Contents

1 Introduction **1**
1.1 Study of forested areas . . . 2
1.2 Remote sensing for forested areas . . . 2
1.3 Context of the thesis . . . 4
1.4 Objectives . . . 4
1.5 Strategy . . . 5
1.6 Structure of the thesis . . . 6

2 State of the art **9**
2.1 Stand segmentation . . . 10
2.2 Segmentation . . . 13
2.2.1 Traditional segmentation methods . . . 14
2.2.2 Segmentation of point cloud . . . 14
2.2.3 Superpixels methods . . . 15
2.3 Classification . . . 15
2.3.1 Supervised classification . . . 16
Random Forest . . . 16
2.3.2 Dimension reduction and feature selection . . . 18
Dimension reduction . . . 18
Feature selection . . . 18
2.4 Smoothing methods . . . 20
2.4.1 Local methods . . . 20
2.4.2 Global methods . . . 21

3 Method **25**
3.1 General flowchart . . . 26
3.2 Over-segmentation . . . 26
3.2.1 Segmentation of lidar data . . . 28
3.2.2 Segmentation of optical images . . . 28

3.3 Feature extraction . . . 29
3.3.1 Point-based lidar features. . . . 29
3.3.2 Pixel-based multispectral features. . . . 30
3.3.3 Pixel-based lidar features. . . . 31
3.3.4 Object-based feature map. . . . 31
3.4 Classification . . . 32
3.4.1 Training set design . . . 32
Feature selection . . . 33
3.5 Smoothing . . . 34
3.5.1 Local methods . . . 35
Filtering . . . 35
Probabilistic relaxation . . . 35
3.5.2 Global smoothing . . . 36
Unary term . . . 37
Prior . . . 37
Energy minimization . . . 39

4 Results **41**

5 Conclusion and perspectives **43**

A Publication **45**
A.1 Journal articles . . . 46
A.2 Peer-reviewed conference papers . . . 46

Bibliography **49**

1 Introduction

1.1 Study of forested areas 2
1.2 Remote sensing for forested areas 2
1.3 Context of the thesis 4
1.4 Objectives 4
1.5 Strategy 5
1.6 Structure of the thesis 6

1.1 Study of forested areas

Forests are an important component of planet's life. They represent 30% of the world land surface. They also hold about 90% of terrestrial biodiversity. Forests are also benefit for the environment; they capt and store the CO_2 [22]. About 45% of the total global carbon is held by forests. They also filter dust and microbial pollution of the air [94]. Finally, They also play an important role in hydrological regulation and water purification [65].

Forest are complex structures [86], for which informations are needed for management. The information can be the tree species or the tree maturity of the forest. There are two ways to extract such information from forest; field inventory or remote sensing. The field inventories are very expensive to set up and are also not adapted for a national study. A more adapted to obtain such information is remote sensing since it allows to extract them at a large scale.

1.2 Remote sensing for forested areas

The analysis of forested areas from a remote sensing point of view can be performed at three different levels: pixel, object (mainly trees) or stand. In statistical national forest inventory (NFI), an automated and accurate tree segmentation is needed in order to extract tree level features (basal area, dominant tree height, etc., [77, 48]). However, the tree level is not the only reliable level of analysis for forest studies. When a joint mapping and statistical reasoning is required (e.g., land-cover (LC) mapping and forest inventory), forest stands remain the prevailing scale of analysis [77, 114]. A stand can be defined in many different ways in terms of homogeneity: tree specie, age, height, maturity, and its definition varies according to the countries.

From a remote sensing point of view, the delineation of the stands is a segmentation problem. Forest stands are interesting in order to extract reliable and statistically meaningful features and to provide an input for multi-source statistical inventory. For land-cover mapping, this is highly helpful

for forest database updating [50], whether the labels of interest are *vegetated areas* (e.g., *deciduous/evergreen/mixed/non-forested)*, or, more precisely, the tree species. Most of the time in national forestry inventory institutes, for reliability purposes, each area is manually interpreted by human operators with very high resolution (VHR) geospatial images focusing on the infrared channel [48]. This work is extremely time consuming and subjective [117]. Furthermore, in many countries, the wide variety of tree species (e.g., $>$20) significantly complicates the problem. The design of an automatic procedure based on remote sensing data would fasten such process. Additionally, the standard manual delineation procedure only takes into account the species, and few characteristics (alternatively height, age, stem density or crown closure), while an automatic method could offer more flexibility and would allow to combine characteristics extracted from all complementary data sources.

The use of remote sensing data for the automatic analysis of forests has been growing in the last 15 years, especially with the synergistic use of airborne laser scanning (ALS) and optical VHR imagery (multispectral imagery and hyperspectral imagery) [102, 114]. They appear to be both well adapted and complementary inputs for stand segmentation [16, 14, 64]. ALS provides a direct access to the vertical distribution of the trees and to the ground underneath. Hyperspectral and multispectral images are particularly relevant for tree species classification: spectral and textural information from VHR images can allow a fine discrimination of many species, respectively. Multispectral images are often preferred due to their higher availability, and higher spatial resolution.

A prerequisite for data fusion is the most accurate alignment of the two data [102]. A frequently used technique is to geo-rectify images using ground controls points (GCPS). A geometric transformation is established between the coordinates of GCPs and their corresponding pixels in the image. It is then applied for each pixel, so that coordinate differences on those checkpoints are reduced to the lowest possible level. This method can be easily applied and is relatively fast in terms of computation time. However the use of GCPs can still cause that the unknowns in the trajectory of the platforms

produce some remarkable residual errors. Automatic methods for data registration have also been developed [37, 75].

1.3 Context of the thesis

In France, the study of forests is two fold. They need to be mapped and inventoried. The forest inventory allows to obtain the wood stock at a national scale. Statistics such as volume per hectare, deciduous volume or conifer volume can then be derived. The inventory is performed through field inventory and extrapolated using the forest mapping. Thus, the mapping of forest is very important in order to derive accurate statistics.
The forest mapping is given by a national forest LC database. It is manually interpreted by human operators with VHR infra-red colored (IRC) orthoimages. It assigns a vegetation type to each mapped beach of more than 5000 m^2. The nomenclature is composed of 32 classes based on hierarchical criteria such as pure stands of the main tree species of the French forest. The forest LC should be updated in a 10 years cycle.

1.4 Objectives

Currently, the forest LC is obtained through remote sensing (namely photo-interpretation), an method could be developed to update it automatically. Since an old version of the forest LC is available, it can be used as a ground truth input subsequent classification [34]. However, the learning process should be carried out carefully [35]. Indeed, some area might have change (e.g. forest cuts). Furthermore, the database is designed generalized [93]. A simple classification would then not be sufficient in order to retrieve homogeneous patches similar to the forest LC. Such results could be obtained using smoothing methods [90]. Furthermore, an automatic method would allow to enrich the LC, i.e. retrieve homogeneous tree species stands also homogeneous in terms of height [33].

1.5 Strategy

Two remote sensing modalities are available for the mapping of forested areas at IGN; VHR optical images and lidar cloud points. The VHR images are a part of a national database. In this thesis, the images used have a spatial resolution of 50 cm. Two type of ortho-images are available, a color image (3 bands; red: 600-720 nm, green: 490-610 nm and blue: 430-550 nm) and and IRC image (3 bands; near infra-red: 750-950 nm, red and green) captured by the IGN digital cameras [95]. It is then possible to obtain four band ortho-images by the combination of the two ortho-images type.
IGN also process lots of test flight over forested areas with a laser scanning device. The airborne lidar data were collected using an Optech 3100EA device. The footprint was 0.8 m in order to increase the probability to reach the ground. The point density for all echoes ranges from 2 to 4 points/m^2.

The registration between airborne lidar point clouds and VHR multispectral images was performed by IGN itself using ground control points. This is a standard procedure in the French mapping agency since IGN operates both sensors and has also a strong expertise in data georeferencing (this is in fact the national institute responsible for that in France for both airborne and spaceborne sensors).
Data were acquired under leaf-on conditions and fit with the standards used in many countries for large-scale operational forest mapping purposes.

The combination of these two data is very relevant for the study of forest, indeed, optical images provide the major information about the tree species, while lidar give information about the vertical structure of the forest. Furthermore, the lidar allows to extract consistent object such as trees.

In order to extract more information from these two modalities, the fusion should be performed at different levels. 3 levels could be defined:

- Low level: It corresponds to the fusion of the observations, in this case, only the reflectance from the optical images and the height of the lidar points.

- Medium level: It corresponds to the fusion of the features, they are derived at the same level (e.g. the pixel) and merged together. It also corresponds to the cooperative understanding of the data; a feature is derived on a modality (e.g. trees from lidar) and use on the other.
- High level: It corresponds to the fusion of decision. One or many classifications have been performed and the final decision is a smart combination of the classifications and the input data.

1.6 Structure of the thesis

- State of the art: Chapter 2
- Method: Chapter 3
- Results: Chapter 4
- Conclusion and perspectives: Chapter 5

2

State of the art

2.1 Stand segmentation 10
2.2 Segmentation 13
2.2.1 Traditional segmentation methods 14
2.2.2 Segmentation of point cloud 14
2.2.3 Superpixels methods 15
2.3 Classification 15
2.3.1 Supervised classification 16
Random Forest 16
2.3.2 Dimension reduction and feature selection 18
Dimension reduction 18
Feature selection 18
2.4 Smoothing methods 20
2.4.1 Local methods 20
2.4.2 Global methods 21

2.1 Stand segmentation

One should note that the literature remains focused on individual tree extraction and tree species classification, developing site-specific workflows with similar advantages, drawbacks and classification performance. More authors have focused on forest delineation [21], that do not convey information about the tree species and their spatial distribution. Consequently, no operational framework embedding the automatic analysis of remote sensing data has been yet proposed in the literature for forest stand segmentation [18].

