UNIVERSITÄT BAYREUTH

Research Group of Climatology

Master Thesis in Geoecology

LiDAR BASED CLASSIFICATION OF FOREST STRUCTURE OF CONIFER DOMINATED ECOSYSTEMS IN THE BLACK FOREST NATIONAL PARK

LiDAR-BASIERTE KLASSIFIKATION DER WALDSTRUKTUR KONIFEREN-DOMINIERTER ÖKOSYSTEME IM NATIONALPARK SCHWARZWALD

by

Florian Lang

(1245974)

in partial fulfilment of the requirements for the degree of

Master of Science (M.Sc.)

Bayreuth,

August 30th, 2017

submitted to: Faculty of Biology, Chemistry and Earth Science

1. Supervisor: Prof. Dr. Cyrus Samimi
   Research Group of Climatology

2. Supervisor: Prof. Dr. Michael Hauhs
   Chair of Ecological Modelling

Statutory declaration

I hereby declare on oath that the present thesis entitled

LiDAR based classification of forest structure of conifer dominated ecosystems in the Black Forest National Park

was independently authored by myself. I did not use any unnamed, illegitimate assistance from third parties, media or materials than the ones referred to in this thesis. I further declare, that all passages quoted either verbatim or in substance from published or unpublished documents were marked as citation in each individual case.

This thesis has not been submitted, wholly or substantially, to any other examination institution.

Bayreuth, August 30th, 2017

Florian Lang

Eidesstattliche Erklärung

Hiermit erkläre ich eidesstattlich, dass ich die vorliegende Arbeit mit dem Titel

LiDAR-basierte Klassifikation der Waldstruktur Koniferen-dominierter Ökosysteme im Nationalpark Schwarzwald

selbständig verfasst habe. Ich habe keine ungenannte und unerlaubte Hilfe Dritter angenommen, noch habe ich Medien oder Hilfsmittel verwendet die nicht angegeben sind. Des Weiteren erkläre ich, dass alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, in jedem Einzelfall durch Angabe der Quelle als Entlehnung kenntlich gemacht sind.

Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Bayreuth, 30. August 2017

Florian Lang

Acknowledgement

Although this master thesis is published under my single authorship, a lot of preparation, inspiration, assistance, and advices accompanied the various processes leading from the first idea to the completion. Acknowledgment of each contribution is a matter of particular concern to me. Therefore, I would like to open with my heartfelt thanks to everybody who has supported me during this final stage of my geoecology study.

I would like to explicitly express my deep sense of gratitude to my principal supervisor and mentor, Prof. Dr. Cyrus Samimi. Without his commitment to initialize the contact to the Black Forest National Park, his unrestricted support during the development of the topic, his constant guidance and encouragement, this thesis would not have been possible. I'm truly grateful for offering the possibility of working on a very exciting topic in an extraordinary working environment and the trust he placed in me to represent the Research Group of Climatology (University of Bayreuth) at the Black Forest National Park.

My special thanks go to the Black Forest National Park for the provision of the comprehensive and detailed database as well as for the possibility to autonomously execute my field work in this beautiful national park. I thank profusely Dr. Marc Förschler, head of the Department 2, Ecological Monitoring, Research and Species Protection for fully integrating me in his team. For their continually and preciously scientific, operational, and interpersonal support I express my deep gratitude to Dr. Stefanie Gärtner, Sönke Birk, Dr. Christoph Dreiser and all team members of the department.

Many volunteers assisted me during the field campaign regardless of weather or terrain conditions and with resilient positive attitude. I would particularly like to mention Patricia Mannßhardt and Carl Seitz who accompanied me through most of the time. My sincere thanks for your intensive engagement go to all of you.

I would like to thank Katja Reinhardt of the Research Group of Climatology for her help in computationally processing a crucial step within the data preparation. In addition, I thank all members of the research group for scientific discourse about methodologic questions and their advices throughout the preparation of my thesis.

My thanks also go to Ulrike Höpfl for her patience in assisting by the illustration of the forest structure silhouettes.

Just mentioned at the end due to chronological reasons, I am very thankful to Janice Chaddock for her thorough editing of my master thesis and her valuable suggestions to give it the final touch.

Abstract

Forests are more than an aggregation of trees, demonstrated by constituting habitat for more than 75% of the world's terrestrial biodiversity. Structural richness increases habitat suitability for a major part of dwelling species. In the Black Forest National Park diverse mosaic forest structures are supposed to supersede large-scale structures under conservation of natural dynamics with considerable effects on biodiversity. A comprehensive ecosystem monitoring is planned to document these processes and to serve as basis for scientific research, management and stakeholder information.

This study aims to adapt a stratification of forest structure to local, conifer dominated forests and classify the forest strata across the national park by the means of remote sensing. The resulting forest structure map should build the basis for a long-term forest structure monitoring. Six forest strata, differentiated by tree dimensions and their heterogeneity, and a complementing opening stratum were defined. A high-density, discrete return LiDAR data set (30 returns/m^2) was used for an area-based forest structure analysis. A set of LiDAR metrics describing forest structure was calculated for each cell of a gridded national park raster of a 20 m x 20 m resolution. For model training and test purpose 86 forest observations across the six forested strata were selected. Additional 15 opening plots were located on meadows by aerial image analysis. Stratum classification by Random Forests modelling was chosen due to its positive results in studies under comparable conditions.

Random Forests integrated recursive feature elimination reduced the metrics used in the final model to 13, mainly density and height related predictors, which were partially highly correlated. The classification reached an overall OOB accuracy of 81.47% ($K_{HAT} = 0.78$) and a test accuracy of 90.00% ($K_{HAT} = 0.88$). The intra-stratum accuracies varied. Low vegetated strata were almost perfectly classified whereas multi-story strata showed moderate accuracies. Stratum membership probability analysis confirmed the differences in classification performance but indicated on average good stratum identification.

The study proved that the introduced stratification defined well differentiable forest structure strata. The classification revealed basically well strata predictions with respect to accuracy measures and to spatial distribution. Reduced accuracies relating to multi-story strata were supposed to result from underrepresented field observations. Future acquisition of reference plots is necessary to extend the presented classification approach to mixed or deciduous dominated forests. In conclusion, the study shows reliable forest strata classification, it provides a comprehensive approach for repeated monitoring assessments and offers various aspects for further research activities.

Zusammenfassung

Wälder sind mehr als eine Ansammlung von Bäumen. Dies zeigt sich daran, dass Wälder den Lebensraum für mehr als 75% der terrestrischen Biodiversität der Erde darstellen. Dabei erhöht strukturelle Vielfalt die Lebensraumqualität für viel waldbewohnende Arten. Im Nationalpark Schwarzwald wird erwartet, dass mosaikartige Waldstrukturen unter dem Schutz der natürlichen Walddynamik vorhandene großflächige Strukturen ablösen, mit Folgen für die Biodiversität. Ein umfassendes Ökosystemmonitoring soll diese Prozesse dokumentieren und als Basis für Forschung und Management sowie zur Information von Interessengruppen dienen.

Ziel dieser Studie ist eine Stratifikation der lokalen, koniferen-dominierten Waldstruktur und die Klassifikation der Nationalparkfläche gemäß dieser Straten auf Basis von Fernerkundungs-Informationen. Eine zusammenfassende Karte soll als Ausgangspunkt für ein langzeitliches Monitoring der Waldstruktur dienen. Sechs Waldstraten wurden anhand der Dimension und der Homogenität der Bäume unterschieden, ergänzt durch ein Stratum für Offenflächen. Die flächenbasierte Strukturanalyse wurde auf Basis eines hoch aufgelösten, diskreten LiDAR-Datensatzes (30 Punkte/m^2) durchgeführt. Aus den LiDAR-Daten wurden Kennzahlen zur Beschreibung der Waldstruktur für jede Zelle der mit einem 20 m x 20 m Raster überzogenen Nationalparkfläche berechnet. Es wurden 86 Beobachtungs-flächen der Waldstraten ausgewählt, um die Klassifikation zu trainieren und zu testen. 15 zusätzliche Offenflächen wurden per Luftbildanalyse auf Wildwiesen verortet. Die Klassifikation erfolgte anhand eines Random Forests-Verfahrens, das in vergleichbaren Studien gute Ergebnisse erzielte.

Eine in Random Forests integrierte rekursive Auswahlroutine reduzierte die Kennzahlen für das endgültige Modell auf 13, zumeist dichte- oder höhenbezogene Prädiktoren, die teilweise stark korreliert waren. Die Klassifikation erzielte eine OOB-Gesamtgenauigkeit (engl: out-of-bag) von 81.47% ($K_{HAT} = 0.78$) und auf den Testdaten 90.00% ($K_{HAT} = 0.88$). Die Genauigkeiten der einzelnen Straten fielen unterschiedlich aus. Niedrig bewachsene Straten wurden nahezu fehlerfrei klassifiziert, während mehrschichtige moderate Genauigkeit aufwiesen. Die Auswertung der Stratenzugehörigkeit nach Wahrscheinlichkeit bestätigte diese Unterschiede, zeigte aber im Schnitt gute Identifikation der Straten.

Die Studie konnte belegen, dass die vorgestellte Stratifikation gut differenzierbare Waldstrukturstraten einteilt. Die Klassifikation zeigte im Grundsatz gute Vorhersagen sowohl hinsichtlich der Genauigkeit als auch der räumlichen Verteilung der Straten. Verringerte Genauigkeiten im Zusammenhang mit den mehrschichtigen Straten sind vermutlich auf die zu geringe Anzahl an Beobachtungsflächen zurück zu führen. Um die Klassifikation zukünftig auf Misch- und Laubwälder ausdehnen zu können, müssen zusätzlich spezifische Beobachtungsflächen aufgenommen werden. Zusammenfassend liefert die Studie eine verlässliche Klassifikation der Waldstrukturen, zeigt eine umfassende Vorgehensweise für wiederholtes Monitoring auf und bietet vielfältige Ansätze für weitere Forschung

Table of contents

List of figures

List of tables

List of abbreviations

3D	three-dimensional
ALS	airborne laser scanning
a.s.l.	above sea level
AUC	area under the receiver operating characteristics curve
BFNP	Black Forest National Park
BMU	Federal Ministry for the Environment, Nature Conservation and Nuclear Safety
CDP	conditional density plot
CRS	coordinate reference system
DBH	diameter at breast height
DHDN	Deutsches Hauptdreiecksnetz
DTM	digital terrain model
E	East
EPSG	European Petroleum Survey Group Geodesy
FAO	Food and Agriculture Organization of the United Nations
FeLis	Chair of Remote Sensing and Landscape Information Systems
GFZ	German Research Centre for Geosciences
GPS	Global Positioning System
IUCN	International Union for Conservation of Nature
Ka	kilo-annum
LGL-BW	Landesamt für Geoinformation und Landentwicklung in Baden-Württemberg
LiDAR	Light Detection and Ranging
Ma	mega-annum
MV	missing value
N	North
OOB	out-of-bag
PA	producer’s accuracy
RF	Random Forests
RFE	recursive feature elimination
ROC	receiver operating characteristics
S	South
UA	user’s accuracy
WRB	World Reference Base

1 Introduction

"To protect natural biodiversity along with its underlying ecological structure and supporting environmental processes", is the primary objective of category II protected areas according to the International Union for Conservation of Nature (IUCN; Dudley, 2013, p. 16). In this sense, it is, inter alia, part of the guiding principle of national parks in general and of the Black Forest National Park (BFNP) in particular (Nationalpark Schwarzwald, 2014). The basic need to conserve biodiversity in terms of diversity within species, between species and of ecosystems is almost worldwide accepted, at least since the signing of the Convention on Biological Diversity in 1992 (Hooper et al., 2005; Secretariat of the Convention on Biological Diversity, 2005). Comprehensive analyses of the interaction of biodiversity and human well-being showed multi-dimensional relationships (Mace et al., 2012). In conclusion, it is evident that loss of biodiversity will negatively affect global material and immaterial wealth from a long-term perspective (Millennium Ecosystem Assessment, 2005). The predictions of the Millennium Ecosystem Assessment (2005) show a progressive loss of species and ecosystems in all scenarios and across terrestrial, freshwater and marine biomes till 2050. Within the terrestrial biomes forested areas are of particular importance: 31% of the global land area is covered by forests (Keenan et al., 2015), home of more than 75% of the world's terrestrial biodiversity (FAO, 2016). Recognising this key role of forests, international and national policies intensified their efforts to preserve forest ecosystems. This led to an increase of protected forest areas from 7.7% to 16.3% at global scale between 1990 and 2015 (Morales-Hidalgo et al., 2015). Designations of protected areas constitute a core element in conservation efforts with direct effect on biodiversity maintenance as well as indirect impact by rising awareness of their benefits (Rands et al., 2010). Abandonment of forest management and protection of natural dynamics is expected to foster biodiversity as it is indicated for species richness by Paillet et al. (2010). In this context, the German government strives to protect natural forest development in 5% of Germany's wooded area till 2020 to improve diversity in structure and enable forest development dynamics, as laid down in the "National Strategy on Biological Diversity" (BMU, 2007) .

The implementation of protected areas, aiming the conservation of biodiversity, ecological structures, and environmental processes creates challenges for planning, management and assessment of conservation actions. Therefore, ecological monitoring and evaluation systems gain in importance for scientific research as well as for political and societal responsibilities (Stem et al., 2005; Lindenmayer and Likens, 2010). Comprehensive monitoring frameworks involve compositional aspects of biodiversity at organizational levels from landscapes to ecosystems, from populations down to genes. Structural and functional components accompany a broad monitoring system at each of the four different levels (Noss, 1990). In forests, structural components are often given high priority in monitoring systems as they represent species habitat configurations. Important structural indicators are for example tree age diversity, vertical stratification, horizontal composition, and occurrence of deadwood (Berglund et al.,

2005; Bergen et al., 2009; Paillet et al., 2010). The dynamics of forested ecosystems can be successfully monitored when the process is set on a large temporal scale and frequencies that are adapted to change processes (Spies and Turner, 1999). This particularly pertains to protected forest areas with natural development dynamics where disturbances on different spatial and temporal scales and of various severities create or change mosaic forest patterns of diverse structural patches (Pickett and White, 1985; Remmert, 1992; Oliver and Larson, 1996; Spies and Turner, 1999; Turner, 2010).

Forest monitoring in forestry and forest science had been based on manual, on-site measurements without alternatives till aerial photography appeared in the 1930's (Coppin and Bauer, 1996). Still today, manual inventory is important but it requires to comply the precepts of efficiency. While respectively small areas could be fully inventoried on the ground, the methodology usually changes to sample-based procedures with increasing area. These areal estimations originating from systematic samplings are still complex, time-consuming and error-prone though (McRoberts and Tomppo, 2007). Therefore, remote sensing methods have been tested for and proven their capability to facilitate the acquisition of areal forest information since the development of aerial photography (Coppin and Bauer, 1996; Wulder, 1998; McRoberts and Tomppo, 2007; Latifi, 2012). A major progress has been achieved by the introduction of Light Detection and Ranging (LiDAR) in ecological remote sensing. LiDAR, which is called Airborne Laser Scanning (ALS) if operated on an airplane, provides a highly accurate three-dimensional (3D) measurement of structures and surfaces (Wulder et al., 2014). ALS is an active remote sensing technique based on high frequent emission of laser pulses from an airplane to the ground and recording of reflections by an on-board sensor (Figure 1-1 and Figure 1-2).

The airplane position is precisely controlled by a Global Positioning System (GPS) network of satellites and ground stations which is the essential prerequisite for accurate surface scanning (Figure 1-1). The exact position of the reflecting object is finally derived from the reflected beam angle, the elapsed time between emission and detection, and the speed of the laser light (Vauhkonen et al., 2014). Depending on the measurement unit the ALS system records either discrete reflections per laser beam or a continuous waveform of the entire backscattered signal (Figure 1-2; Hollhaus et al., 2014). Discrete LiDAR systems record reflections with 3D coordinates for each laser pulse. Independent of the differentiation between large-footprint (5 – 10 m) and small-footprint (0.1 – 2.0 m) ALS approaches, one laser beam typically can lead to multiple echoes by partial reflection in rich vertical forest structures (Vauhkonen et al., 2014). The footprint size depends on the divergence of the laser beam and flight altitude. Continuous waveform recordings, also called full-waveform ALS, usually are post-processed to discrete returns as well. The continuity of the signal enables the derivation of a higher average number of returns per beam, and additional information about pulse width and intensity per return (Figure 1-2;

Hollhaus et al., 2014). Due to the 3D arrangement of the single returns the result of a LiDAR measurement campaign is a 3D return cloud (syn. point cloud) which comprises all reflections.

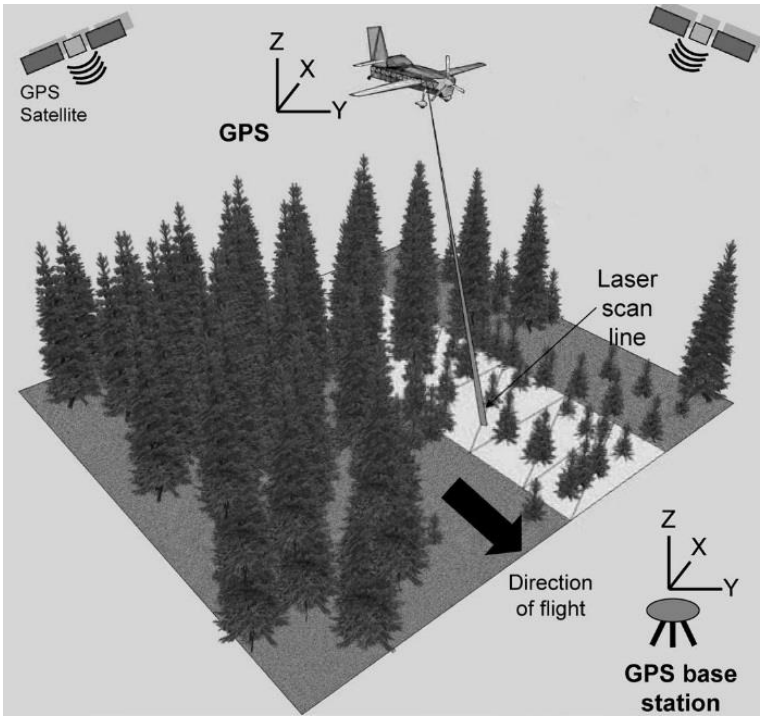

Figure 1-1 GPS controlled ALS (LiDAR) of forest structure; the bright segment symbolises the scanned swath width of one flight strip (White, 2013, p. 3).

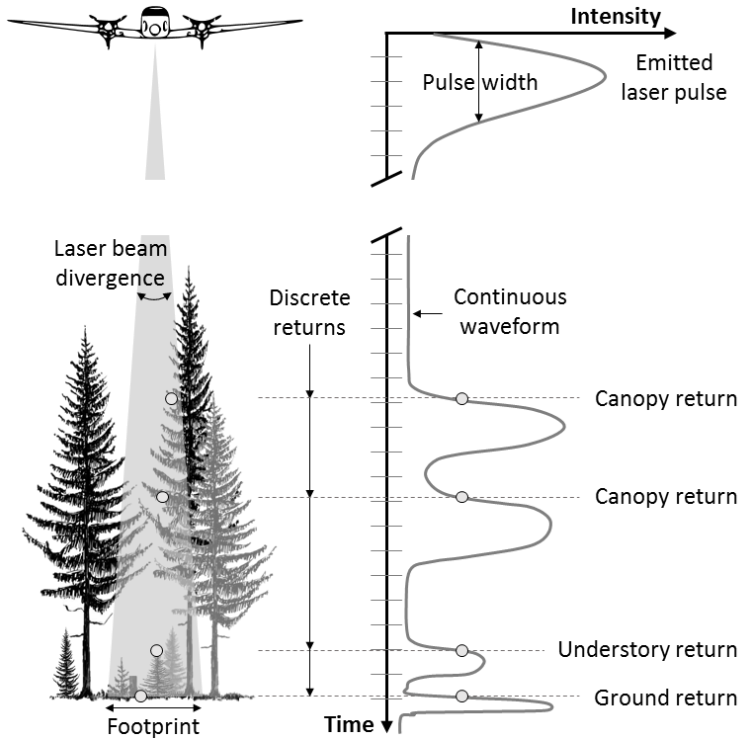

Figure 1-2 Illustrated measurement principle of ALS (LiDAR); backscattered laser pulse energy is recorded by a sensor unit either as discrete returns or as continuous (full) waveform.

The crucial advantage of LiDAR over passive optical systems, such as aerial photography and optical satellite imagery, is its ability to characterize 3D vegetation structure even if it is dense (Lefsky et al., 2001; Næsset et al., 2004; Vauhkonen et al., 2014). Recent comparisons of ALS and return clouds of stereo images from photogrammetric matching confirmed the leadership of LiDAR at least in forested ecosystems (Gobakken et al., 2014; Pitt et al., 2014). This has been the reason for the fast development and the ongoing large popularity of LiDAR measurements in operational forestry (Næsset et al., 2004; Hyyppä et al., 2012), forest science (Shugart et al., 2010), and species ecology as well as habitat modelling (Graf et al., 2009; Müller et al., 2010; Zellweger et al., 2013). In all fields of application small-footprint LiDAR systems are mostly applied, recording either discrete or full-waveform data (Vauhkonen et al., 2014; Sumnall et al., 2016). The return clouds can be analysed by two distinct approaches. At first, area-based methods are used to characterize the ground surface of a determined area and the aggregated vegetation above. Besides directly derived LiDAR return cloud metrics per area (e.g. maximum height, mean canopy height), regression and classification models are typically conducted based on adequate metrics (Vauhkonen et al., 2014). Regression models usually serve to estimate vegetation parameters (e.g. forest stem volume and biomass; Hyyppä et al., 2004; Zolkos et al., 2013) while classification approaches aim to identify pre-defined structures (e.g. habitat and forest stand types; Vierling et al., 2008; Falkowski et al., 2009). Secondly, individual tree approaches facilitate tree-level inventories (Reitberger et al., 2009). They require higher return densities of at least 10 returns/m^2 (Yao et al., 2014) to build the basis for solely remotely sensed forest inventories on single tree scale (Heurich et al., 2015).

Established in January 2014, the Black Forest National Park (BFNP) comprises mainly montane and sub-montane forest ecosystems as well as heathlands at the ridges of the mountain range. The national park area hosted more than 2,500 identified species according to local inventories and incidental identifications till October 2016. This was approx. 17% of the species richness of the federal state of Baden-Württemberg (Germany). As a result of a systematic, long-term biodiversity monitoring system that started in 2017, the recorded species number is expected to further increase (Buse et al., 2016). In addition, the biodiversity monitoring of the BFNP is going to be accompanied by a comprehensive monitoring of forest structure and dynamic change processes (Gärtner et al., 2016). This is of high importance as natural dynamic processes are expected to considerably change the vegetation structure in the BFNP. A large part of present structures of forested areas are predominantly shaped by a long sylvicultural management tradition (PricewaterhouseCoopers & ö:konzept, 2013; Rösch, 2015). For this reason, the administration of the BFNP commissioned an ALS campaign to document the vegetation structure at the initial stage of the national park history for 2015. The flight area covered the entire BFNP and a spacious buffer zone around its border. The on-board LiDAR measurement unit produced high-density, full-waveform data which was processed to discrete return data (MILAN Geoservice GmbH, 2015).

2 Objectives

Today, ecological monitoring needs to measure up to various political and societal interests. Reports must be taken with intensifying specification to a rising number of different stakeholders. The growing complexity and the simultaneous constraint of efficiency is a major challenge to administrations of protected areas (Lindenmayer and Likens, 2010). In this environment, the BFNP is currently building up its detailed and comprehensive biodiversity and ecosystem monitoring (Buse et al., 2016; Gärtner et al., 2016). This study generally aims to classify and map forest structure stages from discrete LiDAR data, as a substantial part of the BFNP's ecological monitoring system.

The study area was limited to conifer dominated areas based on forest management data from 2014. The definition of structure stages follows the idea of stratifying coniferous forest ecosystems along ecologically relevant structural features such as tree diameters, canopy layering, canopy closure and height parameters (O'Hara et al., 1996; Oliver and Larson, 1996; Falkowski et al., 2009). Six forested strata and one open stratum have been defined. This stratification represents structures adapted to the regional forest characteristics in consideration of potential changes under protection of natural dynamics. The strata descriptions and names have been deliberately defined without references to specific forestry or forest succession concepts or terminologies. They solely emphasize structural aspects.

An area-based classification approach was used due to its superior suitability in characterising forest structures on patch scales. A wide range of LiDAR metrics were calculated from rasterized return clouds as potential predictors for a Random Forests (RF) classification (Breiman, 2001). By similar means Falkowski et al. (2009) did achieve a high accuracy classification of a structurally diverse, conifer dominated forest in Idaho, USA. Representative observation plots for training and test purposes had been inventoried during a comprehensive field campaign in 2016. Visual on-site assessment of the forest structure in addition with detailed forest inventories on the plots led to their stratum assignment.

On this basis, the major focus of this study can be differentiated into three detailed objectives as follows:

(1) Establishment of a forest structure stratification scheme covering regional characteristics and integrating expected structural developments.
(2) Identification of and limitation to the most important forest structure predictors achieving a highly parsimonious classification model.
(3) Accurate classification of forest structure strata solely by ALS as basis for a future forest structure monitoring system in the BFNP.

3 Material and methods

3.1 Study area

The Black Forest National Park is located at the northern part of the Black Forest mountain range in the South-West of Germany (Figure 3-1). It was founded by the state parliament of Baden-Württemberg according to the IUCN category II criteria in January 2014 (Landtag von Baden-Württemberg, 2013). The park has an area of 10,062 ha divided into two parts of which the national park region *Hoher Ochsenkopf/Plättig* (2,447 ha) lies to the north of the region *Ruhestein* (7,615 ha). It comprises mainly higher elevations between 472 to 1151 m above sea level (Figure 3-1). The protected area covers approx. 1.7% of the entire Black Forest mountain range (PricewaterhouseCoopers & ö:konzept, 2013).

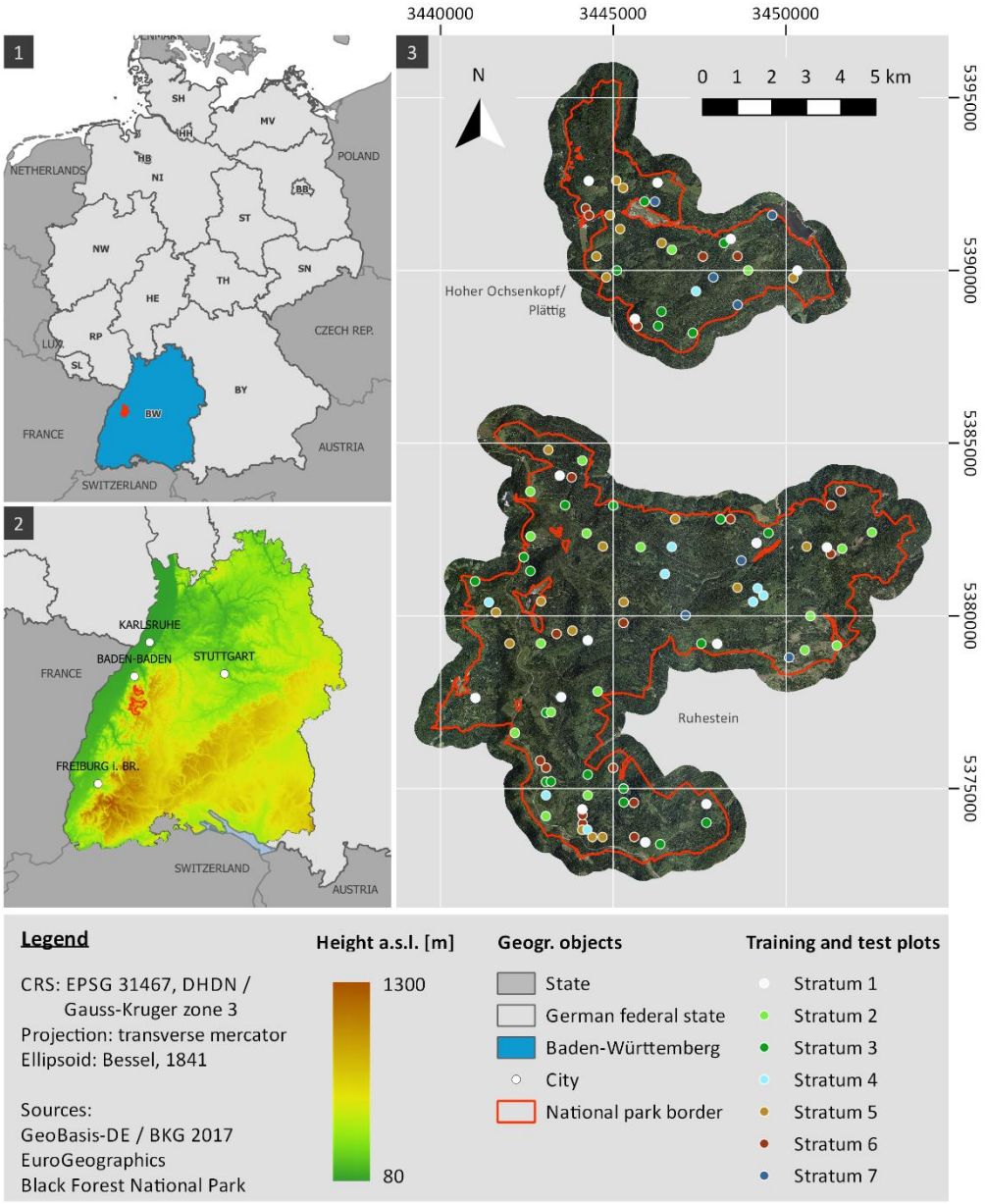

Figure 3-1 Location of the BFNP in Baden-Württemberg, Germany (1); Two-part area of the BFNP at higher altitudes in the north of the Black Forest mountain range (2); Training and test plot distribution (3)

The Black Forest mountain range is geologically part of the central crystalline Moldanubian zone of the European Variscan belt (Echtler and Chauvet, 1991-1992). In the pre-variscan era (approx. 900 Ma – 420 Ma) huge sediment series of greywacke sandstones, pelite mudstones and vulcanites deposited which were metamorphologically transformed to various gneiss types during the Assyntic orogeny ending approx. 500 Ma ago (Schöttle, 2005). Between 500 Ma and 450 Ma the gneisses were partially transformed to anatexites by re-melting (Schöttle, 2005). During the Variscan orogeny (approx. 420 Ma – 300 Ma) magmatic activity led to the formation of extensive granite stocks of different composition (Echtler and Chauvet, 1991-1992). The crystalline basement of the Black Forest consists of these gneisses and granites (Schöttle, 2005). The basement was once covered by Mesozoic sedimentary rocks. But ever since the tectonic uplift of the Black Forest and the Vosges started in the Upper Cretaceous period (approx. 100 Ma) the erosion process scoured off these sedimentary rock formations. At present, only the Lower Triassic sandstone formation of *Buntsandstein* overlies regionally the crystalline basement in the northern part of the Black Forest (Schöttle, 2005). In the centre and the south of the Black Forest, where the overlying rock formations were completely eroded, gneisses and granites form the present bedrock. The National Park is mainly located on the remaining *Buntsandstein* formations. Granites build only locally the underlying bedrock at lower altitudes. Limited to the valley of the Rotmurg a narrow area at the bottom of the valley shows anatexite bedrock (Schöttle, 2005). The recent geomorphology of the Black Forest has essentially been formed during the Pleistocene ice ages. Especially the Riss (300 Ka – 130 Ka) and Wurm (115 Ka – 10 Ka) glaciations covered largely the south of the Black Forest as well as north and east oriented slopes of the higher ranges in the centre and north (Schöttle, 2005).

