



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
International Journal of Applied Earth Observation and Geoinformation

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa32941>

Paper:

Varo-Martínez, M., Navarro-Cerrillo, R., Hernández-Clemente, R. & Duque-Lazo, J. (2017). Semi-automated stand delineation in Mediterranean *Pinus sylvestris* plantations through segmentation of LiDAR data: The influence of pulse density. *International Journal of Applied Earth Observation and Geoinformation*, 56, 54-64.
<http://dx.doi.org/10.1016/j.jag.2016.12.002>

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/iss/researchsupport/cronfa-support/>

23 **Abstract**

24 Traditionally, forest-stand delineation has been assessed based on orthophotography.
25 The application of LiDAR has improved forest management by providing high-spatial-
26 resolution data on the vertical structure of the forest. The aim of this study was to
27 develop and test a semi-automated algorithm for stands delineation in a plantation of
28 *Pinus sylvestris* L. using LiDAR data. Three specific objectives were evaluated, i) to
29 assess two complementary LiDAR metrics, Assmann dominant height and basal area,
30 for the characterization of the structure of *P. sylvestris* Mediterranean forests based on
31 object-oriented segmentation, ii) to evaluate the influence of the LiDAR pulse density
32 on forest-stand delineation accuracy, and iii) to investigate the algorithms' effectiveness
33 in the delineation of *P. sylvestris* stands for map prediction of Assmann dominant height
34 and basal area. Our results show that it is possible to generate accurate *P. sylvestris*
35 forest-stand segmentations using multiresolution or mean shift segmentation methods,
36 even with low-pulse-density LiDAR - which is an important economic advantage for
37 forest management. However, eCognition multiresolution methods provided better
38 results than the OTB (Orfeo Tool Box) for stand delineation based on dominant height
39 and basal area estimations. Furthermore, the influence of pulse density on the results
40 was not statistically significant in the basal area calculations. However, there was a
41 significant effect of pulse density on Assmann dominant height [$F_{2, 9595} = 5.69$, $p =$
42 0.003].for low pulse density. We propose that the approach shown here should be
43 considered for stand delineation in other large *Pinus* plantations in Mediterranean
44 regions with similar characteristics.

45 **Key words**

46 LiDAR, pulse density, mean shift segmentation, multiresolution segmentation, forest-
47 stand delineation, automatic stand delineation.

48

49 **Introduction**

50 In forest management, a stand defines an area occupied by a group of trees that is
51 homogeneous - in terms of species composition, size, age, arrangement, and condition -
52 and distinguishable from other growth forms on adjoining areas ([O'Hara and Nagel,
53 2013](#)). Precise stand delineation is needed to manage different uses of the forested area
54 and its expected ecological and economic benefits and revenue. Moreover, strategic
55 decisions, such as when, where, or how to apply a particular silvicultural treatment, are
56 taken at stand level ([Dechesne et al., 2016](#)).

57 Forest-stand delineation traditionally has been assessed based on orthophotography
58 ([Burnett and Blaschke, 2003](#)). However, the combination of field forest inventory and
59 remote sensing data in cartographic and silvicultural stand delineation is becoming more
60 common ([McRoberts et al., 2014](#)). Satellite imagery (e.g. SPOT, IKONOS, or
61 QuickBird) or Color-Infrared (CIR) images ([Leckie et al., 2003](#); [Pekkarinen, 2004](#)) and
62 Light Detection And Ranging (LiDAR) data ([Bouvier et al., 2015](#)) are currently used in
63 this process.

64 Additionally, the uncertainty introduced by traditional methods of stand delineation has
65 been tackled by automatic algorithms based on image segmentation methods ([Radoux
66 and Defourny, 2007](#)). These techniques run an algorithm which generates partitions of
67 the image with similar properties ([Blaschke et al., 2014](#)). Most of the segmentation
68 techniques are based on statistical methods ([Webb, 2003](#)), where object classes are
69 represented by probability density functions. Those functions are defined over a
70 predetermined attribute space using methods based on machine learning ([Chi and Ersoy,
71 2005](#); [Zhong et al., 2008](#)), directed towards the learning of complex relationships among

72 sample patterns, and structural methods ([Centeno et al., 2003](#); [Sagerer and Niemann,](#)
73 [2013](#)) based on spatial patterns ([Costa et al., 2010](#)).

74 The application of LiDAR provides high-spatial-resolution data on the vertical structure
75 of the forest ([Wu et al., 2013](#)) and it has been used to make precise measurements of
76 forest inventory attributes (e.g., to estimate biomass, timber volume, basal area, stem
77 number, mean diameter, or dominant height) ([Næsset, 2002](#)) in order to improve forest
78 management ([Ruiz et al., 2014](#)). With the introduction of LiDAR into forest inventory
79 assessment, an increasing number of studies have involved stand segmentation detection
80 ([Bouvier et al., 2015](#)). Through time, these studies have shown increased analytical
81 complexity, increased accuracy of results, and a focus on the use of LiDAR data alone.
82 However, the specification of the fieldwork, sensor, and flight parameters for laser data
83 acquisition must be optimized to develop accurate forest inventories and mapping
84 ([Magnussen et al., 2012](#)). The LiDAR data acquisition specifications, such as scan
85 angle, pulse density, footprint size, and scan pattern, influence directly the ability to
86 derive information on the forest structure. However, such attributes must be decided
87 before the forest survey ([Ruiz et al., 2014](#)). Among these parameters, the LiDAR data
88 pulse density is one of the most significant with regard to accurate estimation of forest-
89 stand attributes ([Magnusson et al., 2007](#)).

90 In Mediterranean pine forests, canopy cover metrics are the forest management priority
91 variables (e.g., stem number, diameter, basal area, or dominant height). However, in
92 addition to yield metrics ([Lopatin et al., 2015](#); [Martín-Alcón et al., 2015](#)), other metrics
93 such as horizontal canopy heterogeneity, open canopy forest, and stand patterns must be
94 taken into account to assess silvicultural alternatives. To overcome these drawbacks, the
95 selection of LiDAR acquisition parameters (e.g., pulse number) and meaningful metrics
96 to describe stand structure, as well as delimitation stand segmentation techniques, could

97 help to develop models for specific *Pinus* Mediterranean forests. Therefore, the
98 objective of this study was to develop and test a semi-automated algorithm for stands
99 delineation in a plantation of *Pinus sylvestris* L. using LiDAR data. Three specific
100 objectives were identified, i) to assess two complementary LiDAR metrics, Assmann
101 dominant height and basal area, for characterization of the structure of *P. sylvestris*
102 Mediterranean forests based on object-oriented segmentation (e.g. eCognition software
103 and Orpheo ToolBox software), ii) to evaluate the influence of the LiDAR pulse density
104 on forest stand delineation accuracy, and iii) to investigate the algorithms' effectiveness
105 with regard to delineation of *P. sylvestris* stands, by evaluating its performance in map
106 prediction of Assmann dominant height and basal area.

107 **Materials and Methods**

108 *Study area*

109 The study area is located in "Sierra de Los Filabres" (37°13'20" N, 2°35'40" W,
110 between 1600 and 2186 m.a.s.l.), hereafter abbreviated as Filabres, south-eastern Spain
111 (Fig. S1, Supporting Material). The forest under study is a 40-year-old *Pinus sylvestris*
112 (hereafter Scots pine) plantation covering 409 ha, established using subsoiling as ground
113 preparation between 1970 and 1976. The planting density was 2000 trees ha⁻¹ and the
114 current density ranges between 342 and 1473 trees ha⁻¹. The basal area ranges from
115 11.05 to 47.31 m² ha⁻¹ (Table S1, Supporting material). Overall, the area experiences
116 typical semi-arid Mediterranean climate conditions with annual precipitation shifting
117 between 300 and 400 mm, with an average of 330 mm. Moderately mild temperatures,
118 with an average over the whole year of 13.1°C, have been reported during the 1940-
119 2007 period, reaching a maximum of 32°C in summer and a minimum of -8°C in winter.
120 The soils have developed on schists and quartzites and have loam and silty loam
121 textures (average composition: 30–35% sand, 40–45% silt, 15–20% clay). The soil

122 depth is 45–150 cm and the available soil water content is between 100 and 150 mm.
123 The soil information was obtained from soil cartography at a scale of 1:100000 ([Alias](#)
124 [and Martinez Sanchez, 1988](#)). The dominant soils are xerorthents regosols and the
125 topography is characterized by steep slopes (>35%)

126 *Field data*

127 The forest survey was carried out in August 2014 using Field-Map instrumentation
128 (<http://www.fieldmap.cz/>): 27 field plots of 11 m radius were established using a
129 systematic, stratified sampling design. In each plot, we measured the diameter at breast
130 height (DBH; 1.3 m above ground level) and the total height of all trees with DBH \geq 10
131 cm. Two measurements, with a precision to the nearest millimeter, of DBH were made
132 at right angles with a tree caliper (Masser BT Caliper) and the arithmetic mean was
133 recorded. The total height was measured using a rangefinder and inclinometer (Laser
134 Technology ForestPro Laser), with a precision to the nearest centimeter. The structure
135 and silvicultural conditions were defined using the following stand parameters: number
136 of trees per hectare (N), basal area per hectare (G), mean arithmetic diameter (d_m) and
137 basal area median diameter (d_g), mean arithmetic heights (H_m), and Assmann dominant
138 height (H_o) ([Assmann, 1970](#)) (Table S1, Supporting material).

