Monnet, J.M. Evaluation of a semi-automated approach for the co-registration of forest inventory plots and airborne laser scanning data. Silvilaser 2013, the 13th International Conference on LiDAR Applications for Assessing Forest Ecosystems

Evaluation of a semi-automated approach for the co-registration of forest inventory plots and airborne laser scanning data

Jean-Matthieu MONNET[1]

[1]Irstea, UR EMGR, F-38402 Saint-Martin-d'Hères, France
jean-matthieu.monnet@irstea.fr

Paper Number: SL2013-053

1. Introduction

Traditional forest inventory practices mainly rely on statistical descriptions obtained from field sample plots (Duplat and Perrotte 1981). This method provides reliable estimates at the forest level while limiting fieldwork cost. However information at the management unit level (compartment) is often not usable due to the low number of observations. In the past decade, airborne laser scanning (ALS) has demonstrated its ability to model the 3D geometrical structure of the canopy even at the single tree level. The area-based approach aims at combining the high resolution of LiDAR data with the sample field plots in order to provide statistically calibrated, continuous maps of forest parameters (Næsset 2004). Relationships between local descriptors of the ALS point cloud and forest parameters such as basal area, stem volume, dominant height are investigated based on the available field data. Once a prediction model is validated, it is applied to the whole ALS dataset in order to produce the map of the desired forest parameters.

Whereas the planimetric accuracy of an ALS point cloud is around 25 cm (Baltsavias 1999), position accuracy of field plots is far lower. Indeed, plots are generally positioned relatively to elements that are visible on aerial pictures, or with Global Positioning System (GPS) receivers. Because of the canopy that reflects and attenuates the GPS signal, GPS measurements accuracy inside forest stands is lower than in open areas and of the order of a few meters (Naesset and Jonmeister 2002). However, position accuracy is difficult to estimate as it depends on acquistion parameters, satellite availability and canopy density. The error in field plot position is problematic as the predictive relationships are investigated between ALS point clouds that do not exactly correspond to the inventoried sample plots. This error is likely to decrease the degree of fit of the final prediction models (Gobakken and Næsset 2009). Meanwhile, the prediction error which is often obtained with cross-validation is likely to be over-estimated as the validation data is also affected by the position error.

In order to both calibrate better prediction models and evaluate their accuracy, it is important to improve co-registration. The ALS data itself can be used to correct the field data position when trees positions are recorded for the sample plots, e.g. by visually comparing the Canopy Height Model (CHM) derived from the ALS data with the trees position and size. However this task is very time-consuming, in particular when the position error is large or when the forest canopy is closed and homogeneous. A few methods have been proposed to automatically adjust the plot position to ALS data. Olofsson *et al.* (2008) first detect single trees in the CHM and then computed single trees position images with trees modelled as Gaussian surfaces. The image of the field trees are cross-correlated with the image of the detected trees in order to estimate the offset to be applied to the field data. High co-registration accuracies are obtained provided that the position errors of single trees are small. Dorigo *et al.* (2010) define a cost function wich takes into account the difference in field and CHM heights, as well as the social status of a tree. Pascual *et al.* (2013) also take into account the match between the ALS digital terrain model and topographic measurements.

The objective of this article is to test a co-registration approach based on the direct correlation between an image of the field trees diameters and the CHM, and to evaluate its robustness to the number of georeferenced trees. Indeed, measures of tree heights and positions are time-consuming and the issue of minimizing inventory cost while maintaining the possibility of a posteriori position correction is of high interest when designing ALS-assisted forest inventory protocols.

2. Material

2.1 Study area

The study area is the public forests of Prénovel and Les Piards (Jura, France, figure 1). Elevation range is quite narrow (900 to 1060 m) but the topography is rough. The forest is mainly constituted of uneven-aged, mixed stands of silver fir (*Abies alba*), Norway spruce (*Picea abies*) and beech (*Fagus sylvatica*). In the framework of a national research program about the impact of roe deer, permanent sample plots were installed in 2005. The inventory was updated between Sept. 9th, and Oct. 6th, 2011 in the framework of the project BGF (Biodiversity and Forest Management).

