A new automated approach for co-registration of national forest inventory and airborne laser scanning data

Wouter Dorigo[1], Markus Hollaus[1,2], Klemens Schadauer[3] & Wolfgang Wagner[1,2]

[1]Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Gusshausstrasse 27-29, 1040 Vienna, Austria; email: wd@ipf.tuwien.ac.at
[2]Christian Doppler Laboratory for "Spatial Data from Laser Scanning and Remote Sensing" at the Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Gusshausstrasse 27-29, 1040 Vienna, Austria; email: mh@ipf.tuwien.ac.at; ww@ipf.tuwien.ac.at
[3]Department of Forest Inventory at the Federal Research and Training Centre for Forests, Natural Hazards and Landscape, Seckendorff-Gudent-Weg, 1130 Vienna, Austria; email: klemens.schadauer@bfw.gv.at

Abstract

Airborne laser scanning (ALS) data are often used for downscaling point based forest inventory (FI) measurements in order to obtain spatially distributed estimates of forest parameters. Such downscaling algorithms usually consist in a direct coupling between selected FI parameters and ALS data collected at the field sampling locations. Thus, adequate co-registration between FI and ALS data is an essential pre-processing step in order to get accurate predictive relationships.
The current paper presents a new, automated co-registration approach which iteratively searches for the best match between an ALS based canopy height model and the tree positions and heights measured during the FI. While the basic principle of the algorithm applies to various types of FI sampling configurations, the co-registration approach has been specifically developed to take into account the tree selection criteria posed by angle count sampling. Several criteria are employed to detect possible ambiguous solutions and to reduce post-processing efforts by an image operator. Model validation was based on National Forest Inventory (NFI) and ALS data of the Austrian Vorarlberg province.

Results show that 67% of the sample plots could be accurately automatically co-registered (i.e, distance to reference data set < 4 m). All solutions with deviations from the reference data set > 4 m were correctly marked by the algorithm as being ambiguous. Applying the automatically co-registered sample plots in a growing stock model provided estimates that were clearly superior to those obtained with the original plot positions and even slightly outperformed those based on manual co-registration. As the developed algorithm will be part of an operational processing chain for Austrian NFI data, it has a high practical relevance.

Keywords: LiDAR, relative orientation, relascope, Austria, mountainous environment

1. Introduction

Forest inventories (FIs) are usually based on field measurements performed at selected sampling units. This way of sampling provides statistically derived measures of forest conditions which, depending on the sampling density, are representative for large to medium size administrative units such as countries or provinces. If information is required for smaller administrative units, like municipalities or forest stands, the available forest information has to be downscaled using additional, spatially distributed information sources such as multi-spectral satellite imagery (Koukal, 2004) or aerial photographs (Holmström *et al.*, 2001). In recent years, airborne laser scanning (ALS) has proven a very promising alternative data basis for spatializing point based forest inventories (Maltamo et al., 2007). Its capability of accurately describing the horizontal and vertical distribution of canopy elements makes ALS well suited for the quantitative assessment of structural forest parameters such as tree density, tree height, and stem volume.

The downscaling procedure generally consists of two consecutive steps: i) establishment of a consistent relationship between selected forest inventory parameters and laser scanning data of the field measurement locations (e.g., by k-nearest neighbours or multiple regression), and ii) deploying the relationship thus obtained to the entire laser scanning data set in order to obtain the spatially distributed forest inventory. Establishing a predictive relationship between FI data and ALS relies on a direct coupling between canopy height information contained in the ALS data and the forest and tree attributes of the FI. Therefore, accurate spatial agreement is of vital importance for accurate calibration of the established relationships (Farid et al., 2006; Hollaus et al., 2007). Nevertheless, the coordinates of sampling locations and tree positions are often still measured with non-differential GPS units, leading to positioning errors up to several meters. This is particularly true in mountainous terrain where due to topography the number of visible satellites is significantly reduced compared to flat terrain. In contrast, ALS data typically have planimetric errors of less than 50 cm, making it very suitable as a geographic reference for the FI data. If tree positions and heights of the trees within the sampling units are known, a data analyst can adapt the positions of the FI data to the ALS data set by visual interpretation. This might, however, be a time-consuming and tedious task, especially if several thousands of sampling units have to be co-registered, such as in the case of national forest inventories.

