Title: A system for the estimation of single tree stem diameter and volume using multireturn LIDAR data

This paper appears in: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING,

Date of Publication: July 2011

Author(s): Michele Dalponte, Lorenzo Bruzzone, and Damiano Gianell

Volume: 49, Issue: 7

Page(s): 2479 - 2490

DOI: 10.1109/TGRS.2011.2107744

A system for the estimation of single tree stem diameter and volume using multireturn LIDAR data

Michele Dalponte[a,b], Lorenzo Bruzzone[a], and Damiano Gianelle[b]

[a] *Dept. of Information Engineering and Computer Science, University of Trento, Via Sommarive, 14, I-38123 Povo, Trento, Italy (E-mails: lorenzo.bruzzone@ing.unitn.it).*

[b] *IASMA Research and Innovation Centre, Fondazione E. Mach, Environmental and Natural Resources Area, Via E. Mach, 1, I-38010 S. Michele all'Adige (TN), Italy (E-mail: michele.dalponte@iasma.it; damiano.gianelle@iasma.it).*

Abstract

Forest inventories are important tools for the management of forests. In this context the estimation of the tree stem volume is a key issue. In this paper we present a system for the estimation of forest stem diameter and volume at individual tree level from multireturn LIDAR data. The proposed system is made up of a preprocessing module, a LIDAR segmentation algorithm (aimed at retrieving tree crowns), a variable extraction and selection procedure and an estimation module based on Support Vector Regression (which is compared with a multiple linear regression technique). The variables derived from LIDAR data are computed from both the intensity and elevation channels of all available returns. Three different methods of variable selection are analyzed, and the sets of variables selected are used in the estimation phase. The stem volume is estimated with two methods: i) direct estimation from the LIDAR variables; and ii) combination of diameters and heights estimated from LIDAR variables with the species information derived from a classification map according to standard height/diameter relationships. Experimental results show that the system proposed is effective and provides high accuracies in both the stem volume and diameter estimation. Moreover, the presented study provides useful indications on the effectiveness of SVR with LIDAR in forestry problems.

Keywords: stem volume estimation, tree diameter estimation, multireturn LIDAR, Support Vector Regression, remote sensing, forestry

I. INTRODUCTION

Light Detection and Ranging (LIDAR) is a powerful and effective instrument for the study of forest areas. It provides information on the vertical structure of forests and directly measures the tree height. From these data it is possible to derive many forest and tree parameters useful for forest management (e.g., [1]-[19]). In the literature, two main approaches on the use of LIDAR data in the study of forest areas are reported: stand and single tree approaches. In the stand approach, forest parameters (e.g., stem volume, basal area, etc.) are estimated on groups of trees (usually on circular areas of a given radius), whereas the single tree approach estimates parameters of each tree (e.g., stem height, diameter at breast height (DBH), stem volume). In this paper we focus our attention on the single tree approach. Many papers have been published in the literature on this topic and it is possible to divide them into four groups.

The first group comprises papers focused on algorithms for single tree detection (e.g., [5],[6]). Among them we recall Palenichka *et al.* in [5] that presented an automatic tree detection technique based on a multiscale isotropic matched filtering and multireturn LIDAR data. The proposed algorithm exhibited a correct detection rate of more than 80% on the analyzed dataset. Another recent study is the one of Koch *et al.* in [6]. In this paper the authors proposed a technique to delineate automatically the single tree crowns in rasterized first return LIDAR data. The segmentation algorithm is based on some assumptions on the shape of trees, and on tree tops detection and searching vector technique for the delineation of crowns.

In the second group we have papers that address the estimation of single tree parameters (e.g., stem diameter, stem volume) starting from a model-based approach based on LIDAR-derived tree height and crown diameter. In the literature many papers are present that adopt this approach (e.g., [7]-[12]). As an example, in [7] Zhao *et al.* developed methods for scale-invariant estimation of forest biomass using first and last return LIDAR data. The authors identified the single trees using the algorithm described in [8] and then applied two models based on LIDAR-derived tree height, crown width and crown base height. The final estimation results show an R^2 of more than 0.9. Recently, Heurich in [10] proposed a new method for delineating individual tree crowns based on a fully automated recognition methodology and first return LIDAR data. Once the tree crowns have been modeled the author estimated the tree stem diameter and volume of the recognized trees through a linear regression based on the above cited variables (e.g., LIDAR-derived height, LIDAR-derived crown radius). The experimental results showed an RMSE between 11.9 and 15.2% in the estimation of DBH and between 27 and 35% in the estimation of stem volume.

In the above-mentioned papers (i.e., [7]-[12]) stem parameters are estimated starting from two specific LIDAR derived variables (i.e., stem height and crown diameter). A third group of papers addresses the estimation of stem parameters starting from a larger set of LIDAR-derived variables and including in the processing architecture also a variable selection phase (e.g., [13]-[16]). It is the case of Vauhkonen *et al.* in [13], who estimated DBH, height and volume using first return LIDAR data. They considered a large number of predictors, and after a selection of the most informative ones, estimated the stem parameters. On a validation set of 1249 trees they obtained an RMSE% of 13% and 31% for the estimation of the stem diameter and volume, respectively. Similarly, Chen *et al.* in [14] estimated basal area and stem volume using a set of variables derived from first and last return LIDAR pulses. The authors segmented the single trees using an algorithm based on tree tops detection and watershed segmentation. In the estimation of the basal area they obtained an R^2 of 0.79 on the best model, while for the stem volume the R^2 was equal to 0.80.

Finally, the fourth group of papers refers to techniques that automatically retrieve stem attributes [17]-[19]. Among the studies that face this problem, we recall Bucksch *et al.* in [17] who present an automatic technique for the estimation of tree stem diameters. This technique is based on a 3-step approach and it applies a skeletonization algorithm. The estimation results obtained on a set of 49 trees showed a deviation from the ground measured diameters of about 10%. On the same topic, Kato *et al.* in [18] proposed a novel method called “wrapped surface reconstruction” for the definition of tree crowns characteristics. In greater detail, they estimated the tree height, the crown width, the live crown base, the height of the lowest branches, and the crown volume. They validated the estimations on a set of 55 trees obtaining R^2 values that range between 0.51 and 0.96.

From this brief analysis of the existing literature it is clear that many papers exist on stem attributes estimation. Despite that only few papers explore the combination of variables extracted from multiple returns data for the estimation of tree parameters. Moreover from our analysis, no study addressed the problem of stem diameter and volume estimation with multireturn LIDAR data by using advanced machine learning techniques, like Support Vector Regression (SVR). This regression technique is increasingly used in remote sensing applications [20]-[23], and we expect that it can be effective also in forestry as it exhibits some properties which are very important for the estimation of forest parameters, i.e., i) it has a good generalization capability; ii) it is effective in problems characterized by a small number of training samples (ill-posed problems); iii) the cost function used in the learning of the estimator is convex (i.e. it is always possible to derive the best solution according to quadratic programming methods); and iv) the kernel-SVR can model also strongly non linear relationships between the predictors and the target. This is very important to better exploit the information contained in LIDAR data, especially when returns different from the first one are considered.

Thus, the goals of this paper are: i) to propose a system that, starting from a set of variables extracted from multireturn LIDAR data, exploits the SVR technique to estimate both tree diameter and volume; ii) to compare different variable selection techniques; iii) to compare the SVR technique (with two different kernel functions) with the standard multiple linear estimator; and iv) to compare different approaches to stem volume estimation. To achieve these goals, we adopt a segmentation-based method that: i) identifies single tree crowns from LIDAR data, ii) extracts and selects the most effective variables, and iii) estimates tree diameter and volume for each crown.

The rest of the paper is organized as follows. In Section II materials and methods are presented. At first the data set used is described (Section II.A), and then Section II.B presents the architecture of the system adopted and the data preprocessing operations. The segmentation algorithm is illustrated in section II.C, while section II.D and II.E present the techniques of variables extraction and selection. The SVR estimator technique is described in section II.F. Section III reports experimental results. Finally, conclusions are drawn in section IV.

