# ROBUST FILTERING OF AIRBORNE LASER SCANNER DATA FOR VEGETATION ANALYSIS

W. Wagner [a], C. Eberhöfer [b], M. Hollaus [a], G. Summer [a]

[a] Christian Doppler Laboratory for "Spatial Data from Laser Scanning and Remote Sensing", Vienna University of Technology, Gusshausstrasse 27-29, 1040 Vienna, Austria - ww@ipf.tuwien.ac.at

[b] Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Gusshausstrasse 27-29, 1040 Vienna, Austria

**KEY WORDS:** Laser scanning, Lidar, Vegetation, DEM/DTM, Forestry, Analysis

**ABSTRACT:**

Airborne laser scanning (ALS), often referred to as lidar or laser altimetry, is a remote sensing technique which was originally designed to measure the topography of the Earth's surface. While the first commercially available airborne laser scanners recorded only the time of one backscattered pulse, state-of-the-art systems record several echoes for each emitted laser pulse. Thereby a 3D data cloud is obtained which conveys valuable information about the vegetation canopy. For the retrieval of vegetation parameters the most common procedure is to 1) calculate a digital terrain model (DTM) by filtering last-pulse ALS data, 2) form a digital surface model (DSM) from first-pulse ALS data to represent the topmost surface (top of vegetation, building roofs, etc.), and 3) calculate a normalised digital surface model (nDSM) to represent the height of the vegetation. The so derived vegetation height model can be used as input for further vegetation analysis. In this study we investigate the quality of DTMs and nDSMs derived from first/last-pulse ALS data in an alpine environment. We use a hierarchic robust filtering technique for DTM generating from ALS data obtained during leaves-off (December) and leaves-on (July) conditions. The derived DTMs compare well for flat, non-vegetated terrain. Over forested terrain it is found that the penetration rates are much higher in winter compared to summer, even for forest patches dominated by spruce. As a result of the low forest penetration in summer, high differences between the summer and winter DTMs were observed over forested, steep terrain. Another consequence of the different penetration rates is that only the summer DSM correctly represents the top of canopy. This shows that, ideally, a vegetation height model is obtained by subtracting a winter DTM from a summer DSM.

## 1. INTRODUCTION

Airborne laser scanning (ALS) is an active remote sensing technique where a laser emits short infrared pulses towards the Earth's surface and a photodiode measures the backscattered echo. When there are several objects within the travel path of the laser pulse, multiple echoes are recorded. Therefore, state-of-the-art laser scanner systems measure the round-trip time of first- and last pulse, some are capable of recording up to five pulses. In the near future, commercial systems that record the full waveform of the backscattered signal will become available (Wagner et al., 2004).

Due to the capability of ALS to provide 3D point clouds over vegetated areas, their application to vegetation mapping appears obvious. Indeed, there have been a number of studies that demonstrated the high potential of ALS for forestry applications. Hyyppä and Hyyppä (1999) used a first/last pulse ALS scanner system, flown at an altitude of 400 m and a measurements density of more than 10 points per m$^2$, to retrieve mean tree height, basal area, and volume for forest stands ranging in size between 0.5 and 5 hectares in Finland. Hyyppä and Ikinen (1999) used the same flight to locate individual trees and to estimate their height and crown area. Næsset (2002) employed statistical procedures to derive stand attributes over a site in Norway. All three studies concluded that ALS can achieve accuracies comparable or even better than conventional field inventories over boreal forest sites. Also in other climatic zones encouraging results have been achieved (Schardt et al., 2002; Clark et al., 2004; Popescu and Wynne, 2004).

For the retrieval of vegetation parameters from ALS 3D data clouds the most commonly used procedure is to calculate a height model of the vegetation by subtracting a digital terrain model (DTM) from a digital surface model (DSM), which includes both terrain and non-terrain points (top of canopy, roofs, etc.). The resulting difference model is commonly referred to as normalised digital surface model (nDSM), and when used for vegetation analysis, digital crown model (Hyyppä and Inkinen, 1999) or canopy height model (Popescu and Wynne, 2004). Normalised digital surface models often reveal a wealth of information about the vegetation cover as the example presented in Figure 1 demonstrates. The figure shows a nDSM with a raster width of 1 m along with an infrared orthophoto of an alpine forest in Vorarlberg, Austria. In the nDSM vegetation patches of different height (trees, bushes, grass) can be readily identified, in some cases it is possible to locate individual trees. While the orthophoto conveys more thematic information (plant species, road, etc.), the lower degree of complexity of the nDSM is also an advantage as it simplifies the task of retrieving vegetation parameters in an automatic fashion. A nDSM is less complex than an orthophoto because height is an unambiguous physical quantity, while surface reflectivity is a complex function of illumination conditions, surface properties, and imaging geometry. Also, since laser and photodiode are at the same position, laser scanning does not suffer from shadow effects.