To overcome this problem, the current paper presents a new, automated approach for the co-registration of FI and ALS data. While the basic principle of the approach applies to various types of FI sampling configurations the study will concentrate on data of the Austrian National Forest Inventory (NFI) which is based on angle count sampling (Bitterlich, 1948). Section 2 describes more in detail the characteristics of the NFI, even as the specifications of the used ALS data. The co-registration procedure is presented in Section 3, while its results are presented and discussed in Section 4. Conclusions and outlook are given Section 5.

2. Study site and data

2.1 Study area

The novel co-registration procedure was developed based on ALS and NFI data of the Vorarlberg province in Austria (Figure 1a). Elevation in the Vorarlberg province ranges from 396 m to 3,312 m asl. The landscape is mainly characterized by high alpine areas, coniferous and mixed forests, shrubs, meadows, and sparsely settled areas in the valley floors. The average timberline ranges between 1,700 and 2,000 m. According to the NFI 2000/2002[1] Vorarlberg is covered with about 97,000 ha of forest, representing a forest cover fraction of 37.3%. The main tree species in Vorarlberg are spruce (*Picea abies*; 53.9% of the total area covered by forests), fir (*Abies alba*; 11.6%) and beech (*Fagus sylvatica*; 9.6%). 66.9% of the forested area can be classified as coniferous forest, 23.8% as deciduous forest, while the rest consists of open spaces, shrubs, and bare surfaces[2].

2.2 Airborne laser scanning data

The ALS data were acquired within the framework of a commercial terrain mapping project covering the entire district of Vorarlberg. Since terrain mapping campaigns require snow-free and

[1] http://web.bfw.ac.at/i7/Oewi.oewi0002?geo=8&isopen=0&display_page=0
[2] http://web.bfw.ac.at/i7/Oewi.oewi0002?geo=0&isopen=3&display_page=22

leaf-off conditions, a prerequisite that is usually not simultaneously met for valley floors and high altitudes, the data were acquired during several flight-campaigns in the years 2002 to 2004. The data were acquired by the company TopScan GmbH, Germany deploying Airborne Laser Terrain Mapper systems (ALTM 1225, ALTM 2050) and the company Terra Digital GmbH, Germany which employed a Leica-Scanner ALS50. The flying heights of the ALS campaigns vary between ~500 and ~2,000 m above ground and minimum point density is 1 point/m². For this study, georeferenced 3D-point clouds and digital terrain (DTM) and surface models (DSM) with a resolution of 1 m were provided by the Land Survey Administration Feldkirch. Canopy height models (CHM) were calculated by subtracting the DTM from the DSM.