139 *LiDAR data and processing*

140 The LiDAR data were acquired on April 10, 2013 by the company Heliografics
141 Fotogrametria S.L. (Alicante, Spain), using an ALS50-II laser scanner (Leica-
142 Geosystems AG, Heerbrugg, Switzerland) with a laser repetition rate of 158.2 kHz, a
143 scan frequency of 100 Hz, illuminated footprint diameter of 0.32 cm, and an FOV of 12
144 degrees. The field was scanned by plane from a flight altitude of 3300 m.a.s.l. The ALS
145 data were acquired with a point density of 10.5 points/m². They were geo-referenced in
146 the European Terrestrial Reference System 1989 (ETRS89) coordinate system. The

147 planimetric coordinates (x and y) and ellipsoidal height values were computed for all
148 echoes. The time gap between the LiDAR data acquisition and the field data collection
149 is considered insignificant according to the annual height and diameter growth in the
150 study area ([Sánchez-Salguero et al., 2012](#)).

151 For this study, three different point densities were achieved, based on a random selection
152 of LiDAR pulses in a grid cell of 1 m², and were used in the segmentation process: 10.5,
153 4, and 0.5 pulses m⁻² (density). The forest-stand homogeneity and geographic distribution
154 make this statistic robust and informative. The minimum density, 0.5 pulses m⁻², exceeds
155 the minimum necessary to create the 3-m DEM required under the proposed USGS
156 specifications (USGS, National Spatial Program, 2009).

157 Recommendations mentioned in [Ruiz et al. \(2014\)](#) were followed to avoid the influence
158 of the Digital Terrain Model (DTM) on the final results. Therefore, separate filtering
159 processes for the three point clouds were produced, using an adapted algorithm from
160 [Kraus and Pfeifer \(1998\)](#), based on linear prediction. Next, these filtered returns were
161 used to generate DTMs with a spacing grid of 1, 2, and 5 m, respectively, for the pulse
162 densities mentioned above (10.5, 4, and 0.5 pulses m⁻²) ([Anderson et al., 2006](#)). In this
163 way, equal conditions for obtaining models are guaranteed, so that point clouds of
164 different pulse densities from different flight planning settings could be mock.

165 Next, the elevation values for the LiDAR data returns were normalized using the ground
166 surface model calculated above. We computed LiDAR metrics to support regression,
167 based on previous research by Næsset (2002). Metrics were calculated using FUSION
168 LIDAR Toolkit ([McGaughey, 2014](#)). In this study, a total of 43 metrics were extracted
169 from LiDAR pulses using the *gridmetric* command. The metrics were calculated from
170 the height distribution of laser returns and they were used as regressors in the statistical
171 analyses. To obtain a complete explanation of the FUSION tools, see [McGaughey](#)

172 [\(2014\)](#). The summary of the LiDAR metrics, with their corresponding descriptions, is
173 shown in Table S2, Supporting material.

174 *LiDAR data modeling*

175 We built predictive models with the forest structural attributes and metrics obtained
176 from the LiDAR data within each field plot. We computed multiple linear, power, and
177 exponential regressions corresponding to all possible combinations. Linearized
178 transformations were performed for the power and exponential regressions. Models
179 were evaluated following the criteria: (a) statistical significance (p value <0.05), (b)
180 minimum root-mean-squared error (RMSE), (c) minimum bias, (d) homoscedasticity,
181 performing a Breusch-Pagan test ([Herwartz, 2006](#)), and normal distribution of residuals,
182 verified with a Shapiro-Wilk test ([Mohd Razali and Wah Yap, 2011](#)), (e) parsimony
183 principle, (f) non-collinearity, when more than one variable were selected, and (g)
184 agreement with current biological knowledge ([Vandekerckhove et al., 2014](#)).

185 Specifically, in points (e) and (f), the variables included in the model were selected
186 through an exhaustive search using the Bayesian information criterion (BIC) method,
187 which performs all possible subset regressions and lists the models in ascending order
188 of BIC. The models with the lowest BIC were selected. In addition, multicollinearity
189 among the explanatory variables was verified with the condition index ([Belsley, 1991](#)).
190 All the variables selected in the models had a condition index lower than 30 and a p -
191 value of less than 5%. The accuracy of the models was assessed by performing a leave-
192 one-out cross-validation. The resulting models were applied to the whole extent of the
193 study area. For each model, LiDAR-based metrics were extracted from the whole point
194 clouds, using a pixel size equivalent to the field plot size, with FUSION ([McGaughey,](#)
195 [2014](#)).

196 R software ([R Core Team, 2015](#)) and the leaps package ([Thomas Lumley using Fortran](#)
197 [code by Alan Miller, 2009](#)) for variable selection were the tools employed.

198 *Segmentation methods applied for stand delineation*

199 Stands were segmented using two different algorithms that differed in their complexity.
200 The first one was based on multiresolution segmentation using eCognition software
201 ([Trimble, 2007](#)), involving a more complex method; the second one was based on mean
202 shift segmentation using Orpheo ToolBox software (OTB) ([CNES, 2013](#)) for QGIS
203 ([QGIS Development Team, 2009](#)) (Fig. 1), as a less complex approximation to compare
204 with the eCognition results. Both algorithms used basal area per hectare (G) and
205 Assmann dominant height (H_0) as silvicultural variables to identify and group LiDAR
206 data into a single stand. These structural variables were chosen based on our knowledge
207 of the forest in our study areas.

208 The multiresolution segmentation approach was applied as explained in [Hamilton et al.](#)
209 ([2007](#)), using an optimization procedure which locally minimizes the average
210 heterogeneity of image objects for a given resolution. Using multiresolution
211 segmentation, scale parameter determines the average size of the image objects, and
212 shape and form are determined by the input image layers which weights determine the
213 homogeneity ([Hamilton et al., 2007](#)). Segmentations for different scale parameters were
214 tested from a minimum value of two to an increasing number of parameters, until one
215 unique object resulted.

216 The second segmentation methodology used Orpheo ToolBox software (OTB), a non-
217 parametric density estimator based on Parzen window ([Babich and Camps, 1996](#)). It is
218 an adaptive gradient ascent method that works by discovering local maxima in the
219 feature-space, by moving the window towards them incrementally. With the local
220 maxima detected, the data points can be grouped into clusters ([Wu et al., 2013](#)). Three

221 parameters must be set: (1) the spatial radius, to define the neighborhood, (2) the range
222 radius, to define the interval in the spectral space, and (3) the minimum size of the
223 regions to keep after clustering.

224 *Validation of the segmentation method and stand map analysis*

225 The validation of an image segmentation is still a hard task ([Haralick and Shapiro,](#)
226 [1985](#)). An accurate segmentation is one which homogenizes regions according to a
227 specific characteristic and, at the same time, differentiates adjacent regions according to
228 the same characteristic ([Haralick and Shapiro, 1985](#)). Thus, segmentation should be
229 intra-region uniform and inter-region heterogeneous. From the statistics available to
230 validate image segmentation, Global Score - as defined in [Johnson and Xie \(2011\)](#) - has
231 been selected due to its simplicity of calculation and of understanding and its good
232 results.

233 [Johnson and Xie \(2011\)](#) suggested that the global intra-segment goodness measure
234 should be assessed as a variance weighted by each segment area on which each variance
235 is calculated divided by the total area: (Equation 1).

$$236 \quad wVar = \frac{\sum_{i=1}^n a_i \cdot v_i}{\sum_{i=1}^n a_i} \quad (1)$$

237 Where v_i is the variance and a_i is the area of the segment i . Segments with low
238 variance should be relatively homogeneous. A weighted variance was used so that large
239 segments had more impact on the global calculations than small ones.

240 As an inter-segment global goodness measure, Moran's Index ([Moran, 1950](#)) was used.
241 This is a measure of the spatial autocorrelation within the data and indicates the
242 statistical separation between equal spatial objects ([Kim et al., 2008](#)) (Equation 2). The
243 values of MI range from -1, indicating low spatial correlation and perfect dispersion,
244 which is desirable to the resulting segmentation, to +1, representing perfect correlation.
245 A value of zero indicates a random spatial pattern ([Cliff and Ord, 1981](#)).

$$MI = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2 (\sum_{i=j} \sum w_{ij})} \quad (2)$$

Where n is the total number of regions, w_{ij} is a measure of the spatial proximity, y_i is the mean spectral value of region i , and \bar{y} is the mean spectral value of the image. Each weighted w_{ij} is a measure of the spatial adjacent regions.

In order to compare MI and the weighted variance, they were both normalized (Equation 3).

$$\frac{(X - X_{min})}{X_{max} - X_{min}} \quad (3)$$

Where X_{min} and X_{max} are the minimum and maximum values of weighted variance or MI from all the calculations computed for every layer. Normalized values range between 0 and 1.

Finally, the Global Score is defined as the sum of the normalized weighted variance and the normalized MI (Equation 4).

$$GS = V_{norm} + MI_{norm} \quad (4)$$

Where V_{norm} is the normalized weighted variance and MI_{norm} the normalized MI. Therefore, in segmentation results, GS will range between 0 and 2, the closer to zero the better; that is, with a low weighted variance as a measure of intra-segment heterogeneity and a low MI value as a measure of inter-segment homogeneity. As there was more than one layer in the image, the GS values were averaged by the number of bands ([Johnson and Xie, 2011](#)).

Results

Assmann dominant height and basal area modeling

The stand H_0 and BA models based on regression methods provided R^2 values that ranged from 0.81 to 0.97 (Table 1), with a root-mean-squared error of the cross validation (RMSECV) below 1 m for dominant height and $6 \text{ m}^2 \text{ ha}^{-1}$ for basal area. The

270 models showed low values of bias in all cases, with consistency of the prediction
271 models. In all cases, the exponential and power function models performed significantly
272 better than the linear ones.

273 Values of R^2 greater than 0.95 and an MAE value of 0.40 m were obtained for H_o using
274 exponential models, with an RMSECV between 0.53 m (10 pulses m^{-2}), and 0.76 m (0.5
275 pulses m^{-2}) (Table 1). The basal area models had trends similar to those of the dominant
276 height models, with R^2 values higher than 0.84 and low values of RMSECV for the
277 exponential model (3.99 m for 10 pulses m^{-2} ; 5.05 m for 4 pulses m^{-2} ; and 2.90 m for
278 0.5 pulses m^{-2}) (Table 1).