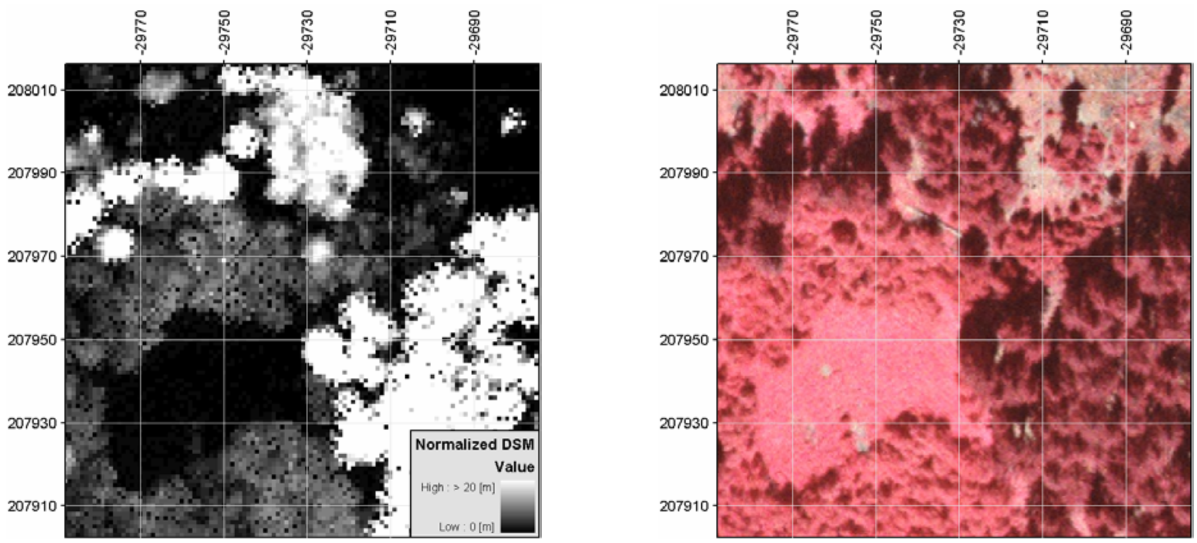


Figure 1. Normalised digital surface model from airborne laser scannering (left) and aerial infrared photography (right) for an alpine forest site in Vorarlberg, Austria. Coordinates are in Gauß-Krüger (reference meridian M28). Data are courtesy of the Landesvermessungsamt Feldkirch.

The usefulness of a nDSM for vegetation retrieval depends on the accuracy of the DTM that is derived from the 3D data cloud by means of filtering techniques. The aim of filtering is to classify the 3D data cloud into terrain and off-terrain points (Pfeifer, 2003). Once the terrain hits are classified, the generation of gridded or triangulated irregular network DTMs is straight forward. In other words, the challenge in the generation of DTMs from laser scanner data is the correct identification of ground hits. This presents a problem because with current ALS systems it is essentially unknown where the recorded echoes are coming from. Therefore physically based approaches are not feasible which necessitates the use of statistical approaches that model, in some way or other, the spatial relationship of the 3D data cloud. A good overview over different filtering strategies is given in Pfeifer (2004). Here we use a hierarchic robust filtering technique described in Kraus and Pfeifer (1998) and Briese et al. (2002).

