Relevance assessment of full-waveform lidar data for urban area classification

Clément Mallet \textsuperscript{a,}* Frédéric Bretar \textsuperscript{b}, Michel Roux \textsuperscript{c}, Uwe Soergel \textsuperscript{d}, Christian Heipke \textsuperscript{d}

\textsuperscript{a}Université Paris Est, Institut Géographique National, Laboratoire MATIS, 73 Av. de Paris, 94165 Saint-Mandé Cedex, France
\textsuperscript{b}Public Works Regional Engineering Office – CETE Normandie Centre, 10, Chemin de la Poudrière, BP 245, 76121 Grand Quevilly Cedex, France
\textsuperscript{c}GET/Télécom Paris – CNRS UMR 5141 – Département TSI, 46 Rue Barrault, 75013 Paris, France
\textsuperscript{d}Institute of Photogrammetry and Geoinformation Leibniz Universität Hannover, Nienburger Str. 1, D-30167 Hannover, Germany

\begin{abstract}
Full-waveform lidar data are increasingly being available. Morphological features can be retrieved from the echoes composing the waveforms, and are now extensively used for a variety of land-cover mapping issues. However, the genuine contribution of these features with respect to those computed from standard discrete return lidar systems has been barely theoretically investigated. This paper therefore aims to study the potential of full-waveform data through the automatic classification of urban areas in building, ground, and vegetation points. Two waveform processing methods, namely a non-linear least squares method and a marked point process approach, are used to fit the echoes both with symmetric and asymmetric modeling functions. The performance of the extracted full-waveform features for the classification problem are then compared to a large variety of multiple-pulse features thanks to three feature selection methods. A support vector machines classifier is finally used to label the point cloud according to various scenarios based on the rank of the features. This allows to find the best classification strategy as well as the minimal feature subsets allowing to achieve the highest classification accuracy possible for each of the three feature selection methods.

The results show that the echo amplitude as well as two features computed from the radiometric calibration of full-waveform data, namely the cross-section and the backscatter coefficient, significantly contribute to the high classification accuracies reported in this paper (around 95%). Conversely, features extracted from the non Gaussian modelling of the echoes are not relevant for the discrimination of vegetation, ground, and buildings in urban areas.
\end{abstract}

\section{1. Introduction}

\subsection{1.1. Context and objectives}

The new technology of full-waveform (FW) lidar systems has appeared in the last twenty years, and has become popular in the last five years, with the emergence of small-footprint airborne topographic commercial devices (Hug et al., 2004; Wagner et al., 2004). It permits to record with high-frequency sampling (typically 1 GHz) the received signal for each transmitted laser pulse, called “waveform” referring to the shape of the return signal (Mallet and Bretar, 2009). Traditional discrete return systems (also called multiple-pulse systems) provide 3D point clouds with echo number and amplitude information.\textsuperscript{1} Conversely, full-waveform data consist of the temporal succession of 1D profiles of the landscape. Due to diffraction, the laser beam penetrates a cone-shaped volume, and each object inside this volume hit by the laser pulse may show-up as distinct echo in the waveform. With signal processing methods, FW data therefore provide additional information on the reflecting characteristics of the objects, which may be relevant for land-cover mapping, and especially for urban area classification. Many classification workflows based on full-waveform data have been established (Ducic et al., 2006; Mallet et al., 2008; Alexander et al., 2010), showing the interest of waveform recording, but the theoretical assessment of the potential of such data has been rarely tackled (Chehata et al., 2009).

This paper therefore aims to study the specific contribution of FW lidar data with regards to traditional 3D lidar point clouds in the specific context of the automatic classification of urban areas. Indeed, urban areas are of simpler geometry than natural and forested areas in terms of light interaction. Without modeling the physics of the signal (Wagner et al., 2006; Jutzi and Stilla, 2006), lidar waveform analysis and contribution can be addressed indirectly through the study of the morphological features extracted from the waveforms. These features are called full-waveform features. A standard approach consists in considering a waveform as a Gaussian mixture model: each echo is modeled by a specific

\begin{table}
\centering
\begin{tabular}{|c|c|c|}
\hline
Feature & Description & Example
\hline
\hline
Intensity & Echo energy & $E$
\hline
Amplitude & Value of the echoes at the maximum location & $a$
\hline
Cross-section & The area under the waveform & $A$
\hline
Backscatter coefficient & Reflecting characteristics of the objects & $\beta$
\hline
\end{tabular}
\caption{Examples of features extracted from waveform data.}
\end{table}
Gaussian curve (Hofton et al., 2000; Persson et al., 2005; Jutzi and Stilla, 2006; Wagner et al., 2006). This allows to compute the amplitude and the width of each echo.

Our study is focused on three main classes: building, ground, and vegetation. They are main classes of interest for large scale mapping, and have been extensively labeled in the literature. Furthermore, to propose an objective evaluation of the contribution of FW data, ground surface and roof material classes have not been considered because of the impossibility to classify them with the simple knowledge of the 3D point locations. They may be discriminated using FW data (Jutzi and Stilla, 2003; Alexander et al., 2010), but it would be more difficult using only discrete return data.

1.2. Potential of full-waveform features

As mentioned above, the full-waveform features may discriminate many objects of interest for various applications. Their behavior has been extensively studied in the last five years (Mallet and Bretar, 2009).

The amplitude of the echoes varies with the radiometric and geometric properties of the targets. High amplitudes are noticed on grass and bare earth, low on streets, and variable amplitudes on the roofs of buildings, depending on the material. Such feature can therefore be useful for surface discrimination (Alexander et al., 2010). Besides, the amplitude value is, for instance, used jointly with geometrical criteria to segment tree crowns (Reitberger et al., 2009).

Lidar echoes are, in general, wider on the canopy or ploughed fields compared to roads or meadow areas. This is therefore a relevant information for road detection or vegetation discrimination (Holzhauser et al., 2011). For digital terrain model generation, the echo width has proved to be useful for filtering ground points from low above-ground points lying for instance on bushes (Doneus et al., 2008; Lin and Mills, 2009; Mücke et al., 2010).

Nevertheless, the classification of objects is not always straightforward using such features. A wide echo with low amplitude does not necessarily come from vegetation. Roads and building roofs are made of different types of material and, therefore, have different characteristics. The features currently extracted for man-made structures often have values similar to natural objects (Gross et al., 2007). In order to cope with area-specific issues, attributes derived from the echo amplitude and width have been proposed. They are presented in Section 3.2. For instance, the backscatter cross-section, which describes the scattering properties of the targets, seems valuable for retrieving these physical characteristics and improve segmentation/classification results (Wagner et al., 2008).