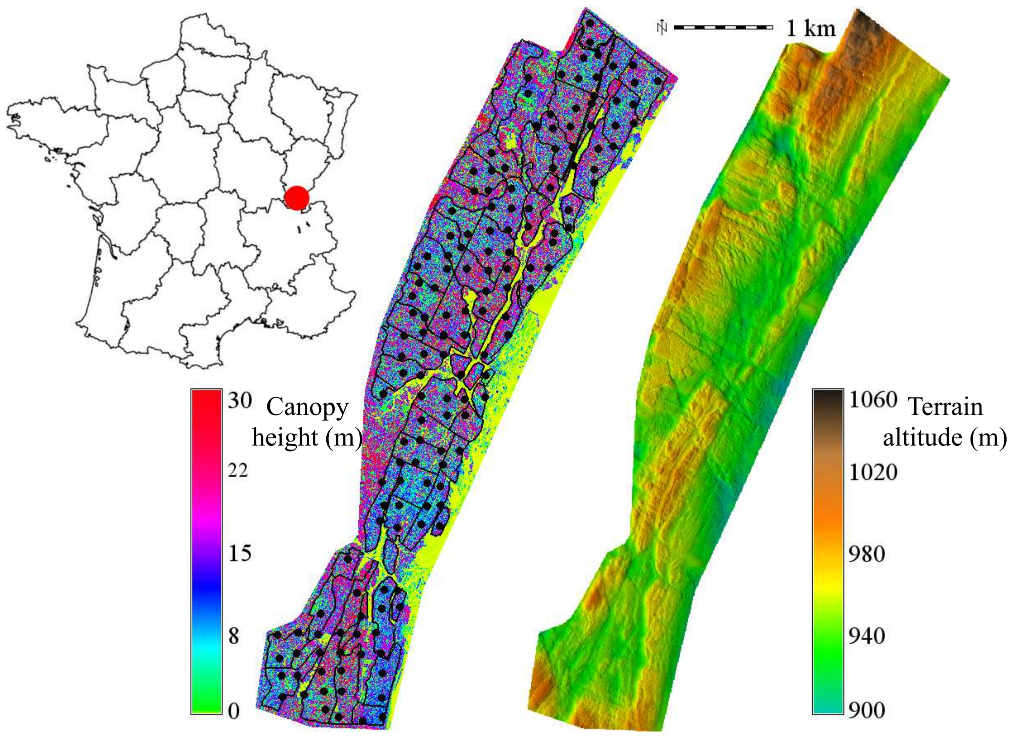


Figure 1: Study area location (left), plot positions over the ALS canopy height model (center), shaded ALS digital terrain model (right).

The sampling design is a grid pattern of size 200 m x 200 m, which results in 139 plots inside the forest (583 ha). Plot centers were positioned with a GARMIN commercial receiver. On each plot, all trees with diameter at breast height (DBH) above 7.5 cm were inventoried up to a radius of 10 m, and all trees with DBH above 27.5 cm up to 17 m. For each tree, the azimuth and soil distance from the plot center, the DBH and species were recorded. Tree heights were measured on sample trees. Among the 139 plots, basal area ranges from 13.3 to 60.5 $m^2.ha^{-1}$, with a mean of 31.9. Stem density ranges from 110 to 1520 stems per hectare, with a mean of 510.

2.2 ALS data

Airborne laser scanning data were acquired in the framework of the NEWFOR research project. The flight took place on Sept. 16-17th, with a full waveform RIEGL LMS-Q560 scanner on a fixed-wing aircraft. Flight speed was 180 km.h^{-1} at 500 m above ground, with a strip overlap of 60%. Pulse frequency was 180 kHz with a scan angle of ± 30 degrees. The obtained mean pulse density on the 9.2 km^2 is 9.3 m^{-2}.
Pre-processing of the raw files was done by the contractor. Echoes were extracted and georeferenced with the RIEGL software suite. The resulting point cloud was classified in two classes (ground and vegetation) with TerraScan.