During the initial years of lidar mapping efforts the accuracy of DTMs was often quoted to be 15-20 cm. Pilot studies have shown that even in wooded terrain accuracies in the order of ±25 cm can be achieved (Kraus and Pfeifer, 1998). Hodgson and Bresnahan (2004), who analysed the accuracy for six different land cover classes (pavement, low and high grass, brush/low tree, evergreen and deciduous forest), also report low root mean square errors in the range from 17 to 26 cm. However, they point out that such good accuracies are only achievable under ideal conditions and call for numerous research efforts to quantify the accuracy of ALS-derived DTMs in dependence of flight parameters and environmental conditions. Also Raber et al. (2002) demonstrate the potentially adverse impacts of different vegetation classes on DTM accuracy and argue for developing adaptive filtering techniques that account for variable vegetation cover.

To explore the suitability of ALS data for vegetation mapping in alpine environments this study investigates the robust filtering technique in an alpine valley in Vorarlberg, Austria. Our methodological approach, as described in Section 3, is to use ALS data from leaves-off (December 2002) and leaves-on (July 2003) conditions and to compare the winter and summer models to identify problem areas. The hierarchic robust filtering is described in the next section, the available data and study area in Section 4. The results and discussion are presented in Section 5, followed by the conclusions in Section 6.

## 2. HIERARCHIC ROBUST FILTERING

The hierarchic robust filtering technique used in this study is implemented in the commercial software package SCOP++, a joint development of the company Inpho, Germany, and the Institute of Photogrammetry and Remote Sensing of the Vienna University of Technology, Austria. The robust filtering procedure was introduced in Kraus and Pfeifer (1998). Its extension into an hierarchic framework is described in detail at http://www.ipf.tuwien.ac.at/eurosdr/ and in Briese et al. (2002).

The hierarchic filtering strategy employs four different processing steps referred to as *thin out*, *interpolate*, *filter*, and *sort out*. These four steps can be applied consecutively, whereby there are few rules that restrict the order of application or the number of iterations. *Thin out* refers to a raster based thinning algorithm which lays a grid over the complete data domain and selects one point (e.g. the lowest) for each cell. In the *interpolate* step a terrain model is derived from the current data set by interpolation without differentiating data points. Also in the *filter* step a terrain model is computed, but this time a weighting function designed to give low computational weight to likely off-terrain points and high weight to likely terrain points is used. In the *sort out* step only data points within a certain distance from a previously calculated DTM are retained. Finally, it is noted that manual editing (e.g. manual elimination of off-terrain points) is possible.

Depending on the characteristics of the ALS data set different filtering strategies can be devised. The flexibility in the design of the strategy, combined with possibility to select a number of parameters in each processing step, has the advantage that, with the exception of few problem areas, satisfactory results can be achieved without manual editing. The disadvantage is that it is difficult to predict how changes in the filtering strategy or parameter setting affects the final DTM. Should one of the default strategies offered in SCOP++ not produce satisfactory results then even experienced interpreters sometimes need many working hours to experiment with different filtering strategies. A sensitivity analysis, i.e. to study the variation in the DTM output to variations in the filtering strategy and parameter settings, may at least give some confidence in converging solutions, while filtering strategies yielding singular results would appear suspicious. Unfortunately, a sensitivity analysis is extremely difficult to realise with SCOP++ because of the high degree of flexibility of the filtering strategy and the large number of parameters (size of analysis area, interpolation method, number of data pyramids, filter weight function, sort out tolerance range, number of iterations, etc.).

Since current ALS systems do not allow to distinguish terrain from off-terrain points based on physical observables, there is no possibility to define physically-based criteria to signal how many of the laser echoes have been correctly classified. Therefore the decision on which filtering strategy to use best depends on the expert judgement of the human interpreter. New ALS systems, which provide calibrated intensities of the laser echoes or which record the full echo-waveform, may provide crucial information to guide the human interpreter in the selection. Once all off-terrain points have been removed from the ALS data set, quality measures as defined by Kraus et al. (2004) can be computed to visualise the spatially variable accuracy of DTMs.

## 3. METHODOLOGY

In order to be useful for vegetation analysis the quality of the derived DTM must be comparable for different vegetation cover types, from low grass to dense forest. Two critical questions are: 1) Are there enough ground hits to reconstruct the DTM? 2) Is the filtering method capable of correctly classifying these ground hits? The number of ground hits does not just depend on the physical penetration of infrared pulses into the vegetation, but also on the technical characteristics of the ALS system (laser point density, range resolution, power, footprint size, frequency, etc.) and the employed pulse detection algorithm (Wagner et al., 2004). With respect to the filtering method, it should be expected that there is no unique filter which does equally well for all vegetation classes. Rather, as pointed out by Raber et al. (2002), it is advisable to use vegetation-adaptive filters. However, in the current implementation of SCOP++ (version 5.2) it is only possible to apply one filter globally for the selected working unit, which normally represents an area of about 1-3 km$^2$.

