algorithm and the additional ellipses construction.

Height class (m)	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	Total
Detected (local maxima)	0	0	2	0	0	5	16	12	35
Detected (point cloud)	0	2	4	0	0	6	16	12	40
Inventoried	0	22	12	3	6	9	17	13	82

The crown of the main tree is also better modelled by the ellipses than with the sole watershed segmentation, particularly when a second tree is detected (figure D.3).

D.3 Discussion and perspectives

Limits of the approach Figure D.4 presents typical cases where the algorithm succeeds or fails in correctly detecting additional trees within watershed segments. Figure D.4a presents two cases where the second tree is detected. The crown of the second tree is sufficiently separated from the crown of the first one to induce a shift of the center of the crown slice at some given level.

When the arrangement of tree crowns results in an almost symmetric segment crown, the presence of a second tree remains unnoticed (figure D.4b). On the right side, the crown of the second tree is completely fused in the crown of the other. On the left side, the small tree is well separated from the dominant one, but the segment crown is globally symmetric due to the presence of low branches on the other side of the main tree.

The last figure (D.4c) exemplifies false detection cases. On the left side, a false second tree will

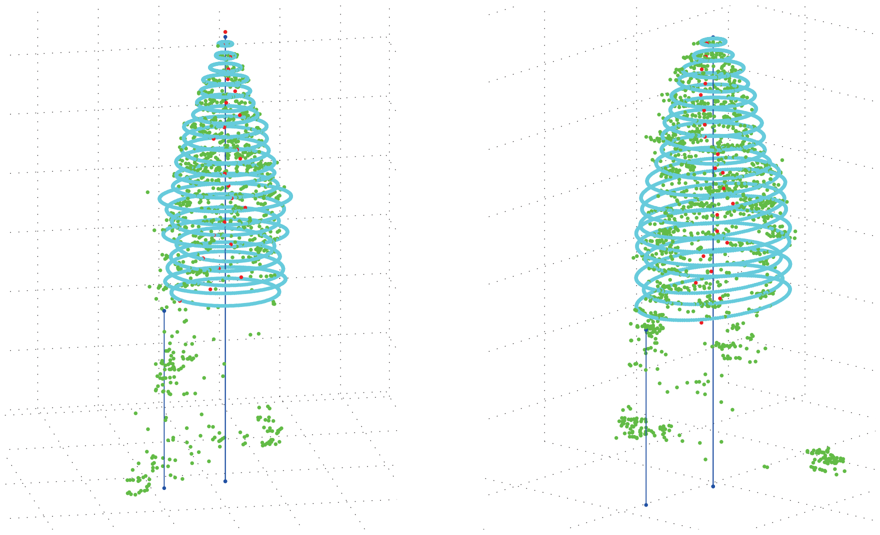

(a) (b)

Figure D.3: Examples of modelled crowns of main trees.

be detected where the crown begins to be asymmetric. Trees will indeed develop a larger crown on the side where they have no close neighbor [239, p. 71]. On the right side, point extraction after the watershed segmentation results in an erroneous affectation of part of the crown of the round tree to the lower part of the large tree. The algorithm will detect a second tree in this segment.

Point cloud based approaches applied on single segments obtained from local maxima filtering are indeed penalized by the fact that adjacent trees are not considered, whereas the shape of a tree is greatly influenced by its neighbors. Due to the complexity of heterogeneous forests, even with a more refined approach, the correct detection of additional trees requires a trade-off with false detections.

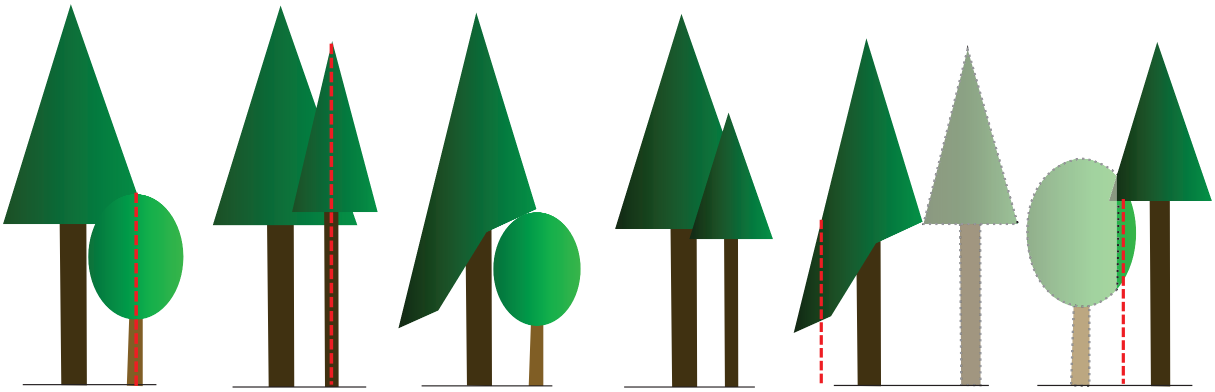

(a) True positives (b) False negatives (c) False positives

Figure D.4: Typical cases encountered by the algorithm.

Advantages The main output of this approach might finally be the modelling of the tree crown of the dominant tree. Extracted parameters such as crown volume, length and area will be useful to improve the models based on LiDAR height for tree diameter and volume estimation. Crown base height can be estimated by analysing the vertical point distribution, but the presence of additional trees in the segment increases the error. Combining height distribution and crown symmetry as proposed with this algorithm are likely to improve the modelling of tree crowns in complex forests.

Algorithm improvement The tested algorithm showed that this approach is successful in detecting additional trees and modelling the shape of dominant trees. However, it can still be improved on several points.

The ellipse construction is based on points range along two axes, hence it is not rotation invariant. With trees that are generally isotropic, a more robust method should be implemented, e.g. by fitting ellipses with a least-squares algorithm on points located on the convex hull of each crown slice. Slice width should also be automatically adapted in order to ensure that enough points are included for computation.

The criterion for the presence of the second tree is only based on the distance between the coordinates of the treetop and the center of the ellipse at each level. The distances to the centers of the ellipses at the previous levels should also be considered.

The algorithm should also be tested on more field data, in order to test its performance on various forest structures. Special attention should also be paid to the influence of LiDAR point density on detection results.

Single tree detection in complex forests Detection algorithm can be improved by making assumptions about forest features. However, additional criteria will allow the detection of more trees, but also raise numerous particular cases that the algorithm will not handle correctly. In complex forests, the variability of tree shapes and arrangements makes many simplifying hypotheses about forest structures hazardous. Improving detection performance requires a comprehensive modelling of the interactions between laser pulses and the vegetation cover, which implies taking both the real shape of tree crowns and the effects of scanning pattern into account. An alternative is to automatically adapt algorithms to forest structures, e.g. with the training approach presented in chapter 8. This last approach is likely to be more robust but less powerful regarding the attained detection rate.

Glossary

ALS Airborne Laser Scanning. 9, 12, 15, 26, 73, 80, 104, 121, 122

CHM Canopy (or Crown) Height Model: image of the vegetation height. 11, 30, 46, 47, 53, 58, 77, 80, 83, 129, 141

CV_{RMSE} Coefficient of Variation of the Root Mean Square Error. 95

DBH Diameter at Breast Height: tree diameter measured at 1.3 m height. 25, 28–30, 33, 34, 67, 97, 99

DSM Digital Surface Model: image of the Earth surface altitude (including vegetation and buidings). 44–48, 52, 58, 68, 72, 73, 76, 77

DTM Digital Terrain Model: image of the terrain altitude. 11, 20, 26, 30, 34, 45–48, 52, 58, 77, 127–129

GLAS/ICESat Geoscience Laser Altimeter System, carried on the Ice, Cloud and land Elevation Satellite. 10

GPS Global Positioning System. 9, 10, 19, 20, 22, 23, 26, 27, 30, 33–35, 53, 54, 57, 69, 90, 105, 106, 127, 128

ICA Independent Component Analysis. 94, 107, 110–112, 114, 115

IMU Inertial Measurement Unit. 9, 10, 19, 22

IQR Inter Quartile Range. 99

LiDAR Light Detection And Ranging. 9–15, 19–21, 23, 25, 26, 28, 34, 36, 41–45, 49, 50, 53, 57, 58, 60, 61, 67–71, 73, 75, 76, 79, 83, 84, 89–91, 97, 105, 107, 115, 128, 129, 141, 144, 145

LOO Leave-One-Out (cross-validation). 95, 101, 112, 114, 115

LVIS Laser Vegetation Imaging Sensor. 10–12, 14

NASA National Aeronautics and Space Administration. 11

***ols*-MR** Ordinary Least Squares Multiple Regression. 91, 93–95, 107, 110–115

PCA Principal Component Analysis. 94, 107, 110, 111, 114, 115

PDOP Positional Dilution Of Precision. 26, 127

$RMSE$ Root Mean Square Error. 95, 99, 101, 102, 105, 111, 112, 114, 115

SE Structuring Element. 48

SLICER Scanning Lidar Imager of Canopies by Echo Recovery. 10, 11

SVR Support Vector Regression. 90, 91, 93–95, 111, 114–116, 122

Bibliography

[1] "Project Interreg Alpine Space MANFRED," http://www.manfredproject.eu/.

[2] "Project Interreg Forêts de Protection (IFP)," http://www.interreg-forets-protection.eu/.

[3] "Mountain areas in Europe: Analysis of mountain areas in EU member states, acceding and other European countries. Final report. European Commission contract no. 2002.CE.16.0.AT.136," Nordregio, Tech. Rep., 2004.

[4] "Projet Interreg III A ALCOTRA no 165. PROVIALP, protection de la viabilité alpine," Cemagref and Arpa, Tech. Rep., 2006.

[5] "CLC2006 technical guidelines," European Environment Agency, Tech. Rep., 2007.

[6] N. A. Akel, S. Filin, and Y. Doytsher, "Orthogonal polynomials supported by region growing segmentation for the extraction of terrain from lidar data," *Photogrammetric Engineering and Remote Sensing*, vol. 73, pp. 1253–1266, 2007.

[7] C. Alexander, "Delineating tree crowns from airborne laser scanning point cloud data using delaunay triangulation," *International Journal of Remote Sensing*, vol. 30, no. 14, pp. 3843–3848, 2009.

[8] C. Alexander, K. Tansey, J. Kaduk, D. Holland, and N. J. Tate, "Backscatter coefficient as an attribute for the classification of full-waveform airborne laser scanning data in urban areas," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 5, pp. 423–432, Sep. 2010.

[9] H. Algan, "Tarifs unifiés," *Revue des Eaux et Forêts*, pp. 555–562, 1901.

[10] P. Ancelin, C. Barthelon, F. Berger, M. Cardew, C. Chauvin, B. Courbaud, L. Descroix, L. Dorren, J. Fay, P. Gaudry, X. Gauquelin, J.-R. J., Genin, D. Joud, P. Loho, É. Mermin, F. Plancheron, A. Prochasson, F. Rey, D. Rubeaud, and L. Wlérick, *Guide des sylvicultures de montagne: Alpes du Nord françaises.* Cemagref-ONF, 2006.