II. MATERIAL AND METHODS

A. Data set description

The study area is a forest site in the Italian Alps located at Lavarone (near the city of Trento) in the Trento Province (Italy) (see Figure 1 for the Digital Canopy Height Model of the area). The central point of the area has the following coordinates: 45° 57’ 30.09” N, 11° 16’ 25.17” E. The topography of this area is complex: it includes hill sides having different inclinations with an altitude that ranges from 1200 to 1600 meters on the sea level. The area has a size of approximately 495 ha. This site is a typical Alpine forest with three main species (i.e., Norway Spruce, Silver Fir and European Beech) and less common species (i.e., European Larch and Scots Pine).

Figure 1. Digital Canopy Model of the investigated area.

TABLE I. DISTRIBUTION OF REFERENCE POINTS (TREES) IN THE TRAINING, TEST AND VALIDATION SETS. THE SPECIES COMPOSITION OF EACH SET AND THE VALUES OF HEIGHT, DIAMETER AT BREAST HEIGHT (DBH) AND STEM VOLUME ARE ALSO PRESENTED (N= TOTAL NUMBER OF SAMPLES).

	N	%	Height (m)		DBH (cm)		Stem Volume (m^3)	
			Mean	Range	Mean	Range	Mean	Range
Training Set	174	100	26.2	7.5 – 38.1	45.2	9 – 90	2.34	0.04 – 10.93
Silver Fir	74	42.5	26.3	7.5 – 38.0	45.8	13 – 90	2.37	0.05 – 10.93
Norway Spruce	79	45.4	28.1	15.4 – 37.7	49.0	25 – 74	2.68	0.37 – 7.12
Other specie	21	12.1	19.6	11.6 – 28.8	29.7	9 – 63	0.95	0.04 – 3.74
Test Set	147	100	25.8	11.1 – 37.0	45.0	13 – 74	2.27	0.08 – 7.20
Silver Fir	71	48.3	25.8	12.9 – 36.8	44.2	19 – 73	2.20	0.20 – 6.11
Norway Spruce	59	40.1	27.4	15.9 – 37.0	48.6	21 – 74	2.66	0.31 – 7.20
Other specie	17	11.6	20.6	11.1 – 29.2	36.3	13 – 60	1.21	0.08 – 3.51
Validation Set	160	100	26.44	9.4 – 38.1	45.2	9.4 – 71	2.30	0.05 – 7.21
Silver Fir	67	41.9	26.1	12.4 – 35.0	43.7	16 – 71	2.17	0.19 – 6.32
Norway Spruce	79	49.4	28.3	9.4 – 38.1	49.6	9.4 – 38.1	2.72	0.05 – 7.21
Other specie	14	8.7	17.6	12.4 – 24.5	27.7	14 – 44	0.53	0.10 – 1.47

The LIDAR data were acquired on September 4th, 2007, between 11:29 AM and 12:07 AM. These data were taken by a sensor Optech ALTM 3100EA, with a mean density of 8.6 points per square meter for the first return. The laser pulse wavelength and the laser repetition rate were 1064 nm and 100 kHz, respectively. The number of recorded returns for each laser pulse is up to four. A Digital Terrain Model (DTM) with a ground resolution of 1 m was derived from these data and it was used in the preprocessing of the LIDAR returns.

Reference data were collected on the ground in summer 2007. We analyzed 481 sample trees distributed in 50 plots selected in order to obtain a statistical representation of the study area. Sample plots were distributed randomly across the analyzed forest area according to an unaligned systematic sampling design (see [24]). A square grid with cells of 250 x 250 m was laid over an ortophoto of the area and in each square a plot was randomly inserted and classified as inside or outside the forest study area. 50 of these plots were interpreted as forest within the boundaries of the study area. Among these plots we measured parameters (e.g., height, DBH) of some trees according to the relascopic technique. These sample trees where then divided intro three sets: training, test and validation sets. Table I shows the distribution of the points in these sets, the species composition, the values of height, Diameter at Breast Height (DBH) and stem volume.

B. Data preprocessing and architecture of the system

Figure 2 presents the architecture of the system proposed. The goal of the system is to obtain a map of tree stem volume by integrating the information provided by both multireturn LIDAR data and a classification map.

Regarding the preprocessing phase we have two steps: i) rasterization of the raw LIDAR data, and ii) subtraction of the Digital Terrain Model (DTM) to the elevation information of the LIDAR data. The rasterization was performed with a ground resolution of 1 m. The average values of elevation and intensity of each return included in a square meter were assigned to each pixel (we obtained two images for each return). In the first return images, pixels with no value were interpolated with the nearest neighbor technique, while for images of returns different from the first we left value 0. After that, the Digital Terrain Model (DTM) of the area considered was subtracted from the elevation image of each return. This allowed us to correct the raw LIDAR elevations from the topography of the scene. In particular subtracting the DTM from the elevation of the first LIDAR return we obtained the digital Canopy Height Model (CHM) (see Figure 1).

In the proposed architecture hyperspectral data are used for the identification of the tree species. The classification map obtained from the classification of these data is aggregated at crown level in order to have an information on the species for each crown. In the rest of this study we focus our attention only on the estimation part. We refer the reader to [25] for more details on the classification architecture adopted for the hyperspectral images.

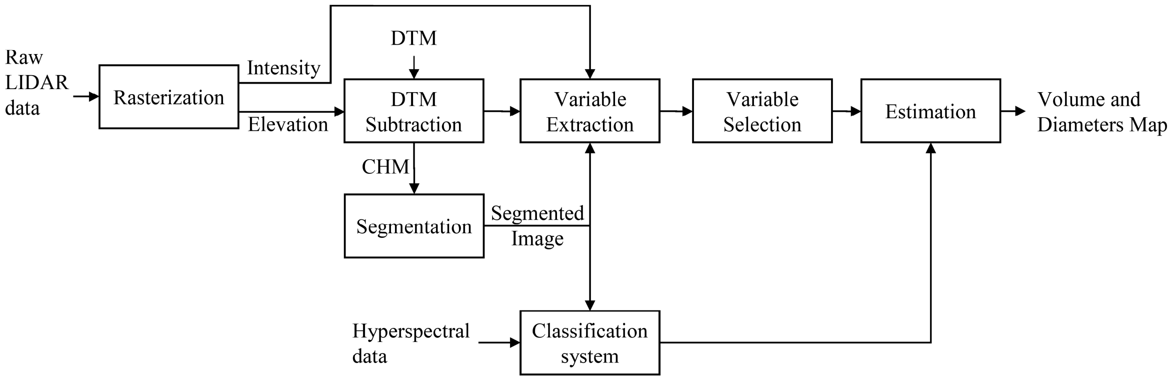

Figure 2. Architecture of the proposed system.

C. Segmentation

The rationale of this step is to identify and delineate individual tree crowns. The segmentation algorithm used in this paper is derived from the algorithm presented by Hyyppä *et al.* in [12]. This algorithm, is divided into three main steps: i) prefiltering; ii) seed point extraction; iii) seed region growing.

According to [12], in the prefiltering phase the CHM was filtered with a convolution filter for emphasizing local maxima and tree crowns. The coefficients of the filter using a 3x3 window are defined as follows:

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} /16 \qquad (1)$$

After this phase it is necessary to identify the seed points (which correspond to the tree tops) for initializing the region growing procedure. In order to consider only trees higher than a given value, seeds are defined as the local maxima higher than a certain threshold $thSeed$. A moving window of a predefined size (selected by the user) is used to detect them. If $\mathbf{W}$ is the analysis window and $I(i, j)$ the elevation value of the pixel at position (i, j), we can write as follows:

$$I(i,j)\ is\ a\ seed\ if\ \begin{cases} \arg\max_{(i,j)\in \mathbf{W}} [I(i,j)] \\ I(i,j) > thSeed \end{cases} \tag{2}$$

At the end of this process we obtain the set of the seed points $S = \{s_1,...,s_N\}$, where s_n identifies the *n*-th seed point. Figure 3 shows an example of seed points extracted from a prefiltered CHM image.

