



Modelling bird richness and bird species presence in a boreal forest reserve using airborne laser-scanning and aerial images

Katrine Eldegard, John Wirkola Dirksen, Hans Ole Ørka, Rune Halvorsen, Erik Næsset, Terje Gobakken & Mikael Ohlson

To cite this article: Katrine Eldegard, John Wirkola Dirksen, Hans Ole Ørka, Rune Halvorsen, Erik Næsset, Terje Gobakken & Mikael Ohlson (2014) Modelling bird richness and bird species presence in a boreal forest reserve using airborne laser-scanning and aerial images, Bird Study, 61:2, 204-219, DOI: [10.1080/00063657.2014.885492](https://doi.org/10.1080/00063657.2014.885492)

To link to this article: <https://doi.org/10.1080/00063657.2014.885492>



Published online: 14 Mar 2014.



Submit your article to this journal [↗](#)



Article views: 714



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 9 View citing articles [↗](#)

Modelling bird richness and bird species presence in a boreal forest reserve using airborne laser-scanning and aerial images

KATRINE ELDEGARD^{1*}, JOHN WIRKOLA DIRKSEN¹, HANS OLE ØRKA¹, RUNE HALVORSEN², ERIK NÆSSET¹, TERJE GOBAKKEN¹ and MIKAEL OHLSON¹

¹Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, PO Box 5003, NO-1432 Ås, Norway; ²Department of Botany, Natural History Museum, University of Oslo, PO Box 1172 Blindern, NO-0318 Oslo, Norway

Capsule Variables obtained from airborne laser-scanning (ALS) enabled slight or fair predictions of bird presence, and including multispectral data further improved predictions slightly.

Aims To assess the usefulness of ALS as a tool for predicting species richness and single-species presence, and to investigate if including information from multispectral aerial images further improved predictability of bird presence.

Methods Bird presence data were sampled in a Norwegian boreal forest reserve. Prediction models were developed for species richness and presence of the eight most abundant species by the use of two different modelling approaches: generalized linear models and the machine learning method random forest. Predictor variables were descriptors of three-dimensional forest structure obtained by ALS, and descriptors of tree species composition obtained from multispectral aerial images.

Results Cross-validation of the prediction models indicated overall slight or fair predictive capability. Best predictions were obtained for Goldcrest, Wren, and Willow Warbler. Inclusion of spectral variables derived from the aerial imagery slightly improved the predictive performance of several models, most notably for Willow Warbler.

Conclusion We suggest that predictability of species richness and presence of single bird species can be improved by better matching of the scale of recording for birds and the predictor variables obtained by remote sensing.

Management actions with the aim of preventing biodiversity loss require knowledge about how species are related to their habitats, both in human-modified habitats, and in habitats modulated by natural disturbance dynamics (Whittaker *et al.* 2005, Franklin 2009). Species distribution models combine observations of species presence and abundance with estimates of habitat characteristics (Guisan & Zimmermann 2000, Whittaker *et al.* 2005, Pearson *et al.* 2007, Elith & Leathwick 2009, Gogol-Prokurat 2011), and such models have become increasingly popular during the last decade as a tool for predicting species occurrence patterns (Lobo *et al.* 2010). Field-based studies of birds that address patterns on relatively fine spatial scales have revealed that the three-

dimensional arrangement of the habitat strongly influences bird habitat-use (MacArthur & MacArthur 1961, Brokaw & Lent 1999). For forest-dwelling birds, the three-dimensional structural complexity of the forest, i.e. canopy height, stem density, and tree species composition, influences the presence of single bird species as well as bird richness (MacArthur & MacArthur 1961, Karr & Roth 1971, Willson 1974, Holmes & Robinson 1981, Peck 1989). However, no general agreement has so far been reached with respect to up-scaling of results from fine-scale studies to predict responses on broader, population, or landscape scales. A main reason for this is that traditional field-recording methods are often labour-intensive and preclude collection of detailed information over wide areas (Miller *et al.* 2004, Johnson & Gillingham 2008, Vierling *et al.* 2008).