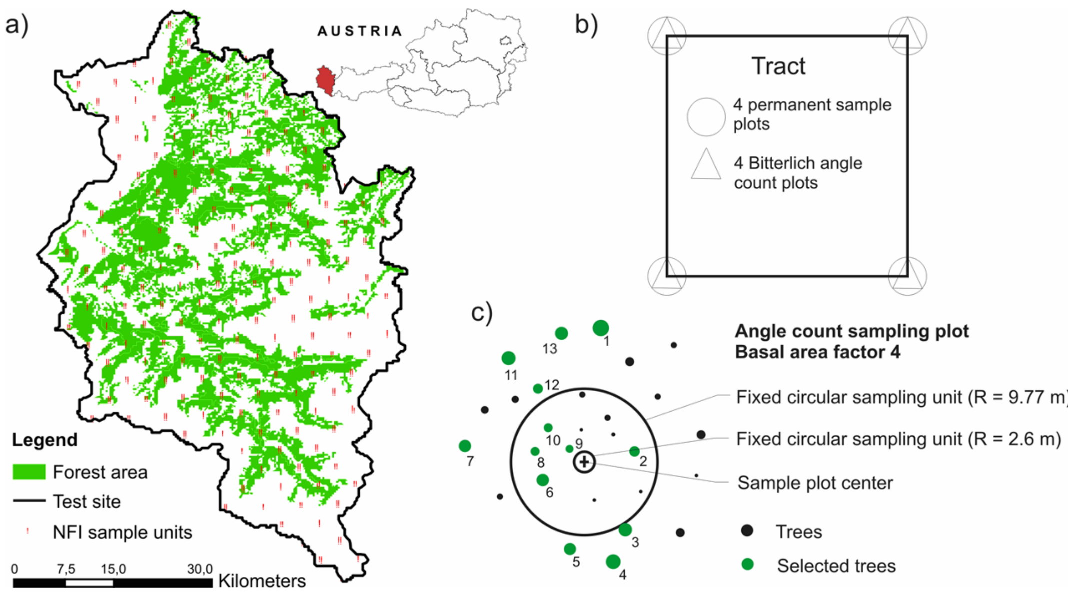

Figure 1: a) Location of Vorarlberg study site. Shown is the forested area overlain with the NFI sample units. b) Configuration of sampling units within a tract as employed at the Austrian NFI. c) Configuration of a sampling unit.

2.3 Forest inventory data

The development of the co-registration procedure was based on Austrian NFI data from the assessment period 2000/2002. The NFI is carried out in regular time intervals of six to eight years and comprises more than 170 attributes that provide information on quantity, quality and trends of the Austrian forests. The attributes relevant for this study are given in Table 1. The sampling design of the NFI is a permanent sampling grid pattern where tracts are regularly distributed (3.89 km grid size) over Austria. Each tract is made up of four sampling units spaced in a square at a distance of 200 m (Figure 1b). The single sampling units comprise a fixed large circular sampling area of 300 m² (R=9.77 m), a fixed small sampling area of 21 m² (R=2.60 m), and an angle count sampling plot (also called Bitterlich plot). While the fixed large circular plot is used to capture site specific properties, within the small sampling circle every tree with a diameter at breast height (DBH) between 50 and 105 mm is characterized. Within the angle count sampling the selection of trees is based on a relascopic measurement of DBH and consequently the plot has a variable size. A basal area factor of 4 was employed. For a subset of the sample trees heights were measured with a VERTEX III[3], while data models were used to estimate heights of the remaining sample trees (Gschwantner and Schadauer, 2004).

[3] http://www.haglofsweden.com/products/VertexIII/

Within the forested area of Vorarlberg 132 sampling units are available (Figure 1). Since no reliable dGPS measurements were available to test the accuracy of the automated co-registration results, reference centre coordinates of each sample plot were determined by manually seeking the optimum fit between tree positions and heights measured by the NFI and the CHM. To do this, the absolute positions of the trees within each plot were calculated from the geographical coordinates of the sample plot centres and the polar coordinates of the individual trees. These coordinates were then converted into ArcGIS shapefiles which, in combination with the NFI heights of each tree, facilitated a visual comparison with the CHM and finally a manual adaptation (Figure 2). 98 of the 132 sampling units could be unambiguously co-registered in this way. The errors of the measured NFI centre coordinates thus established ranged between 0.00 and 54.00 m with an average of 8.50 m.

Table 1: Attributes of the Austrian NFI that are relevant for the presented co-registration procedure.

Variable	Unit	Measurement principle
Center coordinates (X,Y) of individual sample plots	m (GK Austria meridian 28 coord. system)	Non-differential GPS. In case of bad receiving computed from GPS measurement in a nearby open space and eccentric
Polar Azimuth from plot centre	gon	compass
Distance from plot centre	cm	ultrasonic range instrument
Diameter at breast height (DBH)	mm	Calliper, Measuring tape
Tree height	dm	Ultrasonic measurement with VERTEX III
Tree type and tree class (indicating vitality, growth stage, and relationship with neighbouring trees)	-	Using key proposed by (Schieler and Hauk, 2001)

