Can biodiversity study benefit from information on the vertical structure of forests? Utility of LiDAR remote sensing

Matthias Dees*, Christoph Straub and Barbara Koch

Department of Remote Sensing and Landscape Information Systems, Albert-Ludwigs Universität Freiburg, 79085, Freiburg, Germany

Light detection and ranging (LiDAR) remote sensing offers new and improved capabilities for vertical and structural characterization of objects, such as plant height and different growth stages/strata, etc. Climate change will change tree and plant growth conditions and thus will change forest structure and distribution, which can potentially be studied using LiDAR remote sensing. LiDAR data were acquired during leafless season that was subjected to generation of digital terrain model, digital surface model, vegetation height model, forest height map, top height of forest stands, detection of gaps, detection and mapping of tree stands, mapping of density classes of middle layer and understorey detection of coniferous trees. The stand height and structural information derived from the LiDAR imagery would greatly contribute to the characterization of biodiversity through vertical stratification.

Keywords: Biodiversity, forest gaps, LiDAR forest monitoring, stand height, vertical structure.

Introduction

LIGHT detection and ranging (LiDAR) remote sensing is a breakthrough technology for forest studies and holds promises for biodiversity characterization in combination with traditional optical sensor-based assessments. Typical traditional remote sensing images allow analysing various attributes of forests, but are limited in their ability to represent spatial patterns only in two-dimensional space. The advantage of using LiDAR remote sensing for forestry applications is that it provides data on three-dimensional forest structures characterizing vegetation height, vertical distribution of canopy material, crown volume, sub canopy topography, vertical foliage diversity and multiple layers, height to live crown, large tree density, leaf-area index and physiographic or life-form diversity through direct and indirect retrievals[1–5]. As vertical component (z-axis) measurement is the backbone of LiDAR technology, this characteristic has been exploited in a straightforward way for tree height estimation in comparison to the ground. Tree canopy height is obtained by subtracting the elevation of the first and last returns. Vegetation height is a function of species composition, climate and site quality, and can be used for land cover classification or in conjunction with vegetation indices[2,3]. Coupled with species composition and site quality information, height serves as an estimate of stand age or successional stages. Since the vertical components of stands change with age, older stands can be characterized by canopy gaps. And trees of multiple ages and sizes exhibit an even distribution of canopy components.

In the past, several experiments were carried out to determine various tree height metrics and stem numbers by different air-borne laser profiling and scanning systems; wherein majority of the forest stand characteristics have focused on old forest stands or forests where the mean tree height exceeds about 15 m. Næsset and Bjerknes[4] found that the maximum height value of laser canopy hits for a certain fixed area could be used to estimate the mean tree height. Since crown shape, tree species and density that affect the relationship between the distribution of canopy height and tree height tend to vary among different plots, it is probably useful to model the mean height of dominant trees by means of several such quantiles. Other variables of the distribution of canopy heights such as the mean and median values, standard deviation divided by the mean (coefficient of variation) and various quantiles have been found to be correlated with mean tree height, dominant height and other biophysical properties[4]. The procedure for estimating dominant height of entire forest stands was based on the assumption that for a certain cell size, i.e. a fixed area, there exists a certain relationship between the tree height of interest and predictor variables derived from the height distribution of laser pulses classified as canopy hits[2]. There are many difficulties in determining tree height using LiDAR data: (i) Determining the exact elevation of the ground surface poses difficulties for both discrete and waveform LiDAR. (ii) In complex canopies, elevation returned from what appears to be the ground level in fact may be from understorey, if the understorey is dense enough to substantially obstruct the ground surface. (iii) Each type of LiDAR system represents difficulties in detecting the uppermost portion of the plant canopy.

*For correspondence. (e-mail: matthias.dees@felis.uni-freiburg.de)

(iv) Underestimation of canopy height – (a) with discrete return LIDAR, very high footprint densities are required to ensure that the highest portion of individual tree crowns is sampled, and (b) with waveform sampling system, a large footprint is illuminated increasing the probability that treetops will be illuminated by the laser[5]. Here, we have derived the tree height ranges and indicated them to different growth stages with ground-based validation, thereby prescribing the products for various conservation purposes and arguing their utility in the study and characterization of forest biodiversity, whereas the specific context was a specific nature-protection scheme established in the European Union under the name NATURA 2000 network[6].