*Correspondence author. Email. katrine.eldegard@nmbu.no

Continuous technological developments make remote sensing an increasingly promising tool for mapping and quantifying selected habitat characteristics over large areas (Horning *et al.* 2010). Accordingly, in order to expand the spatial extent of their studies, ecologists commonly use aerial or satellite images to distinguish between ecosystems, vegetation types, and tree-species dominance types, but such two-dimensional representations of the forest provide limited information about the forest structure (Turner *et al.* 2003, Bergen *et al.* 2009, Horning *et al.* 2010). In contrast, airborne laser-scanning (ALS) provides an accurate three-dimensional representation of both the forest canopy and the terrain surface (Lefsky *et al.* 2002, Næsset *et al.* 2004), based upon a three-dimensional point cloud, each point being an echo with its own unique position in the x, y, and z coordinate system. These coordinates are obtained by recording the time taken for a laser pulse emitted from an aircraft-mounted instrument to travel to the canopy or the Earth's surface and to return back to the sensor. Typically, in broad-scale acquisitions, one to four echoes are recorded from a single emitted pulse (discrete return laser), but also the complete backscatter signal can be obtained with so-called full-waveform systems. The laser pulses are distributed in different directions by a scanning mechanism that enables coverage of a several hundred metres wide swath along the aircraft's flying direction. Typically, from 0.5 to 10 laser measurements are acquired per square metre, with an accuracy of the coordinates of each measurement at decimetre level. Because the cloud of point measurements contains echoes both from the canopy and from the ground, it is possible at the same time to produce a detailed terrain model and to derive a vertical distribution of the canopy. These terrain models and canopy distributions can, in turn, be used to extract variables that characterize (bird) habitats. Since 2002 ALS has been used operationally in Scandinavia for forest management inventories of extensive areas (Næsset 2004) and important forest characteristics like mean tree height (Næsset 1997a), timber volume (Næsset 1997b), and canopy height (Næsset & Bjerknes 2001) are now routinely predicted from ALS data with high spatial resolution.

The number of ecological studies that have used variables obtained by ALS to describe the three-dimensional structure of forests has increased considerably during the last decade (Turner *et al.* 2003, Aplin 2005, McDermid *et al.* 2005, Bässler *et al.* 2011). Ecological applications of ALS data include

classification of forest successional stages (Falkowski *et al.* 2009), detection of understory shrubs (Hill & Broughton 2009, Martinuzzi *et al.* 2009, Wing *et al.* 2012), quantifying properties of canopy gaps (Koukoulas & Blackburn 2004), and estimation of standing dead trees (Bater *et al.* 2009, Kim *et al.* 2009). In bird ecology, ALS has been used to identify forest structures important for single bird species (Hill *et al.* 2004, Broughton *et al.* 2006, Hinsley *et al.* 2006, Graf *et al.* 2009, Smart *et al.* 2012), whole bird communities (Goetz *et al.* 2007, Clawges *et al.* 2008, Müller *et al.* 2009, Seavy *et al.* 2009), and bird guilds (Goetz *et al.* 2007, Lesak *et al.* 2011). Interestingly, a generally valid result from studies using ALS data seems to be that vegetation height is one of the most important variables influencing bird distribution (Hinsley *et al.* 2009, Seavy *et al.* 2009, Smart *et al.* 2012, Tattoni *et al.* 2012).

Studies of bird-habitat relationships by use of data obtained by ALS have been conducted in many different types of forests, e.g. broadleaf forests in England (Hinsley *et al.* 2002, Broughton *et al.* 2006, Bellamy *et al.* 2009), spruce (*Picea abies* L.), fir (*Abies alba* Mill), and beech (*Fagus sylvatica* L.) forests in Germany (Müller *et al.* 2009, Müller *et al.* 2010), pine (*Pinus sylvestris* L.) and aspen (*Populus tremula* L.) forests in the USA (Clawges *et al.* 2008, Smart *et al.* 2012), and spruce forests in Italy (Tattoni *et al.* 2012). However, despite the fact that ALS is increasingly applied for operational forest inventories in Scandinavia, no studies have so far been performed in this region to explore the degree to which bird species richness and the presence of single bird species can be modelled from habitat characteristics derived from ALS. Moreover, to our knowledge, no studies have assessed the usefulness of ALS for predicting presence of bird species in natural boreal forests.

Given that ALS and multispectral imagery record very different characteristics of the forest (Wulder 1998, Lim *et al.* 2003), recent applications of remote sensing in bird studies indicate that using these two remote sensing techniques together may improve predictions (Goetz *et al.* 2007, Swatantran *et al.* 2012). Previous attempts to predict bird presence and habitat use from remote sensing have mainly applied parametric modelling techniques such as generalized linear modelling. However, a number of techniques are now available for prediction modelling (Elith & Graham 2009), including fairly recently developed machine-learning methods such as random forest (RF) (Breiman 2001).

An advantage of the RF algorithm is that it performs an internal weighting of variables, making variable selection redundant (Cutler *et al.* 2007). So far, few attempts have been made to apply RF for prediction of bird-habitat relationships from remote sensing data (Goetz *et al.* 2010), but the technique is widely used with good results for prediction modelling in forest ecosystems described by variables obtained from remote sensing (Pal 2005, Falkowski *et al.* 2009, Ørka *et al.* 2012, Dalponte *et al.* 2013). Some ecological studies have concluded that this method is often better, sometimes substantially better, than alternative classifiers (including logistic regression) for complex ecological data sets (Cutler *et al.* 2007). In the current study, RF was therefore applied as one of the alternative modelling techniques.

The main objective of this study was to assess the usefulness of discrete return ALS metrics for predicting bird species richness and single-species presence in a natural boreal Scandinavian forest. A secondary objective was to find out if including multispectral data improved the predictive performance of the models.

