remote sensing

MDPI

Identification of Linear Vegetation Elements in a Rural Landscape Using LiDAR Point Clouds

Chris Lucas [1,2], Willem Bouten [3] , Zsófia Koma [3] , W. Daniel Kissling [3] and Arie C. Seijmonsbergen [3,*]

[1] Geodan, President Kennedylaan 1, 1079 MB Amsterdam, The Netherlands; c.lucas@vu.nl
[2] Spatial Information Laboratory (SPINLab), Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
[3] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands; w.bouten@uva.nl (W.B.); z.v.koma@uva.nl (Z.K.); wdkissling@gmail.com (W.D.K.)
* Correspondence: a.c.seijmonsbergen@uva.nl; Tel.: +31-20-525-8137

check for
updates

Abstract: Modernization of agricultural land use across Europe is responsible for a substantial decline of linear vegetation elements such as tree lines, hedgerows, riparian vegetation, and green lanes. These linear objects have an important function for biodiversity, e.g., as ecological corridors and local habitats for many animal and plant species. Knowledge on their spatial distribution is therefore essential to support conservation strategies and regional planning in rural landscapes but detailed inventories of such linear objects are often lacking. Here, we propose a method to detect linear vegetation elements in agricultural landscapes using classification and segmentation of high-resolution Light Detection and Ranging (LiDAR) point data. To quantify the 3D structure of vegetation, we applied point cloud analysis to identify point-based and neighborhood-based features. As a preprocessing step, we removed planar surfaces such as grassland, bare soil, and water bodies from the point cloud using a feature that describes to what extent the points are scattered in the local neighborhood. We then applied a random forest classifier to separate the remaining points into *vegetation* and *other*. Subsequently, a rectangularity-based region growing algorithm allowed to segment the *vegetation* points into 2D rectangular objects, which were then classified into linear objects based on their elongatedness. We evaluated the accuracy of the linear objects against a manually delineated validation set. The results showed high user's (0.80), producer's (0.85), and total accuracies (0.90). These findings are a promising step towards testing our method in other regions and for upscaling it to broad spatial extents. This would allow producing detailed inventories of linear vegetation elements at regional and continental scales in support of biodiversity conservation and regional planning in agricultural and other rural landscapes.

Keywords: rural landscapes; classification; LiDAR; linear vegetation; object recognition; point cloud; rectangularity; segmentation

1. Introduction

The European landscape has dramatically changed during the Holocene as a result of human impact and climatic change [1,2]. Especially since the industrial revolution, landscapes have been deforested and reshaped into rural and agricultural landscapes. These are dominated by a mosaic of grasslands, forests, and urban areas, separated or connected by linear landscape elements such as roads, ditches, tree lines, vegetated lynchets, and hedgerows [3–5]. The distribution, abundance and richness of species in these landscapes is related to the amount, height, length, and quality of linear vegetation

elements [6–8]. The same holds true for the dispersal of seeds and the flow of matter, nutrients, and water [1,9]. Additionally, linear infrastructures such as roads and railways form barriers which lead to habitat fragmentation. In contrast, green lanes, which are flanked by hedges and/or tree lines may form connecting corridors. Hence, linear vegetation elements are of key importance for biodiversity in agricultural landscapes. Nowadays there is also awareness that historic agricultural practices are part of the cultural heritage [10,11] and need to be conserved. However, the occurrence of green lanes and hedgerows has strongly diminished in many countries [12,13]. This is mostly a consequence of larger agricultural fields, monocultures and a reduction in non-crop features which reduces the complexity and diversity of landscape structure [8]. Detailed knowledge of the spatial occurrence, current status, frequency and ecological functions of linear vegetation elements in a landscape is therefore of key importance for biodiversity conservation and regional planning.

The mapping of linear vegetation elements has traditionally been done with visual interpretations of aerial photographs in combination with intensive field campaigns [14]. However, this approach is time-consuming and has limited transferability to large areas. New methods have, therefore, been developed that use raster images to map linear vegetation elements by using their spectral properties in visible or infrared wavelengths, e.g., from the French SPOT satellite system, the ASTER imaging instrument and Landsat imagery [15–17]. This allows an automated and hierarchical feature extraction from very high-resolution imagery [14]. Despite these developments, comprehensive high-resolution inventories of linear vegetation elements such as hedgerows and tree lines are lacking at regional and continental scales. The lack of such high-resolution measurements of 3D ecosystem structures across broad spatial extents hampers major advancements in animal ecology and biodiversity science, e.g., for predicting animal species distribution in agricultural landscapes [18]. On a European scale, density maps of linear vegetation elements (and ditches) have been produced at 1 km^2 resolution through spatial modeling of 200,000 ground observations [5]. However, these maps strongly depend on spatial interpolation methods as well as regional environmental and socio-economic variation and therefore contain a considerable amount of uncertainty in the exact spatial distribution of linear vegetation elements in the landscape. High-resolution measurements of 2D and 3D ecosystem structures derived from cross-national remote sensing datasets are therefore needed to identify and map linear vegetation elements across broad spatial extents [18].

An exciting development for quantifying 3D ecosystem structures is the increasing availability of high-resolution remote sensing data derived from Light Detection and Ranging (LiDAR) [19]. LiDAR data have important properties which are useful for the detection, delineation and 3D characterization of vegetation, such as their physical attributes x, y, z, laser return intensity, and multiple return information [20,21]. Vegetation partly reflects the LiDAR signal and usually generates multiple returns, including a first return at the top of the canopy and a last return on the underlying terrain surface. This provides valuable information for separating vegetation from non-vegetation [19]. Moreover, the intensity values describe the strength of the returning light, which depends on the type of surface on which it is reflected and therefore provides information on the surface composition [22]. The shape and internal structure of vegetation can be analyzed by classifying information from the different return values and a variety of features, which can be calculated from the point cloud [19,23]. Some applications of using airborne LiDAR data to quantify linear elements in agricultural landscapes already exist, e.g., the extraction of ditches in a Mediterranean vineyard landscape [3]. The increasing availability of nation-wide and freely accessible LiDAR data from airborne laser scanning in several European countries provides exciting new avenues for characterizing 3D vegetation structures in agricultural landscapes [18].

Here, we present a transparent and accurate method for classifying linear vegetation elements in an agricultural landscape using LiDAR point clouds derived from airborne laser scanning. In this paper, we focus on linear vegetation elements that are predominantly woody, i.e., composed of shrubs and trees. We refer to them as tall vegetation without defining a strict height in the point cloud. We develop the method using free and open source data and analysis tools and apply it for characterizing different