1.3. Related works on FW data classification

Several classification methods have already been applied to FW lidar data for urban scenes. Supervised classifiers are preferred since they offer a higher flexibility (Ducic et al., 2006; Alexander et al., 2010). For such methods, in addition to the feature vector, a set of data samples have class labels associated with them. This set is called the training dataset, and is used to estimate the parameters of the classifier. An important underlying assumption of supervised classifiers is that the whole dataset is similar in terms of distribution of features to the training dataset. This means that the classifier must have observed similar features in the training in order to perform a good classification (the so-called generalization performance). To prevent overfitting, advanced techniques can be used such as the bagging technique employed in the random forest classifier (Guo et al., 2011). In particular, since classifiers based on statistical learning theory have shown remarkable abilities to deal with both high-dimensional data and a limited training set, support vector machines (SVM) have been naturally investigated for the classification of airborne lidar data in urban areas (Secord and Zakhor, 2007; Samadzadegan et al., 2010), but barely for FW data (Mallet et al., 2008). Rule-based methods have been favored since many features (such as the echo number or the height above ground) have a physical meaning, and may be discriminative using simple thresholds. Decision trees (DT) have therefore been adopted in Ducic et al. (2006), Rutzinger et al. (2008), Alexander et al. (2010), and Höfle and Holzhauser (2010). This is an efficient approach to fuse spatial features and full-waveform features. On the one hand, DT may allow to discriminate a large number of classes: for instance, trees, shrubs, grass, roofs, and road are labeled in Ducic et al. (2006), ground surfaces are discriminated in Alexander et al. (2010), whereas in Höfle and Holzhauser (2010), in an oriented-object approach, six classes of vegetation are distinguished. On the other hand, such heuristic method may suffer from area-specific and sensor-specific thresholds (Alexander et al., 2010). Besides, DT can be easily computed with few number of features, but high classification accuracy may not be achieved. Conversely, a large feature vector may lead to very complex trees where the optimal splitting order may not be easy to find. Feature selection has naturally been investigated to cope with this issue.

For feature selection purposes, the lidar features are, for instance, fed into a classification tree algorithm with distinct complexity parameters in Rutzinger et al. (2008). The best discriminative features and the associated thresholds are automatically selected, and accuracies slightly superior (>90%) to those reported in Ducic et al. (2006) are reached. Neuenschwander et al. (2009) evaluate the discrimination performance of the Gaussian parameters and other waveform metrics for land cover classification. With a supervised Bayesian pairwise classifier, eight features are used to discriminate seven classes (grass, trail, dead trees, oaks, junipers, etc.). The method allows to select class dependent features based on their incremental contribution to a given relevance function. All the pairwise classifiers are computed and the ultimate class is selected within the outputs based on a maximum Bayesian posterior probability rule. The most frequently selected features are the peak amplitude and the energy ratio, and the rise time and fall time of the waveforms are also discriminant for different tree types and densities.

1.4. Contributions

A feature selection step embedded in the classification procedure has already been proved to be a suitable approach for full-waveform lidar data labelling (Neuenschwander et al., 2009; Chehata et al., 2009; Guo et al., 2011). Thanks to this, the paper presents three interesting characteristics:

- Evaluation of a large set of full-waveform features, computed from recent waveform processing and calibration methods, with respect to a large variety of multiple-pulse features, extracted both from spatial and echo-based considerations.
- Maximization of the potential of waveform features for urban classification.
- Flexible approach: one may change the input features without modifying the selection and classification steps.

The paper is organized as follows. Section 2 presents the processing procedure and describes the adopted algorithms. Then, a detailed description of the feature vector, used for the labeling problem, is provided in Section 3. The dataset is described in Section 4. Results are presented and discussed in Section 5, and the conclusions are drawn in Section 6.
2. Methods

The following section describes the processing chain. An overview is depicted in Fig. 1. Lidar waveforms are first processed to derive a 3D point cloud with two distinct sets of morphological features. Then, the 3D point cloud is classified using the supervised support vector machines classifier. Each point is associated to a set of features comprising full-waveform features as well as features derived from discrete return data. The relevance of these features is evaluated through three feature selection algorithms. According to their classification performance, different subsets of features are selected and the classification results compared.

SVMs have been selected because a simple feature vector can be used as input for the discrimination task. As mentioned above, the feature vector partly stems from the modeling step. Therefore, even if the modeling functions are modified, the proposed workflow will not have to be adapted. This would not be the case, for instance, for decision trees, the choice of support vector machines is rather objective. Other methods, such as classification and regression trees (CART) or ensemble classifiers (e.g., random forests), would be as suitable as SVMs, and are known to provide the same level of performance.

2.1. Waveform processing

An end-user has access to the raw lidar data but now a processing step is mandatory to turn 1D signals into 3D point clouds. The objective of waveform processing is to extract more robust and additional information from raw lidar data than provided by multiple-pulse lidar systems. The advantage of waveform processing is twofold. By designing his own signal fitting algorithm, an end-user can:

1. Maximize the detection rate of relevant peaks within the waveforms. Waveform decomposition allows to find the echo positions: a waveform is a series of components assuming that the position of each component within the waveform can be used to calculate the mean elevation of a specific object within the laser footprint. Therefore, maximum locations should be better determined, and close objects better discriminated (Stilla et al., 2007; Lin et al., 2010).
2. Model each echo with a suitable parametric function. The echo shape can be retrieved, providing relevant features (the so-called full-waveform features) for subsequent segmentation and classification purposes (Rutlinger et al., 2008; Lin and Mills, 2009; Neuenschwander et al., 2009). This step is called waveform modeling.

More points can indeed be extracted in a more reliable and accurate way. This concerns weak and complex overlapping echoes, occurring for instance in vegetated areas (tree canopy as well as points lying on the ground beneath) and building edges. The detection of low energy echoes provides more complete information about the target which is useful for classification. Weak echoes correspond to echoes with amplitudes below the threshold level of the hardware real-time detection tool of standard multiple-pulse sensors. Complex waveforms are composed of partly overlapping echoes (i.e., a distance between successive targets close to the sensor resolution), which shapes may significantly vary. Hardware echo detection methods have difficulties to resolve these targets correctly.

As depicted in Fig. 1, two distinct methods are used to extract full-waveform features.

First, the procedure presented in Chauve et al. (2007) is adopted. This is a slight modification of the Gaussian decomposition method adopted in the literature: the waveform is considered as a sum of generalized Gaussian (GG) curves, which allows to retrieve the echo amplitude A, width, w, and flatness α, α = v/2 gives the standard Gaussian case. Furthermore, calibration procedures allow to retrieve additional features (Wagner, 2010), based on the echo amplitude, and width: the cross-section σ, the cross-section per illuminated area σ², and the backscatter coefficient γ (see Section 3.2 for more details). The cross-section per illuminated area is the cross-section normalized by the illuminated surface area, in order to take into account the variation of the footprint size (Wagner, 2010). Finally, to both integrate knowledge of the footprint size and the local target geometry, the backscatter coefficient is introduced: this is the cross-section normalized by the illuminated surface area and the incidence angle.