Study area and methods

A region of 23.1 sq. km, mainly covered by forest was selected for the study. It is located in the Rhine Valley, southwest of Germany, that lies in the temperate climatic zone at about 200 m amsl with a flat terrain (Figure 1). Here the floodplain forests consist of mixed stands of beech, oak, poplar, ash, maple, willow and alder. Several types of reference data were used for geometric correction of the LiDAR data and independent reference data for verification.

The digital terrain model (DTM) and the digital surface model (DSM) were derived using last pulse and first pulse data respectively. The normalized digital surface model (nDSM) was derived by a simple subtraction of DTM from DSM. Single LiDAR signals were measured with an accuracy of ± 0.15 m standard deviation, which is the standard specification of the LiDAR flight campaign in the state of Baden-Württemberg in Germany[7]. DTM, DSM and nDSM were checked for their consistency and errors using the TreesVis software, which enabled an excellent capability of 3D visualization of raw data, DTM, DSM and other derived products[8]. In addition, the nDSM was qualified by comparing with the tree height measurements (Figures 2 and 3). Tree height measurements were available from a sample-based forest inventory ('Betriebsinventur') selected in systematic grid of 100 m × 100 m per sample plot, wherein the height of 1 to 2 trees within a circle of 12 m radius was measured. Trees with relatively large diameters were also selected for measurement to accommodate optimum ranges.

Forest height map

The forest height map was derived by utilizing nDSM using raster/grid-based approach. Size of 5 m was selected for forest height characterization based on the observation from field experience and visual nDSM analysis. Technically, the mean and maximum height was calculated based on nDSM per grid cell of 5 m × 5 m to generate the respective maps. The rationale for including maximum value was based on the assumption that the maxima might represent the local forest height better than the mean during the winter season for the deciduous stands. The forest height map was compared with height measurements from the sample-based forest inventory ('Betriebsinventur') as described earlier (Table 1).

Top stand height

Since the generation of this product requires stand boundaries, it can be designed for forests where stand boundaries exist and are digitally available. The top height of a forest stand was traditionally defined as the mean height of 100 trees with the largest diameter per hectare. Height is highly correlated with the diameter, and this is usually equivalent to the mean height of the 100 largest diameter trees per hectare. This top height is

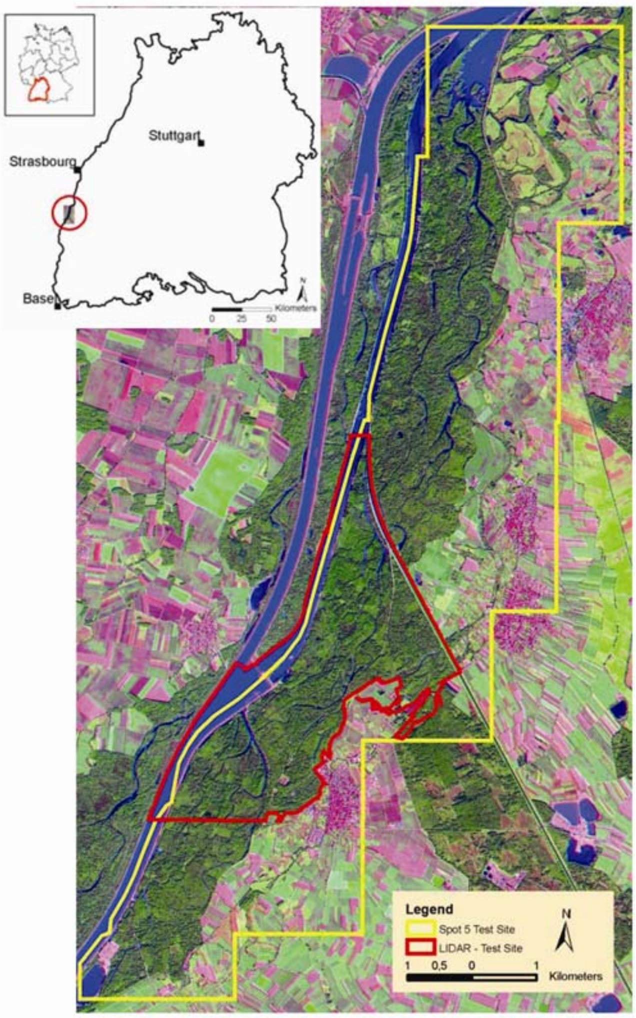

Figure 1. Extent of study area (shown in red outline) in the Rhine Valley, southwest Germany.