linear vegetation elements in a rural landscape of the Netherlands composed of agricultural fields, grasslands, bare soil, roads, and buildings (Figure 1). Our method provides a promising first step for upscaling the detection of linear vegetation objects in agricultural landscapes to broad spatial extents.
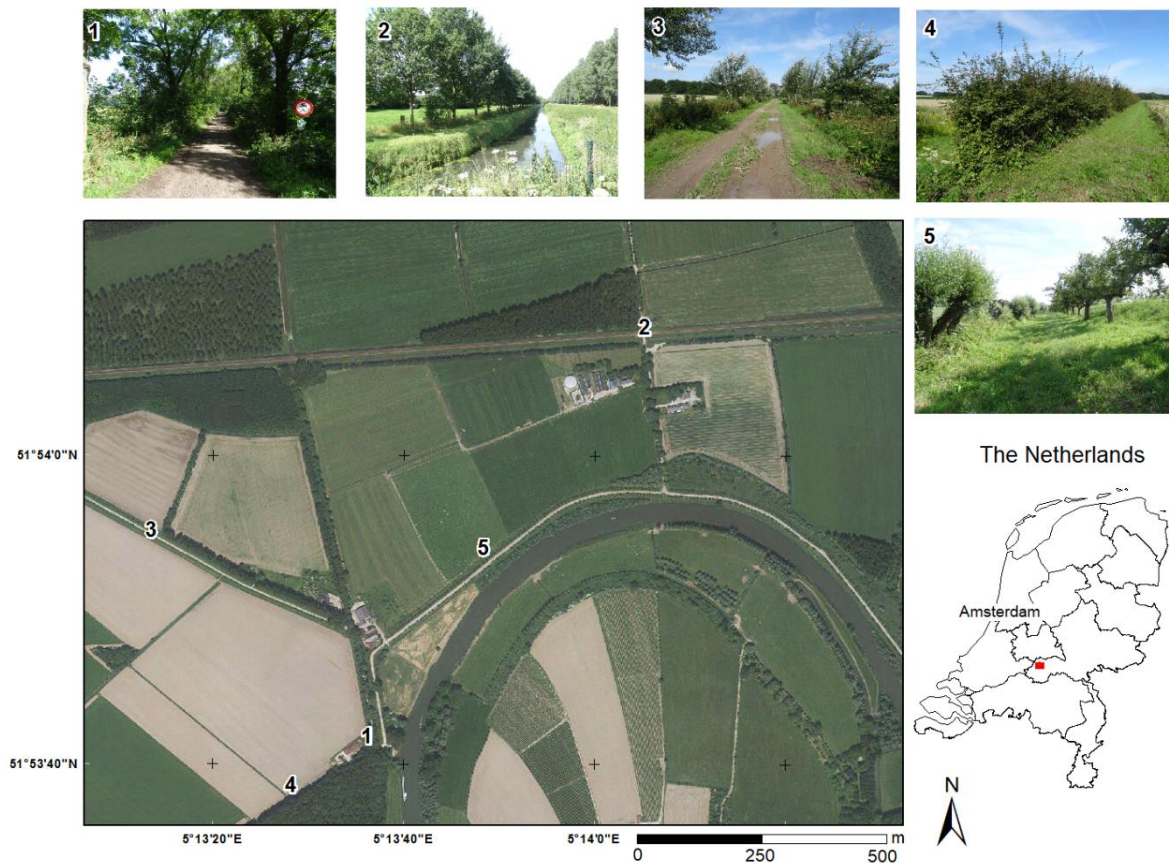
Figure 1. Location of the rural landscape in the central part of the Netherlands. The true color aerial photograph [24] shows various objects identified as agricultural fields, grasslands, bare soil, and infrastructure such as paved and unpaved roads and farmhouses. The numbered photos show a selection of the variety of linear landscape elements such as (1) green lanes (2) planted tall tree lines along ditches, (3) low and high shrubs/copse, (4) hedges and (5) rows of fruit trees and willows.

2. Data and study area

2.1. LiDAR and Orthophoto Data

Raw LiDAR point cloud data were retrieved from "Publieke Dienstverlening op de Kaart" [24], an open geo-information service of the Dutch government. The data are part of the 'Actueel Hoogtebestand Nederland 3' (AHN3) dataset, which was collected between 2014 and 2019. The density of the LiDAR data is around 10 pulses/m^2 and includes multiple discrete return values (which can result into effective point densities between 10 and 20 points/m^2) as well as intensity data. The dataset is collected in the first quarter of each year when deciduous vegetation is leafless [25]. Nevertheless, the return signal is sufficiently strong to retrieve a useful scan of the vegetation cover. Freely available very high resolution (VHR) true color orthophotos [24] with a resolution of 25 cm were consulted for location purposes.

2.2. Study Area

The study area is located in a rural landscape in the center of the Netherlands (Figure 1). The area is about 1.6 km from east to west and 1.2 km from north to south, spanning an area of almost 2 million

square meters. The area contains numerous linear vegetation elements of varying geometry, ranging from completely straight to curved, isolated or connected to other linear or nonlinear objects. Examples of vegetation and non-vegetation elements are planted forest patches, hedges, green lanes, isolated farms, ditches, a river, dykes and a road network (Figure 1). This landscape heterogeneity within a small area ensured that both the classification of vegetation and the segmentation of linear objects can be efficiently trained and tested.

3. Method

The workflow for the classification of linear vegetation objects (Figure 2) consisted of three main routines: 1. Feature extraction, 2. Vegetation classification, and 3. Linear object segmentation. In the first routine we computed features and added these as attributes to each point in the point cloud. In the second routine we trimmed unnecessary data points to improve computational efficiency and we applied a supervised classification machine learning algorithm [26] to classify the vegetation points in the point cloud using features based on echo, local geometric and local eigenvalue information of the point cloud (Table 1). In the third routine two preprocessing steps were applied, after which a rectangularity-based region growing algorithm was used to segment the classified vegetation points into rectangular objects, and an elongatedness criterion was applied to identify linear vegetation objects. The accuracy of the vegetation classification and linear object segmentation was tested against manually annotated datasets.
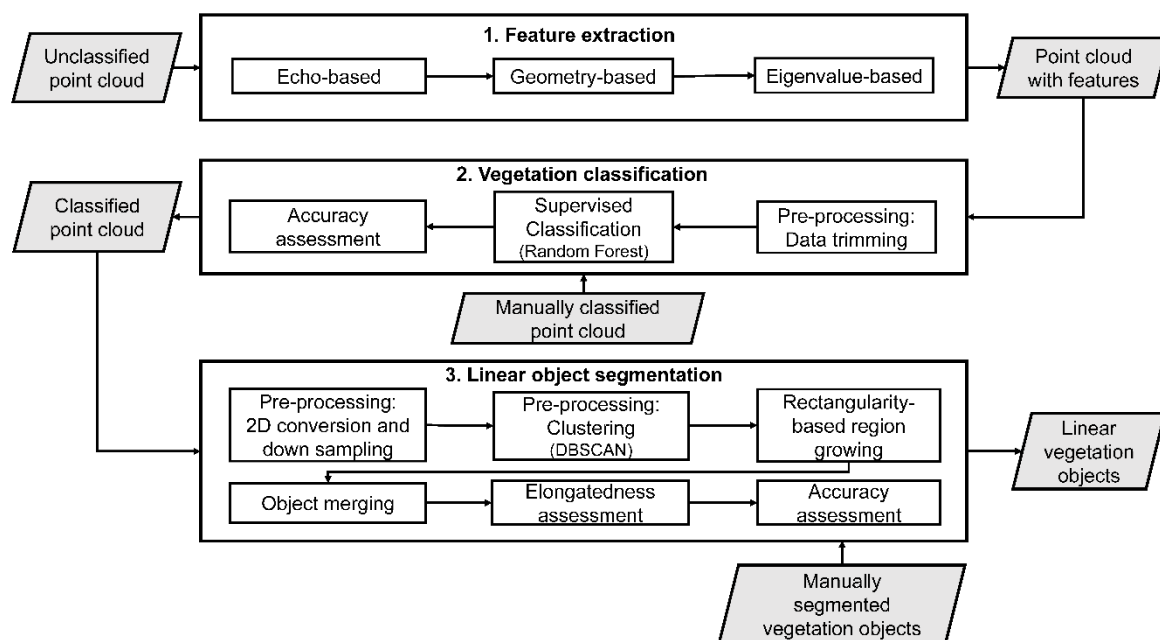
Figure 2. Workflow detailing three routines for the identification of linear vegetation objects, 1. Feature extraction 2. Vegetation classification, and 3. Linear object segmentation. The computational steps are represented as rectangles and datasets as grey parallelograms.