279 Following the independent variable data selection, the models using height variables
280 (i.e., ElevP99, ElevCURT mean CUBE, and ElevP90 for H_o ; and ElevP50, ElevMAD
281 mode CUBE, and Elev mean for G), together with a descriptor for the density of the
282 forest canopy (CanopyReliefRatio), were the most successful models (Fig. S2,
283 Supporting Material).

284 Figure 2 presents the scatter plots of the best estimates of H_o and BA for the selected
285 regression model versus the LiDAR values for the densities 10, 4, and 0.5 pulses m^{-2} .
286 The predicted H_o was in near perfect agreement with the observed measurements, the R^2
287 value (> 0.95) being higher than that of the regression between the modeled and
288 observed G (> 0.84).

289 *eCognition multiresolution and OTB mean shift segmentation*

290 The response of Multiresolution Segmentation to forest-stand delineation, described by
291 the number of segments created, varied with the scale parameter (Tables 2 and 3). Using
292 the eCognition segmentation algorithm, a total of 1628 segments or stands were
293 delineated with a scale parameter of 2 (Moran's Index $MI_{norm} = 0.98$ for H_o and MI_{norm}
294 $= 1.00$ for G; average stand area = 0.25 ha), of which 11 were classified at the 36 scale

295 ($MI_{norm} = 0.06$ for H_o and $MI_{norm} = 0.22$ for G; average stand area = 37.18 ha) (Table 2).
296 A total of 221 segments or stands were delineated using Orpheo ToolBox (OTB), with a
297 spatial radius and a range radius of 2 and a minimum size region of 20 ($MI_{norm} = 0.00$
298 for H_o and $MI_{norm} = 0.88$ for G; average stand area = 1.85 ha), and 11 were classified at
299 a spatial radius of 8, a range radius of 2, and a minimum size region of 400 ($MI_{norm} =$
300 0.05 for H_o and $MI_{norm} = 0.79$ for G; average stand area = 37.18 ha) (Table 3).

301 General trends in the behavior of the method might be detected when the Global Score
302 is represented together with the normalized MI and the normalized variance in the range
303 parameter (Fig. 3). Figure 3a shows the Global Score, normalized weighted variance,
304 and normalized MI of segmentations assessed at different scale parameters for the
305 eCognition Multiresolution segmentation (10 pulses m^{-2}). Segmentations with low
306 normalized MI values had, at the same time, high normalized variance or *vice versa* due
307 to the characteristics of the definition of the variables (Equations 1, 2, 3, and 4).

308 The Global Scores for eCognition Multiresolution segmentation were better than the
309 results obtained with OTB Mean Shift segmentation (Fig. 4). When compared with
310 manual delineation, eCognition Multiresolution segmentations also performed better
311 (Fig. 4).

312 *LiDAR pulse density effects on the segmentation process*

313 As expected, the LiDAR pulse density affected stand delineation. We summarized these
314 effects by analyzing the normalized weighted variance and MI values. The best point
315 density for *P. sylvestris* stand delineation was 10 points m^{-2} , which provided the lowest
316 values of weighted variance and the highest values of normalized MI (Fig. 5). Both
317 methods predicted the dominant heights of the stands better than the basal area.

318 A one-way, between subjects ANOVA was conducted to compare the effects of pulse
319 densities of 10, 4, and 0.5 m^{-2} on plot-measured basal area and dominant height. There

320 were no statistically significant effects of pulse density on basal area when comparing
321 group means at the $p < 0.05$ level [$F_{2, 9595} = 1.15$, $p = 0.317$]. However, there was a
322 significant effect of pulse density on Assmann dominant height [$F_{2, 9595} = 5.69$, $p =$
323 0.003]. *Post hoc* comparisons using the Tukey HSD test indicated that the mean
324 dominant height of the segments for the pulse density of 0.5 ($M = 10.56$, $SD = 1.21$)
325 differed significantly from that of the other groups. However, the dominant heights in
326 the segmentation with a pulse density of 10 ($M = 10.48$, $SD = 1.07$) did not differ
327 significantly from those of the segmentation with a pulse density of 4 ($M = 10.47$, $SD =$
328 1.26).

329 **Discussion**

330 Our results show that, given the conditions set in this study, it is possible to generate
331 accurately *P. sylvestris* forest-stand segmentations using multiresolution or mean shift
332 segmentation methods, even with low-pulse-density LIDAR - which is an important
333 economic advantage for forest management. However, eCognition multiresolution
334 methods provided better results than OTB mean shift segmentation methods for stand
335 delineation based on dominant height and basal area estimations. Furthermore, the
336 influence of pulse density on the results was not statistically significant in basal area
337 calculations. However, for low pulse density, dominant height results could be affected.

338 *Assmann Dominant Height and Basal Area Modeling of the stand*

339 The performance of the Assmann dominant height and basal area models compares
340 favorably with the results of other studies in which stand height and G were modeled
341 using LiDAR data. The coefficient of determination for the final dominant height model
342 developed in this study ($R^2 > 0.95$) was in the range of previously reported values (0.82
343 to 0.98; [Means et al., 2000](#); [Næsset, 2002](#); [Coops et al., 2007](#); [Stone et al., 2011](#);
344 [González-Ferreiro et al., 2012](#); [González-Ferreiro et al., 2013](#); [Watt and Watt, 2013](#)).

345 The RMSE was similar to or lower than the values reported for other coniferous species
346 ([Means et al., 2000](#); [González-Ferreiro et al., 2013](#); [Watt and Watt, 2013](#)); in these
347 studies, it was also found that exponential functions performed better than linear
348 regression models.

349 Model predictions of basal area had a precision ($R^2 > 0.84$) comparable to that found in
350 similar studies within coniferous forests in boreal and temperate regions. The
351 coefficients of determination ranged from 0.62 to 0.94 for models predicting basal area
352 in coniferous forests in the United States of America ([Means et al., 2000](#)), Norway
353 ([Næsset, 2002](#); [Næsset et al., 2005](#)), and Denmark ([Nord-Larsen and Schumacher,
354 2012](#)). Additionally, the type of explanatory variable used might cause the main
355 differences. [González-Ferreiro et al. \(2013\)](#) generated models to estimate biomass,
356 which firstly was calculated as a combination of heights and diameter. However, in our
357 study, first order connections were assessed, as we worked directly with dominant
358 heights, considered as the combination of the tree heights and basal area - as a diameter
359 dependent variable, but not a combination of both variables.

360 Finally, the precision of dominant height and basal area models will be affected by the
361 errors in the generation of DTM, at the height at which the point clouds are normalized,
362 the errors of the sensor, and the pulse density ([Bollandsås et al., 2013](#)). However,
363 variations in LiDAR pulse density did not have a significant effect on the modeling
364 process. In fact, neither bias nor percentage error varied meaningfully ([González-
365 Ferreiro et al., 2013](#); [Ruiz et al., 2014](#)).

366 *LiDAR pulse density effects on prediction models*

367 The accuracy of the forest structure metrics slightly increased as a function of pulse
368 density (Table 1). The determination coefficients of dominant height and basal area did
369 not increase significantly from the lowest to the highest pulse density (i.e., from 0.5 to

370 10.5 pulses m²). Further, accuracy seemed to be related more to model selection than to
371 point density. For example, the accuracy of the exponential models for dominant height
372 remained approximately equal to its maximum ($R^2 > 0.94$). Similarly, the accuracy of
373 the power models for basal area rose steadily up to 4 pulses m⁻² ($R^2 = 0.90$), then
374 decreased to its lowest determination coefficient ($R^2 = 0.81$). Collectively, these results
375 indicate that beyond a certain density level (even as low as 0.5 points m²) accuracy does
376 not increase significantly. These are good examples of forest metrics that require a low
377 density to achieve reasonable accuracy, but do not benefit significantly from very high
378 LiDAR density ([Jakubowski et al., 2013](#)). Other authors reported similar results for the
379 modeling process, with high correlations between LiDAR metrics and forest inventory
380 attributes at the plot level, based on low-pulse-density LiDAR (<2 pulses m⁻²) ([Thomas](#)
381 [et al., 2006](#); [Næsset, 2009](#); [González-Ferreiro et al., 2013](#); [Ruiz et al., 2014](#)). Our results
382 were not as accurate, most likely due to the more complex study area. This indicates
383 that stand allometry requires a relatively lower number of LiDAR returns to be mapped
384 accurately.

385 *eCognition Multiresolution and Mean Shift OTB Segmentation*

386 We found that both eCognition and OTB segmentation could be automatically
387 segmented to produce spatial *P. sylvestris* stands and that an interpreter could label the
388 stands in a manner similar to traditional photography (Fig. 4). These results are
389 consistent with those achieved in other studies using LiDAR for automation of stand
390 delineation applied to forest inventory practices ([Mora et al., 2013](#); [Dechesne et al.,](#)
391 [2016](#)). LiDAR data have demonstrated the utility of within-stand forest structural
392 attributes (e.g., current dominant stand height and basal area) as a subset of attributes
393 required for characterization of forest stands.

394 The high Global Score, normalized weighted variance, and normalized MI of the
395 segmentations indicate that both eCognition Multiresolution and OTB segmentations
396 detected homogenous stands well, in concordance with previous work ([Espindola et al.,](#)
397 [2006](#); [Johnson and Xie, 2011](#)). Further, the eCognition Multiresolution segmentation at
398 10 pulses m^{-2} was the best scored segmentation, giving intra-region uniformity but inter-
399 region heterogeneity. The reason for that could be that the eCognition Multiresolution
400 segmentation algorithm was formulated to search for both intra-region homogeneity and
401 inter-region heterogeneity, while the Mean-Shift algorithm was designed to search only
402 for homogenous regions ([Baatz and Schäpe, 2000](#)). As deduced from Figure 3, the value
403 of the scale parameter had a direct effect on the number of polygons produced by the
404 resulting segmentation. In contrast, OTB Mean-Shift segmentation stands were usually
405 detected and delineated correctly, but the number of segments was lower and did not
406 always match those of the ground reference data.