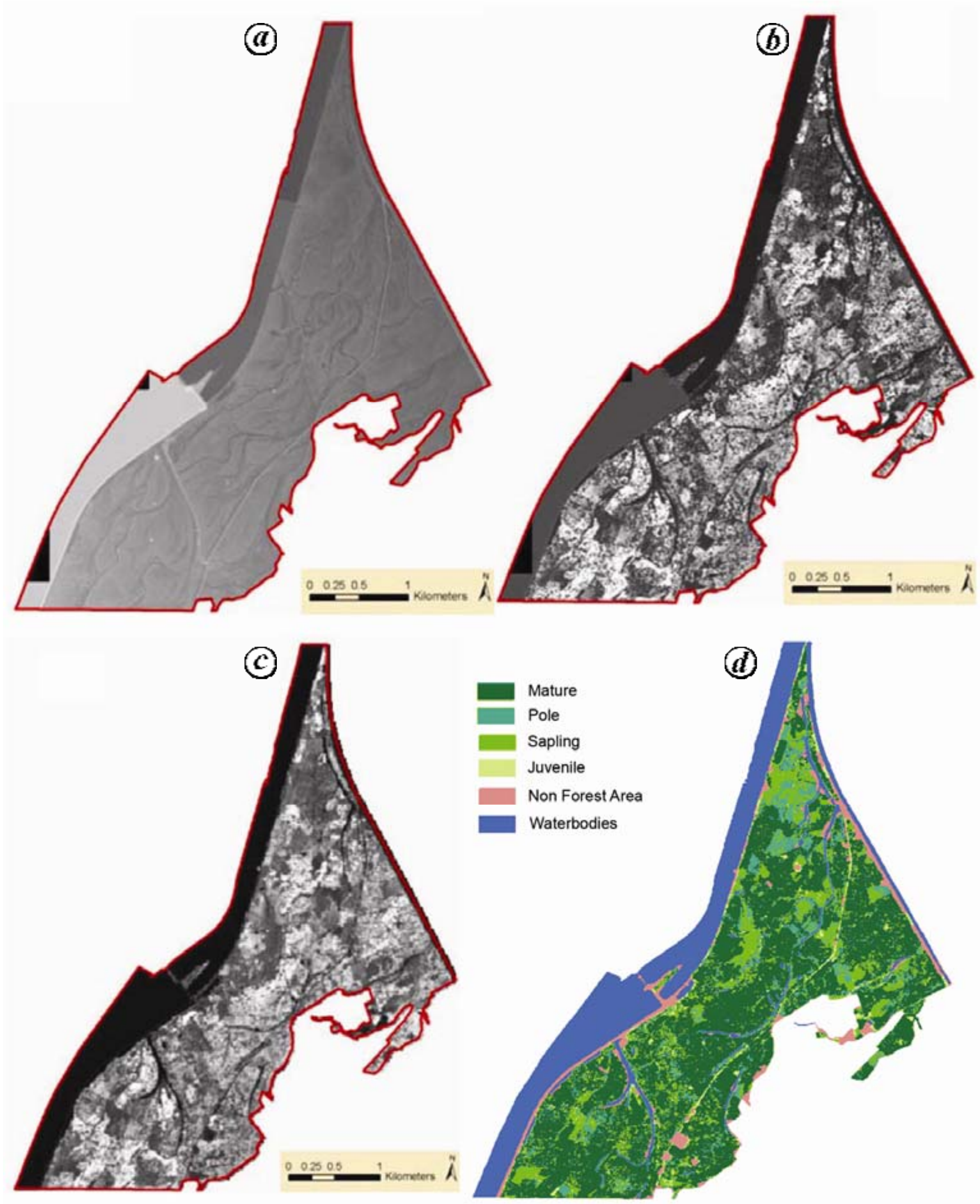

Figure 2. (***a***) Digital terrain model, (***b***) Digital surface model (DSM) and (***c***) normalized digital surface model (nDSM) derived from LiDAR data. (***d***) Forest tree height map.