All data were analyzed using free and open source software. The scripting was performed in Python (3.6.5) using the NumPy (1.14.2) [27], SciPy (1.1.0) [28], pandas (0.22.0) [29], scikit-learn (0.19.1) [30], and CGAL (4.12) [31] libraries. PDAL (1.7.2) [32] was used for preprocessing and downsampling data. CloudCompare (v2.10alpha) [33] was used for visualizing the point cloud and for the manual classification. The full code is available via GitHub: https://github.com/chrislcs/linear-vegetation-elements/tree/master/Code.

3.1. Feature Extraction

The relevance of the various input features has been extensively studied, especially to separate urban objects from vegetation [34–37]. After reviewing the relevant literature, we selected and computed fourteen features (Table 1), which were used for vegetation classification in the second routine of the workflow. These features represent both point-based and neighborhood-based features and reflect information from echo and local neighborhoods (geometric and eigenvalue based), respectively. The qualities of these features are considered to be efficient for discriminating vegetation objects from point clouds [34]. The features were computed for the total extent of the study area (Figure 1), and added as attributes to form a point cloud dataset with features.

Table 1. Overview of point-based and neighborhood-based features used in the vegetation classification. The point-based features are based only on return number information stored with each point (i.e., echo information) whereas neighborhood-based features are based on the local geometry and eigenvalue characteristics derived from the x, y and z coordinates of the point cloud.

Feature Group	Feature	Symbol	Formula	Reference
Point				
Echo	Number of returns	R_t	-	-
	Normalized return number	R_n	R/R_t	[35]
Neighborhood				
Geometric	Height difference	Δ_z	$\max\limits_{j:\mathcal{N}_i}\left(q_{Z_j}\right) - \min\limits_{j:\mathcal{N}_i}\left(q_{Z_j}\right)$	[23]
	Height standard deviation	σ_z	$\sqrt{\frac{1}{k}\sum\limits_{j=1}^{k}\left(q_{z_j} - \overline{q_z}\right)^2}$	[23]
	Local radius	r_1	$\max\limits_{j:\mathcal{N}_i}\left(\left|p_i - q_j\right|\right)$	[23]
	Local point density	D	$\frac{k}{\frac{4}{3}\pi r_{l_i}^3}$	[23]
Eigenvalue	Normal vector Z	N_z		[38]
	Linearity	L_λ	$\frac{\lambda_1 - \lambda_2}{\lambda_1}$	[39]
	Planarity	P_λ	$\frac{\lambda_2 - \lambda_3}{\lambda_1}$	[39]
	Scatter	S_λ	$\frac{\lambda_3}{\lambda_1}$	[39]
	Omnivariance	O_λ	$\sqrt[3]{\lambda_1 \lambda_2 \lambda_3}$	[39]
	Eigentropy	E_λ	$-\lambda_1 \ln(\lambda_1) - \lambda_2 \ln(\lambda_2) - \lambda_3 \ln(\lambda_3)$	[39]
	Sum of eigenvalues	$\sum\lambda$	$\lambda_1 + \lambda_2 + \lambda_3$	[36]
	Curvature	C_λ	$\frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}$	[38]

3.1.1. Point-Based Features

The point-based features represent information from each single point (Table 1). The point cloud $\mathcal{P}$ is a set of points $\{p_1, p_2, \ldots, p_n\}$ $\left(\in R^3\right)$, where each point p_i has x, y and z coordinates. In addition, an intensity value (I), a return number (R), and a number of returns (R_t) of the returned signal are stored. We used R_t as well as the normalized return number R_n as echo-based features (Table 1). The R_n highlights vegetation, since vegetation can be characterized by multiple returns [35]. Since the available LiDAR data were available without the information such as flying height, plane trajectory data and sensor-related parameters, required to do a radiometric correction of the intensity data, we omitted this feature for the classification [40].

3.1.2. Neighborhood-Based Features

We computed point-based local-neighborhood features. A neighborhood set N_i of points $\{q_1, q_2, \ldots, q_k\}$ was defined for each point p_i, where $q_1 = p_i$, by using the *k*-nearest neighbors method with $k = 10$ points. We used a spherical neighborhood search with a fixed number of points instead of a fixed radius to calculate the features in Table 1, because of the homogeneous point density across our study area [23,39]. In this way a *k* of 10 results in a neighborhood of 10 points, one of which is the focal point itself. Based on these neighborhoods we then computed four geometric features: Height difference, height standard deviation, local radius and local point density (Table 1).

In addition to the geometric features, we calculated eight eigenvalue-based features (Table 1), that describe the distribution of points of a neighborhood in space [34,41]. We used the local structure tensor to estimate the surface normal and to define surface variation [39]. The structure tensor describes the principal directions of the neighborhood of a point by determining the covariance matrix of the x, y and z coordinates of the set of neighborhood points and by computing and ranking the eigenvalues (λ_1, λ_2, λ_3, where $\lambda_1 > \lambda_2 > \lambda_3$) of this covariance matrix. Hence, the magnitude of the eigenvalues of the matrix describe the spread of points in the direction of the eigenvector. The eigenvector belonging to the third eigenvalue is equal to the normal vector ($\vec{N} = (N_x, N_y, N_z)$) [39]. The points are linearly distributed if the eigenvalue of the first principle direction is significantly larger than the other two ($\lambda_1 >> \lambda_2 \approx \lambda_3$), and planarly distributed if the eigenvalues of the first two principle directions are about equal and significantly larger than the third ($\lambda_1 \approx \lambda_2 >> \lambda_3$). The points are scattered in all directions if all eigenvalues are about equal ($\lambda_1 \approx \lambda_2 \approx \lambda_3$). These properties (linearity, planarity, and scatter), as well as some additional features (omnivariance, eigenentropy, sum of eigenvalues, and curvature), were quantified using the formulas in Table 1.

3.2. Vegetation Classification

In a first preprocessing step, we removed irrelevant points, i.e., those that do not belong to tall vegetation (shrubs and trees). Subsequently, we applied a supervised classification in which the two echo, four geometric, and eight eigenvalue features (Table 1) were used as input for the vegetation classification after which we calculated metrics to assess the accuracy.

3.2.1. Preprocessing

Points that did not belong to tall vegetation were removed from the dataset to facilitate efficient processing. This was done by trimming data points that belonged to either non-vegetation, bare soil or low-stature vegetation (including grasslands and agricultural fields). This simplified the identification of tall linear vegetation elements composed of shrubs and trees. The removed points were characterized by a locally planar neighborhood and were selected on the basis of the scatter feature (Table 1). Points with low scatter values ($S_\lambda < 0.03$) were removed. This threshold was very conservative, but substantially contributed to reduction of the point data size, while still preserving all points that characterize tall vegetation.