The best way to verify classified ground hits is by locating x-y coordinates of ALS points in the field and measuring the z-coordinates with GPS techniques, as done by Hodgson and Bresnahan (2004). Since the collection of statistical representative ground observations is expensive, particularly if several land cover classes are considered, and since access to field sites is not always possible in alpine regions, it would nevertheless be highly desirable to have alternative validation strategies. Here we follow an approach, in which we compare ALS-derived models (DTM, DSM, nDSM) of an area acquired at two different times. If the derivation of the two DTMs would be error free then the difference-DTM would only show small, random deviations. Large or spatially consistent deviations would indicate problem areas, which also propagate into the nDSM. For the DSM, and consequently also for the nDSM, differences are expected which should reflect changes in vegetation between the two acquisitions.

## 4. STUDY AREA AND DATA SETS

As study area a 675 × 475 m$^2$ large area situated in the alpine valley Gargellental in Vorarlberg, Austria, was chosen (Figure 2). The study site covers a part of the village Winkel. Two small mountain streams run through the study area from west to east, where they discharge into the Suggadinbach. The elevation ranges between 847 and 1071 m, with some steep slopes particularly along the river beds. The land cover is characterised by meadows and mixed forest. The dominated tree species in the region is the Austrian spruce. In the Gauß-Krüger coordinate system (reference meridian M28, WGS84) the lower left corner of the study area is at -29125 m easting and 5209425 northing.

The area was chosen because it was covered by two ALS flights performed by the company TopScan, Germany. The first flight took place on December 10, 2002, the second on July 19, 2003. These two flight campaigns were carried out in an effort to map the complete Gargellental. The valley bottom was covered by the winter flight under snow-free conditions, while the higher altitudes up to about 3000 m were covered during the summer campaign. TopScan employed first/last pulse Airborne Laser Terrain Mapper (ALTM) systems from Optech Inc., Canada. During the winter flight the pulse repetition rate was 25 kHz, for the summer flight 50 kHz. Consequently, the average point density on ground was higher for the summer campaign (2.7 points per m$^2$) than for the winter campaign (0.9 points per m$^2$) The mean flying height was 1000 m and the maximum scan angle 20 deg. As a result the swath width was about 725 m. The overlap between flight lines was 425 m. The ALTM systems operate at laser frequencies of about 1 μm (Baltsavias, 1999). The ALTM beam divergence is 0.3 mrad, resulting in a laser footprint size of 0.3 m at a flying height of 1000 m.

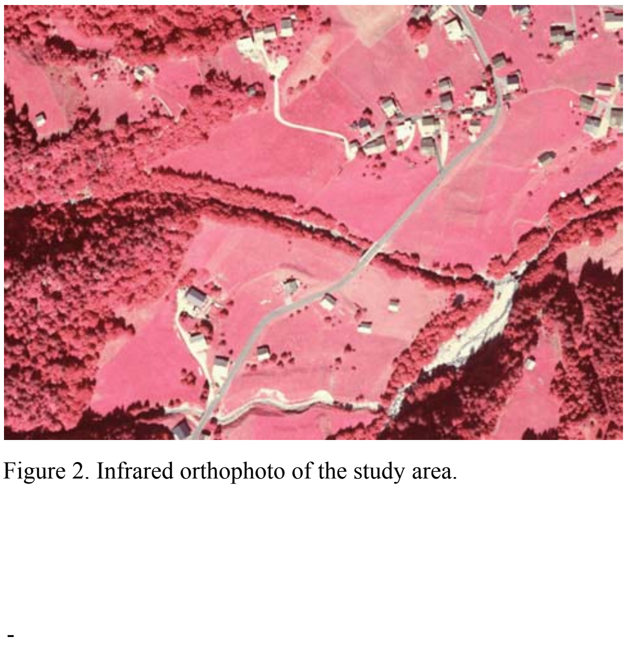

Figure 2. Infrared orthophoto of the study area.

