JOURNAL OF FOREST SCIENCE, *63*, 2017 (2): 88–97

doi: 10.17221/86/2016-JFS

Impact of plot size and model selection on forest biomass estimation using airborne LiDAR: A case study of pine plantations in southern Spain

Rafael M. NAVARRO-CERRILLO[1]*, *Eduardo GONZÁLEZ-FERREIRO*[2], *Jorge GARCÍA-GUTIÉRREZ*[3], *Carlos J. CEACERO RUIZ*[4], *Rocío HERNÁNDEZ-CLEMENTE*[5]

[1]*Department of Forestry Engineering, School of Forest Engineering, University of Cordoba, Córdoba, Spain*

[2]*Department of Agroforestry Engineering, School of Forest Engineering, University of Santiago de Compostela, Lugo, Spain*

[3]*Department of Computer Science Languages and Systems, School of Industrial Engineering, University of Seville, Sevilla, Spain*

[4]*Department of Physiology, Anatomy and Cellular Biology, Faculty of Biology, Pablo de Olavide University, Sevilla, Spain*

[5]*Department of Geography, College of Science, Swansea University, Swansea, UK*

**Corresponding author: irlnacer@uco.es*

Abstract

Navarro-Cerrillo R.M., González-Ferreiro E., García-Gutiérrez J., Ceacero Ruiz C.J., Hernández-Clemente R. (2017): Impact of plot size and model selection on forest biomass estimation using airborne LiDAR: A case study of pine plantations in southern Spain. J. For. Sci., 63: 88–97.

We explored the usefulness of LiDAR for modelling and mapping the stand biomass of two conifer species in southern Spain. We used three different plot sizes and two statistical approaches (i.e. stepwise selection and genetic algorithm selection) in combination with multiple linear regression models to estimate biomass. 43 predictor variables derived from discrete-return LiDAR data (4 pulses per m^2) were used for estimating the forest biomass of *Pinus sylvestris* Linnaeus and *Pinus nigra* Arnold forests. Twelve circular plots – six for each species – and three different fixed-radius designs (i.e. 7, 15, and 30 m) were established within the range of the airborne LiDAR. The Bayesian information criterion and R^2 were used to select the best models. As expected, the models that included the largest plots (30 m) yielded the highest R^2 value (0.91) for *Pinus* sp. using genetic algorithm models. Considering *P. sylvestris* and *P. nigra* models separately, the genetic algorithm approach also yielded the highest R^2 values for the 30-m plots (*P. nigra*: R^2 = 0.99, *P. sylvestris*: R^2 = 0.97). The results we obtained with two species and different plot sizes revealed that increasing the size of plots from 15 to 30 m had a low effect on modelling attempts.

Keywords: airborne laser scanning; forest inventory; regression; survey design; genetic selection methods; *Pinus* sp.

The accurate estimation of forest biomass and the spatially explicit mapping of forest stocks based on field inventory methods or aerial photography have gained considerable interest (Avery, Burkhart 1994). The high cost of establishing field plots has partially been offset by the recent progress made in

---

Supported by the European Commission, Project No. FP7-315165, by the Ministry of Economy, Industry and Competitiveness, Project QUERCUSAT (CLG2013-40790-R), and by the University of Cordoba-Campus de Excelencia (ceiA3).

remote sensing technologies (e.g. airborne LiDAR), and it has been accepted that a modified sampling design with adequate statistical approach is required (Zhao, Popescu 2009). Along with sampling effort, field plot shape and size selection are still among the main challenges of forest inventory design, because it is necessary to minimize potential edge effects and to capture an adequate amount of structural variability in the field (Zenner 2005; González-Ferreiro et al. 2013).

The use of airborne LiDAR as a sampling tool for taking forest inventories has been widely used to provide accurate estimates of aboveground forest biomass (Anderson et al. 2006; Zhao, Popescu 2009). In addition, LiDAR systems reduce ground-based measures such as tree diameters, heights and canopy parameters providing better data processing and positional accuracy than photogrammetric techniques (Maltamo et al. 2004; Næsset 2004; González-Ferreiro et al. 2013). It may be possible to reduce the edge effect associated with LiDAR metrics related to trees located just outside plot boundaries by considering the use of different plot sizes as part of the design parameters in forest surveys. Sample plots with a high perimeter-to-area ratio may substantially reduce the negative impact of the edge effect on LiDAR metrics (Frazer et al. 2011).

Several authors have reported that it is possible to estimate biomass at a tree and stand level in various forest types and conditions using small-footprint LiDAR measurements (Maltamo et al. 2004; González-Ferreiro et al. 2012). Li et al. (2008) demonstrated that some explanatory variables (i.e. mean height, canopy cover, and coefficient of variation of height) are strongly correlated with stand biomass and remain stable across different landscapes.

