RESEARCH PAPER

RegisTree: a registration algorithm to enhance forest inventory plot georeferencing

Maryem Fadili[1] · Jean-Pierre Renaud[2] · Jerome Bock[3] · Cédric Vega[1]

Abstract

• *Key message* The accuracy of remote sensing-based models of forest attributes could be improved by controlling the spatial registration of field and remote sensing data. We have demonstrated the potential of an algorithm matching plot-level field tree positions with lidar canopy height models and derived local maxima to achieve a precise registration automatically.

• *Context* The accuracy of remote sensing-based estimates of forest parameters depends on the quality of the spatial registration of the training data.

• *Aims* This study introduces an algorithm called RegisTree to correct field plot positions by matching a spatialized field tree height map with lidar canopy height models (CHMs).

• *Methods* RegisTree is based on a point (field positions) to surface (CHM) adjustment approach modified to ensure that at least one field tree position corresponds to CHM local maxima.

• *Results* RegisTree has been validated with respect to positioning errors and the performance of lidar-derived estimation of plot volume. Overall, RegisTree enabled to register field plots surveyed in a range of forest conditions with a precision of 1.5 m (± 1.23 m), but a higher performance for conifer plots, and a limited efficiency in homogeneous stands, having similar heights. Improved plot positions were found to have a limited impact on volume predictions under the range of tested conditions, with a gain up to 1.3%.

• *Conclusion* RegisTree could be used to improve the forest plot position from field surveys collected with low-grade GPS and to contribute to the development of processing chains of 3D remote sensing-based models of forest parameters.

Keywords Forest inventory · Lidar · Plot positioning · Registration algorithm · Forest parameter estimation

---

Handling Editor: Tuula Packalen and Klemens Schadauer

Contribution of the co-authors Cédric Vega conceived and designed the study. Maryem Fadili and Cédric Vega contributed to conception and coding of the method, data acquisition, processing, analysis and interpretation of the results, and drafting of the article. Jean-Pierre Renaud contributed to coding, data collection, and drafting of the article. Jérôme Bock contributed to data collection and proofreading of the article.

✉ Cédric Vega
  cedric.vega@ign.fr

[1] Laboratoire d'Inventaire Forestier, Institut National de l'Information Géographique et Forestière (IGN), 14 rue Girardet, 54000 Nancy, France

[2] Département RDI, Office National des Forêts (ONF), 8 allée de Longchamp, 54600 Villers lès Nancy, France

[3] Département RDI, Office National des Forêts (ONF), 42 quai Charles Roissart, 73026 Chambéry, France

INRA
SCIENCE & IMPACT

Springer

1 Introduction

Canopy height models (CHMs) derived from remote sensing methods, such as airborne laser scanning (ALS) and photogrammetric data, are recognized as a reliable and valuable source of information for forest parameter estimation and assessment (White et al. 2015a; Wulder 1998). They are also efficiently used in combination with plot-level field inventory data for both mapping forest parameters (Næsset 2007) and supporting the development of multisource national forest inventories (Tomppo et al. 2008; McRoberts and Tomppo 2007).

One of the key issues of combining remote sensing-based CHMs and field data from national forest inventory for assessing forest parameters relates to the spatial adjustment of both data sources (Nakajima 2016; Johnson et al. 2014). While 3D remote sensing data might provide a sub-metric accuracy in both the vertical and horizontal dimensions (Favorskaya and Jain 2017; Baltsavias 1999), the accuracy of field plots positioning is constrained by the interaction of the Global Navigation Satellite System (GNSS) signal with the canopy elements, resulting in accuracies of a few meters (∼ 5 m) with low-cost differential GNSS measurements (Ransom et al. 2010; Valbuena et al. 2010; Danskin et al. 2009; Wing and Eklund 2007). Overall, the positioning accuracy was found to vary predominantly with the GNSS device, the forest type and density, as well as the integration time (Andersen et al. 2009; Næsset and Jonmeister 2002). To a certain extent, those errors in field plot positioning affect the precision of forest inventory based on 3D remote sensing data (McRoberts et al. 2018; Gobakken and Næsset 2009) highlighting the need to enhance the spatial positioning of field inventory data (Johnson et al. 2014).

Different approaches have been proposed to solve this issue. They mainly consisted in matching algorithms aiming to spatially adjust height information from field inventory with a remote sensing-based CHM. State-of-the-art approaches rely on either surface-to-surface, point-to-surface, or point-to-point matching algorithms. An example of surface-to-surface approach was introduced by Olofsson et al. (2008). The method requires a segmentation of the CHM. Both field positions and CHM segments are converted into Gaussian surfaces using diameter at breast height (DBH) and height information. The match relies then on an image cross-correlation approach. In point-to-surface approaches, the field positions are directly matched with the CHM surface using correlation or error minimization approaches applied to heights or DBH (Monnet and Mermin 2014; Dorigo et al. 2010). Hauglin et al. (2014) proposed a point-to-point scoring approach matching ground-derived positions with local maxima (LM) extracted from the CHM (Hauglin et al. 2014; Dorigo et al. 2010; Olofsson et al. 2008). Each approach has pros and cons. Surface to surface depends on the method used to generate the surfaces.

The method proposed by Olofsson et al. (2008) further requires a segmentation of the CHM, which is sensitive to omission and commission errors, and the point cloud density. Conversely, point-to-surface approaches do not lead to any hypothesis that tree position on the CHM could lead to inappropriate matches. Finally, point-to-point approaches lead to strong assumptions on the spatial arrangement of trees. Korpela et al. (2007) used a network of photogrammetric measurements of tree tops to establish the absolute position of large field plots through triangulation and least squared adjustment, leading to decimeter-level accuracy. Those point-to-point methods are however sensitive to the quality of the LM, in terms of position, omission, and commission. For example, large tree crowns with multiple LM or slanting trees might lead to wrong match.

We suggest an algorithm that takes advantage of point-to-surface and point-to-point approaches. We hypothesize that point-to-surface approaches are more reliable than point-to-point ones, because LM detection is prone to commission and omission errors and can further be affected by horizontal displacements of the LM with respect to the measured field position at the tree base (Khosravipour et al. 2015; Vega et al. 2014). That being said, one drawback of point-to-surface approaches is their lack of priors, making it possible to match objects of different nature. In heterogeneous areas, a common mistake is matching small trees with the lower crown part of tall trees in a close neighborhood of the CHM. Our intuition is that this issue could be solved by taking the LM as spatial priors into account. Building on this idea, the main objective of this paper was to introduce a new point-to-surface field plot registration algorithm constraint by LM and to evaluate its performance in a range of forest conditions. The purpose of the performance evaluation was threefold: (1) analyzing the effect of algorithm parameters and forest structure and composition on the registration errors, (2) assessing the performance of the algorithm with respect to a state-of-the-art algorithm (Monnet and Mermin 2014), and (3) evaluating the impact of the plot registration accuracy on the performance of ALS-derived prediction model of forest volume.

2 Material and methods

2.1 Study sites

The study was conducted over four broadleaved dominated forests (Bure, Saint-Gobain, Compiègne, Darney) and two conifer-dominated forests (Vosges, Aillon). The data collection for Bure, Vosges, and Aillon was part of the research project Foresee (http://foresee.fcba.fr, 2010–2014) (Dataset 1). Over Saint-Gobain, Compiègne, and Darney, the data was acquired by Office National des Forêts (ONF) for management purposes (Dataset 2).