In the large amount of literature in the field, only few papers focus on the issue of stand segmentation or delineation. They can be categorized with regard to the type of data processed.
First, stand segmentation can be achieved with a single remote sensing source. A stand delineation technique using VHR airborne multispectral imagery is proposed in [63]. The trees are extracted using a valley following approach and classified into 7 tree species (5 coniferous, 1 deciduous, and 1 non-specified) with a maximum likelihood classifier. A semi-automatic iterative clustering procedure is then introduced to generate the forest polygons.
A hierarchical and multi-scale approach for the identification of stands is adopted in [40]. The data inputs were the 4 bands of an airborne 0.5 m orthoimage (Red, Green, Blue, and Near Infra-Red) allowing to derive the Normalized Difference Vegetation Index (NDVI). The stand mapping solution is based on the Object-Based Image Analysis concept. It is composed of two main phases in a cyclic process: first, segmentation, then classification. The first level consists in over-segmenting the area of interest and performing fine-grained land cover classification. The second level aims to transfer the vegetation type provided by a land cover geodatabase in the stand polygons, already retrieved from another segmentation procedure. The multi-scale analysis appears to have a significant benefit on the stand labeling but it is highly heuristic and requires a correct definition of the stand while we consider it is an interleaved problem.
A seminal stand mapping method using low density airborne lidar data

is proposed in [53]. It is composed of several steps of feature extraction, creation and raster-based classification. Forest stands are created by grouping neighboring cells within each class. Then, only the stands with a predefined minimum size are accepted. Neighboring small areas of different forest types that do not reach the minimum size are merged together to an existing forest stand. The approach offers the advantage of detecting 15 forest types that match very well with the ground truth but to the detriment of simplicity: the flowchart has to be highly reconsidered to fit to other stand specifications. Additionally, the tree species discrimination is not addressed.

The forest stand delineation proposed in [98] also uses low density airborne lidar still coupling an object-oriented image segmentation and a supervised classification procedure. Three features are computed and rasterized. The segmentation is performed using a region growing approach. Spatially adjacent pixels are grouped into homogeneous discrete image objects or regions. Then, a supervised discrimination of the segmented image is performed using a Battacharya classifier, in order to determine the maturity of the stands. The tree species are ignored and the procedure requires a careful inspection of the raw data both for feature generation and model training.

Following the work of [117] with IKONOS images, Quickbird-2 panchromatic images are used in [79] to automatically delineate forest stands. A standard image segmentation technique is used and the novelty mainly lies on the fact that its initial parameters are optimized with respect to NFI protocols. They show that meaningful stand heights can be derived, which are a critical input for various modeled inventory attributes.

The method proposed in [21] aims to generate a forest mask (*forested area* label only) using low density airborne lidar. A Canopy Height Model (CHM) with a spatial resolution of 1 m is derived. The positions and heights of single trees are determined from the CHM using a local maximum filter, based on a moving window approach. Only detected positions with a CHM height superior to 3 m are considered. The crown radii are estimated using an empirical function. The three neighboring trees are connected using a Delaunay triangulation applied to the previously-detected tree position. The crown cover is then calculated using the crown areas of three neighboring trees and the area of their convex hull for each tree triple. The forest mask

is derived from the canopy cover values. While this is not a genuine stand delineation method, this approach could be easily extended to a multi-class problem and enlightens the necessity of individual tree extraction even with limited point densities as a basis for the stand-level analysis.
A forest stand delineation also based on airborne lidar data is proposed in [116]. Three features are first directly extracted from the point cloud. A coarse forest stand delineation is then performed on the feature image using the unsupervised Mean-Shift algorithm, in order to obtain under-segmented raw forest stands. A forest mask is then applied to the segmented image in order to retrieve forest and non-forest raw stands. It may create some small isolated areas, iteratively merged to their most similar neighbor until their size is larger than a user-defined threshold in order to product big raw forest stands. They are then refined into finer level using a seeded region growing based on superpixels. The idea is to select several different superpixels in a raw forest stand and merge them. This method provides a coarse-to-fine segmentation with relatively large stands. The process was only applied on a small area of a forest in Finland, thus, general conclusions can not be drawn.

Secondly, several methods fusing various types of remote sensing data have also been developed. The analysis of the lidar and multispectral data is performed at three levels in [100], following a given hierarchical nomenclature of classes in forested environments. The first level represents small objects (single tree scale, individual trees or small groups of trees) that can be differentiated by spectral and structural characteristics using a rule-based classification. The second level corresponds to the stand level. It is built using the same classification process which summarizes forest development phases by referencing to small scale sub-objects at level 1. The third level is generated by merging objects of the same classified forest-development into larger spatial units. The multi-scale analysis offers the advantage of alleviating the standard issue of individual tree crown detection and proposing development stage labels. Nevertheless, the pipeline is highly heuristic, under-exploits lidar data and significant confusion between classes are reported.
The automatic segmentation process of forests in [20] is also supplied with

lidar and VHR multispectral images. The idea is to divide the forests into higher and lower sections with lidar. An unsupervised classification process is applied to the two new images. The final stand delineation is achieved by segmenting the classification results with pre-defined thresholds. The segmentation results are improved using morphological operators such as opening and closing, which fill the gaps and holes at a specified extent. This method is efficient if the canopy structure is homogeneous and requires a strong knowledge on the area of interest. Since it is based on height information only, it cannot differentiate two stands of similar height but different species.
In [66] a stand segmentation technique for a forest composed of *Scots Pine*, *Norway Spruce* and *Hardwood* is defined. A hierarchical segmentation on the Crown Height Model followed by a restricted iterative region growing approach is performed on images composed of rasterized lidar data and Colored Infra-Red images. The process was only applied on a limited area of Finland and prevents from drawing strong conclusions. However, the quantitative analysis carried out by the authors shows that lidar data can help to define statistically meaningful stands (here the criterion was the timber volume) and that multispectral images are inevitable inputs for tree species discrimination.

2.2 Segmentation

The direct segmentation of the optical image and/or the lidar point clouds is not sufficient in order to retrieve forest stands. However, with adapted parameters, segmentations algorithms might be useful to obtain relevant over-segmentation of the data [18]. They can be divide in two categories:

- The pure segmentation methods, in theses methods, a specific attention must be paid to the choice of the parameter in order to obtain a relevant over-segmentation. Such segmentation can be applied on an image or a point cloud. Specific methods have also been developed for the segmentation of lidar point cloud.

- The superpixels segmentation methods, they natively produce an over-segmentation of the image. The parameters control the size and the shape of the resulting segments.

2.2.1 Traditional segmentation methods

The segmentation of an image can be performed using number of techniques [83].
The easiest way to segment an image is the thresholding of a gray level histogram of the image [99]. When the image is noisy or the background is uneven and illumination is poor, such thresholding might be not sufficient. Thus, adaptive thresholding methods have been developed [118].

The segmentation can be considered as an unsupervised classification problem. Algorithms dealing with such problems use iterative process. The most popular algorithm is the k-means algorithm. Segmentation methods using the spatial interaction models like Markov Random Field (CRF) [38] or Gibbs Random Field (GRF) [19]. Neural networks are also interesting for image segmentation [28] as they take into account the contextual information.

Lastly, the segmentation of an image can be obtained by the detection of the edges of the image [85]. The idea is to extract points of significant changes in depth values. Edges are local features and are determined based on local information.

2.2.2 Segmentation of point cloud

The segmentation of point cloud has been highly assessed [81]. The aim is to extract meaningful objects. Such extraction has two principal objectives:

- Objects are detected so as to ease or strengthen subsequent classification task. A precise extraction is not mandatory since the labels would be refined after.

- Objects are precisely delineated in order to derive features from these objects. A high spatial resolution is therefore expected.

In forested areas, the only reliable objects to extract are trees. The first way to extract trees from lidar data is to rasterize the point cloud and use image-based segmentation techniques to obtain trees. Several methods have been developed for single tree delineation [15, 108, 47].

2.2.3 Superpixels methods

Several superpixels algorithms have been developed [1]. They group pixels into perceptually meaningful atomic regions. Many traditional segmentation algorithms have been employed with more or less success to generate superpixels [92, 24, 13, 107, 110]. These algorithms produce satisfactory results, however, they may be relatively slow and the number, size and shape of the superpixels might not be specified.

Superpixels algorithms have then been developed. One can control the number of superpixel, their size and their shape. [78] creates superpixels based on a grid. Optimal path are found using graph cut methods. [109] proposes a generation of superpixels based on a global optimization. They are obtained by stitching together overlapping image patches such that each pixel belongs to only one of the overlapping regions. [67] generate superpixels by a dilatation of a set of seed locations using level-set geometric flow. Resulting superpixels are constrained to have uniform size, compactness, and boundary adherence. Finally, [1] proposes a generation of superpixels based on the k-means algorithms. A weighted distance that combines color and spatial proximity is introduced in order to control the size and the compactness of the superpixels.

2.3 Classification

A classification is a process that aim to categorize observation. The idea is to assign an observation to one or more classes. This can be done manually or algorithmically. The classification can be unsupervised, the classes

need to be learned and the observation assigned. Such classification is similar to segmentation (see section 2.2). The classification can be supervised, the target classes are known and observations with labels are available.

2.3.1 Supervised classification

A great number of supervised classification algorithms have been developed and used for remote sensing issues [60, 70, 76]. There are two kind of algorithms: the parametric and the non-parametric methods.

The parametric method assume that each class follow a specific distribution (mainly gaussian). The parameters of the distribution are estimated using the learning set. This is the case for the maximum likelihood [96], maximum a posteriori [23] or in [103].

The non parametric method do not make any assumption on the classes distribution. In this category of algorithms, the most popular are the Support Vector Machines (SVM) [7, 91] and the Random Forest (RF) [9]. The artificial neural networks are also efficient algorithms [39, 3]. Simpler methods exist, such as the k-nearest neighbor [42] or the decision trees [10]. The non parametric methods are more efficient for the discrimination of complex classes [84, 27], and are considered as a basis for land cover classification [11].

We chose to use the RF, which besides their widespread use, since they also offer the possibility of obtaining the probability of belonging of a pixel to a class. This posterior probabilities can be then integrated into a smoothing process. The RF are described in the following paragraph.

Random Forest

The RF have been introduced by [9] and are defined by the aggregation of collection of predictor (decision trees). Here, we refer to the RF with random inputs proposed in this article.

The idea is to create an ensemble of samples $\mathcal{S}_n^{\Theta_1}$, ..., $\mathcal{S}_n^{\Theta_k}$. A Classification and Regression Tree (CART) [10] is built on each sample $\mathcal{S}_n^{\Theta_i}$. Each tree is built using a a random pool of m features. The final classification is obtained by majority vote; each tree vote for a class, the class with the most vote wins (see Figure 2.1. This algorithm has two parameters: the number of trees k

and the number of features m used to build a tree. The first parameter is arbitrary fixed to a high value. The second is generally fixed to the square root of the total number of feature [29].