LiDAR-based models for estimating plot-level forest attributes (such as height, basal area, and biomass) are usually based in the relationships between ground data and selected LiDAR metrics generated using regression analyses (González-Ferreiro et al. 2012). Concerning biomass, ground reference values are obtained by destructive sampling or more often using allometric equations (Nelson et al. 2004). Design inventories of forest biomass at different scale resolutions by considering individual tree data at various spatial scales is a way to assess the effect of scale on the quantification of biomass using airborne LiDAR.

Moreover, in the context of applying LiDAR to the assessment of forest structure, most estimation models explored in previous studies used a stepwise selection combined with multiple linear regression (MLR), and the models were fitted using data collected at a given plot size (Næsset 2002, 2004). However, many problems related to stepwise selection have been reported in the literature. The main concerns involve the dependence of results on the order of parameter entry (or deletion) and the absence of the control of the inflation of Type I errors in the sequence of statistical tests (Whittingham et al. 2006). Recently, some authors have proposed to overcome the classic limitations of stepwise MLR on LiDAR by using evolutionary techniques (Latifi et al. 2010; García-Gutiérrez et al. 2014). Specifically, genetic algorithm (GA) selection appears to be a suitable tool to overcome the above-mentioned problems related to variable selection.

The impact of sample plot size on the prediction accuracy of least-squares regression estimators of forest biomass using airborne LiDAR has often been explored in boreal forests (Gobakken, Næsset 2008; Mauro et al. 2009). Yet, further studies are needed to obtain additional information in order to apply this knowledge to pine plantations in Mediterranean areas. In this paper we explored the optimization, under some cost or time constraints, of plot size to improve the quantification of forest biomass based on LiDAR and field inventory data in forests of *Pinus sylvestris* Linnaeus and *Pinus nigra* Arnold in southern Spain. We compared two statistical techniques for the selection of the variables: stepwise selection and GA selection.

MATERIAL AND METHODS

Study area. The experimental area was located in the Sierra de los Filabres mountain range (Almeria province, southeastern Spain) (37°13'27''N, 2°32'54''W) (Fig. S1). The elevation of the study area ranged from 1,540 to 2,000 m a.s.l. and annual rainfall ranged between 300 and 400 mm. The annual average temperature was 11°C, with a maximum of 32°C in summer and minimum of –8°C in winter. The vegetation was a 40-year-old mixed pine stand of *P. nigra* and *P. sylvestri*s. The forest stands contained sparse evergreen shrubs (*Adenocarpus decorticans* Boissier and *Cistus laurifolius* Linnaeus). Parent material was composed of siliceous rock with quartz mica schists, which form Eutric Cambisol-Regosol soils.

Field sampling. In July 2008, 12 plots – 6 per species – were established within the range of the airborne LiDAR strips (Fig. S1) in *P. sylvestris* and *P. nigra* plantations. The plots were randomly located considering similar canopy structural parameters in areas with slopes < 20%. In each plot, all the trees with DBH (1.3 m a.g.l. – DBH at 1 cm intervals) larger than or equal to 10 cm were measured in circular plots with a radius of 7 (153.9 m^2), 15 (706.9 m^2), and 30 m (2,827.3 m^2) overlapped (i.e. the plot centre is 7 m, which then expands to 15 and 30 m).

Total tree height was measured with a Vertex III hypsometer (Haglöf Sweden AB, Sweden) (Table 1). Total aboveground biomass – W_T (mg·ha^{-1}) was estimated as the sum of the fractions calculated using nationwide *P. sylvestris* and *P. nigra* biomass (kg) equations, as Eqs. 1–8 (Ruiz Peinado et al. 2011):

P. nigra:

$$W_\mathrm{s} = 0.0403 \times d^{1.838} \times h^{0.945} \times \mathrm{RMSE} = 71.13 \tag{1}$$

where:
W_s – biomass weight of the stem fraction,
d – DBH (cm),
h – tree height (m),
RMSE – root mean square error.

$$W_\mathrm{b7} = \left[0.228 \times \left(d - 32.5\right)^2\right] \times Z \times \mathrm{RMSE} = 25.33 \tag{2}$$

where:
W_b7 – biomass weight of the thick branch fraction with a diameter larger than 7 cm (kg),
d – DBH (cm), if $d \leq 32.5$ cm, then $Z = 0$, if $d > 32.5$ cm, then $Z = 1$.

$$W_\mathrm{b2-7} = 0.0521 \times d^2 \times \mathrm{RMSE} = 31.48 \tag{3}$$

where:
$W_\mathrm{b2-7}$ – biomass weight of the medium branch fraction with a diameter between 2 and 7 cm (kg).

$$W_{\mathrm{b}2+n} = 0.0720 \times d^2 \times \mathrm{RMSE} = 26.86 \tag{4}$$

where:
$W_{\mathrm{b}2+n}$ – biomass weight of the thin branch fraction with a diameter smaller than 2 cm with needles (kg).