The soils of the southern Black Forest are strongly dominated by brown soils (WRB soil group: Cambisols; FAO, 2014). To the north, the occurrence of brown soils decreases while the area covered by podsols (WRB soil group: Podzols; FAO, 2014) and stagnogleys (WRB soil group: Stagnosols; FAO, 2014) increases. Sparsely dispersed moors (WRB soil group: Histosols; FAO, 2014) are locally occurring at specific, glacially formed structures (Landesamt für Geologie, Rohstoffe und Bergbau, 2016). The soils in the National Park are mainly podsol and stagnogley. Brown soil can locally be found at the western boundary of the park as well as in valleys (Landesamt für Geologie, Rohstoffe und Bergbau, 2016).

The climate of the Black Forest is considerably influenced by west cyclonic atmospheric conditions transporting maritime, humid air from the Atlantic over France to the Rhine valley where the humidity might increase additionally (Schönwiese, 2008). Due to the north-south orientation of the Black Forest the air gets lifted and intensive orographic rainfall is frequently induced at the western windward slopes of the mountain range. This leads to mean annual precipitation rates of up to 2200 mm a^{-1}. After passing the ridge the mean annual precipitation declines considerably at the leeward slopes. The highest precipitation rates can be found at higher altitudes in the northern part of the Black Forest (Deutscher Wetterdienst, 2017). This maximum, at the orographically lower northern Black Forest, results from the

absence of higher elevations lifting the westerly streams on their way over France. Further south, the Vosges lift up the streaming air before it reaches the southern Black Forest reducing the precipitation there (Gebhardt, 2008). Regarding the climate zone the annual temperature ranges are basically temperate. Due to the higher altitudes of the mountain range, the temperature is colder than the average temperature of Baden-Württemberg. The highest peak of the Black Forest is the Feldberg (1493 m a.s.l.) with a 30-year mean monthly temperature between -3°C and +12°C (1981 – 2010) and a mean annual temperature of 3,9°C during the same period. The mean monthly temperature of Baden Württemberg shows a minimum of 0°C and a maximum of 18°C with a mean annual temperature of 8.7°C (Deutscher Wetterdienst, 2017). The 30-year mean annual temperature of the BFNP (1981 – 2010) shows values between 5.6 and 8.1 °C and the respective mean annual precipitation ranges from 1700 to 2200 mm a^{-1} (Deutscher Wetterdienst, 2017).

The German Federal Agency for Nature Conservation describes the priority protected habitats of the BFNP as montane mixed forests of beech, fir, and spruce as well as heathlands (Bundesamt für Naturschutz, 2016). Focussing on forested areas Norway spruce (*Picea abies* (L.) H. KARST.) is extensively distributed with a share of 70% of the tree species composition at the region Ruhestein and 60% at Hoher Ochsenkopf/Plättig (Figure 3-1). Furthermore, the composition of the Ruhestein region comprises 11% silver fir (*Abies alba* MILL.), 4% beech (*Fagus sylvatica* L.) and 4% Scots pine (*Pinus sylvestris* L.). At Hoher Ochsenkopf/Plättig the percentages of silver fir (18%), beech (7%) and Scots pine (11%) are considerably higher than in the southern part (PricewaterhouseCoopers & ö:konzept, 2013). The remaining proportions are mainly divided among Douglas fir (*Pseudotsuga menziesii* (MIRBEL) FRANCO), Japanese larch (*Larix kaempferi* (LAMB.) CARRIÈRE) and European larch (*Larix decidua* MILL.) with respect to conifers. Less represented broadleaf tree species are sycamore maple (*Acer pseudoplatanus* L.), mountain ash (*Sorbus aucuparia* L.), birch (*Betula pendula* ROTH) and oak (*Quercus robur* L. and *Quercus petraea* (MATTUSCHKA) LIEBL.). Considering the mixing of Norway spruce with silver fir, beech and Scots pine the share of the forested area with at least 20% of the latter three species is almost 80% at Hoher Ochsenkopf/Plättig and 41% at Ruhestein (Pricewaterhouse Coopers & ö:konzept, 2013). However, due to minor modifications of the boundary line in comparison between the expert report (PricewaterhouseCoopers & ö:konzept, 2013) and the finally designated area the figures may vary in detail. Despite these small uncertainties, the dominance of coniferous tree species, especially of Norway spruce is indicated across large parts of the study area. The current tree species composition as well as the predominant structure of the forests result from the long tradition of forest management in the northern Black Forest (Rösch, 2015). The protective forest 'Wilder See' appears as a notable contrast in structure. It has been removed from human usage since 1911 and got extended from 86 ha to 150 ha in 1998 (Schlund, 2003). In addition, disturbances like windthrow, bark beetle infestation and snow-break created structurally rich parts on different scales across the area of the national park (Nationalpark Schwarzwald, 2016).

The study site and subsequently the forest structure classification was limited to conifer dominated areas, openings, and gaps, due to the under-representation of equally mixed and deciduous dominated forest stands (see chapter 3.7.9). This restriction was necessary because a sufficient number of training observations in deciduous dominated stands could not have been achieved.

3.2 Forest structure stratification scheme

In order to monitor the natural, long-term development of forest structures in the BFNP a stratification scheme of six forest strata has been created, complemented by a non-forested open stratum following the experiences and recommendations of O'Hara et al. (1996) and Falkowski et al. (2009). The stratification numbers and names were as follows:

(1) Opening
(2) Gap
(3) Thicket
(4) Low- to mid-dimensional multi-story
(5) Mid-dimensional single-story
(6) Mid- to high-dimensional single-story with regeneration
(7) Multi-dimensional multi-story

These strata describe discrete development stages in a dynamic system but not the processes creating forest structures itself (Oliver and Larson, 1996). Thus, their naming has been derived from spatial characteristics, tree-dimensions, and vertical configuration. Terminology influenced by succession based concepts (O'Hara et al., 1996; Scherzinger, 1997) or by age class schemes of forest management systems (Gadow, 2005) has been avoided to emphasize structural attributes. Sequential order of the strata cannot be assumed in general, because some of the strata might develop out of various precursor stages and disturbances may break up any sequences (Foster et al., 1998).

On the one hand, the seven strata were specified to comprise a broad range of the recent forest structures as well as the potential developments under protection of natural processes. On the other hand, they should enable a differentiation by a classification algorithm, which required recognisable boundaries between the strata for remote sensing data analysis (Foody, 1999). The proposed seven stratum scheme combines these aspects while being aware that transition structures occur that either connect the discrete strata or could build up new strata in future. Considering that natural structures within artificially determined spatial units meet rarely the ideal of a stratum (Foody, 1999), the strata specifications consisted of various aspects (Table 3-1). The basic framework for stratum identification was set by a combination of verbal description and reference values for the main tree height range and the main diameter at breast height (DBH) range. Based on these criteria the stratum assignment was made in the field (see chapter 3.7.2). Forest structure silhouettes complemented the technical descriptions with

visual impressions of the strata 2 to 7 (Figure 3-2). A silhouette for openings is not provided, as this stratum was introduced to represent meadows or low vegetated areas (< 1 m).

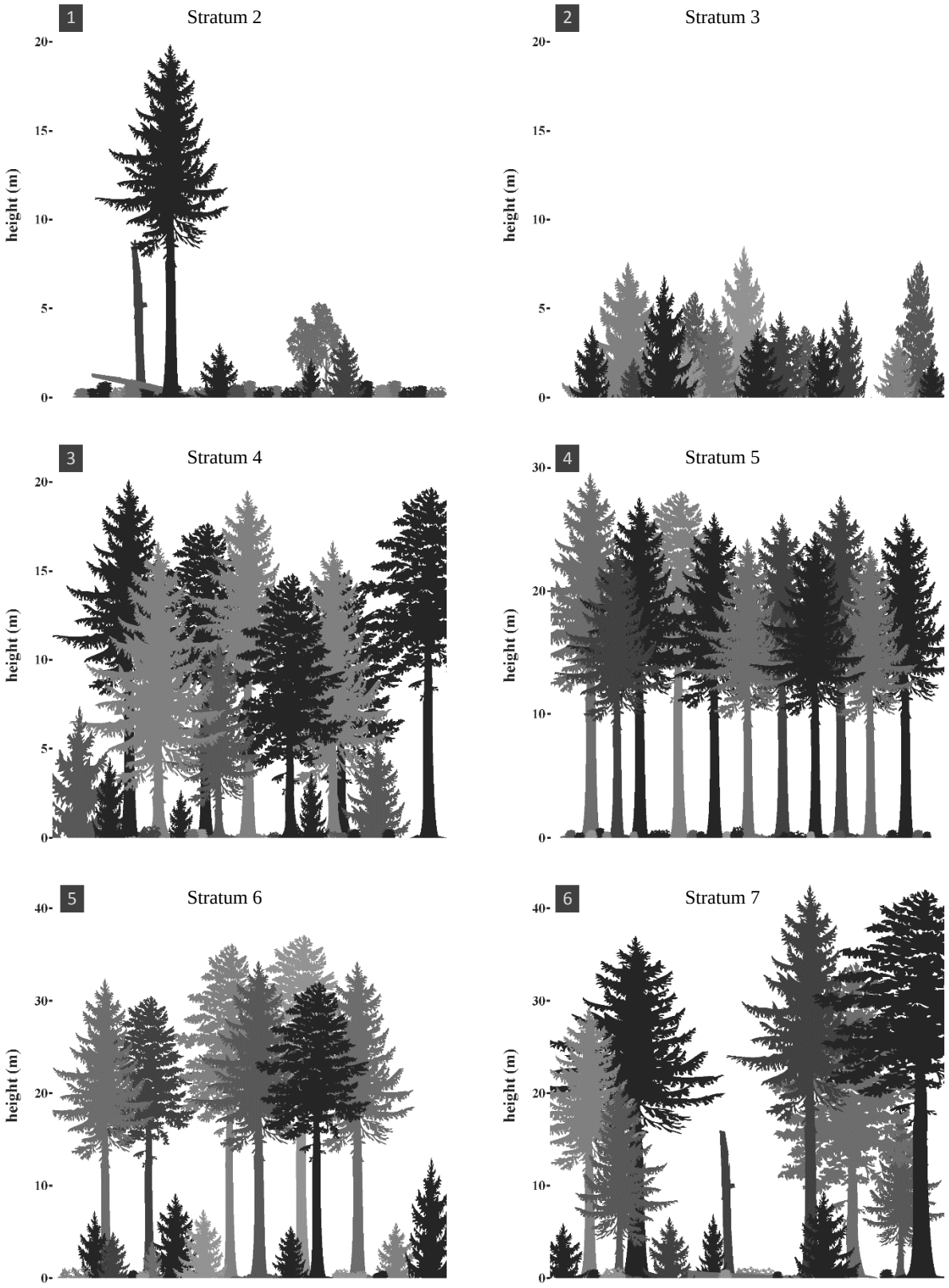

Figure 3-2 Forest structure silhouettes; gap (1), thicket (2), low- to mid-dimensional multi-story (3), mid-dimensional single-story (4), mid- to high-dimensional single-story with regeneration (5), multi-dimensional multi-story (6)

Table 3-1 Forest structure stratification adapted from O'Hara et al. (1996) and Falkowski et al. (2009)

No.	Stratum	Description	reference values		Number of plots
			main tree height range (isolated trees excluded)	main DBH range (isolated trees excluded)	
1	opening	Treeless grass, forb and dwarf shrub vegetation; regeneration, low snags, small rootplates as well as low-dimensional lying deadwood possible (e.g. meadows, cleared stand-replacing windthrows, etc.)	≤ 1 m	low regeneration	15
2	gap	Patchy regeneration or mountain pine (*Pinus mugo* TURRA) habitats with grass, forbs and shrubs; standing and lying deadwood, snags, rootplates and isolated trees possible (e.g. recent windthrows)	> 1 m	regeneration and > 7 cm	18
3	thicket	Homogeneous, quite closed regeneration or young trees; branches close to the ground; deadwood, snags and isolated trees possible	≤ 10 m	≤ 15 cm	22
4	low- to mid-dimensional multi-story	Low- to mid-dimensional trees of different height in closed stand with inhomogeneous canopy surface	≤ 30 m	≤ 50 cm	9
5	mid-dimensional single-story	Mid-dimensional trees, ± even-aged with tightly closed, homogeneous canopy; no understory; sparse regeneration below 1 m possible	15 – 30 m	25 – 60 cm	21
6	mid- to high-dimensional single-story with regeneration	Mid- to high-dimensional trees with opened but homogeneous overstory canopy; regeneration above 1 m or patchy to closed single-layered understory	≤ 5 m & 25 – 40 m	≤ 10 cm & 35 – 90 cm	19
7	multi-dimensional multi-story	Low- to high-dimensional trees of different height in ± closed stand with inhomogeneous canopy surface	≤ 50 m	≤ 90 cm	7

3.3 Spatial resolution

The foundation of the BFNP induced a conversion in forest management from profit-oriented cultivation to widely uninfluenced protection of natural processes (Landtag von Baden-Württemberg, 2013). Consequently, it was and is to be expected that disturbances on different scales, from stand replacing to fine-scale gap dynamics, will force a development towards mosaic forest patterns (Scherzinger, 1997; Brumelis et al., 2011). Detailed documentation of these processes in a forest structure monitoring system requires the optimum selection of spatial resolution. In this context, recent technologies offer a broad range of fine to coarse scales.

The finest resolution of LiDAR analysis in forest management and forest science is the individual tree approach. It aims on separating single trees within the LiDAR return cloud and subsequent derivation of respective tree attributes (Vauhkonen et al., 2014). Promising results in the separation have been achieved in the upper and with limitations in the intermediate layer of forest stands (Heurich, 2008; Reitberger et al., 2009). Identifying single trees is contrasted by area-based approaches characterizing groups of trees or forest stands within determined areas. They have developed to a well proven and standardized concept for describing the structure of forested ecosystems (Næsset, 2014; Wulder et al., 2014). A recent comparison between both methods by Latifi et al. (2015, p. 173) yielded "outstanding performance" of the area-based approach whereas the individual tree approach performed poorly in the understory and in young, dense forest stands. For this reason, an area-based approach appeared to be reasonable for a forest structure monitoring system and has been selected for this study. Shugart et al. (2010) distinguished two scale categories regarding the resolution of a rasterized grids. In their review, they concluded that resolutions of more than 1.0 ha capture equilibrium states of forest structure by reducing the internal variation due to aggregation of different mosaic structures. Conversely, plot sizes of 0.1 ha and less should meet extents of natural dynamics and display nonequilibrium states. They advised against choosing intermediate resolutions because of difficulties in interpretation (Shugart et al., 2010). This is consistent with resolutions of area-based studies which classified forest structure (Falkowski et al., 2009) or calculated forest stand parameters (Maltamo et al., 2004; Coops et al., 2007; Heurich and Thoma, 2008; Latifi et al., 2016). Their plot dimensions ranged between 10 m x 10 m (Latifi et al., 2016) and 30 m x 40 m (Maltamo et al., 2004).

According to the objective of analysing and documenting natural processes, their dynamic, and influence on biotic communities a small-scale grid with a cell size of 20 m x 20 m (400 m^2) was chosen for the forest structure classification in this study. The eastings and northings of grid intersections given in Gauss-Krueger projection (Krüger, 1912; Hooijberg, 1997) were integer multiples of the cell width of 20 m. Water areas have been excised beforehand as they showed almost exclusively zero returns per grid cell. Deciduous dominated areas and special, non-forested sites (e.g. main roads, car parks etc.) were excluded afterwards, based on the most recent forest management data (personal communication Gärtner, S., Birk, S., & Dreiser, C., 2017).

3.4 Field data acquisition

Before the BFNP was founded in 2014 the forest administration of Baden-Württemberg periodically collected statistical data in forest inventories. They took place at fixed points to assess profitability by controlling growth and condition of forest stands. The exact field locations were marked with a metal stick underneath the ground surface in the centre of each plot. From more than 3,500 points 245 plots have been selected as future monitoring plots by stratified sampling across the BFNP. A minimum distance to roads and paths of 20 m, diverse environmental conditions, inclusion of the forest structure strata from stratum 2 to stratum 7 (Table 3-1) and LiDAR return projections of potential areas were considered in the sampling process (Gärtner et al., 2016). Old forest management data and LiDAR return visualisation around the respective centre coordinates provided indications for the expectable forest structure. The major aim for the sampling process was the selection of a sufficient number of plots per stratum under consideration of the stratum area occurrence probability in accordance to the recommendations of Stehman and Czaplewski (1998) and Stehman (2009). This should have ensured their representativeness for classification purpose as well as in a long-term monitoring system. In addition to these 245 monitoring plots 15 sites representing stratum 1, opening, were determined by choosing forest glades using digital orthophotos (resolution: 0.4 m). There, artificial plot centres were digitally located without subsequent acquisition of field data due to the absence of considerable structural elements. The spatial distribution of the 15 plots depended solely on the size of the glades which should have ensured a buffer around the plot border of at least ten meters to avoid edge effects of the surrounding forest. These 260 plots were established as a basic set of potential observations from which the final training and test observations were chosen for the classification (see chapter 3.7.2).

Between 27 June and 25 November 2016 field data had been acquired on the 245 circular plots of strata 2 to 7 within a horizontal radius of 11.3 m (401.15 m^2). Horizontal distances were used to ensure identically projected base areas of each plot. A circular plot design was chosen to facilitate field data collection. Consequences of divergent shapes between the observation plots and the rectangular cells of the classification grid were considered. Potential edge effects due to the shape discrepancy were expected to be small and to be highly overcompensated by accelerated data acquisition and increased accuracy of plot position by avoiding orientation of quadratic plots to compass directions in the field (Næsset et al., 2013b). The coordinates of the plot centre were measured using a GPS handheld device, type Garmin eTrex 30 (Garmin Ltd., Southampton, Hampshire, UK), fixed to a pole at a height of approx. two meters perpendicularly above the centre. This allowed measurements by the evaluator at any point of the plot without influencing signal transmission to the GPS receiver. During the field campaign two GPS handhelds, both of type eTrex 30, were used. To improve accuracy of the centre measurement waypoint tracking was activated for at least 20 minutes per plot. The main device (GPS1), used at 218 plots, set waypoints every second and GPS2, used at 27 plots, operated in automatic mode. Thus, GPSI recorded at least 1,200 entries, whereas GPS2 set approx. 10% of the waypoints in comparison to GPS1 within equivalent timespan. The eastings and northings of the waypoints were

averaged in the geographic information system QGIS, version Essen 2.14.3 (QGIS Development Team, 2017) and mean eastings and northings represented the final centre coordinates. The two devices were referenced in parallel against a control point of the *Landesamt für Geoinformation und Landentwicklung in Baden-Württemberg* (LGL-BW) located in the town centre of Freudenstadt (Baden-Württemberg). By this the GPS measurement accuracy and precision was controlled. The results showed that although GPS2 recorded considerably less waypoints its accuracy was higher (Figure 3-3). The deviations of GPS1 showed a tendency to North-East and ranged between one and two meters. GPS2 yielded no tendency and all waypoint averages were within one meter around the reference point. Regarding precision both devices reached similar distributions. With respect to the different environmental conditions between the widely open town square and the mostly forested monitoring plots this referencing only served as an indicator for measurement quality. Because of the uncertainty about the measurement performance under forested conditions neither the setup of GPS1 was adjusted nor its results have been corrected.

Figure 3-3 GPS referencing at LGL-BW control point; Average waypoint coordinates of five reference sessions with secondly tracking (secTRK) of GPS1 and automatic tracking (autoTRK) of GPS2 (1); Information board at GPS control point (2)

At the study sites four categories of data have been collected (Table 3-2) on a field data form (Appendix C). At first basic information about the plot and its surrounding area has been noted. Secondly standing trees, independent of their vitality, snags with a minimum height of 2 m, and both with a DBH of at least 7 cm were recorded. The horizontal distance between plot centre and central axis of each trunk or snag ($\leq$ 11.3 m) was used as the threshold for inclusion in or exclusion from data collection. As third category lying dead wood has been recognized if the diameter at half-length was 12 cm or more representing coarse woody debris (Bässler et al., 2010). In addition, its total length or residual length inside the plot border had to be at least 2 m. And, fourth, rootplates divided into three diameter classes completed the data collection (Table 3-2).

Table 3-2 Field data category and attribute description

Data Category	Attribute	Description
basic information	general site description	specification of slope, exposition, surface soil characteristics, dominating and minor tree species, ground vegetation, deadwood, recent or old disturbances
	mean tree height upper third [m]	average height of 5 randomly chosen trees of the upper third of trees
	canopy cover [%]	percentage of ground surface covered by tree canopy; differentiated in total coverage (regeneration included) and percentage covered by the uppermost canopy stratum
	number of forest layers	number of differentiable canopy layers within the plot (between 1 and not differentiable)
	regeneration	presence or absence of regeneration below 1 m
	stratum	on-site assessment of forest structure stratum
standing trees and snags	species	in case of standing deadwood only if clearly identifiable
	vitality	distinction between vital, vital with broken top or dead
	distance [m]	horizontal distance from plot centre to central axis of trunk [b]
	azimuth [°]	angle between magnetic North and central axis of trunk measured from the plot centre [a]
	DBH [cm]	diameter at breast height [b]
	crown base height	perpendicular distance between ground and lowest vital primary branch [b]
	total height	perpendicular height from ground to top of dead trees and snags [b]
lying dead wood	species	only if clearly identifiable
	distance [m]	horizontal distances from plot centre to each of both trunk ends
	azimuth [°]	angles between magnetic North and each trunk end measured from the plot centre [a]
rootplates	distance [m]	horizontal distance from plot centre to rootplate centre
	azimuth [°]	angle between magnetic North and rootplate centre measured from the plot centre [a]
	class	dependent upon longest distance (d) of two opposite points through the centre of the rootplate 1: $d < 1.5$ m 2: $1.5 \text{ m} \leq d < 3$ m 3: $3 \text{ m} \leq d$

[a] subsequently corrected by declination (GFZ Potsdam, 2015); [b] Kramer and Akça, 2008

Various measuring instruments were used for data acquisition. The DBH of standing trees was measured with calibrated diameter tapes at a height of 1.3 m assuming a cylindrical shape of the trunk (Kramer and Akça, 2008). The tapes were also used for the diameter at half-length of lying dead wood. The resolution of diameter measurement was 0.001 m which corresponds to the estimated measurement error of manual use of a diameter tape.

Two Vertex IV ultrasound measurement instruments (Haglöf Sweden, Langsele, Västernorrland, Sweden) were used to determine distances and tree heights. Ultrasound technology is especially useful in dense structures with high regeneration and thick understory where measuring tapes and laser instruments cannot be operated. The resolution of distance values was 0.01 m while height was measured in steps of 0.1 m. With respect to distance measurement the Vertex IV operated with an accuracy of 1% or less whereas no manufacturer's information about the height measurement accuracy was available (Haglöf Sweden AB, 2014). Considering that height measurement with a Vertex IV depends on the trigonometric principle, requiring one distance, at least two angle measurements as well as strict position holding of the user, an accumulation of errors up to 5% seemed to be reasonable. This assumption was supported by the results of Heurich et al. (2003) comparing ground based height measurements using a Vertex III with a LiDAR based canopy height model. As an alternative to the Vertex IV a TruPulse 360°R laser range finder (Laser Technology Inc., Centennial, Colorado, USA) was applied to measure distances and heights in open forest structures where undisturbed visual contact from the plot centre to the individual trees existed. Besides the distance, the azimuth was measured by the internal electronic compass of the TruPulse 360°R in some open strata. Its distance resolution was 0.1 m and the azimuth was measured to 0.1° with a respective accuracy of 0.3 m and 1.0° (Laser Technology Inc., 2011). In the case of height measuring, the accuracy of the TruPulse 360°R was assumed to be comparable to the Vertex IV due to the similar trigonometric measurement principle and user application. At most of the plots two analogue compasses were used to record the azimuth (type Suunto KB-14/360R G, Suunto Oy, Vantaa, Finland). Their resolution was 0.5° with an accuracy of 0.33° (Suunto Oy, 2017).

All electronic devices were regularly calibrated according to the manufacturer's recommendations. The compasses were compared to each other to avoid malfunction. In the course of data post-processing the azimuth values, till then measured against magnetic North, were corrected by the magnetic declination which is caused by the dynamic deviation between geographic and magnetic North (Kahmen, 2006). For this reason, the German Research Centre for Geosciences (GFZ) offers an online declination calculator (GFZ Potsdam, 2015). It calculates the magnetic declination by entering the coordinates (easting, northing, and elevation above sea level) and the month of interest. The corrections used in this study were referred to the Ruhestein mountain pass (48° 34' N, 8° 13' E, 915 m a.s.l.), fairly in the centre of the BFNP. The month was chosen according to the respective date of field data acquisition. For June 2016 a value of 2.08° was added to the azimuth which changed from 2.10° in July to 2.12° in August and September, 2.13° in October, to finally 2.15° in November (GFZ Potsdam, 2015).

3.5 LiDAR data acquisition and pre-processing

Full waveform LiDAR data were acquired by MILAN Geoservice GmbH (Kamenz, Germany) in seven flight sessions between 29 April and 6 Mai 2015 (MILAN Geoservice GmbH, 2015). Flight strips were flown at an intended speed of 110 knots (203.72 km h^{-1}), an altitude of 1969 feet (600 m) above ground level, and a strip overlap of 80%. A RIEGL LMS-Q780 long-range laser scanner served as on-board measurement unit (Table 3-3). After the flight campaign MILAN firstly filtered the full waveform raw data with respect to quality criteria and secondly extracted discrete returns (MILAN Geoservice GmbH, 2015). The resulting dataset covered an area of 266 km^2 with an average density of discrete returns of 30.5 returns per square meter. The main information linked to each return was its easting, northing, and the height above sea level where the laser was backscattered. Mean standard deviation of height accuracy measurements ranged between 0.015 and 0.030 m, referenced at twelve ground control points provided by the LGL-BW. The relative position accuracy, measured as standard deviation of GPS position, has been given as mostly between 0.02 and 0.15 m (MILAN Geoservice GmbH, 2015). Besides the coordinates, number of targets, target number, intensity and signal length were documented for each return (MILAN Geoservice GmbH, 2015). The number of targets indicated how many reflections of a specific laser beam had been recognized. The target number gave the sequential number of the respective reflection within the number of targets. Intensity and signal length characterized the shape of the full waveform signal from which the discrete return was derived (Hancock et al., 2015).

