

Usability of citizen science observations together with airborne laser scanning data in determining the habitat preferences of forest birds

Mononen, L.^{ab*}, Auvinen, A.-P.^a, Packalen, P.^c, Virkkala, R.^a, Valbuena, R.^d, Bohlin, I.^e, Valkama, J.^f, Vihervaara, P.^a

- a) Finnish Environment Institute (SYKE), Joensuu, Oulu, and Helsinki Offices, P.O. Box 140, 00251 Helsinki, Finland.
- b) Department of Geographical and Historical Studies, University of Eastern Finland (UEF), P.O. Box 111, 80101 Joensuu, Finland
- c) School of Forest Sciences, University of Eastern Finland (UEF), P.O. Box 111, 80101 Joensuu, Finland.
- d) University of Cambridge, Department of Plant Sciences, Forest Ecology and Conservation, Downing Street, Cambridge CB2 3EA, United Kingdom.
- e) Swedish University of Agricultural Sciences (SLU), Department of Forest Resource Management, SE-901 83, Umeå, Sweden.
- f) Finnish Museum of Natural History (Luomus), P.O. Box 17, FI-00014 University of Helsinki, Finland.

SUMMARY

1. Citizens' field observations are increasingly stored in accessible databases, which makes it possible to use them in research. Citizen science (CS) complements the field work that must necessarily be carried out to gain an understanding of any of bird species' ecology. However, CS data holds multiple biases (e.g. presence only data, location error of bird observations, spatial data coverage) that should be paid attention before using the data in scientific research.
2. The use of Airborne Laser Scanning (ALS) enables investigating forest bird species' habitat preferences in detail and over large areas. In this study the breeding time habitat preferences of 25 forest bird species were investigated by coupling CS observations together with nine forest structure parameters that were computed using ALS data and field plot measurements. Habitat preferences were derived by comparing surroundings of presence-only observations against the full landscape. Also, in order to account for bird observation location errors, we analysed several buffering alternatives.
3. The results correspond well with the known ecology of the selected forest bird species. The size of a bird species' territory as well as some behavioural traits affecting detectability (song volume, mobility etc.) seemed to determine which bird species' CS data could be analysed with this approach. Especially the habitats of specialised species with small or medium sized territories differed from the whole forest landscape in the light of several forest structure parameters. Further research is needed to tackle issues related to the behaviour of the observers (e.g. birdwatchers' preference for roads) and characteristics of the observed species (e.g. preference for edge habitats), which may be the reasons for few unexpected results.
4. Our study shows that coupling CS data with ALS yield meaningful results that can be presented with distribution figures easy to understand and, more importantly, that can cover areas larger than what is normally possible by means of purpose-designed research projects. However, the use of CS data requires an understanding of the process of data collection by volunteers. Some of the biases in the data call for further thinking in terms of how the data is collected and analysed.

Keywords: Citizen Science (CS), Airborne laser scanning (ALS), Light detection and ranging (LiDAR), forest bird habitats, habitat modelling

56 1 INTRODUCTION

57

58 Citizen science (CS) involves the collaboration of professionals and non-professionals in scientific
59 research. During the past decades citizen participation has become a common practice in collecting
60 ecological data for environmental monitoring (Conrad & Hilchey 2011; Dickinson *et al.* 2012). Due to
61 their detectability and the high level of ornithological expertise among non-professionals, birds are
62 among the species groups of which CS observations hold the most potential to be used in research.
63 There are several well-established procedures for sampling birds that contain elements of CS (e.g.
64 Sullivan *et al.* 2009; Laaksonen & Lehtikoinen 2013). Observation schemes involving a strong CS
65 component have been used to study the timing of migration (Jonzén *et al.* 2006; Saino *et al.* 2010;
66 Lehtikoinen *et al.* 2013). However, few if any previous attempts have been made to use CS data in
67 studying the habitat preferences of forest birds, especially in connection with remote sensing.

68

69 Information on habitat characteristics has improved both in terms of accuracy and spatial extent over the
70 past century along with the development of remote sensing (RS) techniques (Kerr & Ostrovsky 2003;
71 Cohen & Goward 2004). In particular, habitat and species distribution modelling have benefited a great
72 deal from the development of RS, which supplements or sometimes even replaces traditional field work
73 (Pettorelli *et al.* 2014). Airborne Laser Scanning (ALS) provides three-dimensional information which
74 greatly advances the spatial analysis of habitat structures (e.g. Lefsky *et al.* 2002; Hill & Thompson
75 2005; Davies & Asner 2014; Valbuena *et al.* 2017) and helps detecting changing patterns of habitat use
76 in a changing climate (Melin *et al.* 2014). Since human activity has a strong effect on the structural
77 complexity of forests (e.g. Brokaw & Lent 1999), ALS derived information has been acknowledged
78 valuable for biodiversity assessments (Vierling *et al.* 2008). ALS parameters assist in the detection of
79 those species that depend on or benefit from the structural heterogeneity of canopy structure (Goetz *et al.*
80 *et al.* 2007; Vierling *et al.* 2008; Palminteri *et al.* 2012).

81

82 Birds are a species group that respond to environmental changes relatively promptly (Barbet-Massin *et al.*
83 *et al.* 2012; Frishkoff *et al.* 2014; Virkkala & Lehtikoinen 2014) which makes them, along with several other
84 characteristics (including ecological traits such as position in a food chain and non-ecological traits such
85 as popularity), good indicators of biodiversity (Butchart *et al.* 2010; Gregory & van Strien 2010).
86 Traditionally, forest bird-habitat relations have been studied in the field by measuring certain habitat
87 variables (e.g. tree species, height, diameter etc.) and connecting these measurements with bird
88 observations (see e.g. MacArthur & MacArthur 1961, Wiens 1989a and references therein). This kind of
89 analysis produces detailed information of species' habitat selection, but can usually be carried out only
90 over relatively small areas. Importantly, there is also stochasticity in species occurrence, which makes
91 difficult to extrapolate results of bird-habitat studies conducted at fine scales (Wiens *et al.* 1987; Haila *et al.*
92 *et al.* 1996; Virkkala & Rajasärkkä 2006). Bird-habitat relationships must therefore be studied on different
93 spatial scales (Wiens *et al.* 1987; Wiens 1989b). On the other hand, ALS and other RS datasets
94 connected with species observations can be used over large areas, and thus they are not susceptible to

95 small scale variation in the occurrence of bird species. In fact, the Group on Earth Observations
96 Biodiversity Observation Network (GEO BON) has identified the potential of remote sensing and in situ
97 data combinations to contribute for extensive and cost-efficient biodiversity monitoring (GEOBON 2015).
98 Using high quality CS observation could greatly advance this goal.

99

100 Several previous studies have proven the capability of ALS derived parameters to predict the species-
101 richness of habitats (reviewed in Simonson *et al.* 2014) and, more recently, the differentiation of diversity
102 among habitats (e.g. Zellweger *et al.* 2017). Further, some studies have successfully examined the
103 specific species-habitat relations by using ALS, but these have focused only on a few habitat indicators
104 or species or both (e.g. Graf *et al.* 2009; Goetz *et al.* 2010; Hagar *et al.* 2014; Melin *et al.* 2016). To date,
105 only few papers have examined the habitat preferences of multiple forest song birds with ALS. E.g. in
106 Hinsley *et al.* (2009) and Müller *et al.* (2009) the observation data was surveyed by professional
107 ecologists. To our knowledge, no previous studies have examined the use of CS data as rigorously (but
108 see Vihervaara *et al.* 2015). In this study the relatively high number of species (25) was achieved by
109 using CS data – collecting such a large dataset over such a large area would have been out of our reach
110 by means of a purpose-designed research project.

111

112 In this study we explore the extent to which CS data can be used to assess the habitat preferences of
113 forest birds, and identify potential pitfalls when doing so. We use positioned observations from 25 forest
114 bird species and nine ALS derived parameters to: I) Explore whether CS observations in connection with
115 ALS based forest structure parameters can provide information that is in line with the known ecological
116 characteristics (e.g. habitat preferences) of the bird species included in the study, II) Investigate which
117 bird species' habitats could be best modelled by using the combination of CS and ALS data, and III)
118 Examine which forest structure parameters are most suitable for predicting bird species' habitats in this
119 connection. In order to facilitate the replicability of the method, the low-density ALS data were used as
120 today they are typically acquired at national scale. However, we used a selection of ALS derived forest
121 parameter layers that can be computed with field plot measurements to examine whether they can offer
122 more detailed or complementary information for research. The potential applications of our study relate
123 not only to the field of animal ecology, but can also help in determining where to focus conservation
124 activities (Rose *et al.* 2015). Potentially, the combination of CS observations and ALS data could enable
125 us to cover areas as large as administrative regions or even nations, and, in the future, also to model
126 and predict the occurrence of species of conservation interest.