Nevertheless, the Gaussian (symmetric) assumption does not hold for a large range of waveforms. Asymmetric echoes are observed within some waveforms corresponding to ground surface or tree canopy. This is due to the canopy attenuation, the height distribution of the targets or the target geometry. Thus, many waveforms exhibit heavier tails and require a more flexible parametric characterization. The generalized Gaussian model only allows to handle symmetric echoes. The approach developed in Mallet et al. (2010) proposes a versatile framework for modeling lidar waveforms based on the decomposition of the signal as a mixture of various functions. This stochastic approach is based on a marked point process which allows to fuse the decomposition and modeling steps. A priori knowledge on the shape of the lidar waveforms is in particular introduced to both propose
meaningful reconstructions and handle complex waveforms. Each mode of lidar waveforms is fitted by the most relevant parametric function, selected in a given library of models. Three parametric functions are selected, including the GG function. In order to model soft right-skewed and left-skewed behaviors, the Nakagami function has been introduced. For heavy right-skewed modes, the Burr function has been added. The Nakagami function has three parameters (the amplitude $A$, $\omega$ and $\zeta$, namely a skewness and a scale parameter). The Burr function models asymmetric modes with four parameters (the amplitude $A$, $a$, $b$, and $c$, which are one scale parameter and two shape parameters, respectively). The fitting issue is finally turned into an energetic issue and a non-convex optimization problem in a high dimension and variable space, since the three models have not the same number of parameters.

A reversible jump Markov Chain Monte Carlo sampler, coupled with simulated annealing allows to find the global minimum of the formulated energy, which finally gives the most relevant mixture of models for each waveform. An example of the decomposition process is given in Fig. 2.

Practically, all waveforms are processed twice: (1) the first run is dedicated to the modeling process using the single generalized Gaussian function (to get the amplitude, width, flatness, and cross-section based attributes); (2) the second run exploits the entire library of functions, and may provide other shape features. Thus, at the end of the modeling process, each echo of a waveform is represented by a vector of attributes: it contains the 3D position of the georeferenced peak, the parameters of the generalized Gaussian decomposition, and the parameters of the selected model of the stochastic approach. Since the number of parameters varies from three (GG and Nakagami) to four (Burr), the echo shape is finally condensed to a vector of three parameters: the echo amplitude, the selected model (GG, Nakagami, or Burr), and the echo asymmetry, which is computed from the other model parameters (see Section 3.2).

Once each echo of the waveforms synthesized as a parametric form, a vector of features is calculated (cf. Section 3).

### 2.2. Support vector machines classification

Subsequent to waveform processing, point-based classification is performed in order to discriminate ground, vegetation, and building points. Support vector machines (SVM) are a supervised classifier. In our experiments, a vector of features, described in Section 3, is introduced in the classifier to discriminate the three classes of interest: building, vegetation, and ground.

#### 2.2.1. Background on SVM

Support vector machines are among the most used kernel-based statistical learning algorithms. SVM are non-parametric methods, i.e., they work in cases where the data distribution cannot properly be modeled by a standard parametric probability density function. It performs robust non-linear classification of samples using the so-called kernel trick. For non linearly separable data, SVMs can map the data onto a higher dimensional space where they are linearly separable. This allows to improve classification abilities. Finding a separating hyperplane in some feature space, induced by the kernel function, permits to perform all the computations in the original space. This is one of the main reasons why SVMs are well suited for high dimensional classification problems. They aim to find the separating hyperplane maximizing the distance between the closest training samples for two classes. This explains their very high capability of generalization. Consequently, only a few training samples are required. Such properties are particularly suited for remote sensing classification problems, and explains their recent adoption (Mountrakis et al., 2011). This is particularly the case for airborne lidar data (Lodha et al., 2006; Samadzadegan et al., 2010; Braun, 2010). Moreover, since a simple feature vector is required as input, many authors are coupled lidar...
Given a training set \( A = \{x_i, y_i\} \), where \( x_i \in \mathbb{R}^d \) are the \( l \) training samples available with their labels \( y_i \in \{1, -1\} \), the decision function is found by solving the following convex optimization problem:

\[
\max_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^{l} y_i (\mathbf{w}^T \mathbf{x}_i + b)
\]

subject to \( \sum_{i=1}^{l} y_i \alpha_i = 0 \) and \( 0 \leq \alpha_i \leq C \)

\( \alpha \) are the Lagrange coefficients, \( K \) is the kernel function, and \( C \) is a constant which determines the trade-off between margin maximization and training error minimization. Once the optimal solution of (2) (i.e., \( \alpha^* \)) is found, the \( x_i \) are labeled by looking to which side of the hyperplane they belong. More details on the problem formulation, and on the optimization procedure can be found for instance in Schölkopf and Smola (2002).

2.2.2. Kernel selection and parameter fitting

Without sufficient a priori knowledge of the influence of geometric and radiometric parameters of the targets on the echo shape, the design of a kernel is a very difficult task. Therefore, a generic kernel was selected, the Gaussian kernel, defined as:

\[
K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad \text{with} \quad \sigma > 0
\]

where \( \sigma \) tunes how similar to the training data the test data is expected to be. Adopting the Gaussian kernel has been stated by many authors (Fauvel, 2007; Hsu et al., 2008; Kavzoglu and Colkesen, 2009), mainly for practical issues. First, more sophisticated kernels do not necessarily improve the results. Secondly, selecting a kernel with a few number of parameters \( (C \) and \( \sigma \) for the Gaussian kernel) minimizes the associated optimization computing time. Thirdly, the performances of the Gaussian kernel have been extensively analysed, and domains where the two parameter values can lead to over or underfitting are well known (Keerthi and Lin, 2010). Finally, the number of instances (ground truth points) remains much larger than the number of features. This is therefore still interesting to map data into a higher dimensional space and not to prefer the linear kernel (Hsu et al., 2008).

There is no criterion concerning the choice of the two parameters \( C \) and \( \sigma \). A simple grid search approach has been adopted. It remains efficient in case of few parameters (Duan et al., 2003). A grid search exhaustively sweeps the 2D parameter space, and for each point the cross-validation accuracy (CVA) is computed. The CVA represents the percentage of samples correctly classified averaged over all the subsets when they were used as the testing subset (average value between the true positive rate and the true negative rate). After the coarse grid search, a finer one is computed in a smaller range around the optimal parameters found in the first step. The values of the kernel parameters can have a significant impact on the learning capacity but their optimum values are not critical: there is a range of values which gives almost the same cross-validation accuracies.

SVMs are designed to solve binary problems (for a two-class problem, the hyperplane separates the space into two half-spaces, such as each instance \( y_i \in \{-1; +1\} \)). Having more than two classes of interest involves the simultaneous discrimination of these classes which cannot be handled with a unique decision function. Various approaches are possible to address the problem, usually combining a set of binary classifiers, e.g., in a hierarchical manner (Melgani and Bruzzone, 2004). The parallel ‘one-against-one’ approach is selected because it has been shown to be more suitable for large problems and few classes (Hsu and Lin, 2002). Three SVMs are involved to model all possible pairwise classifications of such three-class problem. The LIBSVM software is used to perform the SVM classification (Chang and Lin, 2001).

2.3. Feature selection strategies

Since achieving a high classification accuracy is coupled with assessing the relevance of the FW features, a feature selection step is performed jointly with the classification process. Among the feature vector, certain parameters have a stronger influence on the classification result than the others. Although theoretically considered to be insensitive to the dimensionality of the discrimination problem, SVM classifiers have proved to be influenced by the redundancy, and by the irrelevance of input features. This is also the case of a large variety of classifiers, and such statements have lead towards the development of various feature selection strategies (Kohavi and John, 1997; Guyon and Elisseeff, 2003).