3. Methods

3.1 Co-registration algorithm

The workflow for the co-registration of a plot is presented in figure 2. The required inputs are:

- tree positions and value (diameter or height);
- plot center position and radius;
- canopy height model.

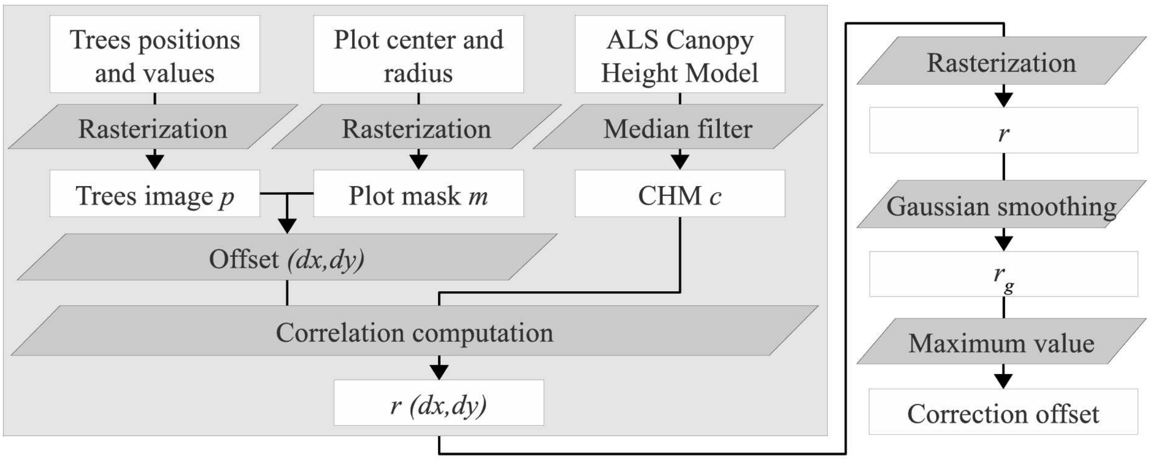
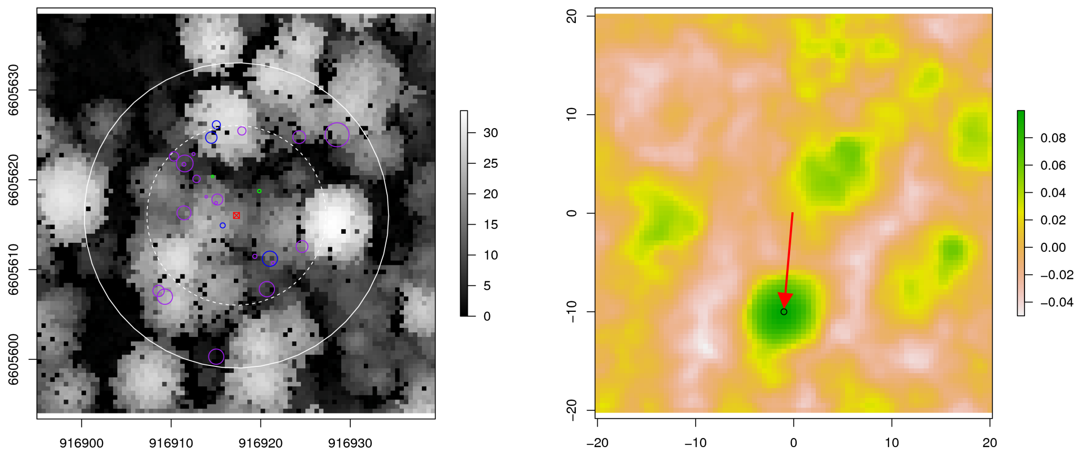


Figure 2: Workflow of the co-registration algorithm.