127

128

129

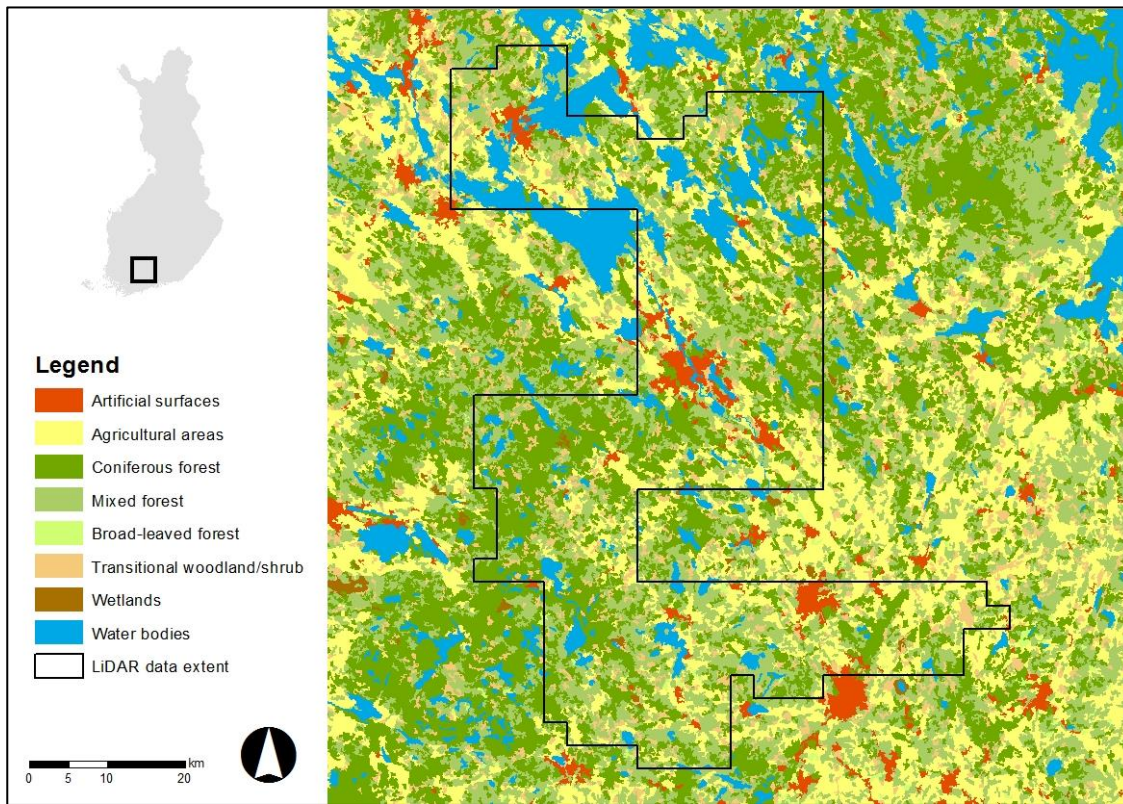
130 **2 MATERIALS AND METHODS**

131

132 **2.1 Study area**

133

134 The study area is located in Southern Finland in the Lake Vanajavesi catchment area of 3 000 km²
135 (Figure 1). The area belongs to the southern boreal taiga vegetation zone and the landscape is
136 dominated by boreal forests. The majority of the forests are commercially managed. Large lakes and
137 small rivers, agricultural areas and wetlands are also typical in the area. The study area was selected
138 based on the available ALS data and CS bird observations.



139
140 **Figure 1.** The study area is located in southern Finland. The area is determined by the coverage of ALS
141 data (black line). Data: Corine Land Cover 2012.

142 143 2.2 Species observations

144
145 In our study, we included species of conservation concern, such as the European Union's Birds Directive
146 Species (Annex I) and redlisted species in Finland, species preferring old-growth or mature forests, and
147 species of herb-rich, lush, and deciduous forests (see Vihervaara *et al.* 2015). We also included species
148 occurring in boreal agricultural-forest mosaics. The forest bird observation data were acquired from two
149 sources; (i) Bird Atlas data from the database of the Finnish Museum of Natural History and (ii) faunistic
150 observations from the Tiira database maintained by BirdLife Finland. Both of these data have been
151 collected by mostly non-professional volunteers and, although the Bird Atlas was more goal-oriented and
152 structured, can be described with good reason as citizen science. All observations were recorded in
153 years 2006–2012. Only the observations that had higher location accuracy than 100 m (marked by the
154 observers) were included in the final dataset. For each species only the observations during their known
155 breeding time were included in the data. In the end, 25 bird species were included and the numbers of
156 observations per species vary between 31 and 355 (Table 1).

157

158
159

Table 1. Bird occurrence data points (N of points), separated by data source (Atlas and Tiira), each species territory size and observation distance.

	Abbrevia tion	Latin name	English name	N of point s	Tiira data	Atlas data	Home range size*	Observat ion distance **
1	ACCGEN	Accipiter gentilis	Goshawk	90	70	20	L	396
2	ACCNIS	Accipiter nisus	Sparrowhawk	81	73	8	L	171
3	AEGCAU	Aegithalos caudatus	Long-tailed Tit	121	111	10	S	20
4	AEGFUN	Aegolius funereus	Boreal Owl	54	48	6	L	483
5	ASIOTU	Asio otus	Long-eared Owl	148	128	20	L	266
6	BONBON	Bonasa bonasia	Hazel Grouse	315	303	12	M	21
7	BUBBUB	Bubo bubo	Eagle Owl	31	25	6	L	487
8	BUTBUT	Buteo buteo	Buzzard	57	47	10	L	167
9	CAPEUR	Caprimulgus europaeus	Nightjar	39	35	4	M	234
10	DENMIN	Dendrocopos minor	Lesser Spotted Woodpecker	232	221	11	M	62
11	DRYMAR	Dryocopus martius	Black Woodpecker	330	319	11	L	169
12	FICPAR	Ficedula parva	Red-breasted Flycatcher	42	33	9	S	50
13	GLAPAS	Glaucidium passerinum	Pygmy Owl	57	51	6	L	166
14	HIPICT	Hippolais icterina	Icterine Warbler	119	109	10	S	60
15	JYNTOR	Jynx torquilla	Wryneck	92	82	10	S	77
16	LUSLUS	Luscinia luscinia	Thrush Nightingale	355	342	13	S	254
17	PERAPI	Pernis apivorus	Honey Buzzard	77	69	8	L	165
18	PHYDES	Phylloscopus trochiloides	Greenish Warbler	34	29	5	S	40
19	PICCAN	Picus canus	Grey-headed Woodpecker	216	203	13	M	120
20	PICTRI	Picooides tridactylus	Three-toed Woodpecker	44	36	8	M	28
21	STRALU	Strix aluco	Tawny Owl	150	126	24	L	382
22	STRURA	Strix uralensis	Ural Owl	78	71	7	L	450
23	TETRIX	Tetrao tetrix	Black Grouse	155	151	4	L	302
24	TETURO	Tetrao urogallus	Capercaillie	133	127	6	L	35
25	TURVIS	Turdus viscivorus	Mistle Thrush	42	35	7	M	58

160

161

162

163

164

165

166

167

* The size of the area over which the species is likely to be observed during the breeding period (normally the territory that the birds defend by singing, but in some cases a larger area used for hunting etc.); L = large, > 100 ha; M = medium, 10–100 ha; S = small, < 10 ha.

** Species-specific maximum distance between the bird and observer that covers 75 % percent of Tiira observations (data contain a number of outliers). See Vihervaara *et al.* (2015).

168 2.3 ALS data

169

170 ALS data were collected on May–June 2008 using Optech ALTM GEMINI laser scanning system. The
171 ALS data point is offset at most by four years from the bird data (2006–2012). In a previous study a gap
172 of this size between the acquisition of ALS and CS data was found to be of marginal impact on the
173 results (Vierling *et al.* 2014). The area was measured from an altitude of 2 000 metres above ground
174 level using half angle of 20°. This resulted in a swath width of 1 450 metres and a nominal sampling
175 density of about 0.5 measurements per square metre. A digital terrain model (DTM) was generated from
176 the ALS data by classifying ALS echoes as ground points and other points as explained in Axelsson
177 (2000). Finally, the orthometric heights of ALS echoes (Z) were converted to above ground heights (dZ)
178 by subtracting the DTM at the corresponding location.

179

180 2.4 Forest sample plots

181

182 The sample plot data were acquired during the summers of 2007 and 2008. A network of circular sample
183 plots with a radius of 9 meters was established on the area. Sample plots were placed over forest stands
184 with different development stages and dominant tree species. Only one sample plot was placed to a
185 stand so usually the distance between nearest plots is at least hundreds of meters. A total of 249 sample
186 plots were measured. The Global Positioning System with differential correction was used to determine
187 the position of the centre of each plot to an accuracy of about 1 meter (the accuracy of the positioning
188 system was tested in a comparable forest area; unpublished data). The diameter at breast height (DBH)
189 and tree species were recorded for all trees with DBH above 5 cm. Tree height was measured from
190 basal area median tree by tree species and storey class on each plot. Näslund's (1937) model was used
191 to predict the height for the rest of the trees. The size of each individual tree was described by its basal
192 area, stem volume and the biomass stocked above ground. Stem volume was predicted using models by
193 Laasasenaho (1982) and above ground biomass (AGB) using models by Repola *et al.* (2007). Plot level
194 stem volumes and AGBs were computed by summing up individual trees by plot and finally species
195 proportions (*Pine*, *Spruce* and *Deciduous*) were calculated based on plot volume. Dominant height
196 (*Hdom*) was determined at the plot level as the mean height of the 100 trees with the largest diameter at
197 breast height per hectare. Forest structure was described by two parameters: the Gini coefficient (*GC*)
198 and the proportion of basal area larger than mean (*BALM*), respectively describing tree size dispersion
199 and asymmetry (Valbuena *et al.* 2013). The *GC* measures the relative inequality in individual tree basal
200 areas, and therefore it evaluates the heterogeneity of tree sizes in the forest population (Weiner &
201 Solbrig 1984). Using *GC* to describe forest structural heterogeneity is similar to the more common
202 parameter for foliage height diversity (*FHD*) based on Shannon's diversity (MacArthur & MacArthur
203 1961), which is more commonly used in ALS-based studies (e.g. Clawges *et al.* 2008). Valbuena *et al.*
204 (2012) explained the mathematical links between these two parameters and why *GC* was preferred
205 above *FHD*. On the other hand, *BALM* was used to evaluate relative dominance among various tree

206 storeys, since understorey ingrowth decreases the proportion of basal area stocked above its mean, and
 207 hence *BALM* (Gove 2004; Valbuena *et al.* 2015).

208 **2.5 Forest structure parameters**

209

210 We considered two main types of ALS derived forest structure parameters (Table 2): (i) those estimated
 211 using the field sample plot data together with metrics computed from the ALS data (Næsset 2002;
 212 Maltamo *et al.* 2005; Valbuena *et al.* 2013; Asner & Mascaro 2014) and (ii) those calculated directly from
 213 the ALS data (Nelson *et al.* 1984; Miura & Jones 2010; Valbuena *et al.* 2017). The variables of former
 214 group were first modelled using field sample plots as response variables and ALS metrics as predictor
 215 variables (Vihervaara *et al.* 2015). This group consists of parameters: *AGB*, *Hdom*, *GC*, *BALM*, *Pine*,
 216 *Spruce* and *Deciduous*. Error rates and biases were reported in Vihervaara *et al.* (2015). The latter group
 217 were metrics calculated directly from the heights above ground of ALS echoes (dZ), which have been
 218 demonstrated to be directly related to the structural properties of forests. One such metric was the
 219 proportion of ALS echoes backscattered from of vegetation (*F_veg*), in other words, those echoes that
 220 are backscattered from vegetation above half a meter. *F_veg* is regarded as a good proxy for canopy
 221 cover (Nelson *et al.* 1984). Another metric employed was the coefficient of L-skewness (*Lskew*) (Hosking
 222 1990), which has been regarded as a good proxy for light availability and therefore understorey
 223 regeneration (Valbuena *et al.* 2017). Finally, forest parameters were either predicted or calculated to the
 224 whole study area using a grid with 15 m cell size.

225

226 **Table 2.** Forest structure parameters that were used in this study.