Table 3-3 Performance and parameterization of RIEGL LMS-Q780 laser scanner

Parameter	Performance
Scan pattern	parallel scan lines
Laser wavelength	near infrared
Technical scan angle range	± 30° = 60° total
Eff. scan angle range	59.976° total
Mean swath width	692.485 m
Eff. laser pulse repetition rate	400.000 kHz
Eff. scan frequency	151.111 Hz
Laser wavelength	near infrared
Laser beam divergence	0.25 mrad
Footprint size	0.15 m^2
Intensity measurement	16-bit

Further pre-processing of discrete return LiDAR data was performed by the Chair of Remote Sensing and Landscape Information Systems (FeLis) at the University of Freiburg. A digital terrain model (DTM) was calculated which represented the unvegetated ground surface. In addition the file format

was changed from ASCII, as used by MILAN (MILAN Geoservice GmbH, 2015), to RWB. RWB is a compressing binary file format for laser scanning data which enables accelerated processing and visualization. It is the operating format of TreesVis, a LiDAR data processing software developed by the FeLis department (Weinacker et al., 2004).

3.6 LiDAR data analysis

3.6.1 Normalization and subsets of LiDAR data

LiDAR return heights, originally stored in height above sea level, were normalized to ground level by subtracting perpendicular height values of the underlying DTM in TreesVis. Resulting negative height values were manually set to zero where the DTM interpolation overestimated the natural relief. Derived from the entire normalized data set (data_0) four subsets were created depending on return height in combination with target number. In the data_0_f dataset only the first returns, defined as returns with target number 1, were separated. The data_0.5 dataset included returns equal to or above a height of 0.5 m. Thus, ground returns were excluded. Data_0.5 was interpreted as vegetation returns although boulders, hunting constructions, huts, etc. usually exceed 0.5 m. But returns of shrubs and trees dominated this dataset by far. A second threshold was at 2 m which separated returns of the tree canopy ($\geq$ 2 m) from ground and shrub canopy returns (Næsset, 2002). Finally, data_2_f reduced data_2 to the first returns in the tree canopy.

3.6.2 Forest structure LiDAR metrics

Metrics characterizing the three-dimensional LiDAR return distribution are essential in an area-based approach (Vauhkonen et al., 2014). Therefore, the return datasets were segmented in orthogonal columns of which the 400 m^2 cells of the BFNP grid (see chapter 3.3) defined the respective bases. Each segment was identified by the coordinates of its lower left base corner (i.e. minimum easting and minimum northing). Returns lying exactly on or above the western or southern border of a grid cell were included into the respective data segment whereas returns precisely on the opposite borderlines were excluded. This procedure ensured one-time consideration of each return during the metric calculation process. The mean return number per cell was 22,279 while the number of returns per cell ranged between 2,568 and 64,010. In contrast, most of the metrics of the field plots were calculated on circular plots with a radius of 11.3 m (401.15 m^2) around the centre coordinates (see chapter 3.4). Consequently, this led to cylindrical columns extracted from the LiDAR data set. The only exception were metrics resulting from subdividing each column into voxels of one cubic meter of volume (Lefsky et al., 1999; Coops et al., 2007) which required squared base areas (Table A-1). The respective squared bases were constructed by using the measured plot centre coordinates as intersection of their diagonals while the eastern and western borderlines were oriented to the north. The mean number of returns across the 245 cylindric field plot data segments was 23,062. The return numbers ranged between a minimum of 12,057 and a maximum of 40,417.

For each data segment 87 height- and density-related LiDAR metrics were calculated to characterize forest structure using the statistical software R, version 3.2.2 (R Core Team, 2015). Two separated codes calculating LiDAR metrics are attached in Appendix C because of differences in processing regular grid cell data and single field plot data (LiDARmetrics_grid.R and LiDARmetrics_traintest.R, Appendix C). The metrics based largely on the return height values (i.e. z-coordinates) and only slightly on their number of targets and the target number. Most of the metrics have already been tested in studies analysing forest structure by area based approaches (Lefsky et al., 1999; Coops et al., 2007; Falkowski et al., 2009; Zellweger et al., 2013) or by regression analyses of forest key values (Næsset, 2002; Heurich and Thoma, 2008; Latifi et al., 2016). In this study, they were divided into five groups summarized in Table A-1 which also assigns the sources of the following metrics in detail (Appendix A). The first group contained return counts per area (grid cell, plot and m^2) based on different datasets as well as the mean ($n.tar_{mean}$) and maximum ($n.tar_{max}$) of the number of targets. The latter two metrics described the average number of reflections per beam and its maximum per data segment respectively. The second group characterized the vertical distribution of returns over the continuous height of each data segment (Table A-1). These metrics were alternatively derived from the entire returns (data_0), the vegetation returns (data_0.5), and the canopy return data (data_2). This mostly led to three variations of each parameter. To characterize the vertical distribution basic height metrics (maximum, mean, quadratic mean, median, mode, standard deviation and variance), relative metrics (relative median, coefficient of variation), and return distribution characteristic measures (skewness and kurtosis) were calculated (Table A-1). The calculation of the 99th height percentile of first returns was specifically introduced to take into account the sensitivity of the maximum height return to outliers (van Kane et al., 2010). But in contrast to van Kane et al. (2010), the recommended height percentile was changed from the 95th to 99th of first returns due to an approximately 30-fold higher mean return density in the present dataset (0.97 m^{-2} to 30.5 m^{-2}). This buffer was ought to exclude outliers caused by birds and objects without connection to the underlying vegetation (e.g. extended branch ends of neighbouring segments). Thus the relative median height ($h_{med.r}$ and $h_{med.r.99}$) was calculated by dividing the median heights of the entire dataset, the vegetation, and the canopy dataset ($h_{med}, h_{med}^{veg}, h_{med}^{can}$) by the maximum height (h_{max}) and by the 99th percentile of first returns (hp_{99}^{f}) as well.

$$h_{med.r} = \frac{h_{med}}{h_{max}} \quad (1) \qquad h_{med.r}^{veg} = \frac{h_{med}^{veg}}{h_{max}} \quad (2) \qquad h_{med.r}^{can} = \frac{h_{med}^{can}}{h_{max}} \quad (3)$$

$$h_{med.r.99} = \frac{h_{med}}{hp_{99}^{f}} \quad (4) \qquad h_{med.r.99f}^{veg} = \frac{h_{med}^{veg}}{hp_{99}^{f}} \quad (5) \qquad h_{med.r.99f}^{can} = \frac{h_{med}^{can}}{hp_{99}^{f}} \quad (6)$$

The coefficient of variation (cov) has been calculated as the quotient of the standard deviation (sd) of heights and the mean height (h_{mean}) for the entire return cloud and for the vegetation and canopy return cloud separately.

$$cov = \frac{sd}{h_{mean}} \quad (7) \qquad cov^{veg} = \frac{sd^{veg}}{h_{mean}^{veg}} \quad (8) \qquad cov^{can} = \frac{sd^{can}}{h_{mean}^{can}} \quad (9)$$

Skewness ($skew$) and kurtosis ($kurt$) were measured for the distribution of n return height values $h_1, \dots, h_n$ by using the third and fourth standardized sample moment (Falk et al., 2002; Komsta and Novomestky, 2015). The respective formulas were applied on data_0, data_0.5, and data_2 with h_{mean}, h_{mean}^{veg}, and h_{mean}^{can} for h_{avg} respectively (Table A-1).

$$skew = \frac{\frac{1}{n}\sum_{i=1}^{n}(h_i - h_{avg})^3}{(\frac{1}{n}\sum_{i=1}^{n}(h_i - h_{avg})^2)^{\frac{3}{2}}} \quad (10)$$

$$kurt = \frac{\frac{1}{n}\sum_{i=1}^{n}(h_i - h_{avg})^4}{(\frac{1}{n}\sum_{i=1}^{n}(h_i - h_{avg})^2)^2} \quad (11)$$

The third group consisted of metrics characterising the return distribution within height sections (Table **A-1**). According to Heurich and Thoma (2008) height percentiles between the 10th and the 90th ($hp_{10}, \dots, hp_{90}$) based on data_0 were calculated. A second segmentation was determined by artificial breaks at 0, 1, 2, 10, 20, 30, and 60 m. As recommended by Falkowski et al. (2009), the number of returns were counted per height layer ($n.hl$). The last partitioning was defined by the height quartiles of data_0. The standard deviation of the resulting four sections (sd_{hq}) completed the group of sectioned vertical distribution metrics (Zellweger et al., 2013).

Density metrics represent another important group to characterize forest structures from LiDAR data in area-based approaches (Vauhkonen et al., 2014). Based on the study of Falkowski et al. (2009) height layer densities ($dens.hl$) were calculated by dividing the return numbers within height intervals ($n.hl$) by the number of total returns (n) based on data_0.

$$dens.hl_i = \frac{n.hl_i}{n} \quad \text{with } i = [0,1],]1,2],]2,10],]10,20],]20,30],]30,60] \text{ in meter} \quad (12)$$

Næsset (2002) proposed canopy density parameters ($dens.can$) per height layer limited to first canopy returns (data_2_f). They can be interpreted as laser beams fully penetrating different height layers. The respective height layers were defined by the 10% height percentiles of first canopy returns (hp_i^f).

$$dens.can_{hp_i}^f = \frac{hp_i^f}{n^f} \quad \text{with } i = 10^{th}, 20^{th}, \dots, 90^{th} \quad (13)$$

The quotient of the total number of first canopy returns ($n^{can.f}$) divided by the number of first returns (n^f) is interpreted as first return canopy density ($dens.can^f$; Table A-1).

In contrast, Latifi et al. (2016) and Heurich and Thoma (2008) measured densities by penetration rates. For this purpose, they divided return numbers between the ground and the lower boundary of a layer ($n_{<lb}$; with $lb = lower\ boundary$) by return numbers between the ground and the upper boundary ($n_{<ub}$; with $ub = upper\ boundary$). This resulted in a parameter which indicated the share of laser beams penetrated through the layer of interest. The difference between their approaches lied in the definition of the height layers. While Latifi et al. (2016) defined fixed heights to distinguish between the penetration rate through the canopy ($pr.can$), the shrub layer ($pr.shrub$) and the understory ($pr.ust$), Heurich and Thoma (2008) proposed flexible heights to measure penetration through an upper ($pr.ul$), intermediate ($pr.il$) and lower layer ($pr.ll$). The maximum height was again replaced by the 99th height percentile (hp_{99}^{f}) of first returns to reduce the influence of outliers. Furthermore a vegetational penetration rate ($pr.veg$) was calculated by the quotient of returns below a height 0.5 m, interpreted as ground returns, and the total number of returns (Heurich and Thoma, 2008). All return numbers based on the entire return dataset (data_0) while n corresponded to the total number of returns per segment.

$$pr.can = \frac{n_{<2m}}{n} \quad (14) \qquad pr.shrub = \frac{n_{<0.5m}}{n_{<5m}} \quad (15) \qquad pr.ust = \frac{n_{<0.5m}}{n_{<2m}} \quad (16)$$

$$pr.ul = \frac{n_{<0.8 \times hp_{99}^{f}}}{n} \quad (17) \qquad pr.il = \frac{n_{<0.5 \times hp_{99}^{f}}}{n_{<0.8 \times hp_{99}^{f}}} \quad (18) \qquad pr.ll = \frac{n_{<1m}}{n_{<0.5 \times hp_{99}^{f}}} \quad (19)$$

$$pr.veg = \frac{n_{<0.5m}}{n} \quad (20)$$

The last group of metrics characterized the vertical and horizontal distribution of returns by subdividing the grid cells or squared field plot areas in 1 m x 1 m subplots. Subsequently the return data segments were divided in cubic voxels of 1 m^3 of volume. The basic approach was introduced by Lefsky et al. (1999) with full waveform LiDAR data on 5 m x 5 m subplots. The size was reduced due to higher return density in this study. Coops et al. (2007) extended this idea to discrete return data which was the basis for the following voxel metrics focussing on the vegetation dataset (data_0.5). First, the voxels were distinguished in the categories empty and filled with respect to the occurrence of returns inside. The empty voxels were again differentiated in open gaps and closed gaps (Coops et al., 2007). They were categorized as open gaps if no filled voxel was located perpendicularly above in the respective column of voxels. Open may be interpreted as unshaded. If there was at least one filled voxel above an empty voxel, it was assigned to the category closed gap representing a shaded volume. Filled voxels were differentiated in the categories euphotic and oligophotic. Both designations refer to the energy distribution within a canopy and were suggested by Richards (1984). In the original context, the term euphotic referred to the part of the canopy which intercepts the bulk of available light (Richards, 1984) and was further interpreted as the uppermost 65% of the canopy (Lefsky et al., 1999). In terms of discrete LiDAR returns Coops et al. (2007) used the 35th height percentile of the subplot returns as threshold.

The voxel including the 35th percentile return and the filled voxels perpendicularly below were assigned to the oligophotic zone while those above were classified as euphotic.

The result was a four-class classification of open gaps (og), closed gaps (cg), euphotic zone voxels (ez) and oligophotic zone voxels (oz). The metrics derived represent the absolute voxel numbers of each class and the respective relative percentage in relation to the total number of voxels (Table A-1). The total number was calculated by multiplying 400 voxels per layer by the maximum height (h_{max}) rounded down to the nearest integer. In this case, maximum height was preferred as the 99th height percentile would have led to a disproportional loss of voxels especially in the open gap category.

3.6.3 Preselection of LiDAR metrics

The number of metrics had been reduced in a preselection from 87 to 60 considering their general suitability for classifying the forest development strata ([a] in Table A-1). In a first step, all metrics directly depending on the flight paths and the resulting overlap per area were excluded. Thus, variables simply counting returns over the entire height (e.g. n, n^f, n^{veg}, n^{can} etc.) or within height layers ($n.hl_{[0,1]}$, $n.hl_{]i,j]}$) have not been further processed. Secondly the 99th height percentile of first returns (hq_{99}^f) and metrics derived from this parameter were preferred in comparison to maximum height metrics to reduce sensitivity to outliers (van Kane et al., 2010). This led to an exclusion of the maximum height (h_{max}) and the relative medians that derived from it ($h_{med.r}$, $h_{med.r}^{veg}$, $h_{med.r}^{can}$). Furthermore, metrics solely based on the canopy return datasets (data_2 and data_2_f) were excluded because its explanatory power was mainly limited to higher strata whereas openings, gaps, and thickets were poorly characterized by returns above a height of 2 m. Finally, the number of voxels in the oligophotic zone as well as their share in relation to the total number of voxels were excluded as constituting the residuals of the complementing three voxel categories.

3.7 Random Forests classification of forest strata

3.7.1 The Random Forests algorithm and settings

Random Forests (RF) is a nonparametric classification algorithm consisting of a collection of tree-structured classifiers (Breiman, 2001). Each of the trees is grown by using an independent, identically distributed, random vector of predictors and a randomly selected subset of training observations. The final vote for a class depends on the most popular vote within the group of multiple trees (Breiman, 2001). RF has been successfully used for classifying land cover from remote sensing data in general (Gislason et al., 2006; Rodriguez-Galiano et al., 2012) and for classification and modelling of forested ecosystems from LiDAR data in particular (Falkowski et al., 2009; Martinuzzi et al., 2009; Leutner et al., 2012; Latifi et al., 2016). Besides the high classification accuracies (Cutler et al., 2007; Belgiu and Drăguţ, 2016), the popularity of RF rests on several advantages compared to other classification algorithms which led to its use in this study.

(i) Independence from distribution of predictor or response variables (Cutler et al., 2007)
(ii) Robustness against predictor collinearity (Dormann et al., 2013)
(iii) Low bias due to random predictor selection (Breiman, 2001; Prasad et al., 2006)
(iv) No overfitting as more trees are added to the forest (Breiman, 2001)
(v) Integrated method for determining variable importance and estimating classification accuracy (Breiman, 2001; Cutler et al., 2007)

In addition, the RF algorithm is parsimonious with respect to internal settings. Only two variables have to be determined. On the one hand, the number of trees (n_{tree}) has to be specified, which implies the number of bootstrap samples drawn from the observation data. One sample comprises approximately 63% of the observations selected for classification tree development (Cutler et al., 2007). The remaining observations are called out-of-bag (OOB) and serve as test set for the respective tree classifier. They enable the derivation of intrinsic test set error computation (Breiman, 2001; Liaw and Wiener, 2002). On the other hand, the size of the randomly chosen subset of predictors at each split (m_{try}) must be set. While growing an unpruned classification tree RF chooses the best splits per node among these m_{try} variables instead of using all predictors. Random sampling of predictors for each tree construction process reduces tree correlation (Breiman, 2001; Kuhn and Johnson, 2013).

Within the framework of RF classification in this study n_{tree} was set to 500 and m_{try} to the square root of the number of input variables according to the most common recommendations (Belgiu and Drăguţ, 2016). The 260 field plots served as a pool of observations from where the final training and test observations were selected (see chapter 3.7.2). Similarly, the preselected 60 LiDAR metrics ([a] in Table A-1) were the potential predictors for classifying forest structures. They were further reduced to reach a more parsimonious classification model (see chapter 3.7.7). In preparation, missing values were

treated appropriately because the RF algorithm is not able to process those (Kuhn and Johnson, 2013). Chapter 3.7.5 gives detailed information about the approaches applied to missing values.

3.7.2 Selection of training and test observations

RF is basically a supervised classification algorithm (James et al., 2015). It obtains well-labelled observation data for training and test purpose, which is a crucial step towards high quality in classification (Congalton and Green, 2009). Although surveying reference data physically in the field is considered as the most secure possibility of determining reference areas, it is prone to subjectivity and subsequent to errors and inaccuracies (Foody, 2002). Moreover, land cover complexity would cause mistakes if the strata are not clearly defined and if the inter-stratum variability is low or the intra-stratum variability is high (Pelletier et al., 2017). This stratum label noise may bias the learning process of supervised classification algorithms (Nettleton et al., 2010). Even if Pelletier et al. (2017) showed the comparable sturdiness of RF to this type of noise, its performance decreased as noise increased. In addition, position deviation of field data and remote sensing data may intensify the noise at forest structure boundaries (Dicks and Lo, 1990).

Next to adjusting stratum definition to balance inter- and intra-stratum variability of common forest structures in the BFNP (Table 3-1), a four-step selection approach of field plots was implemented to minimize stratum label noise.

(i) Field stratum attribution by identical evaluator
(ii) Dominance of conifers
(iii) Homogeneous structural plot environment
(iv) Reduction to well-labelled observations

The basis for consistency in stratum labelling was an attribution of the stratum by an identical evaluator in the field and in post-evaluation of plot data. Field evaluation basically had a stronger influence on the final stratum decision due to the simultaneous, not simplified perception of structural elements. Considered in the field, the basic acceptance criterion for stratum 3 to 7 observations was the dominance of coniferous trees that was defined as an estimated coverage of more than 50% of the forested plot area. The on-site stratum labelling of the plot was accompanied by a summarizing assessment of the forest structure environment at a distance of at least 10 meters around the plot boundary. Inhomogeneity between inside and outside the plot resulted in exclusion from training and test observations. This buffer ensured that deviation of the measured plot centre from its true position did not consequently lead to a divergence between field plot forest structure and LiDAR characteristics. Additionally, this avoided boundary situations as recommended by Dicks and Lo (1990).

Field situations usually differ from ideal forest structure conditions described in theoretical stratum definitions. In some points, descriptions were intentionally vague to cover a broader range of structural features within one stratum (Table 3-1). Both aspects created variation in the confidence of stratum

labelling, which reasonably influenced classification accuracy (Foody, 2002). To limit the variation, only plots with clear membership after on-site evaluation were admitted to further selection.

However, a representation of intra-stratum variability of structural features and a simultaneous emphasis of differences to the other strata had to be assured (Pelletier et al., 2017). Particularly important was the inclusion of observations with or without optional structural elements, referred to as 'possible' in Table 3-1. The general method to regard optional structural elements has been the aspiring of the ideal balance to train the whole range of recognizable structural patterns of one stratum. A frequent example posed isolated trees occurring in the strata gap and thicket. The training and test observations had to represent both subgroups of each stratum in a sufficient number, with and without isolated trees. Simultaneously, the boundaries between strata were aimed to be defined as clearly as possible. Sometimes unambiguous stratum decisions emerged to be a challenge in the field, clear memberships by strict exclusion of ambiguous plots was preferred. Inevitably, this led to broadened transition zones between strata.

By this means, 96 training and test observations were chosen from 245 field plots within stratum 2 to 7. Together with 15 opening plots (stratum 1) 111 observations were finally introduced to the RF classification (Table 3-1).

3.7.3 Evaluation of stratum labelling

Following the selection of training and test observations the labelling of strata 2 to 7 was evaluated by descriptive statistics. Stratum 1 was excluded, as no field data existed. Also, due to the low vegetation height threshold, the LiDAR data was supposed to assure the differentiation to the other strata. The evaluation of the remaining six strata based on the distribution of DBHs that were measured during the field campaign, and the 90th height percentiles (hp_{90}) of the cylindrical LiDAR-segments per plot. Diameter measurements were compared to the DBH reference values whereas hp_{90} was interpreted as the upper boundary of the main tree height interval (Table 3-1). Thus, they were appropriate parameters to evaluate the compliance with the references as well as for estimating the differentiation capacity of the selected observations.

In principle, the DBH distribution of each stratum was left skewed and mean diameters ranged approximately between 8 and 25 cm. This shows the impact of larger numbers of low-dimensional trees than of mid- to high-dimensional trees (Figure 3-4). The distributions of stratum 3 (thicket) and stratum 5 (mid-dimensional single-story) differed considerably from the other strata. The lower 75% of stratum 3 diameters lied within a narrow range between 7 and approx. 15 cm which was in line with the reference value. Stratum 5 diameters showed an average of approx. 25 cm while their range was comparatively broad and the skewness was only moderately directed to the left. In comparison to the reference values this effect was expected but in a larger extent (Table 3-1). Two reasons might have caused the mitigation. Firstly, plots perfectly meeting the stratum description but with diameters

between 15 and 20 cm were included in the training and test observations to broaden the spectrum of stratum 5 and in parallel to narrow the transition zone to stratum 3. The resulting inaccuracy of tree dimension seemed to be acceptable after weighing it against the clearly defined single-layered canopy and respective tree heights of 15 to 20 m. Secondly, many plots of stratum 5 comprised a small number of trees in the understory. They were negligible with respect to the general impression but cumulated they noticeably influenced the DBH distribution. The DBHs of the remaining strata largely lied within the expected main DBH range (Figure 3-4).

The distributions of hp_{90} showed higher variation between the strata in comparison to the DBH distributions (Figure 3-4). All six strata complied with the upper boundary of the reference main tree height range (Table 3-1). Stratum 2 showed a comparatively broad range resulting from a variety of isolated trees. For this reason, its reference height range was not limited upwards (Table 3-1). The trees in stratum 5 tended in majority to the lower heights of the reference height range which is consistent with the DBH distribution. Stratum 4 and 7 showed a compact range between the 1st and 3rd quartile indicating a quite similar canopy surface structure at different height levels within each of the strata. In summary, the hp_{90} distributions confirmed a well labelled classification of field observations with respect to the upper height limit.

The distribution of DBHs and 90th height percentiles led to the assumption that height measures have a considerably higher differentiation capacity. This supported the use of LiDAR remote sensing, where height information is a central aspect, and respective height metrics in area-based classification approaches (Falkowski et al., 2009).

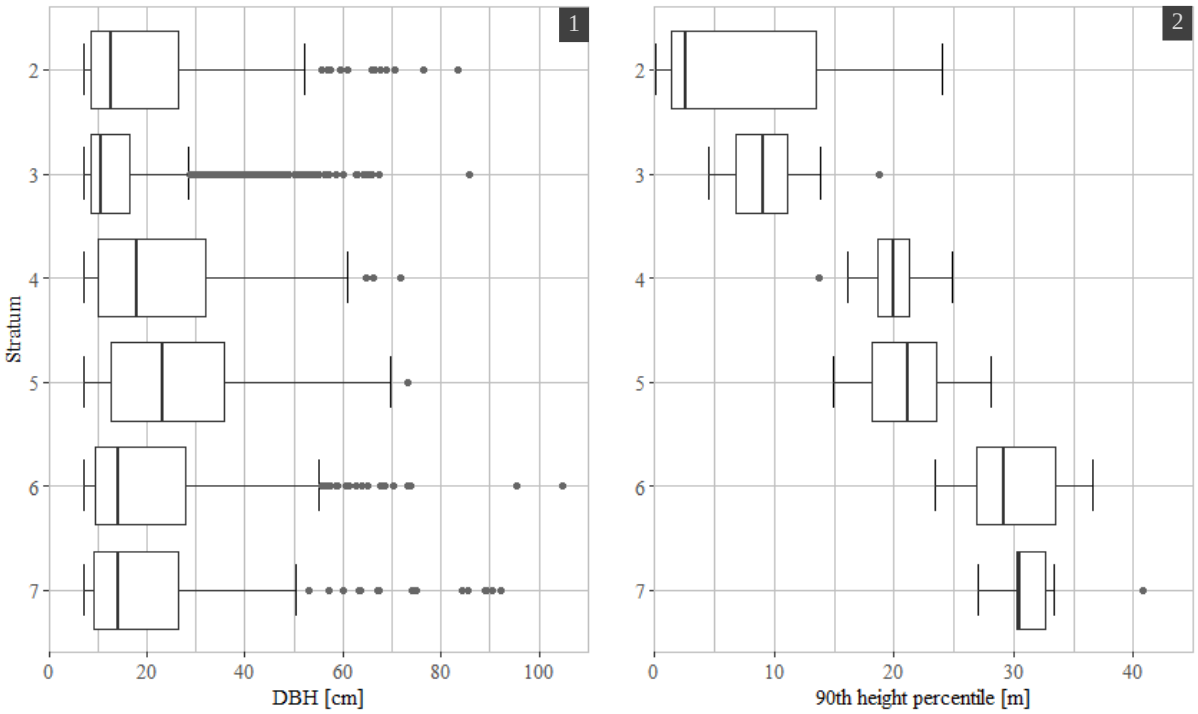

Figure 3-4 Boxplots of DBH distribution (1) and 90th height percentile (hp_{90}) distribution (2) per stratum

3.7.4 Imbalance of strata

The 111 observations were not evenly distributed across the strata (Figure 3-5). Especially stratum 4 and 7 were underrepresented with 9 and 7 observations respectively. Unbalanced distribution may cause poor classification accuracy for the minority strata, as RF is constructed to minimize the overall error rate. For this purpose RF tends to focus on accurate prediction of majority strata (Chen et al., 2004). Studies classifying imbalanced datasets comprised minority strata of 0.01% to 33% of the total number of observations in two or multiple strata datasets (Chen et al., 2004; Khoshgoftaar et al., 2007; Lin and Chen, 2013). This indicated that a share of 8.2% for stratum 4 and 6.4% for stratum 7 was a level of imbalance which could have had negative influence on the accuracy of the minority strata. Sun et al. (2009) pointed out that until their review, most of the research was done exclusively to binary strata imbalance problems and typical imbalance strategies of binary applications were not directly applicable to multi strata cases. They also stated that the degree of imbalance deteriorating the classification performance depends on additional factors like sample size and separability of strata. Thereby, the negative impact of imbalance increases if the number of observations is small and the level of inter-stratum noise is high (Sun et al., 2009). With an overall observation number of 111 and minority strata numbers of 9 and 7 the training and test set met the requirement to be small (Japkowicz and Shaju, 2002). Although separability was not explicitly measured, it was expectable that the strata 4 and 7 had a high potential for confusion due to their multi-layered structure. For example, sparsely occurring trees in the mid-dimensions could have caused similarity to stratum 6 by resembling a bimodal return height distribution. Therefore, it seemed to be reasonable to adapt RF to imbalanced data. Two basic adaption principles are discussed in literature. Cost-sensitive training introduces weighting factors putting increased emphasis on minority classification accuracy. In contrast, sampling strategies base on balancing observation numbers across strata (Chen et al., 2004; Kuhn and Johnson, 2013). Both approaches can be applied to imbalanced multi strata cases but cost-sensitive learning requires considerably higher tuning effort than sampling techniques (Kuhn and Johnson, 2013). Process transparency and experiences with two strata sampling (van Hulse and Khoshgoftaar, 2009) gave advantage to the application of sampling strategies and in particular to inherent down-sampling during the observation sampling process of RF (Kuhn and Johnson, 2013). Explicitly referred to multi-strata down-sampling unrestricted application was assumed for two reasons. Firstly, stratified down-sampling of multiple strata resulted in reducing each stratum to the number of the weakest stratum which is basically identical to down-sampling of imbalances between two strata. Secondly, the consideration rate of majority strata observations increased by repeated down-sampling within RF bootstrapping. In turn, the risk of information loss decreased if a sufficiently high number of trees in the RF ensemble was chosen (Kuhn and Johnson, 2013). Basing on these assumptions down-sampling of majority strata had been introduced in the training process (line 10, Table 3-6).