The canopy height model c is the difference between the digital surface model, calculated by retaining the altitude of the highest ALS point in each pixel, and the digital terrain model, calculated by bilinear interpolation of ALS ground points at the center of pixels. Spatial resolution is 0.5 m. A 3x3 median filter is applied to the canopy height model in order to fill the blank pixels with no ALS points. The trees map is rasterized into an image p by retaining for each pixel the largest value of the trees contained in this pixel. A plot mask m is computed by retaining only pixels whose centers are located inside the plot circle. An offset (dx, dy) is applied to the position of the mask and trees image. The correlation between the canopy height model and the offset trees image is then calculated over the mask according to equation (1):

$$r_{dx,dy} = \frac{\sum_{i \in m} (p_i - \bar{p})(c_i - \bar{c})}{\sqrt{\sum_{i \in m} (p_i - \bar{p})^2} \sqrt{\sum_{i \in m} (c_i - \bar{c})^2}} \text{, with } \bar{x} = \frac{1}{card(m)} \sum_{i \in m} x_i \quad (1)$$

The correlation is computed for *dx* and *dy* between -20 and +20 with 0.5 m increment. The values are rasterized and a Gaussian filter with sigma=0.5 pixel is applied. The hypothesis is that when the co-registration is correct the correlation between the trees image and the canopy height model is maximal. Therefore the coordinates of the maximum value in the smoothed image r_g are retained as the offset to be applied to trees positions.
The workflow is applied for each of the 139 plots of the study area, with the GPS measurement as the reference plot center position. Corrected positions obtained after applying the offset are manually checked.



a) b)

Figure 3: Plot 8 a) Uncorrected tree positions over the canopy height model b) Image of the correlation between the trees image and canopy height model, red arrow is the offset to be applied to the plot center.

3.2 Prediction models

In order to evaluate the effect of co-registration on the accuracy of ALS prediction models, the area-based method is applied to calibrate a prediction model for basal area.
A Box-Cox transformation is first applied to the dependant variable (basal area) in order to normalize its distribution. Independent variables are selected among a list of ALS metrics by retaining the group which yields the highest adjusted-R^2 in ordinary least square regression when testing all possible combinations with at most three variables.
Validation is based on the repetition of 1000 ten-fold cross-validations. For each repetition, ten groups of equal size are randomly constituted among the dataset. On group is discarded at a time when computing the coefficients of the regression for the selected dependant variables, and then the prediction error is calculated for each plot of the discarded group. The global root mean square error for the repetition is recorded. This procedure is applied to the four cases where the GPS / corrected positions are used for the calibration / prediction steps.

3.3 Robustness assessment

To assess the influence of the number of georeferenced trees on the possibility of automated co-registration, the workflow described in paragraph 3.1 is applied to the plots with the corrected positions, but by using only the *n* nearest trees from the plot center or *n* largest trees on the plot. In the case of the nearest trees, the distance between the plot center and the farthest considered tree is used as the plot radius to compute the plot mask.
When the proposed offset for the co-registration based on this restricted trees list is within two meters from the starting position, the algorithm proposition is considered as acceptable. The number of reference trees *n* is tested for values between 3 and 25. In case the number of trees in the plot is smaller than *n*, all trees are used to compute the trees image.

4. Results

4.1 Co-registration

For 127 plots (91.4%) the correction proposed by the algorithm could be unambiguously validated by the operator. Among the twelve other plots, six (resp. two) were correctly co-registered when the searched window was increased to ± 40 (resp. 100) meters. Three of the four remaining plots could be manually co-registered. For two of them a few trees appeared to have been felled between the inventory and the ALS flight. Those trees were subsequently removed from the inventory and the algorithm was then able to correctly co-register the plots. On one plot the algorithm failed to identify the correct position, and one plot could not be identified. For the 138 identified plots, mean distance between the GPS and the validated position is 9.0 ± 8.7 m. Figure 4 shows the distribution of the difference between GPS and validated positions for the 138 identified plots.