## 5. RESULTS AND DISCUSSION

For both the winter and summer data the same procedures were followed in the computation of the different models. In particular, the same filtering strategy was applied to both last-pulse data sets.

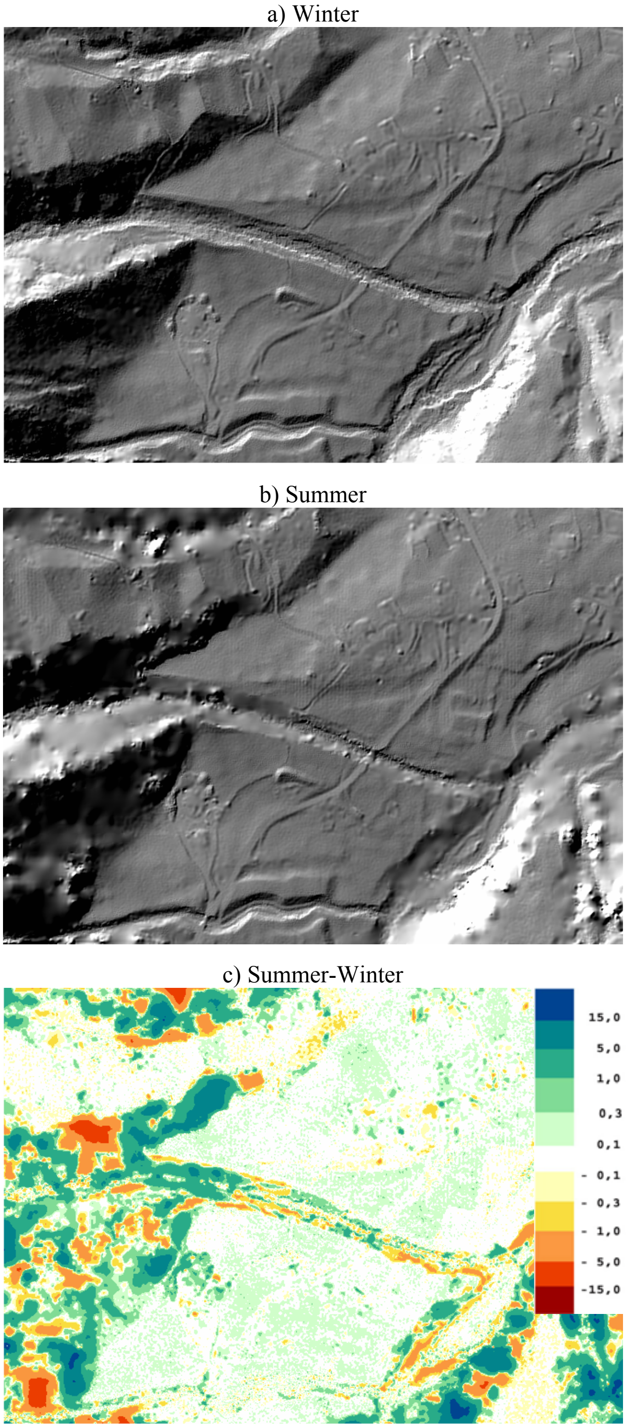


Figure 3. Digital terrain model (DTM) derived from the ALS last-pulse winter (a) and summer (b) data. The summer-winter difference model is shown in (c). Differences are given in meters. Be aware of non-uniform ranges in the colour bar.

Figures 3 and 5 show the derived DTM and nDSM models for both winter (leaves-off) and summer (leaves-on) conditions, together with the difference models (summer-winter). The DSM models are not shown for lack of space.

The comparison of the summer and winter DTMs shows a high degree of agreement for the non-forested areas (Figure 3). There the differences are generally below 0.3 m, large parts are below 0.1 m. In few cases field boundaries can be observed in the difference DTM plot. This suggests that different penetration rates into high and low grass are not entirely accounted for by the non-adaptive filtering strategy.