407 Due to the relatively large size and homogeneity of the study area used, in comparison
408 with other studies ([Espindola et al., 2006](#); [Johnson and Xie, 2011](#)), we found values of
409 the Global Score that represented less than 5% of the difference in score between the
410 first minimum and the next minimum value, suggesting that there is no single best
411 segmentation but multiple ones; which can differ in the number and size of the
412 segments. Because the number of segments depends on the study area surface and the
413 size of the forest stands ([O'Hara and Nagel, 2013](#)), the best results could be identified as
414 the best minimum group of values of Global Scores, normalized weighted variance,
415 normalized MI, and number of segments ([Chen et al., 2014](#)). These values are located
416 around the crossing point of the curves of normalized weighted variance and normalized
417 MI (Figures 3 and 5).

418 Furthermore, there should be agreement between the segmentation with the best Global
419 Score, better normalized weighted variances, and better normalized MI and the resulting
420 number of segments, depending on the forest management objectives. For precision
421 silviculture, foresters may require very precise stand delineation regarding intra-region
422 variance with a high number of segments, despite the similarity of the segments to each
423 other; for example, segmentation at scale parameter 6 in Table 2, even when it is not the
424 one with the best score. However, in Mediterranean pine forests, where protection and
425 water management are the main silvicultural targets, a less precise stand delineation
426 would be demanded, including larger areas of high inter-region heterogeneity - with,
427 consequently, a low number of segments (e.g., segmentation at scale parameter 10 in
428 Table 2), although with higher intra-region variance ([Kim et al., 2008](#); [Johnson and Xie,
429 2011](#)).

430 The Mean Shift segmentation algorithm showed the additional complexity of using
431 three parameters (the range, spatial radius, and minimum region size) in the
432 segmentation process. In the analysis of the Mean Shift segmentations, more
433 unreliability of the validation system was detected. No obvious relationship between the
434 segment number created and the Global Score values could be observed (Fig. 5).
435 Consequently, selection of the best segmentation for forest-stand delineation using OTB
436 methodology is not as straightforward as previously thought. Given that the Mean Shift
437 algorithm has been proven as an adequate method for forest-stand delineation ([Wu et
438 al., 2013, 2014](#)), the disturbances in the detection of the best segmentation may come
439 from the validation system - which, we suggest, should be rethought for this kind of
440 technique with multi-dependent variables.

441 The question that then arises is: what benefit derives from the cost and effort of the
442 eCognition Multiresolution Segmentation approach presented herein, when the results

443 are similar to those obtained using OTB? Moreover, OTB implementation for stand
444 delineation would likely be simpler, more cost-effective, and of similar accuracy. While
445 dominant stand height and basal area do not constitute the entirety of an inventory, each
446 is amongst the more important of the suite of attributes that is generated. Dominant
447 stand height information is important for management purposes and is indicative of site
448 conditions, while basal area (related to volume or biomass) is key to forest management
449 (silvicultural treatments) and carbon-related considerations. A recommendation for the
450 future is that eCognition methodology (and related semi-automated processing
451 approaches) should remain focused on locations where precision silviculture inventory
452 programs persist. On the other hand, segmentation provided by OTB methodology,
453 offering less precision but also compatibility and similarity of stand delineation, should
454 be used in extensive silviculture (e.g., protection, climatic change adaptation, and
455 hydrologically-oriented silviculture). For areas that are not subject to regular
456 management or monitoring activities, it is possible that the more limited precision of
457 stand delineation provided by OTB will prove sufficient for many monitoring and
458 reporting needs. Thus, a stratification of activity may be possible based upon the
459 monitoring requirements associated with a given area.

460 *LiDAR pulse density effects on the segmentation process*

461 The segmentation algorithms were also influenced by the LiDAR pulse density. Our
462 results suggest that basal area is not affected by segmentations based on different
463 LiDAR pulse densities. In contrast, low pulse density affects the estimation of dominant
464 height. Segmentations using medium and high pulse densities do not appear to be
465 significantly affected with respect to dominant height results. However, it should be
466 noted that the values of the means and standard deviations for dominant height were
467 similar among the three pulse density approaches.

468 Overall, our results indicate that a very high LiDAR pulse density may not be necessary
469 to predict typical forest structure metrics at the plot scale or for stand delineation. These
470 findings are particularly important for land managers that need to survey a large area
471 with a specific forest metric and accuracy in mind. Our results, considered in terms of
472 cost, coverage, and density, can help guide this process. For example, if dominant basal
473 area is the most important metric to estimate at a reasonable accuracy level, it may be
474 sufficient to acquire LiDAR data at about 1 pulse m⁻². On the other hand, if it is critical
475 to derive the average dominant height with high accuracy, then it may be advisable to
476 use a much higher pulse density - between 2 and 4 pulses m⁻².

477 **Conclusions**

478 The objective of this study was to use LiDAR data segmentation to produce stand-level
479 predictions of dominant height and basal area as well as to use two different
480 segmentation techniques for stand delineation oriented to *Pinus sylvestris* forest
481 management in Mediterranean mountains. The use of LiDAR data provided a large
482 sample appropriate for model calibration and independent validation of attribute
483 predictions. We have demonstrated the utility of LiDAR data with regard to estimating
484 dominant stand height and basal area with an accuracy suitable for operational
485 activities. We have also noted the differences in stand delineation (number and form)
486 between two different segmentation algorithms (eCognition and OBT), using a semi-
487 automated methodology based on forestry attributes in a Mediterranean environment.
488 We did not find significant differences between high and low LiDAR pulse density,
489 neither in the creation of prediction models for dominant height and basal area nor in
490 the segmentation process. Nevertheless, for further assertions, more comparative studies
491 - varying the radius of the plot sample - should be carried out. The technique developed
492 in this project could be implemented to provide more precise data for forest

493 management. We propose that the approach shown here should be considered for stand
494 delineation in other large *Pinus* plantations in Mediterranean regions with similar
495 characteristics. Further, large-area, wall-to-wall characterization with a high level of
496 attribute detail is difficult to obtain, with sampling offering a practical, robust, and
497 reliable alternative. Future global forest inventory programs may benefit from
498 consideration of the framework and methods presented herein. Also, depending on the
499 location and attributes required, wall-to-wall mapping that integrates high-spatial-
500 resolution sensors (i.e., RapidEye or World-View) with LiDAR data may provide a
501 powerful opportunity for systematic and repeatable monitoring and reporting of
502 silvicultural activities.

503 **Acknowledgements**

504 We are grateful for support from the project ThermoLiDAR "A new tool for sustainable
505 forest management based on LiDAR (Laser Imaging Detection And Ranging) and
506 THERMAL data integration" FP7-SME-2012-315165) (Seventh Framework
507 Programme of the European Union) and the project LIFE13 ENV/ES/001384
508 "Development of technical guidelines for carbon sequestration and dynamization of
509 carbon compensation in forests". We also acknowledge financial and institutional
510 support from the University of Córdoba-Campus de Excelencia CEIA₃. We thank
511 Miguel Lara for his support during field data collection and his insights into the
512 development of the forest structural attributes from the LiDAR metrics.

513 **References**

514 Alias, L.J., Martinez Sanchez, J., 1988. Xerochrepts from the Orce and Maria Sierras
515 (Granada y Almeria). Anales de Edafologia y Agrobiologia (Spain).

516 Anderson, E.S., Thompson, J.A., Crouse, D.A., Austin, R.E., 2006. Horizontal resolution
517 and data density effects on remotely sensed LIDAR-based DEM. *Geoderma* 132, 406-
518 415.

519 Assmann, E., 1970. Principles of forest yield study.

520 Baatz, M., Schäpe, A., 2000. Multiresolution segmentation: an optimization approach
521 for high quality multi-scale image segmentation. *Angewandte Geographische*
522 *Informationsverarbeitung* XII 58, 12-23.

523 Babich, G.A., Camps, O.I., 1996. Weighted Parzen windows for pattern classification.
524 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18, 567-570.

525 Belsley, D., 1991. A Guide to using the collinearity diagnostics. *Computer Science in*
526 *Economics and Management* 4, 33-50.

527 Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R.,
528 van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic Object-
529 Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry*
530 *and Remote Sensing* 87, 180-191.

531 Bollandås, O.M., Maltamo, M., Gobakken, T., Næsset, E., 2013. Comparing parametric
532 and non-parametric modelling of diameter distributions on independent data using
533 airborne laser scanning in a boreal conifer forest. *Forestry* 86, 493-501.

534 Bouvier, M., Durrieu, S., Fournier, R.A., Renaud, J.-P., 2015. Generalizing predictive
535 models of forest inventory attributes using an area-based approach with airborne
536 LiDAR data. *Remote Sensing of Environment* 156, 322-334.

537 Burnett, C., Blaschke, T., 2003. A multi-scale segmentation/object relationship
538 modelling methodology for landscape analysis. *Ecological modelling* 168, 233-249.

539 Centeno, J.A.S., Antunes, A.F.B., Trevizan, S., Correa, F., 2003. Mapeamento de áreas
540 permeáveis usando uma metodologia orientada a regiões e imagens de alta resolução.
541 Revista Brasileira de Cartografia 1.