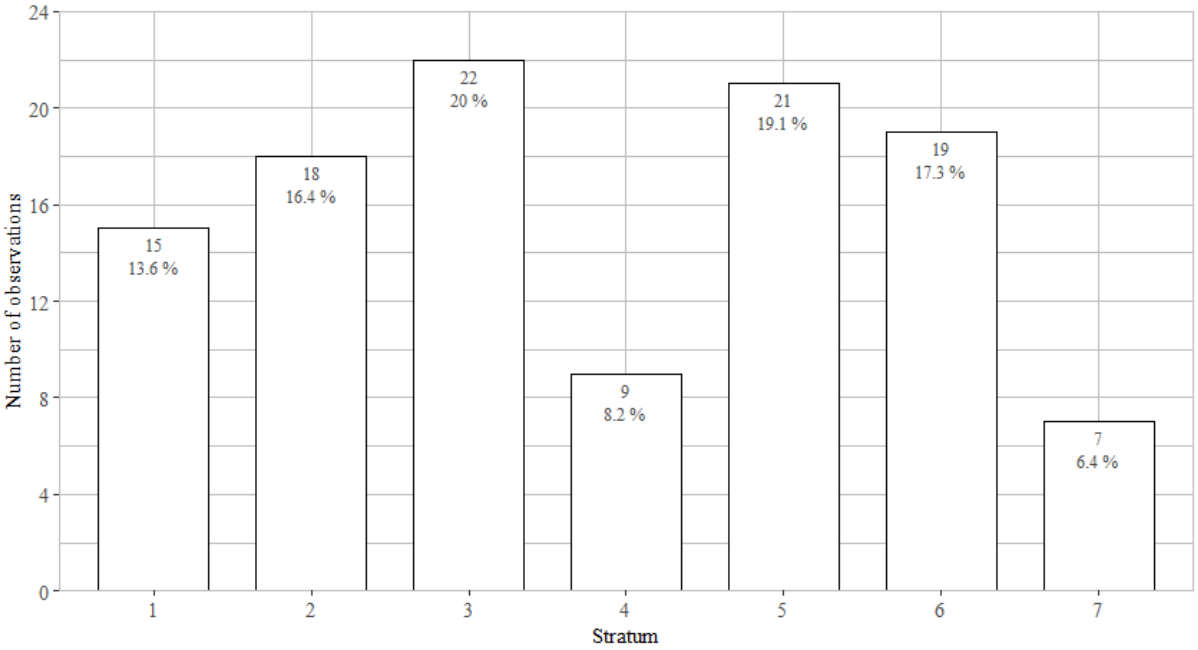

Figure 3-5 Imbalanced numbers of observations across strata

3.7.5 Missing values

Not availbale values and infinite values (NA and Inf) occurred in the LiDAR metrics dataset of the entire national park and in field plot LiDAR metrics. NAs resulted from the absence of returns in specific datasets (e.g. $n^{veg} = 0$ at many openings). Infinite values arose from dividing by zero (e.g. $hp_{99}^{f} = 0$). Although reasons for infinite entries differed from those for NA, both had been treated equally due to not suffering a loss of information by replacing Inf with NA. The term missing value (MV) is collectively used in the further course. Strategies for treating MVs had often been discussed in literature due to their frequent occurrence in datasets and the subsequent difficulties in processing, e.g. in classification routines (Little and Rubin, 2002; Saar-Tsechansky and Provost, 2007; Jerez et al., 2010; Newman, 2014). Explicitly in RF algorithms MVs cannot be processed (Kuhn et al., 2016). Their treatment had to be weighed for this study by analysation of underlying patterns and reasons for occurrence.

In the training and test observation data MVs were limited to 13 observations exclusively of stratum 1 and to 12 predictors (Table 3-4). Overall 120 missing values occurred. The metrics affected by MVs could be split into two groups. The first group comprised all predictors depending on vegetational LiDAR returns. In many of these cases openings showed zero returns at a height of 0.5 m and above. Calculation of vegetational predictors consequently failed and resulted in NA. The MVs of the second group depended on dividing by zero.

The BFNP dataset contained MVs in 2712 observations which were located mostly in openings, as indicated by 2364 cases with zero returns above 0.5 m. Additional 348 observations showed between 1

and 194 returns at 0.5 m or above. Their membership to stratum opening was assumed as well. The residual six observations had in contrast no returns below 2 m of height which led to MVs in the understory penetration rate predictor ($pr.ust$). They were located on the one hand on the roof of a large building and on the other hand in quarries. In both cases the DTM undercut the topology over a wider area which artificially lifted the returns during the normalization process (see chapter 3.6.1). In total 22415 MVs occurred and despite of the six understory penetration rate cases all of them were limited to the identical 12 predictors as at the training and test observations (Table 3-4).

Table 3-4 Missing value occurrence and respective imputation strategy

		Number of missing values			
Group	Predictor	training and test observations [a]	BFNP	Imputation	imputing value/metric
zero returns >= 0.5 m	h_{mean}^{veg}	12	2364	deductive	h_{mean}
	h_{med}^{veg}	12	2364	deductive	h_{med}
	$h_{med.r.99}^{veg}$	12	2512	deductive	0 m
	sd^{veg}	13	2457	deductive	sd
	var^{veg}	13	2457	deductive	var
	cov^{veg}	13	2457	deductive	cov
	$skew^{veg}$	13	2460	deductive	$skew$
	$kurt^{veg}$ [b]	13	2460	deductive	$kurt$
division by 0	$h_{med.r.99}$	6	801	deductive	0 m
	cov	7	1349	missForest	
	$skew$	3	364	missForest	
	$kurt$ [b]	3	364	missForest	
	$pr.ust$		6	missForest	
Sum		120	22415		

[a] limited to openings, [b] excluded after missForest imputation

The limitation of MVs primarily to openings showed that they were systematically distributed in both datasets. Hence, deletion of incomplete observations was not expedient because of disproportionate sample size reduction (Little and Rubin, 2002; Newman, 2014). Considering the training and test observations 13 of 15 opening plots were affected by MVs. Deletion would have inhibited the classification of this stratum. For similar reason dropping of the 13 predictors with MVs was not applied to maintain their explanatory power. Instead, imputation methods were used to avoid discarding valuable information. To achieve a feasible compromise between imputation accuracy and computation time a combination of deductive and algorithmic imputation was applied (Tang and Ishwaran, 2017).

The algorithmic imputation was executed before deductive imputation of MVs because missForest imputation results were used to manually impute remaining MVs. The missForest algorithm, an iterative imputation method based on RF regression and classification, was used to predict MVs by ensemble learning (Stekhoven and Buhlmann, 2012; Stekhoven, 2013). Due to its RF roots the advantages mentioned in chapter 3.7.1 applied to missForest as well. In particular, missForest offers OOB error estimates to assess the quality of the imputation procedure (Stekhoven and Buhlmann, 2012). In comparison with alternative algorithms missForest frequently outperformed compared imputation results (Stekhoven and Buhlmann, 2012; Tang and Ishwaran, 2017).

A reduction of MV affected predictors imputed by missForest was necessary to limit the computational time for the BFNP dataset. Selecting cov, $skew$, $kurt$ and $pr.ust$ for algorithmic imputation resulted from the absence of logical rules for deductive imputation. The number of calculated ensemble trees (n_{tree}) per iteration was set to 50. These measures limited the computational time to around three days instead of the estimated 75 days for all MVs (personal communication Reinhardt, K., 2017). In addition, the maximum iteration value was set to 6, which defined the maximum number of iterative variable reduction if the stopping criterion was not met beforehand (Stekhoven and Buhlmann, 2012). In fact, the missForest imputation stopped after the third iteration. Thereby, the number of predictor variables randomly sampled at each split (m_{try}) was the default square root of the total predictor number of each iteration (Stekhoven and Buhlmann, 2012). The MVs of the trainings and test observation set were treated equally in a separate imputation process which stopped after four iterations.

The OOB imputation error showed divergent results in comparison to the respective means and standard deviations (Table 3-5). The predictors cov and $pr.ust$ showed small OOB errors within an interval of one standard deviation. Skewness imputation ($skew$) provided ambivalent results. In the training and test set the OOB error was almost 250% of the standard deviation. In contrast, it stayed beyond 10% of the standard deviation of skewness in the BFNP data. Overall, the errors were considered acceptable and $skew$ remained as predictor in both datasets. The imputation of $kurt$ produced estimated errors largely exceeding the respective interval of one standard deviation in both datasets (Table 3-5). Kurtosis was consequently omitted from further processing. Because $kurt$ was intended to impute $kurt^{veg}$ after itself was imputed, both parameters were removed from the pool of LiDAR metrics ([b] in Table A-1). Thus, the number of potential RF predictors decreased to 58.

Table 3-5 Out-of-bag (OOB) error estimates of missForest in comparison to respective mean and standard deviation (SD)

Predictor	Training and test observations			BFNP		
	OOB error	Mean	SD	OOB error	Mean	SD
cov	0.12	1.01	0.84	0.01	0.96	0.75
$skew$	24.11	2.14	10.17	0.17	0.53	2.60
$kurt$	275552.16	120.32	806.88	3724.22	9.87	182.24
$pr.ust$				0.00	0.79	0.17

Deductive imputation bases on the principle that representative metrics may reasonably substitute missing values (Waal et al., 2011). The ideal form of imputation is given, when the analysis of the underlying error mechanism would lead to only one possible value (Kalton and Kasprzyk, 1986). But deductive imputation is reasonable as well, if the mechanisms can be guessed with high accuracy level (Waal et al., 2011). In the case of $h_{med.r.99}$ the MVs resulted from dividing by 0 m (hp_{99}^{f}) while the median was 0 m as well. The manual imputation with 0 m was reasonable because hp_{99}^{f} would start to become positive when 1% of first returns exceeds 0 m of height. From this point, the median would remain at 0 m till 50% of all returns lie above the ground. Thus a $h_{med.r.99}$ of 0 m should be representative for openings where vegetation height is low.

The remaining predictors assigned for deductive imputation contained MVs because of the absence of returns at or above 0.5 m which resulted in empty vegetation datasets (Table 3-4). This situation occurred only in low vegetated areas. To impute these predictors a replacing of vegetational metrics was consequent, depending on data_0.5 by the respective metrics along the entire height range (data_0). Metrics based on data_0 comprised ground surface and close to ground vegetation which closely approximated the vegetational indices in these specific situations.

3.7.6 Correlation of imputed LiDAR metrics

The effect of predictor correlations on classification algorithms and especially on importance measures was object of various studies (Gregorutti et al., 2017). Special focus was set on RF and its internal variable importance due to increasing relevance in science. However, summarising the results revealed ambiguous effects of predictor correlation on RF (Gregorutti et al., 2017). Because of this controversy, the risk of biasing variable importance by including correlated predictors against potential loss of valuable predictors after predictor selection should be carefully weighed. This was particularly important for the residual 58 predictors because pairwise correlation analysis revealed strongly correlated and anti-correlated variables (Figure 3-6). Analysing the Pearson correlation coefficients yielded only nine predictors with overall moderate to low correlation below an absolute value of 0.8 (Fahrmeir et al., 2011). These were $h^{veg}_{med.r.99}$ (21), cov (30), hp_{10} (39), sd_{hq3} (55), $dens.hl_{]1,2]}$ (58), $dens.hl_{]2,10]}$ (59), $dens.hl_{]10,20]}$ (60), $dens.hl_{]30,60]}$ (62) and ez_{rel} (86).

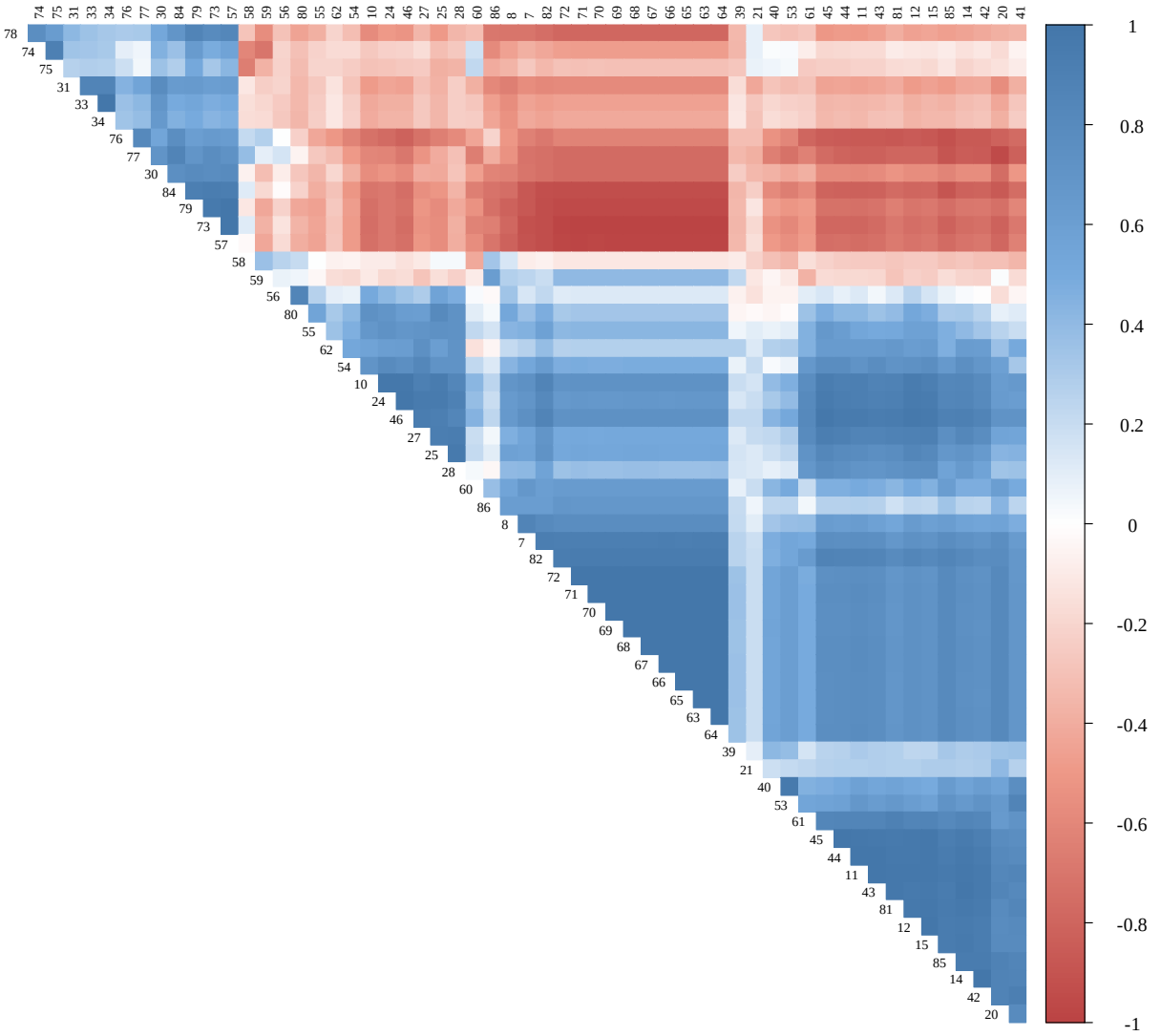

Figure 3-6 Correlation matrix of 58 LiDAR metrics after missing value imputation (Table A-1)

In their detailed introduction to feature selection Guyon and Elisseeff (2003) pointed out that only perfectly correlated variables are truly redundant. They stated that even highly correlated predictors might gain additional information (Guyon and Elisseeff, 2003). This is supported by the results of Cutler et al. (2007) who did not recommend an elimination of predictors solely for reduction of correlation. Therefore, feature selection methods processed before and independent from the classification algorithm were not applied in this study. Instead, integrated recursive feature elimination was used due to presumed redundancies indicated by the occurrence of perfectly correlated metrics.

3.7.7 Recursive feature elimination and final model selection

A feature selection routine reducing LiDAR metrics was used for two reasons. At first, perfectly correlated predictors do not lead to information gain (Guyon and Elisseeff, 2003) but might have negative effects on the model accuracy (Gregorutti et al., 2017). Secondly, keeping classification models parsimonious increases the interpretability of their results by selecting the relevant predictors (Nilsson et al., 2007). To prevent information loss due to elimination of relevant LiDAR metrics, recursive feature elimination (RFE) was applied which uses variable importance measures of classification algorithms for the selection (Guyon et al., 2002). In the present study RF's inherent variable importance measurement served as the RFE ranking criterion (Breiman et al., 1984; Kuhn et al., 2016). The applied RFE approach comprised an iterative three-stage procedure. At first the RF classifier was trained. Secondly, the importance variables were computed and the predictors were ranked accordingly. As the last step predictors with the weakest importance were removed (Guyon et al., 2002). Gregorutti et al. (2017) showed that RFE provides reliable results in identifying parsimonious RF models of high accuracy even under the condition of correlated predictors. The implementation of RFE in this study was wrapped in an outer resampling (Table 3-6, line 3) considering the improvement suggestion of Kuhn and Johnson (2013). They criticised that the control of the RFE results by leave-one-out cross-validation took place after the feature selection in the originally published RFE procedure by Guyon et al. (2002). This would lead to biased model results with expectable low error rates (Ambroise and McLachlan, 2002). To decrease the likelihood of selection bias an outer resampling was implemented by a 10 times repeated, 5-fold cross-validation (Kuhn and Johnson, 2013). By this, variation of feature selection was induced as only 80% of the observations were used for training purpose (Table 3-6, line 4 and 10) and 20% for performance measurement (Table 3-6, line 5 and 11).

The last control level for biased model results was established in advance by random partitioning of the original data. Hereby, independent test observations were separated which comprised 30% of each stratum (test_1; Table 3-6, line 2). This subset was not included in the RF training and prediction procedures. It represented the last stage for testing the final model based on the RFE selected predictor variables (Kuhn and Johnson, 2013). The resulting performance parameters of this prediction were used to evaluate the final model (Table 3-6, line 18).

One important setting within the RFE procedure is the subset of predictors (S_i). It determines the numbers of predictors (i) which should be tested to find the best model under the secondary condition of model parsimony (Kuhn and Johnson, 2013). The interval for i ranged between 1 and 25 while this range was incremented by two at the beginning, before it continued in steps of one from 9 to 25 (Table 3-6, line 8). This interval limited computational effort by inducing 21 iterations on the one hand and covered the most likely range of subset sizes on the other hand. A maximum of 25 predictors seemed to be reasonable under the objective of a parsimonious model.

Table 3-6 Resampled RF classification incorporating recursive feature elimination and class down-sampling (according to Kuhn and Johnson, 2013)

1	Missing value imputation
2	Class-stratified, random partitioning of basic data in training_1 (70%) and test_1 set (30%)
3	**for** *each resampling iteration (10 times repeated, 5-fold cross-validation)*
4	Partitioning of training_1 data into training_2 and test_2 set via resampling
5	Random Forests training on down-sampled training_2 data using all predictors
6	Prediction on test_2 data
7	Calculation of variable importance or rankings
8	**for** *each subset* S_i, $i = 1, 3, 5, 7, 9, 10, \ldots, 25$
9	Keeping the S_i set of the i most important predictors
10	Random Forests training on down-sampled training_2 data using S_i
11	Calculation of model performance by predicting the test_2 data
12	**end**
13	**end**
14	Calculation of the performance profile over S_i
15	Determination of the appropriate number of predictors
16	Estimation of the final list of predictors to keep in the final model
17	Fitting the final model based on the optimal S_i using the training_1 set
18	Prediction on the test_1 data

The model performance for each subset was measured by the area under the receiver operating characteristics curve (AUC) due to its advantages to scalar measures, e.g. classification accuracy or error rates (Fawcett, 2006). Receiver operating characteristics (ROC) are widely used in classifier selection due to representing the trade-off between true positive and false positive classification decisions. The AUC is a single scalar value representing the expected performance of a classifier by combining ROC values of resampled training and test runs (Fawcett, 2006; Kuhn and Johnson, 2013). The selection of the most important predictor subset depended on the exceedance of an AUC threshold (Table 3-6, line 15). This value was set to 98.5% of maximum performance reached by RF models within the subset interval and by the full model (i.e. all predictors included). Assuming a degressively increasing

development of AUC with increasing number of predictors, the threshold approach should fix the subset size at an early stage of constant performance levelling. Following the determination of the appropriate number of predictors the final list of most important predictors had to be estimated (Table 3-6, line 16). An estimation was necessary due to the wrapping resampling process which produced multiple lists of favourite predictors. Therefore, a consensus ranking routine, implemented in the RFE algorithm, enabled the predictor selection (Kuhn et al., 2016). At last, the final model was fitted to the original training observations (training_1) based on the most important predictors (Table 3-6, line 17).

3.7.8 Model performance evaluation

The model performance was assessed by several accuracy statistics. The evaluation statistics were divided into the OOB estimates produced during the model training process (Breiman, 2001; Liaw and Wiener, 2002), and the final model control on the independent test observations (Table 3-6, line 18). The statistics comprised summarising performance indicators and those calculated for each classified stratum. The summarising statistics represented the error rate, the overall accuracy and the kappa coefficient (K_{HAT}). The error rate is defined as the percentage of false classifications in relation to the sum of all classified observations. Vice versa, the accuracy calculates the percentage of correct classifications (Congalton, 1991; Kuhn and Johnson, 2013; Kuhn et al., 2016). This study additionally reported the K_{HAT} as a more informative, summarizing indicator, because the suitability of error rates and accuracies for model evaluations are controversially discussed (Congalton, 1991; Kuhn and Johnson, 2013). Its calculation followed equation 21 where r is the number of rows, x_{ii} the number of observation in row i and column i with x_{i+} and x_{+i} as the marginal totals of the respective row and column. At last, N is the total number of observations (Congalton, 1991).

$$K_{HAT} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} * x_{+i})} \quad (21)$$

The indicators per stratum, producer's accuracy (PA) and user's accuracy (UA) supplemented the evaluation of model performance. They were calculated on the basis of the equations 22 and 23 (Kuhn and Johnson, 2013; Kuhn et al., 2016). PA is the relative share of correctly classified observations within one stratum while UA reports the ratio of correctly predicted observations to all observations classified in one stratum.

$$Producer's\ accuracy = \frac{TP}{TP+FN} \quad (22)$$

$$User's\ accuracy = \frac{TP}{TP+FP} \quad (23)$$

The underlying confusion matrices were calculated as a result of the RF algorithm executed in R (R Core Team, 2015; Kuhn et al., 2016). They showed estimated classification results for the OOB predictions and final results for the prediction of the independent test observations. Basically, all performance indicators mentioned above based on confusion matrices. However, in the case of the OOB estimates

the indicators can seem to be unrelated to the confusion matrix on the trainings observations. This results from various background calculations of the RF algorithm to average the OOB indicators over the bootstrapped classification trees (Breiman, 2001; Kuhn et al., 2016). In contrast, the confusion matrix at the end of the RF training process displays the prediction result of the final model on the training observations (Table 3-6, line 17; Kuhn et al., 2016).

3.7.9 Classification and accuracy assessment

The segmented and imputed BFNP dataset was classified by means of the final classification model (Table 3-6, line 17). Afterwards the classified area was limited to areas representing either conifer dominated forests or areas of opening and gap characteristics (Table 3-1). The selection based on forest stand descriptions of the most recent forest management data of 2014 (personal communication Gärtner, S., Birk, S., & Dreiser, C., 2017). These descriptions were automatically filtered for expressions indicating dominantly coniferous tree species. Specific sites like forest glades, heathlands with mountain pine vegetation (*Pinus mugo* TURRA) and forest gap structures were added. Deciduous and equally mixed forest stands were excluded as well as sections of main roads, car parks, and water surfaces.

Across the strata, sets of conditional density plots (CDP) were compared. Thus, an indicator for the differentiation capacity of predictors and also a basis for ecological interpretation of classification results could been gathered (Falkowski et al., 2009). The highest listed predictor of each metric group according to the variable importance of the final model was chosen as a representative sample (Figure 4-2 and Table A-1). The CDPs were created for each stratum summarising all final classified grid cells. The ecological interpretation focused on the comparison of forest structure definitions of the strata (Table 3-1) and the CDP characteristics. The assessment of the differentiation capacity based on the degree of dominant occurrence probabilities. In the case of well differentiable strata the patterns of the CDPs would show clearly noticeable differences and dominant probability maxima. Increasing similarity in combination with reduced probability indicates poor identifiability by the given set of predictors.

The classification results could not be evaluated by confusion matrices, due to the lack of references. An analysis of stratum membership information was used instead to assess the quality of the classification. The stratum membership probability was analysed to identify possible delimitation difficulties between the strata (Foody, 2002; Falkowski et al., 2009). The required membership information was automatically produced by the RF classification algorithm, which stored each vote probability per stratum per cell. With increasing maximum vote probability the classification decision becomes more unambiguous (Foody, 2002). Assuming a rare occurrence of misclassifications with high probability for the wrong stratum, the confidence of the classification is supposed to rise with increasing probability of the maximum vote.

4 Results

4.1 Predictor selection and variable importance

The recursive feature elimination, inherently applied in the RF classification, reduced the number of predictors involved in the final model from 58 to the 13 most important LiDAR metrics. The criterion for the model selection was the exceedance of the ROC threshold defined as 98.5% of the highest occurring ROC value over the tested number of variables. The maximum ROC was 0.980 by using all variables in the model (not shown in Figure 4-1). The resulted threshold of 0.966 was firstly reached by the 13-predictor model (filled red circle, Figure 4-1). The model, selected by the RFE algorithm, contained the following predictors according to the RF variable importance (Figure 4-2): mean height (h_{mean}), median height (h_{med}), standard deviation (sd) and variance (var) of heights, skewness of vegetation heights ($skew^{veg}$), 60th, 70th, 80th, and 90th height percentile ($hp_{60/70/80/90}$), standard deviation of return heights between the 1st and 2nd height quartile (sd_{hq2}), relative percentage of returns between 2 m and 10 m of height ($dens.hl_{]2,10]}$), penetration rate through the understory ($pr.ust$), and number of closed gap voxels (cg) within or below the canopy ([c] in Table A-1).

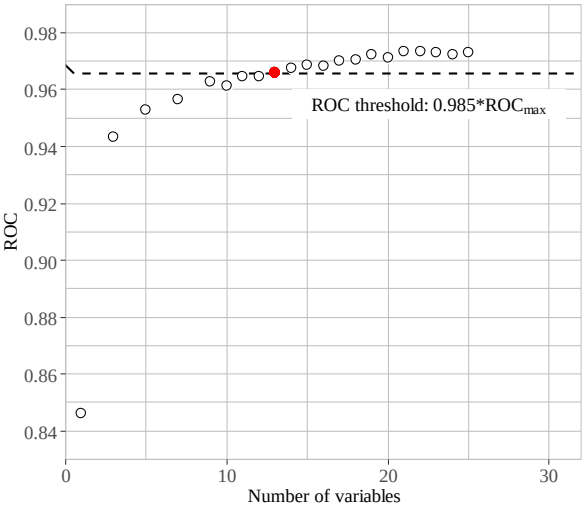

Figure 4-1 RFE calculated ROC values for RF models of variable numbers between 1 and 25; the filled red circle highlights the first threshold (dashed line) exceedance

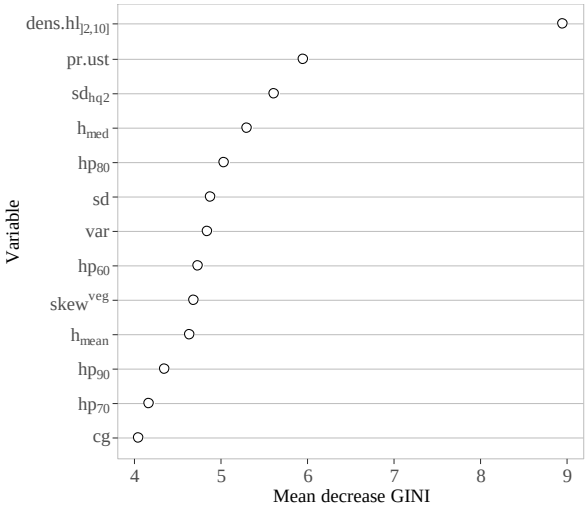

Figure 4-2 Variable importance of the 13 predictors selected for the final model

The specific return density $dens.hl_{]2,10]}$ clearly proved to be the most important predictor with a mean decrease of the GINI index of approximately 9 index-points. The remaining twelve predictors ranged between 6 and 4 index-point decreases (Figure 4-2). The two most important predictors characterized return densities in heights between 0.5 m and 10 m in total. This height interval excluded the influence of ground returns and set the focus on heights where usually returns occurred across all forested strata.

Above ten meter of height a lot of training and test plots of the strata gap and thicket showed no returns due to their low vegetation heights. The occurrence of four height percentile predictors, mean height, and median height within the list of the most important predictors emphasized the influence of height metrics in the classification of forest structure. Standard deviation over the entire height range and within the second height quartile, variance, and skewness of vegetation returns indicates that metrics characterizing the vertical return distribution were of considerable importance as well. The number of closed gap voxels introduced the 3D return distribution aspect in the model, although it had the lowest impact of the 13 most important predictors (Figure 4-2).

4.2 Correlation of final model predictors

Analysis of pairwise Pearson correlation coefficients of the 13 variables included in the final model showed mostly high correlations or anti-correlations. As the only predictors $pr.ust$ and $dens.hl_{]2,10]}$ yielded low pairwise correlation to the remaining ten predictors. The correlation coefficient between these two predictors met the border to medium correlation (Figure 4-3; Fahrmeir et al., 2011). The height related metrics (h_{med}, h_{mean}, $hp_{60/70/80/90}$), standard deviations (sd, sd_{hq2}), var, and cg built a group of very high to perfectly correlated predictors. Every metric of this group showed in turn high anti-correlation to $skew^{veg}$ (Figure 4-3).