The four broadleaved forests are located on a flat to rolling terrain, with mean terrain slope varying from 2.9% (± 1.6%) for Compiègne to 3.2% (± 7.3%) for Bure (Table 1). The forest of Bure (48.53° N, 5.37° E) is multilayered and intensively managed, covering an area of around 50 km^2, dominated by European beech trees (*Fagus sylvatica* L.), European hornbeams (*Carpinus betulus* L.), and sycamore maple trees (*Acer pseudoplatanus* L.). The forest of Compiègne (49.40° N, 2.90° E) covers 143 km^2 and consists mainly in regular stands of European beech and common oak (*Quercus robur* L.). Saint-Gobain's site (49.59° N, 3.37° E) covers 8.5 km^2 and is made of regular stands of oaks and European beech along with coppices of European hornbeams. The site of Darney (48°08 N, 6.04° E, 150 km^2) is also dominated by oaks, mixed with beeches, hornbeams, silver firs (*Abies alba* Mill.), and Norway spruces (*Picea abies* Karst.). The forest structure is regular, forming a mosaic of age classes at the landscape level. The Vosges forest (48.01° N, 7.13° E) ranges between 200 and 1425 m elevation and extends over a hilly relief, characterized by a mean slope of 44.4% (± 19.4%). The site is dominated by silver fir, beech, and Norway spruce, in a variety of forest structures. Finally, the mountain mixed forest of Aillon, located in the Combe d'Aillon area in Eastern France (45.61° N, 6.08° E), covers an area of 25 km^2. It is an old growth heterogeneous and uneven-aged forest, growing between 1055 and 1432 m elevation, dominated by silver fir and Norway spruce, mixed with European beech and sycamore maple. The mean terrain slope is 44.3% (± 11.6%).

2.2 Field and lidar surveys

A total of 246 plots of 15 m radius (706.9 m2) were inventoried over the different sites: 165 over the broadleaved forests of the Bure forest (41 plots surveyed in February 2010), the Compiègne forest (33 plots, surveyed during winter 2014), the Saint-Gobain forest (43 plots, winter 2015), and the Darney forest (48 plots, winter 2014) and 81 plots were

surveyed over coniferous forests in the Vosges (53 plots, during winter 2011) and the Aillon (28 plots, in April–May 2011). All trees with a DBH equal or above 17.5 cm were mapped using distance and angle measurements from the plot center. A threshold of 7.5 cm was used for the Bure forest. Along with tree position, DBH and species were recorded for each tree. Tree height was measured either for each tree of the plots (Aillon) or for the six trees having the largest diameters (dominant trees) (other sites). Over the Bure forest, both height measurement protocols were applied on distinct plot subsets. DBHs were deduced from the circumference measured using a tape and both distances from the plot center and tree heights were recorded with either a digital hypsometer or a laser range finder (Haglöf, Sweden; Laser Technology, USA). The plot volume (AGV, m^3 ha^{-1}) was derived from site-specific allometric equations based on tree species, DBH, and height (Deleuze et al. 2014). Overall, the sites dominated by conifers show higher stem densities and stocks. The mean plot density was 351 ± 191 trees per hectare over the Aillon and reached 465 ± 206 stems per hectare over the Vosges. For broadleaved forests, it varied from 187 ± 98 stems per hectare for Saint-Gobain to 243 ± 89 over Darney. The mean plot volume was maximal over Aillon (727.1 ± 233.3 m^3 ha^{-1}) and minimum over Saint-Gobain (278.3 ± 189.9 m^3 ha^{-1}) (Table 1).

The lidar data was acquired by different providers. Over the Bure and the Aillon forests, the data was acquired by Sintegra (France) using a Riegl LMS-Q560 flown at 550 m above ground level (AGL) using a scan angle of 29.5° and a 50% line overlap, leading to an average point density of 12 points m^{-2}. The Bure was flown over in October 2010, i.e., 8 months after the field survey. The mountain mixed forest of Aillon was flown over 3 months after the field survey, during August 2011. The broadleaved forest of Darney was flown over by the same provider in March 2014, during the field measurement period, using a Riegl LMS-Q560i flown at 450 m AGL. The average point density was 21 points per

Table 1 Mean and standard deviation (SD) of the field inventory data for stem density, mean diameter at breast height (DBH), basal area (BA), height of the dominant trees (*H*), and aboveground volume (AGV)

	Dataset 1			Dataset 2		
	Bure	Aillon	Vosges	Saint-Gobain	Compiègne	Darney
Plot number	41	28	53	43	33	48
Terrain slope (%)	3.2 ± 7.3	44.3 ± 11.7	44.47 ± 19.4	3.06 ± 5.2	2.9 ± 1.6	2.7 ± 7.3
Stem density (tree ha^{-1})	224 ± 174	351 ± 191	465 ± 206	187 ± 98	237 ± 76	243 ± 89
DBH (cm)	30.2 ± 9.3	39.6 ± 9.3	31.0 ± 7.8	32.1 ± 8.1	33.4 ± 9.0	32.5 ± 7.5
BA (m^2 ha^{-1})	22.3 ± 25.6	45.1 ± 12.4	37.6 ± 16.6	18.5 ± 11.6	21.4 ± 8.6	20.9 ± 6.6
H (m)	20.1 ± 5.0	25.1 ± 3.4	23.9 ± 6.1	23.9 ± 5.8	25.3 ± 4.5	27.2 ± 4.0
AGV (m^3 ha^{-1})	365.6 ± 498.3	727.1 ± 233.3	488 ± 254.3	278.3 ± 189.9	325.5 ± 170.2	320.6 ± 129.0

Only trees having a DBH greater or equal to 17.5 cm were considered. Terrain slope was derived from the lidar DTM

squared meters (pts. m^{-2}). The site of Saint-Gobain was flown over by ECARTIP in September 2014. Compiègne was flown over by AERODATA (France) in March 2015. Both flights involved the same Riegl LMS-Q680i sensor flown at 530 m AGL during the field inventory periods. The average point density was 18 and 24 pts. m^{-2}, respectively. Finally, the Vosges area was surveyed by the Institut National de l'Information Géographique et Forestière (IGN), in March and April 2011, 8 to 10 months before the field survey. An Optech ALTM3100 sensor was used at 1500 m AGL with a 71-kHz scan frequency and a 16° scan angle. The average point density was 2.6 pts. m^{-2}.

2.3 Data preparation

For the Riegl datasets, the waveform processing was carried out by the data providers using RiANALYZE ©Software (Riegl, Austria). For all the datasets, the ground points were classified by the data provider, using the TIN-iterative algorithm (Axelsson 2000) implemented within TerraScan software (Terrasolid, http://www.terrasolid.fi/en/products/terrascan), and manually controlled. The resulting triangulated irregular network was then converted into a 1-m cell grid to obtain the final digital terrain. Mean plot-level terrain slope was generated from the digital terrain model (DTM) using R and the Raster package (https://www.R-project.org/, last consulted January10, 2019), by considering an 8-pixel neighborhood. The points above the DTM were processed to generate a CHM using a four-step procedure: (1) a 1-m resolution digital surface model (DSM) is generated by rasterizing the point cloud and selecting the maximum elevation value of the points within each cell, (2) an inverse distance weighting (IDW) method was used to compute empty cell values by interpolating the point values selected at the previous stage, (3) a hole-filling algorithm (Véga and Durrieu 2011) was applied to the resulting DSM to remove outliers, and (4) a CHM was finally computed by subtracting the DTM from the DSM.

In the field, plots were georeferenced using a differential GNSS device (Trimble GeoXT, Trimble, USA), sometimes in association with a total station (Bure) (Leica, Switzerland; Trimble, USA). The position of all plots was refined in the laboratory by an experienced operator who manually matched the plot-level field tree map with the lidar data. This was done by visually matching the field tree map (x, y, and height) with the ALS point cloud in 3D. The operator only considered plots with a positioning accuracy below 1 m. Note that for the Foresee data (Bure, Aillon, Vosges), the task was performed using the CHM data and mostly consisted in matching field tree position with the lidar apparent tree crowns. For the three other sites (Darney, Compiègne, Saint-Gobain), the high point density and leaf-off conditions allow us to match the field tree map directly with the tree trunks visible in the point cloud.

While such an approach is expected to provide more accurate field plot positions, it might lead to larger errors with respect to automatic matching and Foresee data, as it relies on the CHM and not on the trunks.

2.4 RegisTree registration algorithm

Let us define N as the number of field-measured trees within a plot and n as the subset of trees the height of which height has been collected ($n \leq N$). The subset n of (x, y, h) triplets defines the field tree map (FTM) used by the registration algorithm. The algorithm fits the FTM with a CHM in a defined search distance (S_d) in x and y directions.