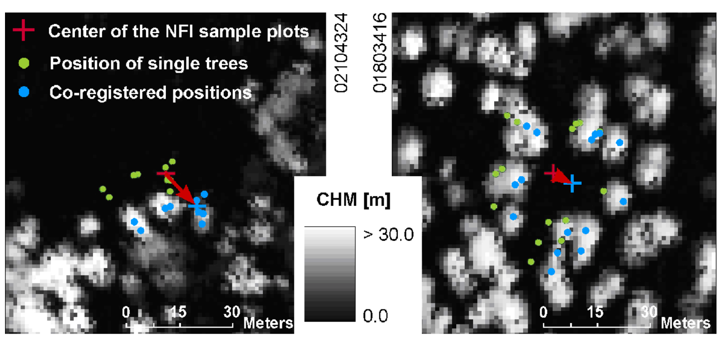

Figure 2: Relative orientation between tree positions, sample plot centre coordinates and the CHM for sample plots 02104324 and 01803416. The red vector indicates the manual shift applied to co-register NFI data to the CHM

3. Automated co-registration

3.1 Model description

An automated co-registration procedure was developed in order to overcome the manual adjustment step between NFI data and CHM described in the previous paragraph. The approach searches iteratively within a specified search window for the best fit between the tree heights measured during the NFI and the heights contained in the CHM (Figure 3-a). Thus, the height difference (*D*) for a given sample plot centre coordinate *x,y* within the search window can be given by:

$$D_{x,y} = \sum_{t=1}^{N} c_t \cdot \left| H_{NFI,t} - H_{CHM,t} \right| \tag{1}$$

where N is the number of trees measured in one NFI sample plot, $H_{NFI,t}$ is the height assessed during the NFI for tree t, and $H_{CHM,t}$ is the value of the canopy model at the location of tree t. Tree class parameter c is introduced in the cost function to account for the vigorousness of a tree and its social status with respect to the surrounding trees (Schieler and Hauk, 2001). It is thus an indicator for its "visibility" in the CHM. The tree class parameter can take a value of 1 (e.g. tree crown is part of bottom canopy layer), 2 (e.g. tree crown belongs to middle canopy layer), or 4 (e.g. predominant or solitary tree). The c factor is normalized for the total number of trees in the sample plot.

It is assumed that within an angle count plot the measured tree positions have an accuracy of ±1.0 m relative to the sample unit centre. To allow both for these small measurement errors and for the uncertainties resulting from rasterizing the ALS data the NFI tree height is compared with the highest CHM value in a 3×3 pixels (i.e. 3×3 m^2) window around the tree location.

Calculating the height difference in the proposed way only considers tree height differences but does not account for the configuration of the angle count sampling. In fact, the angle count sampling only includes those trees that at a certain distance from the sample plot centre have a minimum DBH, defined by the basal area factor. To avoid solutions that conflict with this sampling principle, the minimum tree height required to fall within the sampling was introduced. This is done as follows: For every distance from the centre coordinate the minimum required DBH is calculated. Through an empirical relationship between DBH and height (Table 2), and correcting for the uncertainty in this function, the minimum required tree height for each distance from the sample plot centre is calculated (Figure 3-c). By subtracting the minimum required tree height from the CHM subset (which is defined by the position of the sample plot centre in the search window and by the distance of the outermost tree to the sample plot centre) one obtains the parts of the tree crowns that should be included in the angle count sampling (Figure 3-b). The hypothetical tree crowns that are actually included in the angle count sampling are derived from the NFI parameters by relating crown shape and extension to BHD according to the allometric functions proposed by (Hemery et al., 2005) (Figure 3-d). Subtracting the minimum required tree height (Figure 3-c) from the simulated tree crown model provides the image that is directly comparable with Figure 3-b (Figure 3-e) and in the ideal case would look identical.

As can be seen in Figure 3-d and e the simulated crown shapes are only a rough approximation of the actual crown shapes. For this reason we decided not to compare the complete simulated and measured "visible" tree crowns but, instead, only compare the apexes of the trees while the rest of the simulated tree crown pixels were excluded in the cost function (Figure 3-f). Tree and non-tree pixels are equally weighted in the cost function, i.e. the sum of the weights attributed to the tree apices (while still accounting for social stand differences) equals the sum of all non-tree pixels (Figure 3-f). Hence, equation (1) can now be written as:

$$D_{x,y} = \sum_{p=1}^{N} c_p \cdot \left| H_{NFI,p} - H_{CHM,p} \right| \qquad (2)$$

where $H_{NFI,p}$ is pixel p in the adapted tree crown model (Figure 3-e), $H_{NFI,p}$ the equivalent pixel in the adapted CHM subset (Figure 3-b) and c the weight of the pixel according to Figure 3-f. The D-values of one sample plot is scaled between 0-1 and the coordinate x,y within the search window providing the smallest D value is eventually assumed the new, co-registered sample plot centre coordinate.