In the forested areas, which are often associated with steep slopes, large differences up to ±15 m are observed (Figure 3c). While, due to reduced penetration rates in summer, one would expect the summer DTM to be in general higher than the winter DTM, negative and positive differences are almost equally abundant. This can be explained by considering the number of classified ground hits, as shown in Figure 4. One can observe that while for winter conditions there is a sufficient number of ground hits also in the forested area (blue dots), there are very few ground hits in summer (green dots). As a result there were too few points for the computation of the summer DTM, which caused under- and overshooting of the interpolated DTM particularly in steep, forested terrain. This is reflected by the smooth appearance of the summer DTM in these areas (Figure 3b).

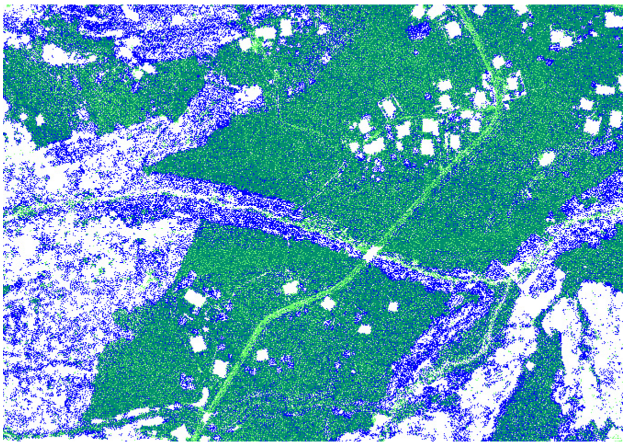

Figure 4. Number of ground hits estimated by robust filtering. Blue indicates ground hits acquired during the winter flight, green during the summer flight.

The large difference between the number of estimated ground hits in winter and summer surprises somewhat, given that the forests are mixed deciduous-evergreen. The fact that the same filtering strategy was applied to both data sets excludes the possibility that a too rigid filtering strategy was selected for the summer data. The much higher penetration rate in winter is also apparent in the nDSM models derived from first-pulse DSM and last-pulse DTM models (Figure 5). The difference-nDSM model exhibits height differences of generally more than 5 m, which reflects the change from leaves-off to leaves-on conditions. However, also errors contained in the DTM models, particularly for the summer data, are propagated through to the nDSM models. For example, the negative values in the lower-right corner of the difference nDSM model are a reflection of the positive differences at this position in the DTMs (compare Figures 3c and 5c). These observations not only suggest that a large percentage of the trees in the study area are deciduous, but also that, possibly, evergreen trees are more transparent for infrared laser pulses in winter than in summer.