The principles of RegisTree are presented in Fig. 1. The algorithm fits an FTM with a CHM in a defined search distance (S_d) in x and y directions, ensuring that at least one of the field trees is associated with a LM of the CHM. To account for spatial uncertainties in matching a field position defined at the DBH position and a remote sensing position defined at the tree top, an iterative procedure was implemented, constrained by two additional user-defined parameters: N_{ref} as the number of reference trees corresponding to a subset of the field trees (FTM) and N_c as the number of candidate positions selected for each reference tree. The selected N_{ref} trees could vary from 1 to n and be processed in a height descending order to give more weight to the taller trees in the plots. To avoid inconsistent matches, only the dominant and co-dominant trees are considered in the field tree map.

For each reference tree, the FTM is shifted throughout the search distance S_d, ensuring that the considered reference tree is located over a LM ($LM \in S_d$). Two statistics are computed for every candidate position: the weighted root mean square error (wRMSE) (Eq. 1) and the weighted correlation (wCorr) (Eq. 2) (Pinto da Costa 2015) as follows:

$$wRMSE = \sqrt{\sum_{i=1}^{n} \left(H_{FTM,i} - H_{CHM,i}\right)^2 \times W_i} \qquad (1)$$

$$wCorr(H_{CHM}, W \times H_{FTM})$$

$$= \frac{cov(H_{CHM}, W \times H_{FTM})}{\sqrt{cov(H_{CHM}, W \times H_{CHM}) \times cov(H_{FTM}, W \times H_{FTM})}} \qquad (2)$$

$$cov(x, W \times y) = \frac{\sum_{i=1}^{n} W_i \times \left(x_i - \overline{x_w}\right)\left(y_i - \overline{y_w}\right)}{\sum_{i=1}^{n} W_i} \qquad (3)$$

where $H_{FTM,i}$ is the field measured height of the tree i and $H_{CHM,i}$ is its height in the CHM; $cov(x, w \times y)$ is defined according to Eq. 3 with $\overline{x_w}$ and $\overline{y_w}$ the weighted means, and $W = (W_1, W_2, \ldots, W_n)$ is the vector of weight accounting for the status of dominance of the tree (see below).
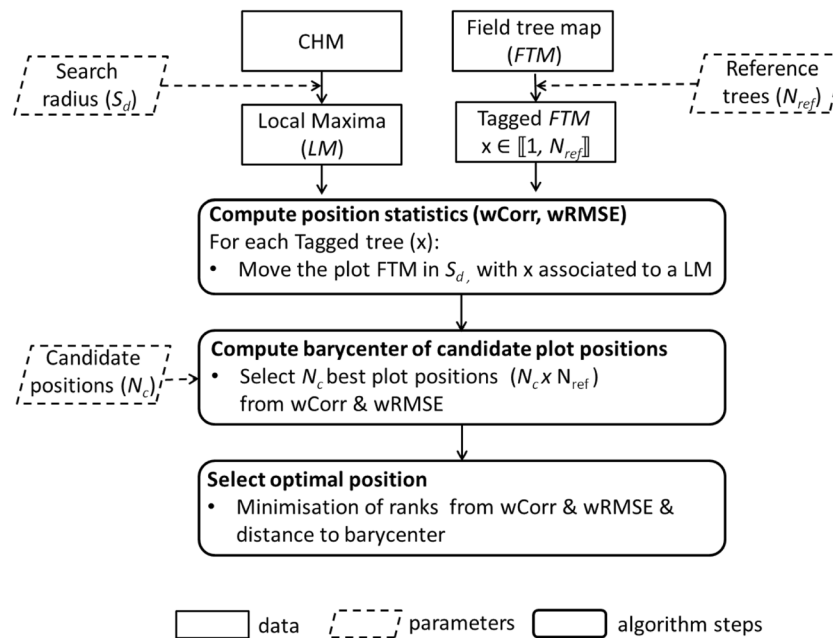
Fig. 1 Principles of RegisTree algorithm

Indeed, for each reference tree, the number of candidate positions computed equals the number of LM detected in the search area. The N_c best positions are then selected based on the minimization of wRMSE and maximization of wCorr. As stated above, such a selection of multiple candidate positions is done to avoid uncertainties linked to the unsupervised detection of local maxima as well as inaccurate field tree positions due to operating errors or cases of bended trees. The $N_c \times N_{ref}$ positions obtained are merged and a third statistic is computed as the distance between each candidate position and their barycenter. The optimal position is selected through a minimization of rank using three statistics.

RegisTree enables the estimation of a vector of weights W in the absence of information about the status of tree dominance, in order to give more weight to the largest trees. The approach is not an attempt to model crown visible area with respect to plot structure and composition but rather a way to discard understory trees from the matching process. This is done by estimating the portion of the tree crown visible on the CHM according to the FTM. Crown diameters of 2, 3, and 5 m are respectively assigned to small (7.5–22.5 cm), medium (22.5–47.5 cm), and large (47.5–67.5 cm) tree diameters classes according to the French NFI standards. These crown radii correspond to 1/3 of plot radius where the tree is measured by the NFI. The portion of the tree crown expected to be visible on the CHM is calculated by subtracting the parts that intersect trees of higher class with respect to the distance between trees in the FTM. If two trees of the same class overlap, their visible crown surface is considered unchanged. Finally, the remaining surface is normalized by dividing it by the sum of all the visible—i.e., non-null—crown surfaces in the plot. Null surfaces define

suppressed trees which are removed from the FTM. W was computed using this approach in the framework of this study.

2.5 Accuracy assessment

2.5.1 Sensitivity analysis and benchmarking

A sensitivity analysis was conducted to test the impact of (1) the number of reference trees (N_{ref}) and candidate positions (N_c), (2) the initial position, and (3) the size of the FTM. The major results were further analyzed with respect to the forest type and the accuracy requirements of the positioning. The resolution of the CHM was not considered in the study. Preliminary experiments indicated that while a higher CHM resolution (i.e., 0.5 m) significantly increased the amount of LM, thus introducing noise, a lower resolution (i.e., 2.0 m) leads to a high number of omissions in the LM. Despite its importance, the search area was also discarded from the sensitivity analysis. Not only did we consider that a 10-m search distance in both x and y directions was a good compromise to account for relatively large positioning errors while minimizing the number of LM to test but also reducing the search distance to the expected GNSS measurement error would result in an improved positioning accuracy. In the following, the positioning accuracy was computed as the difference between the reference position obtained by the operator and that of the algorithm.

The quality of the positioning was investigated by shifting the initial plot coordinates in both x and y directions with the following values $\{(1, 1), (-1, -1), (2, -2), (-2, 2), (-3, 3), (3, -3), (-4, 4), (4, -4), (0, 5), (0, -5), (5, 0), (-5, 0)\}$ and by

measuring the frequency at which identical resulting final positions occur.

The effect of N_{ref} and N_c on the accuracy of the registration was tested in the following area: $\{N_{ref}, N_c\} \in [\![1, 6]\!]^2$. N_{ref} was limited to 6 as only the six dominant trees were measured in five out of the six study sites. The analysis was extended to the Aillon forest and a subset of plot of that of Bure, where all the trees were measured for height. For this subset of plots, N_{ref} values of 6, 9, 12, 15, 18 (Aillon only), and above (i.e., full set of tree) were considered.

The effect of forest type on the accuracy of the result was investigated by analyzing the previous experiments according to the following three classes: broadleaved, mixed, and coniferous plots. Furthermore, to account for the effect of plot structure and topography, the positioning accuracy was analyzed against various forest attributes and the mean plot slope, respectively.

Finally, in order to test the performance of the algorithm with respect to a state-of-the art algorithm, the performance of the algorithm was compared with the one achieved with a point-to-surface approach (Monnet and Mermin 2014). Both algorithms were coded into R software (Version x64 3.4.1).