542 Cliff, A.D., Ord, J.K., 1981. Spatial processes: models & applications. Pion London.

543 CNES, 2013. Orfeo Toolbox in: Program, O.A. (Ed.), 3.18.1 ed.

544 Coops, N.C., Hilker, T., Wulder, M.A., St-Onge, B., Newnham, G., Siggins, A., Trofymow,
545 J.T., 2007. Estimating canopy structure of Douglas-fir forest stands from discrete-
546 return LiDAR. *Trees* 21, 295-310.

547 Costa, G., Feitosa, R.Q., Fonseca, L.M.G., Oliveira, D.A.B., Ferreira, R.S., Castejon, E.F.,
548 2010. Knowledge-based interpretation of remote sensing data with the InterIMAGE
549 system: major characteristics and recent developments, Proceedings of the 3rd
550 GEOBIA.

551 Chen, J.-Y., Lin, C.-H., Hsu, P.-C., Chen, C.-H., 2014. Point Cloud Encoding for 3D
552 Building Model Retrieval. *IEEE Transactions on Multimedia* 16, 337-345.

553 Chi, H.-M., Ersoy, O.K., 2005. A statistical self-organizing learning system for remote
554 sensing classification. *Geoscience and Remote Sensing, IEEE Transactions on* 43, 1890-
555 1900.

556 Dechesne, C., Mallet, C., Le Bris, A., Gouet, V., Hervieu, A., 2016. Forest Stand
557 Segmentation Using Airborne LIDAR Data and Very High Resolution Multispectral
558 Imagery. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and*
559 *Spatial Information Sciences*, 207-214.

560 Espindola, G., Câmara, G., Reis, I., Bins, L., Monteiro, A., 2006. Parameter selection for
561 region-growing image segmentation algorithms using spatial autocorrelation.
562 *International Journal of Remote Sensing* 27, 3035-3040.

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

566 González-Ferreiro, E. M., Miranda, D., Barreiro-Fernandez, L., Bujan, S., Garcia-
567 Gutierrez, J., & Dieguez-Aranda, U. (2013). Modelling stand biomass fractions in
568 Galician *Eucalyptus globulus* plantations by use of different LiDAR pulse densities.
569 *Forest Systems*, 22(3), 510-525.

570 Hamilton, R., Megown, K., Mellin, T., Fox, I., 2007. Guide to automated stand
571 delineation using image segmentation. RSAC-0094-RPT1. Salt Lake City, UT: USDA
572 Forest Service, Remote Sensing Applications Center.

573 Haralick, R.M., Shapiro, L.G., 1985. *Image Segmentation Techniques*, pp. 2-9.

574 Herwartz, H., 2006. Testing for random effects in panel data under cross sectional
575 error correlation—A bootstrap approach to the Breusch Pagan test. *Computational*
576 *statistics & data analysis* 50, 3567-3591.

577 Jakubowski, M.K., Guo, Q., Kelly, M., 2013. Tradeoffs between lidar pulse density and
578 forest measurement accuracy. *Remote Sensing of Environment* 130, 245-253.

579 James, G., Witten, D., Hastie, T., 2014. *An Introduction to Statistical Learning: With*
580 *Applications in R*. Taylor & Francis.

581 Johnson, B., Xie, Z., 2011. Unsupervised image segmentation evaluation and
582 refinement using a multi-scale approach. *ISPRS Journal of Photogrammetry and*
583 *Remote Sensing* 66, 473-483.

584 Kim, M., Madden, M., Warner, T., 2008. Estimation of optimal image object size for the
585 segmentation of forest stands with multispectral IKONOS imagery, *Object-based image*
586 *analysis*. Springer, pp. 291-307.

587 Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with
588 airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing* 53,
589 193-203.

590 Leckie, D.G., Gougeon, F.A., Walsworth, N., Paradine, D., 2003. Stand delineation and
591 composition estimation using semi-automated individual tree crown analysis. *Remote*
592 *Sensing of Environment* 85, 355-369.

593 Lopatin, J., Galleguillos, M., Fassnacht, F.E., Ceballos, A., Hernandez, J., 2015. Using
594 a Multistructural Object-Based LiDAR Approach to Estimate Vascular Plant Richness in
595 Mediterranean Forests With Complex Structure. *Geoscience and Remote Sensing*
596 *Letters, IEEE* 12, 1008-1012.

597 Magnussen, S., Næsset, E., Gobakken, T., Frazer, G., 2012. A fine-scale model for area-
598 based predictions of tree-size-related attributes derived from LiDAR canopy heights.
599 *Scandinavian Journal of Forest Research* 27, 312-322.

600 Magnusson, M., Fransson, J.E., Holmgren, J., 2007. Effects on estimation accuracy of
601 forest variables using different pulse density of laser data. *Forest Science* 53, 619-626.

602 Martín-Alcón, S., Coll, L., De Cáceres, M., Guitart, L., Cabré, M., Just, A., González-
603 Olabarría, J.R., 2015. Combining aerial LiDAR and multispectral imagery to assess
604 postfire regeneration types in a Mediterranean forest. *Canadian Journal of Forest*
605 *Research* 45, 856-866.

606 McGaughey, R., 2014. FUSION/LDV: software for LIDAR data analysis and visualization.
607 USDA Forest Service. Pacific Northwest Research Station.

608 McRoberts, R.E., Liknes, G.C., Domke, G.M., 2014. Using a remote sensing-based,
609 percent tree cover map to enhance forest inventory estimation. *Forest Ecology and*
610 *Management* 331, 12-18.

611 Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., Hendrix, C.J., 2000.
612 Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric*
613 *Engineering and Remote Sensing* 66, 1367-1372.

614 Mohd Razali, N., Wah Yap, B., 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-
615 Smirnov, Lilliefors, and Anderson-Darling test. *Journal of Statistical Modeling and*
616 *Analytics* 2, 21-33.

617 Mora, B., Wulder, M.A., White, J.C., Hobart, G., 2013. Modeling stand height, volume,
618 and biomass from very high spatial resolution satellite imagery and samples of
619 airborne LiDAR. *Remote Sensing* 5, 2308-2326.

620 Moran, P.A., 1950. Notes on continuous stochastic phenomena. *Biometrika*, 17-23.

621 Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser
622 using a practical two-stage procedure and field data. *Remote Sensing of Environment*
623 80, 88-99.

624 Næsset, E., 2009. Effects of different sensors, flying altitudes, and pulse repetition
625 frequencies on forest canopy metrics and biophysical stand properties derived from
626 small-footprint airborne laser data. *Remote Sensing of Environment* 113, 148-159.

627 Næsset, E., Bollandsås, O.M., Gobakken, T., 2005. Comparing regression methods in
628 estimation of biophysical properties of forest stands from two different inventories
629 using laser scanner data. *Remote Sensing of Environment* 94, 541-553.

630 Nord-Larsen, T., Schumacher, J., 2012. Estimation of forest resources from a country
631 wide laser scanning survey and national forest inventory data. *Remote Sensing of*
632 *Environment* 119, 148-157.

633 O'Hara, K.L., Nagel, L.M., 2013. The Stand: Revisiting a Central Concept in Forestry.
634 *Journal of Forestry* 111, 335-340.

635 Pekkarinen, A., 2004. Image segmentation in multi-source forest inventory. Finnish
636 Forest Research Institute, Vantaa Research Centre.

637 QGIS Development Team, 2009. QGIS Geographic Information System. Open Source
638 Geospatial Foundation.

639 R Core Team, 2015. R: A Language and Environment for Statistical Computing, 3.2.1 ed.
640 R Foundation for Statistical Computing, Vienna, Austria.

641 Radoux, J., Defourny, P., 2007. A quantitative assessment of boundaries in automated
642 forest stand delineation using very high resolution imagery. ForestSAT Special Issue
643 ForestSAT 2005 Conference "Operational tools in forestry using remote sensing
644 techniques" 110, 468-475.

645 Ruiz, L., Hermosilla, T., Mauro, F., Godino, M., 2014. Analysis of the Influence of Plot
646 Size and LiDAR Density on Forest Structure Attribute Estimates. Forests 5, 936-951.

647 Sagerer, G., Niemann, H., 2013. Semantic networks for understanding scenes. Springer
648 Science & Business Media.

649 Sánchez-Salguero, R., Navarro-Cerrillo, R.M., Swetnam, T.W., Zavala, M.A., 2012. Is
650 drought the main decline factor at the rear edge of Europe? The case of southern
651 Iberian pine plantations. Forest Ecology and Management 271, 158-169.

652 Stone, C., Penman, T.D., Turner, R., 2011. Determining an optimal model for processing
653 lidar data at the plot level: results for a Pinus radiata plantation in New South Wales,
654 Australia.

655 Thomas Lumley using Fortran code by Alan Miller, 2009. leaps: regression subset
656 selection.

657 Thomas, V., Treitz, P., McCaughey, J.H., Morrison, I., 2006. Mapping stand-level forest
658 biophysical variables for a mixedwood boreal forest using lidar: an examination of
659 scanning density. *Canadian Journal of Forest Research* 36, 34-47.

660 Trimble, 2007. *Definiens Developer*, 7 ed.

661 Vandekerckhove, J., Matzke, D., Wagenmakers, E.-J., 2014. Model comparison and the
662 principle of parsimony. *Oxford Handbook of Computational and Mathematical*
663 *Psychology*. Oxford University Press, Oxford.

664 Watt, P., Watt, M.S., 2013. Development of a national model of *Pinus radiata* stand
665 volume from LiDAR metrics for New Zealand. *International Journal of Remote Sensing*
666 34, 5892-5904.

667 Webb, A.R., 2003. *Statistical pattern recognition*. John Wiley & Sons.

668 Wu, Z., Heikkinen, V., Hauta-Kasari, M., Parkkinen, J., Tokola, T., 2013. Forest Stand
669 Delineation Using a Hybrid Segmentation Approach Based on Airborne Laser Scanning
670 Data, in: Kämäräinen, J.-K., Koskela, M. (Eds.), *Image Analysis*. Springer Berlin
671 Heidelberg, pp. 95-106.