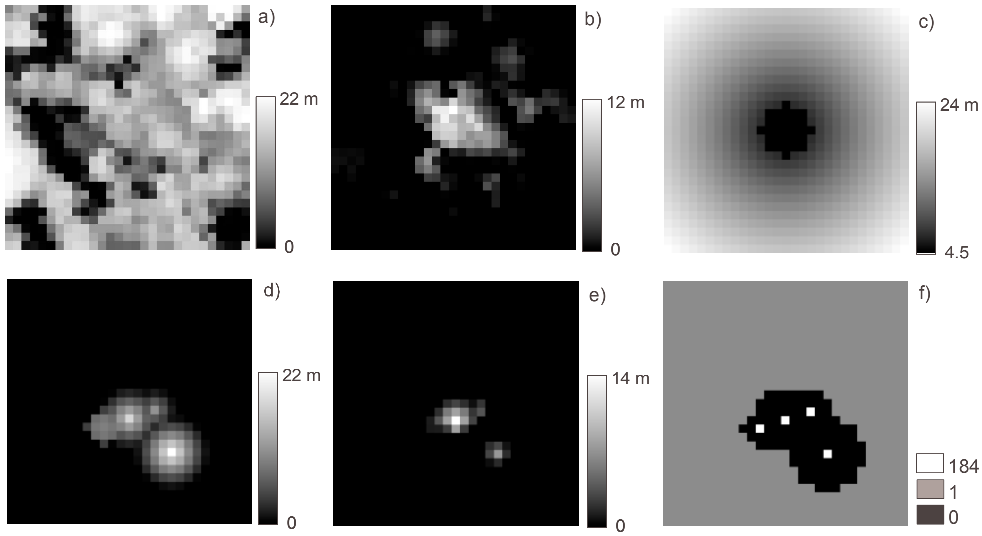

Figure 3: Example of data sets used for co-registration of sample plot 00504100: a) subset extracted from the input CHM around position *x,y* within search window; b) Difference between CHM subset and minimum tree height required to be sampled in Bitterlich plot (shown in c); d) simulated tree crown model; e) difference between simulated tree crown model and minimum required tree height (shown in c); f) weight attributed to every pixel in cost function.

Table 2: Empirical regression functions between BHD and tree height based on Austrian NFI 2000/2002.

Type	Regression function	# observations	R^2
Coniferous	Tree height = 7.3677 * $BHD^{0.5957}$	25201	0.73
Deciduous	Tree height = 14.455 * $BHD^{0.4695}$	7459	0.66
Mixed	Tree height = 8.9211 * $BHD^{0.5604}$	32660	0.70

3.2 Quality flagging

Even if the proposed iterative procedure leads to a global minimum, it is possible that due to errors in the CHM, NFI measurements, and model approximations the obtained minimum does not correspond to the actual optimum position. Identifying those sample plots that potentially have an ambiguous solution is a key element in the workflow since these samples may require manual post-processing by an image processor. Following criteria were considered when marking a solution as ambiguous. In this respect all sample plots that require manual post-processing should be included whereas as few as possible accurately co-registered samples should be included in order to reduce unnecessary quality controls by the image processor:

1. When among the smallest residuals more than one spatial cluster exists (Figure 4 – middle).
2. When residuals are sorted and plotted, a steep slope stands for an unambiguous solution while a flat slope suggests several plausible solutions. (Figure 4 – right)
3. When sample plot has a predominance of deciduous trees, since these have less pronounced tree apexes than conifers and were acquired under leaf-off conditions
4. When distance between original centre coordinate and co-registration result is larger than 20 m.