3.2.2. Supervised Classification

For vegetation classification, we used a random forest classifier because it provides a good trade-off between classification accuracy and computational efficiency [23,42]. The random forest algorithm creates a collection of decision trees, where each tree is based on a random subset of the training data [43]. Random forest parameters such as the maximum number of features, minimal samples per leaf, minimal samples per split, and the ratio between minority and majority samples were optimized using a grid search [44].

As a result of data trimming, the remaining point cloud contained a lot more *vegetation* points than *other* points. Since imbalanced training data can lead to undesirable classification results [45], we used a balanced random forest algorithm. In this algorithm, the subsets are created by taking a bootstrap sample from the minority class and a random sample from the majority class with a sample size similar to the minority class [46]. By employing enough trees all majority class data are eventually used, while still maintaining a balance between the two classes. The decision trees were created using a Classification and Regression Tree (CART) algorithm [47].

3.2.3. Accuracy Assessment

For the accuracy assessment of the vegetation classification, a manual annotation of the trimmed point cloud into *vegetation* (trees and shrubs) and *other* (buildings, ditches, railroad infrastructure)

classes was done using an interpretation of both the point cloud and the high-resolution aerial photos. This resulted in a ground truth dataset of 997085 *vegetation* points and 56170 *other* points. By taking into account the dataset imbalance, we used the receiver operating characteristic (ROC) curve [48], the Matthews correlation coefficient (MCC) [49] and the geometric mean [50] as accuracy metrics. These metrics properly evaluate the performance of a classifier, even when dealing with an imbalanced dataset [51–53]. The area under a ROC curve (AUC) is a measure for the performance of the classifier [48] (Bradley, 1997). To create a ROC curve, the true positive (TP) rate is plotted against the false positive (FP) rate at various decision thresholds. The MCC analyzes the correlation between the observed and the predicted data and is defined as:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{1}$$

where TN are the true negatives and FN the false negatives. The geometric mean is defined as:

$$\text{Geometric mean} = \sqrt{\frac{TP}{TP + FP} \times \frac{TN}{TN + FN}} \tag{2}$$

The MCC, AUC and the geometric mean were obtained using a 10-fold cross validation. This is done by splitting the data into 10 randomly mutually exclusive subsets and using a subset as testing data and a classifier trained on the remaining data [51].

3.3. Linear Object Segmentation

For the linear object segmentation, we applied two preprocessing steps, clustered the points, applied a region growing algorithm, merged nearby and aligned objects, evaluated their elongatedness, and finally assessed the accuracy (Figure 2).

3.3.1. Preprocessing

Since we use linearity as a purely two-dimensional property (Table 1), we first converted the point cloud to 2D by removing the z-coordinate of the vegetation points. In a second step, the data were spatially down-sampled to 1-meter distance between vegetation points using Poisson sampling. This step preserved precision, but substantially facilitated computational efficiency. After down-sampling, we clustered the remaining points using a density based spatial clustering of application with noise (DBSCAN) algorithm [54]. This algorithm generates clusters, based on point density, and removed outliers. The clustering reduced the amount of possible neighboring points and therefore decreased the processing time during the next region growing step.

3.3.2. Rectangularity-Based Region Growing

Region growing is an accepted way of decomposing point clouds [55,56] into homogeneous objects. Normally, regions are grown based on similarity of the attributes of the points. Here, we used an alternative way and grew regions based on a rectangularity constraint (Figure 3). The rectangularity of an object is described as the ratio between the area of an object and the area of its minimum oriented bounding box (MOBB) [57]. The MOBB is computed using rotating calipers [58]. First a convex hull is constructed using the QuickHull algorithm [59] and then the MOBB can be found by rotating the system by the angles which are made by the edges of the convex hull and the x-axis. The algorithm then checks the bounding rectangles of each rotation because the minimum oriented bounding box has a side collinear with one of the edges of the convex hull [60]. The area of the object can be calculated by computing the concave hull of the set of points belonging to the object. This hull is constructed by computing an alpha shape of the set of points [61]. This shape is created by computing a Delaunay triangulation of the points [62] and by removing the triangles with a circumradius higher than $1/\alpha$, where α is a parameter which influences the number of triangles removed from the triangulation and

thus the shape and area of the alpha shape. Higher alphas lead to more complex shapes, while lower ones lead to smoother shapes.
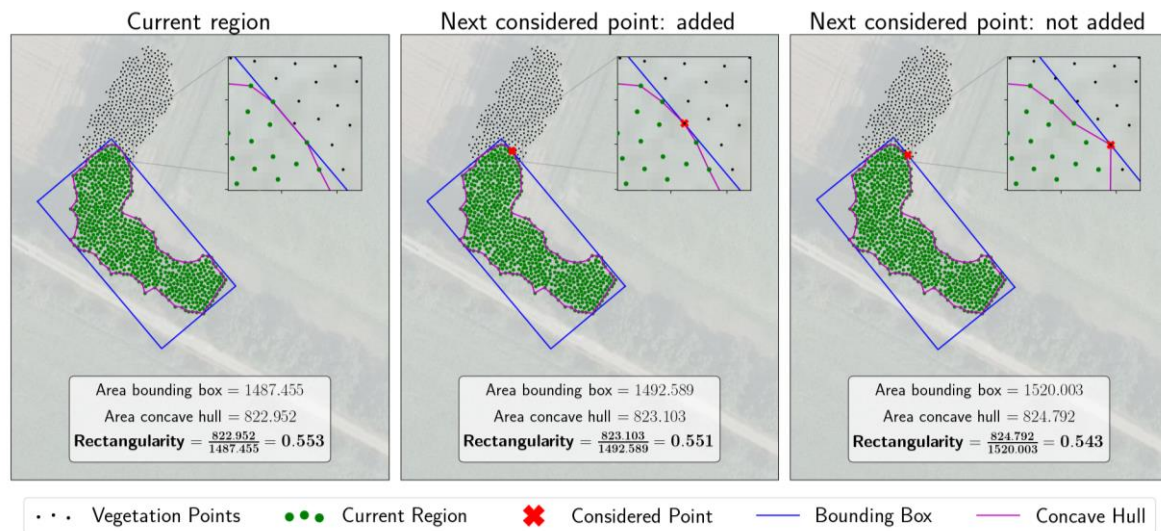
Figure 3. Visualization of the region growing process based on rectangularity in three steps. Step a shows how a current region (in green) is grown in which the ratio between the concave hull and bounding box is above the threshold of 0.55. Step b considers a next point (in red) to be included to the region, which is also accepted. Step c considers a next point, which does not meet the rectangularity constraint of 0.55 and is not added to that region.

For each cluster, points with the minimum x-coordinate and its ten closest neighbors were used as the starting region. Subsequently, the eight nearest neighbors of each point were considered for growth (Figure 3). Points were added as long as the region's rectangularity did not drop below a threshold of 0.55. We determined this threshold value on a subset of the data which showed that the best performance of the algorithm was between 0.5 and 0.6, with only marginal differences in performance. After a region is grown, the growing procedure was repeated for the next region until the entire cluster is segmented into rectangular regions.