Feature selection is the technique of selecting a subset of relevant attributes among a more global set. This allows to enlighten the real added value of full-waveform features with regards to geometric and echo-based features, that can be extracted directly from a discrete-return 3D point cloud. Moreover, the removal of the less relevant features may also improve the class separability, and the classification accuracy. SVMs do not directly obtain the feature importance, and several techniques exist to evaluate the relevance of full-waveform features. Both unsupervised (such as the principal component analysis) and supervised (such as the linear discriminant analysis) methods exist. Since training datasets are available, the selection step is performed according to the properties of the training set. A very large body of the literature has addressed the issue with both strategies (Mitra et al., 2002; Liu and Yu, 2005; Song et al., 2007; Campedel et al., 2008), and many feature selection challenges exist on various kinds of datasets (Guyon et al., 2006). This is particularly the case for SVM-based classification (Fröhlich et al., 2004; Chen and Lin, 2006; Chapelle and Keerthi, 2008).

Three classification strategies have therefore been selected:

- **No selection**: it consists in using the SVM classifier without selecting any feature. The classification accuracies obtained with the full set of features \( f_x \) and with a subset including only multiple-pulse features (i.e., excluding all full-waveform features, \( f_x \setminus \{A, w, x, \sigma, \phi, \gamma, M, s\} \), see Section 3), are compared.
- **Filter selection**: it requires the computation of a score for each feature. The ranking of the features is directly linked to the order of the scores. Two techniques are investigated:
  - **Relief**: uses the K nearest neighbors of each instance per class to calculate a weight for each feature (Robnik-Šikonja and Kononenko, 2003). If the data from the same label have close values, and the data coming from different classes are well discriminated, the feature will exhibit a significant weight. \( K \) controls the locality of the estimation, it should not be set to high values. Indeed, with increasing \( K \) values, informative features may become less and less distinguishable from useless attributes. We follow Kononenko (1994) and set \( K \) to 10.
  - **F-score** measures the discrimination of two sets of real numbers (Chen and Lin, 2006). The F-score of a feature is the ratio between the interclass variance over the intraclass variance. A large value of the F-score means that this feature is more likely to be discriminative.
- **Wrapper-based selection**: here, the SVM classifier is directly used to conduct feature selection. For this kind of family, the recursive feature elimination (SVM–RFE) has been adopted. The selection algorithm has been initially presented in Guyon
et al. (2002), and extended in Rakotomamonjy (2003). The goal is to find a subset of a given size \( v \) among \( D \) features which maximizes the performance of the predictor. The method is based on a backward sequential selection using a linear SVM. One starts with all the features, and removes one feature at a time until \( v \) features are left. At each iteration, the feature \( d \) to be removed is the one whose removal minimizes the variation of ||\( w_d ||^2 \) (which is a direct link to the margin). The cost function \( C(d) \) is therefore:

\[
C(d) = ||w||^2 - ||w_{\overline{d}}||^2
\]

\[
= \frac{1}{2} \sum_{k} j_{k}^{d} z_{j_{k}}^{d} y_{k} K(x_{k}, x_{j}) - \sum_{k} j_{k}^{d} z_{j_{k}}^{d} y_{k} K_{\overline{d}}^{d}(x_{k}, x_{j})
\]

(4)

where \( K^{d} \) is the matrix of the training data when feature \( d \) is removed, and \( z^{d}_{j_{k}} \) is the solution of Eq. 2.

Contrary to the F-score and ReliefF strategy, the recursive feature elimination method does not provide any relevance metric but a simple feature ranking.

The data mining software Weka has been used to compute the ReliefF metric, and the SVM–RFE feature ranking (Hall et al., 2009). Each of the three methods provide in fact a feature ranking, and not directly a subset of relevant features. The ranking allows to sequentially feed the classifier, one feature by one, starting from the best ranked. The analysis of the evolution of the classification accuracy will lead to the selection of a subset of features providing performances similar or better to the full feature vector. This is interesting for enhancing the generalization capability of the SVM as well as for speeding up the learning process. Finally, since the F-score and ReliefF methods provide a relevance measure for each feature, the latter results are also compared to the accuracy that can be achieved using the full set of weighted features, the weight being the relevance measure.

3. Features of interest

The SVM classifier assigns a label to each 3D point, based on a 27-component feature vector (see Eq. 10). These features are all computed from the 3D lidar point cloud retrieved from the georeferencing process, subsequent to the decomposition and modeling step (see Fig. 1). Part of them are derived from spatial considerations, based on the point height or on the geometrical description of the point neighborhood (eigenvalues or local plane estimate). Other features are related to the echo number, whilst the last ones are specially computed from waveform-based considerations. These features are indeed specific to full-waveform lidar data whereas the spatial-based and echo-based features can be computed from traditional multiple-pulse data. Nevertheless, since waveform processing allows to retrieve additional 3D points and correct information on the position of the current echo/point within the waveform are now available, spatial-based and echo-based features are slightly modified and improved with FW data. Such contribution is also quantified in Section 5.

3.1. Geometrical features

3.1.1. The local environment

For each 3D point \( P \), spatial features are computed using a volumetric approach, i.e., lidar points included in a restricted 3D neighborhood \( V \) are used in addition to the current lidar point \( P \). \( V \) is called environment. It can be a sphere \( V_{1} \) centered on point \( P \); or a vertical cylinder, \( V_{2} \). \( V_{1} \) includes all points inside a 3D environment without restriction on the vertical position. It favors the measure of the dispersion in height of the neighbors. These two environments have a single parameter, the radius \( r \). It has been empirically set up to \( r = 1.25 \) m, according to the point density of the data set. Designing neighborhood environments specific to each feature of interest (Filin and Pfeifer, 2005) as well as finding the optimal neighbors for each point would provide more robust features. However, this point has not been investigated in this paper.

3.1.2. Height features

The height features are related to the altitude of the current 3D point. They may discriminate both above-ground points and points with a significant altimetric scattering. Three features are selected.

- \( \Delta z \): height difference between the lidar point and the lowest point found in a large cylindrical volume whose radius has been experimentally set to 20 m.
- \( \Delta z_{0} \): the height difference between the first and the last echoes of the waveform.
- \( \sigma z \): the height variance of the 3D points included in \( V_{1} \).

3.1.3. Eigenvalue features

A covariance matrix of the 3D coordinates is computed in \( V_{1} \). Such matrix can be considered as a 3D structure tensor that brings information about the point distribution within \( V_{1} \). All three eigenvalues \( \lambda_{1}, \lambda_{2}, \lambda_{3} \) (descending order) are a direct link to the dimensions of the ellipsoid that represents the local tridimensional structure. When \( \lambda_{1} \gg \lambda_{2}, \lambda_{3} \), we face a linear structure (e.g., building edge). A flat ellipsoid \( \lambda_{1} \approx \lambda_{2} \approx \lambda_{3} \) means that the point cloud is locally plane, whereas \( \lambda_{1} \approx \lambda_{2} \gg \lambda_{3} \) corresponds to volumetric structures such as vegetated areas. Many combinations of these features are conceivable to describe one, two, or three dimension structures. They may provide discriminant features, especially in classification urban areas (Carlberg et al., 2009). Due to our three-class problem, our interest is focused on 2D and 3D structures. A visual assessment has allowed us to keep the most promising ones: the larger range of values possible, and a good discrimination between vegetation and ground/buildings.