[11] H. E. Andersen, "Using airborne light detection and ranging (lidar) to characterize forest stand condition on the Kenai peninsula of Alaska," *Western Journal Of Applied Forestry*, vol. 24, no. 2, pp. 95–102, Apr. 2009.

[12] H. E. Andersen, R. J. McGaughey, and S. E. Reutebuch, "Estimating forest canopy fuel parameters using lidar data," *Remote Sensing of Environment*, vol. 94, no. 4, pp. 441–449, 2005.

[13] H. E. Andersen, S. E. Reutebuch, and R. J. McGaughey, "A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods," *Canadian Journal of Remote Sensing*, vol. 32, no. 5, pp. 355–366, 2006.

[14] H.-E. Andersen, S. E. Reutebuch, and G. F. Schreuder, "Bayesian object recognition for the analysis of complex forest scenes in airborne laser scanner data," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIV, Part 3A. ISPRS Commission III, Symposium 2002 September 9 - 13, 2002, Graz, Austria*, 2002, p. 7.

[15] A. S. Antonarakis, K. S. Richards, and J. Brasington, "Object-based land cover classification using airborne lidar," *Remote Sensing of Environment*, vol. 112, no. 6, pp. 2988–2998, 2008.

[16] P. Axelsson, "DEM generation from laser scanner data using adaptive TIN models," in *XIXth ISPRS Congress, IAPRS*, vol. XXXIII, 2000, pp. 110–117.

[17] E. P. Baltsavias, "Airborne laser scanning: basic relations and formulas," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 54, no. 2-3, pp. 199–214, Jul. 1999.

[18] ——, "Airborne laser scanning: existing systems and firms and other resources," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 54, no. 2-3, pp. 164–198, Jul. 1999.

[19] Y. F. Bao, C. X. Cao, C. Y. Chang, X. W. Li, E. X. Chen, and Z. Y. Li, "Segmentation to the point clouds of lidar data based on change of kurtosis," *International Symposium on Photoelectronic Detection and Imaging 2007: Image Processing, Sep. 09-12, 2007 Beijing, China*, vol. 6623, p. N6231, 2008.

[20] A. Barbati, G. Chirici, P. Corona, A. Montaghi, and D. Travaglini, "Area-based assessment of forest standing volume by field measurements and airborne laser scanner data," *International Journal of Remote Sensing*, vol. 30, no. 19, pp. 5177–5194, 2009.

[21] C. Bassler, J. Stadler, J. Muller, B. Forster, A. Gottlein, and R. Brandl, "Lidar as a rapid tool to predict forest habitat types in Natura 2000 networks," *Biodiversity and Conservation*, vol. 20, no. 3, pp. 465–481, Mar. 2011.

[22] K. M. Bergen, S. J. Goetz, R. O. Dubayah, G. M. Henebry, C. T. Hunsaker, M. L. Imhoff, R. F. Nelson, G. G. Parker, and V. C. Radeloff, "Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions," *Journal of Geophysical Research-Biogeosciences*, vol. 114, p. G00E06, Dec. 2009.

[23] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard, "Support vectors machines regression for estimation of Mars surface physical properties," in *Proceedings of the 17th European Symposium on Artificial Neural Networks, Bruges, Belgium*, 2009, pp. 195–200.

[24] G. A. Blackburn, "Remote sensing of forest pigments using airborne imaging spectrometer and lidar imagery," *Remote Sensing of Environment*, vol. 82, no. 2-3, pp. 311–321, 2002.

[25] J. B. Blair, D. L. Rabine, and M. A. Hofton, "The Laser Vegetation Imaging Sensor (LVIS): A medium-altitude, digitization-only, airborne laser altimeter for mapping vegetation and topography," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 54, pp. 115–122, 1999.

[26] J.-G. Boureau, *Manuel d'interprétation des photographies aériennes infrarouges : application aux milieux forestiers et naturels.* Inventaire Forestier National, 2008.

[27] T. Brandtberg, “Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 61, no. 5, pp. 325–340, 2007.

[28] T. Brandtberg, T. A. Warner, R. E. Landenberger, and J. B. McGraw, “Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America,” *Remote Sensing of Environment*, vol. 85, no. 3, pp. 290–303, 2003.

[29] J. Breidenbach, B. Koch, G. Kandler, and A. Kleusberg, “Quantifying the influence of slope, aspect, crown shape and stem density on the estimation of tree height at plot level using lidar and InSAR data,” *International Journal of Remote Sensing*, vol. 29, no. 5, pp. 1511–1536, 2008.

[30] J. Breidenbach, E. Naesset, V. Lien, T. Gobakken, and S. Solberg, “Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data,” *Remote Sensing of Environment*, vol. 114, no. 4, pp. 911–924, Apr. 2010.

[31] J. Breidenbach, A. Nothdurft, and G. Kändler, “Comparison of nearest neighbour approaches for small area estimation of tree species-specific forest inventory attributes in central Europe using airborne laser scanner data,” *European Journal of Forest Research*, vol. 129, no. 5, pp. 833–846, 2010.

[32] R. Brennan and T. L. Webster, “Object-oriented land cover classification of lidar-derived surfaces,” *Canadian Journal of Remote Sensing*, vol. 32, no. 2, pp. 162–172, 2006.

[33] F. Bretar, A. Chauve, J. S. Bailly, C. Mallet, and A. Jacome, “Terrain surfaces and 3-D landcover classification from small footprint full-waveform lidar data: application to badlands,” *Hydrology And Earth System Sciences*, vol. 13, no. 8, pp. 1531–1544, 2009.

[34] F. Bretar and N. Chehata, “Terrain modeling from lidar range data in natural landscapes: A predictive and bayesian framework,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1568–1578, Mar. 2010.

[35] C. Briese, B. Höfle, H. Lehner, W. Wagner, M. Pfennigbauer, and A. Ullrich, “Calibration of full-waveform airborne laser scanning data for object classification - art. no. 69500h,” *Laser Radar Technology and Applications XIII*, vol. 6950, 2008.

[36] C. Briese, N. Pfeifer, and P. Dorninger, “Applications of the robust interpolation for DTM determination,” in *International Archives of Photogrammetry and Remote Sensing*, ser. Part 3A, vol. 34, 2002, p. 55–61.

[37] G. Camps-Valls, J. Munoz-Mari, L. Gomez-Chova, K. Richter, and J. Calpe-Maravilla, “Biophysical parameter estimation with a semisupervised support vector machine,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 6, no. 2, pp. 248–252, Apr. 2009.

[38] R. M. N. Cerrillo, M. S. de la Orden, J. G. Bonilla, A. Garcia-Ferrer, R. H. Clemente, and S. Lanjeri, “Lidar-based estimation of leaf area index on holm oak [quercus ilex l. subsp ballota (desf.) samp.] trees,” *Forest Systems*, vol. 19, no. 1, pp. 61–69, Apr. 2010.

[39] L. D. Chang, K. C. Slatton, and C. Krekelera, “Bare-earth extraction from airborne lidar data based on segmentation modeling and iterative surface corrections,” *Journal of Applied Remote Sensing*, vol. 4, Aug. 2010.

[40] L. Chasmer, C. Hopkinson, and P. Treitz, "Investigating laser pulse penetration through a conifer canopy by integrating airborne and terrestrial lidar," *Canadian Journal of Remote Sensing*, vol. 32, no. 2, pp. 116–125, 2006.

[41] A. Chauve, C. Vega, S. Durrieu, F. Bretar, T. Allouis, M. P. Deseilligny, and W. Puech, "Advanced full-waveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest," *International Journal of Remote Sensing*, vol. 30, no. 19, pp. 5211–5228, 2009.

[42] A. Chauve, C. Mallet, F. Bretar, S. Durrieu, M. Pierrot-Deseilligny, and W. Puech, "Processing full-waveform lidar data: modelling raw signals," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. 36 (Part 3/W52). Espoo, Finland, September 2007*, 2007.

[43] Q. Chen, D. Baldocchi, P. Gong, and M. Kelly, "Isolating individual trees in a savanna woodland using small footprint lidar data," *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 8, pp. 923–932, 2006.

[44] X. X. Chen, L. Vierling, E. Rowell, and T. DeFelice, "Using lidar and effective LAI data to evaluate IKONOS and Landsat 7 ETM+ vegetation cover estimates in a ponderosa pine forest," *Remote Sensing of Environment*, vol. 91, no. 1, pp. 14–26, 2004.

[45] M. L. Clark, D. B. Clark, and D. A. Roberts, "Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape," *Remote Sensing of Environment*, vol. 91, no. 1, pp. 68–89, 2004.

[46] R. Clawges, K. Vierling, L. Vierling, and E. Rowell, "The use of airborne lidar to assess avian species diversity, density, and occurrence in a pine/aspen forest," *Remote Sensing of Environment*, vol. 112, no. 5, pp. 2064–2073, 2008.

[47] N. C. Coops, T. Hilker, M. A. Wulder, B. St-Onge, G. Newnham, A. Siggins, and J. A. Trofymow, "Estimating canopy structure of douglas-fir forest stands from discrete-return lidar," *Trees-Structure and Function*, vol. 21, no. 3, pp. 295–310, 2007.

[48] N. C. Coops, A. Varhola, C. W. Bater, P. Teti, S. Boon, N. Goodwin, and M. Weiler, "Assessing differences in tree and stand structure following beetle infestation using lidar data," *Canadian Journal of Remote Sensing*, vol. 35, no. 6, pp. 497–508, Dec. 2009.

[49] N. C. Coops, M. A. Wulder, D. S. Culvenor, and B. St-Onge, "Comparison of forest attributes extracted from fine spatial resolution multispectral and lidar data," *Canadian Journal of Remote Sensing*, vol. 30, no. 6, pp. 855–866, 2004.

[50] B. Courbaud, F. Goreaud, P. Dreyfus, and F. Bonnet, "Sylviculture et modèle de croissance dépendant des distances : mise en oeuvre du logiciel capsis 3.0 sur des pessières irrégulières de montagne," *Revue Forestière Française*, vol. 5, pp. 425–440, 2000.

[51] D. S. Culvenor, "TIDA: an algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery," *Computers & Geosciences*, vol. 28, no. 1, pp. 33–44, Feb. 2002.

[52] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1416–1427, May 2008.

[53] M. Dalponte, N. C. Coops, L. Bruzzone, and D. Gianelle, "Analysis on the use of multiple returns lidar data for the estimation of tree stems volume," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 4, pp. 310–318, Dec. 2009.

[54] T. J. Dean, Q. V. Cao, S. D. Roberts, and D. L. Evans, "Measuring heights to crown base and crown median with lidar in a mature, even-aged loblolly pine stand," *Forest Ecology and Management*, vol. 257, no. 1, pp. 126–133, 2009.

[55] J. Dhote and E. Deherce, "Hyperbolic model for adjustment of sets of height diameter curves," *Canadian Journal of Forest Research*, vol. 24, no. 9, pp. 1782–1790, Sep. 1994.