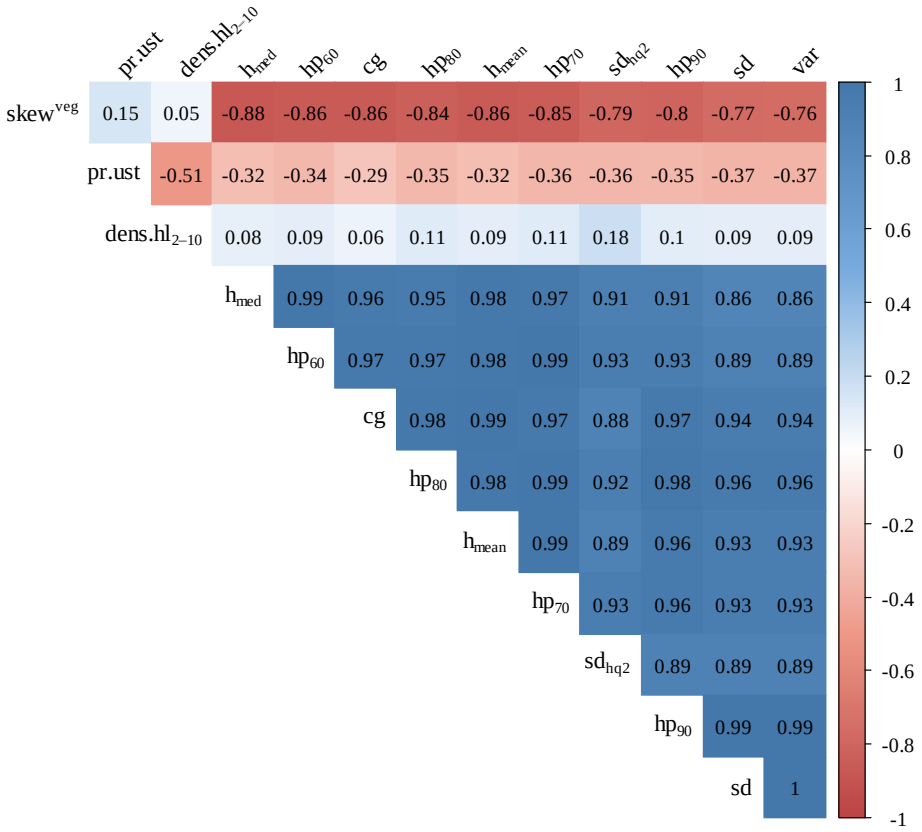

Figure 4-3 Correlation matrix of the 13 most important predictors of the final RF model ($dens.hl_{]2,10]}$ is replaced by $dens.hl_{2\text{-}10}$ in this figure)

4.3 Classification accuracy and performance

4.3.1 OOB estimates

The OOB error statistics, inherently calculated during the RF bootstrap process, revealed an OOB estimated classification error rate of 12.35%. The total accuracy over all bootstrap samples was 81.47% while the kappa coefficient (K_{HAT}) was 0.78.

On a stratum individual basis, the opening observations (stratum 1) were perfectly classified with a producer's and user's accuracy of 1.00 (Table 4-1). The lowest PAs were reported for the two multi-story strata, low- to mid-dimensional (stratum 4) and multi-dimensional (stratum 7), with 0.71 and 0.80 respectively. Stratum 4 showed confusion with thickets (stratum 3) as well as with mid-dimensional single-story forest structures (stratum 5). The classification of stratum 7 showed a tendency to forest structures of mid- to high-dimensional, single-storied characteristic with regeneration (stratum 6). Pairwise confusion was revealed between gaps (stratum 2) and thickets as well as between stratum 5 and stratum 6 (Table 4-1). Although this had reduced PAs as a consequence they reached accuracy values from 0.86 to 0.92. The UAs divided the strata in two groups. The openings and the two multi-story strata were free of erroneously classified plots of other strata with UAs of 1.00 (Table 4-1). The remaining strata either showed pairwise confusion (stratum 2 and stratum 3; stratum 5 and stratum 6) or misclassification of multi-story strata. Thus, misclassification shifts from multi-story strata to structurally adjacent strata occurred more frequent than shifts towards multi-story strata. The UAs of the respective, confused strata ranged between 0.80 and 0.88 (Table 4-1).

Table 4-1 OOB confusion matrix and classification performance measures per stratum

		Reference							User's accuracy
	Stratum [a]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	
Prediction	**1**	11	0	0	0	0	0	0	1.00
	2	0	12	2	0	0	0	0	0.86
	3	0	1	14	1	0	0	0	0.88
	4	0	0	0	5	0	0	0	1.00
	5	0	0	0	1	13	2	0	0.81
	6	0	0	0	0	2	12	1	0.80
	7	0	0	0	0	0	0	4	1.00
	Producer's accuracy	1.00	0.92	0.88	0.71	0.87	0.86	0.80	

[a] for stratum information see Table 3-1

4.3.2 Test observations

The test_1 set of observations, separated before the model training process started, was used to perform an independent assessment of classification accuracy and performance of the final model (Table 3-6, line 2 and 18). The prediction of these 30 observations revealed an overall classification accuracy of 90.00%. The increased accuracy in combination with a kappa coefficient rise to 0.88 revealed a considerably better model performance on the test data in comparison to the OOB estimates.

Besides stratum 1, as already reported by the OOB estimates, stratum 2 and stratum 3 were additionally classified with PA and UA of 1.00 on the test observations (Table 4-2). Regarding the PAs the confusion analysis reported accuracies for stratum 5 and stratum 6 of 0.83 and 0.80 respectively. In the case of stratum 5 one observation was misclassified as a stratum 6 and one stratum 6 reference plot was attributed to stratum 7. Based on only two test observations, stratum 7 showed a split result with a PA of 0.50 due to confusion with stratum 4. Consequently, the UA of stratum 4 was reduced to 0.67. The UAs of the strata 6 and 7 were influenced by the before mentioned, single misclassifications from the strata 5 and 6. Again stratum 7 seemed to be strongly influenced with respect to its UA of 0.50 but this resulted again from one incorrectly classified reference plot. No observation was erroneously classified as stratum 5 which led to a UA of 1.00 (Table 4-2).

Table 4-2 Test observation confusion matrix and classification performance measures per stratum

		Reference							User's accuracy
	Stratum [a]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	
Prediction	**1**	4	0	0	0	0	0	0	1.00
	2	0	5	0	0	0	0	0	1.00
	3	0	0	6	0	0	0	0	1.00
	4	0	0	0	2	0	0	1	0.67
	5	0	0	0	0	5	0	0	1.00
	6	0	0	0	0	1	4	0	0.80
	7	0	0	0	0	0	1	1	0.50
	Producer's accuracy	1.00	1.00	1.00	1.00	0.83	0.80	0.50	

[a] for stratum information see Table 3-1

4.4 Classification results

By the restriction to conifer dominated forest stands the overall classified area was reduced from 10,062 ha to 9,804 ha (Table 4-3; Figure 4-4). A lot of deciduous dominated forest stands were excluded from the classification especially in the north of the northern part.

Almost half of the BFNP was classified as stratum 3 and 5 (23.0% and 25.2%). Both strata form larger entities which were clearly identifiable in the overview map (Figure 4-4). Stratum 2 represented 17.5% of the classified area and appeared often in coherent areas as well. They were mostly adjacent to stratum 3. Both strata clearly traced windthrow structures of open to densely regenerated character and the typical heathland zone along the main road, B500, on the N-S ridge of the southern part (Figure 4-4). Stratum 6 was attributed to 15.5% of the grid cells. It showed ambiguous patterns as some larger units were almost exclusively classified as stratum 6 whereas a lot of small, dark brown patches were spread over the BFNP. However, a spatial connection between stratum 6 and stratum 5 became obvious (Figure 4-4). The lower multi-storied stratum 4 was quite frequently classified with a share of 13.6%. Its grid cells were distributed without clearly forming aggregated units. A detailed visual analysis revealed that stratum 4 was frequently identified at transitions between lower and higher forested areas. This occurred, when stratum 3 was adjacent to stratum 5 or 6. Furthermore, stratum 4 was attributed to grid cells at borders of forest edges to open structures like forest roads or glades (Figure 4-4). With a share of 4.3% the higher multi-storied stratum 7 was weakly represented. Usually these grid cells occurred in combination with stratum 6 and stratum 4. Zones with accumulation of stratum 7 cells were dispersed across the national park but they didn't form coherent areas. Only 0.9% of all grid cells were attributed to stratum 1. The areas classified as openings traced mainly forest glades and meadows according to the strictly defined training observations. A second emphasis of occurrence was across the heathland zones, where loosely spread stratum 1 cells or small patches were recognized in between gap and thicket structures (Figure 4-4).

A differentiated consideration of the two parts of the BFNP showed that stratum 5 clearly dominated the northern part (31.0%) where stratum 3 followed as second (19.5%; Table 4-3). In the southern part stratum 3 and 5 were almost equally dominant with 24.1% and 23.4% respectively. The strata 4 and 7 were approx. proportionally distributed to the area relation of northern to southern part. Stratum 1 and stratum 2 were underrepresented in the North whereas stratum 6 was overrepresented there (Table 4-3).

Table 4-3 Classified area [ha] per stratum

	Stratum							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**total**
Area North [ha]	7.8	298.2	456.5	333.2	727.0	411.6	110.4	2,344.7
Area South [ha]	77.6	1,413.0	1,800.0	1,003.2	1,744.0	1,106.3	315.0	7,459.1
Area BFNP [ha]	85.4	1,711.2	2,256.5	1,336.4	2,471.0	1,517.9	425.4	9,803.8

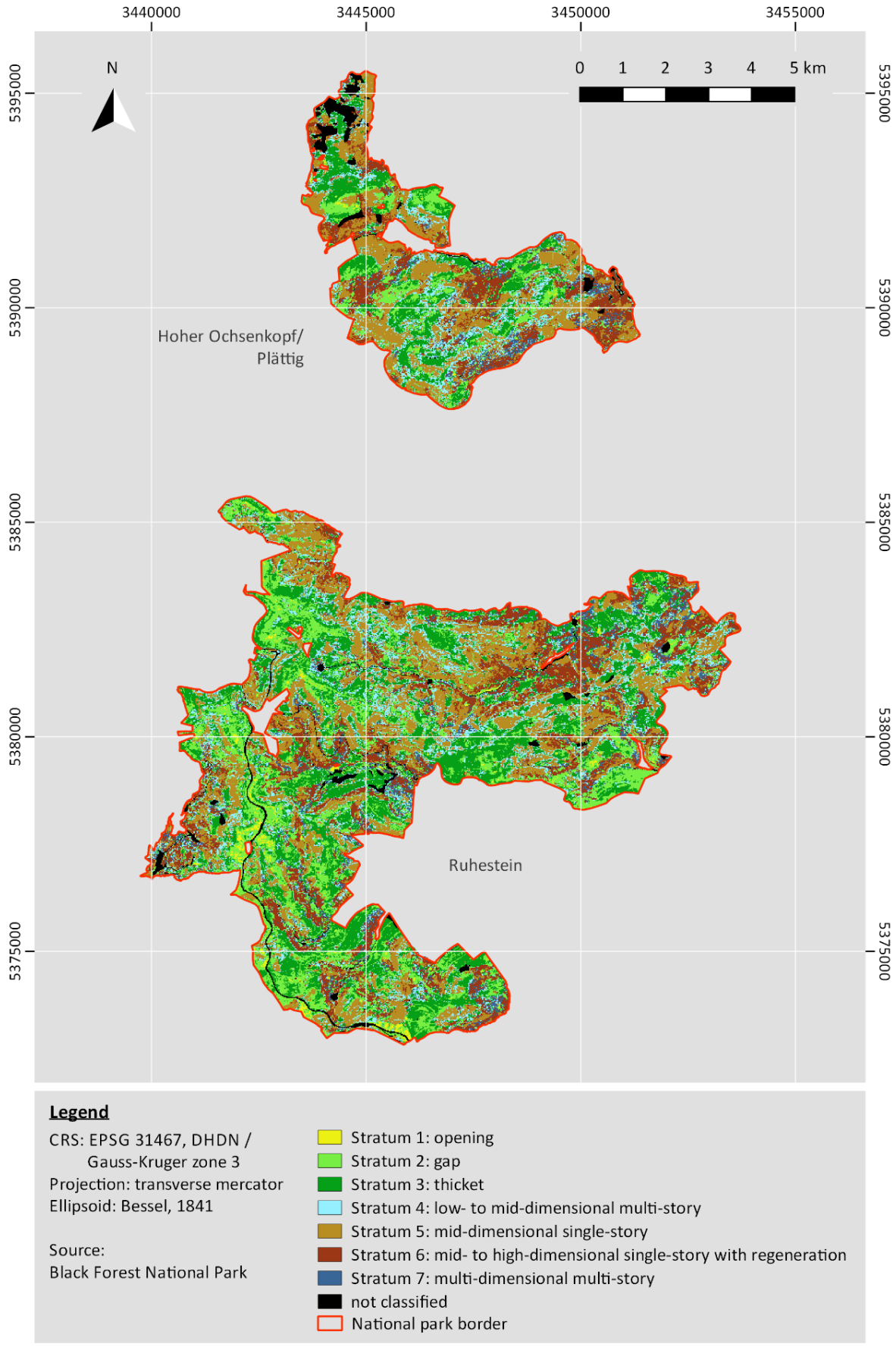

Figure 4-4 Map of forest structure classification

4.5 Stratum membership analysis

The analysis of the stratum membership probability showed divergent results across the strata which were comparable to the confusion matrices of the OOB and test statistics (Figure 4-5, Table 4-1 and Table 4-2). The large majority of classified stratum 1 structures were attributed with high confidence above 80%. The remaining opening classifications based on votes of at least 40% probability which proved the high identification potential of stratum 1. The classification confidence of stratum 2 and 3 only slightly differed from stratum 1 by the occurrence of votes with a probability of less than 40%. With a share of 3.5% and 1.6% they collectively represented only a small area of uncertain attribution (Figure 4-5). Even if the classification of stratum 5 and 6 showed comparable results for the lowest membership probability interval (3.1% and 5.3%), the share of grid cells classified with high confidence above 80% was considerably lower in comparison to the strata 1 to 3. The multi-story strata 4 and 7 were on average classified with the lowest confidence. The proportions of high probability classifications were only 5.8% and 2.0%. The fractions of low probability were 12.6% and 23.6% in contrast (Figure 4-5).

The map of classification confidence (Figure 4-6) underlined that votes with a probability of less than 40% were the minority. They covered only 5.3% of the study area. In addition, they were loosely distributed over the study area. The two intermediate intervals covered almost equal area proportions with 27.9% (40% – 60%) and 27.3% (60% – 80%). Classification votes of high confidence (80% – 100%) were shown on 39.6% of the area which was the largest proportion of the four intervals.

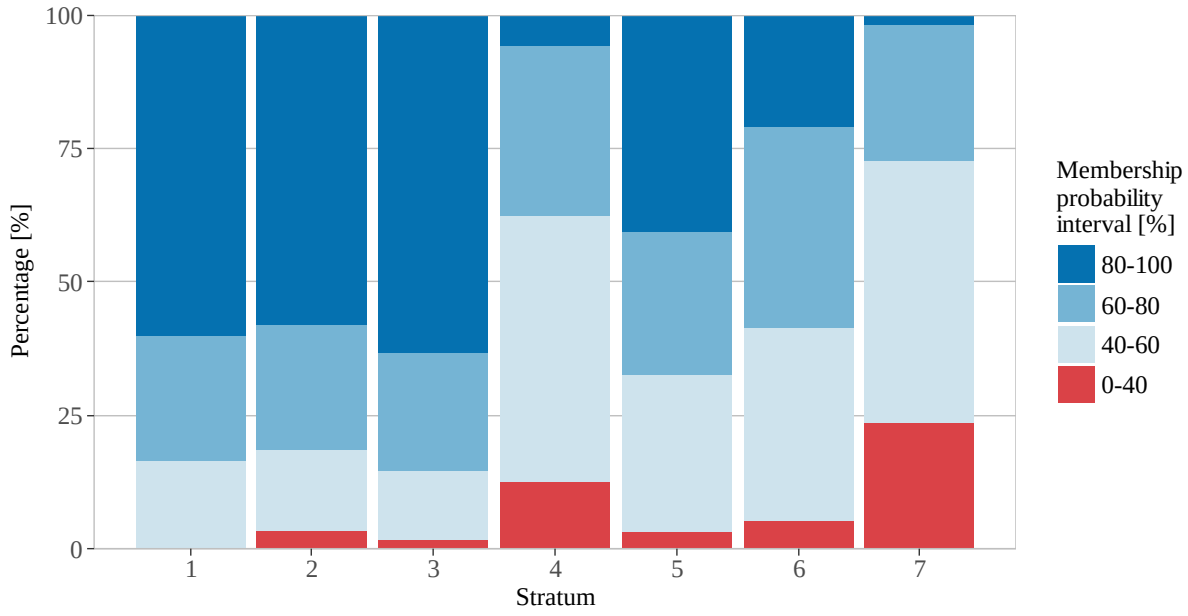

Figure 4-5 Classification confidence per stratum divided in four membership probability intervals

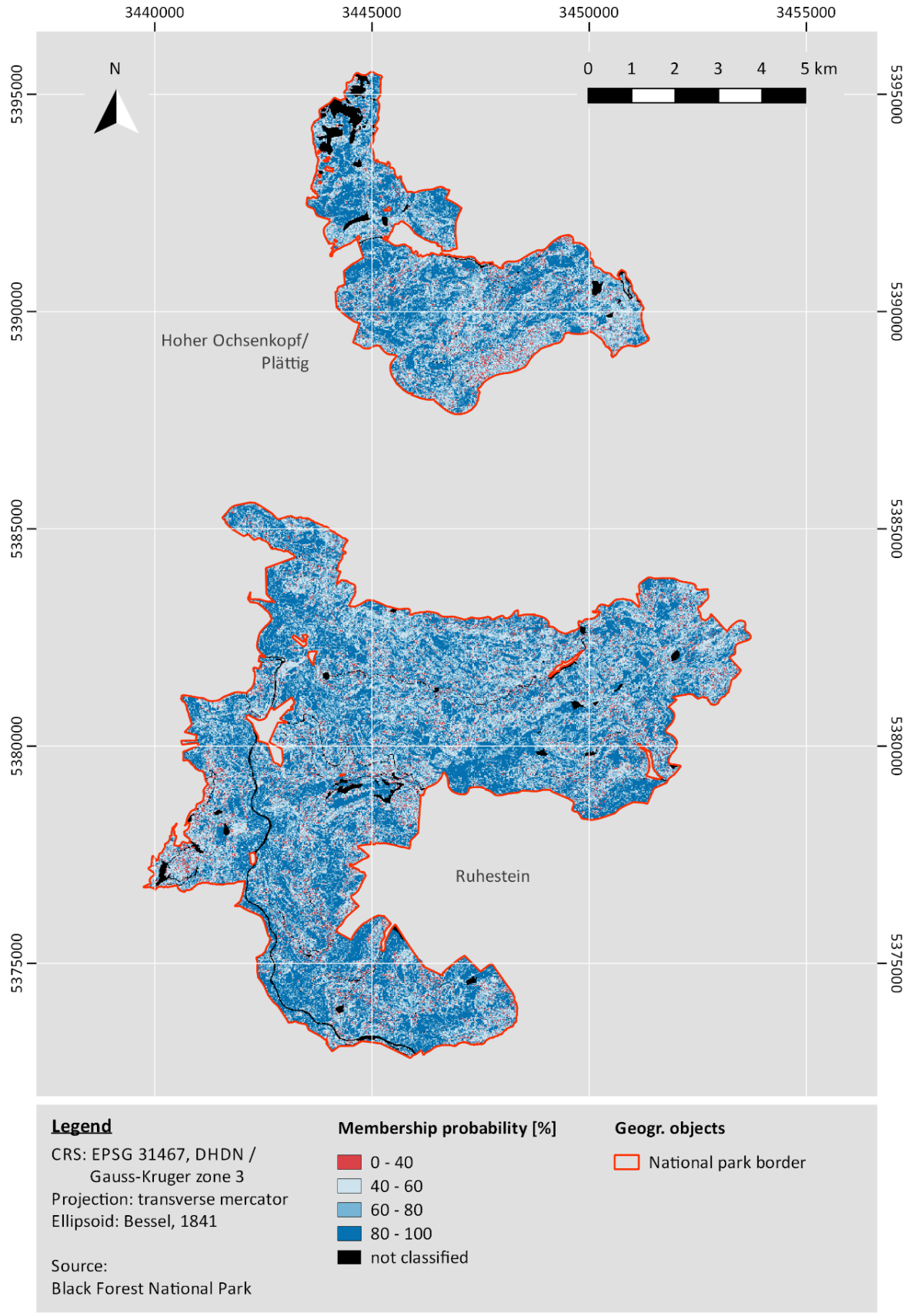

Figure 4-6 Map of classification confidence based on four membership probability intervals

4.6 Differentiation capacity of LiDAR predictors

The relative percentage of returns between 2 m and 10 m of height ($dens.hl_{]2,10]}$) reached the highest variable importance (Figure 4-2). Therefore, it was chosen as example for the density metrics group to be analysed by stratum-wise conditional density plots (CDP; Figure 4-7). The CDP of stratum 3 showed a high differentiation capacity for $dens.hl_{]2,10]}$. In comparison to the remaining strata a density value above 0.5 could be interpreted nearly as a proof for thicket structures. This was ecologically reasonable as the height interval between 2 m and 10 m is only sparsely or even not reached by vegetation of stratum 1 and 2 while the strata 4 to 7 were defined to do not or only dispersedly show branched trees within this height segment (Table 3-1). Stratum 1 revealed almost no considerable occurrence of $dens.hl_{]2,10]}$ which provided identification potential by the absence of dense vegetation. The probability distribution of strata 6, 5, 2 and 4 showed little differences due to slightly increasing maxima between densities of 0 to 0.5. However, maximum probabilities around 0.4 revealed that none of these strata dominated at any density and thus couldn't be clearly identified by $dens.hl_{]2,10]}$. Stratum 7 was poorly characterized by $dens.hl_{]2,10]}$ (Figure 4-7).

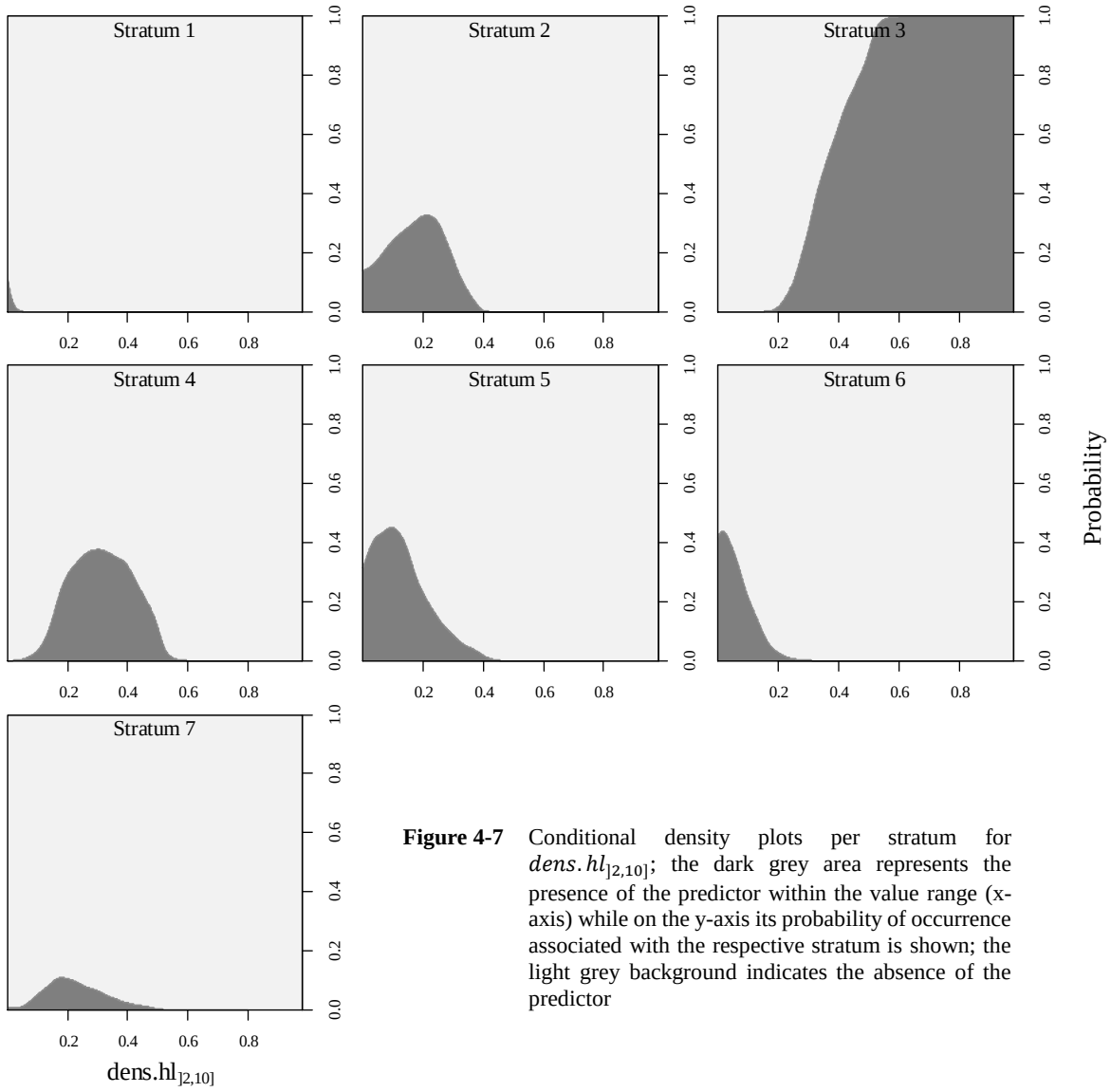

Figure 4-7 Conditional density plots per stratum for $dens.hl_{]2,10]}$; the dark grey area represents the presence of the predictor within the value range (x-axis) while on the y-axis its probability of occurrence associated with the respective stratum is shown; the light grey background indicates the absence of the predictor

Three additional predictors out of the twelve most important LiDAR metrics remaining were chosen to analyse the plausibility of the classification results and the underlying stratification. The highest ranked predictor out of each metric group were sampled to gain a comprehensive assessment of the differentiation capacity between the strata. The second most important representative was the standard deviation of return heights between the 1st and 2nd height quartile (sd_{hq2}; Figure B-1) of the group of sectioned vertical distribution metrics. Median height (h_{med}; Figure B-2) was the highest ranked predictor of the vertical distribution metric group while the number of closed gap voxels (cg; Figure B-3) represented the 3D return distribution metrics. The further descriptions of the following section refer to the mentioned figures without repeated referencing.

The CDP patterns of stratum 1 of the three additional predictors were identical to that of $dens.hl_{]2,10]}$. Recognisable probability values occurred only at the lowest edges of the predictor ranges. The accumulation of zero or near to zero probability values over various predictors supported high differentiation capacity by absence. Gap structures of stratum 2 revealed dominant probability values at the lower edge of sd_{hq2} and h_{med} as well as high probabilities for the lowest values of cg. The predominantly low vegetation of forest gaps (Table 3-1) caused sharp maxima for the standard deviation and median height. The right skewed character of the closed gap distribution resulted from occasionally occurring isolated trees. In combination, the high probabilities at the lower edges of the value ranges indicated a good distinguishability compared to the other strata. Stratum 3 showed three similar probability distributions for the additional predictors. They reached maximum probabilities between approximately 0.75 and 0.85. Their occurrence probability distribution ranged between 0 and approximately the first quartile of each predictor value range. The relatively low vegetation height of thickets (Table 3-1) caused the left oriented maximum. The increase in biomass accumulation from the top of the highest trees to the densely branched bottom led to the slight skewness to the right. In combination with the probability distribution for $dens.hl_{]2,10]}$ the predictor set indicated a very good differentiation capacity for thicket classifications. The probability distributions of stratum 4 revealed only moderate occurrence probabilities over all four predictors with similarities in their CDP characteristics. On the one hand, the broad character of $dens.hl_{]2,10]}$ and sd_{hq2} resulted from the higher intra-stratum variability of multi-story strata with respect to ideally homogeneous biomass distribution over the height range. On the other hand, the sharper character of h_{med} and cg showed intra-stratum similarity. The probability distributions of stratum 4 showed overlapping with other strata and reduced differentiation capacity with respect to the sampled predictors. Stratum 5 revealed moderate probability for low $dens.hl_{]2,10]}$ values. This possibly resulted from partial branching down to heights below 10 m in classified grid cells due to the exclusion of regeneration and understory by its definition (Table 3-1). The forked maximum of the standard deviation probability (sd_{hq2}) underlined the interpretation of partially densely and partially sparsely branched forest structures up to the median height. A third indication for deep branching was the below average orientation of the probability curve of closed gap

voxels. At last the maximum of h_{med} close to the average of the entire medium height range proved accordance with the mid-dimensional aspect of the stratum 5 definition (Table 3-1). The differentiation capacity of the four sample predictors regarding stratum 5 was moderate due to no clear dominance at any interval across the value ranges. Stratum 6 showed moderate probability at the lower edge of the $dens.hl_{]2,10]}$ value range as well. Considering the definition of stratum 6, this was probably caused by regeneration and low understory (Table 3-1). High probabilities in the upper half of the closed gap value range indicated negligible contribution of deep branches at tall trees to vegetation density at lower heights. A maximum probability of the height median around 25 m and general high probabilities along the h_{med} value range above the height median verified the mid- to high-dimensionality of stratum 6 (Table 3-1). The dominant probability values of sd_{hq2}, h_{med} and cg proved high differentiation capacity regarding stratum 6. The multi-storied stratum 7 reached almost no dominant probability across the four sample predictors. Only the sharp peak at the right edge of the cg distribution revealed a narrow interval of specific stratum 7 identification potential. Overall, the CDP characteristics of stratum 7 showed broad overlapping of different intensity with the probability occurrences of the remaining six strata. This verified its ecological definition because multi-storied forest structures and especially those of high dimension were supposed to show quite homogeneous biomass distribution across the entire height range without specific emphasis.