Code	Description	Resolution	Data used
<i>AGB</i>	Above ground biomass of trees (Mg/ha)	15 m	ALS + sample plots
<i>BALM</i>	Basal area larger than mean: Relative development of overstorey and understorey (Valbuena <i>et al.</i> 2013).	15 m	ALS + sample plots
<i>Deciduous (%)</i>	Proportion of deciduous trees (%) with respect to stem volume	15 m	ALS + sample plots
<i>F_veg</i>*	Proportion of vegetation ALS echoes (%). An ALS proxy for amount of canopy cover (Nelson <i>et al.</i> 1984).	15 m	ALS only
<i>GC</i>*	Gini coefficient. Inequality of sizes among trees (Valbuena <i>et al.</i> 2012).	15 m	ALS + sample plots
<i>Hdom (m)</i>*	Dominant height of trees (m). Average upper canopy	15 m	ALS + sample plots
<i>Lskew</i>*	L-Skewness of ALS echo heights. An ALS proxy for light availability and regeneration (Valbuena <i>et al.</i> 2017).	16 m	ALS only
<i>Pine (%)</i>	Proportion of pine (%) with respect to stem volume	15 m	ALS + sample plots
<i>Spruce (%)</i>	Proportion of spruce (%) with respect to stem volume	15 m	ALS + sample plots

227 * These parameters could be computed from ALS data without field measurements.

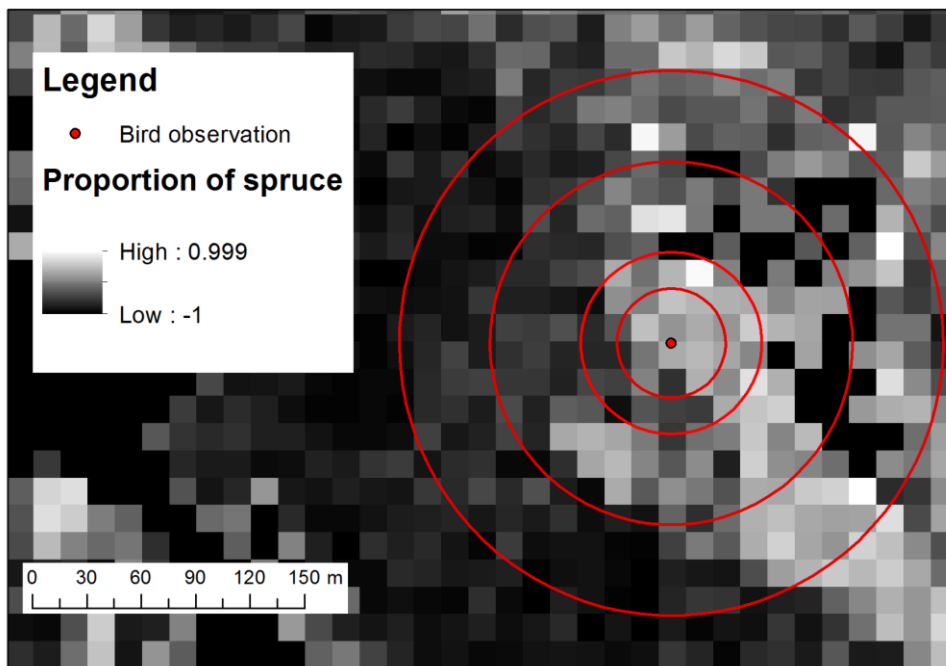
228

229 **2.6 Linking forest parameters to bird observations**

230

231 We generated buffers around bird observation points (Figure 2). To define a suitable buffer size that
232 captures relevant information on the surroundings of bird observation point, we generated buffers with
233 30, 50, 100 and 150 m radii. The forest parameter layers were intersected with the buffers assuming that
234 the cells having a centre point within the buffer are inside.

235



236

237 **Figure 2.** Buffers around observation points. As an example a forest parameters in the background (15
238 m grid).

239

240 The overlapping buffers were considered only once to avoid over-representation of areas where the
241 same species had been observed many times. The cells that included urban or water areas were
242 excluded from both datasets. For each buffer, data was processed for 25 birds and for 9 forest
243 parameters ($25 \times 9 = 225$). Additionally, a dataset that represents the landscape of the entire study area
244 was created. Density curves were created for each bird and parameter combination and the landscape
245 and they were normalized with respect to landscape. It was therefore employed as a measure of the
246 divergence between the habitat preferences of a given species relative to what it is available in the whole
247 study area.

248

249 **2.7 Tests for differences**

250

251 We computed overlap (OL) and Kolmogorov-Smirnov's D (KS-D) metrics to all bird and forest parameter
252 combinations and for the landscape. These were computed in the R environment using 'overlap'
253 (Meredith & Ridout 2016) and 'stats' packages (R Core team 2016). The overlapping values indicate how
254 much the distribution of a certain forest parameter of a certain bird species sample and the landscape of

255 the same forest parameter overlap (%). Smaller percentages would indicate bigger differences between
256 the distributions whereas an overlap value of 100 would mean perfect similarity.

257

258 We used Kolmogorov-Smirnov D statistics to measure the maximum distance between the empirical
259 distribution function of the sample and the cumulative distribution function of the landscape distribution.
260 Higher Kolmogorov-Smirnov D values indicate higher maximum difference between the two distributions
261 and therefore help detecting the species-specific habitat preferences “peaks” in relation to the chosen
262 forest parameters. The actual Kolmogorov–Smirnov significance test was not used because CS bird
263 observations cannot be considered as probability samples. Finally, profiles for each bird species with all
264 forest parameter were generated (see figure 4 as an example).

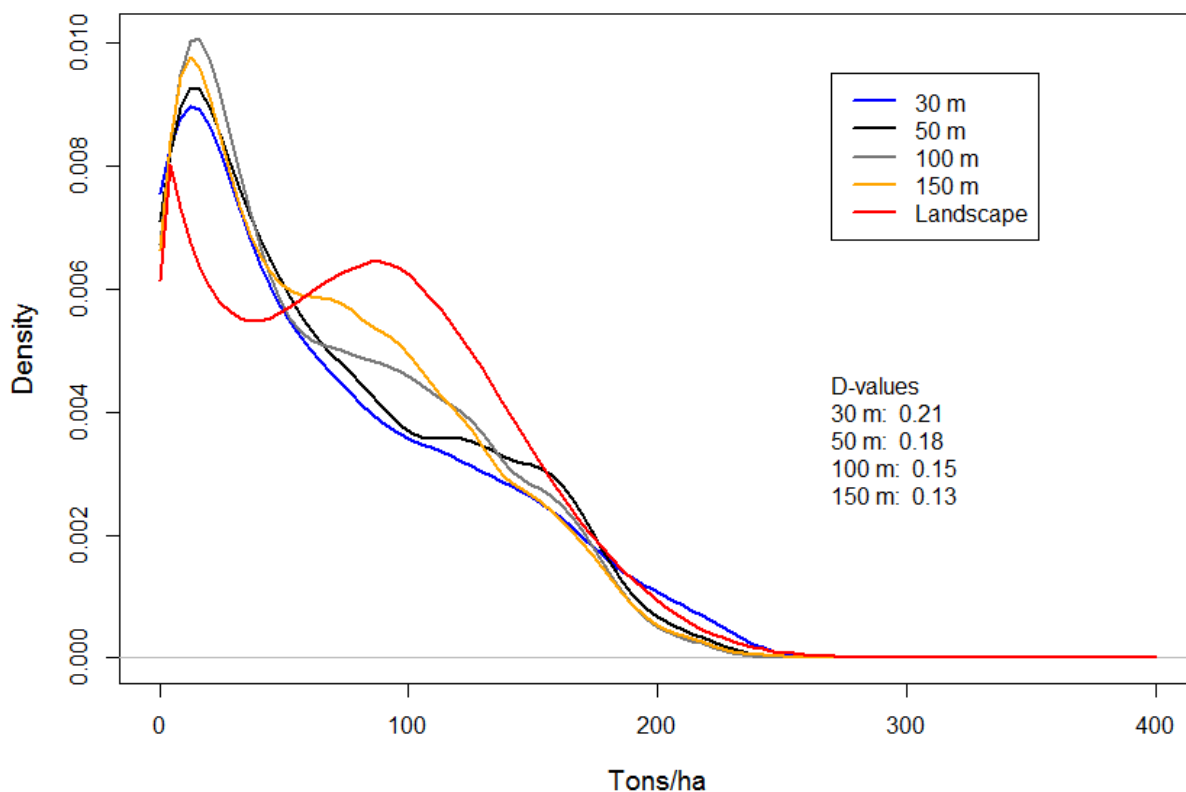
265

266 3 RESULTS

267

268 The Kolmogorov-Smirnov D statistics showed greater differences between the sample and landscape
269 when we used smaller buffers (30 m and 50 m). The results with two of the largest buffers (100 m and
270 150 m) appeared to lose information on bird-specific habitat preferences. An example of this effect is
271 shown in Fig. 3.). The D statistics captured the greatest differences with 30 m buffer. Across all species
272 and forest structure parameters, 50 m, 100 m and 150 m buffers respectively obtained differences 11%,
273 22% and 29% smaller than 30 m buffers. Therefore, the results in this paper are further reported only for
274 30 m buffer for each bird species, because this buffer most accurately measures bird-habitat relations
275 (see Discussion).