relevant for generating stand-based yield tables that allow the estimation of various quantitative stand properties. Since it is hardly possible to detect single trees with the LiDAR dataset, the question arises as to how one can estimate top height based on nDSM, when the top height is defined only by the height of single trees. The technical approach used to estimate the top 100 heights was to assess the maximum heights, per 10 m × 10 m grid cells within a stand-polygon. By calculation of an area-weighted average per grid cell, an estimate of the 100 highest points per hectare of nDSM was achieved. The assumption is that this mean height of the 100 highest points is equivalent to the height of the 100 highest trees per hectare (Table 1; Figure 3).

Detection of gaps

Gaps in forests are important ecological elements and the existence of a relevant number of gaps in forest stands supports ecological and biological diversity of forest ecosystems. The automatic detection of gaps used a technical approach of several steps based on the nDSM image, i.e. identification of areas with [nDSM] < 3 m, grouping of pixel to potential gap polygons; application of minimum and maximum area (100–900 m^2), filling of single pixels or a group of pixels within potential gap polygons and assessment of the height (mean of nDSM) of a circular buffer area with a radius of 10 m larger compared to that of the smallest enclosing circle. If the height of the buffer

Table 1. Evaluation of options for estimating the 'top height 100' of forest stands by winter LIDAR data in deciduous stands

Option	Mean difference of the estimates versus plot measurements of top height 100 (m)	r	R^2 (linear regression model)	Slope (linear regression model)	Intercept (linear regression model; m)
Estimate based on the mean of five maxima, equivalent to a top height assessment of 110.5 maxima per hectare	–1.81	0.861	0.741	0.939	3.221
Estimate based on the mean of four maxima, equivalent to a top height assessment of 88.4 maxima per hectare	–1.58	0.863	0.745	0.941	2.964
Estimate based on the mean of three maxima, equivalent to a top height assessment of 66.3 maxima per hectare	–1.32	0.863	0.744	0.940	2.759
Estimate based on the mean of two maxima, equivalent to a top height assessment of 44.2 maxima per hectare	–1.03	0.859	0.738	0.931	2.701
Estimate based on the mean of one maximum, equivalent to a top height assessment of 44.2 maxima per hectare	–0.67	0.855	0.731	0.923	2.551
Estimate based five maximum heights of cells of the forest height map	–2.44	0.852	0.726	0.964	3.267

is ≥10 m, the potential gap polygon is classified as a gap. The result was verified by a sample-based field campaign that was conducted in a set of 13 randomly selected stands within which 34 sample plots were distributed in totality.

Density of middle and understorey

Characterizing the density of the middle and understorey provides important baseline information for management planning[5]. Two relative classes (highly developed middle and under layers, less developed middle and under layers) were assigned using the following technical approaches: (i) classification of LiDAR raw data between DTM (ground/forest floor) and DSM (upper crown level of top layer) as reflections from the middle and under layers if two conditions are fulfilled – (a) height at least 1 m above DTM and (b) height at least 5 m below DSM; (ii) assessment of the proportion of reflections from middle and understorey to total number of reflections per circular area with a radius of 10 m; (iii) classification of areas with $X \geq 15\%$ of reflections from middle and under layers as stands with intensive middle and under layers and $X < 15\%$ of reflections from middle and under layers as stands with extensive middle and under layers; (iv) filtering and extraction of polygons based on the proportion of reflection from the middle and under layers differing class definitions. The result was verified by a sample-based field campaign as described above (Table 2; Figure 3).

Detection of coniferous trees in deciduous stands

The approach was based on an nDSM image that was derived using a DSM based on last pulse data instead of first pulse data. In deciduous stands, in winter data acquisitions, there are nearly no reflections from the crown regions in the last pulse dataset since there are no leaves in the crown, while in coniferous stands a considerable number of reflections from coniferous trees (with needles) from the crown regions exist in the last pulse data. Since very few coniferous trees occur in the test site, the method was simply used for visualization of the coniferous areas (Figure 4). The result was cross-checked with existing forest GIS data.