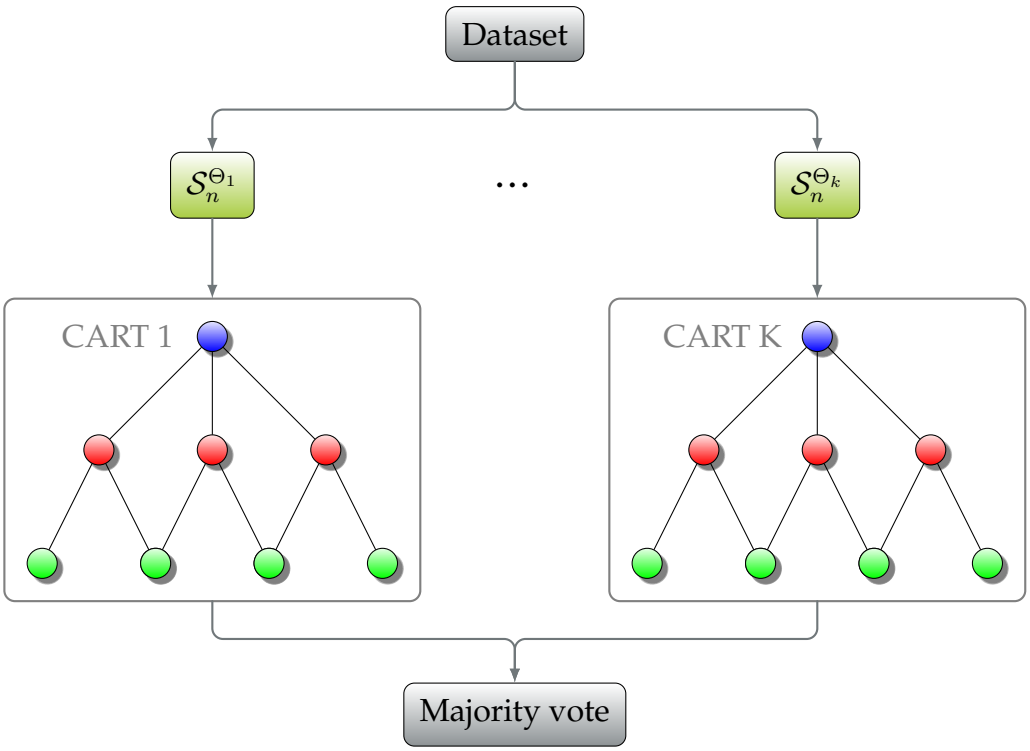

FIGURE 2.1: General diagram of the operation of the Random Forest

RF have shown better classification performances than traditional Boosting methods [9] or SVM [82]. They are also able to handle big dataset with large number of feature. Furthermore, a measure of feature importance have been introduced in [9]. It allows to qualify the relevance of the feature in the classification process [97].

The importance of a feature $\mathbf{X}_j$, $j \in \{1, ..., q\}$ (with q the number of feature) is defined as follow. Let $\mathcal{S}_n^{\Theta_i}$ be a sample and OOB_i all the observations that does not belong to $\mathcal{S}_n^{\Theta_i}$. $errOOB_i$, the error on OOB_i using $\mathcal{S}_n^{\Theta_i}$, is then computed. A random permutation on the value of the j^{th} feature of OOB_i is performed in order to obtain $\widetilde{OOB_i}^j$. $err\widetilde{OOB_i^j}$ is then computed. The importance of the feature j, $FI(\mathbf{X_j})$ is the mean of the difference of the errors (see Equation 2.1).

$$FI(\mathbf{X_j}) = \frac{1}{k}\sum_{i=1}^{k}(err\widetilde{OOB}_i^j - errOOB_i) \tag{2.1}$$

where k is the number of CART.

2.3.2 Dimension reduction and feature selection

The feature selection methods try to overcome the curse of high dimensionality [6, 41]. Indeed, the increasing number of feature available tends to decrease the accuracy of the classifiers. Furthermore, the computation times increase with the number of feature. Thus, reducing the feature dimension is beneficial for the classification task.

Two approaches exist, first the ones based on the transformation of the data, generally using a projection in a space of lower dimensionality. Secondly, there are approaches based on the feature selection, that aim to search for an optimal subset of the features.

Dimension reduction

The most popular dimension reduction method is the Principal Component Analysis (PCA). It is an unsupervised method that aim to maximize the variance between data [44]. However, it has been demonstrated that PCA is not optimal for the purpose of classification [12]. Other methods have been developed based on the PCA: the Independent Component Analysis (ICA) [46] maximize the statistical independence between data, and the Maximum Autocorrelation Factor (MAF) [61] maximize the spatial auto-correlation. When training samples are available, the linear discriminant analysis try to maximize the intra-class homogeneity and the inter-class variance [26, 62].

Feature selection

The feature selection aims to search for an optimal subset of features without modifying them. To obtain such subset, one can explore the subsets of features or define a criteria to evaluate the subsets. Furthermore, the selection can be supervised or unsupervised. The first aims to discriminate the better the classes while the second are looking for an optimal subset that

regroup the data into homogeneous classes while maximizing the distance between classes. The unsupervised methods are more related to clustering algorithms.

Many exploration methods for feature selection have been proposed in the literature. The naive exhaustive exploration of all the subsets can be envisaged when the number of feature is not important. Thus, many non-exhaustive methods for the exploration of subsets have been proposed. There are two kinds of methods : the heuristic methods and the random methods.

Using heuristic rules for feature selection allows to quickly converge to an optimal subset. It could be approaches that add features step by step (forward approaches), also called Sequential Forward Selection (SFS) [74]. It could also be methods that start from the entire feature set and remove feature step by step (backward approaches), also called Sequential Backward Selection (SBS) [115]. A generalization of these methods have been proposed in [52]. Finally, the forward and backward methods could be combined in order to improve the process. The Sequential Floating Forward Selection (SFFS) and the Sequential Floating Backward Selection (SFBS) [87] propose such improvement.

The search can be performed randomly. The generation of the subset can be totally random [69]. Genetic algorithms propose a ponderation of the subsets according to their importance [30]. They allow a faster convergence to a more stable solution.

The other methods for feature selection are the evaluation methods. They estimate a feature or a subset of feature in order to evaluate the different methods of feature selection. Three kinds of methods have been proposed : the filter, wrapper and embedded methods.

The filter methods weight the features according to their capacity to bring together elements of the same class and separate the different elements [43]. Such methods are independent from a classifier and are used as preliminary step to classification. Such methods are very fast but, since they are independent from the classifier, they are not adapted for a specific

classification task.

The wrapper methods weight the features according to their pertinence for the prediction [54]. This weighting is related to the performance of a classifier. Data are separated into two subset. The first is used for the training, while the second evaluate the classifier. The use of a classifier is a big advantage as it fits more to the envisaged problem. However, the use of a classifier significantly increase the computation times.

Eventually, the embedded methods select the features during the training process. They have two advantages: since they use the data as training, they are robust. Furthermore, the feature selection and the classification are performed together, thus, they are faster than the wrapper methods. Many methods have been proposed. The RF allow to assess to the feature importance [9]. Other methods are based on the SVM classifiers, the SVM-RFE (Recursive Feature Elimination) [105] recursively removes the less pertinent features according to a weight estimated with a SVM.

2.4 Smoothing methods

Pixel-wise classification is not sufficient for both accurate and smooth land-cover mapping with VHR remote sensing data. This is particularly true in forested areas: the large intra-class and low inter-class variabilities of classes result in noisy label maps at pixel or tree levels. This is why various regularization solutions can be adopted from the literature (from simple smoothing to probabilistic graphical models).
According to [89], both local and global methods can provide a regularization framework, with their own advantages and drawbacks.

2.4.1 Local methods

In local methods, the neighborhood of each element is analyzed by a filtering technique. The labels of the neighboring pixels (or the posterior class probabilities) are combined so as to derive a new label for the central pixel. Majority voting, Gaussian and bilateral filtering can be employed if it is not targeted to smooth class edges.
The probabilistic relaxation is an other local smoothing method that aims

at homogenizing probabilities of a pixel according to its neighboring pixels. The relaxation is an iterative algorithm in which the probability at each pixel is updated at each iteration in order to have it closer to the probabilities of its neighbors [31]. It reports good accuracies with decent computing time and offers an alternative to edge aware/gradient-based techniques that may not be adapted in semantically unstructured environments.

2.4.2 Global methods

Global methods consider the full area of interest at the same time. They are based on Markov Random Fields (MRF), the labels at different locations are not considered to be independent. The optimal configuration of labels is retrieved when finding the Maximum A Posteriori over the entire field [80]. The problem is therefore considered as the minimization procedure of an energy E over the full image I. Despite a simple neighborhood encoding (pairwise relations are often preferred), the optimization procedure propagates over large distances. Depending on the formulation of the energy, the global minimum may be reachable. However, a large range of optimization techniques allow to reach local minima close to the real solution, in particular for random fields with pairwise terms [57]. For genuine structured predictions, in the family of graphical probabilistic models, Conditional Random Fields (CRF) have been massively adopted during the last decade. Interactions between neighboring objects, and subsequently the local context can be modeled and learned. In particular, Discriminative Random Fields (DRF, [58]) are CRF defined over 2D regular grids, and both unary/association and binary/interaction potentials are based on labeling procedure outputs. Many techniques extending this concept or focusing on the learning or inference steps have been proposed in the literature [55, 59]. A very recent trend even consists in jointly considering CRF and deep-learning techniques for the labeling task [51].
In standard LC classification tasks, global methods are known to provide significantly more accurate results [89] since contextual knowledge is integrated. This is all the more true for VHR remote sensing data, especially in case of a large number of classes (e.g., 10, [2]), but presents two disadvantages. For large datasets, their learning and inference steps are expensive to

compute. Furthermore, parameters should often be carefully chosen for optimal performance, and authors that managed to alleviate the latter problem still report a significant computation cost [71].

3

Method

3.1 General flowchart 26
3.2 Over-segmentation 26
3.2.1 Segmentation of lidar data 28
3.2.2 Segmentation of optical images 28
3.3 Feature extraction 29
3.3.1 Point-based lidar features. 29
3.3.2 Pixel-based multispectral features. 30
3.3.3 Pixel-based lidar features. 31
3.3.4 Object-based feature map. 31
3.4 Classification 32
3.4.1 Training set design 32
Feature selection 33
3.5 Smoothing 34
3.5.1 Local methods 35
Filtering 35
Probabilistic relaxation 35
3.5.2 Global smoothing 36
Unary term 37
Prior 37
Energy minimization 39

3.1 General flowchart

With respect to the methods mentioned above, it appears that there are no forest stand segmentation method, based on tree species, that can satisfactorily handle a large number of classes (>5). The proposed framework is a fully automatic and modular method for species-based forest stand segmentation. The method is composed of four main steps; over-segmentation feature computation, vegetation type (mainly tree species) classification and regularization (see Figure 3.1).