METHODS

Study area

The study was carried out in the Østmarka Nature Reserve (59°50'N, 11° 02'E, 190–368 m above sea level), which covers 12.5 km² and is located in the southern boreal zone of Southeast Norway (Moen 1999). The study area comprises approximately 6 km² in the core of the reserve. Selective logging was performed in some parts of the study area until ca. 1940, but no clear-cutting has been done in the reserve. Today, the forest in the reserve is multi-layered and has a structure typical of forests in the natural state. Old-growth Norway spruce forests dominate whereas Scots pine forests occur on drought-exposed ridges. Deciduous trees, mainly birch (*Betula* sp.), occur scattered throughout the study area. The topography of the study area is dominated by north/south facing ridges and valleys. On a fine scale, the landscape is dissected by minor ridges, valleys and stream beds, and large boulders and steep cliffs are common. This results in a broken topography. A total of 64 bird species have been recorded in the Østmarka reserve (Norwegian Ornithological Society, unpubl. species list).

Stratification and placement of bird counting stations

We applied a stratified random point-sampling technique (Bibby *et al.* 2000) to obtain counting stations from which data on the presence of single bird species were obtained. Initially the study area was divided into five strata that capture variation in vegetation height and tree species composition, which are known to be particularly important for the local distribution of birds (Karr & Roth 1971, Willson 1974, Holmes & Robinson 1981, Peck 1989). Swamp forests larger than 100 m² were mapped in the field in 2010 (Dirksen 2013) and these 121 patches of swamp forest made up stratum 1. Swamp forests have the highest proportion of deciduous trees among forest types encountered in the study area (Dirksen 2013). The other four strata were identified by use of the ALS data. Stratum 2 comprised areas located on ridges (as determined from an ALS derived terrain model), usually dominated by Scots pine and exposed to periodic drought (Økland & Eilertsen 1993). The rest of the study area, which was dominated by Norway spruce, was divided into three strata according to vegetation height, obtained by the following method: the area was first tessellated into 5 m × 5 m grid cells; the laser data for each cell were used to calculate the mean height of echoes with height values greater than 0.5 m above ground within each cell. This information was used to manually delineate polygons with a minimum area of 250 m² based upon the dominant vegetation height category: low (0–10 m), middle (10–20 m), and tall (20–40 m). The sets of polygons for the three height categories made up strata 3 ('low forest'), 4 ('intermediate forest'), and 5 ('tall forest'). Open water and open mires were excluded from the study.

A standardized procedure was used to allocate counting stations of 50 m radius to each stratum. We first allocated random points to the study area until 50 points or the maximum possible number of points was obtained for each of the five strata. Within each stratum, the order by which the points were selected was taken as the priority order of potential bird counting stations representing that stratum. In order to minimize edge effects, points were moved if more than 50% of the area of a 50-m radius circle, centred on the point, fell outside of the stratum polygon. Points were then moved in the direction of the polygon centroid until 70% of the 50-m radius circle fell within the polygon or, if this criterion could not be met, the point was removed.

Table 1. Number of presence observations of the eight most abundant bird species in each of the five sampling strata, number of point counting stations per stratum, number of species observed per stratum (#1,#2,#3,#4,#5), and total values for all strata (All).

Stratum:	#1	#2	#3	#4	#5	All
Percent of total study area	2	4	11	58	25	100
No. of counting stations	20	26	30	43	38	157
Species richness per station (median)	3	2.5	2	3	3	3
Species richness per station (mean)	2.3	2.6	2.2	3.1	2.8	2.7
Chaffinch	10	17	13	29	30	99
Robin	12	7	8	23	19	69
Goldcrest	6	5	7	25	21	64
Willow Warbler	5	10	15	22	4	56
Great Tit	6	11	7	17	14	55
Tree Pipit	2	7	12	7	3	31
Wren	4	3	0	3	13	23
Siskin	1	6	3	9	4	23

Note: See text for additional details about the strata and sampling design.

In order to ensure that counting stations were located at least 150 m apart, points situated closer than 150 m to a previously selected point were moved as explained above. Alternatively, if the 70% and 150-m criteria could not both be met by moving, the point was removed and replaced by the next, lower-priority point. This procedure was used to select ten counting stations from each of the strata, in the order of increasing area covered by each stratum: 1, 2, 3, 5, and 4. Thereafter the procedure was repeated to obtain another ten stations from each stratum. After this no further counting stations could be fitted in stratum 1 that met the selection criteria. Additional counting stations were selected for strata 2, 3, 5, and 4 until no additional stations could be selected without violating one or both of the 70% and 150-m criteria. A total of 157 counting stations were selected, and the number of stations per stratum ranged from 20 to 43 (Table 1).