[56] R. Dianat and S. Kasaei, "Dimension reduction of optical remote sensing images via minimum change rate deviation method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 198–206, Jan. 2010.

[57] O. Diedershagen, B. Koch, and H. Weinacker, "Automatic segmentation and characterisation of forest stand parameters using airborne lidar data, multispectral and fogis data," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXVI, Part 8/W2. Proceedings of the international Conference. Laser-Scanners for Forest and Landscape Assessment - Instruments, Processing Methods and Applications. Freiburg im Breisgau. Germany.*, 2004.

[58] D. Dionne, J.-G. Boureau, M. Deshayes, D. Gratton, and D. Marceau, "Étude de la résolution spatiale optimale dans un milieu forestier de moyenne montagne," *Bulletin - Société française de photogrammétrie et télédétection*, no. 141, pp. 51–55, 1996.

[59] P. L. Dong, "Characterization of individual tree crowns using three-dimensional shape signatures derived from lidar data," *International Journal of Remote Sensing*, vol. 30, no. 24, pp. 6621–6628, 2009.

[60] W. Dorigo, M. Hollaus, W. Wagner, and K. Schadauer, "An application-oriented automated approach for co-registration of forest inventory and airborne laser scanning data," *International Journal of Remote Sensing*, vol. 31, no. 5, pp. 1133–1153, 2010.

[61] W. Dorigo, M. Hollaus, K. Schadauer, and W. Wagner, "A new automated approach for co-registration of national forest inventory and airborne laser scanning data," in *SilviLaser 2008, Sept. 17-19, 2008 - Edinburgh, UK*, 2008.

[62] L. K. A. Dorren, F. Berger, and U. S. Putters, "Real-size experiments and 3-D simulation of rockfall on forested and non-forested slopes," *Natural Hazards and Earth System Sciences*, vol. 6, no. 1, pp. 145–153, 2006.

[63] J. B. Drake, R. O. Dubayah, D. B. Clark, R. G. Knox, J. B. Blair, M. A. Hofton, R. L. Chazdon, J. F. Weishampel, and S. D. Prince, "Estimation of tropical forest structural characteristics using large-footprint lidar," *Remote Sensing of Environment*, vol. 79, no. 2-3, pp. 305–319, 2002.

[64] J. B. Drake, R. O. Dubayah, R. G. Knox, D. B. Clark, and J. B. Blair, "Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest," *Remote Sensing of Environment*, vol. 81, no. 2-3, pp. 378–392, 2002.

[65] K. Dralle and M. Rudemo, "Stem number estimation by kernel smoothing of aerial photos," *Canadian Journal of Forest Research*, vol. 26, no. 7, pp. 1228–1236, 1996.

[66] ——, “Automatic estimation of individual tree positions from aerial photos,” *Canadian Journal of Forest Research*, vol. 27, no. 11, pp. 1728–1736, NOV 1997.

[67] R. O. Dubayah and J. B. Drake, “Lidar remote sensing for forestry,” *Journal of Forestry*, vol. 98, no. 6, pp. 44–46, 2000.

[68] R. O. Dubayah, S. L. Sheldon, D. B. Clark, M. A. Hofton, J. B. Blair, G. C. Hurtt, and R. L. Chazdon, “Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica,” *Journal Of Geophysical Research-Biogeosciences*, vol. 115, p. G00E09, Apr. 2010.

[69] P. Duplat and G. Perrotte, *Inventaire et estimation de l’accroissement des peuplements forestiers.* Office National des Forêts, 1981.

[70] P. Duplat and M. Tran Ha, “Tarifs de cubage pour l’estimation des sapins et des épicéas dans les alpes du nord,” *ONF Bulletin Technique*, vol. 6, pp. 3–17, 1974.

[71] T. Eid, T. Gobakken, and E. Næsset, “Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses,” *Scandinavian Journal of Forest Research*, vol. 19, no. 6, pp. 512–523, 2004.

[72] M. Elmqvist, E. Jungert, F. Lantz, Å. Persson, and U. Söderman, “Terrain modelling and analysis using laser scanner data,” *International Archives of Photogrammetry and Remote Sensing*, vol. XXXIV-3/W4, pp. 219–226, 2001.

[73] T. L. Erdody and L. M. Moskal, “Fusion of lidar and imagery for estimating forest canopy fuels,” *Remote Sensing of Environment*, vol. 114, no. 4, pp. 725 – 737, 2010.

[74] J. S. Evans and A. T. Hudak, “A multiscale curvature algorithm for classifying discrete return lidar in forested environments,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 4, pp. 1029–1038, 2007.

[75] M. J. Falkowski, A. T. Hudak, N. L. Crookston, P. E. Gessler, E. H. Uebler, and A. M. S. Smith, “Landscape-scale parameterization of a tree-level forest growth model: a k-nearest neighbor imputation approach incorporating lidar data,” *Canadian Journal of Forest Research*, vol. 40, no. 2, pp. 184–199, Feb. 2010.

[76] M. J. Falkowski, A. M. S. Smith, P. E. Gessler, A. T. Hudak, L. A. Vierling, and J. S. Evans, “The influence of conifer forest canopy cover on the accuracy of two individual tree measurement algorithms using lidar data,” *Canadian Journal of Remote Sensing*, vol. 34, pp. S338–S350, 2008.

[77] M. J. Falkowski, A. M. S. Smith, A. T. Hudak, P. E. Gessler, L. A. Vierling, and N. L. Crookston, “Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data,” *Canadian Journal of Remote Sensing*, vol. 32, no. 2, pp. 153–161, 2006.

[78] A. Farid, D. C. Goodrich, R. Bryant, and S. Sorooshian, “Using airborne lidar to predict leaf area index in cottonwood trees and refine riparian water-use estimates,” *Journal of Arid Environments*, vol. 72, pp. 1–15, 2008.

[79] A. Farid, D. C. Goodrich, and S. Sorooshian, “Using airborne lidar to discern age classes of cottonwood trees in a riparian area,” *Western Journal of Applied Forestry*, vol. 21, no. 3, pp. 149–158, 2006.

[80] A. Farid, D. Rautenkranz, D. C. Goodrich, S. E. Marsh, and S. Sorooshian, “Riparian vegetation classification from airborne laser scanning data with an emphasis on cottonwood trees,” *Canadian Journal of Remote Sensing*, vol. 32, no. 1, pp. 15–18, 2006.

[81] S. Filin, “Surface classification from airborne laser scanning data,” *Computers & Geosciences*, vol. 30, no. 9-10, pp. 1033–1041, 2004.

[82] S. Filin and N. Pfeifer, “Segmentation of airborne laser scanning data using a slope adaptive neighborhood,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 2, pp. 71–80, 2006.

[83] G. Forzieri, L. Guarnieri, E. Vivoni, F. Castelli, and F. Preti, “Multiple attribute decision making for individual tree detection using high-resolution laser scanning,” *Forest Ecology and Management*, vol. 258, no. 11, pp. 2501–2510, 2009.

[84] G. W. Frazer, S. Magnussen, M. A. Wulder, and K. O. Niemann, “Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of lidar-derived estimates of forest stand biomass,” *Remote Sensing of Environment*, vol. 115, no. 2, pp. 636–649, Feb. 2011.

[85] I. Fujisaki, D. L. Evans, R. J. Moorbead, D. W. Irby, M. J. Mohammadi-Aragh, S. D. Roberts, and P. D. Gerard, “Stand assessment through lidar-based forest visualization using immersive virtual environment technology,” *Forest Science*, vol. 54, pp. 1–7, 2008.

[86] M. García, D. Riaño, E. Chuvieco, and F. M. Danson, “Estimating biomass carbon stocks for a Mediterranean forest in central Spain using lidar height and intensity data,” *Remote Sensing of Environment*, vol. 114, no. 4, pp. 816 – 830, 2010.

[87] R. Gaulton and T. J. Malthus, “Lidar mapping of canopy gaps in continuous cover forests: A comparison of canopy height model and point cloud based techniques,” *International Journal of Remote Sensing*, vol. 31, no. 5, pp. 1193–1211, 2010.

[88] D. L. A. Gaveau and R. A. Hill, “Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data,” *Canadian Journal of Remote Sensing*, vol. 29, no. 5, pp. 650–657, 2003.

[89] C. Ginisty, P. Vallet, S. Chabé-Ferret, C. Levesque, and C. Chauvin, “Disponibilités en biomasse forestière pour des usages énergétiques et industriels en France. Note de synthèse,” Cemagref, Tech. Rep., 2007.

[90] T. Gobakken and E. Næsset, “Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data,” *Scandinavian Journal of Forest Research*, vol. 19, no. 6, pp. 529–542, 2004.

[91] ——, “Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data,” *Canadian Journal of Forest Research*, vol. 39, no. 5, pp. 1036–1052, 2009.

[92] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2008.

[93] N. R. Goodwin, N. C. Coops, and D. S. Culvenor, “Assessment of forest structure with airborne lidar and the effects of platform altitude,” *Remote Sensing of Environment*, vol. 103, no. 2, pp. 140–152, 2006.

[94] ——, “Development of a simulation model to predict lidar interception in forested environments,” *Remote Sensing of Environment*, vol. 111, pp. 481–492, 2007.

[95] F. Gougeon, “A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images,” *Canadian Journal of Remote Sensing*, vol. 21, no. 3, pp. 274–284, 1995.

[96] R. F. Graf, L. Mathys, and K. Bollmann, “Habitat assessment for forest dwelling species using lidar remote sensing: Capercaillie in the Alps,” *Forest Ecology and Management*, vol. 257, no. 1, pp. 160–167, 2009.

[97] T. G. Gregoire, G. Stahl, E. Naesset, T. Gobakken, R. Nelson, and S. Holm, “Model-assisted estimation of biomass in a lidar sample survey in Hedmark county, Norway,” *Canadian Journal of Forest Research-revue Canadienne De Recherche Forestiere*, vol. 41, no. 1, pp. 83–95, Jan. 2011.

[98] S. A. Hall, I. C. Burke, D. O. Box, M. R. Kaufmann, and J. M. Stoker, “Estimating stand structure using discrete-return lidar: an example from low density, fire prone ponderosa pine forests,” *Forest Ecology and Management*, vol. 208, no. 1-3, pp. 189–209, 2005.

[99] J. A. Hamid, P. M. Mather, and R. A. Hill, “Mapping of conifer forest plantations using airborne hyperspectral and lidar data,” *Remote Sensing in Transition*, pp. 185–190, 2004.

[100] D. J. Harding, M. A. Lefsky, G. G. Parker, and J. B. Blair, “Laser altimeter canopy height profiles: methods and validation for closed-canopy, broadleaf forests,” *Remote Sensing of Environment*, vol. 76, no. 3, pp. 283–297, Jun. 2001.

[101] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*, corrected ed. Springer, July 2003.