2.5.2 Impact on forest volume prediction models

The performance of the algorithm was also tested with respect to lidar-based models of forest volume. Two models were built using both the initial and the registered coordinates. As the initial coordinates were derived from differential GNSS data, thus having a good accuracy, a third model was built after shifting the initial coordinates by 5 m in the northern direction. This test was performed to emulate the positional accuracy of a consumer level GNSS device and assess the effect of the local forest structure on the performance of the models.

For the three sets of plot positions, the lidar CHM was clipped to the plot surface and various metrics were computed. These include maximum height (H_{max}), height of the 90th percentile (H_{90}), inner (V_{in}) and outer (V_{out}) canopy volumes (Véga et al. 2016), gap area (G_a) and its corresponding inner (VG_{in}) and outer (VG_{out}) volumes, variance (H_{var}) and standard deviation (H_{sd}) of CHM height, and Rumple index (R_i) (Kane et al. 2010). The forest type (i.e., deciduous, conifers, or mixed) (F_{type}) derived from field surveys was also considered. From this set of metrics, predictive models of plot volume were built using a best-subset linear regression approach. The models taken on were the ones having the minimum Bayesian Information Criteria (BIC) (Neter et al. 1985), with a maximum variance inflation factor (VIF) lower than 5 (O'Brien 2007), all predictor variables being significant (at a level of 0.01), and a non-significant normality test on model residuals (Shapiro test at a level of 0.05). The accuracy of the predictive models was assessed by performing a leave-one-out cross-validation (LOOCV) (Picard and Cook 1984) and computing corresponding root mean square error (RMSEcv) and

adjusted determination coefficient (Adj. R^2). The statistical analyses were performed with R using ClustOfVar libraries, leaps, car, and DAAG (https://cran.r-project.org/).

3 Results

3.1 Sensitivity analysis

The analyses were conducted using a search distance (S_d) of 10 m in both x and y directions. Figure 2 shows the results of RegisTree's sensitivity analysis regarding both the number of reference trees (N_{ref}) and the number of candidate positions (N_c), for the whole dataset and according to forest type. Overall, the positioning error decreased when N_{ref} and N_c increased. On the range of tested parameters' values, the best results for the whole dataset were obtained with $N_{ref} = 5$ and $N_c = 6$. The same optimum was achieved for both conifer and broadleaved plots. The trend over the mixed plots seemed to be less clear. The minimum error was obtained with $N_{ref} = 5$ and $N_c = 1$. Despite this, the results appeared to be stable in all cases with values of both N_{ref} and N_c over 4, highlighting the robustness of the algorithm to forest type.

Figure 3 shows an example of registration results achieved using the optimal set of parameter values ($N_{ref} = 5$, $N_c = 6$). Using an *FTM* of 6 trees, the algorithm achieved registration errors of 1.56 m (± 1.32 m SD), with minimum and maximum values of 0.08 and 9.06 m, respectively (Table 2). The best results were obtained with mixed plots, with a mean error of 1.05 m (± 0.91 m SD) and a maximum error of 4.18 m. The mean error was 1.06 m (± 1.26 m SD) in conifer plots and reached 1.53 m (± 1.33 m SD) in broadleaved plots. The results for Dataset 1 outperformed those of Dataset 2, with mean errors of 1.37 m (± 1.40 m SD) and 1.74 m (± 1.23 m SD). However, this was expected as Dataset 1 contains more than 90% of the conifer plots. Accordingly, the algorithm performed better over the conifer forests of Aillon, showing an error of 0.69 m (± 0.47 m SD).

The effect of the initial position on the robustness of RegisTree is presented in Fig. 4. Overall, higher resilience of positions, defined as the convergence of the algorithm towards the same position, produces lower errors and lower extreme error values (Fig. 4a). Errors appear to have stabilized when the number of identical positions is equal or greater than 4 (i.e., 30% of the initial positions tested). That said, the correlation between the registration error and the resilience of the output remains limited (Pearson correlation coefficient of − 0.36). Figure 4b shows that the registration error remained largely independent from the robustness of the algorithm to the initial position.

The effect of the number of trees on the registration error was investigated for the Aillon and Bure forests, in which all trees were surveyed for height (Table 3). For the coniferous plots of Aillon, *FTM* of 6, 9, 12, 15, 18, and more than 18 trees
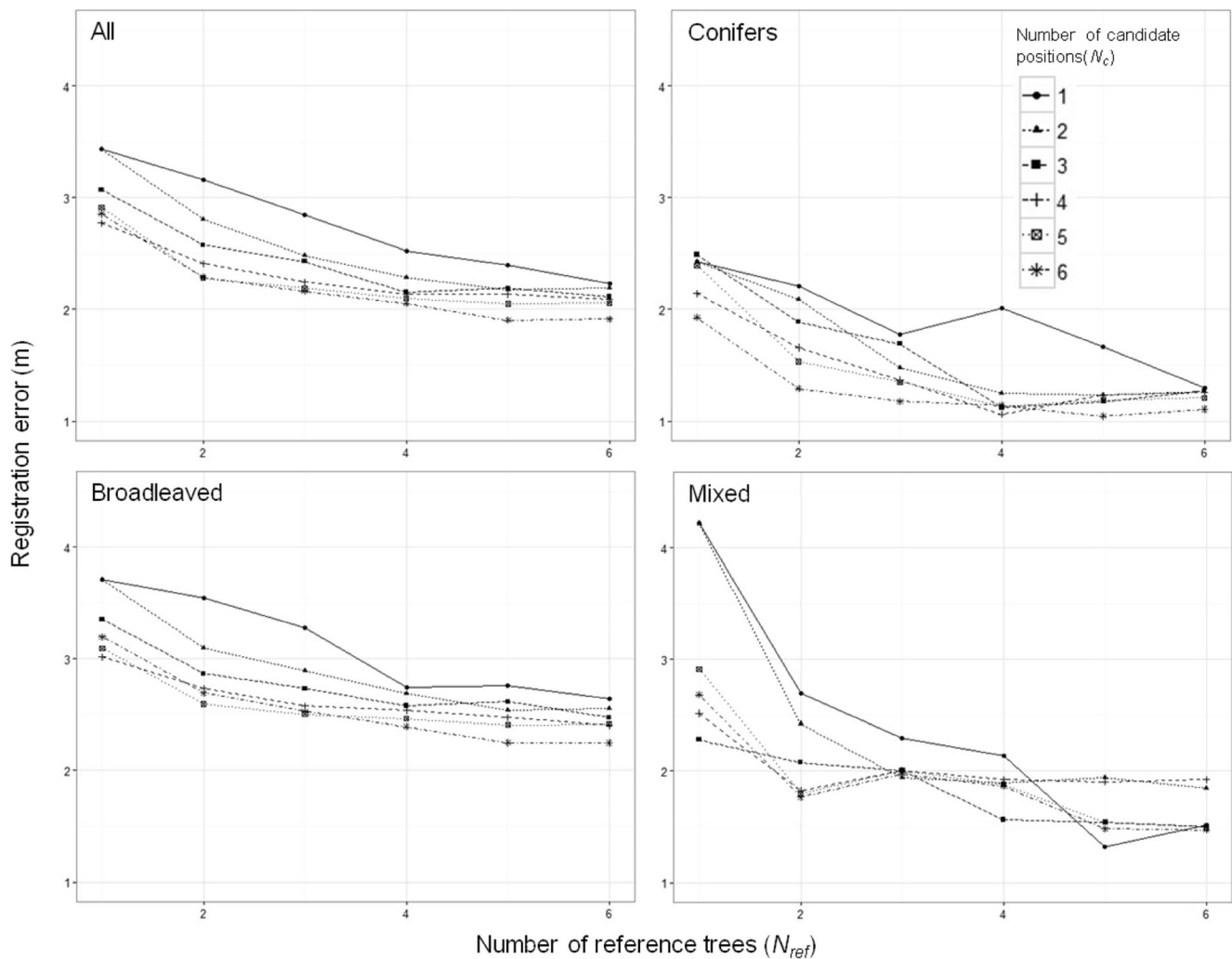
Fig. 2 Sensitivity analysis of RegisTree to the number of reference trees (N_{ref}) and candidate positions (N_c) according to different forest types

were tested, for a total of 15 plots. The best results were obtained with an *FTM* of 12 trees, with a mean error of 0.69 m (± 0.39 m). Beyond this value, the error increased continuously from 0.92 m (± 0.57 m) with an *FTM* of 15 to 1.03 m (± 0.67 m) with an *FTM* greater than 18. Over the broadleaved plots of the Bure forest, *FTMs* of 6, 9, 12, 15, and greater than 15 trees were tested, representing a total of 13 plots. With respect to the Aillon forest, it was not possible to consider an *FTM* of 18, because the number of plots making it possible was limited to 6. Over the Bure forest, the error tended to decrease with an increasing *FTM*. The mean error reached a minimum value of 1.30 m (± 0.47 m) with an *FTM* of more than 15 trees.