Results

The visual analysis of DTM, DSM and nDSM generated using TreesVIS software indicated high quality of DTM than the forest surface in DSM and consequently in nDSM. Some problems occurred with proper representation of the crown shape in the case of deciduous trees, since in winter the highest first pulse reflections were not from the upper crown (as is the case in summer data when the trees have leaves in Germany), but from the branches and trunks of the lower storey. The forest surface is rather rough, where single trees with properly modelled crowns can hardly be found as illustrated in Figure 2 *c*, where a 3D visualization of nDSM together with sample plot heights (based on average tree height of the 1–2 trees with measured tree heights) can be compared (Figure 3 *a*). The nDSM was evaluated qualitatively by a comparison with measured tree heights from ground measurements of the sample-based inventory (Table 1). The mean height per sample plot location was compared from two measurements: (i) mean height of nDSM at sample plot locations: 12.3 m (Figure 3 *b*) and (ii) top height 100 (from ground measurements): 25.4 m, with a correlation of 47.3% (Figure 3 *c*). This high difference supports the assumption that an underestimation of the height of the vegetation surface exists in winter data. The quantitative evaluation of the forest height map was done based on the 'top height 100', assessed by the sample-based inventory. Since the height measurements were

Table 2. Verification result for middle and under layers classification

	Number of samples	Average coverage of middle and under layer (%)	Mean coverage of middle layer (%)	Mean coverage of under layer (%)
High density of middle and under layers	18.00	38.75	34.44	43.06
Low density of middle and under layers	13.00	24.23	30.00	24.62

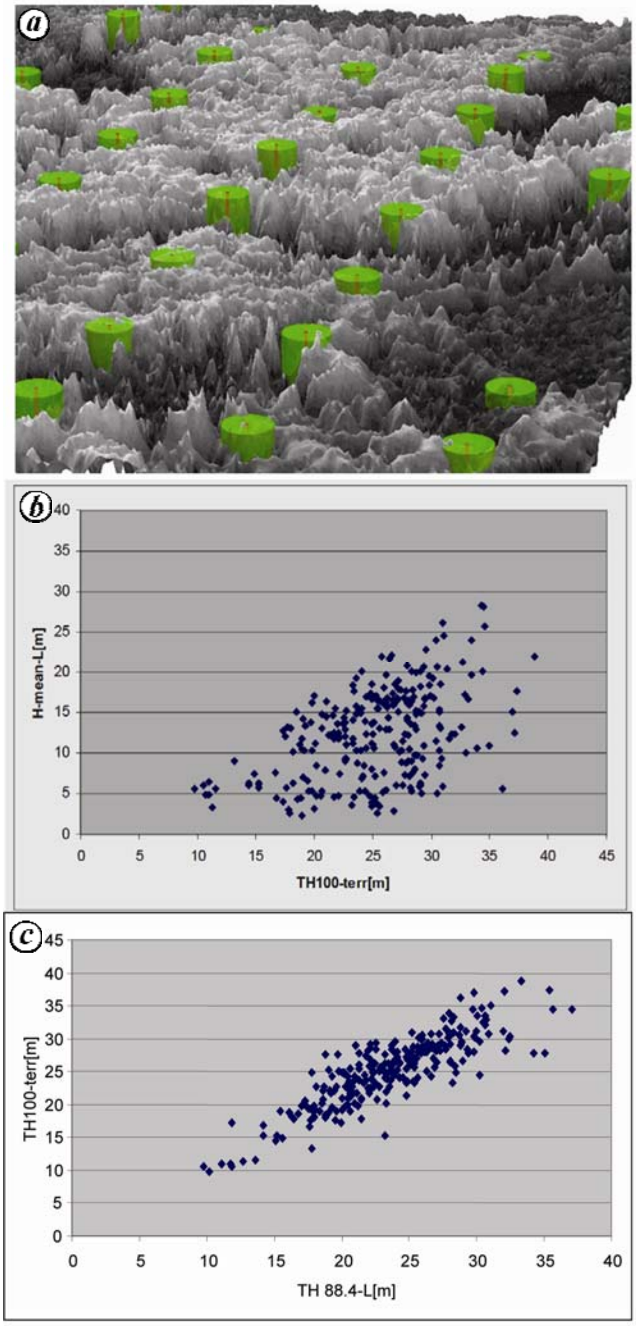

Figure 3. ***a***, Three-dimensional visualization of normalized digital surface model and the sample plots. The cylinders (green colour) at the sample plot locations show the mean tree height. ***b***, Correlation between mean height (derived from LiDAR) of nDSM and 100 top heights (from ground measurements). ***c***, Estimates of 100 top heights based on the mean of 4 maximum heights at sample plot locations, equivalent to an assessment of 88.4 maxima per ha versus the 100 top heights from ground measurements.