- \( \Sigma_{e} \): the sum of the three eigenvalues (highest values for 3D objects such as trees).
- \( e_{1}, e_{2}, \) and \( e_{3} \): the three eigenvalues normalized by \( \Sigma (\forall i \in [1,3], \ e_{i} \equiv e_{i}/\Sigma) \).
- \( A_{i}, O_{i}, P_{i}, E_{i}, S_{i} \): anisotropy, omnivariance, planarity, eigententropy, and scatter, respectively. They are eigen-based features describing the spatial local distribution. Their formulation are taken from Gross and Thoennessen (2006), with slight variations (normalizing coefficient).
  - \( A_{i} = \frac{e_{i}}{\Sigma_{e}} \) and \( P_{i} = 2(e_{2} - e_{1}) \) allow to exhibit locally planar points;
  - \( O_{i} = \sqrt{e_{1} e_{2} e_{3}}, \ E_{i} = -\ln(e_{1})^{\alpha} - \ln(e_{2})^{\alpha} - \ln(e_{3})^{\alpha}, \) and \( S_{i} = \frac{e_{i}^{2}}{\Sigma_{e}^{2}} \) (Toshev et al., 2010) are used to discriminate 3D structures.

3.1.4. Local plane features

The local plane features are related to the scattering of the points with regards to a 3D local plane \( \Pi \). Such plane has been estimated by a robust M-estimator with norm \( L_{1,2} \) (Xu and Zhang, 1996), using the 3D points included in \( V_{2} \). The deviation angle from the vertical direction is useful for discriminating ground from building roofs \( (N_{z} = 0 \) and \( N_{\eta} \in [-45, 45], \) respectively). The two other features are good hints for enhancing the altimetric scattering around the current point \( P \).

- \( N_{z} \): deviation angle of the local normal vector from the vertical direction.
- \( \sigma_{n_{z}} \): variance of the deviation angles, computed for all points in \( V_{1} \).
• \( R_z \): residuals of the local plane \( \Pi \) such as \( R_z = \sum_{i \in V_i} \frac{p_i d_i}{3} \), where \( d_i \) is the distance between the lidar point \( i \in V_i \) and the plane \( \Pi \), and \( p \) is the norm (here \( p = 1.2 \)).

Many other geometrical features are conceivable and may be relevant for discriminating the classes of interest. For instance, attributes used for the specific problem of building detection from lidar data may also be useful in our context. Hence, border gradients or texture features (Elberink and Maas, 2000) such as grey level co-occurrence matrix features (Matkainen et al., 2007), slope indices (Elmqvist et al., 2001; Tóvári, 2006) or roughness measures (Rottensteiner et al., 2005) can be useful. They are often computed from raster data such as digital surface models but may be easily adapted to 3D multiple-pulse data.

3.1.5. Echo-based features

The echo features enhance the multiple scatterings of the laser pulse (i.e., vegetation points). They are:

- \( n \): the echo number.
- \( N \): the total number of echoes within the waveform of \( P \).
- \( N_z \): the normalized number of echoes \( N_z = \frac{N}{w} \).
- \( PDR \): the point density ratio. This is ratio between the number of points within \( V_i \) and within \( V_r \) (Rutzinger et al., 2008).

3.2. Full-waveform features

These features are derived from the double waveform processing step (see Section 2.1): the generalized Gaussian decomposition and the stochastic approach. Other waveform shape descriptors should be used. Waveform metrics such as the rise time to the first peak are used in Neuenschwander et al. (2009). The waveform coefficients presented in Laky et al. (2010) are also conceivable.

- \( A, w, \alpha \): the echo amplitude \( A \), width \( w \), and flatness \( \alpha \) are the three parameters of the generalized Gaussian model. Thanks to the recording of the emitted pulse, the amplitude of each echo \( A_e \) has been corrected from the fluctuations of the emitted peaks along the flight line, normalized by the ratio of the average amplitude value of all emitted pulses \( A_e \), and the amplitude of the emitted pulse of the current peak \( A_{current} \) (Bretar et al., 2009), as well as from the range to the sensor, and the incidence angle, following:

\[
A = \frac{A_e - \bar{A}_s}{R^2} \frac{\bar{A}_s}{R_{currently}^2}
\]

where \( R, R_i \) and \( \theta \) are the range between the current point \( P \) and the sensor, a standard range for the whole survey area, and the incidence angle between the laser beam and the estimated local plane, respectively (Hölle and Pfeifer, 2007). Such correction procedure, coupled with a suitable estimation method allows us to consider the amplitude as a full-waveform feature, however an amplitude value is sometimes available with discrete return data.

- The generalized Gaussian model allows to derive three other features: the cross-section \( \sigma \), the cross-section per illuminated area \( \sigma^0 \), and the backscatter coefficient \( \gamma \). Recent works have highlighted their relevance for classification purposes (Alexander et al., 2010). For a generalized Gaussian model, the cross-section can be formulated as:

\[
\sigma = C_{cal} R \pi Aw R^2 \Gamma \left( \frac{1}{2} \right)
\]

where \( C_{cal} \) is the calibration constant and \( \Gamma \) the Gamma function. This is a slight modification of the Gaussian-based formulation presented in Wagner et al. (2006). The calibration constant is calculated by selecting small asphalt areas close to the nadir view (±2°), for each lidar strip where \( A_{asphalt} \) and \( w_{asphalt} \) are extracted (Wagner et al., 2006). According to the Auster spectral library (Baldridge et al., 2009), the reflectance of asphalt is \( R_{asphalt} = 0.15 \). If \( \beta \) is the laser beam divergence, we have:

\[
C_{cal} = \frac{\pi R_{asphalt} w^2}{PDR A_{asphalt} w_{asphalt}}
\]

Utilizing asphalt targets is not optimal. Firstly, low reflectance values means low signal-to-noise ratio which will provide a calibration of low quality. Secondly, the Lambertian assumption of asphalt areas is wrong: significant backscattering variations (>20%) have been noticed for various incident angles, and the optical properties of such materials also spatially vary a lot. However, no suitable ground surface has yet been found.

Besides, \( \sigma^0 \) is the cross-section normalized by the laser footprint area:

\[
\sigma^0 = \frac{\sigma}{A_i}
\]

where \( A_i \) is the illuminated surface area (Wagner, 2010).

Finally, the backscatter coefficient \( \gamma \) is \( \sigma^0 \) corrected from the incidence angle. When assuming a circular transmitter aperture,

\[
\gamma = 4\sigma \frac{\pi^2}{R^2} \beta^2
\]

where \( R \) is the range between the 3D point and the sensor. Two last features are derived from the stochastic approach presented in Mallet et al. (2010).