Bird counting procedures

The centre-point of each counting station was located in the field with a Garmin 60CSx handheld global positioning system (GPS) receiver with an expected horizontal accuracy of 5 m. All the 157 bird counting stations were visited twice in 2012. The first survey was carried out in the period from 8 May to 24 May, and the second survey in the period from 25 May to 10 June. The minimum and maximum number of days between bird surveys at the same counting station were

8 and 30 days, respectively (median 17). All bird counting was done by the same observer, who had previous experience from bird surveys in which the same sampling and distance estimation techniques were used. Distance estimation was calibrated before data collection started, in order to increase the precision of distance assessments. Bird counting was carried out between sunrise and 10:00 AM (local summer time) and was neither carried out if it was raining nor if the wind speed was higher than 8 m/s. At the centre of each counting station, the observer recorded birds for five minutes at each visit. All observed birds (>90% were singing males) were recorded, but non-territorial birds (e.g. overflying birds) were later excluded from the data set before statistical analyses. Only birds recorded by visual or audio observations within the 50 m radius were thus contained in the data set.

Airborne laser-scanning

ALS data covering the entire study area were acquired on 2 July 2010. Flight and acquisition parameters are shown in Table 2. For all echoes, *x* and *y* coordinates and ellipsoidal height values were computed. Ground echoes were identified using the Progressive Triangular Irregular Network Densification Algorithm (Axelsson 2000, Sithole & Vosselman 2004) of the proprietary TerraScan software by Terrasolid, Finland. A terrain model represented by a triangular irregular network was created from the *x*, *y*, and *z* coordinates of the laser echoes classified as ground echoes. The relative height of each echo was then computed as the difference between the height of the individual echo and the terrain model. Due to the accuracy of the terrain model, which was expected to be ca. 20–30 cm (Kraus & Pfeifer 1998, Reutebuch *et al.* 2003), only echoes higher than 0.5 m were used in the subsequent analyses.

A total of 25 continuous variables were obtained from the first echoes within the 50 m circular buffer around the 157 counting stations. From the first echoes above 0.5 m we computed maximum height (*H*_{max}), mean height (*H*_{mean}), standard deviation (*H*_{sd}), coefficient of variation (*H*_c*v*), kurtosis (*H*_{kurtosis}), skewness (*H*_{skewness}), and nine selected percentiles, i.e. the 10th percentile (*H*₁₀), the 20th percentile (*H*₂₀), ..., and the 90th percentile (*H*₉₀). Furthermore, canopy-density variables were calculated as the proportion of echoes above a certain height layer to total number of echoes. Ten height layers were defined by dividing the

Table 2. Summary of instrument specifications and flight parameter settings used for ALS and multispectral image data acquisitions.

Technical specifications		
Platform	Piper Navajo fixed-wing aircraft	
Sensor	Optech ALTM-Gemini	Vexcel UltraCam Xp
Mean flying speed (m s^{-1})	80	80
Flying altitude above ground (m)	900	2850
Number of flightlines	18	1
Side overlap (%)	55	30
Forward overlap (%)	—	80
Scan frequency (Hz)	55	—
Half-scan angle (degrees)	13.8	—
Swath width (m)	440	2940
Pulse repetition frequency (kHz)	100	—
Max. number of echoes per pulse	4	—
Pulse density (m^{-2})	5	—
Spatial resolution (cm) ^{a,b}	19 ^a	51 ^b
Spectral resolution (nm) ^c	1064	410–570 (blue band) 480–630 (green band) 580–700 (red band) 690–1000 (infrared band)

^aFootprint diameter (cm) computed after Baltsavias (1999) based on mean acquisition settings.

^bGround sampling distance.

^cPart of the optical spectra covered by the sensor.

height for the 95th percentile minus 0.5 m into ten equally vertical height layers (D0, D1,..., and D9). These variables are commonly used to describe forest properties such as mean tree height, stem density, and stem volume using ALS (Næsset 2004). The motivation for including canopy-density variables in the analyses of bird distributions, was that vegetation architecture often affects birds by providing perches, song posts and cover, and by limiting the bird's field of view and ability to fly and to find and capture prey (Sutherland et al. 2004).

Multispectral images

Fourteen multispectral images were taken with a Vexcel UltraCam XP sensor on 8 August 2010. Flight and acquisition parameters are shown in Table 2. The Vexcel sensor acquired both panchromatic and four multispectral bands (red, green, blue, and infrared). In the current study, only multispectral bands were used. The images were delivered according to the processing level 2 defined by Vexcel Imaging GmbH. To co-

register the ALS and the image data, each first echo from the laser was mapped onto the image planes of the digital images using a rotation matrix, co-linear equations and principles of digital photogrammetry (Mikhail et al. 2001). The digital numbers from all four bands of the pixel position of the laser echo were added to the respective laser echo. If a laser-echo pixel position occurred on more than one image, the mean band value from all images was assigned to the respective echo.

The rationale behind including image data in the analysis was to add information that better describes differences in tree species composition than ALS data. Therefore, variables which have been used in other studies in which laser and imagery have been combined for species estimation and classification, were used (Packalén et al. 2009, Waser et al. 2011, Ørka et al. 2012). For each counting station, the descriptive variables, mean, standard deviation, coefficient of variation, skewness, kurtosis, and percentiles (10th, 20th, ..., and 90th), were derived from the image values added to first echoes within the 50-m radius for all the four bands of the imagery. Before computation, the band values were normalized to relative band values by dividing each individual band value by the sum of all band values of the echo in question (Packalén et al. 2009). Thus, a total of 60 continuous spectral variables were computed.