Generally, the performance of RegisTree was slightly impacted by forest attributes (Table 4). The higher correlations were obtained for mean terrain slope and plot basal areas, with values of -0.17. The standard deviation of height was also negatively correlated with error, with a value of -0.16. Interestingly, the worst registrations were obtained for plots having a standard deviation of height below 4 m (Fig. 5),

illustrating the challenges associated with the registration of plots in homogeneous forest conditions, like plantations or less textured canopies (Fig. 3).

3.2 Performance against a state-of-the-art method

The performance of RegisTree against a state-of-the-art point-to-surface approach is presented in Table 5. Both methods performed almost similarly with respect to the size of the *FTM*. The state-of-the-art method provided its best results using an *FTM* of 6 trees and the weighted mean absolute height error (wmae) criteria (mean error of 2.18 m $\pm$ 1.99 m). With both *FTM* sizes, RegisTree performed slightly better than the state-of-the-art method. The best result was obtained using the full *FTM*, with a mean error of 1.5 m (± 1.23 m).

3.3 Effect of position of volume prediction

The results of the predictive models of total plot volume are presented in Table 6. The same independent variables were
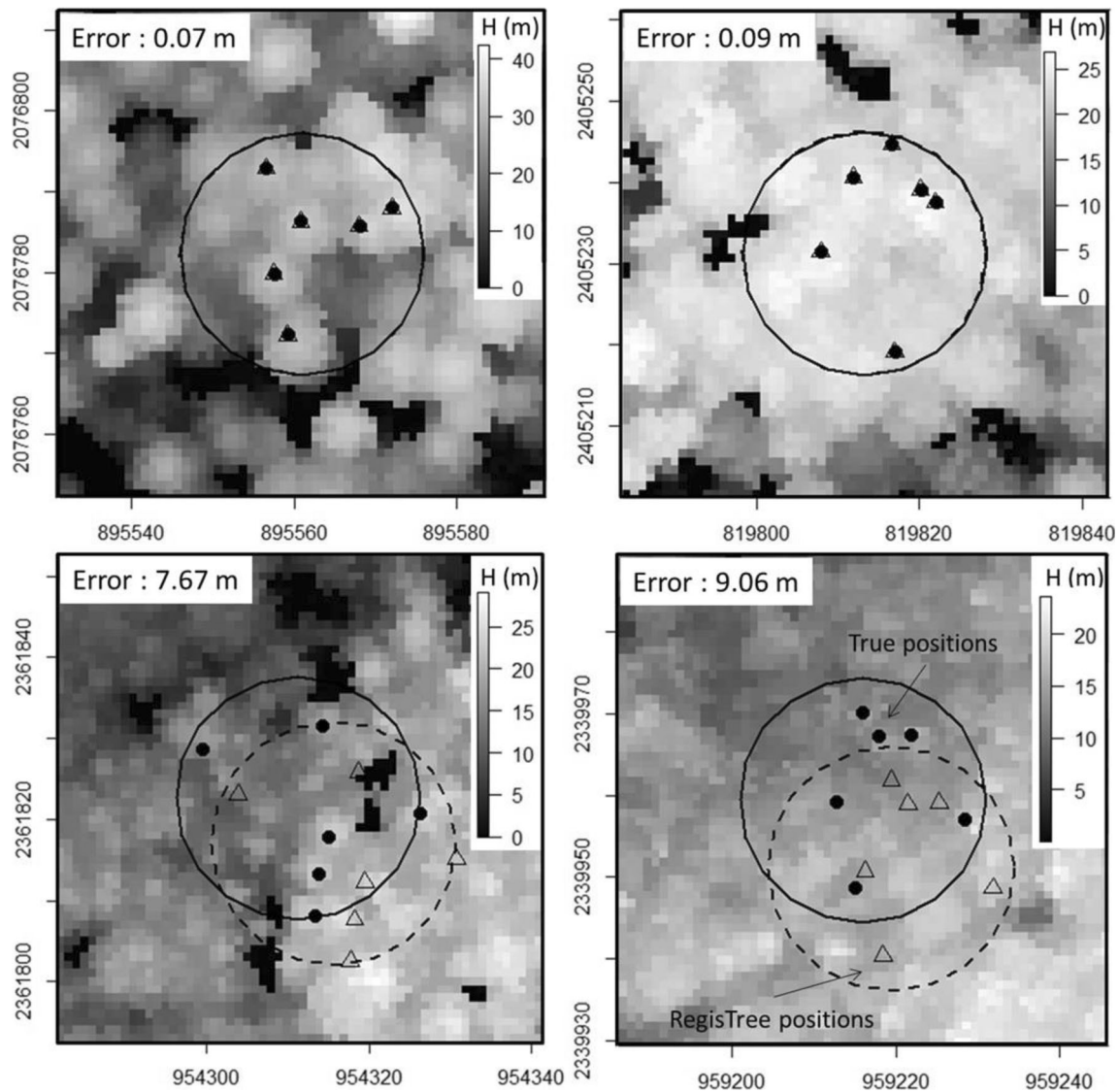
Fig. 3 Best and worst registration results for conifers (left) and deciduous (right) plots, obtained using a *FTM* of 6 trees and the optimal parameters (N_{ref} = 5, N_c = 6)

Table 2 Performance of the registration according to datasets and forest types

Data	Plot number	Min	Mean	Max	SD
All	246	0.08	1.56	9.06	1.32
Dataset 1	122	0.08	1.37	9.06	1.40
Dataset 2	124	0.08	1.74	6.91	1.23
Bure	41	0.09	1.45	7.67	1.26
Aillon	28	0.08	0.69	2.27	0.47
Vosges	53	0.16	1.66	9.06	1.71
Saint-Gobain	43	0.16	1.57	6.58	1.15
Compiègne	33	0.10	2.20	6.91	1.47
Darney	48	0.08	1.60	4.51	1.05
Conifers	60	0.08	1.06	7.13	1.26
Hardwood	168	0.09	1.53	9.06	1.33
Mixed	18	0.26	1.05	4.18	0.91