The last phase consists in the seed region growing and it is aimed at the identification of tree crowns. Seed region growing starts from each seed and grows iteratively. A pixel $I(i,j)$ is added to the considered region if it satisfies two conditions that take into account both the dimension and the shape of the crown. If we define the set of regions $R = \{r_1,...,r_N\}$, where r_n identifies the region around the seed point s_n, we can write as follows:

$$I(i,j) \in r_n \quad if \quad \begin{cases} I(i,j) > P * I_{s_n} \\ D[r_n + I(i,j)] < thDiameter \\ I(i,j) \notin r_k \quad with \quad k \neq n \end{cases} \tag{3}$$

where I_{s_n} is the height of the considered seed point, $P \in (0,1]$, $D[r_n + I(i,j)]$ is the diameter of the considered region including the new pixel $I(i,j)$, and $thDiameter$ is the maximum acceptable diameter of a region. In our experiments we chose a window size of 5x5 and a *thSeed* of 3 m.

The algorithm stops when no more pixels are added to any region.

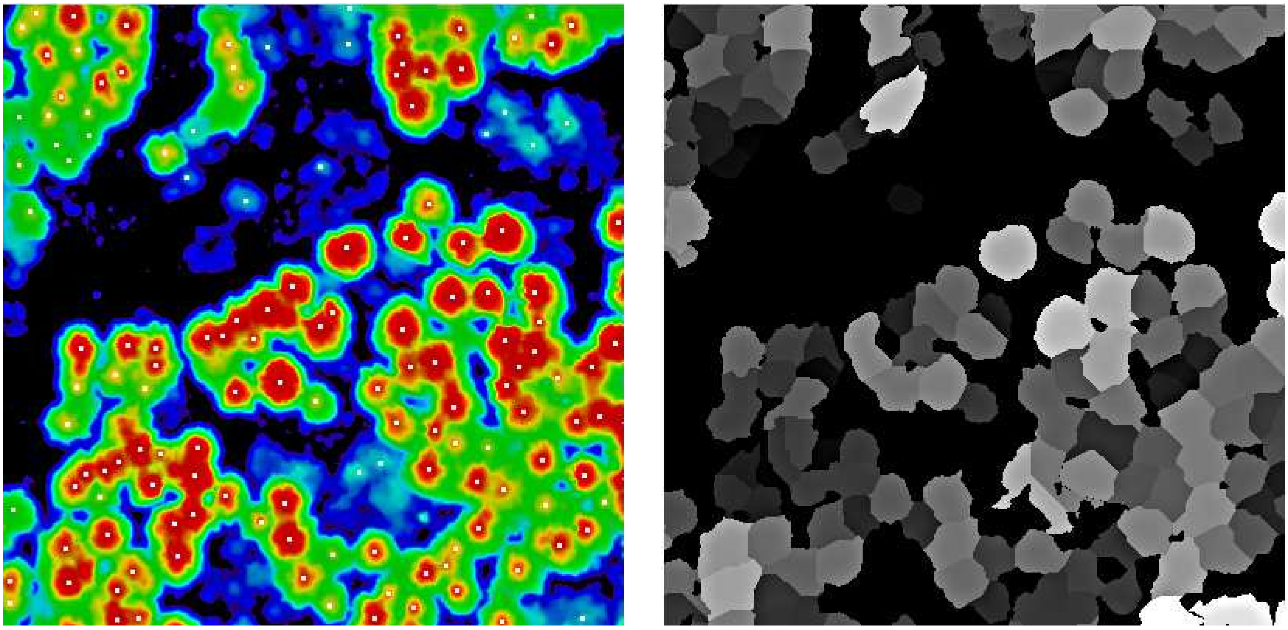

Figure 3. Example of i) seed points extracted from the elevation of the first LIDAR return, and ii) regions associated with the crowns.

D. Variable definition

In order to make a detailed characterization of each tree, some variables were defined and extracted from each segmented region from the elevation and intensity channels of the four returns available in our dataset. These variables were selected according to previous studies reported in the literature (e.g., [13],[16]). They describe the structure and the characteristics of the trees. Table II reports a summary of the variables extracted from each region. Some variables are related to the distribution of the LIDAR pulses within the crown (e.g., "*mean*", "*maximum*", "*minimum*", etc.), while other variables model the crown geometry (e.g., "*area*", "*radius*"). As an example, the variable "*area*" for the first LIDAR return represents the surface of the region in pixels, while for the other returns represents the number of pixels for which a return is present. We extracted two radii: "*radius 1*" is the radius of the circle with area equal to the area of the region, while "*radius 2*" is computed as $(a_x + a_y)/4$, where a_x is the length of the region along the *x* axis and a_y is the length of the region along the *y* axis.

Each extracted variable is related to one or more physical properties of the tree. We defined five different groups of variables: i) tree height; ii) crown horizontal shape; iii) crown vertical shape; iv) crown internal structure; and v) species.

In general, each variable can be related to more than one characteristic. In Table III variables are divided according to their expected physical meaning.

TABLE II. VARIABLES EXTRACTED FROM EACH SEGMENTED REGION.

Return	Variable	Return	Variable
1^{st}	maximum	3^{rd}	maximum
	minimum		minimum
	mean		mean
	range		range
	variance		variance
	area		area
	mean intensity		mean intensity
2^{nd}	maximum	4^{th}	maximum
	minimum		minimum
	mean		mean
	range		range
	variance		variance
	area		area
	mean intensity		mean intensity
	mean 1^{st} – mean 2^{nd}		max 1^{st} – min 3^{rd}
	mean 1^{st} – mean 3^{rd}		radius 1
	mean 1^{st} – mean 4^{th}		radius 2
	mean 2^{nd} – mean 3^{rd}		
	mean 2^{nd} – mean 4^{th}		
	mean 3^{rd} – mean 4^{th}		

TABLE III. PHYSICAL MEANING ATTRIBUTED TO EACH VARIABLE.

Physical characteristic of the tree	Variables	
Tree Height	1^{st} maximum	1^{st} minimum
	1^{st} mean	
Crown Horizontal Shape	1^{st} area	radius 2
	radius 1	
Crown Vertical Shape	1^{st} variance	mean 1^{st} – mean 3^{rd}
	1^{st} range	mean 1^{st} – mean 4^{th}
	mean 1^{st} – mean 2^{nd}	max 1^{st} – min 3^{rd}
Crown Internal Structure	2^{nd} maximum	4^{th} mean
	2^{nd} mean	4^{th} minimum
	2^{nd} minimum	4^{th} variance
	2^{nd} variance	4^{th} range
	2^{nd} range	4^{th} area
	2^{nd} area	mean 1^{st} – mean 2^{nd}
	3^{rd} maximum	mean 1^{st} – mean 3^{rd}
	3^{rd} mean	mean 1^{st} – mean 4^{th}
	3^{rd} minimum	mean 2^{nd} – mean 3^{rd}
	3^{rd} variance	mean 2^{nd} – mean 4^{th}
	3^{rd} range	mean 3^{rd} – mean 4^{th}
	3^{rd} area	max 1^{st} – min 3^{rd}
	4^{th} maximum	
Species	1^{st} intensity	3^{rd} intensity
	2^{nd} intensity	4^{th} intensity

E. Variable selection

In problems with many potentially useful predictors, the selection of variables to give as input to the regression model becomes a critical step (e.g. [26]-[33]). The importance of variable selection depends on many reasons, the most important of which are: i) the degradation of the generalization ability of the regression model when increasing the number of parameters to estimate; ii) the noise introduced by some variables; and iii) the high computational cost caused by a large number of input variables. Thus, variable selection becomes mandatory to improve the regression results both in terms of computational complexity and generalization ability. The goal of this selection is to find the smallest set of variables that provides estimates similar (or better) to those obtained with the whole set of available variables. It is worth noting that a small set of predicting variables results in a simple predicting model characterized by good generalization ability.