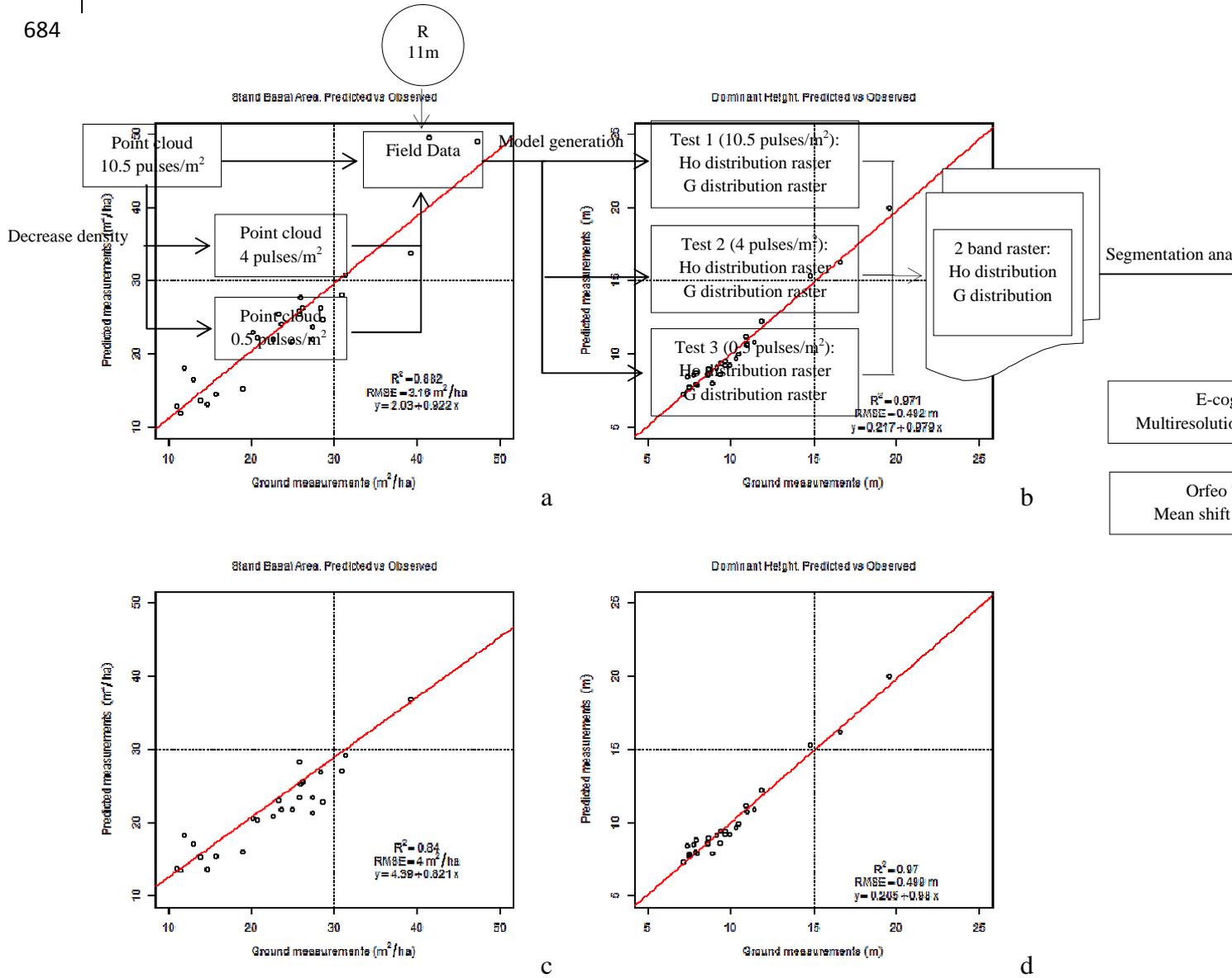
672 Wu, Z., Heikkinen, V., Hauta-Kasari, M., Parkkinen, J., Tokola, T., 2014. ALS data based
673 forest stand delineation with a coarse-to-fine segmentation approach, *Image and*
674 *Signal Processing (CISP)*, 2014 7th International Congress on. IEEE, pp. 547-552.

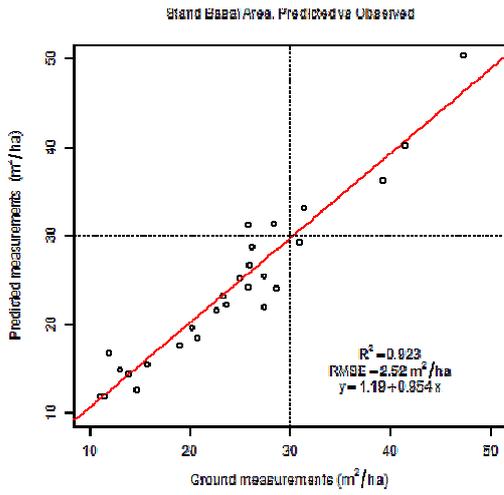
675 Yogamangalam, B.K.R., Karthikeyan, B., 2013. Segmentation Technique Comparison in
676 *Image Processing*. *IJET (ISSN: 0975-4024)* 5, 307-313.

677 Zhong, P., Zhang, P., Wang, R., 2008. Dynamic learning of SMLR for feature selection
678 and classification of hyperspectral data. *Geoscience and Remote Sensing Letters*, IEEE
679 5, 280-284.

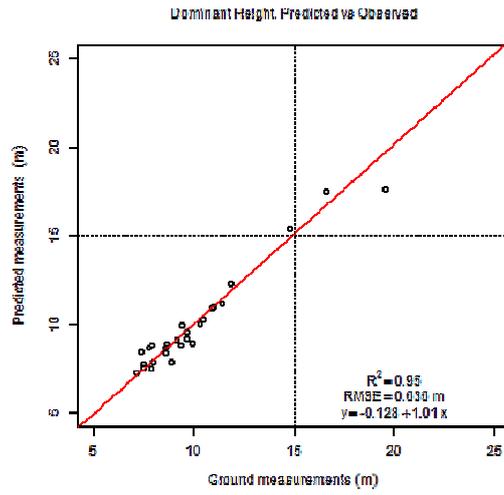
680

682 Figure 1: Flowchart of the modeling and image processing for proposed stand
 683 delimitation methodology based dominant height and basal area using LiDAR data.
 684



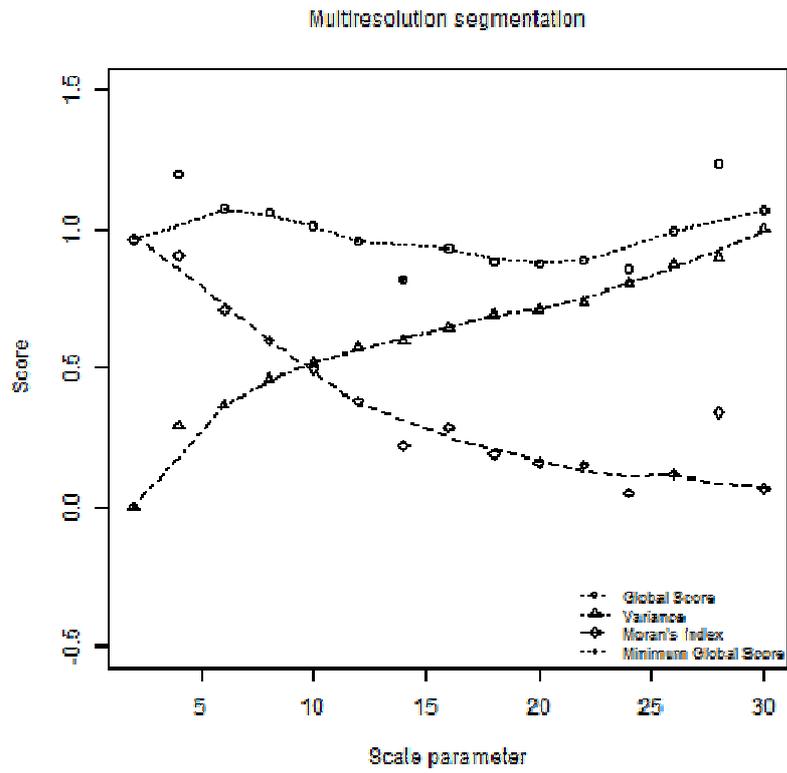


e

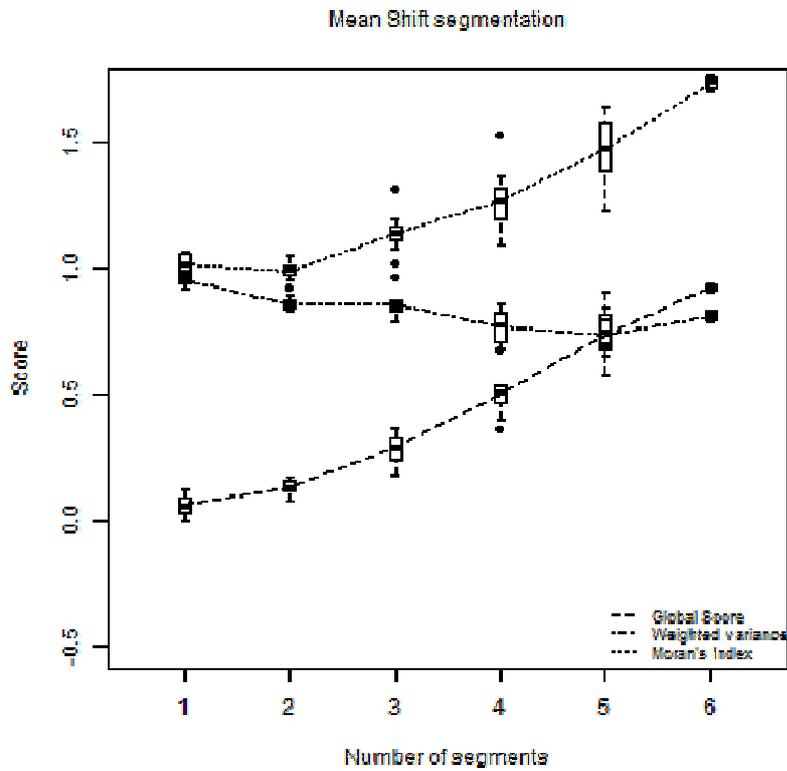


f

685 Figure 2: Scatter plot of the best estimated stand basal area (AB) (a, b, e) and dominate heights
 686 (H_0) (b, c, f) versus LiDAR values according to density pulses (10, 4 and 0.5 pulses/m² from
 687 upper to lower figure). 1/1 Red line.
 688
 689