276



277

278 **Figure 3.** An example showing how the size of the buffer around each bird observation point affects the
 279 density curves and D statistics as calculated for nightjar in relation to above ground biomass. The D
 280 statistic expresses the maximum distance between the distribution curve representing the landscape
 281 (red line) and those of the nightjar observation points surrounded by different sized buffers).
 282

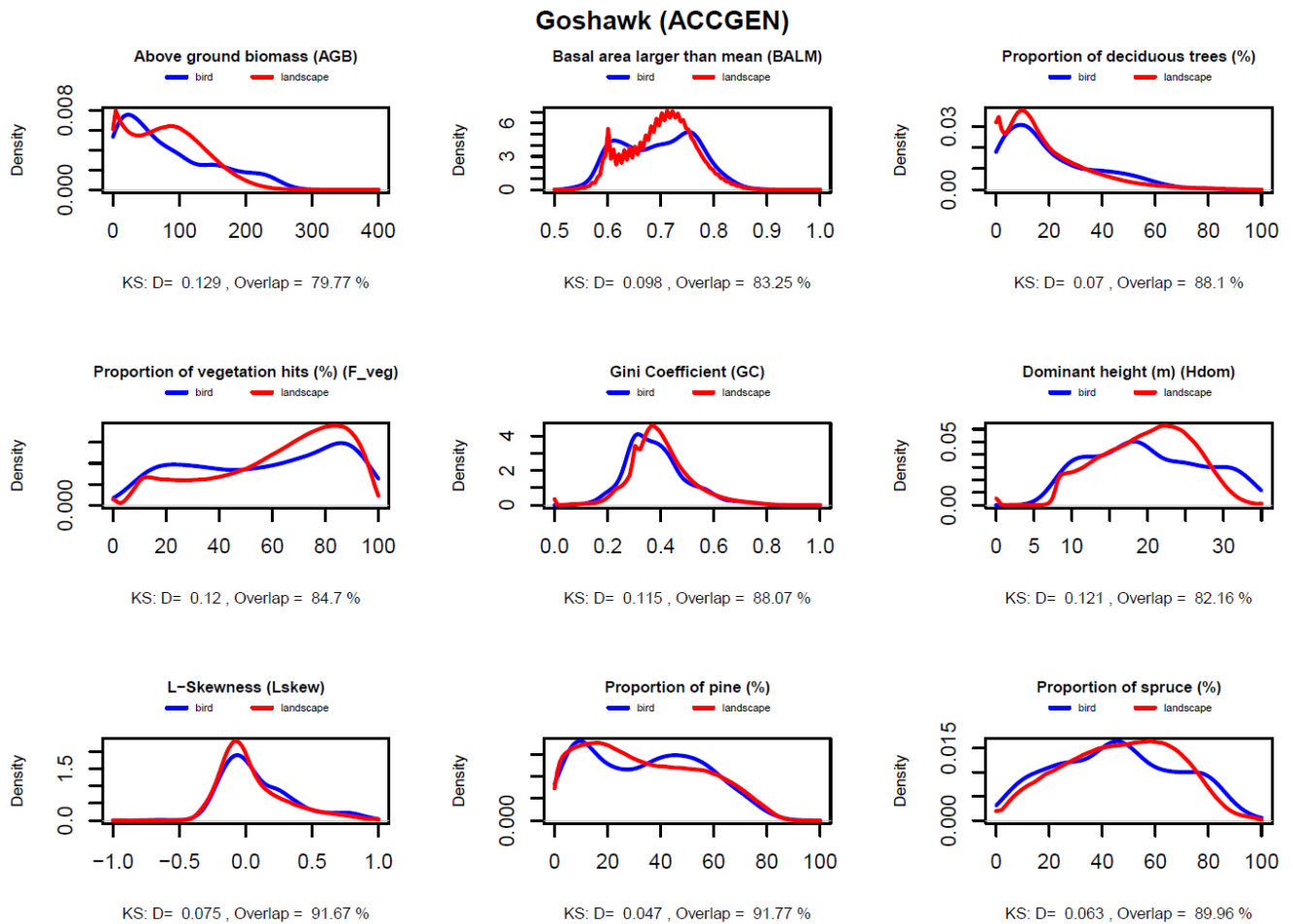
283 3.1 Landscape structure according to forest parameters

284
 285 The density curves computed at the landscape level are presented in red in figure 4 and ALS based
 286 forest parameters presented in Table 3. The landscape of the study area is highly covered by vegetation
 287 (*F_veg*). The largest part of the study area has quite a closed canopy structure (*Lskew*). Tree biomass is
 288 less than 100 ton/ha in the majority of the area (*AGB*) and tree height is typically between 20–25 meters
 289 (*Hdom*). Spruce is the dominating tree species (*Spruce %*), but pine trees are also common (*Pine %*).
 290 The proportion of deciduous trees is low (*Deciduous %*). The two three-dimensional parameters show
 291 that the largest part of the forests is even-sized (*GC*), but also that the area includes structurally complex
 292 forests with a heterogeneous tree composition (*BALM*).
 293
 294
 295
 296

297 **Table 3.** Description of the landscape (excluding water and urban areas) in the light of the ALS based
 298 forest parameters (landscape curve in red in Figure 4). See Table 2 for parameters descriptions.

Parameter	How indicator defines study area?
AGB	Significant amount of areas where there is no tree biomass (fields, clear cuts). Forest biomass is mostly below 100 ton/ha.
BALM	Two peaks show that there are structurally complex forests with understorey developed underneath the dominant canopies (lower peak), but that the greater share of forests have a single-layered vertical structure of vegetation (higher peak).
Deciduous (%)	The proportion of deciduous trees is quite low. There are only a few purely deciduous forest patches (> 80%) pixels.
F_veg	The proportion of vegetation hits shows that vegetation cover is high across the study area. The high peak shows that vegetation is dense in a large share of the area.
GC	Values below 0.5 indicate that the forests are mostly structurally homogenous.
Hdom (m)	The largest share of the forests is 20–25 meters in height.
Lskew	Negative skewness indicates closed canopy structure and applies to majority of the study area.
Pine (%)	There is rather high amount of forest with low percentage of pine. Pure pine forests (< 80%) are rare, but not as rare as pure deciduous forests.
Spruce (%)	Spruce is the dominant tree species in the area.

299



300

301 **Figure 4.** An example of a bird profile that was created for each species. In this case the 30 meter
 302 buffers around goshawk observation points (blue line) are compared with the landscape (red line).

303

304

305 3.2 Bird species' habitat preferences

306

307 Differences between the bird habitat preferences and the landscape characteristics were detected by
 308 calculating the D-values (Table 4) and overlapping metrics (Table 5). The results can be interpreted from
 309 several angles. One is to look over all the nine forest parameters and see which species' habitat use
 310 differ the most from the range of available habitat. The thrush nightingale's (LUSLUS) average D-value is
 311 clearly the highest (0.27) and its overlapping percentage lowest (72.3). The next two species ranked the
 312 same way are honey buzzard (PERAPI; 0.22 / 78.6%) and three-toed woodpecker (PICTRI; 0.20 /
 313 80.2%). At the other end of the spectrum the two grouse species, western capercaillie (TETURO; 0.07 /
 314 91.2%) and hazel grouse (BONBON; 0.05 / 93.3%), do not show much deviation from the available
 315 habitat at all. The D-values and overlapping percentages rank the species mostly in the same order.

316

317 The species can also be ranked by looking at one forest parameter at a time. Four species, for example,
 318 receive a D-value of over 0.2 for dominant height. These are three-toed woodpecker (PICTRI; 0.34), red-
 319 breasted flycatcher (FICPAR; 0.29), thrush nightingale (LUSLUS; 0.23) and greenish warbler (PHYDES;
 320 0.20). However, to understand how these species' habitat preference differs from the available habitat,

321 one needs to look at the species specific distribution curves (Figure 5 and Appendix I). The curves show
 322 that three of the species deviate from the available habitat distribution (landscape) towards taller trees
 323 and one, the thrush nightingale, towards shorter trees.

324

325

326 **Table 4.** D-values for a 30 m buffer. Higher values indicate higher maximum difference between the
 327 distributions. Table cells in darker colours indicate greater differences between the habitat used by the
 328 bird species and that of the whole study area.

329

Bird (30 m)	AGB	BALM	Deciduous %	f_veg	GC	Hdom	LSKEW	Pine %	Spruce %
ACCGEN	0,129	0,097	0,07	0,12	0,115	0,121	0,075	0,047	0,063
ACCNIS	0,222	0,202	0,105	0,198	0,072	0,153	0,064	0,106	0,111
AEGCAU	0,203	0,163	0,179	0,198	0,038	0,075	0,141	0,048	0,133
AEGFUN	0,153	0,062	0,037	0,197	0,119	0,09	0,195	0,072	0,052
ASIOTU	0,265	0,214	0,11	0,249	0,172	0,137	0,201	0,122	0,216
BONBON	0,038	0,052	0,022	0,052	0,082	0,054	0,062	0,054	0,044
BUBBUB	0,204	0,182	0,14	0,237	0,16	0,132	0,124	0,186	0,133
BUTBUT	0,192	0,149	0,178	0,199	0,057	0,11	0,088	0,147	0,086
CAPEUR	0,206	0,076	0,212	0,208	0,067	0,106	0,219	0,237	0,171
DENMIN	0,156	0,091	0,313	0,137	0,097	0,061	0,089	0,084	0,231
DRYMAR	0,146	0,086	0,11	0,169	0,066	0,041	0,138	0,041	0,081
FICPAR	0,241	0,199	0,137	0,151	0,035	0,293	0,175	0,224	0,138
GLAPAS	0,103	0,043	0,156	0,134	0,092	0,114	0,092	0,065	0,071
HIPICT	0,17	0,129	0,184	0,19	0,114	0,033	0,084	0,084	0,234
JYNTOR	0,264	0,223	0,107	0,258	0,117	0,119	0,124	0,064	0,13
LUSLUS	0,367	0,322	0,167	0,336	0,285	0,232	0,206	0,179	0,354
PERAPI	0,316	0,234	0,174	0,335	0,154	0,193	0,251	0,141	0,142
PHYDES	0,053	0,068	0,095	0,069	0,059	0,201	0,066	0,155	0,061
PICCAN	0,118	0,075	0,12	0,12	0,101	0,066	0,116	0,053	0,054
PICTRI	0,292	0,287	0,186	0,128	0,078	0,339	0,174	0,224	0,13
STRALU	0,153	0,084	0,15	0,179	0,069	0,058	0,124	0,063	0,103
STRURA	0,188	0,088	0,124	0,222	0,081	0,057	0,148	0,05	0,086
TETRIX	0,262	0,177	0,152	0,305	0,069	0,142	0,222	0,109	0,091
TETURO	0,101	0,025	0,082	0,064	0,052	0,076	0,044	0,101	0,063
TURVIS	0,094	0,05	0,144	0,121	0,067	0,092	0,095	0,172	0,104