P. sylvestris:

$$W_\mathrm{s} = 0.0154 \times d^2 \times h \times \mathrm{RMSE} = 34.01 \tag{5}$$

$$W_\mathrm{b7} = \left[0.540 \times \left(d - 37.5\right)^2 - 0.0119 \times \left(d - 37.5\right)^2 \times h\right] \times \times Z \times \mathrm{RMSE} = 12.63 \tag{6}$$

d – DBH (cm), if $d \leq 37.5$ cm, then $Z = 0$, if $d > 37.5$ cm, then $Z = 1$.

$$W_\mathrm{b2-7} = 0.0295 \times d^{2.742} \times h^{-0.899} \times \mathrm{RMSE} = 10.83 \tag{7}$$

$$W_{\mathrm{b}2+n} = 0.530 \times d^{2.199} \times h^{-1.153} \times \mathrm{RMSE} = 11.41 \tag{8}$$

LiDAR processing. LiDAR data were acquired with an Optech Airborne Laser Terrain Mapper (small-footprint, multiple returns; LaserMap, USA) sensor operated at a laser wavelength of 1,064 nm from a flight altitude of 1,500 m a.s.l. in August 2008. The beam divergence was 0.3 mrad, pulsing frequency 33 kHz, scan frequency 50 Hz, and the maximum scan angle ± 10°. The first and the last return pulses were recorded. The whole study area was flown over in 18 strips and each strip was flown over three times, which yielded an average measurement density of about 4 pulses per m^2. Vertical and horizontal accuracy was determined to be about 0.15 and 0.45 m, respectively.

For each synthetic LiDAR plot cloud (i.e. circular plots with a fixed radius of 7, 15, and 30 m, respectively), we computed LiDAR metrics to support a multiple regression model, based on previous research by Næsset (2002). Metrics were calculated using a FUSION LIDAR Toolkit (McGaughey 2009). In summary, the LiDAR point clouds were first filtered to generate a Digital Elevation Model (DEM) of 0.5 m cell size. LiDAR metrics were computed for each LiDAR plot after normalising the data by subtracting the DEM. In this study, a total of 43 metrics were extracted from LiDAR pulses and used as regressors in the statistical analyses. For further details of the procedure used to obtain such LiDAR metrics see the steps described in González-Ferreiro et al. (2012). To obtain a complete explanation of the FUSION tools see McGaughey (2009). The summary of the LiDAR

Table 1. Main structural parameters of *Pinus nigra* Arnold and *Pinus sylvestris* Linnaeus forests (mean ± standard error) in the training areas, plot radius R = 15 m

Main species	Age (yr)	Density (trees per hectare)	h (m)	DBH (cm)	G (m^2·ha^{-1})	W_T (mg·ha^{-1})
Pinus sp.		1,169 ± 351	8.30 ± 1.06	16.58 ± 2.54	24.94 ± 4.78	85.61 ± 21.22
P. sylvestris	35 ± 2	1,088 ± 223	8.51 ± 1.19	17.04 ± 1.83	25.50 ± 6.50	87.25 ± 38.12
P. nigra	40 ± 3	1,262 ± 466	8.05 ± 0.92	16.01 ± 3.28	24.29 ± 1.82	83.50 ± 18.10

h – tree height, G – basimetric area, W_T – total aboveground biomass metrics with their corresponding descriptions is shown in Table S1.

Linear regression model and genetic selection methods. Two selection procedures (i.e. stepwise and GA) were used to objectively choose the best linear models for predicting biomass from this suite of predictors.

The stepwise selection model was primarily used to define the empirical relationships between LiDAR and fieldwork adding (forward mode) or dropping (backward mode) predictor variables one at a time until the Bayesian information criterion (BIC) statistic indicative of the relative model fit decreased (Means et al. 2000). According to the principle of parsimony, out of 43 initial variables the independent variables were reduced to fewer than 10 in all cases, including intensity variables. Variables were validated based on Pearson's correlation coefficients between independent and dependent variables. Only the variable with the highest correlation with the dependent variable (> 0.8) was retained. A value of 0.8 was chosen based on the characteristics of the data, with the knowledge that all LiDAR-derived variables are height-related, resulting in inherently high correlations. This method was selected to avoid statistically invalid models (i.e. overfitting) in the final regression step, namely linear regression using Mallow's C_p and adjusted R^2 as selection criteria. The comparison of the estimates for the selected models was based on the following three statistics: R^2, RMSE, and BIC. Although the BIC was used as the final criterion to select the best model because of its better suitability for following the parsimony principle (González-Ferreiro et al. 2012), it does not provide an intuitive idea of model precision. The R^2 and RMSE were therefore calculated to provide additional information. For this study, the data were processed using stepwise selection implemented with SPSS (Version 15.0, 2006).