Prediction modelling

Response variables used in the prediction modelling were: bird species richness (y_n = number of species) and the presence of each of the eight most abundant bird species ($y_{p,i}$ = 1 means that species i was observed while $y_{p,i}$ = 0 means that the species was not observed). Lack of observed presence was considered as absence, even though detectability of birds is below 100% (Bibby et al. 2000). For each response variable, we applied two modelling techniques in parallel: generalized linear models (GLMs; e.g. Venables & Ripley 2002) and RF (Breiman 2001). Both the GLM and RF prediction models were built by separate use of two different sets of predictors: (1) all ALS variables and (2) all ALS variables plus all spectral variables (ALS + spectral).

We modelled presence and absence of each of the eight species with logistic regression (GLM models with binary response and binomial distribution of errors; Venables & Ripley 2002). Because our data

contained considerably more zeroes than ones, we used the clog-log link function (Zuur *et al.* 2013). However, repeating all our analyses using the logit link function yielded almost identical results (not reported). For bird richness, we fitted a GLM model with log link function, assuming a Poisson distribution of errors. Since we adopted a proof of principle approach in this study, i.e. testing if any of the ALS variables allow prediction of bird presence in a Scandinavian context, and if including spectral variables improves the predictive performance of the models, we deliberately wanted to include a large number of variables in the full models. Model reduction was performed by best-subset selection, as implemented in the R-package *glmulti* (Calcagno 2013). In order to reduce the number of predictor variables to a number that could be handled by the best-subset selection method, we performed a pre-selection of variables. To balance the equally desirable goals of including a large number of predictors, and making the best-subset selection work, we used models with 20 predictors as a starting point for best-subset selection. For each response variable, these 20 predictors were selected by first fitting individual models for each of the ALS and spectral variables and then selecting the 20 variables with the lowest Akaike's information criterion (AIC) values in single-variable models. Thereafter, we selected the best subset of predictors among the 20 variables based on AIC. To avoid collinearity in the selected models, we calculated and inspected the variance inflation factors (VIFs) (Zuur *et al.* 2013). We used a VIF value of 5 as a cut-off for dropping a model and instead selecting as the best model the model with the lowest AIC among models with a VIF value <5 (Montgomery & Peck 1992). Interactions between predictors were not included, since adding all possible interactions would lead to overly complex models relative to the quality of the bird data, and since we had no a priori reasons for including specific interactions.

In addition to GLMs, we applied the RF algorithm (Breiman 2001) to predict bird presence and richness. Classification of presence or absence, or observations into two groups is frequently encountered in ecological studies, so rules are required by which current and new observations can be assigned to the two classes using numerical predictor variables (Cutler *et al.* 2007). In logistic regression, this is accomplished by finding linear combinations of the predictor variables that are used to classify the observations. In contrast, RF is a non-parametric method that uses randomly selected subsets of the observations and variables to create

multiple regression trees, and then aggregates information from all the trees to make predictions of, e.g. single bird species or bird richness. We used the *randomForest* package in R (Liaw & Wiener 2002) to build the models for regression (bird species richness) and classification (presence or absence of single species). To account for unbalanced binary data, i.e. considerably more zeroes than ones, we used a balanced RF procedure (Chen *et al.* 2004).

Assessing model adequacy and predictive performance

The GLM models were checked for over-dispersion by inspecting the generalized Pearson statistic, $gPs = \chi^2 / (N - p)$, where N = number of observations and p = number of parameters (Crawley 2013). The final logistic regression after best-subset selection were not over-dispersed, i.e. had gPs values close to 1 (ranging from 0.78 to 1.02). The Poisson regression models were under-dispersed, and we therefore refitted the final models using generalized Poisson regression. However, graphical validation (inspection of plots of residual versus fitted values and residuals versus predictors) showed that that generalized Poisson regression did not improve the model specification, and we therefore kept the Poisson regression models.

Finally, the predictive abilities and performance of both the GLM and the RF models were assessed by cross-validation. Cross-validation was preferred rather than splitting the data into one set for training and one set for evaluation of models because of relatively sparse data and because such splitting of the data does not solve the problem that effects of potential sampling bias will be left undetected (Halvorsen 2012). For each of the models, we carried out a 10-fold cross-validation by splitting the data into ten roughly equal-sized parts (folds), and then performed a leave-one-fold-out cross-validation, i.e. for the j th fold the fit was computed using all the data except for the j th fold and predictions were made for the j th fold (Hastie *et al.* 2009). As measures of predictive performance of the bird presence models, we used Cohen's kappa (Kappa) (Cohen 1960) and the area under the receiver operating characteristic (ROC) curve (area under the curve (AUC), Pearce & Ferrier 2000) in parallel. For the GLM models we optimized the threshold for splitting the continuous predictions into binary classes using the Cohen's kappa. For the Poisson regression models (bird species richness), we used root mean squared error from the cross-validation

(RMSECV) (Equation 1) as a measure of model performance.