In the literature (i.e., [21],[26]-[33]) it is possible to find three main approaches to variable selection: i) the filter approaches; ii) the wrapper approaches; and iii) the embedded approaches. Filter approaches perform the variable selection as a preprocessing step independent on the algorithm used for model construction. These kinds of methods are widely used especially in the forestry community in combination with multiple linear estimators. Wrapper approaches select an optimized set of variables for the specific estimation technique considered. They are sometimes considered “brute force” methods and require a very high computational load, while exhibit the advantage to select the set of variables that optimizes the final predictive model (e.g., [21],[28],[29]). Embedded approaches incorporate the variable selection step in the training of the algorithm. This means that the variable selection is performed during the definition of the estimation model. These methods are mainly used in advanced machine leaning techniques [33].

As we explained in the introduction of this paper, we focus our attention on the Support Vector Regression and thus we adopted a variable selection method that can exploit the characteristics of this estimator. From the previous analysis, it emerges that both wrapper and embedded approaches can be effectively combined with SVR. For practical reasons, and according to previous works in the literature (e.g., [21],[27],[28],[30]), we decided to use a wrapper method. In particular, we adopted as search algorithm the Sequential Forward Selection and as performance measure the Mean Square Error on the test set (e.g., [22],[23]). At each step of the selection process, we applied a model selection for identifying the best parameters of the SVR using the training set (i.e., for the creation of the model) and the test set for its evaluation.

For comparisons we considered also a simple filter method based on the same search strategy and performance evaluation used in the wrapper one.

F. Non linear ε-Insensitive Support Vector Regression

In this section we briefly summarize the main principles of the non-linear and multiple ε-Insensitive Support Vector Regression (ε-SVR) algorithm used in our estimation system. For more details on the SVR theory we refer to [34].

Let $\{(\mathbf{x}_i, y_i), i = 1,...,T\}$ be a training set, where $\mathbf{x}_i \in \Re^d$ is the d-dimensional vector of selected input variables, $y_i \in \Re$ is the target tree attribute to be estimated, and T is the number of training samples. The rationale of the SVR is to map the original variable space into a higher dimensional space $\Phi(\mathbf{x})$ using a non linear transformation function Φ, and to find a linear regression function $f(\mathbf{x})$ in this new space, as:

$$f(\mathbf{x}) = \langle \mathbf{w} \cdot \Phi(\mathbf{x}) \rangle + b \tag{4}$$

where $\mathbf{w} \in \Re^d$ is the weight vector, $b \in \Re$ is the bias, and $\langle w \cdot \Phi(\mathbf{x}) \rangle$ represents the dot product between $\mathbf{w}$ and $\Phi(\mathbf{x})$. This function ($f(\mathbf{x})$) should have at most deviation ε from the real targets y_i for all the training samples and, at the same time, should be as flat as possible. In other words, we neglect errors smaller than ε whereas we penalize errors larger than ε (ε-insensitive tube), and we seek a small $\mathbf{w}$ (flatness).

The optimal function $f(\mathbf{x})$ can be obtained solving the following constrained minimization problem:

$$\begin{cases} \min\limits_{w,b,\xi} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\} \\ y_i - (\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b) \le \varepsilon, \quad \forall i = 1,...,T \\ (\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b) - y_i \le \varepsilon, \quad \forall i = 1,...,T \end{cases} \tag{5}$$

The main assumption in (5) is that a function $f(\mathbf{x})$ exists for each $\{(\mathbf{x}_i, y_i), i = 1,...,T\}$ with precision ε. In many problems this assumption does not hold or we want to allow for some errors in order to increase the generalization ability of the estimator. To this purpose, we can introduce the concept of soft margin, rewriting the minimization problem in (5) as:

$$\begin{cases} \min_{w,b,\zeta} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{T}\left(\zeta_i + \zeta_i^*\right) \right\} \\ y_i - \left(\mathbf{w}\cdot\Phi(\mathbf{x}_i) + b\right) \le \varepsilon + \zeta_i, \quad \forall i = 1,\ldots,T \\ \left(\mathbf{w}\cdot\Phi(\mathbf{x}_i) + b\right) - y_i \le \varepsilon + \zeta_i^*, \quad \forall i = 1,\ldots,T \\ \zeta_i, \zeta_i^* \ge 0, \quad \forall i = 1,\ldots,T \end{cases} \tag{6}$$

The variables ζ_i, ζ_i^* are called slack variables and are used to consider the patterns outside of the ε-insensitive tube. They have been introduced in order to cope with otherwise infeasible constraints of the original optimization problem. The values of the slack variables depend on the kind of penalization function adopted: linear or quadratic (see Figure 4 for an example of linear penalization function). C is a regularization constant that should be tuned in the model selection phase in order to reach the best trade-off between the smoothness of the function $f(\mathbf{x})$ and the tolerance to the errors (due to the patterns outside the ε-insensitive tube).

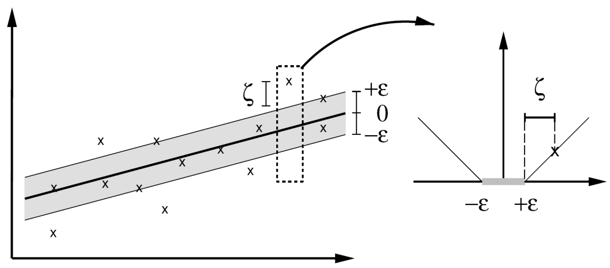

Figure 4. ε-insensitive tube of a linear SVR with a linear penalization function (figure source: [34]).

This minimization problem can be rewritten in a dual formulation and solved according to standard methods of quadratic programming based on the Lagrange multipliers (α_i). In the case of a linear penalty function, we obtain the following Lagrangian function:

$$L(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = \sum_{i=1}^{T} y_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i=1}^{T}(\alpha_i + \alpha_i^*) - \frac{1}{2}\sum_{i=1}^{T}\sum_{j=1}^{T}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{x}_i, \mathbf{x}_j) \tag{7}$$

where $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(x_i)\cdot\Phi(x_j) \rangle$ is a kernel function. This function satisfies the Mercer theorem [35] and allows us to replace the dot product among the transformation functions $\Phi(.)$. This is very important as the kernel functions avoid the need to know explicitly the transformation functions $\Phi(.)$. Common examples of kernel functions are:

1. *Linear kernel function*: $K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \boldsymbol{x}_i \cdot \boldsymbol{x}_j$.

2. *Gaussian kernel function (RBF)*: $K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \exp\left(-\left|\boldsymbol{x}_i - \boldsymbol{x}_j\right|^2 / 2\sigma^2\right)$, where σ^2 is the spread of the Gaussian.

Thus, the minimization problem in the dual formulation becomes:

$$\begin{cases} \max\limits_{\alpha}\left\{L\left(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*\right)\right\} \\ \sum_{i=1}^{T}\left(\alpha_i - \alpha_i^*\right) = 0 \\ 0 \le \alpha_i \le C, \quad \forall i = 1,\ldots,T \\ 0 \le \alpha_i^* \le C, \quad \forall i = 1,\ldots,T \end{cases} \tag{8}$$

The final estimation function in the original variable space is given by:

$$f(\mathbf{x}) = \sum_{i \in U}\left(\alpha_i - \alpha_i^*\right) y_i K\left(\mathbf{x}_i, \mathbf{x}\right) + b \tag{9}$$

The set U represents the set of training patterns corresponding to the Lagrangian multipliers different from zero. Only these patterns, defined as *support vectors* (SVs), affect the definition of the estimation function. The Lagrangian multipliers α_i and α_i^* (with i=1,…,T) allow us to define the contribution that each SV gives to the estimation function. From a geometrical viewpoint the SVs are the training patterns outside of the ε-insensitive tube.