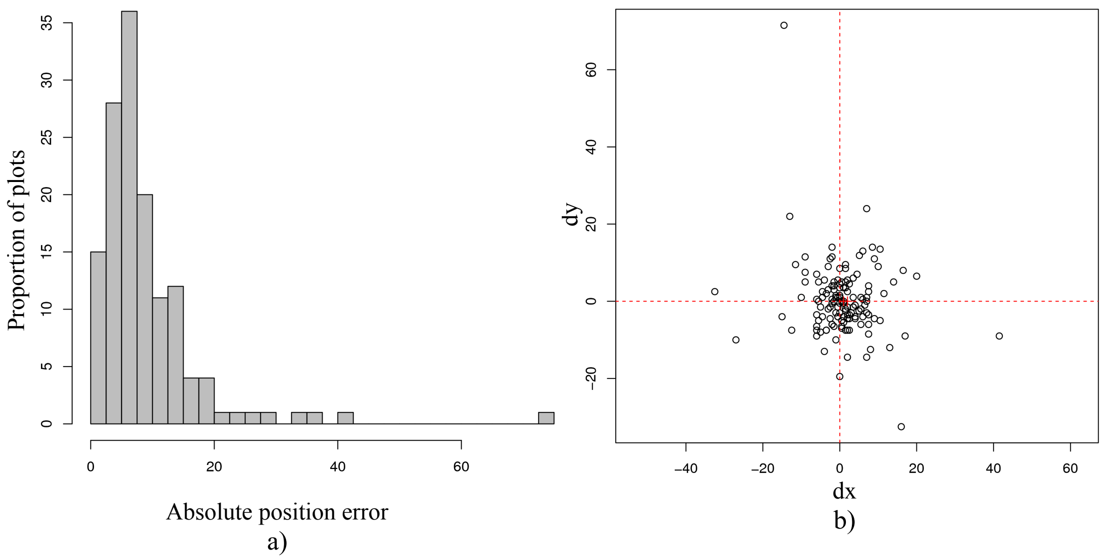


Figure 4: Co-registration error for the 138 identified plots. a) Distribution of the absolute error. b) Plot of correction offsets.

4.2 Prediction models

When the plots are georeferenced with the GPS positions, the selected variables are the sixth decile of first echoes height, the sixth decile of the last echoes height, and the mean of the canopy height model. With the corrected positions, the selected variables are the 99th percentile of the single echoes and the mean of the canopy height model. Statistics for the RMSE obtained in 1000 ten-fold cross-validations are presented in table 1.

Table 1: RMSE mean and standard deviation ($m^2.ha^{-1}$) obtained in 1000 repetitions of ten-fold cross-validations for the basal area prediction model, depending on the plot positions used for calibration (row) and validation (column).

		Validation	
		GPS positions	Corrected positions
Calibration	GPS positions	6.94 ± 0.056	5.83 ± 0.052
	Corrected positions	7.13 ± 0.035	5.83 ± 0.038

4.3 Robustness assessment

Figure 6 presents the proportion of correct offset propositions by the co-registration algorithm when only the *n* nearest or largest trees are used as references. With the three largests trees, 88.4% of plots are correctly georeferenced, but only 60% with the three nearests. All plots are correctly positioned with the six largest trees, whereas the highest percentage attained with the nearest is 99.3% with 29 trees. It is noteworthy that the percentage slightly decreases when more than the sixteen largest trees are used.

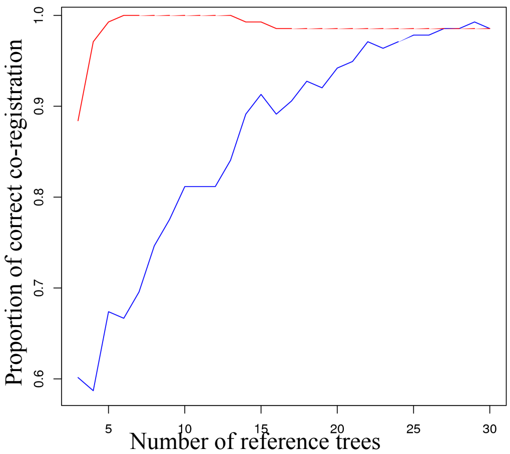


Figure 5: Percentage of correct positions depending on the number of references trees: *n* nearest (blue solid line) or *n* largest (red dotted line).