made 4 years ahead of the LiDAR data assessment, average growth values were applied based on the height growth of beech (Figure 3 *c*). On an average, a height increment of 0.7 m was assumed. Out of 268 sample plots, 7 were excluded from the analysis due to implausible height differences that eventually occurred due to a possible mis-registration of the two datasets. Per circular sample plot with a radius of 12 m the mean of the heights per grid cell was calculated. The results for the two options were obtained using the mean and maximum height per 5 m × 5 m grid cell, based on 261 sample plots, viz. mean height of 'top height 100' (ground measurements): 25.4 m; mean height of mean-based grid cells: 12.3 m; correlation with 'top height 100': 47.3%; mean height of maximum-based grid cells: 17.5 m; correlation with 'top height 100': 61.9% (Table 2). The evaluation indicated that the maximum height is a more appropriate realistic estimate of the forest height. If we consider that 400 cells of 5 × 5 m^2 cover 1 ha, the forest height calculated here can be regarded equivalent to the mean height of the 400 local maximum height elevations per hectare. Although the 400 highest trees are not distributed in a way that one of them occurs in each grid cell, still the measure will be close to the height of the 400 highest trees per ha.

A comparison of the 'top height 100' on the basis of 1, 2, 3, 4 or 5 highest elevations per sample plot based on the mean of the five maximum heights of cells of the forest height map resulted in high correlation (86%). It can be concluded from the fact that true correlation was reduced by mis-registration of the two datasets to one another, that a LiDAR-based top height 100 estimate, based on an estimate of 10 m × 10 m grids per stand, is a reliable one. For such measurement, the average underestimation of the height was about 1.7 m with the correlation of 86% (interpolating from the empirical results of 110.5 and 88.4 maxima per ha). Gaps were mapped automatically for the entire area. The verification based on 34 ground sampling points showed that 9 out of 11 mapped gaps were true gaps; and that on 23 plots, where no gap was mapped, no gap was identified. The overall accuracy thus reached 94.4%. The two density classes for the middle and under layers were produced automatically for the entire area. Verification was based on field assessment. Out of the 34 field samples, 3 were excluded, 1 since harvesting took place and 2 since positioned exactly on the class border, thus 31 out of 34 were used. Very few groups of coniferous stands were detected (Figure 4).

Figure 4. Normalized digital surface model (last pulse, 1 m). Bright areas indicate coniferous trees and yellow lines indicate borderline of coniferous spruce stands.

Formal verification was not possible due to restricted number of coniferous stands in the area. Plausibility check of coniferous stands in Forest GIS of forest management data showed high reliability[6].

Discussion and conclusions

The forest height map derived utilizing the nDSM would fill the information gap for the entire state of Baden-Württemberg, where no age class information is currently available, as expected in private forests. The stand top height product could be utilized for all the Baden-Württemberg forest stands for quantitative assessment of stands/forest height and volume. Both height products are intended to provide information on forest development phases. The height maps can be transferred to development stage maps and can be used to assess as biodiversity indicators. Many objects such as forest roads and other elements of technical infrastructure and objects of fluvial systems such as rivers, streams, water-filled ditches, dams, stream beds and ditches currently not filled with water can be mapped utilizing the LiDAR-derived DTM by visual interpretation. This kind of information can be of importance for management planning, detection of disturbances to biodiversity conservation and for planning of re-naturalization of river systems[6]. Technically, a visualization of DSM is possible in various ways, two-dimensional using black and white or pseudo colours to represent height or a shaded image or three-dimensional using the 3D visualization capability of the TreesVis software[8]. The assessment of coniferous trees within deciduous stands is of high relevance, where coniferous trees or stands are not part of the natural tree species composition and thus are artificial elements.