5 Discussion

5.1 Accuracy and performance measures

The classification accuracies revealed performance levels from satisfactory to perfect. This divergence arose on the one hand between the OOB estimates (see chapter 4.3.1) and prediction of test observations (see chapter 4.3.2) and on the other hand among the stratum specific results (Table 4-1 and Table 4-2).

Almost all performance measures of the test set classification exceeded the RF-internal OOB estimations. These differences may be a result of random sampling of the independent test data before entering the remaining data in the RF classification algorithm for the model training (Table 3-6). If mostly well identifiable observations are selected as the 30% withheld for test purpose and more tendentious observations remain in the training pool, this random selection bias may cause deviations. This presumed effect, in turn, has its roots in the necessity to attribute discrete classes to continuous information in supervised classification approaches. Artificially defined class borders inevitably lead to tendencies in classifications (Rocchini et al., 2013). The performance and reliability of OOB error estimators are critically discussed as well (Dougherty et al., 2010). However, they are basically accepted as estimates for classification accuracy due to their bootstrapped cross-validation character (Fu et al., 2005; Kuhn and Johnson, 2013). Therefore, the classification assessment mainly focuses on the more conservative OOB estimates while the independent test results accompany the comprehensive discussion.

The OOB estimates are preferred in the discussion of inter-stratum variability of classification performance following the identical argumentation. But in this context, stratum membership analysis extends the basis for interpretation (Figure 4-5). It splits up the stratum vote probability, which mitigates the drawback of a hard classification of continuous information (Foody, 1999; Rocchini et al., 2013). Thereby, increasing vote probability displays a rising certainty of correct stratum attribution based on the assumption that misclassifications with high probability are exceptions.

5.2 Classification results

5.2.1 Overall assessment

An overall OOB accuracy of 81.47% with an OOB estimated K_{HAT} of 0.78 proves reasonable classification results (Foody, 2002). In comparison to similar studies they can be evaluated as above average without achieving optimum accuracy. Falkowski et al. (2009) reported for their comparable seven-stratum Random Forests classification an OOB overall accuracy of 90.12% while the K_{HAT} was 0.85 across a structurally diverse, mixed-conifer forest in Idaho (USA). However, they could rise the respective parameters to 95.54% and 0.93 by aggregation of two strata and subsequent reduction to a

six-stratum scheme. Valbuena et al. (2016) reached an overall accuracy of 72.6% with a K_{HAT} of 0.66 in an eight-stratum RF classification approach by using a considerably less dense LiDAR dataset with 0.91 returns per square meter in boreal forests in Finland. Based on a single tree approach and a LiDAR return density of 2.8 m^{-2}, Alberti et al. (2013) distinguished six forest strata and one shrub stratum with an accuracy of 68% and a K_{HAT} of 0.58 in mixed-coniferous forests in the Alps.

The compared studies gave sparse information about the reference data collection and the respective sampling scheme which are essential for an accurate benchmark assessment (Congalton, 1991). But by reaching an overall accuracy of 81.47% the classification of this study only barely missed the typical target accuracy of 85% (Foody, 2002). Considering the impact of relatively high confusion rates of the strata 4 and 7 on underrepresented training observations, discussed in the following chapter, this small gap seems to be negligible. Therefore, the present study shows that the applied LiDAR based forest structure classification approach generally performs well with scope for improvement.

5.2.2 Strata assessment

The non-forested and low vegetated stratum 1 was unaffected by misclassifications. Its introduction with the objective of classifying open areas in addition to forested strata didn't lead to confusion with stratum 2, which represents the structurally most adjacent stratum. For this reason, stratum 1 is a useful supplement to cover open areas like meadows, heathland areas or fully open disturbance patches, and thereby to avoid widespread, not classified pixels or pixel groups.

Within the remaining six forested strata three foci of confusion were revealed by the OOB confusion matrix (Table 4-1). Firstly, classification decisions between stratum 2 and stratum 3 showed interacting confusion. Pursuant to the stratification their differentiation depends largely on canopy cover. Respective descriptions of patchy and quite closed regeneration lead to an ecologically reasonable transition zone between both strata (Table 3-1). Any artificially defined, discrete boundary in between becomes unavoidably blurred by natural dynamics (Rocchini et al., 2013), due to the process of successively increasing coverage of woody species on undisturbed or unmanaged forest gap structures (Oliver and Larson, 1996). Against this background the results of the OOB confusion matrix and in addition of the test prediction indicates high differentiation potential between the selected observations of both strata. Marginal membership probabilities up to 40% in combination with most vote probabilities exceeding 80% shows that accurate differentiation between stratum 2 and stratum 3 can be assumed for the prediction across the BFNP as well (Figure 4-5). This interpretation is consistent with results of Falkowski et al. (2009) and Valbuena et al. (2016), both reporting low confusion for partially open or dense, low vegetated forest structures. In summary, the classification of stratum 2 and stratum 3 can be evaluated as very accurate.

A second pairwise confusion of moderate severity occurred between stratum 5 and stratum 6, which are structurally similar strata primarily differentiated by the coverage and height of regeneration or

understory (Table 3-1). One reason for the confusion may again be the continuous transition between both strata (Rocchini et al., 2013). The natural forest development would lead from stratum 5 to stratum 6 in the sense of a successional concept (Oliver and Larson, 1996). A second reason for confusion might be the partially high density of the dominant canopy which considerably reduces the return numbers below. This may result in insufficient LiDAR information about the regeneration and understory decreasing the differentiation capacity of specific predictors (Goodwin et al., 2007; Latifi et al., 2016). The membership analysis yielded again small proportions of stratum voting probabilities below 40% although the shares of classification decisions based on more than 80% probability were reduced, especially for stratum 6 (Figure 4-5). This is interpreted an increasing vote uncertainty and subsequently rising confusion potential. But most of the observations were still classified by clear votes with probabilities above 60%. Falkowski et al. (2009) reported more severe classification confusion between their strata closed stem exclusion (comparable to stratum 5) and mature multi-story (comparable to stratum 6). In this context, the differentiation of stratum 5 and stratum 6 proved as successful and their classification results can be summarized as accurate.

The most critical confusion occurred between the multi-storied strata 4 and 7. Due to their small observation numbers in the training set the estimated OOB accuracies are less robust to misclassifications. With two observations per stratum in the test set these accuracies are even not representative (Congalton and Green, 2009; Foody, 2009). Therefore, their evaluation is of qualitative character and is focused to the description of tendencies. Both multi-storied strata showed ecologically plausible confusion with structurally related strata or between each other. Stratum 4 observations erroneously attributed to stratum 3 or stratum 5 reflect tendencies in biomass accumulation whereas multi-story strata are supposed to show homogeneous biomass distribution by definition (Table 3-1). Stratum 7 confusion with stratum 6 indicates weak representation of mid-dimensional trees. Accordingly, both multi-storied strata showed the highest proportion of the lowest membership probability interval. Stratum decisions with probabilities above 80% were the small minority (Figure 4-5). Therefore, difficulties by the classification of multi-storied strata are proven. This is in line with recent results of Valbuena et al. (2016). They reported high confusion of a multi-layered regeneration stand stratum which is comparable to stratum 4. Falkowski et al. (2009) even aggregated a multi-story and a single-story stratum to reduce the high rate of misclassification of their stratum mature multi-story. Under consideration of the compared studies and the described deficiencies in the present study, the classification of stratum 4 and stratum 7 can be evaluated as satisfactory.

In conclusion, the levels of the UAs and PAs are on average high for strata of representative observation numbers. All misclassifications occurred between strata of plausible structural transitions. The stratum-wise membership results are in line with the results of the confusion matrices, especially those of the OOB estimated matrix. The commonly recommended request of relatively even levels of accuracy for all strata was widely attained (Foody, 2002). Again, the confusion between stratum 4 and stratum 7 led

to missing this target. Therefore, the per stratum assessment confirms accurate classification results with constraints for multi-story strata.

5.2.3 Spatial distribution

Classification maps of forest structures published in the expert report on the foundation of the BFNP (PricewaterhouseCoopers & ö:konzept, 2013) offer a basis for large-scale spatial comparison of the presented classification result (Figure 4-4 and Table 4-3). As they are available only as geographically non-referenced graphics, the further evaluations base on thorough visual interpretations. The expertise maps were created based on a digital surface model with a resolution of 1 m^2 derived from stereoscopic aerial images of the year 2009. Although this report referred to slightly different borderlines for potential national park areas, large parts are identical (PricewaterhouseCoopers & ö:konzept, 2013). Considering the different methodologies, stratifications and the spatial inaccuracy the following comparisons should not be understood as referencing. However, the interpretation gives indications about patterns and classified area proportions which are reasonably to discuss.

The combination of stratum 5 and stratum 6 revealed large coherent areas which are fairly congruent with the comparable units beginning optimal phase and optimal phase of the expert report (PricewaterhouseCoopers & ö:konzept, 2013). In addition, the covered area of the strata 5 and 6 summarized to a share of 40.7% in the present study whereas the comparative strata covered 35.1%. Both underlines the importance of these strata within the present forest structure stratification.

Stratum 2 and Stratum 3 built the second group of coherent structures. Besides the heathland zones a lot of wind disturbance areas, mostly created by the hurricane Lothar at the end of the year 1999, were classified as gaps and thickets in the present study. Stratum 2 and stratum 3 show congruence with the summarily comparable strata opening, beginning regeneration, thicket and beginning pole wood in the expertise. In the present study both strata covered 40.0% of the BFNP whereas the four compared strata comprised 34.0% with respect to the surveyed area of the expert report (PricewaterhouseCoopers & ö:konzept, 2013). Therefore, these low vegetated strata represent the second very important group of forest structures.

Grid cells attributed to the remaining strata of the present study were mostly loosely spread across the BFNP as well as the remaining strata of the expert report. Thus, further detailed spatial comparison was impossible and only the character of distribution can be in general described as similar. However, comparisons of stratum 1 grid-cells and a high-resolution aerial image of the entire study area proves their classification exclusively to patches free of woody vegetation. In addition, a proportional area of less than 1% confirmes the intention of introducing as a filling stratum for low vegetated areas avoiding competition with the forested strata. The detailed visual analysis of the multi-story strata provided some clear, well interpretable patterns. Stratum 7 occurred frequently next to stratum 6 which is reasonable due to the at least partially high dimensions of comprised trees in both strata. The area proportion of

stratum 7 was 4.3% whereas comparable structures of the expert report (selection phase and old-growth forests) summarized to 4.5%. Finally, grid cells attributed as stratum 4 frequently linked strata of lower and higher vegetation height or they traced borders of higher strata to forest roads and open areas. These occurrences are reasonable with respect to possible forest structures at transition zones between height differing forest stands as well as at forest edges (Oliver and Larson, 1996).

In conclusion, the obtained classification seems to be consistent, although the comparative analysis based only on visual comparisons of structurally similar, but not identical strata. The qualitative spatial analysis yields reasonable distributions of all strata whereas quantitative comparisons of covered areas per stratum show classifications of adequate proportions.

5.3 Classification post-processing

Loosely spread pixels in large-scale remote sensing classification maps are commonly interpreted as noise or salt-and-pepper effect disturbing coherent area mapping. In these cases, moving window filters are mostly used to replace the noise pixels by the majority value of the neighbouring pixels (Lillesand et al., 2015). In the present study, no filtering was applied due to the following argumentation. A small-scale classification grid with a cell size of 20 m x 20 m was determined to classify mosaic forest patterns. Those will gain in importance with progressive dissolution of large, coherent areas of homogeneous forest structure by gap dynamic processes under protection of natural forest development (Oliver and Larson, 1996; Scherzinger, 1997; Brumelis et al., 2011). Shugart et al. (2010) stated that grid-cell sizes of 0.1 ha or less are optimal to display forest dynamics seeking nonequilibrium which is the central objective of this study and the future forest structure monitoring built upon. If a moving window filter would have been applied, the visible grid resolution would indeed have remained at 20 m x 20 m. However, the effective resolution would have shifted, because of increased area of influence on the attribution. In case of a 3 x 3 or 5 x 5 window filter the extent determining the stratum of one grid-cell would get increased to 60 m x 60 m or 100 m x 100 m. However, cell sizes of approx. 1 ha and above are assigned to be suitable for capturing parameters of forest dynamics in equilibrium whereas those between 0.1 ha and 1 ha are not recommended due to their intermediate size (Shugart et al., 2010). Therefore, the application of a moving window filters contradicts the mapping of mosaic forest structural development and should be refrained from further forest structure monitoring approaches.

In the post-processing deciduous dominated forest stands and specific non-forested sites were excluded (Figure 4-4). The latter comprised for example main roads whereas forest roads were not masked after careful consideration. From a formal point of view forest roads are unvegetated areas and had to be excluded from the classification. On the other hand, forest roads do not substantially differ from open vegetated areas like meadows with respect to their vertical and horizontal LiDAR return distribution. Therefore, entire forest road pixels were expected to be classified as openings. However, in most of the cases grid-cells would have captured border situations with forests. Depending on the impact of the tree

structures these pixels are expected to be probably classified as gaps or low- to mid-dimensional multi-story. In rare cases, other classification would be conceivable as well. All these possible structures occur at forest edges to meadows as well, which were regularly classified. In combination with the planned closing of approx. two thirds of forest roads by the BFNP administration (personal communication Birk, S., 2017), the grid-cells that are forest road affected aren't finally masked in the classification. Allowing natural dynamics to reshape these open sites will lead to regrowth of forests in the future. A comprehensive forest structure monitoring should take these developments into account. In addition, the regrowth on forest roads under natural dynamics might possibly be a specific field of interest with respect to renaturation of intensively managed forests.

5.4 Forest structure stratification scheme

The summary of chapter 5.2.1 and 5.2.2 indicates that the presented stratification scheme is generally a reasonable basis for the classification of the forest structures in the BFNP. The low vegetated strata 1 to 3 were predicted with high accuracies and classified by mostly clear votes across the study area. Their ecological differentiation proved as reasonable for classification purpose and therefore adjustments in stratum definitions are not indicated and recommended.

The strata 5 and 6 showed slightly lower classification accuracies and on average reduced vote probability. A rise in stratum specification to refine the distinguishability of both strata don't seem to be achievable due to the clear ecological differentiation criterion of absence or presence of regeneration and understory and due to the sufficient number of well-labelled observations. Merging both strata to reduce classification errors as conducted with comparable strata by Falkowski et al. (2009) seems to be not recommendable. On the one hand, the confusion between stratum 5 and stratum 6 was less severe in the present study than reported by Falkowski et al. (2009) for structurally comparable strata. Furthermore, keeping them distinct preserves stratum 6, the only stratum with quite dense regeneration or understory under a dominating canopy, which is an important factor in the ecological assessment of forested ecosystems (Martinuzzi et al., 2009; Wing et al., 2012). In conclusion, the error rates can be evaluated as acceptable in comparison to the potential loss of information resulting from stratum aggregation. Continuation of separated strata 5 and 6 is clearly recommended.

Considerably reduced performance was shown by the classification the multi-story strata 4 and 7 under the restriction of low observation numbers. Therefore, the aggregation of both multi-story strata might be a possibility to overcome the limitation of underrepresentation and to potentially rise the classification accuracy. On the other hand, multi-storied forest structures are a key indicator for forest naturalness (Oliver and Larson, 1996; Brumelis et al., 2011). At the time of the introduction of a monitoring system a generalization by the aggregation of strata would neglect potential forest structure changes under natural dynamics, because the dissolution of silvicultural shaped forests will increase inhomogeneous, multi-storied structures of different scales in the BFNP (Oliver and Larson, 1996; Pricewaterhouse

Coopers & ö:konzept, 2013). Thus, separated strata covering different stages of multi-storied forest structures are supposed to be essential for forest structure monitoring. For this reason, future enhancement in identification and acquisition of additional reference plots is recommended to broaden the basis for model training and to subsequently refine the multi-story characteristic within the classification process, instead of stratum aggregation (Congalton and Green, 2009; Foody, 2009). Additionally, further discriminatory information could resolve interclass confusion of multi-storied strata. The development of specific multi-story LiDAR metrics and the involvement of additional remote sensing data might be possible approaches to increase the differentiation capacity.

At present the forest structures of the BFNP widely represent relatively young forests (Pricewaterhouse Coopers & ö:konzept, 2013). An increase of tree dimensions with respect to DBH, crown diameter, and height are to be expected under protection of natural dynamics (Oliver and Larson, 1996; Roloff, 2010). Simultaneously senescence and decomposition will rise their impact with the potential of creating multi-story structures (Oliver and Larson, 1996). An expansion of the stratification scheme (Table 3-1) by an additional stratum accounting for multi-story forest structures including trees of very large dimensions will be required under these circumstances. This is only one example of possible future changes affecting the stratification scheme. Indeed, at least the strata 4 to 7 will be influenced by an enlarged tree dimension range. Furthermore, new strata may become reasonable or necessary to cover future forest structures. Therefore, the suitability of the stratification must be reviewed periodically to make timely adjustments.

On the basis of a crisp stratification scheme, the recognised difficulties in stratifying forest structures and recorded confusions were expectable considering the fundamental challenge to divide gradual variability of information into discrete classes of a hard classification (Rocchini et al., 2013). Soft or fuzzy classification approaches were discussed as an alternative to cope with the continuity of nature and remote sensing information (Wang, 1990; Foody, 1999; Xie et al., 2008). First practical applications were demonstrated (Benz et al., 2004) but besides the recognition of their potential and occasional ecological applications the deployment of this alternative is still limited (Li et al., 2017). One application in the field of supervised forest classification was reported by Zhang et al. (2004). Compared to a hard classification, they gained higher accuracies with fuzzy habitat classifications and linear discriminant analysis. But their results can only be an indication for potential applicability in the forest monitoring of the BFNP due to the large methodological differences.

In conclusion, the presented stratification scheme divides the current forest structure of the BFNP in six meaningful forested strata accompanied by an open stratum. The study has proven that it is a reliable, initial basis for ecological long-term monitoring which is consistent with the results of Falkowski et al. (2009). In addition, the consideration of membership probabilities mitigated the drawbacks of hard boundaries considering the continuity of natural structures. By this an applicable and well-balanced approach for forest stratification is presented.

5.5 LiDAR predictors

5.5.1 Predictor selection and correlation

The recursive feature elimination achieved a considerable predictor reduction from 58 to 13 predictors (Figure 4-1). In accordance with the well comparable study of Falkowski et al. (2009) the ranking of predictor importance was dominated by density and height related metrics (Figure 4-2). However, the density metrics in the present study focused on low vegetation heights whereas overall canopy density yielded the highest variable importance in Falkowski et al. (2009). Mean and median height were listed in both studies as important predictors. LiDAR metrics characterising the vertical return distribution were only ranked in the present study.

The higher variability of predictors in the present study is probably caused by the selection of 13 metrics in comparison to 8 and 5 predictors for the respective seven- and six-stratum scheme of Falkowski et al. (2009). The effective importance of all 13 predictors for the classification model can be at least questioned considering their pairwise correlation matrix. Especially the height related metrics showed very strong correlations with correlation coefficients above 0.90 (Figure 4-3). Consequently, further predictor reduction might be possible without considerable decrease in classification accuracy. This assumption is additionally supported by the ROC values of RF models using less predictors. The ROC parameter would only marginally decrease for down to nine predictors (Figure 4-1). Even though strongly correlated predictors can contribute additional information for the classification (Guyon and Elisseeff, 2003) predictor redundancy can negatively affect the classification accuracy of RF (Gregorutti et al., 2017). Therefore, intensified model assessment may lead to an improved balance between predictor selection and classification accuracy.

In summary, RFE reduced the predictors but considerable pairwise correlations remained. This provides potential for further improvement with respect to model parsimony, for example by combining recursive and non-recursive feature elimination strategies to identify and eliminate irrelevant variables more precise (Gregorutti et al., 2017). However, the model results of this study can be evaluated as quite robust due to the principal reduction of correlated predictors by RFE and the basic stability of the model due to bootstrapping in RF (Gregorutti et al., 2017).

5.5.2 Differentiation capacity

The CDP analysis results demonstrate that a predictor set is required to differentiate the strata as no single predictor showed distinct CDP characteristics over all strata. Indeed, the predictor $dens.hl_{]2,10]}$ did reach the highest variable importance but this results mainly from its crucial importance for identifying stratum 3 (Figure 4-7). In contrast, its contribution to the classification of the remaining forested strata can be interpreted as low. The set of four predictor CDPs collectively yielded further indications about the inter-stratum confusion of the final model. Especially the already mentioned weak differentiation capacity regarding strata 4 and 7 is underlined by considerable overlaps between CDP

distributions of different strata. High differentiation capacities of the predictor set can be attributed in the cases of the remaining five strata.

Although nine predictor CDPs have not been analysed the sampling of four predictors of different metric groups under consideration of high pairwise predictor correlations suggest basic representativeness. Therefore, acquired results and interpretations based on this sample should be transferable to final model assessment. However, the analysis of the entire list of important predictors might give additional information about potential predictor redundancies after completion of algorithmic selection methods. In that way, CDP analysis may offer an additional approach to examine the simplification potential in terms of further predictor reduction in the final model.

5.6 Field observations

5.6.1 Sampling design

The 245 plots for field data acquisition representing the forested strata 2 to 7 were selected by a stratified sampling design (see chapter 3.4). The sampling based on most recent forest inventory and forest management data, various environmental and geographic factors, and LiDAR return projections of potential areas. The pre-utilization of LiDAR return projections proved as necessary due to the requirement of detecting structurally appropriate plots within reasonable time. Alternative forest management data turned out to be too coarse and vague for precise stratification of forest structure areas. Consequently, this first sampling phase was stratified but with a purposeful character instead of a random one as recommended by Stehman and Czaplewski (1998). The following reduction to 96 well-labelled observations based on field impressions and field data (see chapter 3.7.2). By this an additional non-random aspect was introduced. Thus, the applied sampling approach evolved almost to a non-probability sampling (Stehman and Czaplewski, 1998; Stehman, 2009). Consequently, the forest structure classification is not erroneously in itself, but the classification accuracy, computed by the OOB testing or revealed by the prediction of the test observations, cannot be transferred to the final prediction without restrictions (Stehman and Czaplewski, 1998; Stehman, 2009).

In summary, the sampling design deviates from the good practice guidelines for unrestricted accuracy assessment by confusion statistics. This deficiency could be mitigated by interpreting the stratum membership probabilities as independent and soft performance indicator (Foody, 1999). Similarities between both evaluations basically indicate an observation selection of reasonable representativeness.

5.6.2 Stratum labelling in the field

The stratum labelling in the field can be a source of classification errors and accuracy bias (Foody, 2002; Rocchini et al., 2013). Although forest structure descriptions (Table 3-1) and silhouettes (Figure 3-2) set an operating framework for identification, the attribution in the field remained challenging. The difficulty in stratum labelling of field observations was reflected by the reduction from 245 potential

plots to 96 well-labelled and finally processed plots. This implies, 149 field observations showed strong tendencies or were even not attributable to any stratum due to unclear stratum characteristic. One method to possibly rise the field attribution rate and foster stratum decisions might be the definition of stratum boundaries for further parameters which are simply to measure or reasonably to estimate in the field. In the present study, such reference values for comparatively hard stratum differentiation are given for the main tree height and the main DBH ranges whereas soft indicators describe vegetation community, structural elements and coverage of tree species. Canopy closure or coverage are characterized for example as patchy, open, closed or homogeneous which leave a wide scope for interpretation. They could alternatively be divided in proportional closure or coverage intervals (Congalton and Green, 2009). However, numerical estimation of parameters in the field is still subjective and hence the labelling of the strata remains sensitive to human interpretation (Congalton and Green, 2009). In addition, an increase in numerical stratum criteria can also lead to conflicts which in turn may inhibit clear stratum attribution.

All 245 field observations of the forested strata 2 to 7 are integrated in the comprehensive monitoring system of the BFNP (Gärtner et al., 2016). Thereby, additional forest structure parameters will be assessed on the plots within the years 2017 and 2019. It seems to be recommendable to analyse the forest structure field data at the end of the first monitoring period to identify possible key measures differentiating the strata. They may improve stratum membership decisions in case of expanding the pool of field observations or in reassessments of existing plots.

5.7 Study area

The limitation of the study area to conifer dominated forests was a consequence of underrepresentation of deciduous dominated forests and resulting difficulties in finding sufficient observations across all strata. Therefore, some areas haven't been classified, predominantly located in the northern part. In addition, focussing on conifer dominated forests is a source for uncertainty of classified grid cells as well. Whereas the exclusion of deciduous dominant to equally mixed forests based on the scale of forest stands field observations indicated that a noticeable share of scattered grid cells might be dominated by groups of deciduous trees within basically coniferous forest stands. This fuzziness is expected to increase in the future due to recently conducted, planned and legally required forest management fostering admixture species according to the natural tree species composition of the Black Forest (Landtag von Baden-Württemberg, 2013; PricewaterhouseCoopers & ö:konzept, 2013; Nationalpark Schwarzwald, 2016). Besides silver fir (*Abies alba* MILL.), beech (*Fagus sylvatica* L.) is predestined to an increase in proportion as it is the characterising deciduous tree of the natural tree species composition (PricewaterhouseCoopers & ö:konzept, 2013). Natural dynamics can possibly reinforce this development in a more distant future (PricewaterhouseCoopers & ö:konzept, 2013). At least from a medium-term perspective, e.g. 15 to 30 years, it seems necessary to expand the remotely sensed forest structure monitoring to deciduous and mixed forest areas.

As a first step to the integration of deciduous forest structures a comparison can be conducted between the classified but subsequently masked, deciduous to equally mixed forest stands and the respective forest structures in the field. A spatially referenced and detailed analysis would provide insights about the applicability of the presented, on coniferous forests calibrated model to these untrained structures. If the results are positive, the current classification model might be operated as an interim approach to classify the entire BFNP. Secondly, the model should get retrained on the entire range of structures as soon as sufficient observations for one or more strata get identified in deciduous dominated areas and corresponding new LiDAR data is available. Following, results of precursor models can be reclassified on the old datasets to ensure consistent comparisons between the classifications at different times. Inclusion of deciduous training observations will finally lead to a robust classification model for a long-term monitoring system comprising the complete range of forest structure in the BFNP.

6 Conclusion and outlook

Based on a high-resolution data set of the BFNP, the present study aimed to establish a comprehensive stratification of forest structures and to accurately classify these strata across the BFNP by considering a minimum set of relevant LiDAR predictors.

The presented stratification scheme proved substantially appropriate and achieved a balanced consideration of recent structures and potential structural changes. Confusion between the forested strata was ecologically reasonable and mostly of moderate intensity. Only the multi-storied strata require enhancement in their differentiation capacity as aggregation of stratum 4 and 7 is not recommended to reduce their level of confusion. Additional training observations and multi-story specific LiDAR predictors may lead to considerable improvement. In addition, the stratification should not be considered concluded. Forest structures establishing by natural dynamics are supposed to lead to the introduction of new strata. Furthermore, technological progress may refine structural identification and enable splits of presented strata. Finally, the extension of the stratification scheme to deciduous forest structures should certainly be sought to cover all types of forest composition in future.

The final classification model revealed highest importance of density and height related metrics. However, high pairwise correlations between numerous final predictors indicated remaining redundancies and this stands in conflict with the objective of high model parsimony achievement. Assessments of CDPs confirmed the assumption of redundant predictors. Therefore, adjustments of the applied recursive feature elimination and further selection approaches can be tested to reduce the predictors to an essential set. Irrespective of improvable model parsimony, RF based classification approaches proved to be sufficiently robust against correlated predictors. Thus, reliable classification results of the presented model can be assumed.

The applied classification based on hard stratification softened subsequently by membership probabilities mitigated drawbacks of solely crisp classifications. Overall OOB accuracy of 81.47% confirmed the general applicability on the suggested stratification. Accuracy assessments per stratum and concurring stratum membership results revealed divergence between low vegetated, multi-story and the remaining strata. However, even the lowest accuracies of the multi-story strata 4 and 7 did neither fundamentally question the stratification nor the classification approach itself due to indication of sufficient performance. With respect to the future reliability of forest structure monitoring, enhancements regarding multi-storied strata are recommended as mentioned above.

The study presented an overall effective and comprehensive forest structure classification approach to initiate and pursue a long-term monitoring in the BFNP. As an optional complement to the area-based approach individual tree segmentation can be processed, fully exploiting the potential of LiDAR analysis (Heurich, 2008; Reitberger et al., 2009; Latifi et al., 2015). Further information may also be

gained by combining LiDAR and optical imagery data. However, scientific studies showed divergent results (Dalponte et al., 2008; Latifi et al., 2012; Buddenbaum et al., 2013).