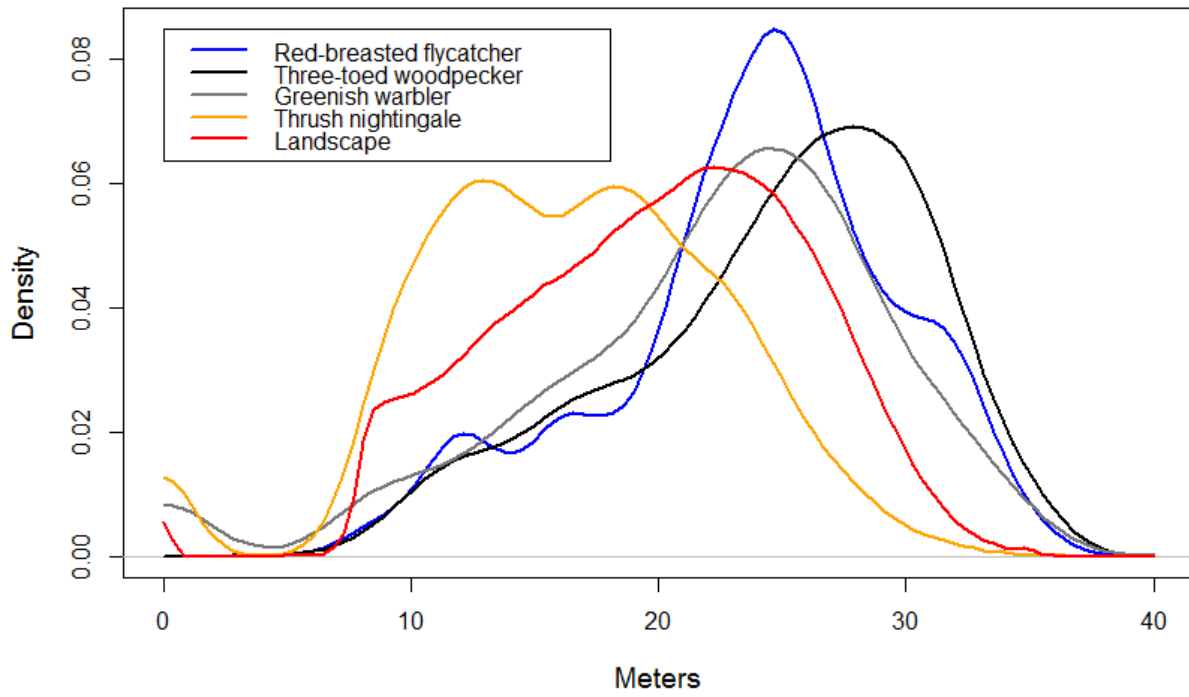
330

331

332 **Table 5.** Overlapping values for a 30 m buffer (%). Lower percentages indicate higher difference
 333 between the distributions. Table cells in darker colours indicate greater differences between the habitat
 334 used by the bird species and that of the whole study area.
 335

Bird (30 m)	AGB	BALM	Deciduous %	F_veg	GC	Hdom	LSKEW	Pine %	Spruce %
ACCGEN	79.77	83.51	88.10	84.70	88.07	82.16	91.68	91.77	89.96
ACCNIS	78.76	78.93	85.56	78.59	90.84	85.56	90.15	87.89	88.80
AEGCAU	80.84	81.59	77.35	79.38	91.16	92.94	86.71	91.64	87.47
AEGFUN	85.86	87.62	89.64	78.93	83.27	89.53	81.02	93.21	90.59
ASIOTU	74.45	78.28	80.76	75.24	84.10	87.65	79.44	84.84	79.04
BONBON	93.18	90.45	94.36	92.34	91.20	93.87	94.14	94.81	95.05
BUBBUB	81.40	73.76	84.20	75.36	83.89	81.30	86.39	85.42	85.75
BUTBUT	82.26	81.17	83.64	78.12	93.31	86.97	90.39	84.59	92.04
CAPEUR	81.12	85.70	80.21	78.63	89.20	84.48	74.97	76.40	83.07
DENMIN	84.93	86.74	64.34	85.27	88.11	92.72	89.61	91.34	77.64
DRYMAR	85.05	85.88	82.79	82.20	92.08	93.60	86.02	91.93	91.12
FICPAR	74.71	79.75	82.39	86.70	94.57	72.60	83.81	78.90	86.62
GLAPAS	90.25	91.33	78.84	84.81	88.39	89.42	91.09	94.26	90.79
HIPICT	84.11	84.25	73.73	80.07	82.96	95.85	92.06	88.74	77.32
JYNTOR	75.05	76.59	81.04	73.74	86.79	88.47	87.23	88.04	87.68
LUSLUS	64.34	68.08	77.30	66.16	73.17	77.50	77.95	80.20	65.58
PERAPI	69.96	76.28	82.51	66.99	84.29	82.91	75.59	83.75	86.55
PHYDES	90.43	86.48	82.80	91.61	90.21	78.78	88.96	86.16	89.90
PICCAN	87.20	86.71	82.74	86.66	87.68	91.87	88.77	94.92	93.70
PICTRI	70.48	70.54	80.42	89.91	91.33	67.55	83.54	79.43	88.37
STRALU	85.48	85.02	79.90	81.41	87.03	94.29	87.01	92.34	90.32
STRURA	82.25	87.77	86.34	77.06	89.60	94.28	86.73	90.04	86.99
TETRIX	74.57	81.86	85.08	69.69	91.14	86.45	77.49	87.22	90.38
TETURO	88.60	91.79	90.82	92.09	91.40	90.54	95.49	88.95	91.67
TURVIS	84.90	88.76	86.28	88.80	90.89	87.33	85.99	82.13	90.83

336
 337



338

339 **Figure 5.** Density curves for dominant height for the habitat used by four species; three-toed
 340 woodpecker, red-breasted flycatcher, thrush nightingale and greenish warbler as well as the landscape
 341 (red line).
 342

343 4 DISCUSSION

344

345 4.1 Success of habitat modelling

346

347 Our results conform to previous knowledge of the bird species' habitat preferences. The habitat
348 characteristics of the buffers around the observation points of specialist species with strict habitat
349 requirements (e.g. three-toed woodpecker; Angelstam & Mikusiński 1994) differed generally more from
350 the characteristics of the entire study area than those of more generalist species or, putting it in other
351 words, species of which preferred habitat was common in all parts of our study area (e.g. hazel grouse;
352 Åberg *et al.* 2003). The three-toed woodpecker clearly chose those forest patches within the study area
353 that are characterised by tall trees, high above ground biomass and heterogeneous structures with
354 structurally complex understoreys (*BALM*). The hazel grouse did not show any clear preference in
355 relation to any of the nine ALS parameters we employed.

356

357 However, in the case of some bird species the analysis did not show habitat specialisation although
358 some specialisation was expected based on previous knowledge. We found several reasons for this.
359 First, species that have a large territory and/or home range (> 100 hectares; see Table 1) showed the
360 least difference between the forest characteristics generally available in the landscape and the habitat
361 they used. In a forest landscape consisting of relatively small habitat parcels, the individuals with a large
362 home range are ultimately observed in many different habitat types. Due to private forest ownership and
363 the dominant even-aged stands forest management practice the forests in our study area have a high
364 level of small-scale heterogeneity. In southern Finland the average size of managed stands is
365 approximately 1.2 hectares (Parviainen & Västilä 2011). As a rule of thumb the home range size of a
366 species increases with its body size, and carnivores have larger ranges than herbivores (e.g. Reiss
367 1988). Accordingly, large birds of prey like hawks and owls (e.g. ACCGEN, AEGFUN) are among
368 species that in our results showed least habitat preferences. An analysis on a larger landscape scale
369 (i.e. larger grids and buffers) would probably be better suited to study birds with large home ranges.

370

371 Second, some characteristics related to the CS observations themselves are quite likely reflected in the
372 results. Compared with other species in the analysis, the majority of the observations of birds of prey
373 were made from afar. The observations of four owl species – Eurasian eagle-owl, Ural owl, tawny owl,
374 and boreal owl – together with northern goshawk had the highest average distances between the bird
375 and observer (75th observation distance percentile approx. 400–500 metres; see Table 1). Further, a
376 large part of the owl observations were auditive; the birds were often observed hooting in the dark. This
377 makes the bird locations even more inaccurate. The finding supports eliminating such inaccurate
378 observations from the dataset.

379

380 All but one large sized species with a large territory were observed at a great distance. In the case of the
381 one exception, western capercaillie, the distance between the bird and observer was normally very short

382 (75th percentile 35 metres). As a result, there should not be any location error arising from a long
383 observation distance, yet our data indicate almost no difference between the characteristics of the
384 available habitat and those of the sites where capercaillies were observed. In contrast with our results
385 that showed no preference for tall trees of high biomass, for example, the western capercaillie has even
386 been considered as a species preferring continuous old-growth forest or even an old-growth forest
387 specialist (Helle 1985; Saniga 2003; Virkkala & Rajasärkkä 2006; Gregersen & Gregersen 2008). This
388 can be due to a number of reasons. According to Sirkiä *et al.* (2011) the capercaillie prefers fine-grained
389 mosaic and heterogeneity of forest in its spring time lekking areas. Fragmentation of very dense forests
390 may produce more space for capercaillie males in their lekking sites. Moreover, capercaillies may, for
391 example, prefer younger stands and moist edge habitats with bilberry during the breeding and fledgling
392 (post-breeding) period (see Miettinen *et al.* 2008) than at other parts of their annual cycle. The result
393 may also be an artefact created by the situation in which both birdwatchers and capercaillies prefer
394 forest roads: birdwatchers for the ease of movement and capercaillies for collecting small pebbles for
395 their gizzard.

396
397 Bird species with a small or medium sized territory, especially those that were most often observed from
398 a short distance, stands out by showing clearest signs of habitat preference in relation to the landscape
399 available. Three bird species, in particular, had high preference for the characteristics of mature or old-
400 growth stands (esp. dominant tree height): the three-toed woodpecker, red-breasted flycatcher and
401 greenish warbler. The two latter ones are small passerine birds with a small territory. Normally they are
402 observed singing close by and at the core of their breeding habitat (75th observation distance percentile
403 50 and 40 metres, respectively). The three-toed woodpecker, in contrast, is a medium sized bird with a
404 medium sized territory. However, it is also normally observed close by (75th observation distance
405 percentile 28 metres) which makes the observations very accurate. These characteristics together with
406 the fact that these species are known as habitat specialists of mature and old-growth coniferous forests
407 (Tiainen 1980; Virkkala *et al.* 1994; Roberge *et al.* 2008; Pakkala *et al.* 2014), makes the identification of
408 their habitats by means of ALS derived parameters successful.

409
410 Other species which results seem to conform to previous knowledge of their ecology include species
411 preferring deciduous stands and edge habitats. Alike three-toed woodpecker, the lesser spotted
412 woodpecker has a medium sized territory and is normally observed at short range. Out of all species in
413 the analysis, the lesser spotted woodpecker showed the greatest preference for deciduous stands, which
414 is well in accordance with previous studies (Roberge & Angelstam 2006; Roberge *et al.* 2008). Other
415 species preferring deciduous forests included the icterine warbler and thrush nightingale. The latter
416 received surprisingly low values for parameters describing the volume of vegetation (*AGB* and *F_veg*,
417 especially). This is likely to be due to the fact that thrush nightingales are normally observed singing
418 close to an edge and therefore the buffer around the observation point is likely to include habitat pixels
419 with zero values (i.e. agricultural lands, urban areas, water and so on). We ran a test of the individual
420 thrush nightingale buffers and found out that nearly half of the pixels had zero values for the number of

421 vegetation hits (F_veg), for example. This is in high contrast with the habitat pixels inside red-breasted
422 flycatcher buffers of which only 15% had zero values. A similar pattern probably holds true for the
423 Eurasian wryneck and long-eared owl both of which are species that also prefer edge habitats (Väisänen
424 *et al.* 1998).