3.3.3. Object Merging

After region growing, some objects might be isolated or very small, for example as the result of minor curves in the linear elements or small interruptions in vegetation. These objects were merged if they were within 5 m of another object, faced a similar compass direction, and were aligned. The compass direction was determined by computing the angle between the long sides of the minimum bounding box and the x-axis. The alignment of objects was checked by comparing the angle of the line between the two center points with the compass directions of the objects. Once merged the lengths of the objects were combined and the maximum of the two object widths taken as the new width.

3.3.4. Elongatedness

All objects were assessed for linearity by evaluating their elongatedness, which is defined as the ratio between its length and width [63]. We used a minimum elongatedness of 1.5 and a maximum width of 60 meters, which was found to be realistic values in this rural landscape. These thresholds produced consistent linear vegetation elements while large forest patches were excluded.

3.3.5. Accuracy Assessment

The accuracy of the delineated linear objects was assessed by calculating the user's, producer's and overall accuracy, as well as the harmonic mean of the precision and recall (F1) and MCC scores [64].

We manually prepared a dataset of linear and nonlinear vegetation objects by means of a field visit in combination with interpretation of high-resolution air-photos. Based on the difference between the automated and manually constructed data a map and confusion matrix detailing the accuracy were created.

4. Results

4.1. Vegetation Classification

The vegetation classification resulted in a map with three classes (Figure 4). A class with points that were removed during the preprocessing (grasslands, agricultural fields, bare soil, water bodies), a class with points that were identified as infrastructure (e.g., building edges, ditches, and railroad infrastructure) and the relevant points that represent tall vegetation. The accuracy assessment of the *vegetation* and *other* classes have a producer's accuracy of 0.98 for *vegetation* and of 0.85 for the *other* classes (Table 2). The AUC of 0.98 showed that the *vegetation* and *other* class were well separated, and this was also supported by an MCC value of 0.76 (indicating of a positive correlation between the predicted and observed classes) and the geometric mean of 0.90.
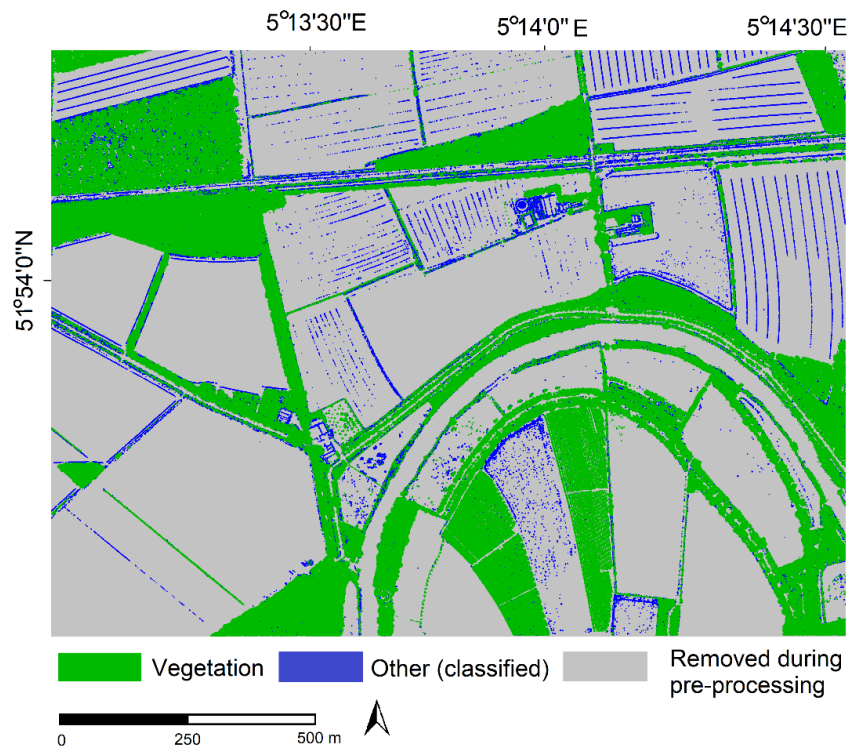

Figure 4. Results of the supervised classification. The green class represents the tall vegetation. The two 'other' classes contained data points that were classified as grasslands, agricultural fields, bare soil and water bodies (grey class), or as infrastructure and ditches (blue class).

Table 2. Confusion matrix of the predicted against the actual classes.

		Actual		
		Vegetation	Other	User's Accuracy
Predicted	Vegetation	974,177	8171	0.99
	Other	22,908	47,999	0.68
Producer's accuracy		0.98	0.85	Overall accuracy: 0.97

4.2. Linear Object Segmentation

The tall vegetation (see Figure 4) was segmented using a rectangularity-based region growing algorithm. The resulting objects were assessed for linearity and the results were compared with the manual segmentation (Figure 5). Areas that were correctly classified as linear vegetation elements (true positives) and areas that were correctly classified as nonlinear vegetation objects (true negatives) could be identified (Figure 5). Some nonlinear areas were misclassified as linear (false positives), and some linear regions were classified as nonlinear (false negatives). However, the confusion matrix (Table 3) shows an overall good accuracy of 0.90, and an F1-score of 0.82 and an MCC of 0.76. Hence, the majority of linear vegetation elements was successfully segmented, which is supported by user's and producer's accuracies of 0.85 and 0.80, respectively (Table 3). Nonlinear objects were successfully separated, with user's and producer's accuracies of 0.92 and 0.94, respectively.
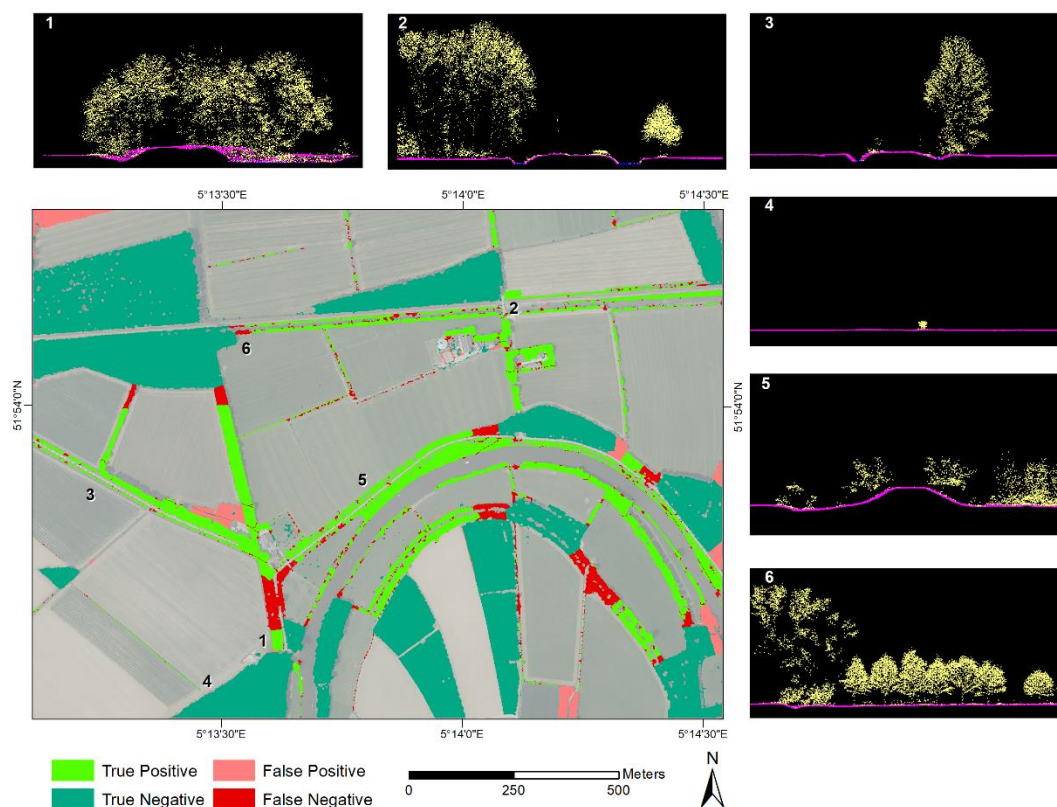
Figure 5. Results of vegetation classification. Green areas have been correctly classified as linear vegetation (light green) and non-linear vegetation (dark green). Red areas have been wrongly classified as linear vegetation (light red), while dark red areas have been wrongly classified as non-linear vegetation. Cross plots 1–6 illustrate the variation in linear vegetation elements (in yellow) and terrain points (in purple), as visible from the LiDAR point cloud.