5. Discussion

Dorigo *et al.* (2010) could automatically co-register 68% of the 98 samples plots from the Austrian National Forest Inventory in Vorarlberg within 5 m of the position previously determined by an operator. Those plots consisted in a combination of fixed-area (radius=2.6 m) and angle count sampling. The proportion of correctly co-registered plots in the present study is higher (91.4%), which could be linked to several factors. First the plots are of fixed area with 17 m radius, so that more trees are inventoried inside smaller surfaces, leading to a better local correspondance between the stand described by the inventory and the canopy height model. Moreover, the plots are concentrated in a small area where the stands are uneven-aged and dominated by coniferous species. The canopy is made of a few dominant trees and several small gaps, so that when the correlation is computed, the correct position results in a clear global maximum. Besides, the time gap between the ALS flight and the field inventory was only one year, so that stand changes lead to errors in only two plots. In Vorarlberg the time gap might be larger, with an ALS data from 2002-2004 and field surveys from 2000-2002.

With a posteriori validation of the co-registration, a bias might exists as the operator would be tempted to consider a proposition as correct even tough he would a priori have preferred a slighlty different position. However, it turned out that the propositions of the algorithm were mostly within one or two pixels from the visually determined solutions. The absolute precision of the co-registration is expected to be around one meter, while the relative position precision of a tree within a plot is around 0.5 m. However, this value is based on the hypothesis that the position of the maxima of the canopy height model is an unbiaised indicator of the position of tree stems. In slope areas, normalizing the digital surface model with the digital terrain model leads to a deformation of the canopy shape and a possible virtual shift of the apices of round crowns (figure 6b-c). Moreover, in slope areas trees tend to bend downslope or to develop a flag-shaped crown so that the planimetric position of the tree apices are located downslope (figure 6a). In such cases, the co-registration is finally biaised because the crowns in the canopy height model are really or virtually shifted downslope compared to tree stems. The position error will then depend on the slope, crown shape and tilt of trees. Methods that rely on the comparison of the positions of local maxima (Olofsson *et al.* 2008) will avoid the deformation of tree crowns if local maxima are detected in the digital surface model, but will fail to handle the case of tilted trees. Anyway, this error relatively to the trees stems leads to a better co-registration with the tree crowns so that this should actually benefit to the prediction models as they are built on relationships with the ALS-described crowns, and not the stems themselves.

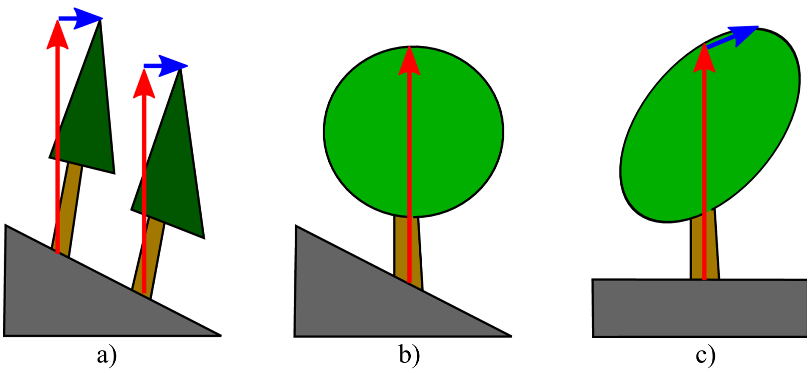


Figure 6: a) Difference between stem position and apex position in tilted trees. (b-c) Shift of the apex position of a round crown when correcting slope.