425

426 There were, however, few species for which the results found were in discordance with their known
427 ecology and habitat preferences. These results are likely to be artefacts caused by the combination of
428 the species' ecological traits and observation factors. Our results regarding the honey buzzard, for
429 example, suggest that the species would prefer habitats with low biomass, low lying trees, and sparse
430 vegetation. This is not in accordance with what is known about the breeding habitats of honey buzzard:
431 the species nests in mature mixed stands with Norway spruce, birches and European aspen (Björklund
432 *et al.* 2015). However, the CS data are mostly of individuals seen hunting or displaying over their
433 territory, and not from their nest site where the species is very hard to observe. Therefore, the results
434 relate more to the characteristic of honey buzzard's hunting grounds than its core territory.

435

436 **4.2 Applicability of the methodological setup**

437

438 Our results indicate the usefulness of structural ALS based forest parameters for studying bird ecology
439 as previous studies have already revealed (Hill *et al.* 2004; Bradbury *et al.* 2005; Clawges *et al.* 2008;
440 Graf *et al.* 2009; Goetz *et al.* 2010). Our set of structural forest parameters revealed habitat preferences
441 in the case of several bird species which implies potential for using them in habitat models. However, CS
442 data should not be used with models/tools that require the assumption of the dataset being independent
443 sample from the unknown probability distribution. Particularly *BALM* (basal area larger than mean) and
444 *Hdom* (dominant height) revealed differences among multiple bird species. While *Hdom* describes the
445 development of the higher canopy, *BALM* expresses the ingrowth of understorey vegetation under the
446 dominant tree crowns. Previous studies have shown that joint variables explain better patterns related to
447 bird habitat selection (Heikkinen *et al.* 2004). These particular parameters, *Hdom* and *BALM*, were both
448 obtained using a combination of ALS and field plots. An alternative to *Hdom* derived from ALS data only
449 could be the height of the CHM which, although we employed it in an earlier study (Vihervaara *et al.*
450 2015), it was ruled out among the parameters reported in this article because *Hdom* showed more
451 relevant results. On the other hand, no ecology studies have considered the use of *BALM* before, and its
452 use has so far been restricted to forest science (Gove 2004). Although many alternatives for estimating
453 *BALM* from ALS have been proved reliable (Valbuena *et al.* 2014), current research has not yet found
454 any alternatives to *BALM* derived from ALS data only.

455

456 Our set of structural forest parameters revealed habitat preferences in the case of several bird species
457 which implies potential for using them in habitat models. However, one essential assumption of
458 modelling species distributions from presence-only records is that the data are an independent sample
459 from the species' unknown probability distribution of occurrence (Gomes *et al.* 2018). Our CS data do

460 not fulfil this requirement, and we believe that in general most CS data of birds do not meet this
461 condition. This may be a serious drawback from the monitoring point of view. It is essential to understand
462 the characteristics and limitations of CS data, and assumptions and pitfalls of the used species
463 distribution modelling method.

464

465 The ALS data are already freely available and the data are expected to cover the entire land area of
466 Finland by 2019 (National Land Survey 2016), which creates an opportunity to apply the same
467 methodology elsewhere. Multiple European countries have already reached the 100% ALS data
468 coverage. Because of the increasing availability of ALS data, we also tested the reliability of indicators
469 derived directly from the ALS data (e.g. Nelson *et al.* 1984; Lefsky *et al.* 2002; Miura & Jones 2010;
470 Valbuena *et al.* 2017) without the support of field information for modelling and estimating forest
471 parameters (e.g. Næsset 2002; Nelson *et al.* 2007; Asner & Mascaro 2014). Among many options that
472 were initially tested (data not shown), we found two variables– *Lskew* and *F_veg* – that could be
473 employed as proxies of forest parameters and linked to the bird species considered in this study. While
474 *F_veg* (the proportion of ALS first echoes reflected from vegetation) can be considered as a ALS proxy
475 for canopy cover (Nelson *et al.* 1984), *Lskew* (a ratio of L-moments showing the skewness of ALS echo
476 height distribution) tells us about the availability of light passing through the dominant tree crowns
477 (Valbuena *et al.* 2017). The reason for obtaining relevant results from these in particular is an indication
478 that these characteristics can be among the most relevant in defining bird habitat preferences. In
479 contrast with the remaining forest parameters requiring the acquisition of field plots, variables obtained
480 directly from the ALS data could substantially increase the potential uses of these data. It can, as well as
481 to decrease costs, enable us to measure the environment directly from ALS for the purpose of habitat
482 characterization (Vierling *et al.* 2008) – in some cases as accurately as when including field data in the
483 analysis. Further research should be devoted to investigating whether the inclusion of field data from
484 forest plots could be critical to the habitat characterization of any particular species.

485

486 One important question to consider is the size of the buffer around each observation point. The buffer
487 has to be large enough to deal with the inaccuracies of the CS observation data. The accuracy by which
488 bird enthusiasts mark a spot on a digital map when reporting a bird observation is rarely very high.
489 Choosing one or only a few pixels as the bird's habitat (location) would quite likely also exceed the
490 accuracy of the ALS data. On the other hand, increasing the buffer radius around each observation point
491 beyond 50 metres (0.8 hectares in area) leads to a decrease in the difference between the buffers and
492 the whole landscape. Mainly due to private ownership, which in Finland typically implies small forest
493 stands (Parviainen & Västilä 2011) and large structural heterogeneity (Valbuena *et al.* 2016), buffers
494 larger than one hectare begin to include parts of other habitat patches than of that where the bird was
495 actually observed. Determining the suitable buffer size would require further investigation.

496

497 It is important to keep in mind that the CS data that we used do not represent a probability sample and
498 therefore it should be carefully considered whether statistical testing should be applied for the data, and

499 whether statistical significance tests would be valid for these purposes. The approach of combining CS
500 and ALS (with or without field plots) should always be considered species-specifically in relation to
501 species ecology and the quality of observations (Pettorelli *et al.* 2014; Vihervaara *et al.* 2015).
502 Nevertheless, we consider that combinatory studies with CS and ALS data holds considerable potential
503 to be used in revealing the suitable habitats of many specialised species across the landscape. Because
504 of this, similar analyses would also be applicable to detecting areas of special conservation value (Farrell
505 *et al.* 2013; Rechsteiner *et al.* 2017), and in biodiversity conservation in general (Butchart *et al.* 2010).
506 Further, proxy maps for biodiversity in landscape-scale ecosystem service assessments (Nelson *et al.*
507 2009; Kopperoinen *et al.* 2014) can be created by similar means. Furthermore, the tested forest
508 parameters have many potential applications in ecosystem service analyses (Pettorelli *et al.* 2014). Our
509 study will be taken further by running the species distribution maps based on the results of this study.

510

511 Citizen science, with careful data refinement and consideration of the limits of CS observations, showed
512 much potential for being used in further analyses in predicting the potential habitats of some forest bird
513 species. These were species that were typically observed close-by and at the core of their territories.
514 Using these types CS observations for predicting the habitats of species with large home ranges, for
515 example, would require a somewhat different approach (larger buffers, more careful consideration of the
516 observation times and other circumstances etc.). Means of coping with the uneven distribution
517 observation effort could also be developed. A compensation factor for roads and nearness of towns and
518 cities could be devised, for example. The quality of CS observations could also be raised by some quite
519 easy means. Location errors in observations could be overcome by in-situ reporting with smart phone
520 applications using GPS (already partly in place) and birdwatchers could be directed to report more useful
521 data by redesigning reporting forms (now, in Tiira, only giving the location of the observer in mandatory,
522 for example).

523

524

525 **5 CONCLUSIONS**

526

527 This study showed the potential of combining CS bird observation data with forest parameters derived
528 from ALS with field measurements or ALS alone in detecting the habitats preferred by several forest
529 dwelling bird species. In general, bird species with small or medium territories reflected better their true
530 habitat preferences than species with large territories. Also the way how CS observations are made and
531 recorded plays a role – if the observation is made from a close distance and if the species is rather seen
532 than only heard the coordinates can be expected to be more accurate. We therefore conclude that the
533 precision of species location in data collected via CS must be in accordance with the spatial resolution of
534 the RS data for this type of analysis to be useful. We emphasize that expertise of species ecology is
535 essential as well as a thorough understanding of the CS data including factors related to reporting
536 observations. On the whole, the use of CS with ALS holds great promise for modelling bird habitats for

537 specialised species and can be expected to contribute to ecosystem mapping and assessment also on a
538 larger scale.

539

540 **ACKNOWLEDGEMENTS**

541

542 Local BirdLife partners Tavastia Proper Ornithological Society/Juhani Kairamo and Pirkanmaa
543 Ornithological Society/Jukka T. Helin. The CLIMES project (Academy of Finland), FRESHABIT-LIFE IP
544 467 project, MAES Finland project, The IBC-CARBON project funded by The Strategic Research Council
545 (SRC) at the Academy of Finland (Grant n:o 312559). We thank anonymous reviewers for very
546 productive feedback in preparation of the manuscript.

547

548

549 **REFERENCES**

550

551 Åberg, J., Swenson, J.E. & Angelstam, P. (2003) The habitat requirements of hazel grouse (*Bonasa*
552 *bonasia*) in managed boreal forest and applicability of forest stand descriptions as a tool to identify
553 suitable patches. *Forest Ecology and Management*, **175**, 437–444.

554

555 Angelstam, P. & Mikusiński, G. (1994) Woodpecker assemblages in natural and managed boreal and
556 hemiboreal forest — a review. *Annales Zoologici Fennici*, **31**, 157–172.

557

558 Asner, G.P. & Mascaro, J. (2014) Mapping tropical forest carbon: Calibrating plot estimates to a simple
559 LiDAR metric. *Remote Sensing of Environment*, **140**, 614–624.

560

561 Axelsson P. (2000) DEM generation from laser scanner data using adaptive TIN models. *International*
562 *Archives of Photogrammetry and Remote Sensing*, **33** (Part B4), 110–117.

563

564 Barbet-Massin, M., Thuiller, W., & Jiguet, F. (2012) The fate of European breeding birds under climate,
565 land-use and dispersal scenarios. *Global Change Biology*, **18**(3), 881–890.

566

567 Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D.,
568 Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F. & Pecl, G.T. (2014) Statistical solutions for error and
569 bias in global citizen science datasets. *Biological Conservation*, **173**, 144–154.