selected for the four tested scenarios and included the canopy volume (V_{in}), the gap area (G_a), and the forest type (F_{type}). For each model, all the variables were significant at 0.95 and the variance inflation factor of all the variables was below 5. As expected, the best results were obtained using the reference plot positions, with a cross-validation adjusted R-squared (cv-Adj-R^2) of 0.81 and a cross-validation root mean squared error (cv-RMSE) of 94.4 m³ ha⁻¹ or 26.4%. The model based on RegisTree positions performed better than the ones using initial or blurred positions. However, the differences were low. The model based on RegisTree positions showed the same cv-Adj-R^2 than the reference position and a slightly higher cv-RMSE of 94.7 m³ ha⁻¹ (26.5%). In contrast, the degraded initial positions generated a cv-Adj-R^2 of 0.79 and a cv-RMSE of 99.1 m³ ha⁻¹ or (27.8%).
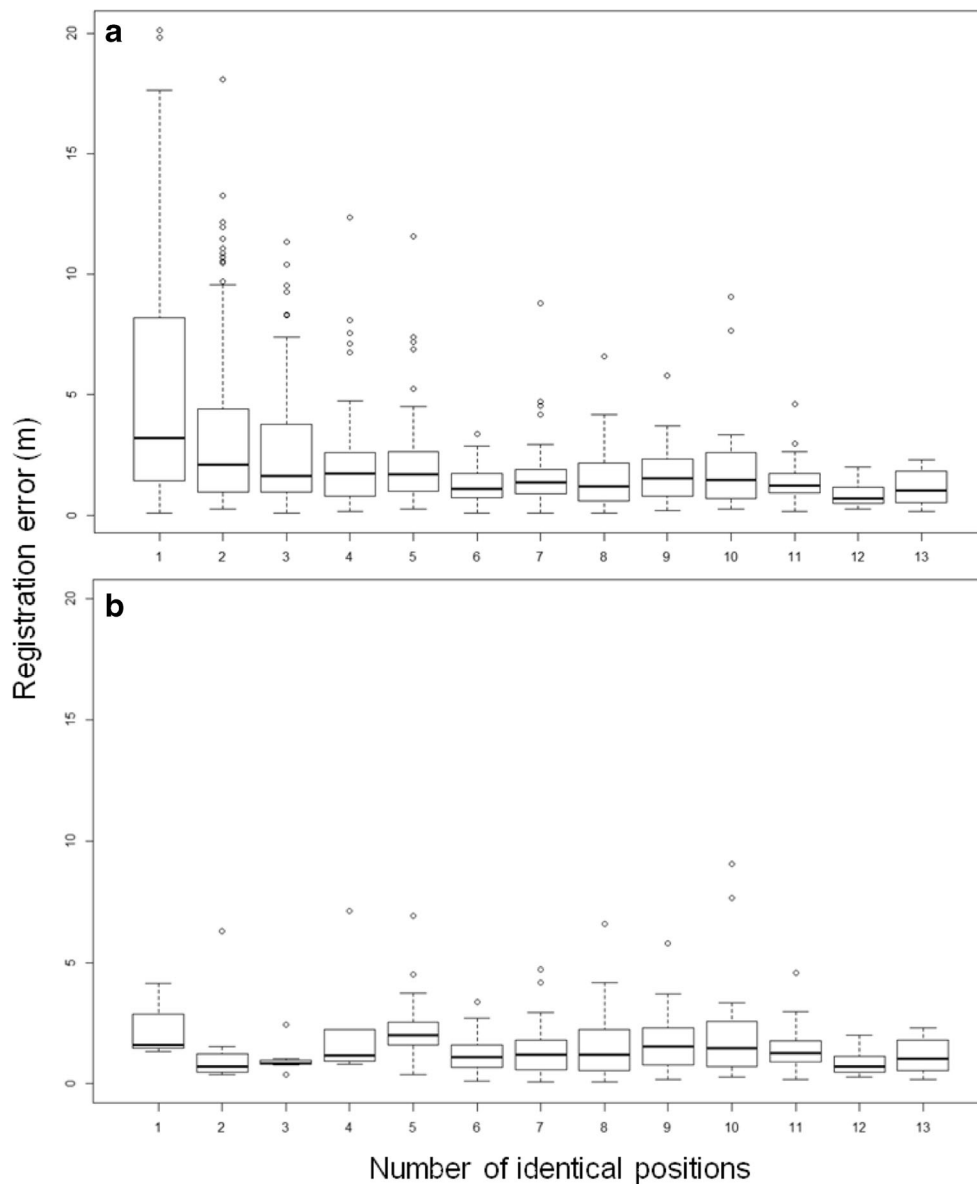
Fig. 4 Boxplot of registration errors as a function of the number of identical positions for the 13 starting position tested. **a** Results of all of the starting positions (13 times 250 plots). **b** Reference position alone (250 plots)

4 Discussion

4.1 Positioning accuracy

With an *FTM* of six trees, the best results were obtained using five reference trees and six candidate positions. This result showed that keeping a large number of potential positions (i.e., 30 in this case) provided more consistent plot positions. In accordance with our working hypothesis, this can be explained by the uncertainties associated with the spatial registration of objects of a different kind and characterized by a distinct level of accuracy and errors. As expected, RegisTree performed better for coniferous plots than for broadleaved plots. Conifers often have more differentiated tree apices than broadleaved trees and show more conical crown shapes, leading to contrasting CHM heights. As a consequence, the 3D positions from the CHM are expected to match better with the *FTM* in coniferous plots. On the other hand, broadleaved species often have multiple local maxima for a single tree crown and are more difficult to measure in the field. Also, the lidar acquisitions considered here were done in leaf-off conditions. Such acquisition conditions are known to produce height underestimations as compared to leaf-on conditions for broadleaved forests (Wasser et al. 2013). All in one, uncertainties in apex positions and height measurements contributed to lower precision in broadleaved forests.

Table 3 Mean registration error (m) and standard deviation (in parenthesis) with respect to the size of the field tree map (*FTM*) with $N_{ref} = 5$ and $N_c = 6$

FTM size	6	9	12	15	18	Total
Aillon (*n* = 15)	0.71 (0.57)	0.75 (0.51)	0.69 (0.39)	0.92 (0.57)	0.96 (0.65)	1.03 (0.67)
Bure (*n* = 13)	1.66 (1.20)	1.81 (1.54)	1.68 (1.21)	1.61 (1.12)	–	1.31 (0.47)

Overall, RegisTree turned out to be quite robust to the initial positions. For the range of positions tested, the algorithm achieved the same result in 65% of cases. That being said, the error was not correlated to the resilience of the algorithm to the initial position, making the criteria relatively significant to qualify the accuracy of the resulting position. However, large errors were systematically obtained in homogeneous plots, as defined by low standard deviation of tree height (Fig. 5), with homogeneous canopy conditions, such as plantations or textureless canopies (Fig. 3). This result is consistent with Dorigo et al. (2010) who reported unreliable registration in very dense deciduous plots with tree apices not clearly defined. Olofsson et al. (2008) and Dorigo et al. (2010) further reported lower registration accuracies in dense plots. Our results did not confirm this finding, with a correlation between density and error of − 0.1 (Table 4). As acknowledged by Dorigo et al. (2010), our results suggested that large errors are attributed to homogeneous canopies, such as textureless canopies or regular plantations. In such plots, both the *FTM* and CHM show little height variability, thus introducing uncertainty in registration. For such plots, it is highly recommended to narrow the search distance to avoid inconsistencies. One may also consider that the positioning accuracy is less of a problem in homogeneous plots, thus minimizing the importance of improving plot position.

Increasing the size of the *FTM* improves the positioning accuracy of deciduous plots by 62% (1.66 to 1.03 m mean error). This result can be explained by the greater uncertainties in defining the tree apices in broadleaved plots and is further exacerbated in our study by the fact that ALS acquisitions were performed in leaf-off conditions. Indeed, increasing the *FTM* size might increase the constraints on the LM and thus helps identify the optimal position. For conifer plots, the results are not as clear, and error tended to increase with *FTM* including more than 12 trees. Because tree apices are often

more easily identifiable in conifer plots, a limited number of dominant trees is sufficient to achieve the optimal plot position. It could be hypothesized that increasing the *FTM* may include co-dominant trees with undefined apices and with apices that could be slightly shifted from the trunk position, thus adding uncertainties. That said, the datasets used for testing this effect were limited in sizes and extended analysis would be required to validate findings on this parameter. Independently from the *FTM*, the results given by RegisTree provide a greater accuracy than that obtained using the state-of-the-art point-to-surface approach (Monnet and Mermin 2014). This highlights the interest of the method and the interest of using LM to constraint the point-to-surface approach.