4. Results

The accuracy of the automated co-registration procedure was investigated by computing the distances between the automatically co-registered sample plot centre coordinates and the centre coordinates that could be unambiguously manually allocated by the image processor. For visual interpretation the distances were sorted in ascending order (Figure 5a). The figure shows that 67 of the 98 sample plots (i.e. 68%) were correctly co-registered, with a distance to the manually obtained results ranging between 0.08 and 3.64 m. The causes of several sample plots not being correctly co-registered (defined as those with a distance included the issues already pointed out in paragraph 3.2, i.e. the presence of multiple solutions and the predominance of deciduous trees (Figure 5c). In addition, two of the incorrectly co-registered sample plots had a manual solution outside the iteration search window (60×60 m^2) and also CHMs with a point density of less than 1 point/m^2 appeared problematic (Figure 5b).

Figure 5 additionally shows the results of the quality flagging. It can be seen that all of the points with a deviation > 4 m were marked "ambiguous", leading to an omission error of 0%. Similarly, all plots with a deviation < 4 m were marked "unambiguous". Hence, the overall accuracy of quality flagging amounts 74%. In contrast, 26 plots showing only small deviations from manual co-registration results were incorrectly tagged as "unambiguous", leading to a commission error of 45%. As a consequence, these 26 samples will be superfluously controlled during post-processing.

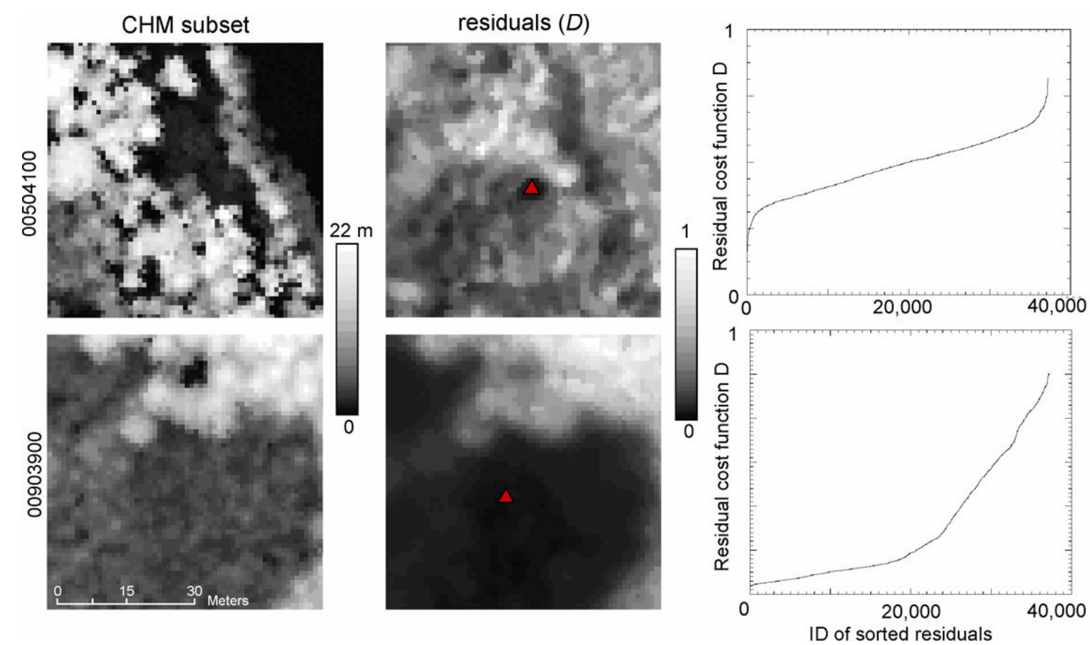

Figure 4: Quality measures considered during co-registration, demonstrated for sample plot 00504100 (top) and 00903900 (bottom): the spatial distribution of residuals elucidates if more than one or very large clusters of minimum residuals exist (middle), the red triangle indicates the absolute minimum; the slope of the sorted residuals at the smallest absolute D-value indicates if the found absolute minimum is likely to be the global minimum or if several local minima exist (right).