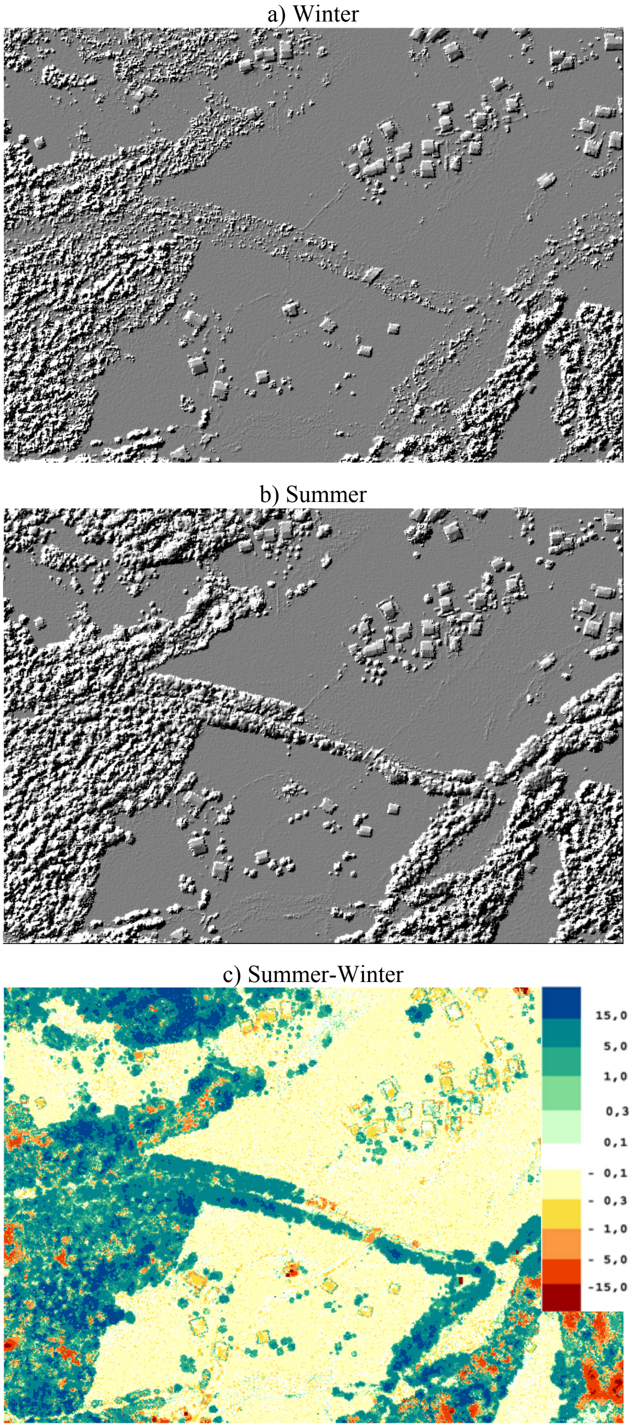


Figure 5. Normalised digital surface model (nDSM) derived from the ALS first-pulse DSM and last-pulse DTM for winter (a) and summer (b) conditions. The summer-winter difference model is shown in (c). Differences are given in meters. Be aware of non-uniform ranges in the colour bar.

Let us come back to Figure 4 which shows the number of estimated ground hits. Over the non-forested areas the higher point density acquired during the summer flight is obvious. In addition, an interesting phenomena is apparent over the main street that runs through the study area in south-northerly direction. While for the summer flight the points density is as large as over the surrounding areas, there are few street echoes in the winter data. Possibly the street was wet during the winter acquisition, thereby further reducing the already low reflectivity of asphalt (the reflectance of asphalt is lower than 0.2 according to Jelalian, 1992). But more likely, ALS system parameters were different for the summer and winter flights (e.g. intensity of laser pulse, flying height, etc.) resulting in a reduced sensitivity to record echoes from the asphalt during the winter flight. This means that over the forested area large differences in penetration rate are observed despite the much higher point density in summer compared to winter (2.7 versus 0.9 points per $m^2$) and despite a presumably higher echo sensitivity in summer.

To accurately represent the height of the vegetation canopy it is clear that neither the winter nor the summer nDSM are the best choice. In the winter case, the DTM is of high quality, but the DSM is poorly defined. In the summer case, the first pulse can be assumed to represent the top of the canopy reasonably well, while the reduced penetration causes a low accuracy of the DTM. For vegetation analysis the best nDSM is the one obtained by subtracting the winter DTM from the summer DSM. The so derived nDSM as shown in Figure 6 is not too different from the summer nDSM shown in Figure 5b. However, it is clear that vegetation height is more accurate, particularly in steep terrain.

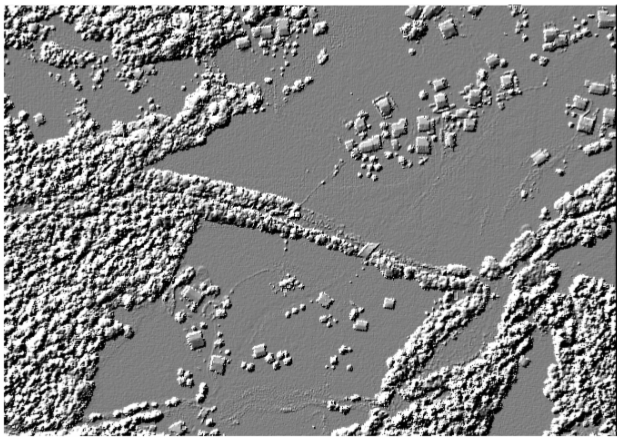

Figure 6. Normalised digital surface model (nDSM) derived by subtracting the winter-DTM from the summer DSM.