Features are first derived at the pixel and at the object level. The most relevant ones are subsequently selected in a supervised way. The objects are extracted using various segmentation methods, since they appear to be sufficient for subsequent steps. A classification is performed at the object level as it significantly improves the discrimination results (about 10% better than the pixel-based approach). This classification is then smoothed. The smoothing may produce homogeneous vegetation type (mainly tree species) areas with smooth borders. The contributions of this method are two-fold:

- Such framework can be fed with specific constraints allowing to tailor the results to specific criteria (height, age, specie, maturity, density, ...).
- Here, the training set is automatically derived from an existing forest land-cover geodatabase. Specific attention is paid to the extraction of the most relevant training pixels, which is highly challenging with outdated and generalized vector databases.

3.2 Over-segmentation

The over-segmentation aims to extract small object that are consistent according to the input data. They are detected so as to ease or strengthen subsequent classification task. A precise extraction is not mandatory since the labels would be refined after.

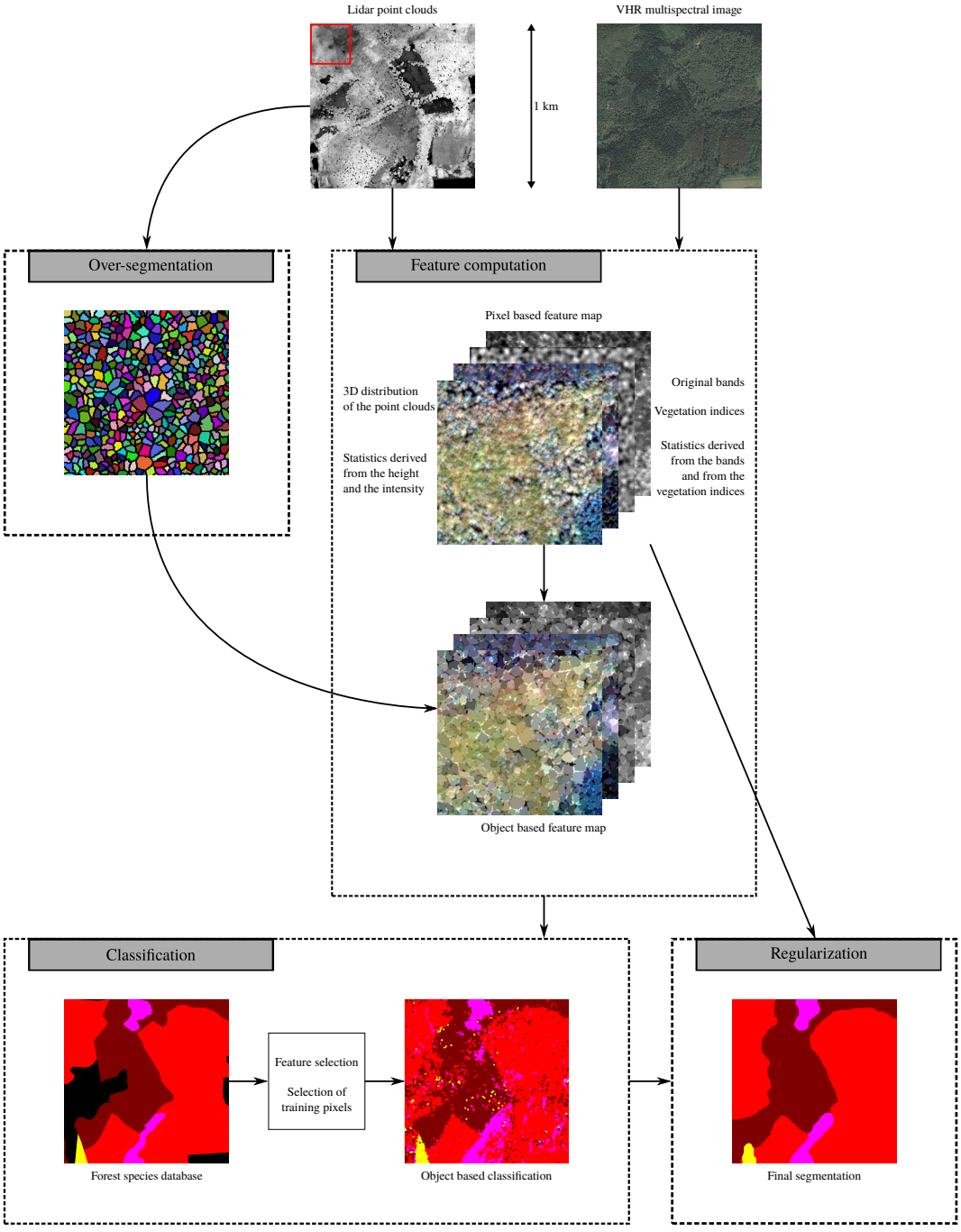

FIGURE 3.1: Flowchart of the proposed method.

3.2.1 Segmentation of lidar data

Two approaches could be envisaged: the direct segmentation of the point cloud or the segmentation of a rasterized lidar feature using image-based segmentation algorithms.

The tree extraction from the point cloud is a complex task that has been widely discussed [15, 108, 47]. However, a precise tree extraction is not needed here, since the extracted trees are only needed to improve the classification task. A coarse method is therefore adopted: the tree tops are first extracted from the lidar point clouds using a local maximum filter. A point is considered as a tree top when it has the highest height value within a 5 meter radius. Only the points above 3 meters are retained as it is a common threshold of the literature [21], and appears to be highly discriminative in non-urban areas. Points belonging to a tree are obtained through two criteria. (i) If the height of a point within a 5 m radius is greater or equal than 80% the height of the closest tree top, it is aggregated to the tree top. (ii) If the distance in the (x, y) plane between an unlabeled point and the closest tree point is smaller than 3 m they are also aggregated. This delineation method allows to discard low vegetation, but buildings might be extracted and considered as trees.

The image-based segmentation are also very efficient for the over-segmentation of lidar data. They are mainly applied on the normalized digital surface model (height). Thus a method using a single band is needed. The watershed algorithm [110] with specific parameters allow to obtain quickly a consistent over-segmentation of the image. A hierarchical segmentation [36] is more adapted since only one parameter that control the segmentation level needs to be provided.

3.2.2 Segmentation of optical images

Several algorithms have been developed for the over-segmentation of optical RGB (Red-Green-Blue) images. The most common are the superpixels methods [1].

3.3 Feature extraction

3.3.1 Point-based lidar features.

Lidar-derived features require a consistent neighborhood for their computation. For each lidar point, 3 cylindrical neighborhoods, aligned with the vertical axis, are used (1 m, 3 m and 5 m radii, infinite height). A cylinder appears to be the most relevant environment in forested areas so as to take into account the variance of altitudes of the lidar points. Three radius values are considered so as to handle the various sizes of the trees and assuming a feature selection step will prune the initial set of attributes. Two vegetation density features, $\mathcal{D}_1$ and $\mathcal{D}_2$, are computed: the first one based on the number of local maxima within the neighborhoods, and the second one related to the number of non-ground points within the neighborhoods (ground points were previously determined by a filtering step). $\mathcal{D}_1$ and $\mathcal{D}_2$ are calculated as follows:

$$\mathcal{D}_1 = \sum_{r_1 \in \{1,3,5\}} \sum_{r_2 \in \{1,3,5\}} Nt_{r_1,r_2}, \tag{3.1}$$

$$\mathcal{D}_2 = \frac{1}{3} \sum_{r \in \{1,3,5\}} \frac{Ns_r}{Ntot_r}, \tag{3.2}$$

where Nt_{r_1,r_2} is the number of local maxima retrieved from a r_1 maximum filter within the cylindrical neighborhood of radius r_2. Ns_r is the number of points classified as ground points within the cylindrical neighborhood of radius r and $Ntot_r$ is the total number of points within the cylindrical neighborhood of radius r. Additionally, the scatter $\mathcal{S}$ and the planarity $\mathcal{P}$ features are computed following Weinmann et al. [113]:

$$\mathcal{S} = \frac{1}{3} \sum_{r \in \{1,3,5\}} \frac{\lambda_{3,r}}{\lambda_{1,r}}, \tag{3.3}$$

$$\mathcal{P} = \frac{1}{3} \sum_{r \in \{1,3,5\}} 2 \times (\lambda_{2,r} - \lambda_{3,r}), \tag{3.4}$$

where $\lambda_{1,r} \geq \lambda_{2,r} \geq \lambda_{3,r}$ are the eigenvalues of the covariance matrix within the cylindrical neighborhood of radius r. They are retrieved with a standard Principal Component Analysis.

Statistical features, known to be relevant for vegetation type (mainly tree species) classification [15, 101], are also derived. For each lidar point, the same 3 cylindrical neighborhoods are used. Two basic information from the lidar data, namely height and intensity, are used to derive statistical features. A statistical feature f_d, derived from an original feature f_o, (height or intensity) is computed as follows:

$$f_d = \frac{1}{3} \sum_{r \in \{1,3,5\}} f_s(\mathbf{p}_{\mathbf{r},\mathbf{f_o}}), \tag{3.5}$$

where f_s is a statistical function (minimum; maximum; mean; median; standard deviation; median absolute deviation from median (medADmed); mean absolute deviation from median (meanADmed); skewness; kurtosis; 10^{th}, 20^{th}, 30^{th}, 40^{th}, 50^{th}, 60^{th}, 70^{th}, 80^{th}, 90^{th} and 95^{th} percentiles), and $\mathbf{p}_{\mathbf{r},\mathbf{f_o}}$ a vector containing the sorted values of the original feature f_o within the cylindrical neighborhoods of radius r. All the statistical functions are used for the height. Only the mean is used for the intensity: it is hard to know how well the sensor is calibrated and a suitable correction of intensity values within tree canopies has not yet been proposed.
24 features are extracted during this step; 2 related to vegetation density, 2 related to the 3D local distribution of the point cloud (planarity and scatter), and 20 statistical features.