Once the stepwise model results were available, we proceeded with the GA selection techniques (Renner, Ekárt 2003; García-Gutiérrez et al. 2014). The GA selection used in this paper was implemented using the Watchmaker framework (García-Gutiérrez et al. 2014). The goodness of fit of each individual model, i.e. a set of predictors selected as the potentially best subset that competes to be the best solution in the evolutionary process, was measured according to the BIC (Schwarz 1978), which provides a quality measurement penalizing the use of too many parameters (and therefore following the principle of parsimony).

In both cases, stepwise selection and GA multicollinearity were taken into account. Among the explanatory variables selected, the condition index (CI) and the variance inflation factor (VIF) were calculated. Subset models with CI > 30 or VIF > 10 were disregarded, as recommended by Belsley (1991) and Stevens (2002).

In the case of GA for individuals with values above these thresholds, the worst possible fitness was assigned to such values and they were removed by the GA itself. This guaranteed the elimination of multicollinear solutions from the genetic algorithm process. Finally, GAs were also affected by a high random influence. We reduced this undesirable effect by running the method five times for each dataset and variable (i.e. plot size and species). Subsequently, the median of the GA quality (i.e. the third best result) was selected in each case to be used in comparison with the results of the stepwise method based on the BIC, R^2, and RMSE (higher R^2 and lower residual error and BIC indicated better predictive models).

RESULTS

Stepwise and genetic selection methods

By applying a forward selection method, the number of independent variables was reduced from the original 43 to fewer than five in each model (Table 2). LiDAR-derived variables were good predictors of biomass. LiDAR height variables were the best predictors, followed by intensity variables, which were all represented across species and plot sizes.

The correlation analysis was a key component of the preprocessing analysis. It was useful to remove unwanted high correlations and reduce the number of independent variables even more, ensuring viable, valid models. Although vegetation percentiles were often well represented, only the most highly correlated percentiles with the dependent variable were ultimately selected (i.e. the 10th, 20th, 30th, 40th, 80th, 90th, 95th, and 99th height percentiles). Factors such as descriptive statistic and skewness were logical selections for distinguishing between different variables and species, based on distribution shapes and height frequencies.

The number of variables selected from the full models via stepwise and genetic selection to predict h, DBH and W_T was similar. Except for a few models (e.g. 7 m plot size), it was possible to limit most models to five or fewer independent variables with no appreciable loss in goodness of fit for each forest type.

Table 2. Multiple linear regression models built using stepwise and genetic algorithm selection procedures to predict total aboveground biomass (mg·ha^{-1}) from LiDAR variable groups according to plot size (7, 15 and 30 m). For details of prediction of LiDAR variables see the FUSION manual (McGaughey 2009)

Plot size (m)	Selection	
	stepwise	evolutionary
	Pinus spp.	
7	TOTAL, TOTAL_R2, H_P05, H_P40, H_P95, I_P40	H_P40, H_P95, I_L_SKEW
15	H_MIN, H_P05, H_P20, H_P95	H_MAD_MED, H_L1
30	H_STD, H_MAD_MODE, H_P20	TOTAL_R1, H_P10, I_L3
	Pinus nigra Arnold	
7	H_MAD_MODE, H_P40,H_P80	H_MAD_MODE, H_P40, H_CUBIC_MEAN
15	H_P10, H_P90, I_P25	H_P40, H_P50, I_P40
30	H_MIN, H_P80	H_IQ, I_L_SKEW, I_L_KURT
	Pinus sylvestris Linnaeus	
7	H_MAD_MODE, H_P99, I_MIN	H_CV, H_L3, H_P99
15	H_P95, H_VAR, I_P05, I_P25	H_P20, H_P99, I_P05
30	H_MAX	TOTAL_R3, I_KURT

Table 3 lists the best-performing models and associated descriptive statistics for both regression models. The final model selection was based on BIC, R^2, and RMSE (higher R^2 and lower residual error and BIC indicated better predictive models). As expected, models including the largest plots (i.e. 30 m) yielded the highest R^2 statistics considering *Pinus* sp. as an individual species. *Pinus* equations with adjusted R^2 values ranged from 0.58 (RMSE = 21.43%, GA, 7 m plot size) to as high as 0.91 (RMSE = 7.81%, GA, 30 m plot size). R^2 values for species biomass considering the species separately were also higher in large plots and GA models (*P. nigra* and *P. sylvestris*, 0.99, RMSE = 0.38% and 0.97, RMSE = 2.89%, respectively; GA, 30 m plot size).

Biomass cartography

The best W_T models (Table 3; GA models using 30-m plots) were applied to the LiDAR layers selected as predictor variables by the models. The total biomass estimated in the study area ranged between 21 ± 2.98 mg·ha^{-1} (RMSE of 8%) and 160 ± 2.98 mg·ha^{-1} (RMSE of 8%), and between 21 ± 2.98 mg·ha^{-1} (RMSE of 8%) and 160 ± 2.98 mg·ha^{-1} (RMSE of 8%) for *P. sylvestris*; they ranged between 21 ± 2.98 mg·ha^{-1} (RMSE of 8%) and 160 ± 2.98 mg·ha^{-1} (RMSE of 8%) for *P. nigra*. The W layers for the study area appear fragmented with blue, green, and brown areas related to the topographic and silvicultural status of the different areas, since many of them are composed of pine stands with differing site conditions and productivity (Fig. 1).