III. EXPERIMENTAL RESULTS

A. Design of experiments

We define two main experiments. In the first one we estimate the stem diameters with the proposed system and then combine them with both the stem height (directly measured by the LIDAR) and the species information (derived from a classification map) in order to retrieve the stem volume. In the second experiment, we estimate the stem volume directly from the LIDAR variables. For both experiments we investigated three different variable selection methods (SFS with multiple linear estimator, SFS with SVR estimator and linear kernel function, and SFS with SVR estimator and RBF kernel function) and different estimators (multiple linear, SVR with linear and RBF kernel function). In the following we denote the SVR with linear kernel function with the acronym SVR-linear, and the SVR with RBF kernel function with SVR-RBF.

In the learning of the SVR algorithm we performed a grid search for the values of the parameters ε , C and of the RBF kernel width γ. The grid search range was defined in the neighborhood of the values derived from the following empirical equations [36]:

$$C = \max\left(\left| \bar{y} + 3\rho_y \right|, \left| \bar{y} - 3\rho_y \right| \right) \qquad (10)$$

$$\varepsilon = 3\rho_y \sqrt{\frac{\log(T)}{T}} \qquad (11)$$

where $\bar{y}$ is the mean value of the targets of the training set, ρ_y is their standard deviation, and T the size of the training set.

As mentioned in the data description section, we divided the available reference data in three sets: training (174 points), test (147 points) and validation (160 points) sets. The training set was used for the variable selection and the learning of the estimation algorithm, the test set was used for the model selection, and the validation set for the final estimation of performances in order to asses the generalization ability of the estimators.

B. Results

Let us consider the results of the first experiment. Table IV presents the results obtained for the estimation of stem diameters using different variable subsets and estimators. The order in which variables are presented is the selection order. One can observe that more than half of the variables selected in all the three models are extracted from the first return. These variables are chosen at the first step of the selection. In particular, the variable "*1st maximum*" is the first variable selected in all the models. It explains about 50% of the variance (53%, 54% and 54% in the three models, respectively). This variable, together with the variable "*1st mean*" was selected by all the three algorithms. These two variables and the variable "*1st minimum*" are related to the tree height and hence to the stem diameter. Among the selected variables there is then a group of variables that is connected with the crown internal structure ("*3rd variance*", "*2nd mean*", "*2nd range*", "*4th maximum*", and "*4th mean*"), a variable that can be linked to the crown vertical shape ("*1st mean – 2nd mean*"), and one to the species ("*1st intensity*").

Comparing the sets selected by the three methods one can note that they select almost the same variables. This points out that the use of a slow selection procedure like the one that exploits the SVR in the framework of a wrapper method does not necessarily select variables with physical meanings different from those of the ones selected by the faster technique based on multiple linear estimator.

Regarding the estimation algorithms, SVR-RBF provided the highest accuracy even if the difference with the other estimators is small. In greater detail, all the three estimators provided similar results with all the three variables sets considered. This is probably due to the fact that the variables considered are linearly related to the diameter and thus a simple multiple linear estimator is enough to obtain good results. From the results one can also see that, as expected, each estimator provided the highest accuracy when the selection was performed on the basis of the same estimator.

Concerning the estimation errors, it is worth noting that the MAE is small (about 6 cm on an average value of diameters equal to approximately 45 cm). Moreover the estimation is unbiased as the mean error is almost zero in all the configurations considered.

A first conclusion that it is possible to draw from these results is that with the considered variables it is not necessary to use complex and non linear techniques for the estimation of stem diameters. In particular, by using a simple linear estimator in both phases of selection and estimation it is possible to obtain results comparable to those provided by the more complex SVR, yet with a significantly lower computational cost.

Figure 5 shows a scatterplot of the observed vs. predicted diameters obtained by using the SVR-RBF technique for both variable selection and estimation for the test and validation sets. The correlation between estimated and measured diameters is good. The R^2 on the test and validation samples considered together is of about 0.63. This R^2 is the same that we obtain if we plot the observed height vs. diameters. This shows that the height explains the majority of the variability in diameter.

In order to estimate the tree stem volume we used the diameters derived by the selection and estimation based on SVR-RBF, and the tree heights provided by LIDAR measurements. The estimation was carried out by using standard height/diameter relationships adopted for the estimation of the tree stem volume in forest inventories. These equations estimate the volume combining the tree diameter, the tree height and the species information. We considered as height of a tree the variable "*1st maximum*", as diameter the one estimated with the SVR-RBF, and as species that extracted from a classification map of the considered area. Tables V and VI show the results obtained on the stem volume estimation. As it is possible to see, the results in terms of MAE, MSE and R^2 are good, and in particular they are much better compared to the ones of the diameters. This was expected as the contribution of the tree height to the volume value is much higher than that of the diameter(s) one. It is worth noting that in terms of total volume there is a underestimation of the volume for both the test and validation sets.

TABLE IV. MEAN ABSOLUTE ERROR (MAE), MEAN SQUARE ERROR (MSE) AND COEFFICIENT OF VARIATION (R^2) OF THE ESTIMATES OBTAINED ON THE TEST AND VALIDATION SETS USING DIFFERENT VARIABLE SETS AND ESTIMATORS.

Variables selected	Selection method	Estimator	Test points			Validation points		
			MAE	MSE	R^2	MAE	MSE	R^2
1st maximum, 1st mean, 1st intensity, 1st mean – 2nd mean, 3rd variance	SFS with linear estimator	linear	6.34	68.99	0.621	7.16	78.85	0.589
		SVR-linear	6.29	69.03	0.625	7.00	75.44	0.608
		SVR-RBF	6.30	69.24	0.621	6.98	77.90	0.595
1st maximum, 1st mean, 2nd mean, 1st intensity, 3rd variance, 2nd range	SFS with SVR-linear estimator	linear	6.42	69.77	0.617	7.28	79.71	0.585
		SVR-linear	6.20	67.78	0.630	7.15	78.24	0.594
		SVR-RBF	6.35	67.52	0.631	7.13	79.09	0.588
1st maximum, 1st mean, 1st minimum, 4th maximum, 4th mean	SFS with SVR-RBF estimator	linear	6.41	70.17	0.615	7.20	79.45	0.586
		SVR-linear	6.28	68.86	0.625	7.12	78.97	0.591
		SVR-RBF	6.20	67.70	0.630	7.17	78.19	0.593