3.3.2 Pixel-based multispectral features.

The original 4 spectral bands of the image are kept and considered as multispectral features. The Normalized Difference Vegetation Index (NDVI), [104], the Difference Vegetation Index (DVI), [4] and the Ratio Vegetation Index (RVI) [45] are computed as they are relevant vegetation indices. Indeed, they can provide more information about the species than the original bands alone [119]. As the point-based lidar features, statistical features are also derived from each band and each vegetation index according to Equation 3.5 (3 circular neighborhoods of 1 m, 3 m and 5 m radii). Other statistical functions are used (minimum; maximum; mean; median; standard deviation; mean absolute deviation from median (meanADmed); mean absolute deviation from mean (meanADmean); median absolute deviation from median

(medADmed); median absolute deviation from mean (medADmean). Finally, the pixel-based multispectral feature set is composed of 70 attributes.

3.3.3 Pixel-based lidar features.

The lidar features are rasterized at the same resolution of the multispectral image using a pit-free method proposed in [49]. This rasterization method is interesting because it produces smooth images that will lead to better results for classification and regularization [68]. Such data fusion process at the feature level is valid since both datasets have approximately the same spatial resolution. The Canopy Height Model (CHM) is also computed using this method, at the same spatial resolution using an existing 1 m Digital Terrain Model provided with the filtered point cloud [25]. The CHM is very important as it allows to derive the height above the ground and is known as a very discriminative feature for classification [73, 112].

3.3.4 Object-based feature map.

The pixel-based multispectral and lidar maps are merged so as to obtain a pixel-based feature map. Then, an object-based feature map is created using the over-segmentation and the pixel-based feature map. The value v_t of a pixel belonging to an object t in the object-based feature map is computed as follows:

$$v_t = \frac{1}{N_t} \sum_{p \in t} v_p, \tag{3.6}$$

where N_t is the number of pixels in object t ,and v_p is the value of the pixel p. If a pixel does not belong to a tree, it keeps the value of the pixel-based feature map. Here, only the mean value of the pixels within the tree is envisaged but one can also consider other statistical values (minimum, maximum, percentiles etc.).

Other morphological features could also be directly derived from the lidar cloud point at the object-level. For instance, an alpha-shape can be performed on the individual trees [106] and a penetration feature can be derived as it can help to classify vegetation type (mainly tree species). However, low point densities (1-5 points/m^2) compatible with large-scale lidar

surveys are not sufficient in order to derive a significant penetration indicator.

3.4 Classification

The classification is performed using a supervised classifier, in order to discriminate the vegetation type (mainly tree species) provided by an existing forest land-cover database. The classifier used in this study is the Random Forest (RF), implemented in OpenCV [8], as it has been shown relevant in the literature [5] and in a previous study compared to SVM [17], since it provide similar results while being faster. The outputs of the classification are (i) label map and (ii) probability map (posterior class probabilities for each pixel/object). This probability map is the main input for the subsequent regularization step.
In order to reduce the computation times, a feature selection is carried out to identify an "optimal" feature subset. Additionally, a strategy is proposed in order to select the most suitable training pixels for an existing land-cover forest maps, subsequently improving the classification accuracy.

3.4.1 Training set design

Using an existing forest land-cover (LC) database for training a model is not straightforward [32, 88, 72]. First, locally it can suffer from a lack of information (not all the classes of interest are present). Secondly, this database may also be semantically and, more frequently, geometrically incorrect (see Figure **??**): changes may have happened (forest cut or grow) and the geodatabase may have been generalized, resulting in sharp polygon vertices that do not exactly correspond to the class borders. Thirdly, in many forest LC databases, polygons of a given vegetation type (mainly tree species) may contain other vegetation type (mainly tree species) in a small proportion. Two strategies are employed to overcome these problems:

Firstly, in order to increase the knowledge on existing class labels, the model could be trained on a larger area. The size of the training area has been chosen arbitrarily. However, when it is too large, we observe that the quality of the classification decreases. The optimal choice of the training

area has not been investigated yet. The model has therefore been trained on a larger area than the ones of interest. The areas selected for the training are the ones maximizing the number of classes within a 5 km search zone.

Secondly, in order to correct the potential errors of the LC database, a k-means clustering has been therefore performed on each of the labels in the training area. We assume that erroneous pixels are present in a small proportion and that therefore the main cluster corresponds to the class of interest. Let $p_{i-c,t}$ be the i^{th} pixel of the vegetation type (mainly tree species) t in the cluster c of the k-means. The pixels P_t used to train the model for the vegetation type (mainly tree species) t correspond to the set:

$$P_t = \{p_{i-c,t} \quad | \quad c = \underset{c\in[1,k]}{\text{argmax}}\, \text{Card}(\cup_i p_{i-c,t})\}. \tag{3.7}$$

That is to say, only samples belonging to the main k-mean cluster among training pixels for one class are kept in the training dataset.
In practice, $k = 3$: the main cluster corresponds to the label of interest whereas the two other ones correspond to the ground and minority vegetation type (mainly tree species) within the polygons. 1000 samples per class are then randomly selected in order to design the final training set.

Feature selection

Due to the high number of features involved, an automatic Feature Selection (FS) has been integrated. This selection is composed of two steps: the choice of the number of features to select and the feature selection itself. Indeed, the choice of the number of features is very important because it enables to greatly decrease the computation times.
The Sequential Forward Floating Search (SFFS) [87] algorithm is used for both steps. The SFFS algorithm has two main advantages: (i) it can be used with many classification score (in this study, the Kappa coefficient), (ii) it enables to access to the evolution of the classification score/accuracy according to the number of selected features. The accuracy of the classification is assessed through the Kappa coefficient of the RF classifier. The SFFS algorithm selects p features by maximizing FS score criterion (the Kappa coefficient). In order to retrieve the optimal number of features, the SFFS algorithm is performed n times on different training sets with p equal to the

total number of features (95). The classification accuracy is conserved for each selection of s features ($s \in [1, p]$) and averaged over the n iterations. The number of optimal features n_{opt} corresponds to the size of the selection of s features having the maximal mean accuracy. In order to reduce the computation times, the optimal number of features was computed for a single 1 km^2 area and used for all the areas of interest.
The feature selection is then carried out for each area of interest (one selection for each area) with $p = n_{\text{opt}}$. The selected features are used for both the classification and the energy minimization framework. The feature selection could be carried out only once on a single area or on multiple areas in order to reduce the computation times.

3.5 Smoothing

The proposed method assumes that a label map is provided for the areas of interest, and is accompanied with a class membership probability map, which provides, for each pixel of the image, the posterior class membership for all classes of interest. These are the necessary inputs for all methods described below. In practice, the strategy proposed in [18] is as followed: a supervised classification is performed on a selection of features extracted both from 3D lidar point clouds and aerial multispectral images. The training pixels are selected according to an existing forest LC geodatabase. The used classifier is the Random Forest (RF) classifier. This is an efficient classifier, that directly handles multiple classes, and provides posterior probabilities for each class.
Here, both local and global methods are tested. For local techniques, majority voting and probabilistic relaxation are selected. For global methods, various energy formulations based on a feature-sensitive Potts model are proposed.

3.5.1 Local methods

Filtering

An easy way to smooth a probability map is to filter it. All the pixels in a $r \times r$ pixels moving window $\mathcal{W}$ are combined in order to generate an output label of the central pixel. The most popular filter is the majority filter. Firstly, the class probabilities are converted into labels, assuming that the label of pixel $\mathbf{x}$ is the label of the most probable class.

$$C(\mathbf{x}) = [c_i | P(\mathbf{x}, c_i) \geq P(\mathbf{x}, c_j) \forall j], \tag{3.8}$$

with $i, j \in [1, n_c]$, where n_c is the number of classes. From this label image, the final smoothed result is obtained by taking the majority vote in a local neighborhood.

$$C_{smooth}(\mathbf{x}) = \arg\max_i \left[\sum_{\mathbf{u} \in \mathcal{W}} [C(\mathbf{u}) = c_i] \right]. \tag{3.9}$$

Many other filters have been developed.

Probabilistic relaxation

The probabilistic relaxation aims at homogenizing probabilities of a pixel according to its neighboring pixels. The relaxation is an iterative algorithm in which the probability at each pixel is updated at each iteration in order to have it closer to the probabilities of its neighbors [31]. It was adopted for simplicity reasons. First, good accuracies are reported with decent computing time, which is beneficial over large scales. Secondly, it offers an alternative to edge aware/gradient-based techniques that may not be adapted in semantically unstructured environments like forests. The probability $P_k^t(\mathbf{u})$ of class k at a pixel $\mathbf{u}$ at the iteration t is defined by $\delta P_k^t(\mathbf{u})$ which depends on:

- The distance $d_{\mathbf{u},\mathbf{v}}$ between the pixel $\mathbf{u}$ and its neighbors $\mathbf{v}$ (the pixels that are distant of less than r pixels from $\mathbf{u}$).
- A co-occurrence matrix $T_{k,l}$ defining a priori correlation between the probabilities of neighboring pixels. The local co-occurrence matrix has

been tuned arbitrarily, but can also be estimated using training pixels [111]. The matrix is expressed as follow:

$$T_{k,l} = \begin{bmatrix} 0.8 & p & \cdots & p \\ p & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & p \\ p & \cdots & p & 0.8 \end{bmatrix}, \text{with } p = \frac{0.2}{n_c - 1}.$$

The update factor is then defined as:

$$\delta P_k^t(\mathbf{u}) = \sum_{\mathbf{v} \in \mathcal{N}_\mathbf{u}} d_{\mathbf{u},\mathbf{v}} \sum_{l=1}^{n_c} T_{k,l}(\mathbf{u}, \mathbf{v}) \times P_l^t(\mathbf{v}). \tag{3.10}$$

In order to keep the probabilities normalized, the update is performed in two steps using the unnormalized probability $Q_k^{t+1}(\mathbf{u})$ of class k at a pixel $\mathbf{u}$ at the iteration $t+1$:

$$Q_k^{t+1}(\mathbf{u}) = P_k^t(\mathbf{u}) \times \big(1 + \delta P_k^t(\mathbf{u})\big), \tag{3.11}$$

$$P_k^{t+1}(\mathbf{u}) = \frac{Q_k^{t+1}(\mathbf{u})}{\sum_{l=1}^{n_c} Q_l^{t+1}(\mathbf{u})}. \tag{3.12}$$