[102] T. J. Hawbaker, N. S. Keuler, A. A. Lesak, T. Gobakken, K. Contrucci, and V. C. Radeloff, “Improved estimates of forest vegetation structure and biomass with a lidar-optimized sampling design,” *Journal of Geophysical Research-Biogeosciences*, vol. 114, p. G00E04, Sep. 2009.

[103] J. Heinzel and B. Koch, “Exploring full-waveform lidar parameters for tree species classification,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 1, pp. 152–160, Feb. 2011.

[104] M. Heurich, “Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park,” *Forest Ecology and Management*, vol. 255, no. 7, pp. 2416–2433, 2008.

[105] M. Heurich, F. Fischer, O. Knorzer, and P. Krzystek, “Assessment of digital terrain models (DTM) from data gathered with airborne laser scanning in temperate European beech (fagus sylvatica) and Norway spruce (picea abies) forests,” *Photogrammetrie Fernerkundung Geoinformation*, no. 6, pp. 473–488, 2008.

[106] M. Heurich and F. Thoma, “Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (fagus sylvatica) and Norway spruce (picea abies) forests,” *Forestry*, vol. 81, no. 5, pp. 645–661, Apr. 2008.

[107] T. Hilker, M. A. Wulder, and N. C. Coops, “Update of forest inventory data with lidar and high spatial resolution satellite imagery,” *Canadian Journal of Remote Sensing*, vol. 34, no. 1, pp. 5–12, 2008.

[108] R. A. Hill, S. A. Hinsley, and D. L. A. Gaveau, "Mapping forest pattern and structure at a landscape scale using airborne laser scanning technology," *Avian Landscape Ecology: Pure and Applied Issues in the Large-Scale Ecology of Birds*, pp. 60–67, 2002.

[109] S. A. Hinsley, R. A. Hill, P. E. Bellamy, and H. Balzter, "The application of lidar in woodland bird ecology: Climate, canopy structure, and habitat quality," *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 12, pp. 1399–1406, 2006.

[110] Y. Hirata, "The effects of footprint size and sampling density in airborne laser scanning to extract individual trees in mountainous terrain," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI (Part 8/W2)*, 2004, pp. 102–107.

[111] ——, "Relationship between tree height and topography in a chamaecyparis obtusa stand derived from airborne laser scanner data," *Journal of the Japanese Forest Society*, vol. 87, no. 6, pp. 497–503, 2005.

[112] Y. Hirata, Y. Akiyama, H. Saito, A. Miyamoto, M. Fukuda, and T. Nishizono, "Estimating forest canopy structure using helicopter-borne lidar measurement," *Advances in Forest Inventory for Sustainable Forest Management and Biodiversity Monitoring*, vol. 76, pp. 125–134, 2003.

[113] Y. Hirata, K. Sato, S. Kuramoto, and A. Sakai, "Extracting forest patch attributes at the landscape level using new remote sensing techniques - an integrated approach of high-resolution satellite data, airborne lidar data and GIS data for forest conservation," *EFI Proceedings*, no. No.51, pp. 359–367, 2004.

[114] Y. Hirata, K. Sato, A. Sakai, S. Kuramoto, and Y. Akiyama, "The extraction of canopy-understory vegetation-topography structure using helicopter-borne lidar measurement between a plantation and a broad-leaved forest," in *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, vol. 5, 2003, pp. 3222–3224.

[115] Y. Hirata, N. Furuya, M. Suzuki, and H. Yamamoto, "Airborne laser scanning in forest management: Individual tree identification and laser pulse penetration in a stand with different levels of thinning," *Forest Ecology and Management*, vol. 258, no. 5, pp. 752–760, Aug. 2009.

[116] M. Hofton, J. Minster, and J. Blair, "Decomposition of laser altimeter waveforms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 4, pp. 1989–1996, 2000.

[117] M. Hollaus, W. Dorigo, W. Wagner, K. Schadauer, B. Höfle, and B. Maier, "Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data," *International Journal of Remote Sensing*, vol. 30, no. 19, pp. 5159–5175, 2009.

[118] M. Hollaus, W. Wagner, C. Eberhofer, and W. Karel, "Accuracy of large-scale canopy heights derived from lidar data under operational constraints in a complex alpine environment," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 5, pp. 323–338, 2006.

[119] M. Hollaus, W. Wagner, B. Maier, and K. Schadauer, "Airborne laser scanning of forest stem volume in a mountainous environment," *Sensors*, vol. 7, no. 8, pp. 1559–1577, 2007.

[120] J. Holmgren, M. Nilsson, and H. Olsson, “Simulating the effects of lidar scanning angle for estimation of mean tree height and canopy closure,” *Canadian Journal of Remote Sensing*, vol. 29, no. 5, pp. 623–632, 2003.

[121] J. Holmgren and Å. Persson, “Identifying species of individual trees using airborne laser scanner,” *Remote Sensing of Environment*, vol. 90, no. 4, pp. 415–423, 2004.

[122] J. Holmgren, “Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning,” *Scandinavian Journal of Forest Research*, vol. 19, no. 6, pp. 543–553, Dec. 2004.

[123] H. Holmström, M. Nilsson, and G. Ståhl, “Simultaneous estimations of forest parameters using aerial photograph interpreted data and the k nearest neighbour method,” *Scandinavian Journal of Forest Research*, vol. 16, no. 1, pp. 67–78, 2001.

[124] M. Holopainen, A. Mäkinen, J. Rasinmäki, J. Hyyppä, H. Hyyppä, H. Kaartinen, R. Viitala, M. Vastaranta, and A. Kangas, “Effect of tree-level airborne laser-scanning measurement accuracy on the timing and expected value of harvest decisions,” *European Journal of Forest Research*, vol. 129, no. 5, pp. 899–907, Sept. 2010.

[125] C. Hopkinson, “The influence of lidar acquisition settings on canopy penetration and laser pulse return characteristics,” *2006 IEEE International Geoscience and Remote Sensing Symposium*, vol. 1-8, pp. 2420–2423, 2006.

[126] C. Hopkinson and L. Chasmer, “Testing lidar models of fractional cover across multiple forest ecozones,” *Remote Sensing of Environment*, vol. 113, no. 1, pp. 275–288, 2009.

[127] P. A. Houle, K. Q. Zhang, M. S. Ross, and M. Simard, “Use of airborne lidar for the assessment of landscape structure in the pine forests of Everglades national park,” *2006 IEEE International Geoscience and Remote Sensing Symposium*, vol. 1-8, pp. 1960–1963, 2006.

[128] S. L. Huang, S. A. Hager, K. Q. Halligan, I. S. Fairweather, A. K. Swanson, and R. L. Crabtree, “A comparison of individual tree and forest plot height derived from lidar and InSAR,” *Photogrammetric Engineering and Remote Sensing*, vol. 75, no. 2, pp. 159–167, 2009.

[129] A. T. Hudak, N. L. Crookston, J. S. Evans, D. E. Hall, and M. J. Falkowski, “Nearest neighbor imputation of species-level, plot-scale forest structure attributes from lidar data,” *Remote Sensing of Environment*, vol. 112, no. 1, pp. 2232–2245, 2008.

[130] A. T. Hudak, M. A. Lefsky, W. B. Cohen, and M. Berterretche, “Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height,” *Remote Sensing of Environment*, vol. 82, no. 2-3, pp. 397–416, 2002.

[131] G. C. Hurtt, R. Dubayah, J. Drake, P. R. Moorcroft, S. W. Pacala, J. B. Blair, and M. G. Fearon, “Beyond potential vegetation: combining lidar data and a height-structured model for carbon studies,” *Ecological Applications*, vol. 14, no. 3, pp. 873–883, 2004.

[132] P. Hyde, R. Dubayah, B. Peterson, J. B. Blair, M. Hofton, C. Hunsaker, R. Knox, and W. Walker, “Mapping forest structure for wildlife habitat analysis using waveform lidar: Validation of montane ecosystems,” *Remote Sensing of Environment*, vol. 96, no. 3-4, pp. 427–437, 2005.

[133] P. Hyde, R. Nelson, D. Kimes, and E. Levine, "Exploring lidar-radar synergy - predicting aboveground biomass in a southwestern ponderosa pine forest using lidar, SAR and InSAR," *Remote Sensing of Environment*, vol. 106, no. 1, pp. 28–38, 2007.

[134] R. J. Hyndman and Y. Fan, "Sample quantiles in statistical packages," *The American Statistician*, vol. 50, no. 4, pp. 361–365, Nov. 1996.

[135] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999.

[136] J. Hyyppä, "Feasibility for estimation of single tree characteristics using laser scanner," in *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, vol. 3, 2000, pp. 981–983.

[137] J. Hyyppä, O. Kelle, M. Lehikoinen, and M. Inkinen, "A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 5, pp. 969–975, 2001.

[138] J. Hyyppä, U. Pyysalo, H. Hyyppä, H. Haggrén, and G. Ruppert, "Accuracy of laser scanning for DTM generation in forested areas," *Laser Radar Technology and Applications V*, vol. 4035, pp. 119–130, 2000.

[139] K. Ioki, J. Imanishi, T. Sasaki, Y. Morimoto, and K. Kitada, "Estimating stand volume in broad-leaved forest using discrete-return lidar: plot-based approach," *Landscape And Ecological Engineering*, vol. 6, no. 1, pp. 29–36, Jan. 2010.

[140] M. Islam, M. Kurttila, L. Mehtätalo, and T. Pukkala, "Inoptimality losses in forest management decisions caused by errors in an inventory based on airborne laser scanning and aerial photographs," *Canadian Journal of Forest Research*, vol. 40, no. 12, pp. 2427–2438, Dec. 2010.

[141] D. Jaskierniak, P. N. J. Lane, A. Robinson, and A. Lucieer, "Extracting lidar indices to characterise multilayered forest structure using mixture distribution functions," *Remote Sensing of Environment*, vol. 115, no. 2, pp. 573–585, Feb. 2011.

[142] J. L. R. Jensen, K. S. Humes, L. A. Vierling, and A. T. Hudak, "Discrete return lidar-based prediction of leaf area index in two conifer forests," *Remote Sensing of Environment*, vol. 112, no. 10, pp. 3947–3957, 2008.

[143] A. Jochem, M. Hollaus, M. Rutzinger, and B. Höfle, "Estimation of aboveground biomass in alpine forests: A semi-empirical approach considering canopy transparency derived from airborne lidar data," *Sensors*, vol. 11, no. 1, pp. 278–295, Jan. 2011.

[144] T. Jolliffe, I., *Principal Component Analysis*, 2nd ed. Springer, 2002.

[145] B. Jutzi and U. Stilla, "Range determination with waveform recording laser systems using a Wiener filter," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 61, no. 2, pp. 95 – 107, 2006.

[146] T. Kakiuchi and H. Chikatsu, "DTM generation in forested area using multiple return pulses from airborne laser scanner," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 7447, 2009.

[147] V. R. Kane, J. D. Bakker, R. J. McGaughey, J. A. Lutz, R. F. Gersonde, and J. F. Franklin, "Examining conifer canopy structural complexity across forest ages and elevations with lidar data," *Canadian Journal of Forest Research*, vol. 40, no. 4, pp. 774–787, Apr. 2010.