Table 3. Confusion matrix of the automatically segmented against the manually annotated set of linear and non-linear vegetation objects (m^2).

		Actual		
		Linear	**Non-Linear**	**Producer's Accuracy**
Predicted	Linear	116,483.76	20,201.53	0.85
	Non-linear	28,385.56	336,754.65	0.92
User's accuracy		0.80	0.94	Overall accuracy: 0.90

5. Discussion

We designed a method to delineate linear vegetation elements in a rural agricultural area from airborne laser scanning point clouds. Our intention was to apply already existing codes (which mainly stem from the classification of complex city scenes) in a new context for the extraction of linear vegetation elements in rural landscapes. Our findings show that LiDAR features, calculated from a pre-processed point cloud, are useful to separate tall vegetation from low-stature vegetation and non-vegetation. Moreover, the subsequent analysis enabled the extraction of linear vegetation elements by applying a rectangularity-based growing algorithm. This application is novel in this context and allows mapping and monitoring of hedges, tree rows, and other linear vegetation elements with high resolution and unprecedented detail. We see great potential to apply this methodology to derive inventories of linear vegetation elements in rural and agricultural landscapes for subsequent use in ecological applications, conservation management and regional planning.

5.1. Feature Extraction

To our knowledge, LiDAR-based features have not been applied in this way to identify linear vegetation objects. Most features have been used to quantify 3D-forest structure or to separate vegetation and infrastructure in urban environments [65,66]. This is supported by similar studies [35,36] that emphasized the importance of feature selection for separating vegetation and infrastructure. We used the 14 features which were reported to be useful for filtering vegetation from other objects in the landscape [23,34–37]. In the preprocessing phases, we used the planarity and scatter features to trim unnecessary non-vegetation data before the actual vegetation classification, since our goal was to only classify tall linear vegetation elements. Approximately 22% of the point cloud, corresponding mainly to smooth and planar areas such as bare soil, grassland and water bodies, was removed, which made the processing of the remaining data much more efficient. Some features, such as the number of returns and point density have clear relations with the vegetation structure, while others, such as omnivariance and the sum of the eigenvalues, are more difficult to interpret in terms of linear vegetation structure. Our results suggest that these LiDAR-based features can successfully be applied to separate (linear) vegetation from other classes.

5.2. Vegetation Classification

Trimming the data proved an efficient step to reduce the computation time needed to classify the vegetation. This trimming preprocessing step forced the dataset to be imbalanced, which was overcome by using a balanced random forest classification algorithm [46] and selecting suitable accuracy metrics to evaluate performance [52,53]. When analyzing the classification statistics it is important to take the effect of the trimming step into consideration. The removed points (Figure 4) do clearly not belong to the tall vegetation class. Consequently, the remaining *other* points share some mutual similarities with the tall vegetation points and are therefore more difficult to classify correctly. If the trimmed points would be included in the accuracy assessment, higher accuracy values would be reached, but less insights would be provided for the classification process. Nevertheless, the majority of points were correctly classified (Table 2).

5.3. Linear Object Segmentation

Our workflow to identify linear vegetation elements has proven successful in a small test area of a typical rural area in the Netherlands. Most of the linear vegetation elements were successfully extracted (Figure 5). In Figure 5 near cross section 1, multiple parallel linear vegetation elements occur, which were incorrectly classified. The cross section near 5 shows very small errors, which are caused by small interruptions within the linear vegetation elements. In cross section 6, linear vegetation elements are adjacent to forest patches, which causes some small errors as well.

These observations leave some space for improving the workflow, and we suggest to apply and test the method in some other rural areas, which show different types and spatial configurations of linear vegetation elements. However, seen in the context of this rural landscape, our method is already accurate, which opens up possibilities to further refine the method and to apply it to other study areas or broader spatial extents. For example, pruning activities, a common practice in the Netherlands, may influence the accuracy of the classifications, because they are difficult to detect in a point cloud dataset. Multi-temporal point cloud change detection could overcome such problems.

For the extraction of linear vegetation elements on national or continental scales, our method would need to be upscaled. This requires us to optimize the computational efficiency and speed. For that, some data requirements and methodological steps are required. LiDAR data need to be available at large spatial extents (e.g., across single or multiple countries) and, due to variability of e.g., point cloud density and collection methods across different countries, preprocessing steps need to be carefully tested. For example, minimum point density for the calculation of features should be tested, and further testing is needed to determine the optimal number of points in a neighborhood [23] for applying our methodology across different point densities. Training data in different environmental areas need to be collected to ensure sufficiently high classification accuracies.

Another issue is that the computation time needs to be reduced by speeding up all individual workflow steps. For example, the number of points could be thinned without affecting the quality needed to extract the features which are necessary for classifying the point cloud. Another possibility would be to use parallel or cloud computing solutions which have been introduced in ongoing LiDAR projects [18]. If such steps are taken, the availability of detailed locational information on linear vegetation elements at national or cross-national scales would advance the analysis, monitoring and conservation of a wide range of species such as insects, birds etc., that depend on linear vegetation elements for shelter, food and survival in a fragmented, rural landscape. Initiatives to develop software for upscaling LiDAR vegetation metrics to national and European scale using efficient software and cloud computing facilities are in progress [18,67].

6. Conclusions

At present, LiDAR datasets differ in quality, content, and accessibility within and across countries. Therefore, identifying robust and scalable LiDAR features for object identification should help to overcome these inconsistencies. The quality of the AHN3 dataset of the Netherlands is high and allows us to correctly identify linear vegetation objects with our method. In addition, multi-temporal LiDAR datasets could be analyzed to monitor changes in the spatial distribution and configuration of linear vegetation objects.