3.5.2 Global smoothing

The global smoothing method uses only a small number of pairwise cliques between neighboring pixels (4-neighbors or 8-neighbors) to describe the smoothness. Over the entire resulting first order random fields, the maximization of the posterior probability leads to a smoothed results. This can be done by finding the minimum of the negative log-likelihood, $\arg\min_C E(I, C, A)$ with

$$\begin{aligned} E(I, C, A) = &\sum_{\mathbf{u} \in I} E_{\text{data}}(\mathbf{u}, P(\mathbf{u})) + \\ &\gamma \sum_{\mathbf{u} \in I, \mathbf{v} \in \mathcal{N}_\mathbf{u}} E_{\text{pairwise}}(\mathbf{u}, \mathbf{v}, C(\mathbf{u}), C(\mathbf{v}), A(\mathbf{u}), A(\mathbf{v})), \end{aligned} \tag{3.13}$$

where $P(\mathbf{u}) = [P(\mathbf{u}, c_i) | P(\mathbf{u}, c_i) \geq P(\mathbf{u}, c_j) \forall j]$, $A(\mathbf{u})$ are the values of the features at pixel $\mathbf{u}$ (such as height, reflectance...) and $\mathcal{N}_\mathbf{u}$ is the 8-connected neighborhood of the pixel $\mathbf{u}$ (only the 8-connected neighborhood is investigated in this paper). When $\gamma = 0$, the pairwise term has no effect in the energy formulation; the most probable class is attributed to the pixel,

leading to the same result as the classification output. When $\gamma \neq 0$, the resulting label map becomes more homogeneous, and the borders of the segments/stands are smoother. However, if γ is too high, the small areas are bound to be merged into larger areas, removing a part of the useful information provided by the classification step. The automatic tuning of the parameter γ has been addressed in [80] but is not used here.
In this paper, two formulations of E_{data} (unary term) and four formulations of E_{pariwise} (prior) are investigated.

Unary term

A widely used formulation for the unary term is the log-inverse formulation using the natural logarithm. It corresponds to the information content in information theory and is formulated as follow:

$$E_{\text{data}} = -\log(P(\mathbf{u})). \tag{3.14}$$

It highly penalizes the low-probability classes but increase the complexity with potential infinite values.
An other simple formulation for the unary term is the linear formulation,

$$E_{\text{data}} = 1 - P(\mathbf{u}). \tag{3.15}$$

It penalizes less than the log-inverse formulation but has the advantage of having values lying in $[0, 1]$.

Prior

In this work, the prior has a value depending on the class of neighboring pixels. In the four formulations, two neighboring pixels pay no penalty if they are assigned to the same class. Two basic and popular priors, the *Potts model* and the *contrast-sensitive Potts model* (called here *z-Potts model*), are investigated. In the *Potts model*, two neighboring pixels pay the same penalty if they are assigned to different labels, the prior for the *Potts model*

is:

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= 0, \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= 1. \end{aligned} \tag{3.16}$$

In the *z-Potts model*, the penalty for a change of label depends on the gradient of height between two neighboring pixels. The *z-Potts model* is a standard *contrast-sensitive Potts model* applied to the height obtained from the point clouds. Here, since we are dealing with forest stands that are likely to exhibit distinct heights, the gradient of the height map (given with the 3D lidar point cloud) is computed for each of the four directions separately. The maximum M_g over the whole image in the four directions is used to compute the final pairwise energy. A linear function has been used: the penalty is highest when the gradient is 0, and decreases until the gradient reaches its maximum value. The prior of the *z-Potts model* is therefore:

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= 0, \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= 1 - \frac{g_{\mathbf{u}\to\mathbf{v}}}{M_g}, \end{aligned} \tag{3.17}$$

where $g_{\mathbf{u}\to\mathbf{v}}$ is the gradient between pixel $\mathbf{u}$ and pixel $\mathbf{v}$, i.e., the absolute value of the height difference of the two pixels.
An other pairwise energy investigated is a global feature sensitive energy (called here *Exponential-features model*). The pairwise energy is computed with respect to a pool of n features. When the features have close values, the penalty is high and decreases when the features tends to be very different. The pairwise energy in this case is expressed as follows:

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= 0, \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= \frac{1}{n}\sum_{i=1}^{n} \exp(-|A_i(\mathbf{u}) - A_i(\mathbf{v})|), \end{aligned} \tag{3.18}$$

where $A_i(\mathbf{u})$ is the value of the i^{th} feature of the pixel $\mathbf{u}$. To compute such energy, the features need to be first normalized (i.e., zero mean, unit standard deviation) in order ensure that they all have the same dynamic.
The last formulation investigated is also a global feature sensitive energy (called here *Distance-features model*). The pairwise energy is still computed

with respect to a pool of n features. In this case, the energy is computed according to the distance between the two neighboring pixels in the feature space, the penalty is high when the pixels are close in the feature space and decrease when they get distant. The pairwise energy in this case is expressed as follow:

$$\begin{aligned} &E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) = 0, \\ &E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) = 1 - ||A(\mathbf{u}); A(\mathbf{v})||_{n,2}, \end{aligned} \tag{3.19}$$

with

$$||A(\mathbf{u}); A(\mathbf{v})||_{n,2} = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^{n} \big(A_i(\mathbf{u}) - A_i(\mathbf{v})\big)^2}. \tag{3.20}$$

To compute such energy, the features need to be first normalized (i.e., zero mean, unit standard deviation) in order ensure that they all have the same dynamic. They are then rescaled between 0 and 1 to ensure that $||A(\mathbf{u}); A(\mathbf{v})||_{n,2}$ lies in $[0; 1]$ $\forall(\mathbf{u}, \mathbf{v})$.

In [18], a high number of features was extracted from available lidar and optical images ($\sim$ 100) but can be selected. They can also be weighted according to their importance, computed through the Random Forest classification process. Since the most important features (20) are almost all equally weighted, it does not bring additional discriminative information for the global feature sensitive energy.

Energy minimization

The energy minimization is performed using graph-cut methods. The graph-cut algorithm employed is the quadratic pseudo-boolean optimization (QPBO). The QPBO is a popular and efficient graph-cut method as it efficiently solves energy minimization problems (such as the proposed ones) by constructing a graph and computing the min-cut [56]. α-expansion moves are used, as they are an efficient way to deal with the multi-class problems [57].

4

Results

5

Conclusion and perspectives

A

Publication

A.1 Journal articles 46
A.2 Peer-reviewed conference papers 46

A.1 Journal articles

C. Dechesne, C. Mallet, A. Le Bris, V. Gouet-Brunet. *Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery*. ISPRS Journal of Photogrammetry and Remote Sensing, 126 (2017), pp.129–145, 2017.

M. Fauvel, C. Dechesne, A. Zullo, F. Ferraty. *Fast forward feature selection of hyperspectral images for classification with gaussian mixture models*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8(6), pp. 2824-2831, 2015.

A.2 Peer-reviewed conference papers

C. Dechesne, C. Mallet, A. Le Bris, V. Gouet-Brunet. *How to combine LIDAR and very high resolution multispectral images for forest stand segmentation?* Proc. of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, USA, July 2017.

C. Dechesne, C. Mallet, A. Le Bris, V. Gouet-Brunet. *Semantic segmentation of forest stands of pure specie as a global optimisation problem*. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017.

C. Dechesne, C. Mallet, A. Le Bris, V. Gouet-Brunet. *Segmentation sémantique de données de télédétection multimodale : application aux peuplements forestiers*. ORASIS, Colleville-sur-Mer, France, Juin 2017.

C. Dechesne, C. Mallet, A. Le Bris, V. Gouet-Brunet, A. Hervieu. *Forest stand segmentation using airborne Lidar data and very high resolution multispectral imagery*. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 41 (B3), pp 207-214 , ISPRS Congress, Prague, Juillet 2016.

Bibliography

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. "SLIC superpixels compared to state-of-the-art superpixel methods". In: *IEEE transactions on pattern analysis and machine intelligence* 34.11 (2012), pp. 2274–2282.

[2] L. Albert, F. Rottensteiner, and C. Heipke. "Contextual land use classification: How detailed can the class structure be?" In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B4 (2016), pp. 11–18.

[3] PM Atkinson, MEJ Cutler, and H Lewis. "Mapping sub-pixel proportional land cover with AVHRR imagery". In: *International Journal of Remote Sensing* 18.4 (1997), pp. 917–935.

[4] Cédric Bacour, François-Marie Bréon, and Fabienne Maignan. "Normalization of the directional effects in NOAA–AVHRR reflectance measurements for an improved monitoring of vegetation cycles". In: *Remote Sensing of Environment* 102.3 (2006), pp. 402–413.

[5] Mariana Belgiu and Lucian Drăguţ. "Random Forest in remote sensing: A review of applications and future directions". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 114 (2016), pp. 24–31.

[6] Richard E Bellman. *Adaptive control processes: a guided tour*. Princeton university press, 2015.

[7] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. "A training algorithm for optimal margin classifiers". In: *Proceedings of the fifth annual workshop on Computational learning theory*. ACM. 1992, pp. 144–152.

[8] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[9] Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[10] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

[11] Gustau Camps-Valls and Lorenzo Bruzzone. *Kernel methods for remote sensing data analysis*. John Wiley & Sons, 2009.

[12] Anil Cheriyadat and L Mann Bruce. "Why principal component analysis is not an appropriate feature extraction method for hyperspectral data". In: *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International*. Vol. 6. IEEE. 2003, pp. 3420–3422.

[13] Dorin Comaniciu and Peter Meer. "Mean shift: A robust approach toward feature space analysis". In: *IEEE Transactions on pattern analysis and machine intelligence* 24.5 (2002), pp. 603–619.

[14] Michele Dalponte, Francesco Reyes, Kaja Kandare, and Damiano Gianelle. "Delineation of Individual Tree Crowns from ALS and Hyperspectral data: a comparison among four methods". In: *European Journal of Remote Sensing* 48 (2015), pp. 365–382.

[15] Michele Dalponte, Hans Ole Ørka, Liviu Theodor Ene, Terje Gobakken, and Erik Næsset. "Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data". In: *Remote Sensing of Environment* 140 (2014), pp. 306–317.

[16] Michele Dalponte, Lorenzo Bruzzone, and Damiano Gianelle. "Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data". In: *Remote Sensing of Environment* 123 (2012), pp. 258–270.

[17] C Dechesne, C Mallet, A Le Bris, V Gouet-Brunet, and A Hervieu. "Forest stand segmentation using airborne lidar data and very high resolution multispectral imagery". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 41-B3* (2016), pp. 207–214.