[148] D. Karsky, D. Patterson, and T. Jasumback, "Evaluation of the Trimble ProXRS GPS receiver. Tech Tip 9971-2318-MTDC." U.S. Department of Agriculture, Forest Service, Missoula Technology and Development Center, Tech. Rep., 1999.

[149] A. Kato, L. M. Moskal, P. Schiess, M. E. Swanson, D. Calhoun, and W. Stuetzle, "Capturing tree crown formation through implicit surface reconstruction using airborne lidar data," *Remote Sensing of Environment*, vol. 113, no. 6, pp. 1148 – 1162, 2009.

[150] T. A. Kennaway, E. H. Helmer, M. A. Lefsky, T. A. Brandeis, and K. R. Sherrill, "Mapping land cover and estimating forest structure using satellite imagery and coarse resolution lidar in the Virgin islands," *Journal of Applied Remote Sensing*, vol. 2, 2008.

[151] L. W. Kenyi, R. Dubayah, M. Hofton, and M. Schardt, "Comparative analysis of SRTM-NED vegetation canopy height to lidar-derived vegetation canopy metrics," *International Journal of Remote Sensing*, vol. 30, no. 11, pp. 2797–2811, 2009.

[152] S. Kim, R. J. McGaughey, H.-E. Andersen, and G. Schreuder, "Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data," *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1575–1586, Aug. 2009.

[153] Y. Kim, Z. Q. Yang, W. B. Cohen, D. Pflugmacher, C. L. Lauver, and J. L. Vankat, "Distinguishing between live and dead standing tree biomass on the north rim of grand canyon national park, USA using small-footprint lidar data," *Remote Sensing of Environment*, vol. 113, no. 11, pp. 2499–2510, Nov. 2009.

[154] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[155] A. Kobler, N. Pfeifer, P. Ogrinc, L. Todorovski, K. Ostir, and S. Dzeroski, "Repetitive interpolation: A robust algorithm for dtm generation from aerial laser scanner data in forested terrain," *Remote Sensing of Environment*, vol. 108, no. 1, pp. 9–23, 2007.

[156] B. Koch, U. Heyder, and H. Weinacker, "Detection of individual tree crowns in airborne lidar data," *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 4, pp. 357–363, 2006.

[157] B. Koetz, F. Morsdorf, G. Sun, K. J. Ranson, K. Itten, and B. Allgower, "Inversion of a lidar waveform model for forest biophysical parameter estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 3, no. 1, pp. 49–53, 2006.

[158] I. Korpela, H. O. Ørka, J. Hyyppä, V. Heikkinen, and T. Tokola, "Range and AGC normalization in airborne discrete-return lidar intensity data for forest canopies," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 4, pp. 369–379, Jul. 2010.

[159] I. Korpela, H. O. Ørka, M. Maltamo, T. Tokola, and J. Hyyppä, "Tree species classification using airborne lidar - effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type," *Silva Fennica*, vol. 44, no. 2, pp. 319–339, 2010.

[160] S. Y. Kotchenova, N. V. Shabanov, Y. Knyazikhin, A. B. Davis, R. Dubayah, and R. B. Myneni, "Modeling lidar waveforms with time-dependent stochastic radiative transfer theory

for remote estimations of forest structure," *Journal of Geophysical Research-Atmospheres*, vol. 108, no. D15, 2003.

[161] S. Y. Kotchenova, X. D. Song, N. V. Shabanov, C. S. Potter, Y. Knyazikhin, and R. B. Myneni, "Lidar remote sensing for modeling gross primary production of deciduous forests," *Remote Sensing of Environment*, vol. 92, no. 2, pp. 158–172, 2004.

[162] S. Koukoulas and G. A. Blackburn, "Quantifying the spatial properties of forest canopy gaps using lidar imagery and GIS," *International Journal of Remote Sensing*, vol. 25, no. 15, pp. 3049–3071, 2004.

[163] ——, "Mapping individual tree location, height and species in broadleaved deciduous forest using airborne lidar and multi-spectral remotely sensed data," *International Journal of Remote Sensing*, vol. 26, no. 3, pp. 431–455, 2005.

[164] K. Kraus and N. Pfeifer, "Determination of terrain models in wooded areas with airborne laser scanner data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 53, no. 4, pp. 193–203, 1998.

[165] D. A. Kwak, J. Chung, W. K. Lee, M. Kafatos, S. Y. Lee, H. K. Cho, and S. H. Lee, "Evaluation for damaged degree of vegetation by forest fire using lidar and a digital aerial photograph," *Photogrammetric Engineering and Remote Sensing*, vol. 76, no. 3, pp. 277–287, Mar. 2010.

[166] D. A. Kwak, W. K. Lee, J. H. Lee, G. S. Biging, and P. Gong, "Detection of individual trees and estimation of tree height using lidar data," *Journal of Forest Research*, vol. 12, pp. 425–434, 2007.

[167] J. Langford, O. Niemann, G. W. Frazer, M. A. Wulder, and T. Nelson, "Exploring small footprint lidar intensity data in a forested environment," *2006 IEEE International Geoscience and Remote Sensing Symposium*, vol. 1-8, pp. 2416–2419, 2006.

[168] D. Leckie, F. Gougeon, D. Hill, R. Quinn, L. Armstrong, and R. Shreenan, "Combined high-density lidar and multispectral imagery for individual tree crown analysis," *Canadian Journal of Remote Sensing*, vol. 29, no. 5, pp. 633–649, 2003.

[169] A. C. Lee and R. M. Lucas, "A lidar-derived canopy density model for tree stem and crown mapping in Australian forests," *Remote Sensing of Environment*, vol. 111, pp. 493–518, 2007.

[170] H. Lee, K. C. Slatton, B. E. Roth, and W. P. Cropper, "Prediction of forest canopy light interception using three-dimensional airborne lidar data," *International Journal of Remote Sensing*, vol. 30, no. 1, pp. 189–207, 2009.

[171] ——, "Adaptive clustering of airborne lidar data to segment individual tree crowns in managed pine forests," *International Journal of Remote Sensing*, vol. 31, no. 1, pp. 117–139, 2010.

[172] J.-B. Lee, K.-Y. Yu, Y.-I. Kim, and A. Habib, "Segmentation and extraction of linear features for detecting discrepancies between lidar data strips," in *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International*, vol. 7, 2005, pp. 4954–4957.

[173] M. A. Lefsky, W. B. Cohen, S. A. Acker, G. G. Parker, T. A. Spies, and D. Harding, "Lidar remote sensing of the canopy structure and biophysical properties of douglas-fir western hemlock forests," *Remote Sensing of Environment*, vol. 70, no. 3, pp. 339–361, 1999.

[174] M. A. Lefsky, W. B. Cohen, D. J. Harding, G. G. Parker, S. A. Acker, and S. T. Gower, "Lidar remote sensing of above-ground biomass in three biomes," *Global Ecology and Biogeography*, vol. 11, no. 5, pp. 393–399, 2002.

[175] M. A. Lefsky, D. Harding, W. B. Cohen, G. Parker, and H. H. Shugart, "Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA," *Remote Sensing of Environment*, vol. 67, no. 1, pp. 83–98, 1999.

[176] M. A. Lefsky, A. T. Hudak, W. B. Cohen, and S. A. Acker, "Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest," *Remote Sensing of Environment*, vol. 95, no. 4, pp. 532–548, 2005.

[177] M. Lefsky, W. Cohen, G. Parker, and D. Harding, "Lidar remote sensing for ecosystem studies," *BioScience*, vol. 1, pp. 19–30, 2002.

[178] Y. Z. Li, H. E. Andersen, and R. McGaughey, "A comparison of statistical methods for estimating forest biomass from light detection and ranging data." *Western Journal of Applied Forestry*, vol. 23, no. 4, pp. 223–231, 2008.

[179] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, C. Thiel, J. Inglada, E. Christophe, J. Chanussot, and P. Gamba, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3857–3865, Nov. 2009.

[180] K. Lim, P. Treitz, K. Baldwin, I. Morrison, and J. Green, "Lidar remote sensing of biophysical properties of tolerant northern hardwood forests," *Canadian Journal of Remote Sensing*, vol. 29, no. 5, pp. 658–678, Oct. 2003.

[181] K. S. Lim and P. M. Treitz, "Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators," *Scandinavian Journal of Forest Research*, vol. 19, no. 6, pp. 558–570, Dec. 2004.

[182] E. Lindberg, J. Holmgren, K. Olofsson, J. Wallerman, and H. Olsson, "Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods," *International Journal of Remote Sensing*, vol. 31, no. 5, pp. 1175–1192, 2010.

[183] R. Loos and O. Niemann, "Identification of individual trees and canopy shapes using lidar data for fire management," *2006 IEEE International Geoscience and Remote Sensing Symposium*, vol. 1-8, pp. 3755–3757, 2006.

[184] J. L. Lovell, D. L. B. Jupp, D. S. Culvenor, and N. C. Coops, "Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests," *Canadian Journal of Remote Sensing*, vol. 29, no. 5, pp. 607–622, 2003.

[185] J. L. Lovell, D. L. B. Jupp, G. J. Newnham, N. C. Coops, and D. S. Culvenor, "Simulation study for finding optimal lidar acquisition parameters for forest height retrieval," *Forest Ecology and Management*, vol. 214, no. 1-3, pp. 398–412, 2005.

[186] R. M. Lucas, A. C. Lee, and P. J. Bunting, "Retrieving forest biomass through integration of CASI and lidar data," *International Journal of Remote Sensing*, vol. 29, no. 5, pp. 1553–1577, 2008.

[187] S. Magnussen and P. Boudewyn, "Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators," *Canadian Journal of Forest Research*, vol. 28, no. 7, pp. 1016–1031, 1998.

[188] S. Magnussen, E. Næsset, and T. Gobakken, "Reliability of lidar derived predictors of forest inventory attributes: A case study with Norway spruce," *Remote Sensing of Environment*, vol. 114, no. 4, pp. 700 – 712, 2010.

[189] M. Maltamo, O. Bollandsås, J. Vauhkonen, J. Breidenbach, T. Gobakken, and E. Næsset, "Comparing different methods for prediction of mean crown height in Norway spruce stands using airborne laser scanner data," *Forestry*, vol. 83, no. 3, pp. 257–268, 2010.

[190] M. Maltamo, K. Eerikainen, J. Pitkänen, J. Hyyppä, and M. Vehmas, "Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions," *Remote Sensing of Environment*, vol. 90, no. 3, pp. 319–330, 2004.

[191] M. Maltamo, J. Hyyppä, and J. Malinen, "A comparative study of the use of laser scanner data and field measurements in the prediction of crown height in boreal forests," *Scandinavian Journal of Forest Research*, vol. 21, no. 3, pp. 231–238, 2006.

[192] M. Maltamo, K. T. Korhonen, P. Packalen, L. Mehtatalo, and A. Suvanto, "Testing the usability of truncated angle count sample plots as ground truth in airborne laser scanning-based forest inventories," *Forestry*, vol. 80, no. 1, pp. 73–81, 2007.