570

571 Björklund H., Valkama J., Tomppo E. & Laaksonen T. (2015) Habitat effects on the breeding
572 performance of three forest-dwelling hawks. *Plos One*, **10**(9), e0137877.

573

574 Bradbury, R.B., Hill, R.A., Mason, D.C., Hinsley, S.A., Wilson, J.D., Balzter, H., Anderson, G.Q.,
575 Whittingham, M.J., Davenport, I.J. & Bellamy, P.E. (2005) Modelling relationships between birds and
576 vegetation structure using airborne LiDAR data: a review with case studies from agricultural and
577 woodland environments. *Ibis*, **147**(3), 443–452.

578

579 Brokaw, N. & Lent, R. (1999) Vertical structure. In Hunter, M. (Eds.) *Maintaining Biodiversity in Forest*
580 *Ecosystems*, pp. 373–399, Cambridge Univ. Press, Cambridge, UK.

581

582 Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie,
583 J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M.,
584 Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory,
585 R.D., Hockings, M., Kapos, V., Lamarque, J.F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L.,
586 Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik,
587 B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.C. & Watson,
588 R. (2010) Global Biodiversity: Indicators of Recent Declines. *Science*, **328**, 1164–1168.

589
590 Clawges R., Vierling K., Vierling L., Rowell E. (2008). The use of airborne lidar to assess avian species
591 diversity, density, and occurrence in a pine/aspen forest. *Remote Sensing of Environment* **112**(5), 2064–
592 2073.

593
594 Cohen, W.B. & Goward, S.N. (2004) Landsat's role in ecological applications of remote sensing.
595 *Bioscience*, **54**(6), 535–545.

596
597 Conrad, C.C. & Hilchey, K.G. (2011) A review of citizen science and community-based environmental
598 monitoring: issues and opportunities. *Environmental monitoring and assessment*, **176**(1-4), 273–291.
599

600 Davies, A.B. & Asner, G.P. (2014) Advances in animal ecology from 3D-LiDAR ecosystem mapping.
601 *Trends in ecology & evolution*, **29**(12), 681–691.

602
603 Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips, T. & Purcell, K. (2012)
604 The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in*
605 *Ecology and the Environment*, **10**(6), 291–297.

606
607 Dickinson, J.L., Zuckerberg, B. & Bonter, D.N. (2010) Citizen science as an ecological research tool:
608 challenges and benefits. *Annual review of ecology, evolution and systematics*, **41**, 149–72.

609
610 Farrell, S.L., Collier, B.A., Skow, K.L., Long, A.M., Campomizzi, A.J., Morrison, M.L., Hays, K.B. &
611 Wilkins, R.N. (2013) Using LiDAR-derived vegetation metrics for high-resolution, species distribution
612 models for conservation planning. *Ecosphere*, **4**(3), 1–18.

613
614 Frishkoff, L.O., Karp, D.S., M'Gonigle, L.K., Mendenhall, C.D., Zook, J., Kremen, C., Hadly, E.A. & Daily,
615 G.C. (2014) Loss of avian phylogenetic diversity in neotropical agricultural systems. *Science*, **345**(6202),
616 1343–1346.

617
618 GEO BON (2015) Global Biodiversity Change Indicators: Model-based integration of remote-sensing & in
619 situ observations that enables dynamic updates and transparency at low cost. V1.1.

620
621 Goetz, S., Steinberg, D., Dubayah, R. & Blair, B. (2007) Laser remote sensing of canopy habitat
622 heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote*
623 *Sensing of Environment*, **108**(3), 254–263.

624
625 Goetz, S.J., Steinberg, D., Betts, M.G., Holmes, R.T., Doran, P.J., Dubayah, R. & Hofton, M. (2010)
626 Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird. *Ecology*, **91**(6),
627 1569–1576.

628
629 Gomes, V.H., Ijff, S.D., Raes, N., Amaral, I.L., Salomão, R.P., Coelho, L.S., Matos, F.D.A., Castilho,
630 C.V., Lima Filho, D.A., López, D.C. & Guevara, J.E. (2018). Species distribution modelling: contrasting
631 presence-only models with plot abundance data. *Scientific reports*, **8**(1), 1003.

632
633 Gove J.H. (2004) Structural stocking guides: a new look at an old friend. *Canadian Journal of Forest*
634 *Research*, **34**, 1044–1056.

635
636 Graf, R.F., Mathys, L., & Bollmann, K. (2009) Habitat assessment for forest dwelling species using
637 LiDAR remote sensing: Capercaillie in the Alps. *Forest Ecology and Management*, **257**(1), 160–167.
638

639 Gregersen, H. & Gregersen, F. (2008) Old bilberry forest increases likelihood of Capercaillie Tetrao
640 urogallus lek occupancy in Southern Norway. *Ornis Norvegica*, **31**, 105–115.

641
642 Gregory, R.D. & van Strien, A. (2010) Wild bird indicators: using composite population trends of birds as
643 measures of environmental health. *Ornithological Science*, **9**, 3–22.

644
645 Hagar, J.C., Eskelson, B.N., Haggerty, P.K., Nelson, S.K. & Vesely, D.G. (2014) Modeling marbled
646 murrelet (*Brachyramphus marmoratus*) habitat using LiDAR-derived canopy data. *Wildlife Society*
647 *Bulletin*, **38**(2), 237–249.

648
649 Haila Y., Nicholls A.O., Hanski I.K. & Raivio S. (1996) Stochasticity in bird habitat selection: year-to-year
650 changes in territory locations in boreal forest bird assemblage. *Oikos*, **76**, 536–552.
651
652 Heikkinen, R.K., Luoto, M., Virkkala, R. & Rainio, K. (2004) Effects of habitat cover, landscape structure
653 and spatial variables on the abundance of birds in an agricultural–forest mosaic. *Journal of Applied*
654 *Ecology*, **41**(5), 824–835.
655
656 Helle P. (1985) Effects of forest fragmentation on bird densities in northern boreal forests. *Ornis Fennica*,
657 **62**, 35–41.
658
659 Hill, R.A. & Thomson, A.G. (2005) Mapping woodland species composition and structure using airborne
660 spectral and LiDAR data. *International Journal of Remote Sensing*, **26**(17), 3763–3779.
661
662 Hinsley, S., Hill, R., Fuller, R., Bellamy, P. & Rothery, P. (2009) Bird species distributions across
663 woodland canopy structure gradients. *Community Ecology*, **10**(1), 99–110.
664
665 Hosking, J. R. (1990) L-moments: analysis and estimation of distributions using linear combinations of
666 order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 105–124.
667
668 Jonzén, N., Lindén, A., Ergon, T., Knudsen, E., Vik, J.O., Rubolini, D., Piacentini, D., Brinch, C., Spina,
669 F., Karlsson, L., Stervander, M., Andersson, A., Waldenström, J., Lehikoinen, A., Edvardsen, E., Solvang,
670 R. & Stenseth, N.C. (2006) Rapid advance of spring arrival dates in long-distance migratory birds.
671 *Science* **312**, 1959–1961.
672
673 Kerr, J. T., & Ostrovsky, M. (2003) From space to species: ecological applications for remote sensing.
674 *Trends in ecology & evolution*, **18**(6), 299–305.
675
676 Kopperoinen, L., Itkonen, P. & Niemelä, J. (2014) Using expert knowledge in combining green
677 infrastructure and ecosystem services in land use planning: an insight into a new place-based
678 methodology. *Landscape Ecology*, **29**, 1361–1375.
679
680 Laaksonen, T. & Lehikoinen, A. (2013) Population trends of boreal birds: continuing declines in long-
681 distance migrants, agricultural and northern species. *Biological Conservation* **168**, 99–107.
682
683 Laasasenaho, J. (1982) Taper curve and volume functions for pine, spruce and birch. *Communicationes*
684 *Instituti Forestalis Fenniae*, **108**, 1–74.
685
686 Lefsky, M.A., Cohen, W.B., Parker, G.G. & Harding, D.J. (2002) Lidar Remote Sensing for Ecosystem
687 Studies Lidar, an emerging remote sensing technology that directly measures the three-dimensional
688 distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of
689 particular interest to forest, landscape, and global ecologists. *BioScience*, **52**(1), 19–30.
690
691 Lehikoinen, A., Jaatinen, K., Vähätalo, A.V., Clausen, P., Crowe, O., Deceuninck, B., Hearn, R., Holt,
692 C.A., Hornman, M., Keller, V., Nilsson, L., Langendoen, T., Tománková, I., Wahl, J. & Fox, A.D. (2013)
693 Rapid climate driven shift in three wintering waterbird species. *Global Change Biology* **19**, 2071–2081.
694
695 MacArthur, R.H. & MacArthur, J.W. (1961) On bird species diversity. *Ecology*, **42**(3), 594–598.
696
697 Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyypä, J., & Pitkänen, J. (2005). Identifying and
698 quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest*
699 *ecology and management*, **216**(1-3), 41–50.
700
701 Melin, M., Matala, J. Mehtätalo, L., Tiilikainen, R., Tikkanen, O.-P., Maltamo, M., Puseenius, J. &
702 Packalén, P. (2014) Moose (*Alces alces*) reacts to high summer temperatures by utilising thermal
703 shelters in boreal forests – an analysis based on airborne laser scanning of the canopy structure at
704 moose locations. *Global Change Biology*, **20**(4), 1115–1125.
705

706 Melin, M., Mehtätalo, L., Miettinen, J., Tossavainen, S. & Packalen, P. (2016) Forest structure as a
707 determinant of grouse brood occurrence—An analysis linking LiDAR data with presence/absence field
708 data. *Forest Ecology and Management*, **380**, 202–211.

709

710 Meredith, M. & Ridout, M. (2016) Package ‘overlap’ - Estimates of Coefficient of Overlapping for Animal
711 Activity Patterns. Version 0.2.6.

712

713 Miettinen J., Helle P., Nikula A. & Niemelä P. (2008) Large-scale landscape composition and
714 capercaillie (*Tetrao urogallus*) density in Finland. *Annales Zoologici Fennici*, **45**, 161–173.

715

716 Miura, N., & Jones, S. D. (2010) Characterizing forest ecological structure using pulse types and heights
717 of airborne laser scanning. *Remote Sensing of Environment*, **114**(5), 1069–1076.

718

719 Müller, J., Moning, C., Baessler, C., Heurich, M. & Brandl, R. (2009). Using airborne laser scanning to
720 model potential abundance and assemblages of forest passerines. *Basic and Applied Ecology*, **10**(7),
721 671–681.

722

723 Næsset, E. (2002) Predicting forest stand characteristics with airborne scanning laser using a practical
724 two-stage procedure and field data. *Remote sensing of environment*, **80**(1), 88–99.