DISCUSSION

This study provides a new statistical approach to taking forest inventories based on regression equations using LiDAR predictor variables to model aboveground biomass in Mediterranean pine forests. The results of this study indicate that a single equation can be used to relate LiDAR data to total aboveground biomass in two areas with plantations of different *Pinus* species over a large area with acceptable precision (*Pinus* 85.61 mg·ha^{-1}; *P. sylvestris* 87.25 mg·ha^{-1} and *P. nigra* 83.50 mg·ha^{-1} RMSE < 21.43%), despite the fact that the use of a LiDAR-based model introduces variability due to sampling and modelling error. Raw LiDAR datasets seem to contain a considerable amount of useful information apart from height measurements. LiDAR-derived height variables are clearly more sensitive to canopy structural variation at the plot scale of sampling (Frazer et al. 2011). This was expected, given the even-age forests with the homogeneous canopy studied, but the intensity variables proved to be useful as predictors in some models.

Two different statistical approaches – stepwise selection and genetic algorithm – were compared to estimate biomass. Although reducing the number of parameters was an important consideration,

Table 3. Summary of regression parameters obtained by stepwise and evolutionary selection methods to predict total aboveground biomass (W_T, mg·ha^{-1}) from LiDAR variable groups according to plot size (7, 15 and 30 m) and species

Plot size (m)	Selection									
	stepwise					evolutionary				
	R^2	RMSE	BIC	VIF	CI	R^2	RMSE	BIC	VIF	CI
					Pinus sp.					
7	0.66	12.89	123.88	6.76	22.38	0.56	14.59	119.41	3.05	13.50
15	0.73	11.52	115.82	4.95	11.01	0.73	15.94	119.14	1.13	46.74
30	**0.89**	**7.47**	**63.38**	**4.54**	**29.29**	**0.91**	**6.69**	**60.51**	**1.94**	**11.16**
					Pinus nigra Arnold					
7	0.81	13.55	490.88	2.48	26.56	0.82	13.34	488.15	2.65	29.87
15	0.89	9.61	121.33	1.73	24.28	**0.93**	**8.04**	**112.77**	**2.24**	**25.73**
30	**0.99**	**2.29**	**15.94**	**1.20**	**24.80**	0.99	0.32	–4.11	2.64	15.52
					Pinus sylvestris Linnaeus					
7	0.49	17.65	619.87	2.29	21.56	0.52	17.13	613.54	6.01	27.69
15	0.78	12.30	158.19	5.49	40.04	**0.74**	**13.55**	**159.12**	**1.41**	**20.33**
30	0.75	7.76	31.71	0.00	15.61	**0.97**	**2.53**	**19.22**	**1.49**	**28.31**

R^2 – coefficient of determination, RMSE – root mean square error, BIC – Bayesian information criterion, VIF – variance inflation factor, CI – condition index, the most robust model is indicated in bold letters

it was not the only one applied to choose the best model for biomass estimation. In order to predict each response variable, we also sought a model with predictor variables similar to those of the other candidate models. Once that had been achieved, we chose the model with the lowest BIC values. Except for biomass models in which both species were used (*Pinus* sp.) it was possible to limit the models to five or fewer independent variables with no appreciable loss in goodness of fit for each specific forest type. *Pinus* equations with adjusted R^2 values ranged from 0.58 (RMSE = 21.43%, GA, 7 m plot size) to as high as 0.91 (RMSE = 7.81%, GA, 30 m plot size). R^2 values for species biomass considering each species separately were also higher in large plot and GA models (*P. nigra* and *P. sylvestris*, 0.99, RMSE = 0.38% and 0.97, RMSE = 2.89% respectively; GA, 30 m plot size).

R^2 values were likely high due to the homogeneous, even-aged coniferous stand structure. Moreover, these results are comparable with those obtained by Means et al. (2000) – adjusted R^2 and Næsset (2002) – R^2. Both authors used a grid-cell based LiDAR distribution approach to biomass modelling. R^2 values for these two studies ranged from 0.91 to 0.97 for *P. sylvestris* in a mature forest (639.8 m^3·ha^{-1}) (Næsset 2002). Lower adjusted R^2 values in small plots were attributed to a narrower range in biomass per-hectare values in this study (83.50 to 87.25 mg·ha^{-1}). RMSE values for conifers in our study (< 18.35 mg·ha^{-1}) compared favourably with those found by Means et al. (2000) – 73 m^3·ha^{-1} and Næsset (2002) – 18.3–31.9 m^3·ha^{-1}.