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}, \quad (1)$$

where y is the observed species richness, \hat{y} is the predicted species richness in the cross-validation and N is the number of observations.

RESULTS

The eight most common species in this study, each observed in 23 or more of the 157 counting stations, were Chaffinch *Fringilla coelebs*, Robin *Erithacus rubecula*, Goldcrest *Regulus regulus*, Willow Warbler *Phylloscopus trochilus*, Great Tit *Parus major*, Tree Pipit *Anthus trivialis*, Wren *Troglodytes troglodytes*, and Siskin *Carduelis spinus* (Table 1). The distribution of observations of these eight species on the five strata differed strongly (Table 1). The remaining species were observed in ≤ 9 counting stations. Species richness varied from 0 to 5 (median 3) per counting station (Fig. 1).

Overall, neither logistic regression nor RF prediction models of bird presence as function of ALS variables had more than 'fair' predictive performance (Fig. 2). The best predictive performance was found for logistic

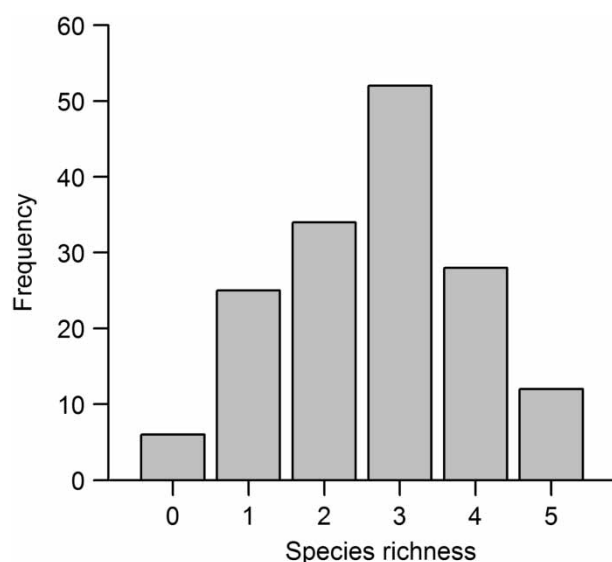


Figure 1. Frequency distribution of bird species richness values. Species richness is the total number of species observed in two surveys of the same 157 counting stations (50 m radius).

regression models for Goldcrest (AUC = 0.69, Kappa = 0.26), Wren (AUC = 0.68, Kappa = 0.26), and Chaffinch (AUC = 0.64 and Kappa = 0.30) and for RF models for Willow Warbler (AUC = 0.70, Kappa = 0.33) and Wren (AUC = 0.70, Kappa = 0.22) (Fig. 2). The logistic regression model for Goldcrest indicated that the probability of presence decreased with increasing skewness (Hskewness) of the laser height distribution, and increased with the 60th height percentile (H60) (Fig. 3, Table 3). For Wren, the probability of presence increased with increasing standard deviation of the laser height distribution (Hsd) and the D0 canopy density (Fig. 4, Table 3). Hsd was positively correlated with the 90th height percentile (H90) ($r = 0.95$), i.e. the spread of the height distribution increased with vegetation height. For Chaffinch, the probability of presence increased with H90 (Table 3). No statistically significant difference in performance was found between logistic regression models and RF models based on the Wilcoxon rank sum test (AUC: $P = 0.88$ and Kappa: $P = 0.33$). The extent of overlap in terms of predictor variables included in the final logistic regression models (Table 3) and the most influential predictors in RF (Fig. 5) was not substantial. Predictor variables that were not included in any of the logistic regression models did have a substantial influence in at least one of the RF models (Table 3, Fig. 5).

Including multispectral data in addition to the ALS variables gave a slight improvement in predictive performance of many of the models (Fig. 2). There was no significant difference between the GLM and the RF models (AUC: $P = 0.57$ and Kappa: $P = 0.16$). The greatest improvement in predictive performance was found for the Willow Warbler using the GLM where the accuracy increased from 0.11 and 0.59 to 0.35 and 0.73 for Kappa and AUC, respectively (Fig. 2). This GLM indicated that in addition to a decreased probability of presence with increasing mean height of the laser height distribution (Hmean) (Fig. 6a, Table 4), the probability of presence was significantly related to three of the spectral variables (Fig. 6b–d, Table 4).

The Poisson regression models for prediction of species richness had generally poor predictive capability (Fig. 7), and including the spectral variables did not improve predictive performance. After best-subset selection, the prediction model built from ALS variables, and the model built from ALS + spectral variables were identical, and included only the 10th height percentile (H10: $\beta = -0.75$, $P = 0.45$) and the D4 canopy density ($\beta = 2.10$, $P = 0.036$) as predictors.