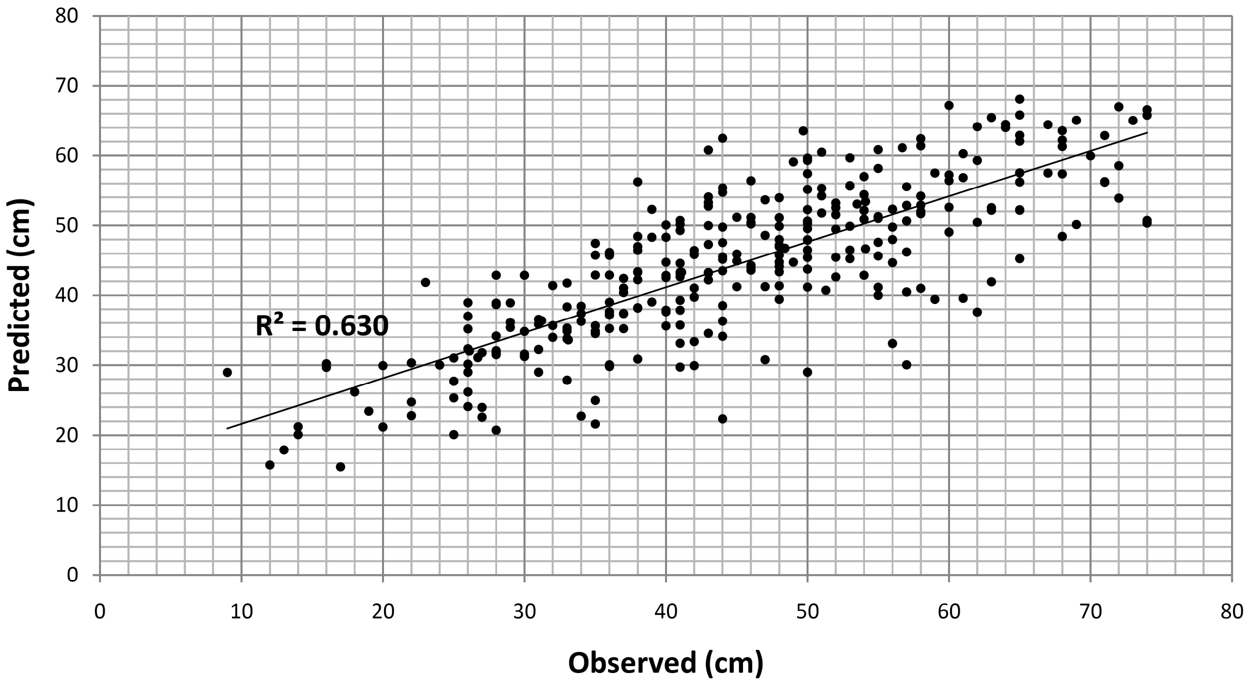

Figure 5. Observed vs. predicted diameters for the 307 trees of the test and validation sets.

Figure 6 shows the distribution of the observed vs. the predicted volume for the test and validation sets. It is possible to see that the R^2 is of about 0.7, with a significant increase compared to that obtained in the diameter estimation (0.63). From the scatterplot it is possible to note an heterogeneity in the residuals variance. This behavior can be related to our variable selection algorithm. We minimized the MSE, thus the areas where the majority of the observations are localized have an advantage in the estimation process. This behavior confirms the results reported in other similar works (e.g., [10]).

TABLE V. MEAN ABSOLUTE ERROR (MAE), MEAN SQUARE ERROR (MSE) AND COEFFICIENT OF VARIATION (R^2) OF THE ESTIMATES OBTAINED ON THE TEST AND VALIDATION SETS FOR THE STEM VOLUME. THE ESTIMATION WAS CARRIED OUT WITH STANDARD HEIGHT/DIAMETER RELATIONSHIPS USING THE DIAMETERS ESTIMATED WITH LIDAR VARIABLES AND AS HEIGHT THE VARIABLE "*1ST MAXIMUM*".

	MAE	MSE	R^2
Test Set	0.59	0.66	0.726
Validation Set	0.65	0.82	0.674

TABLE VI. TREE STEM VOLUME ESTIMATIONS OBTAINED ON THE TEST AND VALIDATION SETS WITH THE PROPOSED SYSTEM (ESTIMATED VOLUME) AND WITH GROUND COLLECTED MEASURES (REFERENCE VOLUME).

	Reference Volume (m^3)		Estimated Volume (m^3)	
	Total	Mean	Total	Mean
Test Set	330.676	2.250	311.647	2.120
Validation Set	368.613	2.304	350.937	2.193

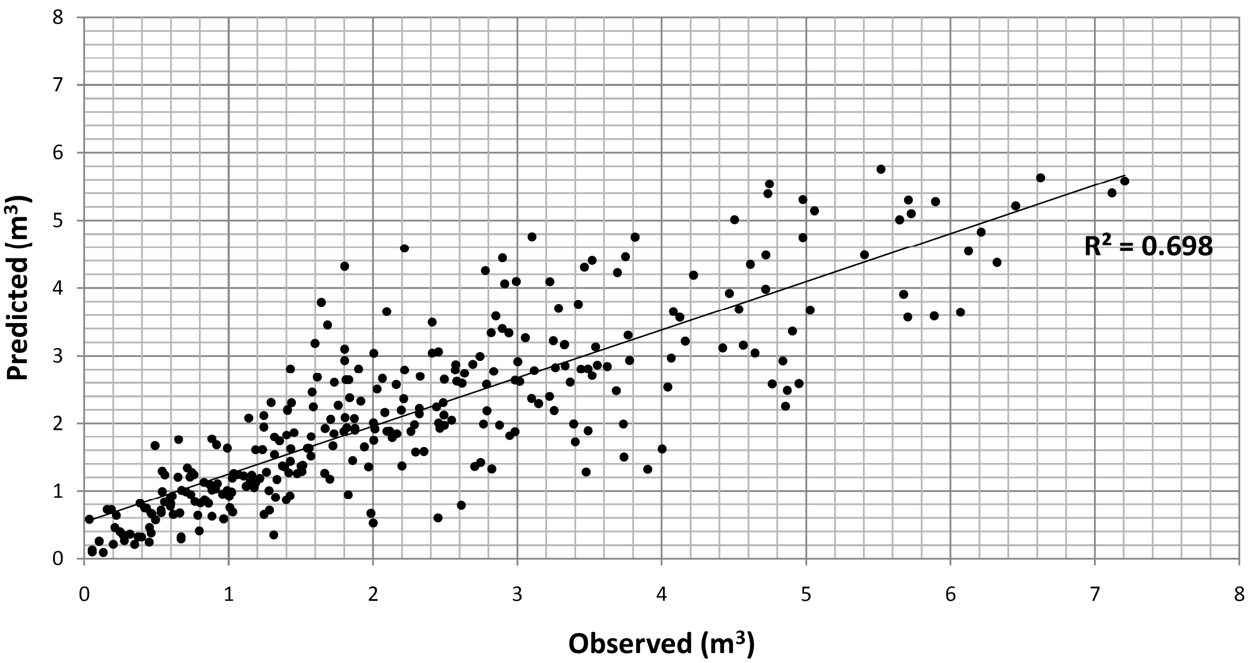

Figure 6. Observed vs. predicted volume for the 307 trees of the test and validation sets.

Let us now consider the results of the second experiment aimed at estimating the stem volume directly from the extracted LIDAR variables.

Table VII shows the estimates of the stem volume by using different variable subsets and different estimators. All the selection algorithms identified the variables "*1st maximum*" and "*1st mean – 3rd mean*". The first one is the tree height, while the second one is connected with the vertical shape of the crown. The remaining variables differ for each selection method. The selection based on multiple linear estimator chooses only variables based on the 1st return, and in particular variables connected with the crown vertical shape ("*1st mean – 2nd mean*", "*1st mean – 3rd mean*", "*1st mean – 4th mean*" and "*1st variance*"). The selection based on SVR selects variables distributed among all the four returns and all the categories considered.

These results show us that many of the variables considered are connected with the stem volume. In particular, some of them have a strong linear relation (e.g., those connected with tree height and crown vertical shape), whereas others have a non linear relation (like the crown horizontal shape and the crown internal characteristics). As expected, the latter are selected only by a non linear model (SVR-RBF).

TABLE VII. MEAN ABSOLUTE ERROR (MAE), MEAN SQUARE ERROR (MSE) AND COEFFICIENT OF VARIATION (R^2) OF THE ESTIMATES OBTAINED ON THE TEST AND VALIDATION SETS USING DIFFERENT VARIABLE SETS AND ESTIMATORS.