The LiDAR-based information products have the potential to provide highly important structural information on entire forest areas of a federal state or country at a moderate cost. This applies for areas where LiDAR data are available at moderate cost from survey administration, as is the case of Germany. It is still unclear in most federal states in Germany as to what cycle the LiDAR assessments will be repeated and such repetitions are essential for continuous monitoring. The large-scale availability and affordability of LiDAR data on regional and national-level scale will depend on repetition cycles for LiDAR mapping by state survey administration that has not yet been established. An alternative could be the development of appropriate sampling schemes. The large-scale feasibility of the assessment of major indicators for biodiversity and climate change, namely forest area, distribution of main tree species, the structural parameters, height, structural elements such as gaps and standards, layer structure, and detailed mapping of coniferous trees in mixed stands has been shown. These capabilities provide the opportunity to use this technology in the context of large-scale monitoring of changes, including those due to climate change and the characterization of biodiversity through vertical stratification that is needed for now and in future[9,10].

1. Bock, M. and Dees, M., Feasibility of remote sensing for assessment of Annex 1 habitats for the monitoring of Natura 2000 SAC sites – A feasibility study based on assessment guidelines of the German Federal Agency for Nature Conservation. Geoland project report, European Commission, FP6-2002-SPACE-1 Proposal No. FP-6-502871, 2007, p. 100.
2. Koch, B. *et al.*, Forestry applications. In *Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences* (eds Li, Z., Chen, J. and Baltsavias, E.), ISPRS Congress Book. Taylor & Francis Group, London, 2008, pp. 439–465.
3. Koch, B., Walentowski, H., Dees, M. and Seitz, R., Untersuchung der Nutzungsmöglichkeiten von amtlichen Laserscannerdaten für den Wald-Forst-Bereich – Entwicklung von Methoden für forstliche Inventur- und Monitoringaufgaben im Rahmen der FFH Kartierung. Waldökologie, Landschaftsforschung und Naturschutz – Forest Ecology, Landscape Research and Nature Conservation, 2010, pp. 23–33.
4. Næsset, E. and Bjerknes, K., Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. *Remote Sensing Environ*., 2001, **78**, 328–340.
5. Weinacker, H., Koch, B., Heyder, U. and Weinacker, R., Development of filtering, segmentation and modelling modules for LIDAR and multispectral data as a fundament of an automatic forest inventory system. In *Laser-Scanners for Forest and Landscape Assessment*, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Freiburg, 2004, vol. XXXVI, pp. 50–55.
6. Dees, M., Straub, C. Langar, P. and Koch, B., Remote sensing based concepts utilising SPOT 5 and LIDAR for forest habitat mapping and monitoring under the EU Habitat Directive. In (eds Lamb, A. *et al.*), ONP 10 Test & Benchmarks Report, Geoland project report, European Commission, FP6-2002-SPACE-1 Proposal No. FP-6-502871, 2007, pp. 53–78.
7. Schleyer, A., Das Laserscan-DGM von Baden-Württemberg. In *Photogrammetric Week* (eds Fritsch, D. and Spiller, R.), Wichman Verlag, Heidelberg, 2001, pp. 217–225.
8. Weinacker, H., Koch, B. and Weinacker, R., TREESVIS: A software system for simultaneous ED-real-time visualisation of DTM, DSM, laser raw data, multispectral data, simple tree and building models. In Proceedings of the ISPRS Working Group VIII/2, Freiburg, 3–6 October 2004, pp. 90–95, ISSN 1682-1750.
9. Parcs, C. and Bernier, P., Adaptation of forests and forest management to changing climate with emphasis on forest health: A review of science, policies and practices. *For. Ecol. Manage.*, 2010, **259**, 657–659.
10. Thompson, I., Mackey, B., McNulty, S. and Mosseler, A., Forest resilience, biodiversity, and climate change. A synthesis of the biodiversity/resilience/stability relationship in forest ecosystems. Secretariat of the Convention on Biological Diversity, Montreal, Technical Series 43, 2009, p. 67.

ACKNOWLEDGEMENTS. The research presented here was funded by the European Union (Project Geoland, FP6-2002-SPACE-1, Proposal No. FP-6-502871). We thank Dr M. D. Behera, IIT Kharagpur for providing useful information during the preparation of this manuscript.