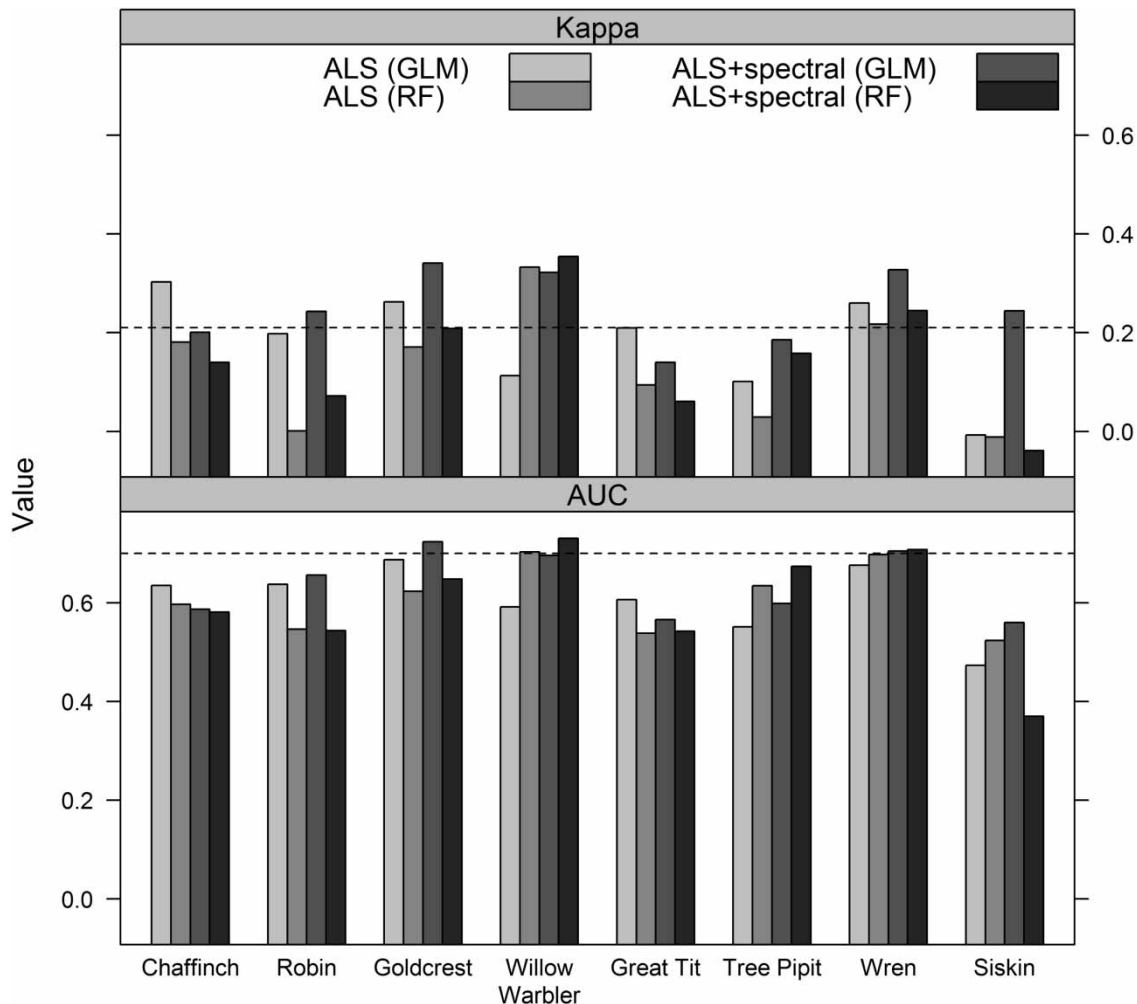


Figure 2. Predictive performance of models predicting presence of eight bird species. Prediction models were constructed from two sets of predictor variables: (1) ALS variables and (2) ALS plus spectral variables (ALS + spectral). For each species and predictor-variable set, two types of prediction models were constructed: a logistic regression model (GLM), and a RF model. Each model was evaluated by 10-fold cross-validation, and area under the curve (AUC) and Cohen's kappa (Kappa) values were used as measures of predictive capability. The horizontal dashed lines indicate a 'fair' prediction for Kappa and AUC as defined by Landis and Koch (1977) and Halvorsen (2013), respectively.

This model was substantially under-dispersed ($gPs = 0.56$). The RMSECV error values ranged from 1.23 to 1.30, which is similar to the standard deviation of the field-measured values.

DISCUSSION

Variables obtained from ALS enabled fair predictions for Goldcrest, Wren, Chaffinch, and Willow Warbler, whereas predictions were poor for Robin, Great Tit, Tree Pipit, and Siskin. Our result that high-resolution remote sensing data and short-duration point counts provide models with limited capability of predicting

bird-habitat associations at landscape scales, accord with results of previous studies (Graf *et al.* 2009, Müller *et al.* 2009). In such studies, the geo-location accuracy of the vegetation data obtained from ALS far exceeds that of the bird data, and this mismatch of data quality and resolution probably explains the limited success of these approaches. High-resolution ALS and aerial imagery are perhaps best regarded as 'precision tools' that may be best-suited for ecological questions for which relevant bird data of a comparable level of detail can be obtained.