[193] M. Maltamo, K. Mustonen, J. Hyyppä, J. Pitkänen, and X. Yu, "The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve," *Canadian Journal of Forest Research*, vol. 34, no. 9, pp. 1791–1801, 2004.

[194] M. Maltamo, P. Packalen, X. Yu, K. Eerikainen, J. Hyyppä, and J. Pitkänen, "Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data," *Forest Ecology and Management*, vol. 216, no. 1-3, pp. 41–50, 2005.

[195] M. Maltamo, K. Eerikainen, P. Packalen, and J. Hyyppa, "Estimation of stem volume using laser scanning-based canopy height metrics," *Forestry*, vol. 79, no. 2, pp. 217–229, 2006.

[196] J. W. McCombs, S. D. Roberts, and D. L. Evans, "Influence of fusing lidar and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation," *Forest Science*, vol. 49, no. 3, pp. 457–466, 2003.

[197] D. O. McInerney, J. Suarez-Minguez, R. Valbuena, and M. Nieuwenhuis, "Forest canopy height retrieval using lidar data, medium-resolution satellite imagery and kNN estimation in Aberfoyle, Scotland," *Forestry*, vol. 83, no. 2, pp. 195–206, Apr. 2010.

[198] A. J. McLane, G. J. McDermid, and M. A. Wulder, "Processing discrete-return profiling lidar data to estimate canopy closure for large-area forest mapping and management," *Canadian Journal of Remote Sensing*, vol. 35, no. 3, pp. 217–229, Jun. 2009.

[199] J. E. Means, S. A. Acker, B. J. Fitt, M. Renslow, L. Emerson, and C. J. Hendrix, "Predicting forest stand characteristics with airborne scanning lidar," *Photogrammetric Engineering and Remote Sensing*, vol. 66, no. 11, pp. 1367–1371, 2000.

[200] J. E. Means, S. A. Acker, D. J. Harding, J. B. Blair, M. A. Lefsky, W. B. Cohen, M. E. Harmon, and W. A. McKee, "Use of large-footprint scanning airborne lidar to estimate forest

stand characteristics in the western cascades of Oregon," *Remote Sensing of Environment*, vol. 67, no. 3, pp. 298–308, 1999.

[201] C. Mei and S. Durrieu, "Tree crown delineation from digital elevation models and high resolution imagery," in *International Archives Of Photogrammetry, Remote Sensing And Spatial Information Sciences, Volume Xxxvi, Part 8/W2. Proceedings of the international Conference. Laser-Scanners for Forest and Landscape Assessment - Instruments, Processing Methods and Applications. Freiburg im Breisgau. Germany.*, 2004.

[202] X. Meng, L. Wang, J. L. Silván-Cárdenas, and N. Currit, "A multi-directional ground filtering algorithm for airborne lidar," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 1, pp. 117 – 124, 2009.

[203] É. Mermin and J. Renaud, "Installation de placettes permanentes dans les forêts résineuses des Alpes du Nord : objectifs, méthode, résultats," Cemagref, Tech. Rep., 1996.

[204] É. Mermin, D. Waszak, and J. Fay, "Sylviculture par collectifs dans les forêts de montagne: principes et exemples d'application," *Revue Forestière Française*, vol. 56, no. 2, pp. 143–154, 2004.

[205] C. J. Miller, "Fusion of high resolution lidar elevation data with hyperspectral data to characterize tree canopies," *Algorithms for Multispectral, Hyperspectral and Ultraspectral Imagery VII*, vol. 4381, pp. 246–252, 2001.

[206] H. Mitasova, L. Mitas, and R. S. Harmon, "Simultaneous spline approximation and topographic analysis for lidar elevation data in open-source GIS," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 2, no. 4, pp. 375–379, 2005.

[207] N. Miura and S. Jones, "Characterizing forest ecological structure using pulse types and heights of airborne laser scanning," *Remote Sensing of Environment*, vol. 114, no. 5, pp. 1069–1076, 2010.

[208] M. Moeur and R. Stage, Albert, "Most similar neighbor: an improved sampling inference procedure for natural resource planning," *Forest Science*, vol. 41, no. 2, pp. 337–359, May 1995.

[209] T. Moffiet, K. Mengersen, C. Witte, R. King, and R. Denham, "Airborne laser scanning: Exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, no. 5, pp. 289–309, 2005.

[210] J.-M. Monnet, N. Clouet, F. Bourrier, and F. Berger, "Using geomatics and airborne laser scanning for rockfall risk zoning: a case study in the French Alps," in *Proceedings of the 2010 Canadian Geomatics Conference and Symposium of Commission I (ISPRS), Calgary Alberta, CAN, 15/06/2010*, 2010, p. 5.

[211] J.-M. Monnet, É. Mermin, J. Chanussot, and F. Berger, "Using airborne laser scanning to assess forest protection function against rockfall," in *Proceedings of the Interpraevent International Symposium in Pacific Rim, Taipei, TWN, 26/04/2010*, 2010, pp. 586–594.

[212] F. Morsdorf, O. Frey, E. Meier, K. I. Itten, and B. Allgower, "Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning," *International Journal of Remote Sensing*, vol. 29, no. 5, pp. 1387–1406, 2008.

[213] F. Morsdorf, E. Meier, B. Kotz, K. I. Itten, M. Dobbertin, and B. Allgower, “Lidar-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management,” *Remote Sensing of Environment*, vol. 92, no. 3, pp. 353–362, 2004.

[214] F. Morsdorf, A. Mårell, B. Koetz, N. Cassagne, F. Pimont, E. Rigolot, and B. Allgöwer, “Discrimination of vegetation strata in a multi-layered mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning,” *Remote Sensing of Environment*, vol. 114, no. 7, pp. 1403–1415, 2010.

[215] F. Morsdorf, E. Meier, B. Allgöwer, and D. Nüesch, “Clustering in airborne laser scanning raw data for segmentation of single trees,” in *Proceedings of the ISPRS working group III/3 workshop "3-D reconstruction from airborne laserscanner and InSAR data" Dresden, Germany 8-10 October 2003, vol. XXXIV, part 3/W13*, 2003.

[216] G. Moser and S. Serpico, “Automatic parameter optimization for support vector regression for land and sea surface temperature estimation from remote sensing data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 909–921, Mar. 2009.

[217] G. Moser and S. B. Serpico, “Modeling the error statistics in support vector regression of surface temperature from infrared data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 6, no. 3, pp. 448–452, Jul 2009.

[218] J. D. Muss, D. J. Mladenoff, and P. A. Townsend, “A pseudo-waveform technique to assess forest structure using discrete lidar data,” *Remote Sensing of Environment*, vol. 115, no. 3, pp. 824–835, Mar. 2011.

[219] J. Müller and R. Brandl, “Assessing biodiversity by remote sensing in mountainous terrain: the potential of lidar to predict forest beetle assemblages,” *Journal of Applied Ecology*, vol. 46, no. 4, pp. 897–905, 2009.

[220] E. Naesset and T. Jonmeister, “Assessing point accuracy of DGPS under forest canopy before data acquisition, in the field and after postprocessing,” *Scandinavian Journal of Forest Research*, vol. 17, no. 4, pp. 351–358, 2002.

[221] R. F. Nelson, P. Hyde, P. Johnson, B. Emessiene, M. L. Imhoff, R. Campbell, and W. Edwards, “Investigating radar-lidar synergy in a North Carolina pine forest,” *Remote Sensing of Environment*, vol. 110, no. 1, pp. 98–108, 2007.

[222] R. Nelson, W. Krabill, and G. MacLean, “Determining forest canopy characteristics using airborne laser data,” *Remote Sensing of Environment*, vol. 15, no. 3, pp. 201–212, Jun. 1984.

[223] M. Nilsson, “Estimation of tree heights and stand volume using an airborne lidar system,” *Remote Sensing of Environment*, vol. 56, no. 1, pp. 1–7, 1996.

[224] H. Niska, J. P. Skön, P. Packalén, T. Tokola, M. Maltamo, and M. Kolehmainen, “Neural networks for the prediction of species-specific plot volumes using airborne laser scanning and aerial photographs,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1076–1085, Mar. 2010.

[225] T. Nord-Larsen and T. Riis-Nielsen, “Developing an airborne laser scanning dominant height model from a countrywide scanning survey and national forest inventory data,” *Scand. J. For. Res.*, vol. 25, no. 3, pp. 262–272, 2010.

[226] E. Næsset, “Determination of mean tree height of forest stands using airborne laser scanner data,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 52, no. 2, pp. 49–56, 1997.

[227] ——, “Estimating timber volume of forest stands using airborne laser scanner data,” *Remote Sensing of Environment*, vol. 61, no. 2, pp. 246–253, 1997.

[228] ——, “Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data,” *Remote Sensing of Environment*, vol. 80, no. 1, pp. 88–99, Apr. 2002.

[229] ——, “Accuracy of forest inventory using airborne laser scanning: Evaluating the first nordic full-scale operational project,” *Scandinavian Journal of Forest Research*, vol. 19, no. 6, pp. 554–557, 2004.

[230] ——, “Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser,” *Remote Sensing of Environment*, vol. 91, no. 2, pp. 243–255, 2004.

[231] ——, “Practical large-scale forest stand inventory using a small-footprint airborne scanning laser,” *Scandinavian Journal of Forest Research*, vol. 19, no. 2, pp. 164–179, 2004.

[232] ——, “Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia,” *Scandinavian Journal of Forest Research*, vol. 22, pp. 433–442, 2007.

[233] E. Næsset and K. O. Bjerknes, “Estimating tree heights and number of stems in young forest stands using airborne laser scanner data,” *Remote Sensing of Environment*, vol. 78, no. 3, pp. 328–340, 2001.

[234] E. Næsset, O. M. Bollandsås, and T. Gobakken, “Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data,” *Remote Sensing of Environment*, vol. 94, no. 4, pp. 541–553, Feb. 2005.

[235] E. Næsset and T. Gobakken, “Estimating forest growth using canopy metrics derived from airborne laser scanner data,” *Remote Sensing of Environment*, vol. 96, no. 3-4, pp. 453–465, 2005.

[236] E. Næsset and R. Nelson, “Using airborne laser scanning to monitor tree migration in the boreal-alpine transition zone,” *Remote Sensing of Environment*, vol. 110, no. 3, pp. 357–369, 2007.

[237] E. Næsset and T. Økland, “Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve,” *Remote Sensing of Environment*, vol. 79, no. 1, pp. 105–115, 2002.

[238] E. Næsset, “Assessing sensor effects and effects of leaf-off and leaf-on canopy conditions on biophysical stand properties derived from small-footprint airborne laser data,” *Remote Sensing of Environment*, vol. 98, no. 2-3, pp. 356 – 370, 2005.

[239] C. D. Oliver and B. C. Larson, *Forest stand dynamics.* John Wiley & Sons, 1996.

[240] P. Ozenda, *La végétation de la chaîne alpine dans l’espace montagnard européen.* Paris: Masson, 1985.