5. Discussion and conclusion

With a correct co-registration of 67% of the sample plots the automated algorithm has the potential of significantly reducing pre-processing efforts in order to obtain more accurate ALS based predictive models. This is best illustrated with a practical example. For this purpose we calibrated and cross validated the growing stock model of (Hollaus et al., 2008) for 3 different co-registration states of the NFI, using i) the original, ii) the automatically co-registered, and iii) the manually co-registered sample plot centre coordinates. The selection of centre coordinates was based on the 41 sample plots that during automated co-registration were marked "unambiguous". Calibration and cross validation was based on in situ growing stock measurements collected at each sample plot within the framework of the Austrian NFI (Gabler and Schadauer, 2006). Four sample plots were excluded from growing stock measurements and model calibration due to the absence of trees with sufficiently large DBH.

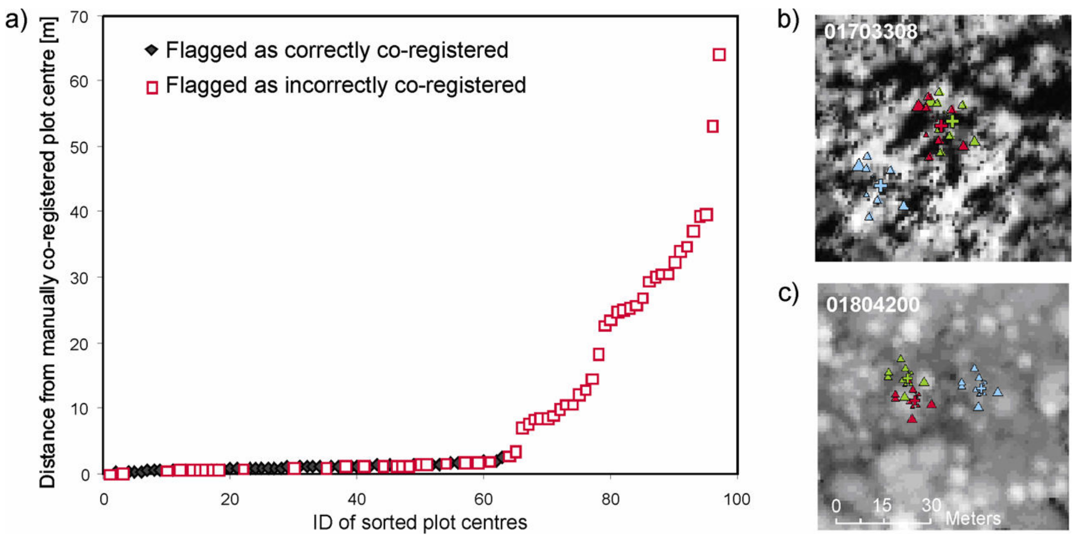

Figure 5: a) Distance between automatically and manually co-registered sample plot centre coordinates, sorted in ascending order. Black diamonds indicate the points that during co-registration were flagged "unambiguous" while red squares were marked "ambiguous". b, c) Examples of incorrectly co-registered sample plots. b) shows a plot dominated by deciduous trees, c) is characterized by an insufficient ALS point density. Red triangles (crosses) show the original tree (centre point) locations, blue the results of the automated, and green the results of the manual co-registration.

Figure 6 shows that using the automatically co-registered data yields significant improvement (both R^2 and relative standard deviations (SD) obtained by cross validation) compared to the original sample plot centre coordinates and even slightly outperforms the accuracy obtained when using the manually co-registered sample plot centre coordinates.

The above example illustrates the practical relevance of adequate co-registration between FI and ALS data in general and the potential of the automated algorithm in performing this task in particular. Moreover, the quality flagging allows the user to identify those results that should be treated with precaution or require manual post-processing. Future efforts will concentrate on testing the developed algorithm on other data sets. In this context, a higher overall accuracy is expected when ALS data with a higher point density is used. Since the developed algorithm will be part of an operational processing chain for Austrian NFI data, it has a high practical relevance.