## 6. CONCLUSIONS

In this study we have investigated the quality of digital terrain models (DTMs) derived from ALS data using the hierarchic robust filtering technique in an alpine environment. High-quality DTMs are a necessary prerequisite to compute vegetation height models, which represent valuable input for further vegetation analysis (biomass retrieval, structural analysis, etc.). The 675 × 475 $m^2$ large study area was covered by two ALS flights, one in winter 2002 and one in summer

2003. The results showed that the robust filtering technique performs well. However, while under winter conditions a sufficient number of ground hits were recorded also in the forested areas, too few ground points were available for the summer flight to properly reconstruct the forest floor. As a result, high differences between the summer and winter DTMs were observed over forested, steep terrain. The best-quality vegetation height model is obtained by subtracting the winter DTM from the summer DSM, which can be assumed to represent the top of canopy well. Visually, the so derived vegetation height model is of astounding quality. In future work we will investigate its utility of vegetation height maps for estimating biomass and other important forest parameters.

The study also showed that further understanding of the physical interaction of the laser pulses and the vegetation is needed. The results suggest that the penetration rate is quite different for winter and summer conditions, even for forest patches dominated by spruce. A better understanding of the penetration rates for different tree species in dependence of phenological state and ALS system parameters would be useful in order to be able to critically assess the correctness of this result. In general, a major disadvantage of current ALS systems is that there is no physical basis for distinguishing terrain from off-terrain points. New ALS systems, which provide calibrated intensities or the full-waveform of the backscattered echoes, may allow to carry out a rough classification of the laser echoes into terrain and off-terrain points, thereby making the ALS processing less dependent on the experience of the human interpreter.

## ACKNOWLEDGEMENTS

We would like to thank the Landesvermessungsamt Feldkirch for granting the use of the ALS data and the infrared orthophoto for the purpose of this study and Karl Kraus and Christian Briese for their critical comments on the manuscript.

## REFERENCES

Baltsavias, E. P., 1999. Airborne laser scanning: existing systems and firms and other resources. *ISPRS Journal of Photogrammetry & Remote Sensing*, 54, pp. 164-198.

Briese, C., N. Pfeifer, and P. Dorninger, 2002. Applications of the Robust Interpolation for DTM determination. In: *International Archives of Photogrammetry and Remote Sensing*, Photogrammetric Computer Vision (PCV'02), 9-13 September 2002, Graz, Austria. http://www.isprs.org/publications/ archives.html.

Clark, M. L., D. B. Clark, and D. A. Roberts, 2004. Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sensing of Environment*, 91(1), pp. 68-89.

Hyyppä, H., and J. Hyyppä, 1999. Comparing the accuracy of laser scanner with other optical remote sensing data sources for stand attribute retrieval. *The Photogrammetric Journal of Finland*, 16(2), pp. 5-15.

Hyyppä, J., and M. Inkinen, 1999. Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, 16(2), pp. 27-42.

Jelalian, A. V., 1992. Laser radar systems. Artech House, Bosten, London, 292 p.

Kraus, K., C. Briese, M. Attwenger, and N. Pfeifer, 2004. Quality measures for digital terrain models. In: *International Archives of Photogrammetry and Remote Sensing*, XXth ISPRS Congress, Istanbul, Turkey, 12-23 July 2004, CDROM.

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

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

Pfeifer, N., 2003. Oberflächenmodelle aus Laserdaten. *Österreichische Zeitschrift für Vermessung & Geoinformation*, 4, pp. 243-252.

Popescu, S. C., and R. H. Wynne, 2004. Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering & Remote Sensing*, 70(5), pp. 589-604.

Raber, G. T., J. R. Jensen, S. R. Schill, and K. Schuckman, 2002. Creation of digital terrain models using an adaptive lidar vegetation point removal process. *Photogrammetric Engineering & Remote Sensing*, 68(12), pp. 1307-1315.

Schardt, M., M. Ziegler, A. Wimmer, R. Wack, and J. Hyyppä, 2002. Assessment of forest parameters by means of laser scanning. In: *International Archives of Photogrammetry and Remote Sensing*, Photogrammetric Computer Vision (PCV'02), 9-13 September 2002, Graz, Austria. http://www.isprs.org/ publications/archives.html.

Wagner, W., A. Ullrich, T. Melzer, C. Briese, and K. Kraus, 2004. From single-pulse to full-waveform airborne laser scanners: Potential and practical challenges. In: *International Archives of Photogrammetry and Remote Sensing*, XXth ISPRS Congress, Istanbul, Turkey, 12-23 July 2004, CDROM.