The ecological value of providing such large datasets of linear vegetation objects lies in the broad extent and fine-scale locational details, which is a powerful quality that can be used in the (3D) characterization of ecosystem structure. Existing ecosystem and biodiversity assessment projects, such as the MAES (Mapping and Assessment of Ecosystems and their Services) project [68], the SEBI (Streamlining European Biodiversity Indicators) project [69], and the high nature value farmland assessment [70] on a European scale and assessments of Planbureau voor de Leefomgeving (PBL) on a national level [71], could benefit from the new details.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Turner, M.G. Landscape ecology: The effect of pattern on process. *Annu. Rev. Ecol. Syst.* **1989**, *20*, 171–197. [CrossRef]
2. Marquer, L.; Gaillard, M.-J.; Sugita, S.; Poska, A.; Trondman, A.-K.; Mazier, F.; Nielsen, A.B.; Fyfe, R.M.; Jöhnsson, A.M.; Smith, B.; et al. Quantifying the effects of land use and climate on Holocene vegetation in Europe. *Quat. Sci. Rev.* **2017**, *171*, 20–37. [CrossRef]
3. Bailly, J.S.; Lagacherie, P.; Millier, C.; Puech, C.; Kosuth, P. Agrarian landscapes linear features detection from lidar: Application to artificial drainage networks. *Int. J. Remote Sens.* **2008**, *29*, 3489–3508. [CrossRef]
4. Meyer, B.C.; Wolf, T.; Grabaum, R. A multifunctional assessment method for compromise optimisation of linear landscape elements. *Ecol. Indic.* **2012**, *22*, 53–63. [CrossRef]
5. Van der Zanden, E.H.; Verburg, P.H.; Mücher, C.A. Modelling the spatial distribution of linear landscape elements in Europe. *Ecol. Indic.* **2013**, *27*, 125–136. [CrossRef]
6. Aguirre-Gutiérrez, J.; Kissling, W.D.; Carvalheiro, L.G.; WallisDeVries, M.F.; Franzén, M.; Biesmeijer, J.C. Functional traits help to explain half-century long shifts in pollinator distributions. *Sci. Rep.* **2016**, *6*. [CrossRef]
7. Spellerberg, I.F.; Sawyer, J.W. *An Introduction to Applied Biogeography*; Cambridge University Press: Cambridge, UK, 1999. [CrossRef]
8. Croxton, P.; Hann, J.; Greatorex-Davies, J.; Sparks, T. Linear hotspots? The floral and butterfly diversity of green lanes. *Biol. Conserv.* **2005**, *121*, 579–584. [CrossRef]
9. Burel, F. Hedgerows and their role in agricultural landscapes. *Crit. Rev. Plant Sci.* **1996**, *15*, 169–190. [CrossRef]
10. Jongman, R.G.H. Landscape linkages and biodiversity in Europe. In *The New Dimensions of the European Landscape*; Jongman, R.G.H., Ed.; Springer: Dordrecht, The Netherlands, 1996; pp. 179–189.
11. Gobster, P.H.; Nassauer, J.I.; Daniel, T.C.; Fry, G. The shared landscape: What does aesthetics have to do with ecology? *Landsc. Ecol.* **2007**, *22*, 959–972. [CrossRef]
12. Boutin, C.; Jobin, B.; Bélanger, L.; Baril, A.; Freemark, K. Hedgerows in the Farming Landscapes of Canada. Hedgerows of the World: Their Ecological Functions in Different Landscapes. Available online: https://www.researchgate.net/publication/264670164_Hedgerows_in_the_farming_landscapes_of_Canada (accessed on 31 January 2019).
13. Stoate, C.; Boatman, N.; Borralho, R.; Carvalho, C.R.; De Snoo, G.; Eden, P. Ecological impacts of arable intensification in Europe. *J. Environ. Manag.* **2001**, *63*, 337–365. [CrossRef]
14. Aksoy, S.; Akcay, H.G.; Wassenaar, T. Automatic mapping of linear woody vegetation features in agricultural landscapes using very high resolution imagery. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 511–522. [CrossRef]
15. Thornton, M.W.; Atkinson, P.M.; Holland, D. Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping. *Int. J. Remote Sens.* **2006**, *27*, 473–491. [CrossRef]
16. Vannier, C.; Hubert-Moy, L. Multiscale comparison of remote-sensing data for linear woody vegetation mapping. *Int. J. Remote Sens.* **2014**, *35*, 7376–7399. [CrossRef]
17. Tansey, K.; Chambers, I.; Anstee, A.; Denniss, A.; Lamb, A. Object-oriented classification of very high resolution airborne imagery for the extraction of hedgerows and field margin cover in agricultural areas. *Appl. Geogr.* **2009**, *29*, 145–157. [CrossRef]
18. Kissling, W.D.; Seijmonsbergen, A.C.; Foppen, R.; Bouten, W. eEcolidar, eScience infrastructure for ecological applications of LiDAR point clouds: Reconstructing the 3d ecosystem structure for animals at regional to continental scales. *Res. Ideas Outcomes* **2017**, *3*, e14939. [CrossRef]
19. Lim, K.; Treitz, P.; Wulder, M.; St-Onge, B.; Flood, M. Lidar remote sensing of forest structure. *Prog. Phys. Geogr.* **2003**, *27*, 88–106. [CrossRef]
20. Lefsky, M.A.; Cohen, W.B.; Parker, G.G.; Harding, D.J. Lidar remote sensing for ecosystem studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *AIBS Bull.* **2002**, *52*, 19–30.

21. Eitel, J.U.; Höfle, B.; Vierling, L.A.; Abellán, A.; Asner, G.P.; Deems, J.S.; Glennie, C.L.; Joerg, P.C.; LeWinter, A.L.; Magney, T.S.; et al. Beyond 3-d: The new spectrum of lidar applications for earth and ecological sciences. *Remote Sens. Environ.* **2016**, *186*, 372–392. [CrossRef]

22. Song, J.-H.; Han, S.-H.; Yu, K.; Kim, Y.-I. Assessing the possibility of land-cover classification using lidar intensity data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2002**, *34*, 259–262.

23. Weinmann, M.; Jutzi, B.; Hinz, S.; Mallet, C. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS J. Photogramm. Remote Sens.* **2015**, *105*, 286–304. [CrossRef]

24. PDOK. Available online: https://www.pdok.nl/viewer/ (accessed on 12 July 2018).

25. AHN Inwinjaren AHN2 & AHN3. Available online: http://www.ahn.nl/common-nlm/inwinjaren-ahn2--ahn3.html (accessed on 12 July 2018).

26. Yan, W.Y.; Shaker, A.; El-Ashmawy, N. Urban land cover classification using airborne LiDAR data: A review. *Remote Sens. Environ.* **2015**, *158*, 295–310. [CrossRef]

27. Van der Walt, S.; Colbert, S.C.; Varoquaux, G. The numpy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30. [CrossRef]

28. SciPy: Open source scientific tools for Python. Available online: http://www.scipy.org/ (accessed on 12 July 2018).

29. McKinney, W. Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 51–56.

30. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

31. CGAL Project. *CGAL User and Reference Manual*, 4.13th ed.; CGAL Editorial Board, 2018; Available online: https://doc.cgal.org/latest/Manual/packages.html (acceseed on 1 February 2019).

32. PDAL. Available online: https://pdal.io/ (accessed on 12 July 2018).

33. Cloud Compare. Available online: http://www.cloudcompare.org/ (accessed on 12 July 2018).

34. Chehata, N.; Guo, L.; Mallet, C. Airborne lidar feature selection for urban classification using random forests. *Remote Sens. Spat. Inf. Sci.* **2009**, *38*, 207–2012.