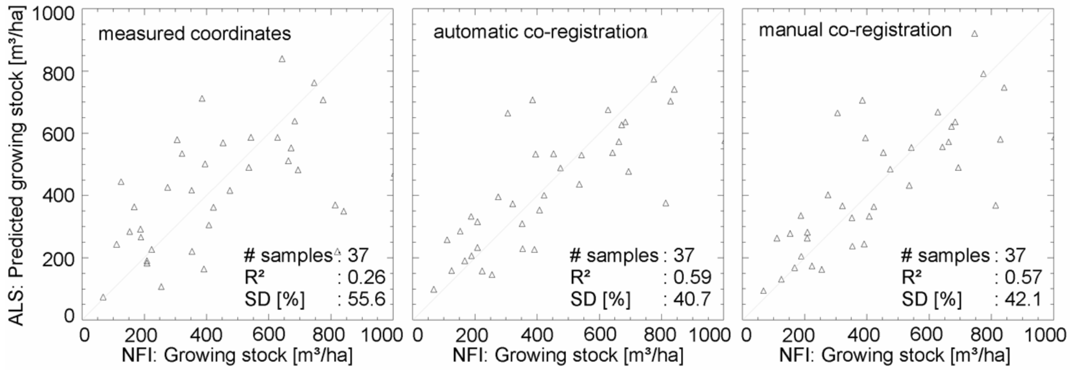

Figure 6: Effect of co-registration on calibration of ALS based growing stock model of (Hollaus et al., 2008). Left plot shows the results when original sample center coordinates measured by GPS are used, the middle (right) plot when automatically (manually) co-registered coordinates are used.

Acknowledgements

We would like to thank the Landesvermessungsamt Feldkirch for granting the use of the ALS data. Parts of this study were funded by the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management (BMLFUW) in the framework of the project ÖWI-Regio.

References

Bitterlich, W., 1948. Die Winkelzählprobe. *Allgemeine Forst-und Holzwirtschaftliche Zeitung*, 59 4-5.

Farid, A., Goodrich, D.C. and Sorooshian, S., 2006. Using airborne lidar to discern age classes of cottonwood trees in a riparian area. *Western Journal of Applied Forestry,* 21(3), 149-158.

Gabler, K. and Schadauer, K., 2006. Methoden der Österreichischen Waldinventur 2000/02 - Grundlagen, Entwicklung, Design, Daten, Modelle, Auswertung und Fehlerrechnung, *BFW-Berichte*, 135, 132 pp.

Gschwantner, T. and Schadauer, K., 2004. Datenmodelle der Österreichischen Waldinventur 2000/2002. BFW Dokumentation, 4, 76 pp.

Hemery, G.E., Savill, P.S. and Pryor, S.N., 2005. Applications of the crown diameter-stem diameter relationship for different species of broadleaved trees. *Forest Ecology and Management*, 215(1-3), 285-294.

Hollaus, M., Dorigo, W., Wagner, W., Schadauer, K., Höfle, B. and Maier, B., 2008. Large-area growing stock estimation based on airborne laser scanning and national forest inventory data. *International Journal of Remote Sensing*, Submitted.

Hollaus, M., Wagner, W., Maier, B. and Schadauer, K., 2007. Airborne laser scanning of forest stem volume in a mountainous environment. *Sensors*, 7(8), 1559-1577.

Holmström, H., Nilsson, M. and Ståhl, G., 2001. Simultaneous Estimations of Forest Parameters using Aerial Photograph Interpreted Data and the k Nearest Neighbour Method. *Scandinavian Journal of Forest Research*, 16, 67-78.

Koukal, T., 2004. Nonparametric Assessment of Forest Attributes by Combination of Field Data of the Austrian Forest Inventory and Remote Sensing Data, PhD thesis, Universität für Bodenkultur, Vienna, 115 pp.

Maltamo, M., Packalén, P., Peuhkurinen, J., Suvanto, A., Pesonen, A. and Hyyppä, J., 2007. Experiences and possibilities of ALS based forest inventory in Finland, *ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*, Espoo, Finland.

Schieler, K. and Hauk, E., 2001. Instruktion für die Feldarbeit - Österreichische Waldinventur 2000/2002, Dienstanweisung, FBVA, Wien, 199.