[241] H. O. Ørka, E. Næsset, and O. M. Bollandsås, “Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data,” *Remote Sensing of Environment*, vol. 113, no. 6, pp. 1163 – 1174, 2009.

[242] ——, “Effects of different sensors and leaf-on and leaf-off canopy conditions on echo distributions and individual tree properties derived from airborne laser scanning,” *Remote Sensing of Environment*, vol. 114, no. 7, pp. 1445–1461, Jul. 2010.

[243] P. Packalén and M. Maltamo, “Predicting the plot volume by tree species using airborne laser scanning and aerial photographs,” *Forest Science*, vol. 52, no. 6, pp. 611–622, 2006.

[244] ——, “The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs,” *Remote Sensing of Environment*, vol. 109, no. 3, pp. 328–341, Aug. 2007.

[245] R. M. Palenichka and M. B. Zaremba, “Multiscale isotropic matched filtering for individual tree detection in lidar images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, pp. 3944–3956, 2007.

[246] Y. Pang, Z.-Y. Li, G.-Q. Sun, M. Lefsky, X.-J. Che, and E.-X. Chen, “Effects of terrain on the large footprint lidar waveform of forests,” *Forest Research*, vol. 20, no. 4, pp. 464–468, 2007.

[247] C. E. Parrish, “Exploiting full-waveform lidar data and multiresolution wavelet analysis for vertical object detection and recognition,” *Geoscience and Remote Sensing Symposium. IGARSS’07 IEEE International*, vol. 1-12, pp. 2499–2502, 2007.

[248] C. Pascual, A. Garcia-Abril, W. B. Cohen, and S. Martin-Fernandez, “Relationship between lidar-derived forest canopy height and Landsat images,” *International Journal of Remote Sensing*, vol. 31, no. 5, pp. 1261–1280, 2010.

[249] C. Pascual, A. Garcia-Abril, L. G. Garcia-Montero, S. Martin-Fernandez, and W. B. Cohen, “Object-based semi-automatic approach for forest structure characterization using lidar data in heterogeneous pinus sylvestris stands,” *Forest Ecology and Management*, vol. 255, no. 11, pp. 3677–3685, 2008.

[250] G. Patenaude, R. A. Hill, R. Milne, D. L. A. Gaveau, B. B. J. Briggs, and T. P. Dawson, “Quantifying forest above ground carbon content using lidar remote sensing,” *Remote Sensing of Environment*, vol. 93, no. 3, pp. 368–380, 2004.

[251] G. Perrin, X. Descombes, and J. Zerubia, “A marked point process model for tree crown extraction in plantations,” in *Proc. IEEE International Conference on Image Processing (ICIP)*, Genova, Sep. 2005.

[252] Å. Persson, J. Holmgren, and U. Soderman, “Detecting and measuring individual trees using an airborne laser scanner,” *Photogrammetric Engineering and Remote Sensing*, vol. 68, no. 9, pp. 925–932, 2002.

[253] J. Peuhkurinen, M. Maltamo, and J. Malinen, “Estimating species-specific diameter distributions and saw log recoveries of boreal forests from airborne laser scanning data and aerial photographs: a distribution-based approach,” *Silva Fennica*, vol. 42, no. 4, pp. 625–641, 2008.

[254] S. C. Popescu, R. H. Wynne, and R. F. Nelson, “Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size,” *Computers and Electronics in Agriculture*, vol. 37, no. 1-3, pp. 71–95, Dec. 2002.

[255] ——, “Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass,” *Canadian Journal of Remote Sensing*, vol. 29, no. 5, pp. 564–577, 2003.

[256] S. C. Popescu, R. H. Wynne, and J. A. Scrivani, “Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA,” *Forest Science*, vol. 50, no. 4, pp. 551–565, 2004.

[257] S. C. Popescu and K. Zhao, “A voxel-based lidar method for estimating crown base height for deciduous and pine trees,” *Remote Sensing of Environment*, vol. 112, no. 3, pp. 767–781, 2008.

[258] S. Popescu and R. Wynne, “Seeing the trees in the forest: using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height,” *Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 5, pp. 589–604, 2004.

[259] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011.

[260] G. T. Raber, J. R. Jensen, S. R. Schill, and K. Schuckman, “Creation of digital terrain models using an adaptive lidar vegetation point removal process,” *Photogrammetric Engineering and Remote Sensing*, vol. 68, no. 12, pp. 1307–1315, 2002.

[261] J. Reitberger, P. Krzystek, and U. Stilla, “Analysis of full waveform lidar data for the classification of deciduous and coniferous trees,” *International Journal of Remote Sensing*, vol. 29, no. 5, pp. 1407–1431, 2008.

[262] J. Reitberger, C. Schnorr, P. Krzystek, and U. Stilla, “3D segmentation of single trees exploiting full waveform lidar data,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 6, pp. 561–574, Nov. 2009.

[263] J. Renaud, C. Rupé, C. Chauvin, É. Mermin, D. Leclerc, and J. Fay, “Les forêts résineuses à fonction de protection dans les Alpes du Nord françaises : typologie générale des structures et méthode descriptive, analyse de la stabilité des peuplements et interventions sylvicoles,” Cemagref, Tech. Rep., 1994.

[264] D. Riano, E. Chuvieco, S. Condes, J. Gonzalez-Matesanz, and S. L. Ustin, “Generation of crown bulk density for pinus sylvestris l. from lidar,” *Remote Sensing of Environment*, vol. 92, no. 3, pp. 345–352, 2004.

[265] D. Riano, E. Meier, B. Allgower, E. Chuvieco, and S. L. Ustin, “Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling,” *Remote Sensing of Environment*, vol. 86, no. 2, pp. 177–186, 2003.

[266] J. J. Riggins, J. A. Tullis, and F. M. Stephen, “Per-segment aboveground forest biomass estimation using lidar-derived height percentile statisticsnaesset,” *Giscience & Remote Sensing*, vol. 46, no. 2, pp. 232–248, Apr. 2009.

[267] S. D. Roberts, T. J. Dean, D. L. Evans, J. W. McCombs, R. L. Harrington, and P. A. Glass, “Estimating individual tree leaf area in loblolly pine plantations using lidar-derived

measurements of height and crown dimensions," *Forest Ecology and Management*, vol. 213, no. 1-3, pp. 54–70, 2005.

[268] C. Salas, L. Ene, T. Gregoire, E. Næsset, and T. Gobakken, "Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models," *Remote Sensing of Environment*, vol. 114, no. 6, pp. 1277–1285, 2010.

[269] T. Sasaki, J. Imanishi, K. Ioki, Y. Morimoto, and K. Kitada, "Estimation of leaf area index and canopy openness in broadleaved forest using an airborne laser scanner in comparison with high-resolution near-infrared digital photography," *Landscape and Ecological Engineering*, vol. 4, no. 1, pp. 47–55, 2008.

[270] A. Schaeffer, A. Gazin, and A. D'Alverny, *Sapinières : le jardinage par contenance (Méthode du contrôle par les courbes).* Les Presses Universitaires de France, 1930.

[271] J. O. Sexton, T. Bax, P. Siqueira, J. J. Swenson, and S. Hensley, "A comparison of lidar, radar, and field measurements of canopy height in pine and hardwood forests of southeastern North America," *Forest Ecology and Management*, vol. 257, no. 3, pp. 1136–1147, Feb. 2009.

[272] K. R. Sherrill, M. A. Lefsky, J. B. Bradford, and M. G. Ryan, "Forest structure estimation and pattern exploration from discrete-return lidar in subalpine forests of the central Rockies," *Canadian Journal of Forest Research*, vol. 38, no. 8, pp. 2081–2096, 2008.

[273] N. S. Skowronski, K. L. Clark, M. Duveneck, and J. Hom, "Three-dimensional canopy fuel loading predicted using upward and downward sensing lidar systems," *Remote Sensing of Environment*, vol. 115, no. 2, pp. 703–714, Feb. 2011.

[274] A. M. S. Smith, M. J. Falkowski, A. T. Hudak, J. S. Evans, A. P. Robinson, and C. M. Steele, "A cross-comparison of field, spectral, and lidar estimates of forest canopy cover," *Canadian Journal of Remote Sensing*, vol. 35, no. 5, pp. 447–459, Oct. 2009.

[275] G. Smith, Å. Persson, J. Holmgren, B. Hallberg, J. E. S. Fransson, and L. M. H. Ulander, "Forest stem volume estimation using high-resolution lidar and SAR data," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International*, vol. 4, 2002, pp. 2084–2086.

[276] P. Soille, *Morphological Image Analysis: Principles and Applications.* Springer-Verlag, 2003.

[277] S. Solberg, E. Næsset, and O. M. Bollandsås, "Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest," *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 12, pp. 1369–1378, 2006.

[278] S. Solberg, A. Brunner, K. H. Hanssen, H. Lange, E. Næsset, M. Rautiainen, and P. Stenberg, "Mapping LAI in a Norway spruce forest using airborne laser scanning," *Remote Sensing of Environment*, vol. 113, no. 11, pp. 2317–2327, Nov. 2009.

[279] B. St-Onge, J. Jumelet, M. Cobello, and C. Vega, "Measuring individual tree height using a combination of stereophotogrammetry and lidar," *Canadian Journal of Forest Research*, vol. 34, no. 10, pp. 2122–2130, 2004.

[280] B. St-Onge, C. Vega, R. A. Fournier, and Y. Hu, "Mapping canopy height using a combination of digital stereo-photogrammetry and lidar," *International Journal of Remote Sensing*, vol. 29, no. 11, pp. 3343–3364, 2008.

[281] G. Stahl, S. Holm, T. G. Gregoire, T. Gobakken, E. Naesset, and R. Nelson, "Model-based inference for biomass estimation in a lidar sample survey in Hedmark county, Norway," *Canadian Journal of Forest Research-revue Canadienne De Recherche Forestiere*, vol. 41, no. 1, pp. 96–107, Jan. 2011.

[282] J. Su and E. Bork, "Influence of vegetation, slope, and lidar sampling angle on DEM accuracy," *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 11, pp. 1265–1274, 2006.

[283] J. C. Suarez, C. Ontiveros, S. Smith, and S. Snape, "Use of airborne lidar and aerial photography in the estimation of individual tree heights in forestry," *Computers & Geosciences*, vol. 31, no. 2, pp. 253–262, 2005.

[284] D. Sun, Y. Li, and Q. Wang, "A unified model for remotely estimating chlorophyll a in lake Taihu, China, based on SVM and in situ hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2957–2965, Aug. 2009.

[285] G. Q. Sun and K. J. Ranson, "Modeling lidar returns from forest canopies," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 6, pp. 2617–2626, Nov. 2000.

[286] G. Q. Sun, K. J. Ranson, D. W. Liu, and B. Koetz, "Simulation studies of forest structure using 3D lidar and radar models," *Geoscience and Remote Sensing Symposium, IGARSS'07. IEEE International*, vol. 1-12, pp. 2562–2565, 2007.