- \( M \): the selected modeling function (generalized Gaussian, Nakagami, Burr),
- \( s \): the echo asymmetry. This is the ratio between the echo width at half-maximum before the echo mode and the echo width at half-maximum after the echo mode (noted \( w_1 \) and \( w_2 \), respectively). The full-width at half-maximum (FWHM) is equal to \( w_1 + w_2 \).

\[
s = \frac{w_1}{w_2} = \frac{\int_{-\infty}^{x} \frac{1}{f(x)} \frac{x}{\sqrt{2 \pi}} dx}{\int_{-\infty}^{\infty} \frac{1}{f(x)} \frac{x}{\sqrt{2 \pi}} dx}
\]

where \( f \) is the modeling function, \( x^* \) is the echo location, and \( 1 \) the characteristic function. \( s < 1 \) means that the echo is left-skewed, \( s = 1 \) corresponds to symmetric echoes, whereas \( s > 1 \) reveals a right-skewness.

Finally, the feature vector \( f_v \) can be written as follows:

\[
f_v = \begin{bmatrix}
\Delta z, \Delta x, \sigma_c, \\
\Sigma_z, \Sigma_e, \Sigma_x, A, O, P, E, S;
\end{bmatrix}
\]

\[
f_v = \begin{bmatrix}
N_z, \sigma_{n_z}, \Sigma_z; \\
n, N, N_e, PDR; \\
A, w, \sigma, \alpha, \sigma^0, \gamma, M, s
\end{bmatrix}
\]

4. The data set

The data set has been acquired over the city of Biberach, Baden-Wuerttemberg, Germany. The survey has been carried out by Toposys on September 2006, with the Riegli LMS-Q560 system, covering approximately 1.1 km² (leaf-on condition) One strip is available, with a point density close to 5 points/m². It results on more than 2.2 million waveforms. The strip covers different kinds of landscapes: a residential area, a wooded area with significant slope (45°), a semi-dense downtown, and an industrial area. Fig. 3
provides an overview of the area and of the 3D point cloud computed from the lidar waveforms. This is the center of a small European city: there are no high buildings, no specific building organization, close buildings can be found but with no comparison with Mediterranean city centers.

The whole waveform processing chain has been carried out. Firstly, range values provided by the on-board signal processing algorithms, emitted pulses, backscattered signals are extracted from raw full-waveform files. Secondly, the waveform processing and the georeferencing steps are subsequently performed to turn the series of estimated ranges into a 3D point cloud.

A 0.25 m resolution orthoimage covers the city of Biberach (except the industrial area). This image is not introduced in the classification scheme. Besides, this image is used to create manually 2D polygons of the classification ground truth, which is necessary for both training the supervised classifier, and for assessing its accuracy: 3D points lying with x, y coordinates lying on such polygons will be labeled according to the polygon semantic. It works well for ground and building regions. For the vegetation class, the third dimension (or multiple-pulse) problem may be solved by labeling only echoes that are not last echoes (e.g., using the normalized number of echoes – \( N_e = 1 \)). Such process is sufficient for the SVM training step, but has one main drawback for accuracy assessment. A large majority of 3D points located on the tree canopy or on the understory, as well as those lying on the building edges, are not labeled, whereas they are likely to be misclassified. As a consequence, it will lead to an overestimation of the accuracy of the classification process. However, this approach is simple and has been selected: approximately 1/3 of the 3D points over Biberach have been labeled.

The training step has been performed with 100 samples per class randomly selected. This corresponds to 0.04% of the training set, and approximately to 0.01% of the test set. Another 0.05% of the samples have been randomly chosen in the ground truth to find the optimal values of \( C \) and \( \nu \) in the cross-validation procedure. For all the tests carried out, the correct classification rate for the training step oscillated between 80% and 90%. It illustrates that the SVM classifier does not over-fit, but is able to generalize, and has been trained sufficiently.

5. Results and discussion

5.1. Feature ranking and relevance assessment

The features of \( f_v \) have been ranked using the three methods: F-score, ReliefF, and SVM-RFE. The results are provided in Table 1. Since the methods have distinct behaviors, and do not behave equally when dealing with redundant features, the rankings may vary significantly. In particular, PDR and \( n \) are among the four most
relevant features for SVM–RFE, whereas F-score and ReliefF methods rank them after the 10th position. For a general statement, the mean ranking has been computed for each feature and is displayed in Fig. 4. Therefore, one can notice that:

- Some discrete return attributes are always ranked among the less discriminant features (E, z, and N), and other ones are always among the most relevant features (A, N, and N). We can conclude that both spatial and echo-based features are discriminative. The detection of additional 3D points using FW data improves the computation of z and N.
- According to Fig. 4, full-waveform features can be classified in three groups. The first one corresponds to well-ranked features (A, N, and N). The second group gathers low-ranked attributes (M, w, s, and a), whereas s is ranked in-between.

The following relevance order of full-waveform features is therefore given:

\[
A > \gamma > \sigma > s > w > s > x
\]

In particular, before mentioning any classification accuracy, one can notice that the flatness x and echo asymmetry s are always ranked among the five last features. They can be considered as irrelevant for our three-class problem. However the generalized Gaussian model has shown that lidar echoes differ from the standard Gaussian assumption [x < 1.5, 1.7]. x only slightly behaves differently between building areas and ground and vegetation areas (Mallet et al., 2008). It does not bring enough information for our discrimination problem. Similarly for the echo asymmetry, it is shown in Mallet et al. (2010) that only 20% of the lidar waveforms in urban areas are not modeled by a symmetric function, and that, for small-footprint data, the asymmetry is never significant [x < 0.9, 1.3] and not linked to specific objects: asymmetry may come from building edges as well as trees and slopped surfaces. This explains its low rank.

In addition to their rank, relevance can also be measured by evaluating the contribution of each feature to the improvement of the classification accuracy. For that purpose, for each of the three ranking methods, according to their rank, the features are progressively added in the feature set. For each feature set, the SVM classification is performed (training and testing), and the accuracy is computed according to the ground truth. The overall accuracy (OA) and average accuracy (AA) are used as quality measures (Congalton and Green, 1998). Fig. 5 provide the evolution of OA