R^2 biomass values were similar in 15- and 30-m plots; however, this trend was not observed in 7-m plots. Additionally, the biomass equations used to model tree biomass were based on tree diameter and height values (Ruiz-Peinado et al. 2011), which may have contributed to better residual distributions due to the use of the same sort of variable between biomass equations and LiDAR-based biomass models. LiDAR data are inherently height-based data sources, while most biomass models do not include height as an independent variable.

The results of applying the genetic algorithm to LiDAR-based estimates using the three alternative plot sizes (7, 15 and 30 m) indicate that using larger plots (30 m) can substantially increase the precision of total biomass estimates in the Mediterranean pine plantations observed in this study. Using a larger plot captures greater variability within a given forest area and reduces the sampling error associated with measuring only small trees in a smaller plot. The authors of previous studies have proposed that the plot size be modified according to different canopy structures and densities. Yet, this is seldom possible in taking operational forest inventory with a standard plot protocol (Gobakken, Næsset 2008). However, considering the range of RMSE values of 30- and 15-m plots, the operational measurement of biomass using a smaller plot size is feasible and provides accuracy of information.

Model fit statistics did not exhibit any large differences between different approaches but generally improved with the increasing average plot size. Although biomass estimation results for 15-m

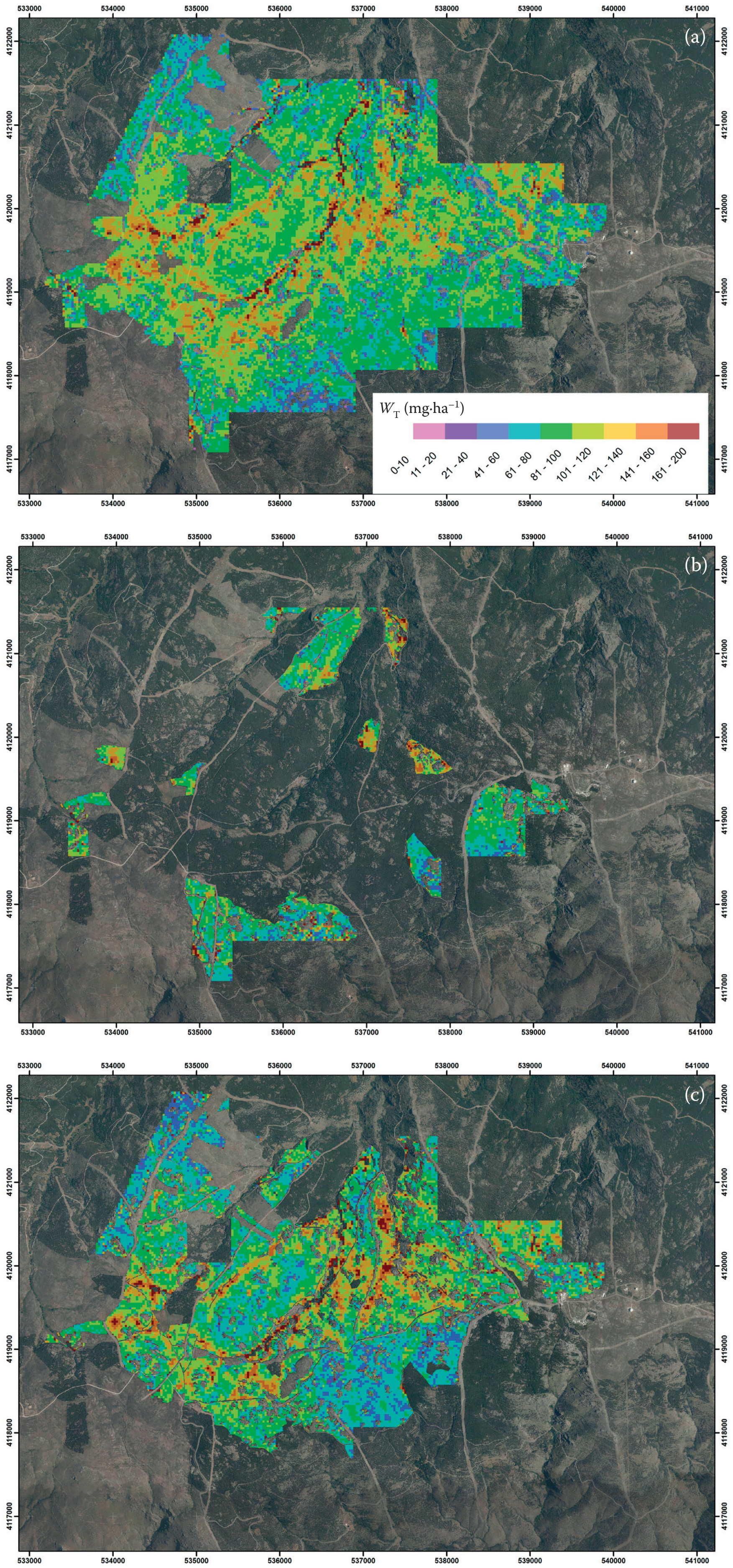

Fig. 1. Predicted total aboveground biomass (mg·ha^{-1}) for the Sierra de los Filabres woodlands using genetic algorithm models including the largest plots (30 m) for *Pinus* sp. (a), *Pinus nigra* Arnold (b), and *Pinus sylvestris* Linnaeus (c) considered separately. Note that the areas along the drainage system concentrate the highest biomass values