Variables selected	Selection method	Estimator	Test			Validation		
			MAE	MSE	R^2	MAE	MSE	R^2
1st maximum 1st mean-2nd mean 1st mean-3rd mean 1st mean-4th mean 1st variance	SFS with linear estimator	linear	0.69	0.76	0.643	0.76	0.89	0.79
		SVR-linear	0.68	0.74	0.652	0.74	0.89	0.79
		SVR-RBF	0.67	0.69	0.661	0.68	0.87	0.76
1st maximum 1st mean-3rd mean 4th mean 3rd minimum 1st mean-4th mean	SFS with SVR-lin estimator	linear	0.69	0.77	0.636	0.77	0.91	0.83
		SVR-linear	0.69	0.71	0.661	0.75	0.93	0.86
		SVR-RBF	0.66	0.64	0.681	0.71	0.88	0.77
1st maximum 1st mean-3rd mean 1st minimum 4th range 1st intensity radius 1	SFS with SVR-RBF estimator	linear	0.70	0.77	0.639	0.78	0.91	0.83
		SVR-linear	0.68	0.72	0.651	0.74	0.92	0.85
		SVR-RBF	0.60	0.49	0.715	0.71	0.86	0.74

One can observe that also in this case the selection based on SVR-RBF combined with the SVR-RBF estimator provided the best results on both the test and validation sets (R^2 is of 0.71 on the test set compared to an average of 0.65 of the other methods). It is also worth noting that the estimations based on the SVR have always higher accuracies than those based on the linear estimator even if in some configurations the differences are relatively small (e.g., in the selection based on linear estimator results obtained with linear estimator and SVR with linear kernel function are quite similar).

Table VIII reports the total and mean predicted volume in m^3 for the test and validation sets, compared to the observed ones. As one can see, the total and mean values are quite similar for both the sets.

TABLE VIII. TREE STEM VOLUME ESTIMATION OBTAINED ON THE TEST AND VALIDATION SETS WITH THE PROPOSED SYSTEM (PREDICTED VOLUME) AND WITH GROUND COLLECTED MEASURES (OBSERVED VOLUME).

	Observed Volume (m^3)		Predicted Volume (m^3)	
	Total	Mean	Total	Mean
Test Set	330.676	2.250	325.601	2.215
Validation Set	368.613	2.304	362.693	2.267

Figure 7 shows the distribution of the observed volume vs. the predicted one with the SVR-RBF algorithm for the test and validation sets.

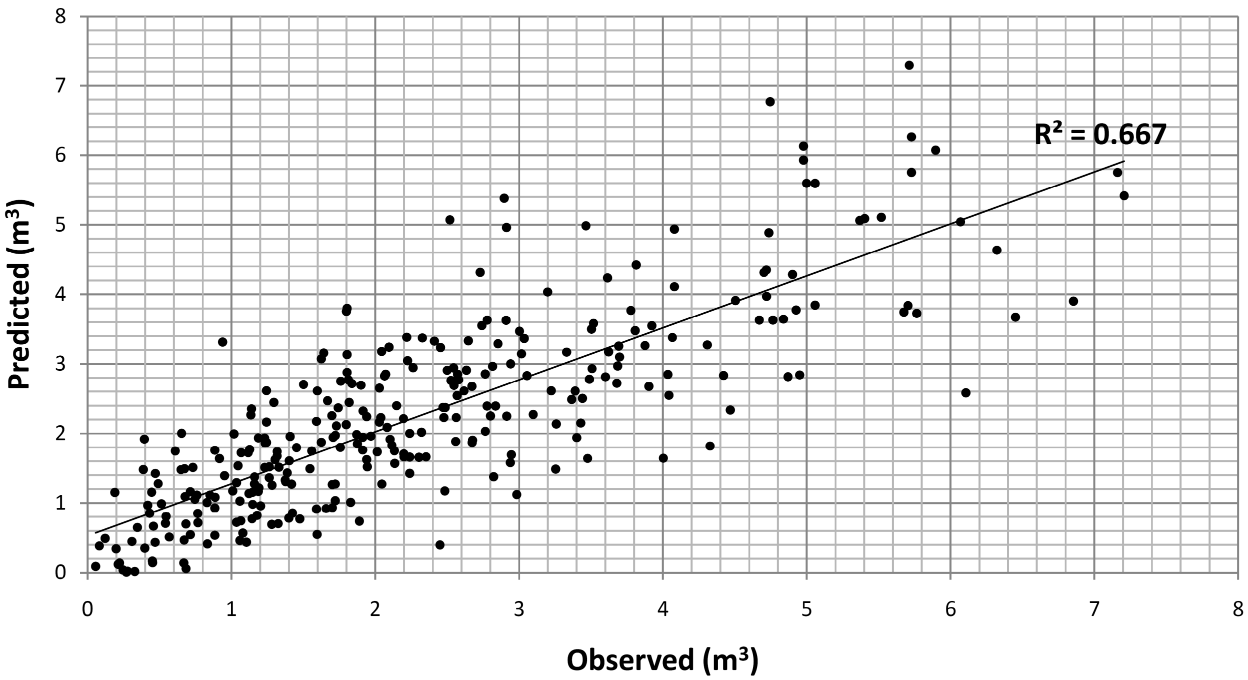

Figure 7. Observed vs. predicted volume for the 307 trees of the test and validation sets.

C. Discussion

By analyzing the results of the two experiments it is possible to draw some interesting conclusions. Firstly, the estimates of the stem volume obtained with the two considered approaches are accurate in both cases, with the approach based on stem diameter estimation that provides better results in terms of estimation accuracy and coefficient of determination (we have higher values of R^2 on both the test and validation sets: i.e., 0.72 and 0.67 vs. 0.71 and 0.65, respectively). These results show the effectiveness of the proposed system. By comparing Tables VI and VIII one can observe that the approach based on stem diameter estimation underestimates the total volume more than the approach based on the direct volume estimation (even if the results are only slightly different).

Another interesting observation regards the comparison between the estimation results of stem diameter and volume starting from LIDAR variables. It seems that the considered variables are more correlated with stem volume than with diameter. In the estimation of stem diameters the variables selected are quite similar for all the estimators considered; thus the results of the different estimators are very similar. This is not the case of volume estimation, where the variables selected differs from an estimator to another. In particular SVR-RBF (a complex non linear estimator) selects a pool of variables with different physical meanings reaching higher accuracies with respect to other estimators.

By analyzing the variables selected, one can see that there are some variables strongly correlated with both the volume and the diameter, like the ones related to tree height. A variable belonging to this set is present in all selection results. This is quite obvious as the tree height is strongly related to both diameter and volume ("*1st maximum*" explains more than half of the variance in both cases). Regarding the other variables, it seems that those related to the crown internal structure are those more correlated to the diameter. In fact almost one third of the selected variables belong to this set. On the contrary, it seems that variables related to the horizontal and vertical shape of the crown do not provide much information in this context. Considering the estimation of stem volume, the situation is quite different. In this case we have very different variables sets changing the estimator considered. It is worth noting that the selection performed by the SVR-RBF provides variables belonging to every set. Moreover, in this case the variables selected by the linear estimator belong all to the 1st return showing that probably these are the only variables that have a linear correlation with the volume.

Comparing the variables selected in this study with works previously presented in the literature one can note that for the estimation of stem diameter a linear relationship between LIDAR-derived height and LIDAR-derived crown diameter is commonplace (e.g., [7]-[12]). Among the papers that use an approach based on a variable selection procedure, the most common variables are: LIDAR-derived tree height, some variables linked to the crown shape (e.g., crown area, crown exterior volume) and percentiles of LIDAR heights. In our case the LIDAR-derived height is always present in our models ("*1st maximum*"); also variables linked to the crown shape and structure are always present in our selection (e.g., "*1st mean- 2nd mean*", "*2nd range*").

In the literature the stem volume is widely computed through standard relationships between LIDAR-derived height and stem diameter (as in the first approach that we adopted). In the cases where it is directly estimated from LIDAR data, the predictors are similar of diameters.

IV. CONCLUSION

In this paper a system for the exploitation of discrete multireturn LIDAR data for the estimation of tree stem diameter and volume is presented. The system proposed is made up of four different blocks: preprocessing, segmentation, variable extraction and selection, and estimation. We presented and analyzed different kinds of variables extracted from LIDAR data, different variable selection algorithms and different estimation techniques. From the experimental results we can draw the following conclusions:

i. the proposed system is effective for the estimation of tree stem diameter and volume;
ii. the approach to the estimation of stem volume based on the estimation of stem diameter seems to be the most effective. The results in terms of MAE, MSE and R^2 are better if compared to the direct estimation of stem volume;
iii. the estimation accuracy is maxima when using the same regression technique in both the phases of variable selection and estimation;
iv. when the relationship between the variables and the targets can be approximated as linear, a simple estimator (like the multiple linear regressor) provides results comparable to complex non-linear estimators (like SVR);
v. a non-linear regression model (like SVR with RBF kernel function) provides always better results when compared to other estimators (like multiple linear regression).