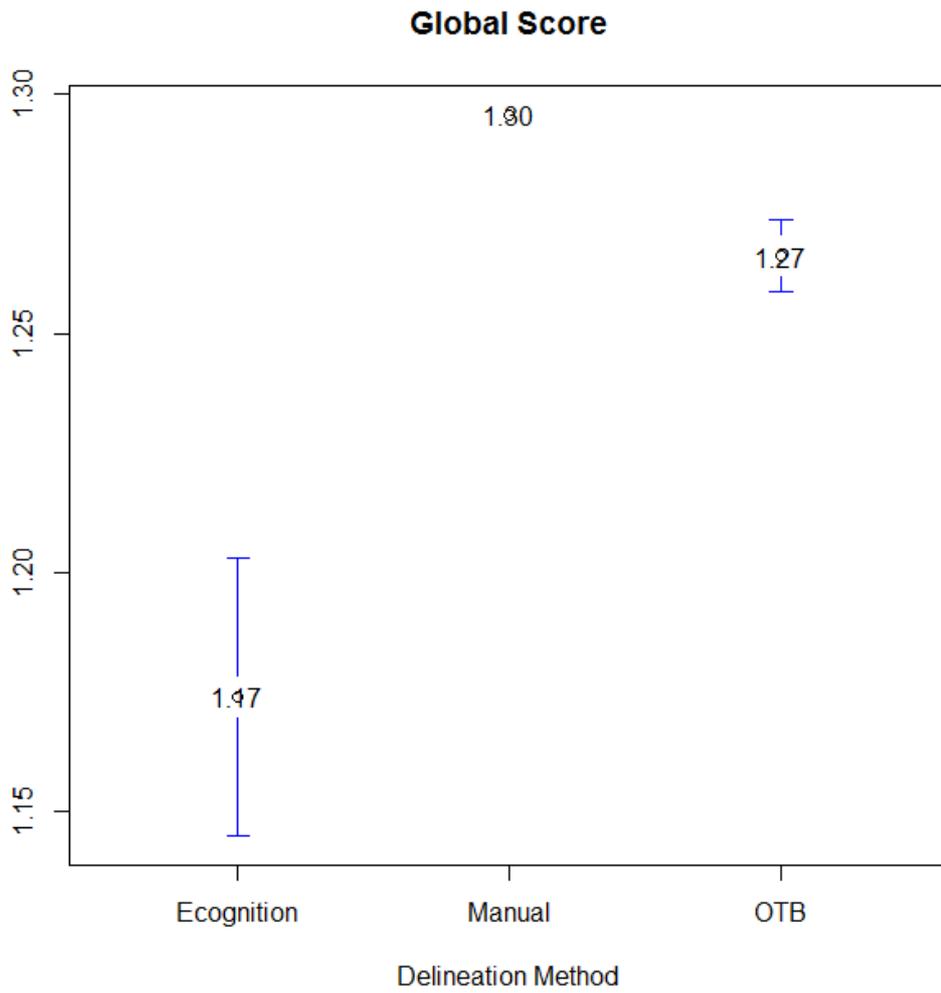
a



b

691 Figure 3: Global Score, normalized weighted variance and normalized Moran's Index
692 for the evaluation of the segmentations a) for the eCognition's Multiresolution

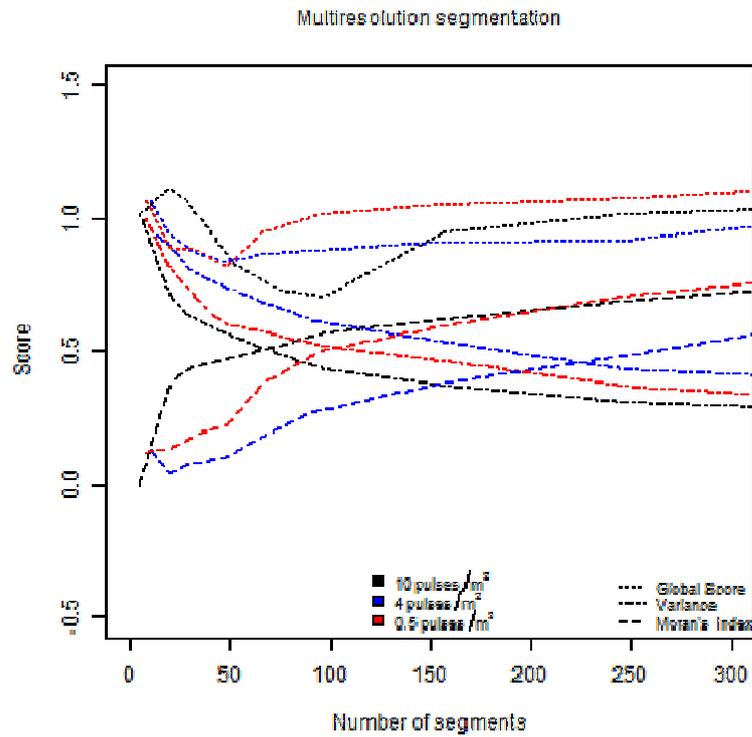
693 segmentation at 10 pulse density; b) for the OTB Mean Shift segmentation at 10 pulse
694 density.
695



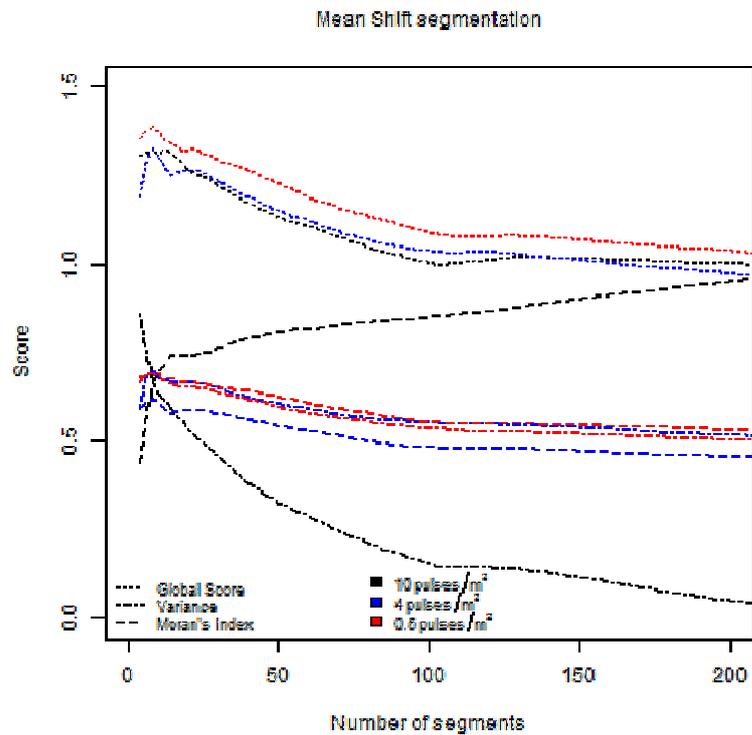
696

697 Figure 4: Mean Global Score values obtained using semi-automatic forest stand
698 delineation at 10 pulse·m⁻² density for Ecognition's Multiresolution segmentation, OTB
699 Mean-Shift segmentation and manual delineation