Table 1

<table>
<thead>
<tr>
<th>Rank</th>
<th>F-score</th>
<th>ReliefF</th>
<th>SVM-RFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N e</td>
<td>Δz</td>
<td>Δz</td>
</tr>
<tr>
<td>2</td>
<td>Δz</td>
<td>σ z</td>
<td>PDR</td>
</tr>
<tr>
<td>3</td>
<td>e z</td>
<td>N e</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>S λ</td>
<td>A</td>
<td>N e</td>
</tr>
<tr>
<td>5</td>
<td>O λ</td>
<td>O λ</td>
<td>N e</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>N e</td>
<td>γ</td>
</tr>
<tr>
<td>7</td>
<td>N</td>
<td>γ</td>
<td>σ</td>
</tr>
<tr>
<td>8</td>
<td>σ z</td>
<td>σ 0</td>
<td>σ z</td>
</tr>
<tr>
<td>9</td>
<td>P λ</td>
<td>σ</td>
<td>N</td>
</tr>
<tr>
<td>10</td>
<td>A λ</td>
<td>N</td>
<td>w</td>
</tr>
<tr>
<td>11</td>
<td>R z</td>
<td>P λ</td>
<td>S λ</td>
</tr>
<tr>
<td>12</td>
<td>PDR</td>
<td>S λ</td>
<td>M</td>
</tr>
<tr>
<td>13</td>
<td>γ</td>
<td>e z</td>
<td>E λ</td>
</tr>
<tr>
<td>14</td>
<td>σ</td>
<td>σ N e</td>
<td>Δ z 0</td>
</tr>
<tr>
<td>15</td>
<td>Δ z 0</td>
<td>e z</td>
<td>e z</td>
</tr>
<tr>
<td>16</td>
<td>N e</td>
<td>P DR</td>
<td>A λ</td>
</tr>
<tr>
<td>17</td>
<td>σ 0</td>
<td>R z</td>
<td>O λ</td>
</tr>
<tr>
<td>18</td>
<td>σ N e</td>
<td>M</td>
<td>N e</td>
</tr>
<tr>
<td>19</td>
<td>n</td>
<td>Δ z 0</td>
<td>σ N e</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>A λ</td>
<td>σ 0</td>
</tr>
<tr>
<td>21</td>
<td>e z</td>
<td>s</td>
<td>α</td>
</tr>
<tr>
<td>22</td>
<td>e z</td>
<td>s</td>
<td>α</td>
</tr>
<tr>
<td>23</td>
<td>Σ λ</td>
<td>w</td>
<td>e z</td>
</tr>
<tr>
<td>24</td>
<td>w</td>
<td>n</td>
<td>e z</td>
</tr>
<tr>
<td>25</td>
<td>s</td>
<td>α</td>
<td>s</td>
</tr>
<tr>
<td>26</td>
<td>E λ</td>
<td>E λ</td>
<td>R z</td>
</tr>
</tbody>
</table>

Fig. 4. Global feature ranking using the three methods. The mean rank is plotted for each feature. Full-waveform features correspond to white bars. The value inside brackets indicates how many times a full-waveform feature has been selected in the optimal subset for the three methods.
and AA for the three methods. It enhances the logarithmic behavior of the classification performance whilst adding less and less discriminative attributes. The advantage of such curves is two-fold. They allow to find: (1) the subset that achieves the highest accuracy; (2) the minimal subset necessary to reach a suitable accuracy. Suitable accuracy can be located one or two features after the inflexion point of the accuracy curves. As mentioned before, the less features are involved, the less complicated the decision function will be, and the more important the generalization capability of the SVM will be. We consider as discriminant, features that appear in the minimal subsets.

- F-score: an almost optimal accuracy is found for seven attributes \( \{N_e, \Delta \zeta, \varepsilon_3, S_e, O_e, A, N\} \), and the highest score is reached for the full set (94.0%, cf. Fig. 5a). This reveals the low performance of such feature selection method.
- Using ReliefF (Fig. 5b), the highest accuracy is reached for a subset of 18 features (94.9%), whereas suitable accuracy is found with only six features (94.6% for \( \{\Delta \zeta, \sigma_\zeta, N_e, A, O_e, N_d\} \)).
- The recursive feature elimination strategy allows to achieve the best accuracy (95.3%, cf. Fig. 5c) with a subset of eight features \( \{\Delta \zeta, PDR, A, n, N_e, \gamma, \sigma, \sigma_\zeta\} \).

Full-waveform features are present in each minimal subset of the three methods, and their occurrence are reported in Fig. 4. \( A, \gamma \), and \( \sigma \) can therefore be labeled as discriminant features with respect to the full feature set. This enhances the necessity of defining and carrying out suitable calibration procedures. \( \sigma, M \), and \( w \) are thus not considered as relevant for our three-class problem. Cross-section and echo width information are already included in \( \gamma \) and \( \sigma^a \) features, that have also the advantage of being normalized. Since the three selected ranking methods are not always able to correctly deal with redundant features, this may explain their low performance. Finally, \( M \) is not discriminant, as it was foreseen.

First, two modeling functions with distinct parameters can lead to almost the same shape, and therefore be similarly selected by the stochastic approach, which results in slightly noisy results (Mallet et al., 2010). Secondly, the generic Gaussian kernel is not particularly adapted to handle features with integer values \( \{M_1; M_2; M_3\} \) in practice: indeed, if \( M_1 = M_2, M_3(x, x) = 1 \).

### 5.2. Classification accuracies

Table 2 summarizes the classification performances achieved with the various strategies that have been tested. First, the simple use of the eight full-waveform features allows to reach an average accuracy of 64.8%, whereas the full feature vector introduced in the SVM classifier allows to achieve an average accuracy of 94.35%. The average accuracy drops to 91.4% when the eight full-waveform fea-

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Average accuracy (%)</th>
<th>( f_d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-waveform features</td>
<td>64.8</td>
<td>8</td>
</tr>
<tr>
<td>Multiple-pulse features</td>
<td>91.4</td>
<td>19</td>
</tr>
<tr>
<td>All features</td>
<td>94.35</td>
<td>27</td>
</tr>
<tr>
<td>F-score + subset selection + SVM</td>
<td>93.8</td>
<td>7</td>
</tr>
<tr>
<td>F-score + weighted SVM</td>
<td>94.4</td>
<td>27</td>
</tr>
<tr>
<td>ReliefF + subset selection + SVM</td>
<td>94.6</td>
<td>6</td>
</tr>
<tr>
<td>F-ReliefF + weighted SVM</td>
<td>94.7</td>
<td>27</td>
</tr>
<tr>
<td>SVM–RFE + subset selection + SVM</td>
<td>95.3</td>
<td>8</td>
</tr>
</tbody>
</table>

Fig. 5. Evolution of the classification accuracies (overall and average accuracies) for the three feature selection methods: forward introduction of features, according to their rank.
tures are discarded, what underlines their positive impact for the classification results, without applying any feature selection algorithm.

Besides, as mentioned in the previous section, the average accuracies achieved with the full set of features and with the minimal subset selected with Fig. 5a can be compared for the three selection strategies. Moreover, since F-score and ReliefF methods also provide a relevance measure, these two accuracies can also be compared to the AA achieved with the full set of features weighted by these measures.

- Using the relevance measure of the F-score allows to improve AA from 93.8% to 94.4%.
- Feature weighting using the ReliefF metric entails a slight improvement the AA (94.6% to 94.7%).
- The highest accuracy is achieved with the selection of the minimal subset with the SVM–RFE method (95.3%).

Figs. 3, 6, and 7 show the classification results for the full area of Biberach, and two regions of interest (a residential area, and a dense urban center). The classification has been achieved with the optimal solution: selection of a subset of features using the SVM–RFE, and afterwards application of SVM over this subset. The confusion matrix is reported in Table 3.