As a future development of the proposed system, we plan: i) to improve the estimation of diameters considering other variables; ii) to compare the results of the SVR with other non-linear multiple parametric regression techniques; and iii) to analyze the impact of the posting density (number of LIDAR measures per square meter) on the estimation results.

ACKNOWLEDGMENTS

We would like to thank Sergio Tonolli for the field work and for the preparation of the ground truth data, and Matteo Tononi for the help in the LIDAR data processing. This work was partially supported by the CARBOITALY project funded by the FISR program of the Italian Ministry of University and Research.

REFERENCES

[1] C. J. Houldcroft, C. L. Campbell, I. J. Devenport, R. J. Gurney, N. Holden, "Measurement of Canopy Geometry Characteristics Using LiDAR Laser Altimetry: A Feasibility Study," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2270-2282, 2005.

[2] R. Hecht, G. Meinel, M. F. Buchroithner, "Estimation of Urban Green Volume Based on Single-Pulse LiDAR Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3832-3840, 2008.

[3] C. Wang, M. Menenti, M. Stoll, A. Feola, E. Belluco, M. Marani, "Seperation of Ground and Low Vegetation Signatures in LiDAR Measurements of Salt-Marsh Environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2014-2023, 2009.

[4] J. S. Evans, A. T. Hudak, "A Multiscale Curvature Algorithm for Classifying Discrete Return LiDAR in Forested Environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 4, pp. 1029-1038, 2007.

[5] R. M. Palenichka, M. B. Zaremba, "Multiscale Isotropic Matched Filtering for Individual Tree Detection in LiDAR Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3944-3956, 2007.

[6] B. Koch, U. Heyder, H. Weinacker, "Detection of Individual Tree Crowns in Airborne Lidar Data," *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 4, pp. 357-363, 2006.

[7] K. Zhao, S. Popescu, R. Nelson, "Lidar remote sensing of forest biomass: A scale-invaraint estimation approach using airborne lasers", *Remote Sensing of Environment*, vol. 113, pp. 182-196, 2009.

[8] A. Persson, J. Holmgren, U. Soderman, "Detecting and Measuring Individual Trees Using an Airborne Laser Scanner," *Photogrammetric Engineering & Remote Sensing*, vol. 68, no. 9, pp. 925-932, 2002.

[9] M. Schardt, M. Ziegler, A. Wimmer, R. Wack, J. Hyyppä, "Assessment of forest parameters by means of laser scanning," *Proceedings of the ISPRS (International Society of Photogrammetry and Remote Sensing) working group part 8/2Int. Arch. Photogramm. Remote Sensing*, vol. 36, pp. 272−276, 2002.

[10] M. Heurich, "Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park, *Forest Ecology and Management*, vol. 255, pp. 2416-2433, 2008.

[11] S. C. Popescu, R. H. Wynne, R. F. Nelson, “Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass,” *Canadian Journal of Remote Sensing*, vol. 29, no. 5, pp. 564-577, 2003.

[12] J. Hyyppä, O. Kelle, M. Lehikoinen, and M. Inkinen, “A Segmentation-Based Method to Retrieve Stem Volume Estimates from 3-D Tree Height Models Produced by Laser Scanners,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 5, pp. 969-975, 2001.

[13] J. Vauhkonen, I. Korpela, M. Malmo, T. Tokola, “Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics,” *Remote Sensing of Environment*, vol. 114, pp. 1263-1276, 2010.

[14] Q. Chen, P. Gong, D. Baldocchi, Y. Q. Tian, “Estimating Basal Area and Stem Volume for Individual Trees from Lidar Data,” *Photogrammetric Engineering & Remote Sensing*, vol. 74, no. 12, pp. 1355-1365, 2007.

[15] Z. J. Bortolot, and R. H. Wynne, “Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data”, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, no. 6, pp. 342-360, 2005.

[16] M. Dalponte, N. C. Coops, L. Bruzzone, D. Gianelle, “Analysis on the use of multiple return LiDAR data for the estimation of tree stems volume,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 2, No. 4, pp. 310-318, December 2009.

[17] A. Bucksch, R. Lindenbergh, M. Menenti, M. Z. Raman, “Skeleton-based botanic tree diameter estimation from dense LiDAR data,” *Proceedings of SPIE*, vol. 7460 (SPIE, Bellingham, WA 2009) 746007.

[18] A. Kato, L. M. Moskal, P. Schiess, M. E. Swanson, D, Calhoun, W. Stuetzle, “Capturing tree crown formation trough implicit surface reconstruction using airborne lidar data,” *Remote Sensing of Environment*, vol. 113, pp. 1148-1168, 2009.

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

[20] F. Yang, M. A. White, A. R. Michaelis, K. Ichii, H. Hashimoto, P. Votava, A-X. Zhu, and R. R. Nemani, “Prediction of Continental-Scale Evotranspiration by Combining MODIS and AmeriFlux Data Through Support Vector Machines,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3452-3461, 2006.

[21] J. Guajardo, J. Miranda, and R. Weber, "A hybrid forecasting methodology using feature selection and support vector regression", *Proceedings of the Fifth International Conference on Hybrid Intelligent Systems 2005*, vol. 6, 2005.

[22] Y. Bazi, and F. Melgani, "Semisupervised PSO-SVM Regression for Biophysical Parameters Estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1887-1895, 2007.

[23] L. Bruzzone, and F. Melgani, "Robust Multiple Estimator Systems for the Analysis of Biophysical Parameters From Remotely Sensed Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 159-174, 2005.

[24] U.S. Environmental Protection Agency, "Guidance on choosing a sampling design for environmental data collection", EPA QA/G-5S, Washington, D.C, 2002.

[25] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1416-1427, 2008.

[26] A. L. Blum, P. Langley, "Selection of relevant features and examples in machine," *Artificial Intelligence*, vol. 97, pp. 245-271, 1997.

[27] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, V. Vapnik, "Feature Selection for SVMs," *Advances in Neural Information Processing Systems 13*, pp. 668-674, 2000.

[28] P. Langley, "Selection of Relevant Features in Machine Learning," *Proceedings of the AAAI Fall symposium on relevance*, pp. 140-144, 1994.

[29] R. Kohavi, G. H. John, "Wrappers for feature subset selection," Artificial Intelligence, vol. 97, pp. 273-324, 1997.

[30] T. Nagatani, S. Abe, "Backward Varilable Selection of Support Vector Regressors by Block Deletion," *International Joint Conference on Neural Networks*, 2007, pp. 2117-2122.

[31] I. Guyon, A. Elisseeff, "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.

[32] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and Regression Trees", *Wadsworth & Brooks*, Monterey, 1984.

[33] J. Weston, A. Elisseff, B. Schoelkopf, and M. Tipping, "Use of the zero norm with linear models and kernel methods", *Journal of Machine Learning Research*, vol. 3, pp. 1439-1461, 2003.

[34] A. J. Smola, and B. Schoelkopof, “A Tutorial on Support Vector Regression”, *NeuroCOLT2 Technical Report Series*, NC2-TR-1998-030, October, 1998.

[35] J. Mercer, “Functions of positive and negative type and their connection with the theory of integral equations,” *Philos. Trans. Roy. Soc. London*, 1909.

[36] V. Cherkassky, and Y. Ma, “Practical Selection of SVM Parameters and Noise Estimation for SVM Regression”, *Neural Networks*, vol. 17, no. 1, pp. 113-126, 2004.