[287] A. Suratno, C. Seielstad, and L. Queen, "Tree species identification in mixed coniferous forest using airborne laser scanning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 6, pp. 683–693, 2009.

[288] T. Takahashi, Y. Awaya, Y. Hirata, N. Furuya, T. Sakai, and A. Sakai, "Stand volume estimation by combining low laser-sampling density lidar data with QuickBird panchromatic imagery in closed-canopy Japanese cedar (cryptomeria japonica) plantations," *International Journal of Remote Sensing*, vol. 31, no. 5, pp. 1281–1301, 2010.

[289] T. Takahashi, K. Yamamoto, Y. Senda, and M. Tsuzuku, "Estimating individual tree heights of sugi (cryptomeria japonica d. don) plantations in mountainous areas using small-footprint airborne lidar," *Journal of Forest Research*, vol. 10, no. 2, pp. 135–142, 2005.

[290] R. Q. Thomas, G. C. Hurtt, R. Dubayah, and M. H. Schilz, "Using lidar data and a height-structured ecosystem model to estimate forest carbon stocks and fluxes over mountainous terrain," *Canadian Journal of Remote Sensing*, vol. 34, pp. S351–S363, 2008.

[291] V. Thomas, P. Treitz, J. H. McCaughey, and I. Morrison, "Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: an examination of scanning density," *Canadian Journal of Forest Research*, vol. 36, no. 1, pp. 34–47, 2006.

[292] S. Tonolli, M. Dalponte, L. Vescovo, M. Rodeghiero, L. Bruzzone, and D. Gianelle, "Mapping and modeling forest tree volume using forest inventory and airborne laser scanning," *European Journal of Forest Research*, vol. 130, no. 4, pp. 569–577, Jul. 2011.

[293] M. Tran-Ha, G. Perrotte, T. Cordonnier, and P. Duplat, "Volume tige d'un arbre ou d'une collection d'arbres pour six essences principales en france," *Revue Forestière Française*, vol. 59, no. 6, pp. 609–624, 2007.

[294] R. N. Treuhaft, F. G. Goncalves, J. B. Drake, B. D. Chapman, J. R. dos Santos, L. V. Dutra, P. M. L. A. Graca, and G. H. Purcell, "Biomass estimation in a tropical wet forest using Fourier transforms of profiles from lidar or interferometric SAR," *Geophysical Research Letters*, vol. 37, p. L23403, Dec. 2010.

[295] R. Treuhaft and S. Cloude, "The structure of oriented vegetation from polarimetric interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2620–2624, 1999.

[296] *GPS Pathfinder Office User Guide Volume 1*, Version 2.80 part number 31310-28-ENG revision B ed., Trimble, March 2001.

[297] M. Törmä, "Estimation of tree species proportions of forest stands using laser scanning," in *International Archives of Photogrammetry and Remote Sensing*, vol. 33, 2000, pp. 1524–1531.

[298] J. Vauhkonen, "Estimating crown base height for scots pine by means of the 3D geometry of airborne laser scanning data," *International Journal of Remote Sensing*, vol. 31, no. 5, pp. 1213–1226, 2010.

[299] J. Vauhkonen, I. Korpela, M. Maltamo, and T. Tokola, "Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics," *Remote Sensing of Environment*, vol. 114, no. 6, pp. 1263–1276, Jun. 2010.

[300] J. Vauhkonen, T. Tokola, M. Maltamo, and P. Packalen, "Effects of pulse density on predicting characteristics of individual trees of Scandinavian commercial species using alpha shape metrics based on airborne laser scanning data," *Canadian Journal of Remote Sensing*, vol. 34, pp. S441–S459, 2008.

[301] C. Vega and B. St-Onge, "Height growth reconstruction of a boreal forest canopy over a period of 58 years using a combination of photogrammetric and lidar models," *Remote Sensing of Environment*, vol. 112, no. 4, pp. 1784–1794, 2008.

[302] M. Vehmas, K. Eerikainen, J. Peuhkurinen, P. Packalen, and M. Maltamo, "Identification of boreal forest stands with high herbaceous plant diversity using airborne laser scanning," *Forest Ecology and Management*, vol. 257, no. 1, pp. 46–53, 2009.

[303] U. Vepakomma, D. Kneeshaw, and B. St-Onge, "Interactions of multiple disturbances in shaping boreal forest dynamics: a spatially explicit analysis using multi-temporal lidar data and high-resolution imagery," *Journal of Ecology*, vol. 98, no. 3, pp. 526–539, May 2010.

[304] U. Vepakomma, B. St-Onge, and D. Kneeshaw, "Spatially explicit characterization of boreal forest gap dynamics using multi-temporal lidar data," *Remote Sensing of Environment*, vol. 112, no. 5, pp. 2326–2340, 2008.

[305] L. Vierling, E. Rowell, X. Chen, D. Dykstra, and K. Vierling, "Relationships among airborne scanning lidar, high resolution multispectral imagery, and ground-based inventory data in a ponderosa pine forest," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International*, vol. 5, 2002, pp. 2912–2914.

[306] W. Wagner, M. Hollaus, C. Briese, and V. Ducic, "3D vegetation mapping using small-footprint full-waveform airborne laser scanners," *International Journal of Remote Sensing*, vol. 29, no. 5, pp. 1433–1452, 2008.

[307] W. Wagner, A. Ullrich, V. Ducic, T. Melzer, and N. Studnicka, "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 2, pp. 100–112, 2006.

[308] J. Wallerman and J. Holmgren, "Estimating field-plot data of forest stands using airborne laser scanning and SPOT HRG data," *Remote Sensing of Environment*, vol. 110, no. 4, pp. 501–508, 2007.

[309] E. D. Wallington and J. C. Suarez, "Evaluation of commercial airborne lidar and SAR products to estimate top height and associated parameters in production forests in Britain," *Sustainable forestry: from monitoring and modelling to knowledge management and policy science*, pp. 298–313, 2007.

[310] C. Wang and N. F. Glenn, "Integrating lidar intensity and elevation data for terrain characterization in a forested area," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 6, no. 3, pp. 463–466, Jul. 2009.

[311] L. Wang and H. Chu, "Graph theoretic segmentation of airborne lidar data," *Independent Component Analyses, Wavelets, Unsupervised Nano-Biomimetic Sensors, and Neural Networks VI*, vol. 6979, p. N9790, 2008.

[312] Y. Wang, H. Weinacker, and B. Koch, "A lidar point cloud based procedure for vertical canopy structure analysis and 3D single tree modelling in forest," *Sensors*, vol. 8, no. 6, pp. 3938–3951, 2008.

[313] T. C. Weber and D. E. Boss, "Use of lidar and supplemental data to estimate forest maturity in Charles county, MD, USA," *Forest Ecology and Management*, vol. 258, no. 9, pp. 2068–2075, Oct. 2009.

[314] C. Weed, M. Crawford, A. Neuenschwander, and R. Gutierrez, "Classification of lidar data using a lower envelope follower and gradient-based operator," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International*, vol. 3, 2002, pp. 1384–1386.

[315] J. F. Weishampel, J. B. Blair, R. G. Knox, R. Dubayah, and D. B. Clark, "Volumetric lidar return patterns from an old-growth tropical rainforest canopy," *International Journal of Remote Sensing*, vol. 21, no. 2, pp. 409–415, 2000.

[316] M. G. Wing, A. Eklund, and J. Sessions, "Applying lidar technology for tree measurements in burned landscapes," *International Journal of Wildland Fire*, vol. 19, no. 1, pp. 104–114, 2010.

[317] M. A. Wulder and D. Seemann, "Forest inventory height update through the integration of lidar data with segmented Landsat imagery," *Canadian Journal of Remote Sensing*, vol. 29, no. 5, pp. 536–543, 2003.

[318] M. A. Wulder, J. C. White, F. Alvarez, T. Han, J. Rogan, and B. Hawkes, "Characterizing boreal forest wildfire with multi-temporal Landsat and lidar data," *Remote Sensing of Environment*, vol. 113, no. 7, pp. 1540–1555, Jul. 2009.

[319] P. Yong, G. Sun, and L. Zengyuan, "Effects of forest spatial structure on large footprint lidar waveform," in *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, vol. 7, 2004, pp. 4738–4741.

[320] X. W. Yu, J. Hyyppä, H. Kaartinen, and M. Maltamo, "Automatic detection of harvested trees and determination of forest growth using airborne laser scanning," *Remote Sensing of Environment*, vol. 90, no. 4, pp. 451–462, 2004.

[321] X. W. Yu, J. Hyyppä, A. Kukko, M. Maltamo, and H. Kaartinen, "Change detection techniques for canopy height growth measurements using airborne laser scanner data," *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 12, pp. 1339–1348, 2006.

[322] H. Zhan, P. Shi, and C. Chen, "Retrieval of oceanic chlorophyll concentration using support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 12, pp. 2947–2951, Dec. 2003.

[323] K. Q. Zhang, "Identification of gaps in mangrove forests with airborne lidar," *Remote Sensing of Environment*, vol. 112, no. 5, pp. 2309–2325, 2008.

[324] K. G. Zhao and S. Popescu, "Lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA," *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1628–1645, Aug. 2009.

[325] K. G. Zhao, S. Popescu, and R. Nelson, "Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers," *Remote Sensing of Environment*, vol. 113, no. 1, pp. 182–196, 2009.

[326] D. A. Zimble, D. L. Evans, G. C. Carlson, R. C. Parker, S. C. Grado, and P. D. Gerard, "Characterizing vertical forest structure using small-footprint airborne lidar," *Remote Sensing of Environment*, vol. 87, no. 2-3, pp. 171–182, 2003.

Résumé

De nombreux travaux ont montré le potentiel de la télédétection par scanner laser aéroporté pour caractériser les massifs forestiers. Cependant, l'application aux forêts complexes de montagne reste encore peu documentée. On se propose donc de tester les deux principales méthodes permettant d'extraire des paramètres forestiers sur des données acquises en zone montagneuse et de les adapter aux contraintes spécifiques à cet environnement. En particulier on évaluera d'une part l'apport conjoint de la régression à vecteurs de support et de la réduction de dimension pour l'estimation de paramètres de peuplement, et d'autre part l'intérêt d'un apprentissage non supervisé pour la détection d'arbres.

Mots-clés: Télédétection, scanner laser aéroporté, LiDAR, forêt, vecteurs de support, apprentissage non supervisé, inventaire forestier, détection d'arbre.

Abstract

Numerous studies have shown the potential of airborne laser scanning for the mapping of forest resources. However, the application of this remote sensing technique to complex forests encountered in mountainous areas requires further investigation. In this thesis, the two main methods used to derive forest information are tested with airborne laser scanning data acquired in the French Alps, and adapted to the constraints of mountainous environments. In particular, a framework for unsupervised training of treetop detection is proposed, and the performance of support vector regression combined with dimension reduction for forest stand parameters estimation is evaluated.

Keywords: Remote sensing, airborne laser scanning, LiDAR, forest mapping, support vector regression, unsupervised training, treetop detection.