Misclassified points are found first on building edges (see Fig. 8). They are labeled as vegetation since such points feature similar behaviors for many attributes, especially the most relevant ones: low amplitude and large echo width, equivalent point density ratio, and multiple reflections. Besides, errors appear very locally on roof tops and on the ground, where points are labeled as vegetation and buildings, respectively. They mainly correspond to low-rise objects (chimneys, cars or poles) that are not taken into account in the classification process, and which do not behave as the large majority of the 3D points belonging to their class. Finally, the major source of errors comes from the labeling of whole buildings as ground points (red rectangular areas in Fig. 8). Such buildings are not very elevated, and their roofs are composed of metal. Their scattering properties are therefore very similar with those of streets. Since full-waveform features $A$, $g$, and $\sigma^0$ are included in the optimal subset, this therefore leads to a certain confusion.

However, these misclassified points can easily be corrected by adding some contextual knowledge at the object level in the regularization algorithm.

### 5.3. Historical assessment

Finding the most relevant features among an unordered set of attributes is the first and most objective way to assess the contribution of full-waveform data. Furthermore, such assessment can also be carried out by adding progressively the attributes within the feature set, and performing the classification with the available set. It has therefore been decided to insert them according to their “historical order of appearance”. The evolution of the AA is reported in Fig. 9.

We first start with all the attributes that can be computed with the simple knowledge on the $\{x, y, z\}$ triplet (i.e., a simple 3D point cloud which corresponds to 14 features – scenario 1). The echo number and the amplitude are then successively added (five new features – scenario 2, and one for scenario 3, respectively), enhancing the development of discrete return sensors providing an uncalibrated amplitude value. This first group of 20 features does not need any waveform recording process. Later, with the development of full-waveform sensors, the Gaussian decomposition method has provided another feature (namely the echo width), as well as those computed for the calibration procedure (three other

<table>
<thead>
<tr>
<th># Reference points</th>
<th>Class</th>
<th>Building</th>
<th>Ground</th>
<th>Vegetation</th>
</tr>
</thead>
<tbody>
<tr>
<td>275,058</td>
<td>Building</td>
<td>94.4</td>
<td>3.8</td>
<td>1.8</td>
</tr>
<tr>
<td>333,506</td>
<td>Ground</td>
<td>1.7</td>
<td>98.1</td>
<td>0.2</td>
</tr>
<tr>
<td>157,281</td>
<td>Vegetation</td>
<td>5.0</td>
<td>0.1</td>
<td>94.9</td>
</tr>
</tbody>
</table>

**Fig. 8.** Qualification of the classification over Area #2 (see Fig. 7).
features – scenario 4). Afterwards, the generalized Gaussian model has allowed to extract the echo flatness (scenario 5), and the stochastic approach jointly the best-fit model and the peak asymmetry (two features – scenario 6).

For each of these six scenarios, a SVM classification is performed without any feature selection step. Both discrete return and full-waveform data have been tested. One can note that discrete return data only allows to reach scenario 3. Discrete return data correspond to the point cloud that is provided with the full-waveform datasets.

When comparing discrete return and FW data, one can first note that the simple knowledge of the 3D position of the points leads to slightly better results with multiple-pulse data than with FW data. This can be explained by the fact that the waveform processing step allows to retrieve points in complex locations, such as building edges, facades, or inside the tree canopies. Such points typically correspond to confusing areas, where the classification procedure may fail. This explains why the results are getting worse. However, the results are reversed with the addition of the amplitude feature: the classification process clearly benefits from its correction procedure (Eq. (5)). From scenario 1 to scenario 4, i.e., until the waveform Gaussian decomposition process, one can note in Fig. 9 a constant improvement of the AA (from 85.2% to 94.75%). Then, the addition of the shape feature \( \alpha \) only allows to reach an AA of 94.9%. Again, it shows that the generalized Gaussian model is at present time more relevant for improving waveform fitting. Finally, the same conclusion can be drawn with the scenario 6, since the knowledge of the asymmetry of the echoes provide worse results (94.9% → 94.35%). However, this can only be stated for our three-class problem, using a standard SVM classifier, and with our limited ground truth.

### 6. Conclusions

This paper has addressed the issue of the classification of 3D point clouds in urban areas, using full-waveform data. The main goal was to design an objective workflow for evaluating the real contribution of full-waveform data in such a classification context. The focus has been therefore made on three main classes: **building**, **ground**, and **vegetation**. Furthermore, this workflow is versatile and flexible since it offers the possibility to change the modeling functions during the waveform processing step (and therefore add new features for the labeling issue) without modifying the classification step. Besides, it still remains valid for other classes of interest since one just has to modify the training step of the SVM classifier to take these new classes into account.

A supervised SVM classifier has indeed been used, and operates on a 27-component feature vector. Nineteen are extracted from the point cloud based both on geometric and echo considerations whereas eight are specific to waveform data. In order to quantify the relevance of full-waveform data for land-cover analysis, the classifier has been coupled with three feature selection step. It has allowed us to design several classification scenarios. Firstly, it may permit to weight the features in the SVM classifier, which improves the accuracy; secondly, the analysis of the evolution of the accuracy with forward selection enables to find the minimal feature set allowing to achieve suitable accuracy. With a maximal average accuracy close to 95%, common trends emerge regarding the relevance of full-waveform data:

- three full-waveform features, namely the amplitude, the cross-section per illuminated area, and the backscatter coefficient are among the most discriminative features;
- the flatness and especially the skewness attributes are among the less discriminative ones.
- the type of function (generalized Gaussian, Nakagami, Burr) used to model the echoes has a limited impact for our three-class problem: a simple symmetric decomposition method is sufficient.
- The amplitude correction and calibration procedure is fundamental in order to extract discriminant features.

Two limitations appear in this work. Firstly, the limited ground truth overestimates the classification accuracy and does not allow to perfectly take into account 3D points located in the border between classes. Secondly, performing feature selection contributes to a better understanding of the genuine contribution of FW data but adding redundant features in a same set (or using methods that are not able to cope with redundant features) prevents from concluding on the contribution of each feature. Data mining

---

**Fig. 9.** Evolution of the average accuracy of the classification procedure according to the progressive insertion of the 27 features of interest. The “historical” order of appearance has been selected for the introduction order. The two plots correspond to the two distinct kinds of lidar data than have been tested (multiple-pulse versus full-waveform data).
approaches would first be carried out before the feature selection step to deal with such issue.

Finally, many perspectives are conceivable. First, classifying other areas with the same ground truth would assess whether normalized and corrected full-waveform features are indeed no longer area-specific neither sensor-specific, and can be relevant at large scales. Besides, testing new symmetric modeling functions in the waveform processing step would provide new morphological parameters that may be relevant for classification purposes. Furthermore, the classification problem has been restricted to three labels, arguing that it was the fairest strategy to evaluate the impact of FW data with respect to discrete return data. Once such a coarse classification achieved, it would be relevant to test whether full-waveform features may help to discriminate ground surfaces, roof materials or tree species. This seems to be the topic of interest with the highest potential for full-waveform data.

Acknowledgments

The authors would like to thank Marine Campedel (Telecom ParisTech) for fruitful discussions on SVM and feature selection, Xavier Briottet (ONERA/DOTA) for discussions on radiometric calibration, and Nicolas David (IGN), for his help on waveform processing and data georeferencing.

References