In addition to the monitoring purpose the classified forest structure map might be useful for a suite of different applications in nature conservation practice and research. According to the definition of a monitoring, change detection can be processed, beginning from the second ALS campaign in the BFNP. Although change detection of forest structure based on LiDAR data has theoretically been mentioned in scientific literature (Dees et al., 2012; Wulder et al., 2014) applications are still rare (Frew et al., 2016; Song et al., 2016). Especially change detection by comparison of multi-temporal area-based LiDAR classifications cannot be found. Therefore, it seems to be recommendable to transfer general good practice principles and methods of remote sensing change detection to the forest structure monitoring system of the BFNP (Olofsson et al., 2014; Willis, 2015).

Another direction of pursuing research are habitat analyses or species distribution models with respect to forest structural classifications or structural metrics (Guo et al., 2017). Habitat preferences of GPS-collared animals can be analysed along different temporal scales from diurnal over seasonal to multi-annual periods by combining GPS-telemetry data and forest structure classification (Melin et al., 2013; Ewald et al., 2014; Melin et al., 2016). In this context, a first study about the temporal and spatial habitat selection of red deer (*Cervus elaphus* L.) was conducted in the BFNP based on the presented classification (Rauscher, 2017). The assessment of potential habitats of avian species based on LiDAR analysis and bird occurrence data is of high potential due to the broad basis of documented bird observations (Clawges et al., 2008; Bergen et al., 2009). Specifically, bird species indicative for diverse forest structures may attract the focus with respect to structural changes under natural forest dynamics (Martinuzzi et al., 2009; Müller et al., 2009; Zellweger et al., 2013). A first habitat model with focus on pygmy owl (*Glaucidium passerinum* L.) and its forest structure requirements was presented by Holderried (2016) for the BFNP. Comparable habitat modelling could be applied to further indicator species, e.g. capercaillie (*Tetrao urogallus* L.) and forest beetle assemblages (Müller and Brandl, 2009). However, it seems to be recommendable to use standardized forest structure classification to keep resulting models spatially comparable. Against the background of the floristic, faunistic and fungal species monitoring, initiated in 2017 (Buse et al., 2016), a general species diversity model depending on forest structure might be a comprehensive approach for further research based on the presented forest structure analysis (Bergen et al., 2009).

LiDAR remote sensing offers the opportunity to model above-ground biomass in forest ecosystems in general (Zolkos et al., 2013) and in conifer dominated forests in particular (Næsset et al., 2013b). For this purpose, biomass data can be subsequently collected at the field observation plots to extend the current database. Hence, a multi-temporal biomass monitoring could be established according to the presented classification to budget changes under natural forest development in the BFNP based on repeated LiDAR flight campaigns (Næsset et al., 2013a). Furthermore, biomass modelling and forest

structure classifications across the BFNP can build the basis for fuel load quantifications and fire risk simulations (Skowronski et al., 2007; González-Olabarria et al., 2012).

In abiotic fields of research, the forest structure analyses may contribute to hydrologic modelling in the BFNP. Forest structure considerably influences snow accumulation and ablation (Varhola et al., 2010; Broxton et al., 2015) and is coupled to the evapotranspiration (Mitchell et al., 2012; Ringgaard et al., 2012; Tang et al., 2017). The presented classification or at least the underlying forest structure metrics can support hydrological or micro-meteorological modelling approaches at different scales in the BFNP.

In conclusion, the introduced forest structure classification fulfils not only the main objective of initiating a forest structure monitoring. It opens a broad field of further application in scientific research and practical nature conservation. Therefore, it facilitates advancement in documentation, planning, management, and assessment of the status and future development of forested ecosystems in the BFNP.

References

Alberti, G., Boscutti, F., Pirotti, F., Bertacco, C., Simon, G. de, Sigura, M., Cazorzi, F., & Bonfanti, P. (2013). A LiDAR-based approach for a multi-purpose characterization of Alpine forests. An Italian case study. *iForest - Biogeosciences and Forestry*, *6*, 156–168.

Ambroise, C., & McLachlan, G.J. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 6562–6566.

Bässler, C., Müller, J., Dziock, F., & Brandl, R. (2010). Effects of resource availability and climate on the diversity of wood-decaying fungi. *Journal of Ecology*, *98*, 822–832.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing. A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, *58*, 239–258.

Bergen, K.M., Goetz, S.J., Dubayah, R.O., Henebry, G.M., Hunsaker, C.T., Imhoff, M.L., Nelson, R.F., Parker, G.G., & Radeloff, V.C. (2009). Remote sensing of vegetation 3-D structure for biodiversity and habitat. Review and implications for lidar and radar spaceborne missions. *Journal of Geophysical Research: Biogeosciences, 114 (G2)*.

Berglund, H., Edman, M., & Ericson, L. (2005). Temporal variation of wood-fungi diversity in boreal old-growth forests: Implications for monitoring. *Ecological Applications*, *15*, 970–982.

Birk, S. (2017). Aktueller Stand zum Wegekonzept im Nationalpark Schwarzwald. personal communication. Bad Peterstal-Griesbach.

BMU (2007). National Strategy on Biological Diversity. Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU). https://biologischevielfalt.bfn.de/fileadmin/NBS/documents/broschuere_biolog_vielfalt_2015_strategie_bf.pdf. 05.06.2017.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and regression trees.* Belmont: Wadsworth International Group.

Broxton, P.D., Harpold, A.A., Biederman, J.A., Troch, P.A., Molotch, N.P., & Brooks, P.D. (2015). Quantifying the effects of vegetation structure on snow accumulation and ablation in mixed-conifer forests. *Ecohydrology*, *8*, 1073–1094.

Brumelis, G., Jonsson, B., Kouki, J., Kuuluvainen, T., & Shorohova, E. (2011). Forest naturalness in northern Europe. Perspectives on processes, structures and species diversity. *Silva Fennica, 45*.

Buddenbaum, H., St. Seeling, & Hill, J. (2013). Fusion of full-waveform lidar and imaging spectroscopy remote sensing data for the characterization of forest stands. *International Journal of Remote Sensing, 34*, 4511–4524.

Bundesamt für Naturschutz (2016). Nationalparke. https://www.bfn.de/0308_nlp+M54a708de802.html. 26.04.2017.

Buse, J., Gärtner, S., Popa, F., & Förschler, M. (2016). How many species can be expected in a protected area? - A case study in the Black Forest National Park. Black Forest National Park. Bad Peterstal-Griesbach. https://www.researchgate.net/publication/309210204_How_many_species_can_be_expected_in_a_protected_area_-_A_case_study_in_the_Black_Forest_National_Park. 22.06.2017.

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. http://oz.berkeley.edu/users/chenchao/666.pdf. 14.03.2017.

Clawges, R., Vierling, K., Vierling, L., & Rowell, E. (2008). The use of airborne lidar to assess avian species diversity, density, and occurrence in a pine/aspen forest. *Remote Sensing of Environment, 112*, 2064–2073.

Congalton, R.G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*, 35–46.

Congalton, R.G., & Green, K. (2009). *Assessing the accuracy of remotely sensed data. Principles and practices*. (2nd ed.). Boca Raton [u.a.]: CRC.

Coops, N.C., Hilker, T., Wulder, M.A., St-Onge, B., Newnham, G., Siggins, A., & Trofymow, J.A. (2007). Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees, 21*, 295–310.

Coppin, P.R., & Bauer, M.E. (1996). Digital change detection in forest ecosystems with remote sensing imagery. *Remote Sensing Reviews, 13*, 207–234.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., & Lawler, J.J. (2007). Random Forests for classification in ecology. *Ecology, 88*, 2783–2792.

Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. *IEEE Transactions on Geoscience and Remote Sensing, 46*, 1416–1427.

Dees, M., Straub, C., & Koch, B. (2012). Can biodiversity study benefit from information on the vertical structure of forests? Utility of LiDAR remote sensing. *Current Science, 8*, 1181–1187.

Deutscher Wetterdienst (2017). CDC - Climate Data Center. ftp://ftp-cdc.dwd.de/pub/CDC/. 27.04.2017.

Dicks, S.E., & Lo, T.H.C. (1990). Evaluation of thematic map accuracy in a land-use and land-cover mapping program. *Photogrammetric engineering and remote sensing, 56*, 1247–1252.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., & Lautenbach, S. (2013). Collinearity. A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*, 27–46.

Dougherty, E.R., Sima, C., Hua, J., Hanczar, B., & Braga-Neto, U.M. (2010). Performance of error estimators for classification. *Current Bioinformatics, 5*, 53–67.

Dudley, N. (Ed.) (2013). Guidelines for applying protected area management categories including IUCN WCPA best practice guidance on recognising protected areas and assigning management categories and governance types. Gland [u.a.]: IUCN.

Echtler, H.P., & Chauvet, A. (1991-1992). Carboniferous convergence and subsequent crustal extension in the southern Schwarzwald (SW Germany). *Geodinamica Acta, 1-2*, 37–49.

Ewald, M., Dupke, C., Heurich, M., Müller, J., & Reineking, B. (2014). LiDAR Remote Sensing of Forest Structure and GPS Telemetry Data Provide Insights on Winter Habitat Selection of European Roe Deer. *Forests, 5*, 1374–1390.

Fahrmeir, L., Künstler, R., Pigeot, I., & Tutz, G. (2011). *Statistik. Der Weg zur Datenanalyse*. (7th ed.). Berlin, Heidelberg: Springer.

Falk, M., Marohn, F., & Tewes, B. (2002). *Foundations of Statistical Analyses and Applications with SAS.* Basel, s.l.: Birkhäuser Basel.

Falkowski, M.J., Evans, J.S., Martinuzzi, S., Gessler, P.E., & Hudak, A.T. (2009). Characterizing forest succession with lidar data. An evaluation for the Inland Northwest, USA. *Remote Sensing of Environment, 113*, 946–956.

FAO (2014). *World reference base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps.* Rome: Food and Agriculture Organization of the United Nations.

FAO (2016). *State of the World's Forests 2016. Forests and agriculture: land-use challenges and opportunities.* Rome: Food and Agriculture Organization of the United Nations.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861–874.

Foody, G.M. (1999). The continuum of classification fuzziness in thematic mapping. *Photogrammetric engineering and remote sensing, 65*, 443–451.

Foody, G.M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80*, 185–201.

Foody, G.M. (2009). Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing, 30*, 5273–5291.

Foster, D.R., Knight, D.H., & Franklin, J.F. (1998). Landscape Patterns and Legacies Resulting from Large, Infrequent Forest Disturbances. *Ecosystems, 1*, 497–510.

Frew, M.S., Evans, D.L., Londo, H.A., Cooke, W.H., & Irby, D. (2016). Measuring Douglas-Fir Crown Growth with Multitemporal LiDAR. *Forest Science, 62*, 200–212.

Fu, W.J., Carroll, R.J., & Wang, S. (2005). Estimating misclassification error with small samples via bootstrap cross-validation. *Bioinformatics (Oxford, England), 21*, 1979–1986.

Gadow, K. von (2005). *Forsteinrichtung. Analyse und Entwurf der Waldentwicklung.* Göttingen: Univ.-Verlag Göttingen.

Gärtner, S., Birk, S., Buse, J., del Val Alfaro, E., Dreiser, C., Kratzer, R., Lang, F., Popa, F., & Förschler, M. (2016). Monitoring der natürlichen Waldentwicklung im Nationalpark Schwarzwald. Nationalpark Schwarzwald. Bad Peterstal-Griesbach. https://www.researchgate.net/publication/309463890_Monitoring_der_naturlichen_Waldentwicklung_im_Nationalpark_Schwarzwald. 12.06.2017.

Gärtner, S., Birk, S., & Dreiser, C. (2017). Selection of conifer dominated forest stands from forest management data of the forest administration of Baden-Württemberg (ForstBW). personal communication. Bad Peterstal-Griesbach.

Gebhardt, H. (2008). *Geographie Baden-Württembergs. Raum, Entwicklung, Regionen.* Stuttgart: W. Kohlhamer GmbH.

GFZ Potsdam (2015). IGRF Deklinationsrechner. http://www-app3.gfz-potsdam.de/Declinationcalc/declinationcalc.html. 04.05.2017.

Gislason, P.O., Benediktsson, J.A., & Sveinsson, J.R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters, 27*, 294–300.

Gobakken, T., Bollandsas, O.M., & Næsset, E. (2014). Comparing biophysical forest characteristics estimated from photogrammetric matching of aerial images and airborne laser scanning data. *Scandinavian Journal of Forest Research, 30*, 73–86.

González-Olabarria, J.-R., Rodríguez, F., Fernández-Landa, A., & Mola-Yudego, B. (2012). Mapping fire risk in the Model Forest of Urbión (Spain) based on airborne LiDAR measurements. *Forest Ecology and Management, 282*, 149–156.

Goodwin, N.R., Coops, N.C., Bater, C., & Gergel, S.E. (2007). Assessment of sub-canopy structure in a complex coniferous forest. In *Proceedings of the ISPR Workshop "Laser Scanning 2007 and SilviLaser 2007"* (pp. 169–172). Espoo, Finland.

Graf, R.F., Mathys, L., & Bollmann, K. (2009). Habitat assessment for forest dwelling species using LiDAR remote sensing. Capercaillie in the Alps. *Forest Ecology and Management, 257*, 160–167.

Gregorutti, B., Michel, B., & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing, 27*, 659–678.

Guo, X., Coops, N.C., Tompalski, P., Nielsen, S.E., Bater, C.W., & Stadt, J.J. (2017). Regional mapping of vegetation structure for biodiversity monitoring using airborne lidar data. *Ecological Informatics, 38*, 50–61.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46*, 389–422.

Haglöf Sweden AB (2014). Vertex IV Manual. www.forestrytools.com.au/manuals. 04.05.2017.

Hancock, S., Armston, J., Li, Z., Gaulton, R., Lewis, P., Disney, M., Mark Danson, F., Strahler, A., Schaaf, C., Anderson, K., & Gaston, K.J. (2015). Waveform lidar over vegetation. An evaluation of inversion methods for estimating return energy. *Remote Sensing of Environment, 164*, 208–224.

Heurich, M. (2008). Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. *Forest Ecology and Management, 255*, 2416–2433.

Heurich, M., Krzystek, P., Polakowsky, F., Latifi, H., Krauss, A., & Müller, J. (2015). Erste Waldinventur auf Basis von Lidardaten und digitalen Luftbildern im Nationalpark Bayerischer Wald. *Forstliche Forschungsberichte München*, 101–113.

Heurich, M., Schneider, T., & Kennel, E. (2003). Laser scanning for identification of forest structures in the Bavarian forest national park. In Swedish University of Agricultural Sciences, Department of Forest Resource Management and Geomatics (Ed.), *Proceedings of the Scandlaser Scientific* (pp. 98–107). Umea, Sweden.

Heurich, M., & Thoma, F. (2008). Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (Fagus sylvatica) and Norway spruce (Picea abies) forests. *Forestry, 81*, 645–661.

Holderried, P. (2016). Using LiDAR derived vegetation metrics for multiscale species distribution mapping of Eurasian Pygmy owl (Glaucidium passerinum) in the Black Forest National Park. Masterthesis. Faculty of Biological Sciences. University of Leeds. Leeds, UK.

Hollhaus, M., Mücke, W., Roncat, A., Pfeifer, N., & Briese, C. (2014). Full-waveform airborne laser scanning systems and their possibilities in forest applications. In M. Maltamo, E. Næsset, & J. Vauhkonen (Eds.), *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies* (pp. 43–62). Dordrecht: Springer Netherlands.

Hooijberg, M. (1997). *Practical Geodesy. Using Computers.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Hooper, D.U., Chapin, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A.J., Vandermeer, J., & Wardle, D.A. (2005). Effects of biodiversity on ecosystem functioning. A consensus of current knowledge. *Ecological Monographs, 75*, 3–35.

Hyyppä, J., Hyyppä, M., Litkey, P., Yu, X., Haggrén, H., Rönnholm, P., Pyysalo, U., Pitkänen, J., & Maltamo, M. (2004). Algorithms and methods of airborne laser scanning for forest measurements. *International Archives of Photgrammetry, Remote Sensing and Spatial Information Sciences, 36*, 82–89.

Hyyppä, J., Yu, X., Hyyppä, H., Vastaranta, M., Holopainen, M., Kukko, A., Kaartinen, H., Jaakkola, A., Vaaja, M., Koskinen, J., & Alho, P. (2012). Advances in Forest Inventory Using Airborne Laser Scanning. *Remote Sensing, 4*, 1190–1207.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2015). *An introduction to statistical learning. With applications in R.* (6th ed.). New York, Heidelberg, Dordrecht, London: Springer.

Japkowicz, N., & Shaju, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis, 6*, 429–449.

Jerez, J.M., Molina, I., Garcia-Laencina, P.J., Alba, E., Ribelles, N., Martin, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial intelligence in medicine, 50*, 105–115.

Kahmen, H. (2006). *Angewandte Geodäsie. Vermessungskunde*. (20th ed.). Berlin: de Gruyter.

Kalton, G., & Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology, 12*, 1–16.

Keenan, R.J., Reams, G.A., Achard, F., Freitas, J.V. de, Grainger, A., & Lindquist, E. (2015). Dynamics of global forest area. Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management, 352*, 9–20.

Khoshgoftaar, T.M., Golawala, M., & van Hulse, J. (2007). An Empirical Study of Learning from Imbalanced Data Using Random Forest. In IEEE Computer Society (Ed.), *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)* (pp. 310–317). Los Alamitos, USA.

Komsta, L., & Novomestky, F. (2015). moments. Moments, cumulants, skewness, kurtosis and related tests. http://CRAN.R-project.org/package=moments. 25.06.2017.

Kramer, H., & Akça, A. (2008). *Leitfaden zur Waldmesslehre*. (5th ed.). Frankfurt am Main: Sauerländer.

Krüger, L. (1912). *Konforme Abbildung des Erdellipsoids in der Ebene.* Leipzig: B. G. Teubner.

Kuhn, M., Contributions from Wing, J., St. Weston, Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2016). caret: Classification and Regression Training. R package version 6.0-73. http://CRAN.R-project.org/package=caret. 22.03.2017.

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling.* New York: Springer New York.

Landesamt für Geologie, Rohstoffe und Bergbau (2016). WMS LGRB-BW BÜK200: Bodenübersichtskarte 1 : 200 000. Regierungspräsidium Freiburg - Abteilung 9 Landesamt für Geologie, Rohstoffe und Bergbau. http://maps.lgrb-bw.de/?view=lgrb_uek350_boden. 01.06.2017.

Landtag von Baden-Württemberg (2013). Gesetz zur Errichtung des Nationalparks Schwarzwald und zur Änderung weiterer Vorschriften. NLPG. In *Gesetzblatt für Baden-Württemberg* (2013), 17 (pp. 449–476). Stuttgart, Deutschland.

Laser Technology Inc. (2011). TruPulse 360°R. User's manual. http://www.lasertech.com/TruPulse-Laser-Rangefinder.aspx. 04.05.2017.

Latifi, H. (2012). Characterizing forest structure by means of remote sensing: a review. In B. Escalante (Ed.), *Remote Sensing - Advanced Techniques and Platforms* (pp. 1–26). Rijeka: InTech.

Latifi, H., Fassnacht, F., & Koch, B. (2012). Forest structure modeling with combined airborne hyperspectral and LiDAR data. *Remote Sensing of Environment, 121*, 10–25.

Latifi, H., Fassnacht, F.E., Müller, J., Tharani, A., Dech, S., & Heurich, M. (2015). Forest inventories by LiDAR data. A comparison of single tree segmentation and metric-based methods for inventories of a heterogeneous temperate forest. *International Journal of Applied Earth Observation and Geoinformation, 42*, 162–174.

Latifi, H., Heurich, M., Hartig, F., Müller, J., Krzystek, P., Jehl, H., & Dech, S. (2016). Estimating over- and understorey canopy density of temperate mixed stands by airborne LiDAR data. *Forestry, 89*, 69–81.

Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., & Harding, D. (1999). Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Remote Sensing of Environment, 70*, 339–361.

Lefsky, M.A., Cohen, W.B., & Spies, T.A. (2001). An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Canadian Journal of Forest Research, 31*, 78–87.

Leutner, B.F., Reineking, B., Müller, J., Bachmann, M., Beierkuhnlein, C., St. Dech, & Wegmann, M. (2012). Modelling Forest αDiversity and Floristic Composition - On the Added Value of LiDAR plus Hyperspectral Remote Sensing. *Remote Sensing, 4*, 2818–2845.

Li, M., Zang, S., Zhang, B., Li, S., & Wu, C. (2017). A Review of Remote Sensing Image Classification Techniques. The Role of Spatio-contextual Information. *European Journal of Remote Sensing, 47*, 389–411.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News, 2*, 18–22.

Lillesand, T.M., Kiefer, R.W., & Chipman, J.W. (2015). *Remote sensing and image interpretation.* New York: John Wiley & Sons.

Lin, W.-J., & Chen, J.J. (2013). Class-imbalanced classifiers for high-dimensional data. *Briefings in bioinformatics, 14*, 13–26.

Lindenmayer, D.B., & Likens, G.E. (2010). The science and application of ecological monitoring. *Biological Conservation, 143*, 1317–1328.

Little, R.J., & Rubin, D.B. (2002). *Statistical analysis with missing data.* New York: John Wiley & Sons.

Mace, G.M., Norris, K., & Fitter, A.H. (2012). Biodiversity and ecosystem services: a multilayered relationship. *Trends in ecology & evolution, 27*, 19–26.

Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J., & Pitkänen, J. (2004). Identifying and quantifying heterogeneous boreal forest structures using laser scanner data. *International Archives of Photgrammetry, Remote Sensing and Spatial Information Sciences, 36*, 153–156.

Martinuzzi, S., Vierling, L.A., Gould, W.A., Falkowski, M.J., Evans, J.S., Hudak, A.T., & Vierling, K.T. (2009). Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sensing of Environment, 113*, 2533–2546.

McRoberts, R.E., & Tomppo, E.O. (2007). Remote sensing support for national forest inventories. *Remote Sensing of Environment, 110*, 412–419.

Melin, M., Matala, J., Mehtätalo, L., Pusenius, J., & Packalen, P. (2016). Ecological dimensions of airborne laser scanning — Analyzing the role of forest structure in moose habitat use within a year. *Remote Sensing of Environment, 173*, 238–247.

Melin, M., Packalén, P., Matala, J., Mehtätalo, L., & Pusenius, J. (2013). Assessing and modeling moose (Alces alces) habitats with airborne laser scanning data. *International Journal of Applied Earth Observation and Geoinformation, 23*, 389–396.

MILAN Geoservice GmbH (2015). Projekt-Dokumentation. Airborne-Laser-Scanning - Forst Schwarzwald.

Millennium Ecosystem Assessment (2005). *Ecosystems and Human Well-being: Biodiversity Synthesis*. Washington, DC: World Resources Institute.

Mitchell, P.J., Lane, P.N.J., & Benyon, R.G. (2012). Capturing within catchment variation in evapotranspiration from montane forests using LiDAR canopy profiles with measured and modelled fluxes of water. *Ecohydrology, 5*, 708–720.

Morales-Hidalgo, D., Oswalt, S.N., & Somanathan, E. (2015). Status and trends in global primary forest, protected areas, and areas designated for conservation of biodiversity from the Global Forest Resources Assessment 2015. *Forest Ecology and Management, 352*, 68–77.

Müller, J., & Brandl, R. (2009). Assessing biodiversity by remote sensing in mountainous terrain. The potential of LiDAR to predict forest beetle assemblages. *Journal of Applied Ecology, 46*, 897–905.

Müller, J., Moning, C., Bässler, C., Heurich, M., & Brandl, R. (2009). Using airborne laser scanning to model potential abundance and assemblages of forest passerines. *Basic and Applied Ecology, 10*, 671–681.

Müller, J., Stadler, J., & Brandl, R. (2010). Composition versus physiognomy of vegetation as predictors of bird assemblages. The role of lidar. *Remote Sensing of Environment, 114*, 490–495.

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

Næsset, E. (2014). Area-Based Inventory in Norway - From Innovation to an Operational Reality. In M. Maltamo, E. Næsset, & J. Vauhkonen (Eds.), *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies* (pp. 215–240). Dordrecht: Springer Netherlands.

Næsset, E., Bollandsås, O.M., Gobakken, T., Gregoire, T.G., & Ståhl, G. (2013a). Model-assisted estimation of change in forest biomass over an 11year period in a sample survey supported by

airborne LiDAR. A case study with post-stratification to provide "activity data". *Remote Sensing of Environment, 128*, 299–314.

Næsset, E., Gobakken, T., Bollandsås, O.M., Gregoire, T.G., Nelson, R., & Ståhl, G. (2013b). Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway. *Remote Sensing of Environment, 130*, 108–120.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., & Söderman, U. (2004). Laser scanning of forest resources. The nordic experience. *Scandinavian Journal of Forest Research, 19*, 482–499.

Nationalpark Schwarzwald (2014). Leitbild Nationalpark Schwarzwald. Orientierende Handlungsleitsätze für die Nationalparkverwaltung. Seebach. http://www.schwarzwald-nationalpark.de/mediathek/pdf-mediathek. 25.06.2017.

Nationalpark Schwarzwald (2016). *Jahresbericht 2015. Nationalpark Schwarzwald - Erste Schritte.* Seebach: Nationalpark Schwarzwald.

Nettleton, D.F., Orriols-Puig, A., & Fornells, A. (2010). A study of the effect of different types of noise on the precision of supervised learning techniques. *Artificial Intelligence Review, 33*, 275–306.

Newman, D.A. (2014). Missing Data. *Organizational Research Methods, 17*, 372–411.

Nilsson, R., Pena, J.M., Bjorkegren, J., & Tegner, J. (2007). Consistent feature selection for pattern recognition in polynomial time. *Journal of Machine Learning Research, 8*, 589–612.

Noss, R.F. (1990). Indicators for Monitoring Biodiversity. A Hierarchical Approach. *Conservation Biology, 4*, 355–364.

O'Hara, K.L., Latham, Penelope A., Hessburg, Paul, & Smith Bradley G. (1996). A Structural Classification for Inland Northwest Forest Vegetation. *Western Journal of Applied Forestry, 11*, 97–102.

Oliver, C.D., & Larson, B.C. (1996). *Forest stand dynamics.* New York: John Wiley & Sons.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., & Wulder, M.A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment, 148*, 42–57.

Paillet, Y., Berges, L., Hjalten, J., Odor, P., Avon, C., Bernhardt-Romermann, M., Bijlsma, R.-J., Bruyn, L. de, Fuhr, M., Grandin, U., Kanka, R., Lundin, L., Luque, S., Magura, T., Matesanz, S., Meszaros, I., Sebastia, M.-T., Schmidt, W., Standovar, T., Tothmeresz, B., Uotila, A., Valladares, F., Vellak, K., & Virtanen, R. (2010). Biodiversity differences between managed and unmanaged forests: meta-

analysis of species richness in Europe. *Conservation biology: the journal of the Society for Conservation Biology, 24*, 101–112.

Pelletier, C., Valero, S., Inglada, J., Champion, N., Marais Sicre, C., & Dedieu, G. (2017). Effect of Training Class Label Noise on Classification Performances for Land Cover Mapping with Satellite Image Time Series. *Remote Sensing, 9*, 173.

Pickett, S.T.A., & White, P.S. (Eds.) (1985). The ecology of natural disturbance and patch dynamics. Orlando: Academic Press.

Pitt, D.G., Woods, M., & Penner, M. (2014). A Comparison of Point Clouds Derived from Stereo Imagery and Airborne Laser Scanning for the Area-Based Estimation of Forest Inventory Attributes in Boreal Ontario. *Canadian Journal of Remote Sensing, 40*, 214–232.

Prasad, A.M., Iverson, L.R., & Liaw, A. (2006). Newer Classification and Regression Tree Techniques. Bagging and Random Forests for Ecological Prediction. *Ecosystems, 9*, 181–199.

PricewaterhouseCoopers & ö:konzept (2013). Gutachten zum potenziellen Nationalpark im Nordschwarzwald. Gutachten zu Händen des Ministeriums für Ländlichen Raum und Verbraucherschutz des Landes Baden-Württemberg. Berlin. http://www.schwarzwald-nationalpark.de/mediathek/pdf-mediathek. 02.05.2017.

QGIS Development Team (2017). QGIS Geographical Information System. Open Source Geospatial Foundation Project. http://www.qgis.org/. 20.03.2017.

R Core Team (2015). R: A language and environment for statistical computing. https://www.R-project.org/. 13.03.2017.

Rands, M.R.W., Adams, W.M., Bennun, L., Butchart, S.H.M., Clements, A., Coomes, D., Entwistle, A., Hodge, I., Kapos, V., Scharlemann, J.P.W., Sutherland, W.J., & Vira, B. (2010). Biodiversity conservation. Challenges beyond 2010. *Science, 329*, 1298–1303.

Rauscher, M. (2017). Individuelle Analyse des Raum-Zeit-Verhaltens zweier Rotwildstücke (Cervus elaphus) im Nationalpark Schwarzwald anhand von GPS-Daten. Bachelorarbeit. Professur für Wildökologie und Jagdwirtschaft. Hochschule für Forstwissenschaft. Rottenburg a. N.

Reinhardt, K. (2017). Estimated time for imputing 22415 missing values by using missForest algorithm. personal communication. Bayreuth.

Reitberger, J., Schnörr, C., Krzystek, P., & Stilla, U. (2009). 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*, 561–574.