[18] Clément Dechesne, Clément Mallet, Arnaud Le Bris, and Valérie Gouet-Brunet. "Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 126 (2017), pp. 129–145.

[19] Haluk Derin and Howard Elliott. "Modeling and segmentation of noisy and textured images using Gibbs random fields". In: *IEEE Transactions on pattern analysis and machine intelligence* 1 (1987), pp. 39–55.

[20] Oliver Diedershagen, Barbara Koch, and Holger 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* 36(8/W2) (2004), pp. 208–212.

[21] Lothar Eysn, Markus Hollaus, Klemens Schadauer, and Norbert Pfeifer. "Forest delineation based on airborne LIDAR data". In: *Remote Sensing* 4.3 (2012), pp. 762–783.

[22] Timothy J Fahey, Peter B Woodbury, John J Battles, Christine L Goodale, Steven P Hamburg, Scott V Ollinger, and Christopher W Woodall. "Forest carbon storage: ecology, management, and policy". In: *Frontiers in Ecology and the Environment* 8.5 (2010), pp. 245–252.

[23] Mathieu Fauvel, Clément Dechesne, Anthony Zullo, and Frederic Ferraty. "Fast forward feature selection of hyperspectral images for classification with Gaussian mixture models". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.6 (2015), pp. 2824–2831.

[24] Pedro F Felzenszwalb and Daniel P Huttenlocher. "Efficient graph-based image segmentation". In: *International journal of computer vision* 59.2 (2004), pp. 167–181.

[25] António Ferraz, Clément Mallet, and Nesrine Chehata. "Large-scale road detection in forested mountainous areas using airborne topographic lidar data". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 112 (2016), pp. 23 –36.

[26] Ronald A Fisher. "The use of multiple measurements in taxonomic problems". In: *Annals of human genetics* 7.2 (1936), pp. 179–188.

[27] Giles M Foody. "Status of land cover classification accuracy assessment". In: *Remote sensing of environment* 80.1 (2002), pp. 185–201.

[28] Ashish Ghosh, Nikhil R Pal, and Sankar K Pal. "Image segmentation using a neural network". In: *Biological Cybernetics* 66.2 (1991), pp. 151–158.

[29] Pall Oskar Gislason, Jon Atli Benediktsson, and Johannes R Sveinsson. "Random forests for land cover classification". In: *Pattern Recognition Letters* 27.4 (2006), pp. 294–300.

[30] David E Goldberg. "Genetic algorithms in search, optimization, and machine learning, 1989". In: *Reading: Addison-Wesley* (1989).

[31] P. Gong and P.J. Howarth. "Performance analyses of probabilistic relaxation methods for land-cover classification". In: *Remote Sensing of Environment* 30.1 (1989), pp. 33–42.

[32] A. Gressin, C. Mallet, N. Vincent, and N. Paparoditis. "Updating land cover databases using a single very high resolution satellite image". In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W2 (2013), pp. 13–18.

[33] Adrien Gressin, Nicole Vincent, Clément Mallet, and Nicolas Paparoditis. "A unified framework for land-cover database update and enrichment using satellite imagery". In: *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE. 2014, pp. 5057–5061.

[34] Adrien Gressin, Clément Mallet, Nicole Vincent, and Nicolas Paparoditis. "Updating land cover databases using a single very high resolution satellite image". In: *The ISPRS Workshop on Image Sequence Analysis*. Vol. 2. 3. 2013, pp. 13–18.

[35] Adrien Gressin, Clément Mallet, Nicole Vincent, and Nicolas Paparoditis. "Updating the new French national land cover database". In: *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*. IEEE. 2014, pp. 3534–3537.

[36] L. Guigues, J.-P. Cocquerez, and H. Le Men. "Scale-sets image analysis". In: *IJCV* 68.3 (2006), pp. 289–317.

[37] Ayman Habib, Mwafag Ghanma, Michel Morgan, and Rami Al-Ruzouq. "Photogrammetric and LiDAR data registration using linear features". In: *Photogrammetric Engineering & Remote Sensing* 71.6 (2005), pp. 699–707.

[38] FR Hansen and Howard Elliott. "Image segmentation using simple Markov field models". In: *Computer Graphics and Image Processing* 20.2 (1982), pp. 101–132.

[39] Georgef Hepner, Thomas Logan, Niles Ritter, and Nevin Bryant. "Artificial neural network classification using a minimal training set-Comparison to conventional supervised classification". In: *Photogrammetric Engineering and Remote Sensing* 56.4 (1990), pp. 469–473.

[40] A Hernando, Dirk Tiede, Florian Albrecht, and Stefan Lang. "Spatial and thematic assessment of object-based forest stand delineation using an OFA-matrix". In: *International Journal of Applied Earth Observation and Geoinformation* 19 (2012), pp. 214–225.

[41] Gordon Hughes. "On the mean accuracy of statistical pattern recognizers". In: *IEEE transactions on information theory* 14.1 (1968), pp. 55–63.

[42] Piotr Indyk and Rajeev Motwani. "Approximate nearest neighbors: towards removing the curse of dimensionality". In: *Proceedings of the thirtieth annual ACM symposium on Theory of computing*. ACM. 1998, pp. 604–613.

[43] George H John. "Enhancements to the data mining process". PhD thesis. stanford university Ph. D. thesis, 1997.

[44] Ian Jolliffe. "Principal component analysis". In: *International encyclopedia of statistical science*. Springer, 2011, pp. 1094–1096.

[45] Carl F Jordan. "Derivation of leaf-area index from quality of light on the forest floor". In: *Ecology* 50.4 (1969), pp. 663–666.

[46] Christian Jutten and Jeanny Herault. "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture". In: *Signal processing* 24.1 (1991), pp. 1–10.

[47] Kaja Kandare, Michele Dalponte, Damiano Gianelle, and Jonathan Cheung-Wai Chan. "A new procedure for identifying single trees in understory layer using discrete LiDAR data". In: *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*. IEEE. 2014, pp. 1357–1360.

[48] Annika Kangas and Matti Maltamo. *Forest inventory: methodology and applications*. Vol. 10. Springer Science & Business Media, 2006.

[49] Anahita Khosravipour, Andrew K Skidmore, Martin Isenburg, Tiejun Wang, and Yousif A Hussin. "Generating pit-free canopy height models from airborne lidar". In: *Photogrammetric Engineering & Remote Sensing* 80.9 (2014), pp. 863–872.

[50] M. Kim, M. Madden, and T.A. Warner. "Forest Type Mapping using Object-specific Texture Measures from Multispectral Ikonos Imagery: Segmentation Quality and Image Classification Issues". In: *Photogrammetric Engineering & Remote Sensing* 75.7 (2009), pp. 819–829.

[51] Alexander Kirillov, Dmitrij Schlesinger, Walter Forkel, Anatoly Zelenin, Shuai Zheng, Philip Torr, and Carsten Rother. "A Generic CNN-CRF Model for Semantic Segmentation". In: *arxiv:1511.05067* (2015).

[52] Josef Kittler. "Feature set search algorithms". In: *Pattern recognition and signal processing* (1978).

[53] B Koch, Ch Straub, M Dees, Y Wang, and H Weinacker. "Airborne laser data for stand delineation and information extraction". In: *International Journal of Remote Sensing* 30.4 (2009), pp. 935–963.

[54] Ron Kohavi and George H John. "Wrappers for feature subset selection". In: *Artificial intelligence* 97.1-2 (1997), pp. 273–324.

[55] Pushmeet Kohli, L'. Ladický, and P. Torr. "Robust Higher Order Potentials for Enforcing Label Consistency". English. In: *International Journal of Computer Vision* 82.3 (2009), pp. 302–324. ISSN: 0920-5691.

[56] Vladimir Kolmogorov and Carsten Rother. "Minimizing non-submodular functions with graph cuts-a review". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.7 (2007), pp. 1274–1279.

[57] Vladimir Kolmogorov and Ramin Zabih. "What energy functions can be minimized via graph cuts?" In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.2 (2004), pp. 147–159.

[58] S. Kumar and M. Hebert. "Discriminative random fields." In: *International Journal of Computer Vision* 68.2 (2006), pp. 179–201.

[59] L'ubor Ladický, Chris Russell, Pushmeet Kohli, and Philip Torr. "Inference Methods for CRFs with Co-occurrence Statistics". In: *International Journal of Computer Vision* 103.2 (2012), pp. 213–225.

[60] David A Landgrebe. *Signal theory methods in multispectral remote sensing*. Vol. 29. John Wiley & Sons, 2005.

[61] Rasmus Larsen. "Decomposition using maximum autocorrelation factors". In: *Journal of Chemometrics* 16.8-10 (2002), pp. 427–435.

[62] L Lebart, A Morineau, and Mr Piron. "Multidimensional exploratory statistics". In: *París: Ed. Dunod* (1997).

[63] Donald G Leckie, François A Gougeon, Nicholas Walsworth, and Dennis Paradine. "Stand delineation and composition estimation using semi-automated individual tree crown analysis". In: *Remote Sensing of Environment* 85.3 (2003), pp. 355–369.

[64] J. Lee, X. Cai, J. Lellmann, M. Dalponte, Y. Malhi, N. Butt, M. Morecroft, C. B. Schönlieb, and D. A. Coomes. "Individual Tree Species Classification From Airborne Multisensor Imagery Using Robust PCA". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.6 (2016), pp. 2554–2567. ISSN: 1939-1404.

[65] TC Lemprière, PY Bernier, AL Carroll, MD Flannigan, RP Gilsenan, DW McKenney, EH Hogg, JH Pedlar, D Blain, et al. "The importance

of forest sector adaptation to climate change". In: *Nat. Resour. Can., Can. For. Serv., North. For. Cent., Edmonton, AB. Inf. Rep. NOR-X-416E* (2008).

[66] VJ Leppänen, T Tokola, M Maltamo, L Mehtätalo, T Pusa, and J Mustonen. "Automatic delineation of forest stands from LIDAR data". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38(4/C1)* (2008), pp. 5–8.

[67] Alex Levinshtein, Adrian Stere, Kiriakos N Kutulakos, David J Fleet, Sven J Dickinson, and Kaleem Siddiqi. "Turbopixels: Fast superpixels using geometric flows". In: *IEEE transactions on pattern analysis and machine intelligence* 31.12 (2009), pp. 2290–2297.