The positioning accuracies obtained using RegisTree are lower than the sub-metric ones that can be achieved under forest canopies using differential GNSS data (Hauglin et al. 2014) under the range of parameter tested. However, it provides a valuable alternative to improve the positioning accuracy of plots surveyed using low-grade GNSS (Valbuena et al. 2010). The algorithm may be particularly relevant to NFIs in the context of multisource forest inventories, as NFIs may not

Table 4   Correlation of registration error with terrain and forest parameters

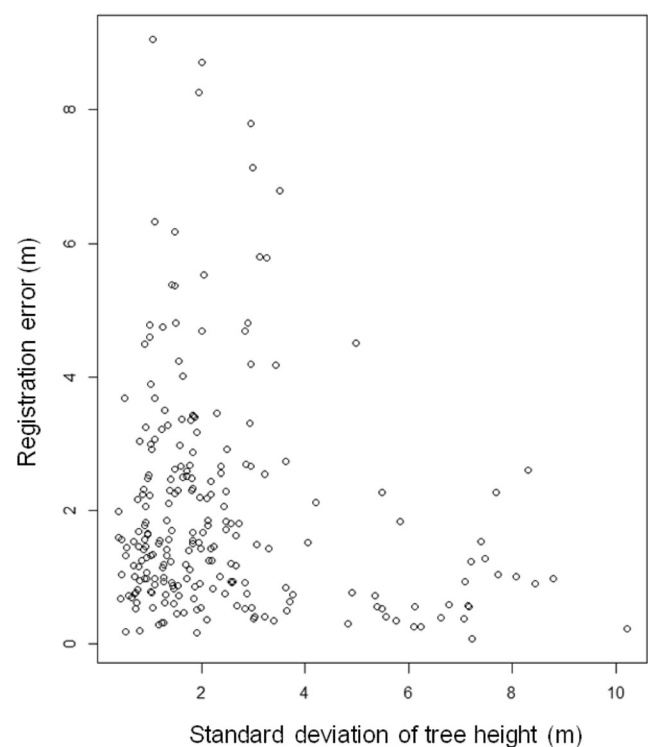|	
Terrain slope (%)	− 0.17
Plot density (stem ha^{-1})	− 0.1
Mean tree height (−m)	0.11
Standard deviation of tree height (m)	− 0.16
Mean tree diameter (cm)	0.02
Standard deviation of tree diameter (cm)	− 0.05
Basal area (m^2 ha^{-1})	− 0.17
Total volume (m^3 ha^{-1})	− 0.11

Fig. 5 Registration error as a function of standard deviation of tree height per plot

Table 5 Registration errors (m) obtained using RegisTree and a state of the art point to surface method using *FTM* of six trees or more

	FTM	Min	Mean	Max	SD
State-of-the-art (correlation)	6	0	3.69	9.89	2.57
State-of-the-art (wmae)	6	0	2.18	8.2	1.99
RegisTree	6	0.07	1.55	9.06	1.33
State-of-the-art (correlation)	Full	0	3.9	12.81	3.09
State-of-the-art (wmae)	Full	0	2.39	14.12	2.48
RegisTree	Full	0.07	1.5	9.06	1.23

The state of the art method was tested with two registration criteria, the correlation and the weighted mean absolute height error (wmae) (Monnet and Mermin 2014). RegisTree was run with $N_{ref} = 5$ and $N_c = 6$

be able to afford the acquisition cost of survey grade GNSS receivers for numerous field crews (McRoberts et al. 2018).

Under high-density lidar acquisitions and leaf-off conditions, the positioning accuracy could be improved by changing the registration paradigm. Using point densities between 30 and 35 pts. m^{-2}, Bock et al. (2017) showed that an individual tree trunk could be identified and isolated. Such an approach makes it possible to use the detected tree positions at DBH height instead of LM. This would solve the issues with both LM positions and height uncertainties and could also improve the performance of algorithms in textureless broadleaved canopies.

4.2 Prediction of plot volume

The accuracy of plot-level predictive models of volume follows the same trends than the registration error did. The best results were achieved with the plot positions that were corrected manually. But RegisTree achieved almost similar results. That said, the gain in performance remained moderated. Compared to the initial conditions, the gain in cross-validated RMSE was only 0.7%, from 27.2% to 26.5%. Even so, in the range of the tested conditions, a systematic shift of 5 m of the differential GNSS position, representing one third of the plot radius, only degraded the cross-validation RMSE by 1.3%. While surprising with respect to Dorigo et al. (2010), Olofsson et al. (2008), or McRoberts et al. (2018) who reported a significant impact of the positioning accuracy on the predictive models of forest parameters, our results are in accordance with those reported by Monnet and Mermin

(2014) for a basal area. A possible explanation for this low performance improvement lies in the quasi-absence of plots falling between stands or at the forest boundaries. Contrary to National Forest Inventory plots that are representative of all forest conditions, the management inventory plots used here were predominantly concentrated inside forest stands, with locally homogeneous forest conditions.

Despite the slight improvement in model performance, the achieved Adj. R^2 are in the same range as those reported in other studies. In similar forest conditions, Bouvier et al. (2015) obtained predictive models of forest volume, explaining 82% of the variance of the volume measured in the field. For Scandinavian areas, Næsset (2007) reported R-squared values within the range of 0.84 and 0.89. While our study sites were quite varied in terms of forest structure and composition, the performance of the models seemed to be reasonable and could be further improved. ALS data were acquired in a range of acquisition conditions, both in terms of density and seasonality. The ALS data was not submitted to any harmonization procedure. The differences in point densities may lead to bias in metrics influenced by the point density such as the percentiles (Magnussen et al. 2012). This being said, no such metrics were included in our models and the spatial resolution of the CHM was largely lower than the point density, thereby limiting the problem with the raster-based metrics. The seasonality might have penalized the performance of the model, as a large majority of broadleaved plots were scanned in leaf-off conditions. Beside the impact of leaf-off conditions on the LM and therefore on the registration, impact could be expected on metrics, like gaps and percentiles. However, Wasser et al. (2013) and White et al. (2015b) reported no effect of leaf-off and leaf-on conditions on the performance of models, minimizing the potential influence of acquisition period on it. Potential improvements may also be achieved by considering metrics describing the vertical canopy structure such as density or penetration metrics (Næsset 2007; Véga et al. 2016). There, a priority was given to metrics describing the canopy volumes, with the underlying objective of applying this approach to photogrammetric models of forest canopies. We do consider that those kinds of CHM and metrics have a high potential for multisource forest inventories and large scale mapping and updating of forest parameters owing to the availability of aerial photographs. That said, it would be worth testing the contribution

Table 6 Adjusted *R*-squared (Adj. R^2), root mean squared error (RMSE) (m^3 ha^{-1} and %) of the predictive model, and volume and corresponding cross-validation results (cv-Adj-R^2, cv-RMSE)

Case	Adj. R^2	RMSE m^3 ha^{-1} (%)	cv-Adj-R^2	cv-RMSE m^3 ha^{-1} (%)
Reference positions	0.82	91.9 (28.6)	0.81	94.4 (26.4)
Initial positions	0.81	94.5 (26.5)	0.80	97 (27.2)
Initial degraded positions	0.80	96.6 (27.1)	0.79	99.1 (27.8)
RegisTree positions	0.82	92.1 (25.8)	0.81	94.7 (26.5)

All the selected models included the same independent variables: V_{in}, G_a, and F_{type}

of density and penetration metrics in order to have a more complete assessment of the effect of plot position on the performance of the models.

5 Conclusion

Improving the spatial registration between field and remote sensing data is important for mapping purposes and inferential estimation using auxiliary data. Many approaches have been proposed to solve the problem with varying degrees of success. The limited efficiency of matching methods comes from both field (plot size, measurement availability, spatial positioning) and remote sensing data (nature, resolution). Here, we suggest an algorithm called RegisTree to improve data registration automatically by matching a spatialized field tree height map with a CHM. RegisTree is based on a modified point (field positions) to surface (CHM) registration approach, ensuring that at least one of the field tree position matches with a local maximum of the CHM. It allowed registering field plots surveyed in a range of forest condition with an accuracy of 1.5 m ($\pm$ 1.23 m). RegisTree was found to be robust to the forest conditions but performed better for conifers than for hardwood, due to their more textured canopies. Due to the complexity of canopy covers and an increase probability of matching errors with increasing search distance, such methods should be used to improve accuracy in the position of plots surveyed with low quality GNSS or to locally improve the position acquired using survey grade GNSS devices for tree-based approaches for modeling forest parameters. Despite this limitation, the development of remote sensing in forest monitoring and management will certainly increase demand for high precision data registration.