Remmert, H. (1992). Das Mosaik-Zyklus-Konzept und seine Bedeutung für den Naturschutz - Eine Übersicht. *Laufener Seminarbeiträge*, 45–57.

Richards, P.W. (1984). The three-dimensional structure of tropical rain forest. In G.T. Prance, S.L. Sutton, T.C. Whitmore, & A.C. Chadwick (Eds.), *Tropical Rain Forest. Ecology and Management* (pp. 3–10). Oxford: Blackwell.

Ringgaard, R., Herbst, M., & Friborg, T. (2012). Partitioning of forest evapotranspiration. The impact of edge effects and canopy structure. *Agricultural and Forest Meteorology, 166-167*, 86–97.

Rocchini, D., Foody, G.M., Nagendra, H., Ricotta, C., Anand, M., He, K.S., Amici, V., Kleinschmit, B., Förster, M., Schmidtlein, S., Feilhauer, H., Ghisla, A., Metz, M., & Neteler, M. (2013). Uncertainty in ecosystem mapping by remote sensing. *Computers & Geosciences, 50*, 128–135.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J.P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 93–104.

Roloff, A. (Ed.) (2010). Bäume Mitteleuropas. Von Aspe bis Zirbel-Kiefer. (1st ed.). Weinheim: John Wiley & Sons.

Rösch, M. (2015). Nationalpark - Natur - Weißtanne - Fichte. Sechs Jahrtausende Wald und Mensch im Nordschwarzwald. *Denkmalpflege in Baden-Württemberg, 3*, 154–159.

Saar-Tsechansky, M., & Provost, F. (2007). Handling Missing Values when Applying Classification Models. *Journal of Machine Learning Research, 8*, 1625–1657.

Scherzinger, W. (1997). Tun oder Unterlassen? Aspekte des Prozeßschutzes und Bedeutung des "Nichts-Tuns" im Naturschutz. *Laufener Seminarbeiträge, 97/1*, 31–44.

Schlund, W. (2003). Naturschutzgebiet und Bannwald Wilder See - Hornisgrinde. http://www4.lubw.baden-wuerttemberg.de/servlet/is/71000/. 29.04.2017.

Schönwiese, C.-D. (2008). *Klimatologie*. (1st ed.). Stuttgart: UTB GmbH.

Schöttle, M. (2005). *Geotope im Regierungsbezirk Freiburg*. (1st ed.). Karlsruhe: LfU.

Secretariat of the Convention on Biological Diversity (2005). *Handbook of the Convention on Biological Diversity. Including its Cartagena Protocol on Biosafety*. (3rd ed.). Montreal.

Shugart, H.H., Saatchi, S., & Hall, F.G. (2010). Importance of structure and its measurement in quantifying function of forest ecosystems. *Journal of Geophysical Research: Biogeosciences, 115 (G2)*.

Skowronski, N., Clark, K., Nelson, R., Hom, J., & Patterson, M. (2007). Remotely sensed measurements of forest structure and fuel loads in the Pinelands of New Jersey. *Remote Sensing of Environment, 108*, 123–129.

Song, Y., Imanishi, J., Sasaki, T., Ioki, K., & Morimoto, Y. (2016). Estimation of broad-leaved canopy growth in the urban forested area using multi-temporal airborne LiDAR datasets. *Urban Forestry & Urban Greening, 16*, 142–149.

Spies, T.A., & Turner, M.G. (1999). Dynamic forest mosaics. In M.L. Hunter (Ed.), *Maintaining biodiversity in forest ecosystems* (pp. 95–160). Cambridge: Cambridge University Press.

Stehman, S.V. (2009). Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing, 30*, 5243–5272.

Stehman, S.V., & Czaplewski, R.L. (1998). Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment, 64*, 331–344.

Stekhoven, D.J. (2013). missForest. Nonparametric Missing Value Imputation using Random Forest. R package version 1.4.

Stekhoven, D.J., & Buhlmann, P. (2012). MissForest-non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*, 112–118.

Stem, C., Margoluis, R., Salafsky, N., & Brown, M. (2005). Monitoring and Evaluation in Conservation. A Review of Trends and Approaches. *Conservation Biology, 19*, 295–309.

Sumnall, M.J., Hill, R.A., & Hinsley, S.A. (2016). Comparison of small-footprint discrete return and full waveform airborne lidar data for estimating multiple forest variables. *Remote Sensing of Environment, 173*, 214–223.

Sun, Y., Wong, A.K.C., & Kamel, M.S. (2009). Classification of imbalanced data: a review. *International Journal of Pattern Recognition and Artificial Intelligence, 23*, 687–719.

Suunto Oy (2017). Suunto Precision Instruments. User Guide KB-14/PM-5/Tandem. http://www.suunto.com/de-DE/Produkte/Kompasse/Suunto-KB-14/Suunto-KB-14360R-DG-Compass/. 04.05.2017.

Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *arXiv preprint arXiv:1701.05305*.

Tang, X., Li, H., Ma, M., Yao, L., Peichl, M., Arain, A., Xu, X., & Goulden, M. (2017). How do disturbances and climate effects on carbon and water fluxes differ between multi-aged and even-aged coniferous forests? *The Science of the total environment, 599-600*, 1583–1597.

Turner, M.G. (2010). Disturbance and landscape dynamics in a changing world 1. *Ecology, 91*, 2833–2849.

Valbuena, R., Maltamo, M., & Packalen, P. (2016). Classification of forest development stages from national low-density lidar datasets. A comparison of machine learning methods. *Revista de Teledetección*, 15–25.

van Hulse, J., & Khoshgoftaar, T. (2009). Knowledge discovery from imbalanced and noisy data. *Data & Knowledge Engineering, 68*, 1513–1542.

van Kane, R., McGaughey, R.J., Bakker, J.D., Gersonde, R.F., Lutz, J.A., & Franklin, J.F. (2010). Comparisons between field- and LiDAR-based measures of stand structural complexity. *Canadian Journal of Forest Research, 40*, 761–773.

Varhola, A., Coops, N.C., Bater, C.W., Teti, P., Boon, S., & Weiler, M. (2010). The influence of ground- and lidar-derived forest structure metrics on snow accumulation and ablation in disturbed forests. *Canadian Journal of Forest Research, 40*, 812–821.

Vauhkonen, J., Maltamo, M., McRoberts, R.E., & Næsset, E. (2014). Introduction to Forestry Applications of Airborne Laser Scanning. In M. Maltamo, E. Næsset, & J. Vauhkonen (Eds.), *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies* (pp. 1–16). Dordrecht: Springer Netherlands.

Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S., & Clawges, R.M. (2008). Lidar. Shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the Environment, 6*, 90–98.

Waal, T.d., Pannekoek, J., & Scholtus, S. (2011). *Handbook of statistical data editing and imputation.* Hoboken, NJ: John Wiley & Sons.

Wang, F. (1990). Improving remote sensing image analysis through fuzzy information representation. *Photogrammetric engineering and remote sensing, 8*, 1163–1169.

Weinacker, H., Koch, B., & Weinacker, R. (2004). TREESVIS: A software system for simultaneous ED-real-time visualisation of DTM, DSM, laser raw data, multispectral data, simple tree and building models. *International Archives of Photgrammetry, Remote Sensing and Spatial Information Sciences, 36*, 90–95.

White, J. (2013). *A best practices guide for generating forest inventory attributes from airborne laser scanning data using the area-based approach.* Victoria, British Columbia: Canadian Forest Service.

Willis, K.S. (2015). Remote sensing change detection for ecological monitoring in United States protected areas. *Biological Conservation, 182*, 233–242.

Wing, B.M., Ritchie, M.W., Boston, K., Cohen, W.B., Gitelman, A., & Olsen, M.J. (2012). Prediction of understory vegetation cover with airborne lidar in an interior ponderosa pine forest. *Remote Sensing of Environment, 124*, 730–741.

Wulder, M. (1998). Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. *Progress in Physical Geography, 22*, 449–476.

Wulder, M.A., Coops, N.C., Hudak, A.T., Morsdorf, F., Nelson, R., Newnham, G., & Vastaranta, M. (2014). Status and prospects for LiDAR remote sensing of forested ecosystems. *Canadian Journal of Remote Sensing, 39*, 1–5.

Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping. A review. *Journal of Plant Ecology, 1*, 9–23.

Yao, W., Krull, J., Krzystek, P., & Heurich, M. (2014). Sensitivity Analysis of 3D Individual Tree Detection from LiDAR Point Clouds of Temperate Forests. *Forests, 5*, 1122–1142.

Zellweger, F., Braunisch, V., Baltensweiler, A., & Bollmann, K. (2013). Remotely sensed forest structural complexity predicts multi species occurrence at the landscape scale. *Forest Ecology and Management, 307*, 303–312.

Zhang, L.J., Liu, C.M., Davis, C.J., Solomon, D.S., Brann, T.B., & Caldwell, L.E. (2004). Fuzzy classification of ecological habitats from FIA data. *Forest Science, 50*, 117–127.

Zolkos, S.G., Goetz, S.J., & DUBAYAH, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment, 128*, 289–298.

Appendix A

Table A-1 Forest structure LiDAR metrics

No.	Metric group Metric abbreviation	Description	Data	Reference
	Basic return counts			
1	n	Number of returns	data_0	Latifi et al., 2016
2	n^f	Number of first returns	data_0_f	-
3	$n^f\ m^{-2}$	Number of first returns per m^2	data_0_f	-
4	n^{veg}	Number of vegetation returns	data_0.5	-
5	n^{can}	Number of canopy returns	data_2	-
6	$n^{can.f}$	Number of first canopy returns	data_2_f	Næsset, 2002
7	$n.tar_{mean}$ [a]	Mean number of targets	data_0	-
8	$n.tar_{max}$ [a]	Maximum number of targets	data_0	-
	Vertical distribution metrics			
9	h_{max}	Maximum height	data_0	Falkowski et al., 2009
10	hp_{99}^f [a]	99th height percentile of first returns	data_0_f	-
11	h_{mean} [a c]	Mean height	data_0	Falkowski et al., 2009
12	h_{mean}^{veg} [a]	Mean height of vegetation	data_0.5	-
13	h_{mean}^{can}	Mean height of canopy	data_2	-
14	h_{med} [a c]	Median height	data_0	Falkowski et al., 2009
15	h_{med}^{veg} [a]	Median height of vegetation	data_0.5	-
16	h_{med}^{can}	Median height of canopy	data_2	-
17	$h_{med.r}$	Relative median height	data_0	Heurich and Thoma, 2008
18	$h_{med.r}^{veg}$	Relative median height of vegetation	data_0	-
19	$h_{med.r}^{can}$	Relative median height of canopy	data_0	-
20	$h_{med.r.99}$ [a]	Relative median height (hp_{99}^f)	data_0 & data_0_f	Heurich and Thoma, 2008
21	$h_{med.r.99}^{veg}$ [a]	Relative median height of vegetation (hp_{99}^f)	data_0.5 & data_0	-
22	$h_{med.r.99}^{can}$	Relative median height of canopy (hp_{99}^f)	data_2 & data_0	-
23	$h_{mode}^{can.f}$	Modal height of first canopy returns	data_2_f	Falkowski et al., 2009

24	sd [a c]	Standard deviation of heights	data_0	Latifi et al., 2016
25	sd^{veg} [a]	Standard deviation of vegetation heights	data_0.5	-
26	sd^{can}	Standard deviation of canopy heights	data_2	-
27	var [a c]	Variance of heights	data_0	Latifi et al., 2016
28	var^{veg} [a]	Variance of vegetation heights	data_0.5	-
29	var^{can}	Variance of canopy heights	data_2	-
30	cov [a]	Coefficient of variation of heights	data_0	Falkowski et al., 2009
31	cov^{veg} [a]	Coefficient of variation of vegetation heights	data_0.5	-
32	cov^{can}	Coefficient of variation of vegetation heights	data_2	-
33	$skew$ [a]	Skewness of heights	data_0	Falkowski et al., 2009
34	$skew^{veg}$ [a c]	Skewness of vegetation heights	data_0.5	-
35	$skew^{can}$	Skewness of canopy heights	data_2	-
36	$kurt$ [a b]	Kurtosis of heights	data_0	Falkowski et al., 2009
37	$kurt^{veg}$ [a b]	Kurtosis of vegetation heights	data_0.5	-
38	$kurt^{can}$	Kurtosis of canopy heights	data_2	-
	Sectioned vertical distribution metrics			
39-46	$hp_{10}, \dots, hp_{40}$ & $hp_{60}, \dots, hp_{90}$ [a c]	10th to 90th height percentiles (median excluded)	data_0	Heurich and Thoma, 2008
47	$n.hl_{[0,1]}$	Number of returns from 0 to 1 m	data_0	Falkowski et al., 2009
48-52	$n.hl_{]i,j]}$	Number of returns within height intervals (]i,j] =]1,2],]2,10],]10, 20],]20,30] and]30,60] in meter)	data_0	Falkowski et al., 2009
53	sd_{hq1} [a]	Standard deviation of return heights between ground and 1st height quartile	data_0	Zellweger et al., 2013
54	sd_{hq2} [a c]	Standard deviation of return heights between 1st and 2nd height quartile	data_0	Zellweger et al., 2013
55	sd_{hq3} [a]	Standard deviation of return heights between 2nd and 3rd height quartile	data_0	Zellweger et al., 2013
56	sd_{hq4} [a]	Standard deviation of return heights between 3rd height quartile and maximum height	data_0	Zellweger et al., 2013
	Density metrics			
57	$dens.hl_{[0,1]}$ [a]	Return density between 0 and 1 m	data_0	Falkowski et al., 2009

58-62	$dens.hl_{]i,j]}$ [a c]	Relative percentage of returns within height intervals (]i,j] =]1,2],]2,10],]10,20],]20,30] and]30,60] in meter)	data_0	Falkowski et al., 2009
63	$dens.can^{f}$ [a]	Absolute canopy density	data_0_f & data_2_f	Næsset, 2002
64-72	$dens.can^{f}_{hp_i}$ [a]	Canopy density above respective height percentile with i = 10th, ..., 90th	data_0_f & data_2_f	Næsset, 2002
73	$pr.can$ [a]	Penetration rate canopy	data_0	Latifi et al., 2016
74	$pr.shrub$ [a]	Penetration rate shrub layer	data_0	Latifi et al., 2016
75	$pr.ust$ [a c]	Penetration rate understory	data_0	Latifi et al., 2016
76	$pr.ul$ [a]	Penetration rate upper layer	data_0	Heurich and Thoma, 2008
77	$pr.il$ [a]	Penetration rate intermediate layer	data_0	Heurich and Thoma, 2008
78	$pr.ll$ [a]	Penetration rate lower layer	data_0	Heurich and Thoma, 2008
79	$pr.tot$ [a]	Total penetration rate	data_0	Heurich and Thoma, 2008
	3D return distribution metrics			
80	og [a]	Number of open gap voxels	data_0.5	Lefsky et al., 1999; Coops et al., 2007
81	cg [a c]	Number of closed gap voxels	data_0.5	Lefsky et al., 1999; Coops et al., 2007
82	ez [a]	Number of euphotic zone voxels	data_0.5	Lefsky et al., 1999; Coops et al., 2007
83	oz	Number of oligophotic zone voxels	data_0.5	Lefsky et al., 1999; Coops et al., 2007
84	og_{rel} [a]	Relative percentage of open gap voxels	data_0.5	-
85	cg_{rel} [a]	Relative percentage of closed gap voxels	data_0.5	-
86	ez_{rel} [a]	Relative percentage of euphotic zone voxels	data_0.5	-
87	oz_{rel}	Relative percentage of oligophotic zone voxels	data_0.5	-

[a] Preselected metrics, [b] excluded after missForest imputation, [c] partly or fully selected by RFE (see chapter 4.1)

Appendix B

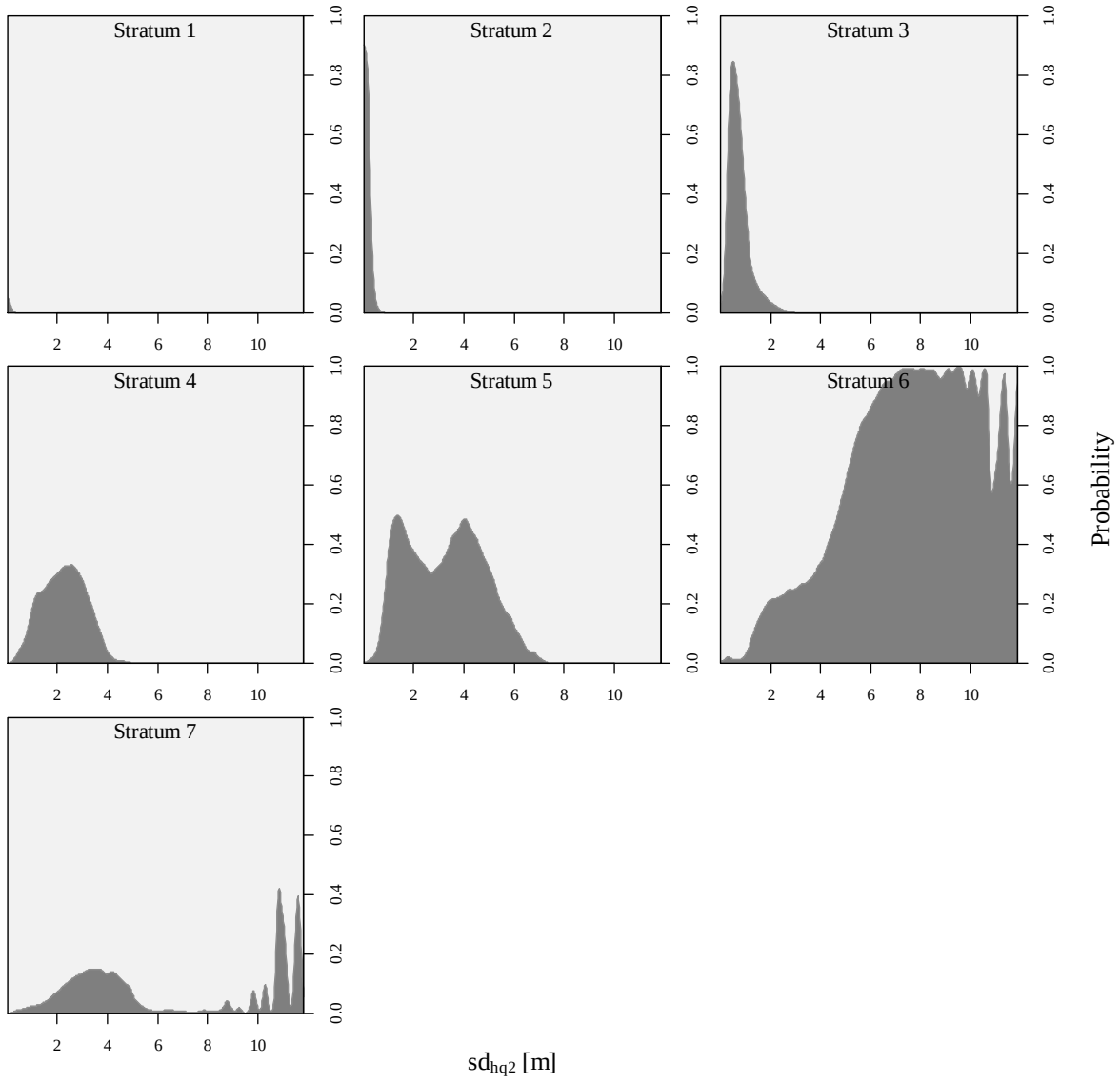

Figure B-1 Conditional density plots per stratum for sd_{hq2}

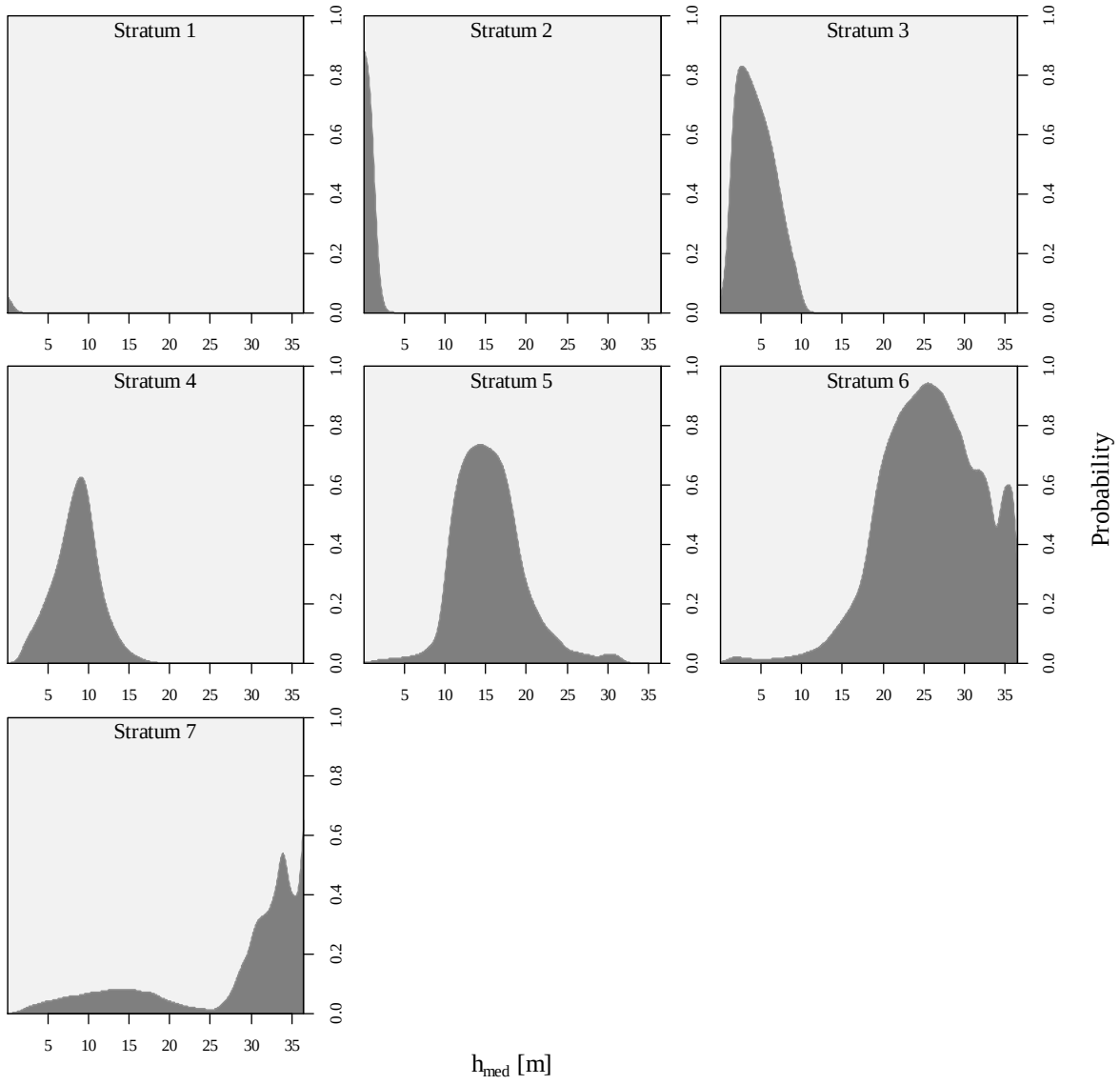

Figure B-2 Conditional density plots per stratum for h_{med}

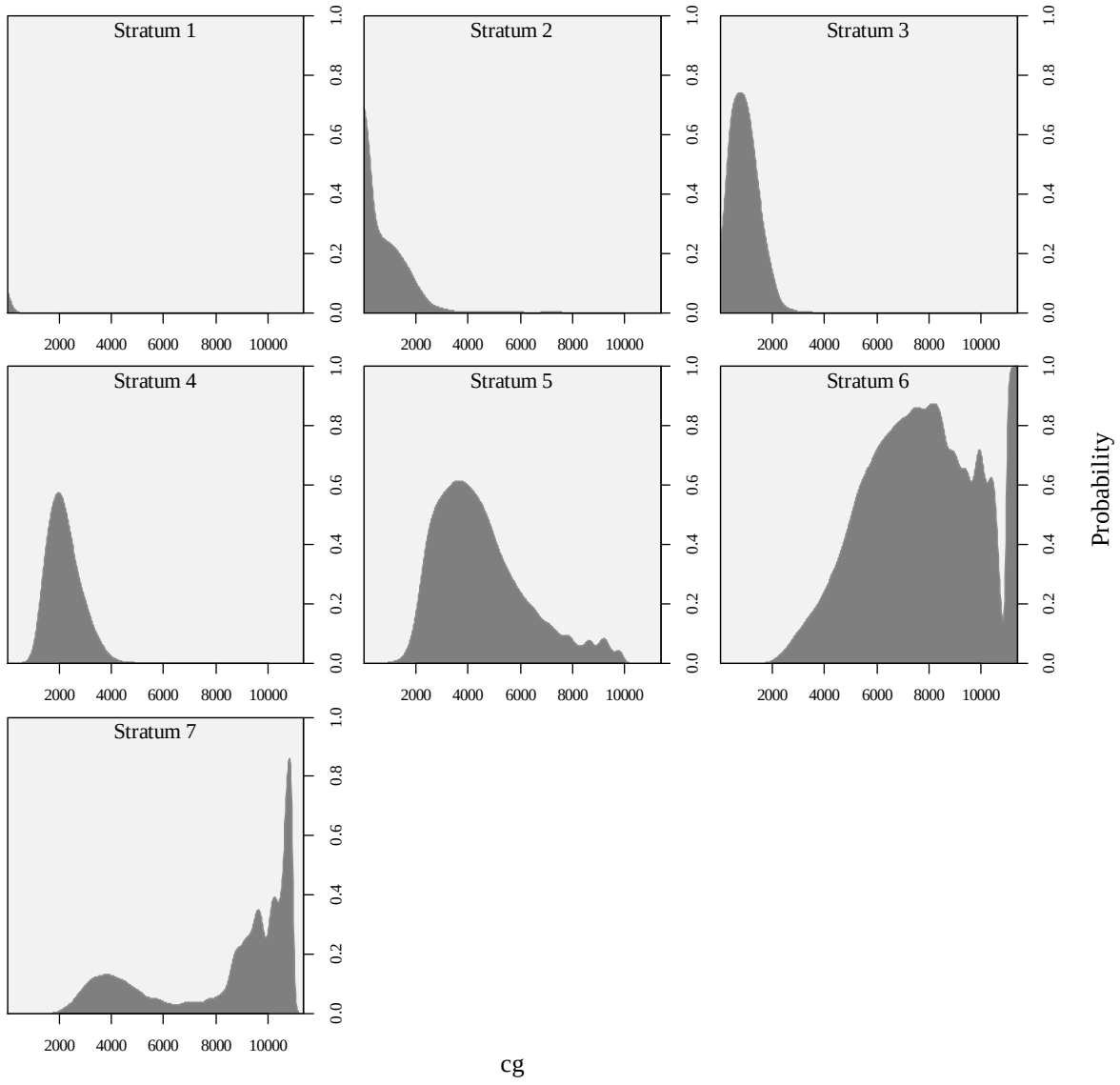

Figure B-3 Conditional density plots per stratum for *cg*

Appendix C

Content of the accompanying DVD

Observations

- Specific observation plot folders
  - GPS centre coordinate measurement data (folder: *etrex*)
  - Panorama photo at plot centre (folder: *Pano*)
  - LiDAR raw data per circular and squared plot base (folder: *RawData_LiDAR*)
  - Various graphics based on plot data (folder: *Graphics*)
  - Field data (file: *PlotXXX_Aufnahmebogen.xls*)
- Overview of plot coordinates and strata (file: *Plots_Str1-7_MA.csv*)

Predictors

- Non-imputed LiDAR metrics data of all potential 260 training and test plots (file: *LiDARmetrics_traintest.csv*)
- Imputed LiDAR metrics data of selected 111 training and test plots (file: *LiDARmetrics_traintest_MA_imputed.csv*)
- Non-imputed LiDAR metrics data of the BFNP grid with water areas excluded (file: *LiDARmetrics_BFNP_exWater.csv*)
- Imputed, predictor reduced LiDAR metrics data of the BFNP grid with water areas excluded (file: *LiDARmetrics_BFNP_exWater.csv*)

R_codes

- Calculation of LiDAR metrics for the BFNP grid (file: *LiDARmetrics_grid.R*); not executable as the entire LiDAR data set of the BFNP is not appended
- Calculation of LiDAR metrics for the training and test plots (file: *LiDARmetrics_traintest.R*); executable on the LiDAR raw data per plot (folder: *Observations\RawData_LiDAR*)
- Data preparation and RF classification (file: *DataPreparation_RFClassification.R*) executable on imputed LiDAR metrics data (folder: *Predictors*)
- R-workspace after final classification (file: *Workspace_FinalClass.RData*)

Results

- Imputed, predictor reduced, and predicted LiDAR metrics data of the BFNP grid with water areas excluded (file: *LiDARmetrics_BFNP_exWater_imputed_predicted.csv*)
- Final classification raster separated for the northern and southern part of the BFNP for GIS (files: *Classification_Coniferous_North.tif*, *Classification_Coniferous_South.tif*)
- Membership probability raster for GIS (file: *MembershipProbabilityRaster.tif*)