plots were poorer than for 30-m plots, adjusted R^2 values for the former may be acceptable according to inventory precision (Sun et al. 2011). A similar value was obtained for *P. nigra* (0.93 in 15-m plots and 0.99 in 30-m plots) but it was lower for *P. sylvestris* (0.74 in 15-m plots and 0.97 in 30-m plots). This latter result is of importance and indicates the potential of the genetic algorithm approach to biomass modelling, specifically when compared to stepwise analysis. Interestingly, 15-m plot size is an established forestry inventory technique in Spain and may account for all variability in a given stand because RMSE values were better than those found by Næsset (2002). This suggests that a LiDAR-based approach using a medium plot size (15 m) and the genetic biomass estimation for two *Pinus* species in similar ecological situations can be operationally applied to forest inventory and forest exploitation activities at a real scale. This was attributed to the relative homogeneity of the conifer stands studied, resulting in a better selection of independent variables that were correlated with dependent variables, as opposed to heterogeneous deciduous stands, where more independent variables were deemed significant for modelling. Most plots were located in uniform conifer stands and captured the homogeneous nature of such stands without including more heterogeneous structural components such as stand boundaries.

These results indicate that increasing plot size can lead to more precise estimates of biomass using LiDAR; however, there are many other factors to consider when determining the appropriate plot size for a forest inventory, including consistency, time spent on the plot, and statistical efficiency. The results presented here are intended to show that there are advantages of using medium-sized (i.e. 15 m) plots and the genetic statistical approach for LiDAR model-based estimation of biomass in Mediterranean pine plantations, an important consideration for planning future projects.

As operational cartography, biomass was mapped across the study areas. The distribution of biomass is characterized by a clear pattern associated with the drainage network. It can be seen (green colour) that the largest amount of biomass is concentrated in the vicinity of the streams where water availability increases throughout the growth season. By contrast, the lowest concentration of biomass (brown colour) is related to places of lower water availability related to defoliation processes (Navarro-Cerrillo et al. 2014). The Sierra de los Filabres is a very sensitive area to global change (Hernández-Clemente et al. 2011), which makes it particularly important to establish inventory protocols at a local scale based on LiDAR and other remote sensing data acquisitions (Navarro-Cerrillo et al. 2014). LiDAR-based cartography becomes more reliable for forest managers' decisions (as frequent as possible).

CONCLUSIONS

Biomass modelling based on LiDAR distributions has been implemented successfully for *P. sylvestris* and *P. nigra* in Mediterranean pine plantations. Our study was limited to two mountain conifer species, but R^2 values above 0.90 bode well for future LiDAR application studies in forest biomass estimations. The improvement observed in R^2 values among species and plot sizes using the genetic algorithm approach indicates that increasing the plot size from 15 to 30 m had a low effect on modelling attempts, which may allow a better sampling design to LiDAR-based forest inventories. The usefulness of the genetic algorithm to estimate biomass was very promising, although R^2 values for biomass were lower when both species were considered together. However, stepwise regression models, which showed the lowest response variables, can also be applied as an alternative in some stands. Additionally, the genetic approach proved successful in reducing the number of independent variables from as many as 43 initial height distributional variables to fewer than 5 used for final modelling. The final model selection was based on R^2, RMSE values, the BIC, and model simplicity. All criteria proved useful and even necessary to select a single best option. Biomass modelling could be a comprehensive approach to taking forest inventory using LiDAR technology and LiDAR equations could be applied to future stands, with periodic verification using plots of different size. Our results can potentially be used to efficiently assess the minimum plot size required to produce a regression estimator with the best statistical properties to monitor biomass in large pine plantation areas in Mediterranean forests.

Acknowledgment

The individual contributions of the authors to the manuscript were R.M. Navarro-Cerrillo conceived the study, carried out the field measurements, performed partially the statistical analysis and wrote the manuscript; E. González-Ferreiro performed partially the statistical analysis and wrote the manuscript; J. García-Gutiérrez performed partially the statistical analysis and wrote the manuscript; C.J. Ceacero Ruiz carried out the field measurements and wrote the manuscript; R. Hernández-Clemente carried out the field measurements, performed partially the statistical analysis and wrote the manuscript. We thank M. de los Ángeles Varo for her support during data processing and map development.

References

Anderson J., Martin M.E., Smith M.L., Dubayah R.O., Hofton M.A., Hyde P., Knox R. (2006): The use of waveform lidar to measure northern temperate mixed conifer and deciduous forest structure in New Hampshire. Remote Sensing of Environment, 105: 248–261.