Statement on data availability   The datasets generated and/or analyzed during the current study are not publicly available due to ownership and funding constraints but are available from the author upon reasonable request and with permission of ONF and IGN. Code and sample data supporting the findings of this study are available in the Zenodo repository (Fadili et al. 2019)

Compliance with ethical standards

Conflict of interest   The authors declare that they have no conflict of interest.

References

Andersen H-E, Clarkin T, Winterberger K, Strunk J (2009) An accuracy assessment of positions obtained using survey- and recreational-grade Global Positioning System receivers across a range of forest conditions within the Tanana Valley of interior Alaska. West J Appl For 24:128–136

Axelsson P (2000) DEM generation from laser scanner data using adaptive TIN models. Int Arch Photogramm Remote Sens Spat Inf Sci 33:110–117 **Part B4/1**

Baltsavias EP (1999) Airborne laser scanning: basis relations and formulas. ISPRS J Photogramm Remote Sens 54:199–214

Bock J, Piboule A, Jolly A (2017) TidALS: trunk identification in dense Airborne Laser Scanner data to estimate. In: Silvilaser conference, October 10–12, 2017, Blacksburg, Virginia, USA

Bouvier M, Durrieu S, Fournier RA, Renaud J-P (2015) Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. Remote Sens Environ 156: 322–334

Danskin SD, Bettinger P, Jordan TR, Cieszewski C (2009) A comparison of GPS performance in a southern hardwood forest: exploring low-cost solutions for forestry applications. South J Appl For 33:9–16

Deleuze C, Morneau F, Renaud J -P, Vivien Y, Rivoire M, Santenoise P, Longuetaud F, Mothe F, Hervé JC, Vallet P (2014) Estimer le volume total d'un arbre, quelles que soient l'essence, la taille, la sylviculture, la station. RDV techniques ONF 44: 22–32

Dorigo W, Hollaus M, Wagner W, Schadauer K (2010) An application-oriented automated approach for registration of forest inventory and airborne laser scanning data. Int J Remote Sens 31:1133–1153

Fadili M, Renaud JP, Bock J, Vega C (2019) RegisTree: a registration algorithm to enhance forest inventory plot georeferencing. V1. Zenodo. [Dataset]. https://doi.org/10.5281/zenodo.2577140

Favorskaya MN, Jain LC (2017) Overview of LiDAR technologies and equipment for land cover scanning In Handbook on advances in remote sensing and geographic information systems: paradigms and applications in forest landscape modeling, intelligent systems reference library. Springer International Publishing, 122, pp 19–68

Gobakken T, Næsset E (2009) Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. Can J For Res 39:1036–1052. https://doi.org/10.1139/X09-025

Hauglin M, Lien V, Næsset E, Gobakken T (2014) Geo-referencing forest field plots by co-registration of terrestrial and airborne laser scanning data. Int J Remote Sens 35:3135–3149

Johnson KD, Birdsey R, Finley AO, Swantaran A, Dubayah R, Wayson C, Riemann R (2014) Integrating forest inventory and analysis data into a LIDAR-based carbon monitoring system. Carbon Balance Manage 9:3

Kane VR, McGaughey RJ, Bakker JD et al (2010) Comparisons between field- and LiDAR-based measures of stand structural complexity. Can J For Res 40:761–773

Khosravipour A, Skidmore AK, Wang T, Isenburg M, Khoshelham K (2015) Effect of slope on treetop detection using a LiDAR Canopy Height Model. ISPRS J Photogramm Remote Sens 104:44–52

Korpela I, Tuomola T, Välimäki E (2007) Mapping forest plots: an efficient method combining photogrammetry and field triangulation. Silva Fenn 41:457–469

Magnussen S, Næsset E, Gobakken T, Frazer G (2012) A fine-scale model for area-based predictions of tree-size-related attributes derived from LiDAR canopy heights. Scand J For Res 27:312–322

McRoberts RE, Tomppo EO (2007) Remote sensing support for national forest inventories. Remote Sens Environ 110:412–419

McRoberts RE, Chen Q, Walters BF, Kaisershot DJ (2018) The effects of global positioning system receiver accuracy on airborne laser scanning-assisted estimates of aboveground biomass. Remote Sens Environ 207:42–49

Monnet J-M, Mermin É (2014) Cross-correlation of diameter measures for the co-registration of forest inventory plots with airborne laser scanning data. Forests 5:2307–2326

Nakajima H (2016) Plot location errors of National Forest Inventory: related factors and adverse effects on continuity of plot data. J For Res 21:300–305. https://doi.org/10.1007/s10310-016-0538-1

Næsset E (2007) Airborne laser scanning as a method in operational forest inventory: status of accuracy assessments accomplished in Scandinavia. Scand J For Res 22:433–442

Næsset E, Jonmeister T (2002) Assessing point accuracy of DGPS under forest canopy before data acquisition, in the field and after postprocessing. Scand J For Res 17:351–358

Neter J, Wasserman W, Kutner MH (1985) Applied linear statistical models (2nd ed.). Irwin, New York

O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41:673–690

Olofsson K, Lindberg E, Holmgren J (2008) A method for linking field-surveyed and aerial-detected single trees using cross correlation of position images and the optimization of weighted tree list graphs In proceeding of Silvilaser 2008, Sept 17-19, 2008 – Edinburgh, UK, pp 95–104

Picard RR, Cook RD (1984) Cross-validation of regression models. J Am Stat Assoc 79:575–583

Pinto da Costa J (2015) Rankings and preferences—new results in weighted correlation and weighted principal component analysis, **SpringerBriefs in Statistics, 95 pp.**

Ransom MD, Rhynold J, Bettinger P (2010) Performance of mapping-grade GPS receivers in southeastern forest conditions. RURALS: Review of Undergraduate Research in Agricultural and Life Sciences: Vol 5: Iss 1, Article 2

Tomppo E, Olsson H, Ståhl G, Nilsson M, Hagner O, Katila M (2008) Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sens Environ 112:1982–1999

Valbuena R, Mauro F, Rodriguez-Solano R, Manzanera JA (2010) Accuracy and precision of GPS receivers under forest canopies in a mountainous environment. Span J Agric Res 8:1047–1057

Véga C, Durrieu S (2011) Multi-level filtering segmentation to measure individual tree parameters based on Lidar data: application to a mountainous forest with heterogeneous stands. Int J Appl Earth Obs Geoinf 13:646–656

Vega C, Hamrouni A, El Mokhtari S, Morel J, Bock J, Renaud J-P, Bouvier M, Durrieu S (2014) PTrees: a point-based approach to forest tree extraction from lidar data. Int J Appl Earth Obs Geoinf 33:98–108

Véga C, Renaud J-P, Durrieu S, Bouvier M (2016) On the interest of penetration depth, canopy area and volume metrics to improve Lidar-based models of forest parameters. Remote Sens Environ 175:32–42

Wasser L, Day R, Chasmer L, Taylor A (2013) Influence of vegetation structure on Lidar-derived canopy height and fractional cover in forested riparian buffers during leaf-off and leaf-on conditions. PLoS One 8:e54776

White JC, Stepper C, Tompalski P, Coops NC, Wulder MA (2015a) Comparing ALS and image-based point cloud metrics and modelled forest inventory attributes in a complex coastal forest environment. Forests 6:3704–3732

White JC, Arnett JTTR, Wulder MA, Tompalski P, Coops NC (2015b) Evaluating the impact of leaf-on and leaf-off airborne laser scanning data on the estimation of forest inventory attributes with the area-based approach. Can J For Res 45:1498–1513

Wing MG, Eklund A (2007) Performance comparison of a low-cost mapping grade global positioning systems (GPS) receiver and consumer grade GPS receiver under dense forest canopy. J For 105:9–14

Wulder M (1998) Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. Prog Phys Geogr 22: 449–476