[68] Jili Li, Baoxin Hu, and Thomas L. Noland. "Classification of tree species based on structural features derived from high density LiDAR data". In: *Agricultural and Forest Meteorology* 171–172 (2013), pp. 104 –114.

[69] Huan Liu and Rudy Setiono. "Feature selection and classification-a probabilistic wrapper approach". In: *Proceedings of 9th International Conference on Industrial and Engineering Applications of AI and ES*. 1997, pp. 419–424.

[70] Dengsheng Lu and Qihao Weng. "A survey of image classification methods and techniques for improving classification performance". In: *International journal of Remote sensing* 28.5 (2007), pp. 823–870.

[71] A. Lucchi, Y. Li, X. Boix, K.Smith, and P. Fua. "Are Spatial and Global Constraints Really Necessary for Segmentation?" In: *Proc. of ICCV*. 2011, pp. 9–16.

[72] A. Maas, F. Rottensteiner, and C. Heipke. "Using label noise robust logistic regression for automated updating of topographic geospatial databases". In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* III-7 (2016), pp. 133–140.

[73] Clément Mallet, Frédéric Bretar, Michel Roux, Uwe Soergel, and Christian Heipke. "Relevance assessment of full-waveform lidar data for

urban area classification". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 66.6 (2011), S71–S84.

[74] Thomas Marill and D Green. "On the effectiveness of receptors in recognition systems". In: *IEEE transactions on Information Theory* 9.1 (1963), pp. 11–17.

[75] Andrew Mastin, Jeremy Kepner, and John Fisher. "Automatic registration of LIDAR and optical images of urban scenes". In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE. 2009, pp. 2639–2646.

[76] Paul Mather and Brandt Tso. *Classification methods for remotely sensed data*. CRC press, 2016.

[77] Joseph E Means, Steven A Acker, Brandon J Fitt, Michael Renslow, Lisa Emerson, and Chad J Hendrix. "Predicting forest stand characteristics with airborne scanning lidar". In: *Photogrammetric Engineering & Remote Sensing* 66.11 (2000), pp. 1367–1372.

[78] Alastair P Moore, Simon JD Prince, Jonathan Warrell, Umar Mohammed, and Graham Jones. "Superpixel lattices". In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE. 2008, pp. 1–8.

[79] Brice Mora, Michael A. Wulder, and Joanne C. White. "Segment-constrained regression tree estimation of forest stand height from very high spatial resolution panchromatic imagery over a boreal environment". In: *Remote Sensing of Environment* 114.11 (2010), pp. 2474 –2484.

[80] G. Moser, S.B. Serpico, and J.A. Benediktsson. "Land-Cover Mapping by Markov Modeling of Spatial Contextual Information in Very-High-Resolution Remote Sensing Images". In: *Proceedings of the IEEE* 101.3 (2013), pp. 631–651.

[81] Anh Nguyen and Bac Le. "3D point cloud segmentation: A survey". In: *Robotics, Automation and Mechatronics (RAM), 2013 6th IEEE Conference on*. IEEE. 2013, pp. 225–230.

[82] M Pal. "Random forest classifier for remote sensing classification". In: *International Journal of Remote Sensing* 26.1 (2005), pp. 217–222.

[83] Nikhil R Pal and Sankar K Pal. "A review on image segmentation techniques". In: *Pattern recognition* 26.9 (1993), pp. 1277–1294.

[84] JD Paola and RA Schowengerdt. "A review and analysis of back-propagation neural networks for classification of remotely-sensed multi-spectral imagery". In: *International Journal of remote sensing* 16.16 (1995), pp. 3033–3058.

[85] Tamar Peli and David Malah. "A study of edge detection algorithms". In: *Computer graphics and image processing* 20.1 (1982), pp. 1–21.

[86] Arne Pommerening. "Approaches to quantifying forest structures". In: *Forestry: An International Journal of Forest Research* 75.3 (2002), pp. 305–324.

[87] Pavel Pudil, Jana Novovičová, and Josef Kittler. "Floating search methods in feature selection". In: *Pattern Recognition Letters* 15.11 (1994), pp. 1119–1125.

[88] Julien Radoux, Céline Lamarche, Eric Van Bogaert, Sophie Bontemps, Carsten Brockmann, and Pierre Defourny. "Automated Training Sample Extraction for Global Land Cover Mapping". In: *Remote Sensing* 6.5 (2014), p. 3965. ISSN: 2072-4292. DOI: 10.3390/rs6053965.

[89] K. Schindler. "An overview and comparison of smooth labeling methods for land-cover classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 50.11 (2012), pp. 4534–4545.

[90] Konrad Schindler. "An overview and comparison of smooth labeling methods for land-cover classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 50.11 (2012), pp. 4534–4545.

[91] Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[92] Jianbo Shi and Jitendra Malik. "Normalized cuts and image segmentation". In: *IEEE Transactions on pattern analysis and machine intelligence* 22.8 (2000), pp. 888–905.

[93] John Miles Smith and Diane CP Smith. "Database abstractions: aggregation and generalization". In: *ACM Transactions on Database Systems (TODS)* 2.2 (1977), pp. 105–133.

[94] William H Smith. *Air pollution and forests: interactions between air contaminants and forest ecosystems*. Springer Science & Business Media, 2012.

[95] Jean-Philippe Souchon, Christian Thom, Christophe Meynard, and Olivier Martin. "A large format camera system for national mapping purposes". In: *Revue Française de Photogrammétrie et de Télédétection* 200 (2012), pp. 48–53.

[96] Alan H Strahler. "The use of prior probabilities in maximum likelihood classification of remotely sensed data". In: *Remote sensing of Environment* 10.2 (1980), pp. 135–163.

[97] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. "Bias in random forest variable importance measures: Illustrations, sources and a solution". In: *BMC bioinformatics* 8.1 (2007), p. 25.

[98] Alicia A Sullivan, Robert J McGaughey, Hans-Erik Andersen, and Peter Schiess. "Object-oriented classification of forest structure from light detection and ranging data for stand mapping". In: *Western Journal of Applied Forestry* 24.4 (2009), pp. 198–204.

[99] Torfinn Taxt, Patrick J Flynn, and Anil K Jain. "Segmentation of document images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.12 (1989), pp. 1322–1329.

[100] Dirk Tiede, Thomas Blaschke, and Marco Heurich. "Object-based semi automatic mapping of forest stands with Laser scanner and Multi-spectral data". In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36(8/W2) (2004), pp. 328–333.

[101] H Torabzadeh, R Leiterer, ME Schaepman, and F Morsdorf. "Optimal structural and spectral features for tree species classification using combined airborne laser scanning and hyperspectral data". In: *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*. IEEE. Milano, Italy, 2015, pp. 5399–5402.

[102] Hossein Torabzadeh, Felix Morsdorf, and Michael E Schaepman. "Fusion of imaging spectroscopy and airborne laser scanning data for characterization of forest ecosystems–A review". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 97 (2014), pp. 25–35.

[103] Roger Trias-Sanz and Didier Boldo. "A high-reliability, high-resolution method for land cover classification into forest and non-forest". In: *Image Analysis* (2005), pp. 417–441.

[104] Compton J Tucker. "Red and photographic infrared linear combinations for monitoring vegetation". In: *Remote sensing of Environment* 8.2 (1979), pp. 127–150.

[105] Devis Tuia, Fabio Pacifici, Mikhail Kanevski, and William J Emery. "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines". In: *IEEE Transactions on Geoscience and Remote Sensing* 47.11 (2009), pp. 3866–3879.

[106] Jari Vauhkonen, Ilkka Korpela, Matti Maltamo, and Timo Tokola. "Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics". In: *Remote Sensing of Environment* 114.6 (2010), pp. 1263–1276.

[107] Andrea Vedaldi and Stefano Soatto. "Quick shift and kernel methods for mode seeking". In: *Computer vision–ECCV 2008* (2008), pp. 705–718.

[108] Cédric Véga, A Hamrouni, S El Mokhtari, J Morel, J Bock, J-P Renaud, M Bouvier, and Sylvie Durrieu. "PTrees: A point-based approach to forest tree extraction from lidar data". In: *International Journal of Applied Earth Observation and Geoinformation* 33 (2014), pp. 98–108.

[109] Olga Veksler, Yuri Boykov, and Paria Mehrani. "Superpixels and supervoxels in an energy optimization framework". In: *Computer Vision–ECCV 2010* (2010), pp. 211–224.

[110] Luc Vincent and Pierre Soille. "Watersheds in digital spaces: an efficient algorithm based on immersion simulations". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 6 (1991), pp. 583–598.

[111] M. Volpi and V. Ferrari. "Semantic segmentation of urban scenes by learning local class interactions". In: *Proc. of CVPR Workshops*. 2015, pp. 1–9.

[112] Martin Weinmann. "Reconstruction and Analysis of 3D Scenes". In: *Springer International Publishing, Cham. doi* 10 (2016), pp. 978–3.

[113] Martin Weinmann, Boris Jutzi, Stefan Hinz, and Clément Mallet. "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 105 (2015), pp. 286 –304.

[114] Joanne C White, Nicholas C Coops, Michael A Wulder, Mikko Vastaranta, Thomas Hilker, and Piotr Tompalski. "Remote sensing technologies for enhancing forest inventories: A review". In: *Canadian Journal of Remote Sensing* 42.5 (2016), pp. 619–641.

[115] A Wayne Whitney. "A direct method of nonparametric measurement selection". In: *IEEE Transactions on Computers* 100.9 (1971), pp. 1100–1103.

[116] Zhengzhe Wu, Ville Heikkinen, Markku Hauta-Kasari, Jussi Parkkinen, and Timo Tokola. "ALS data based forest stand delineation with a coarse-to-fine segmentation approach". In: *IEEE Congress on Image and Signal Processing (CISP)*. Dalian, China, 2014, pp. 547–552.

[117] Michael A Wulder, Joanne C White, Geoffrey J Hay, and Guillermo Castilla. "Towards automated segmentation of forest inventory polygons on high spatial resolution satellite imagery". In: *The Forestry Chronicle* 84.2 (2008), pp. 221–230.

[118] Shimon D Yanowitz and Alfred M Bruckstein. "A new method for image segmentation". In: *Computer Vision, Graphics, and Image Processing* 46.1 (1989), pp. 82–95.

[119] Amin Zargar, Rehan Sadiq, Bahman Naser, and Faisal I Khan. "A review of drought indices". In: *Environmental Reviews* 19 (2011), pp. 333–349.