Avery T.E., Burkhart H. (1994): Forest Measurements. 4th Ed. Boston, McGraw-Hill: 407.

Belsley D. (1991): Conditioning Diagnostics: Collinearity and Weak Data in Regression. New York, John Wiley & Sons: 396.

Frazer G.W., Magnussen S., Wulder M.A., Niemann K. (2011): Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. Remote Sensing of Environment, 115: 636–649.

García-Gutiérrez J., González-Ferreiro E., Riquelme-Santos J.C., Miranda D., Diéguez-Aranda U., Navarro-Cerrillo R.M. (2014): Evolutionary feature selection to estimate forest stand variables using LiDAR. International Journal of Applied Earth Observation and Geoinformation, 26: 119–131.

Gobakken T., Næsset E. (2008): Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. Canadian Journal of Forest Research, 38: 1095–1109.

González-Ferreiro E., Diéguez-Aranda U., Miranda D. (2012): Estimation of stand variables in *Pinus radiata* D. Don plantations using different LiDAR pulse densities. Forestry, 85: 281–292.

González-Ferreiro E., Diéguez-Aranda U., Barreiro-Fernández L., Buján S., Barbosa M., Suárez J., Miranda D. (2013): A mixed pixel- and region-based approach for using airborne laser scanning data for individual tree crown delineation in *Pinus radiata* D. Don plantations. International Journal of Remote Sensing, 34: 7671–7690.

Hernández-Clemente R., Navarro-Cerrillo R.M., Suárez L., Morales F., Zarco-Tejada P. (2011): Assessing structural effects on PRI for stress detection in conifer forests. Remote Sensing of Environment, 115: 2360–2375.

Latifi H., Nothdurft A., Koch B. (2010): Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: Application of multiple optical/LiDAR-derived predictors. Forestry, 83: 395–407.

Li Y., Andersen H.E., McGaughey R. (2008): A comparison of statistical methods for estimating forest biomass from light detection and ranging data. Western Journal of Applied Forestry, 23: 223–231.

Maltamo M., Eerikäinen K., Pitkänen J., Hyyppä J., Vehmas M. (2004): Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing of Environment, 90: 319–330.

Mauro F., Valbuena R., García A., Manzanera J. (2009): GPS admissible errors in positioning inventory plots for forest structure studies. Available at http://skog.for.msu.edu/meeting/proc2/Mauro_Valbuena_Garcia_Manzanera.pdf

McGaughey R. (2009): FUSION/LDV: Software for LIDAR Data Analysis and Visualization. Seattle, USDA Forest Service, Pacific Northwest Research Station: 182.

Means J., Acker S., Fitt B., Renslow M., Emerson L. (2000): Predicting forest stand characteristics with airborne scanning LIDAR. Photogrammetric Engineering and Remote Sensing, 66: 1367–1371.

Næsset E. (2002): Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment, 80: 88–99.

Næsset E. (2004): Practical large-scale forest stand inventory using small-footprint airborne scanning laser. Scandinavian Journal of Forest Research, 19: 164–179.

Navarro-Cerrillo R.M., Trujillo J., Sánchez M., Hernández-Clemente R. (2014): Hyperspectral and multispectral satellite sensors for mapping chlorophyll content in a Mediterranean *Pinus sylvestris* L. plantation. International Journal of Applied Earth Observation, 26: 88–96.

Nelson R., Short A., Valenti M. (2004): Measuring biomass and carbon in Delaware using an airborne profiling LIDAR. Scandinavian Journal of Forest Research, 19: 500–511.

Renner G., Ekárt A. (2003): Genetic algorithms in computer aided design. Computer-Aided Design, 35: 709–726.

Ruiz-Peinado R., del Rio M., Montero G. (2011): New models for estimating the carbon sink capacity of Spanish softwood species. Forest Systems, 20: 176–188.

Schwarz G. (1978): Estimating the dimension of a model. Annals of Statistics, 6: 461–464.

Stevens J. (2002): Applied Multivariate Statistics for the Social Sciences. Hillsdale, Lawrence Erlbaum Associates, Inc.: 699.

Sun G., Ranson K.J., Guo Z., Zhang Z., Montesano P., Kimes D. (2011): Forest biomass mapping from lidar and radar synergies. Remote Sensing of Environment, 115: 2906–2916.

Whittingham M.J., Stephens P.A., Bradbury R.B., Freckleton R. (2006): Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology, 75: 1182–1189.

Zenner E.K. (2005): Investigating scale-dependent stand heterogeneity with structure-area-curves. Forest Ecology and Management, 209: 87–100.

Zhao K., Popescu S. (2009): LiDAR-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA. Remote Sensing of Environment, 113: 1628–1645.

Received for publication June 29, 2016
Accepted after corrections December 19, 2016