725

726 National Land Survey (2016) Laser scanning plan covers year 2019.
727 [http://www.maanmittauslaitos.fi/en/professionals/topographic-data/remote-sensing/laser-scanning/laser-](http://www.maanmittauslaitos.fi/en/professionals/topographic-data/remote-sensing/laser-scanning/laser-scanning-plan-covers-year-2019)
728 [scanning-plan-covers-year-2019](http://www.maanmittauslaitos.fi/en/professionals/topographic-data/remote-sensing/laser-scanning/laser-scanning-plan-covers-year-2019). Accessed 7.11.2016.

729

730 Nelson, R., Krabill, W., & MacLean, G. (1984) Determining forest canopy characteristics using airborne
731 laser data. *Remote Sensing of Environment*, **15**(3), 201–212.

732

733 Nelson, R. F., Hyde, P., Johnson, P., Emessiene, B., Imhoff, M. L., Campbell, R., & Edwards, W. (2007)
734 Investigating RaDAR–LiDAR synergy in a North Carolina pine forest. *Remote Sensing of Environment*,
735 **110**(1), 98–108.

736

737 Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, D., Chan, K., Daily, G.C.,
738 Goldstein, J., Kareiva, P.M. & Lonsdorf, E. (2009) Modeling multiple ecosystem services, biodiversity
739 conservation, commodity production, and tradeoffs at landscape scales. *Frontiers in Ecology and the*
740 *Environment*, **7**(1), 4–11.

741

742 Näslund. M. (1937) Skogsförskansanstaltens gallringsförsök i tall-skog Primärearbetning. *Meddelanden*
743 *från Statens Skogsförsöksanstalt*, **29**, 1–121.

744

745 Palminteri, S., Powell, G.V., Asner, G.P., & Peres, C.A. (2012) LiDAR measurements of canopy structure
746 predict spatial distribution of a tropical mature forest primate. *Remote Sensing of Environment*, **127**, 98–
747 105.

748

749 Pakkala T., Lindén A., Tiainen J., Tomppo E. & Kouki J. (2014) Indicators of forest biodiversity: which
750 bird species predict high breeding bird assemblage diversity in boreal forests at multiple spatial scales?
751 *Annales Zoologici Fennici*, **51**, 457–476.

752

753 Parvinen, J. & Västilä, S. (2011) State of Finland’s Forests 2011. Based on the Criteria and Indicators of
754 Sustainable Management. Ministry of Agriculture and Forestry 5a/2011. 98 pp.

755

756 Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H. & Turner, W. (2014) Satellite
757 remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology*, **51**(4),
758 839–848.

759

760 Pearce, J.L. & Boyce, M.S. (2006) Modelling distribution and abundance with presence-only data.
761 *Journal of Applied Ecology*, **43**, 405–412.

762

763 R Core Team (2016) R: A language and environment for statistical computing. R Foundation for
764 Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

765
766 Rechsteiner, C., Zellweger, F., Gerber, A., Breiner, F.T. & Bollmann, K. (2017) Remotely sensed forest
767 habitat structures improve regional species conservation. *Remote Sensing in Ecology and Conservation*,
768 **3**(4), 247-258.

769
770 Reiss, M. (1988) Scaling of home range size: Body size, metabolic needs and ecology. *Trends in*
771 *Ecology and Evolution*. **3**, 85–86.

772
773 Repola, J., Ojansuu, R., & Kukkola, M. (2007) Biomass functions for Scots pine, Norway spruce and
774 birch in Finland. *Working Papers of the Finnish Forest Research Institute*, **53**. Finnish Forest Research
775 Institute.

776
777 Roberge J.-M. & Angelstam P. (2006) Indicator species among resident forest birds - A cross-regional
778 evaluation in northern Europe. *Biological Conservation*, **130**, 134–147.

779
780 Roberge J.-M., Angelstam P. & Villard M.-A. (2008) Specialised woodpeckers and naturalness in
781 hemiboreal forests - Deriving quantitative targets for conservation planning. *Biological Conservation*,
782 **141**, 997–1012.

783
784 Rose, R.A., Byler, D., Eastman, J.R., Fleishman, E., Geller, G., Goetz, S., Guild, L., Hamilton, H.,
785 Hansen, M., Headley, R. & Hewson, J. (2015) Ten ways remote sensing can contribute to conservation.
786 *Conservation Biology*, **29**(2), 350–359.

787
788 Royle, J.A., Kéry, M., Gautier, R. & Schmid, H. (2007) Hierarchical spatial models of abundance and
789 occurrence from imperfect survey data. *Ecological Monographs*, **77**(3), 465–481.

790
791 Saino, N., Rubolini, D., Von Hardenberg, J., Ambrosini, R., Provenzale, A., Romano, M., & Spina, F.
792 (2010) Spring migration decisions in relation to weather are predicted by wing morphology among trans-
793 Mediterranean migratory birds. *Functional Ecology*, **24**(3), 658–669.

794
795 Saniga, M. (2003) Ecology of the capercaillie (*Tetrao urogallus*) and forest management in relation to its
796 protection in the West Carpathians. *Journal of Forest Science*, **49**, 229–239.

797
798 Simonson, W.D., Harriet, D.A. & Coomes, D.A. (2014) Applications of airborne lidar for the assessment
799 of animal species diversity. *Methods in Ecology and Evolution*, **5** 719–729.

800
801 Sirkiä S., Helle P., Lindén H., Nikula A., Norrdahl K., Suorsa P. & Valkeajärvi P. (2011) Persistence of
802 Capercaillie (*Tetrao urogallus*) lekking areas depends on forest cover and fine-grain fragmentation of
803 boreal forest landscapes. *Ornis Fennica*, **88**, 14–29.

804
805 Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D. & Kelling, S. (2009) eBird: a citizen-based
806 bird observation network in the biological sciences. *Biological Conservation*, **142**(10), 2282–2292.

807
808 Tiainen J. (1980) Habitat of the Greenish Warbler *Phylloscopus trochiloides* (In Finnish with an English
809 summary). *Lintumies*, **15**, 50–53.

810
811 Väisänen R.A., Lammi E. & Koskimies P. (1998) *Distribution, numbers and population changes of*
812 *Finnish breeding birds (In Finnish with an English summary)*. Otava, Helsinki, Finland.

813
814 Valbuena R. (2015) Forest structure indicators based on tree size inequality and their relationships to
815 airborne laser scanning. Dissertationes Forestales 205. Finnish Society of Forest Sciences.

816
817 Valbuena, R., Packalén, P., Martín-Fernández, S., & Maltamo, M. (2012) Diversity and equitability
818 ordering profiles applied to study forest structure. *Forest Ecology and Management*, **276**, 185–195.

819
820 Valbuena, R., Packalén, P., Mehtätalo, L., García-Abril, A., & Maltamo, M. (2013) Characterizing forest
821 structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser
822 scanning. *Canadian Journal of Forest Research*, **43**(11), 1063–1074.

823

- 824 Valbuena R., Vauhkonen J., Packalen P., Pitkänen J. & Maltamo M. (2014) Comparison of Airborne
825 Laser Scanning Methods for Estimating Forest Structure Indicators Based on Lorenz Curves.
826 *ISPRS Journal of Photogrammetry & Remote Sensing* **95**, 23–33.
827
- 828 Valbuena R., Eerikäinen K., Packalen P. & Maltamo M. (2016) Gini Coefficient Predictions from Airborne
829 Lidar Remote Sensing Display the Effect of Management Intensity on Forest Structure. *Ecological*
830 *Indicators*, **60**, 574–585
831
- 832 Valbuena, R., Maltamo, M., Mehtätalo, L. & Packalen, P. (2017) Key structural features of Boreal forests
833 may be detected directly using L-moments from airborne lidar data. *Remote Sensing of Environment*,
834 **194**, 437–446.
835
- 836 Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S. & Clawges, R.M. (2008) Lidar: shedding new
837 light on habitat characterization and modeling. *Frontiers in Ecology and the Environment*, **6**(2), 90–98.
838
- 839 Vierling, K.T., Swift, C.E., Hudak, A.T., Vogeler, J.C. & Vierling, L.A. (2014) How much does the time lag
840 between wildlife field-data collection and LiDAR-data acquisition matter for studies of animal
841 distributions? A case study using bird communities. *Remote Sensing Letters*, **5**(2), 185–193.
842
- 843 Vihervaara, P., Mononen, L., Auvinen, A.P., Virkkala, R., Lü, Y., Pippuri, I., Packalen, P., Valbuena, R. &
844 Valkama, J. (2015) How to integrate remotely sensed data and biodiversity for ecosystem assessments
845 at landscape scale. *Landscape Ecology*, **30**(3), 501–516.
846
- 847 Virkkala, R. & Lehtikoinen, A. (2014) Patterns of climate-induced density shifts of species: poleward shifts
848 faster in northern boreal birds than in southern birds. *Global Change Biology*, **20**, 2995–3003.
849
- 850 Virkkala R. & Rajasärkkä A. (2006) Spatial variation of bird species in landscapes dominated by old-
851 growth forests in northern boreal Finland. *Biodiversity and Conservation*, **15**, 2143–2162.
852
- 853 Virkkala R., Rajasärkkä A., Väisänen R.A., Vickholm M. & Virolainen E. (1994) Conservation value of
854 nature-reserves: do hole-nesting birds prefer protected forests in southern Finland. *Annales Zoologici*
855 *Fennici*, **31**, 173–186.
856
- 857 Weiner, J. & Solbrig, O.T. (1984) The meaning and measurement of size hierarchies in plant
858 populations. *Oecologia*, **61**(3), 334–336.
859
- 860 Wiens, J.A. (1989a) The ecology of bird communities. Volume 1. Foundations and patterns. Cambridge
861 Univ. Press, Cambridge, UK.
862
- 863 Wiens, J.A. (1989b) The ecology of bird communities. Volume 2. Processes and variations. Cambridge
864 Univ. Press, Cambridge, UK.
865
- 866 Wiens, J.A., Rotenberry, J.T. & Van Horne, B. (1987) Habitat occupancy patterns of North American
867 shrubsteppe birds: the effects of spatial scale. *Oikos*, **48**, 132–147.
868
- 869 Zellweger, F., Roth, T., Bugmann, H., & Bollmann, K. (2017) Beta diversity of plants, birds and butterflies
870 is closely associated with climate and habitat structure. *Global Ecology and Biogeography*, **26**(8), 898–
871 906.
872
873
874
- 875 **Appendix I:** Distribution curves for 30 m buffers for 25 birds, 9 forest structure parameters.
876