35. Guo, L.; Chehata, N.; Mallet, C.; Boukir, S. Relevance of airborne LiDAR and multispectral image data for urban scene classification using random forests. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 56–66. [CrossRef]

36. Mallet, C.; Bretar, F.; Roux, M.; Soergel, U.; Heipke, C. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, S71–S84. [CrossRef]

37. Niemeyer, J.; Rottensteiner, F.; Soergel, U. Contextual classification of LiDAR data and building object detection in urban areas. *SPRS J. Photogramm. Remote Sens.* **2018**, *87*, 152–165. [CrossRef]

38. Pauly, M.; Gross, M.; Kobbelt, L.P. Efficient simplification of point-sampled surfaces. In Proceedings of the Conference on Visualization, IEEE Visualization, Boston, MA, USA, 27 October–1 November 2002; pp. 163–170.

39. West, K.F.; Webb, B.N.; Lersch, J.R.; Pothier, S.; Triscari, J.M.; Iverson, A.E. Context-driven automated target detection in 3d data. In Proceedings of the SPIE 5426, Automatic Target Recognition XIX, Orlando, FL, USA, 21 September 2004; pp. 133–143. [CrossRef]

40. Kashani, A.G.; Olsen, M.J.; Parrish, C.E.; Wilson, N. A review of LiDAR radiometric processing: From ad hoc intensity correction to rigorous radiometric calibration. *Sensors* **2015**, *15*, 28099–28128. [CrossRef]

41. Hoppe, H.; DeRose, T.; Duchampt, T.; McDonald, J.; Stuetzle, W. Surface reconstruction from unorganized points. *Comp. Graph.* **1992**, *26*, 2. [CrossRef]

42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

43. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Int.* **1998**, *20*, 832–844.

44. Hsu, C.; Chang, C.; Lin, C. A Practical Guide to Support Vector Classification. Available online: https://www.csie.ntu.edu.tw/~{}cjlin/papers/guide/guide.pdf (accessed on 12 July 2018).

45. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [CrossRef]

46. Chen, C.; Liaw, A.; Breiman, L. *Using Random Forest to Learn Imbalanced Data*; Technik Report 666; Department of Statistics, UC Berkeley: Berkeley, CA, USA, 2004.

47. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.

48. Bradley, A.P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recogn.* **1997**, *30*, 1145–1159. [CrossRef]

49. Matthews, B.W. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochim. Biophys. Acta (BBA) Protein Struct.* **1975**, *405*, 442–451. [CrossRef]

50. Kubat, M.; Holte, R.C.; Matwin, S. Machine learning for the detection of oil spills in satellite radar images. *Mach. Learn.* **1998**, *30*, 195–215. [CrossRef]

51. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; pp. 1137–1145.

52. Sun, Y.; Wong, A.K.; Kamel, M.S. Classification of imbalanced data: A review. *Int. J. Pattern Recognit. Artif. Intell.* **2009**, *23*, 687–719. [CrossRef]

53. López, V.; Fernandez, A.; García, S.; Palade, V.; Herrera, F. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Inf. Sci.* **2013**, *250*, 113–141. [CrossRef]

54. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Kdd-96 Second International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; pp. 226–231.

55. Rabbani, T.; Van Den Heuvel, F.; Vosselmann, G. Segmentation of point clouds using smoothness constraint. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2006**, *36*, 248–253.

56. Vosselman, G. Point cloud segmentation for urban scene classification. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *1*, 257–262. [CrossRef]

57. Rosin, P.L. Measuring rectangularity. *Mach. Vis. Appl.* **1999**, *11*, 191–196. [CrossRef]

58. Toussaint, G.T. Solving geometric problems with the rotating calipers. In Proceedings of the IEEE Melecon'83, Athens, Greece, 24–26 May 1983; pp. 1–8.

59. Preparata, F.P.; Shamos, M. *Computational Geometry: An Introduction*; Springer-Verlag: New York, NY, USA, 1985; ISBN 978-0-387-96131-6.

60. Freeman, H.; Shapira, R. Determining the minimum-area encasing rectangle for an arbitrary closed curve. *Commun. ACM* **1975**, *18*, 409–413. [CrossRef]

61. Edelsbrunner, H.; Kirkpatrick, D.; Seidel, R. On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory* **1983**, *29*, 551–559. [CrossRef]

62. Delaunay, B. Sur la sphere vide. Izv. Akad. Nauk SSSR. *Otdelenie Matematicheskii i Estestvennyka Nauk 7* **1934**, *1–2*, 793–800.

63. Nagao, M.; Matsuyama, T. *A Structural Analysis of Complex Aerial Photographs*; Springer-Verlag: New York, NY, USA, 1980; ISBN13 9781461582960.

64. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2008; ISBN 9781420055139.

65. Eysn, L.; Hollaus, M.; Schadauer, K.; Pfeifer, N. Forest Delineation Based on Airborne LiDAR Data. *Remote Sens.* **2012**, *4*, 762–783. [CrossRef]

66. Yang, H.; Chen, W.; Qian, T.; Shen, D.; Wang, J. The Extraction of Vegetation Points from LiDAR Using 3D Fractal Dimension Analyses. *Remote Sens.* **2015**, *7*, 10815–10831. [CrossRef]

67. Pfeiffer, N.; Mandlburger, G.; Otepka, J.; Karel, W. OPALS—A framework for Airborne Laser Scanning data analysis. *Comput. Environ. Urban Syst.* **2014**, *45*, 125–136. [CrossRef]

68. Maes, J.; Teller, A.; Erhard, M.; Liquete, C.; Braat, L.; Berry, P.; Egoh, B.; Puydarrieux, P.; Fiorina, C.; Santos, F.; et al. *Mapping and Assessment of Ecosystems and Their Services*; Tech. Rep. EUR 27143 EN. Joint Research Center—Institute for Environment and Sustainability, 2013. Available online: http://ec.europa.eu/environment/nature/knowledge/ecosystem_assessment/pdf/102.pdf (accessed on 10 December 2018).

69. Biała, K.; Condé, S.; Delbaere, B.; Jones-Walters, L.; Torre-Marín, A. *Streamlining European Biodiversity Indicators 2020*; Tech. Rep. 11/2012. European Environment Agency, 2012. Available online: https: //www.eea.europa.eu/publications/streamlining-european-biodiversity-indicators-2020 (accessed on 10 December 2018).

70. Paracchini, M.L.; Petersen, J.-E.; Hoogeveen, Y.; Bamps, C.; Burfield, I.; van Swaay, C. *High Nature Value Farmland in Europe*; Tech. Rep. EUR 23480 EN. Joint Research Center—Institute for Environment and Sustainability, 2008. Available online: http://agrienv.jrc.ec.europa.eu/publications/pdfs/HNV_Final_ Report.pdf (accessed on 10 December 2018).

71. Bouwma, I.; Sanders, M.; Op Akkerhuis, G.J.; Onno Knol, J.V.; de Wit, B.; Wiertz, J.; van Hinsber, A. Biodiversiteit Bekeken: Hoe Evalueert en Verkent Het PBL het Natuurbeleid? 2014. Available online: https: //www.pbl.nl/sites/default/files/cms/publicaties/PBL_2014_Biodiversiteit%20bekeken_924.pdf (accessed on 10 December 2018).