700



a

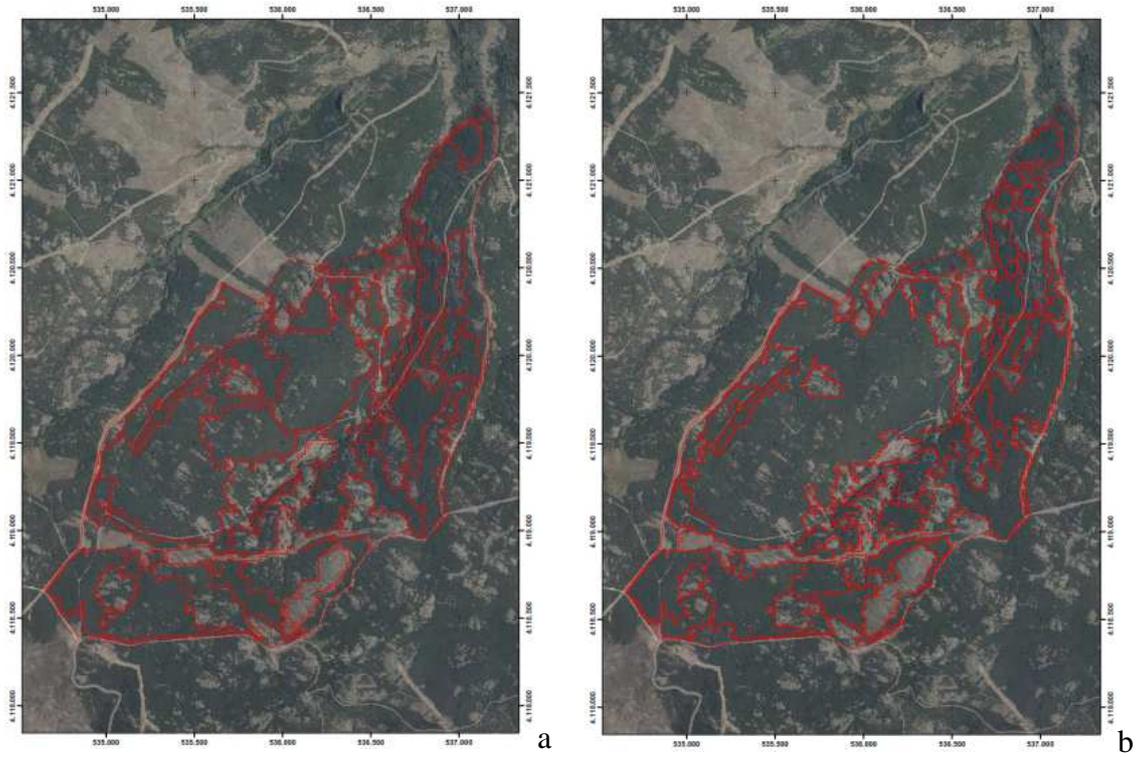


b

701 Figure 5: a) Normalized weighted variance, normalized Moran's Index and global
 702 scores for segmentations at 10, 4 and 0.5 pulse-m⁻² density with eCognition's
 703 Multiresolution Segmentation, b) Normalized weighted variance, normalized Moran's
 704 Index and global scores for segmentations at 10, 4 and 0.5 pulse-m⁻² density with Mean
 705 Shift Segmentation with OTB.

706

707 Figure 6. Maps obtained using semi-automatic forest stand delineation at 10 pulse·m⁻²
708 density a) Multiresolution segmentation; b) Mean-Shift segmentation



709

710

711

712

713

714
715
716
717

Table 1: Summary of the statistical criteria computed for evaluation the models for dominant height (H_0) and basal area (BA).

LiDAR pulse density (m^{-2})	Type of regression	Variable	R^2	BIAS	MAE	RMSE	RMSECV	%ERROR
<i>Dominant height (m)</i>								
10	Lineal	Ho	0.97	$-3.7e^{-5}$	0.35	0.59	0.63	5.65
10	Exponential	Ho	0.97	-0.01	0.40	0.49	0.53	5.35
10	Power	Ho	0.94	0.03	0.50	0.62	0.79	7.89
4	Lineal	Ho	0.96	$-1.6e^{-4}$	0.44	0.56	0.64	6.44
4	Exponential	Ho	0.97	-9.6e⁻³	0.40	0.50	0.54	5.54
4	Power	Ho	0.95	$-2.9e^{-2}$	0.47	0.62	0.73	6.23
0.5	Lineal	Ho	0.93	$3.6e^{-6}$	0.59	0.78	0.94	7.78
0.5	Exponential	Ho	0.95	2.0e⁻²	0.40	0.64	0.76	6.37
0.5	Power	Ho	0.94	$-2.6e^{-2}$	0.56	0.70	0.90	7.04
<i>Basal area ($m^2 ha^{-1}$)</i>								
10	Lineal	G	0.92	$2.3e^{-5}$	3.60	3.48	3.82	10.28
10	Exponential	G	0.88	-0.15	2.43	3.16	3.99	9.85
10	Power	G	0.81	-0.14	3.11	3.90	4.58	16.15
4	Lineal	G	0.92	$-4.3e^{-4}$	3.09	4.46	5.75	11.38
4	Exponential	G	0.84	-9.0e⁻²	2.95	4.00	5.05	10.60
4	Power	G	0.93	-0.13	2.98	4.46	5.67	16.19
0.5	Lineal	G	0.93	$-1.8e^{-4}$	2.70	2.81	2.96	10.20
0.5	Exponential	G	0.92	-9.1e⁻²	2.01	2.52	2.90	10.46
0.5	Power	G	0.87	-0.14	2.63	3.12	3.39	12.95

718 Mean Absolute Error (MAE), Mean Squared Error of Cross Validation (MSECV), Root Mean Squared
719 Error of Cross Validation (RMSECV) and percentage of error (%ERROR)

720
721

722

723 Table 2: Normalized variance (V_{norm}), normalized Moran's Index (MI_{norm}) and global
 724 scores (GS) for all scale-parameter segmentations with their resulting number of
 725 segments for the 10 pulse density approach. eCognition's Multiresolution segmentation
 726 approach.. Ordering indexes for two-band average values are shown in brackets.
 727 Minimum values Global Score are highlighted.
 728

Scale Parameter	Ho 10-pulse/m ² density band			G 10-pulse/m ₂ density band			Two-band average			Number of segments
	V_{norm}	MI_{norm}	GS	V_{norm}	MI_{norm}	GS	V_{norm}	MI_{norm}	GS	
2	0,00	0,98	0,98	0,00	1,00	1,00	0,00	0,99	0,99(11)	1628
4	0,35	1,00	1,35	0,29	0,68	0,98	0,32	0,84	1,17(18)	541
6	0,44	0,54	0,99	0,42	0,40	0,83	0,43	0,47	0,91(6)	246
8	0,50	0,47	0,97	0,52	0,35	0,88	0,51	0,41	0,92(7)	154
10	0,61	0,24	0,85	0,61	0,27	0,89	0,61	0,26	0,87(3)	92
12	0,65	0,23	0,89	0,69	0,20	0,90	0,67	0,22	0,90(5)	68
14	0,71	0,07	0,79	0,76	0,07	0,83	0,73	0,07	0,81(1)	47
16	0,78	0,13	0,91	0,83	0,00	0,83	0,80	0,06	0,87(2)	30
18	0,78	0,11	0,90	0,83	0,04	0,88	0,81	0,07	0,89(4)	29
20	0,85	0,03	0,88	0,92	0,05	0,97	0,88	0,04	0,93(9)	20
22	0,85	0,03	0,88	0,92	0,05	0,97	0,88	0,04	0,93(10)	20
24	0,87	0,02	0,89	0,93	0,03	0,96	0,90	0,02	0,93(8)	19
26	0,89	0,03	0,93	0,98	0,08	1,06	0,94	0,05	0,99(12)	16
28	0,90	0,02	0,92	0,98	0,11	1,10	0,94	0,07	1,01(14)	15
30	0,90	0,00	0,90	0,98	0,12	1,11	0,94	0,06	1,00(13)	14
32	0,92	0,03	0,95	0,99	0,24	1,23	0,95	0,13	1,09(15)	12
34	0,92	0,03	0,95	0,99	0,24	1,23	0,95	0,13	1,09(16)	12
36	1,00	0,06	1,06	1,00	0,22	1,22	1,00	0,14	1,14(17)	11

729

730

731

732 Table 3: Normalized variance (V_{norm}), normalized Moran's Index (MI_{norm}) and global
 733 scores (GS) for all spatial radius, rage radius and minimum size of the region
 734 segmentations with theirs resulting number of segments for the 10 pulse density
 735 approach. OTB Mean Shift segmentation approach. This is an extract of the 20 best
 736 global-scored segmentations out of 278. Ordering indexes for two-band average values
 737 are shown in brackets. Minimum values of Global Score are highlighted.

Spatial radius	Rage Radius	Min size of region	Ho 10-pulse/m ² density band			G 10-pulse/m ₂ density band			Two-band average			Number segments
			V_{norm}	MI_{norm}	GS	V_{norm}	MI_{norm}	GS	V_{norm}	MI_{norm}	GS	
2	2	20	0,00	0,00	0,00	0,97	0,88	1,85	0,48	0,44	0,92(4)	221
2	2	1000	0,55	0,31	0,87	0,27	0,48	0,76	0,41	0,40	0,819(2)	6
2	12	20	0,71	0,91	1,63	0,00	0,00	0,00	0,35	0,45	0,81(1)	5
4	2	20	0,03	0,03	0,06	0,95	0,88	1,84	0,49	0,45	0,95(7)	199
4	2	50	0,23	0,19	0,42	0,81	0,74	1,56	0,52	0,46	0,99(11)	97
6	2	20	0,06	0,00	0,06	0,95	0,88	1,83	0,50	0,44	0,95(6)	198
6	2	50	0,28	0,17	0,46	0,83	0,75	1,58	0,55	0,46	1,02(19)	86
8	2	20	0,10	0,07	0,18	1,00	0,85	1,85	0,55	0,46	1,01(18)	195
8	2	400	0,53	0,30	0,83	0,56	0,61	1,18	0,54	0,46	1,00(15)	11
8	4	20	0,14	0,05	0,19	0,89	0,79	1,69	0,52	0,42	0,94(5)	116
8	4	50	0,29	0,12	0,41	0,80	0,70	1,51	0,55	0,41	0,96(8)	62
10	2	20	0,09	0,03	0,12	0,96	0,91	1,88	0,53	0,47	1,00(14)	187
10	4	20	0,07	0,11	0,18	0,93	0,80	1,74	0,50	0,45	0,96(9)	118
10	4	1000	0,58	0,43	1,02	0,41	0,55	0,97	0,50	0,49	0,99(13)	5
12	2	20	0,08	0,08	0,16	0,99	0,88	1,88	0,53	0,48	1,02(20)	192
12	4	20	0,17	0,12	0,29	0,93	0,76	1,69	0,55	0,44	0,99(12)	108
14	2	20	0,01	0,04	0,06	0,92	0,84	1,76	0,47	0,44	0,91(3)	180
16	2	20	0,12	0,09	0,22	0,95	0,85	1,80	0,53	0,47	1,01(16)	181
16	4	20	0,16	0,15	0,32	0,89	0,72	1,61	0,53	0,43	0,96(10)	94
18	2	20	0,12	0,07	0,20	0,95	0,87	1,82	0,54	0,47	1,01(17)	179

738

739

740