With the corrected positions used in calibration and validation of the prediction models, the RMSE of basal area prediction is 5.83 $m^2.ha^{-1}$. With GPS positions the apparent RMSE is higher (6.94) but the models are actually better as the RMSE for the GPS-calibrated model is the same as the co-registered-calibrated model, when both are validated with the co-registered positions. This shows that the calibration step is quite robust to the position error of field plots, but that this error leads to a significant under-estimation of the precision of the model in the validation step. Using GPS instead of corrected positions only leads to a slightly higher variability of the estimations of RMSE (0.052 instead of 0.03 $m^2.ha^{-1}$). Considering that the GPS position error was 9.0 ± 8.7 m, these results differ from the findings by Gobakken and Næsset (2009), which showed that position errors higher than five meters yielded predictions with lower accuracy. The difference might be explained by the difference in the position error distribution, which was simulated in their study, or by the homogeneity of the forest stands. Indeed in homogeneous forest stands the position error should not result in large differences in the ALS metrics.

A high percentage of co-registration can be achieved with a small number of positioned trees when those selected are the largests on the plot. This result shows that with a limited time spent on georeferencing a few trees, the possibility of automatic co-registration remains. The decrease of successful co-registration when smaller trees are used may be explained by the fact that those trees are often overtopped by the crown of larger trees. When added to the tree image, small value pixels are created where the canopy is actually high because of adjacent dominant trees. To avoid this effect, it is possible to model the crown of each tree and compute the tree image as the maximum of all tree crowns (Olofsson et al., 2008), but this requires additional hypotheses about tree crown shape.

With our dataset, the automatic co-registration failed mainly because of two reasons. First, changes that occurred in the forest and resulted in a CHM that does not reflect the inventoried trees. Second, large GPS errors that are outside the window searched by the algorithm. Those plots might be correctly co-registrated by extending the window size but this requires longer processing times. Moreover, a trade-off must be found between the number of additional correct co-registrations and the higher probability of false co-registrations on other plots due to the wider window. In order to handle those cases, a flagging criterion such as proposed by Dorigo *et al.* (2010) is required to avoid manual checking of all algorithm propositions.

6. Conclusion

The algorithm proposed for co-registration of field inventory with ALS data proved to be efficient as 91.4% of plots could be automatically corrected with only the position and diameter information. Moreover, the automated co-registration also performed successfully when only the five largest trees on the plot are used, which demonstrates the possibility of a posteriori position correction while limiting inventory time. Besides, better co-registration turns out to have only a small effect on ALS prediction models accuracy, but to be necessary to estimate it properly. In order to implement this approach in practice for large datasets, a quality criterion needs to be added to decrease the need for manual checking.

Acknowledgements

This work was funded by the European Commission (project Alpine Space 2-3-2-FR NEWFOR) and by the French National Research Agency (project ANR-2010-BIOE-008 FORESEE).
Many thanks to the French Forest Office (Office National des Forêts) and project BGF for the Prénovel-Les Piards field data.

References

Baltsavias, E. P., 1999. Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, 199-214.

Dorigo, W.; Hollaus, M.; Wagner, W. and Schadauer, K., 2010. An application-oriented automated approach for co-registration of forest inventory and airborne laser scanning data. *International Journal of Remote Sensing*, 31, 1133-1153.

Duplat, P. and Perrotte, G., 1981. *Inventaire et estimation de l'accroissement des peuplements forestiers*. Office National des Forêts.

Gobakken, T. and Næsset, E., 2009. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*, 39, 1036-1052.

Naesset, E. and Jonmeister, T., 2002. Assessing point accuracy of DGPS under forest canopy before data acquisition, in the field and after postprocessing. *Scandinavian Journal of Forest Research*, 17, 351-358.

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

Olofsson, K.; Lindberg, E. and 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: R.A., H.; Rosette, J. & Suárez, J. (Eds.), *Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory. September 2008. Edinburgh, UK*.

Pascual, C.; Martín-Fernández, S.; García-Montero, L. and García-Abril, A., 2013. Algorithm for improving the co-registration of LiDAR-derived digital canopy height models and field data. *Agroforestry Systems*, 1-9.