The relatively weak predictive capability of our species distribution models may in part be due to

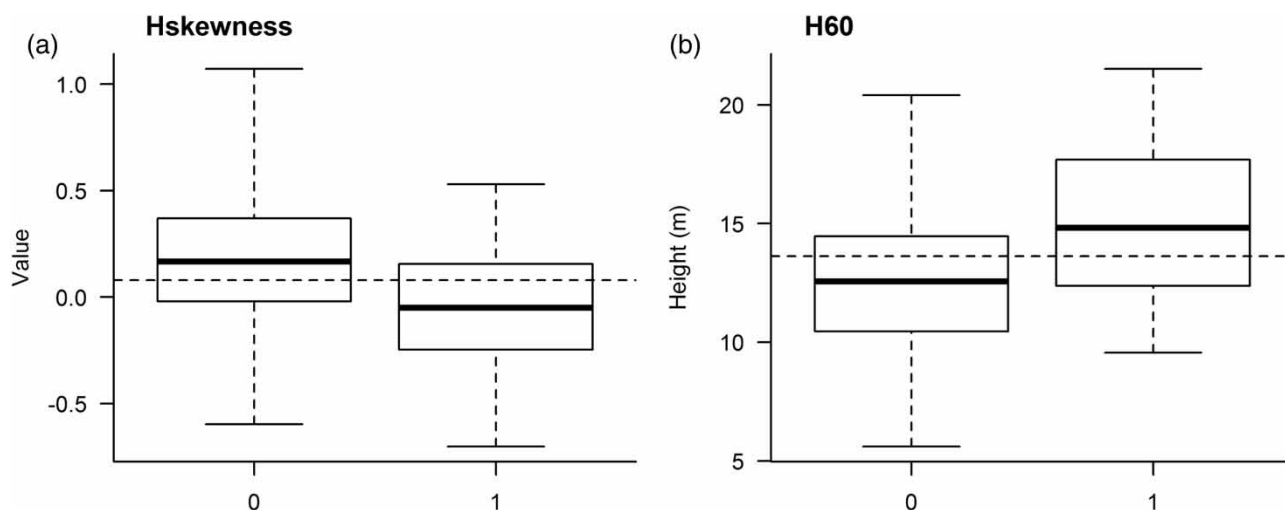


Figure 3. Boxplots for values of the predictor variables (a) skewness of the laser height distribution (Hskewness) and (b) the 60th height percentile (H60), obtained by ALS of counting stations in which Goldcrest was either observed (1) or not observed (0). The dashed horizontal lines show the average value for the predictor variable (i.e. mean value calculated from $N = 157$ counting stations).

shortcomings in our presence/absence data: uncertainty in single observations as well as limitations in the ecological information conveyed by these data. Although we reduced uncertainty in the bird data by using an experienced observer, who carried out the surveys during the time of year and time of day when bird detectability is at its highest, and by performing two surveys per station to capture both early and late breeders, the bird absence data from each counting station should be regarded as pseudo-absences because the field data are indicative rather than exhaustive. Moreover, generalist bird species such as the Great Tit might be expected to occur across the entire field site, and thus it is not surprising that useful predictive models could not be obtained from presence data. More valuable information is likely to be obtained from species abundance data that indicates the strength of habitat suitability (Broughton *et al.* 2012), or from biological activity data that indicates habitat quality (Hinsley *et al.* 2008). For example, Hinsley *et al.* (2006) found that within the same study area, the forest structure conferring 'best' habitat quality for Great Tits differed between years depending on weather conditions. Bird sampling methods that can be used together with high-resolution remote sensing methods to quantify habitat quality include territory mapping (Broughton *et al.* 2012), multi-year sampling (Goetz *et al.* 2010, Swatantran *et al.* 2012), and radio-tracking of individual birds (Trainor *et al.* 2013).

Despite the weaknesses of presence/pseudo-absence bird data, fair predictions were obtained for four out of eight species. Because the final logistic regression models were simple (few predictor variables), they can serve as a basis for discussing possible causal bird-habitat relationships. The laser-scanning variables included in the prediction models for Wren showed that the probability of presence increased with increasing canopy density including the lower layers of the forest, and with variation in vegetation height (which in turn increased with vegetation height). This fits well with results of previous studies, which have found that Wrens prefer old-growth (spruce) forests (Wesołowski 1983, Cramp & Simmons 1988, Wesołowski *et al.* 2010). For Goldcrest, upward shifted laser height distribution increased the probability for presences, and the 60th height percentile (H60) was higher for presence sites. Together, these two variables indicate that Goldcrest prefer higher and denser forest. Müller *et al.* (2009) obtained low predictability of Goldcrest, and found that the most important laser-scanning variable for this species was the standard deviation of canopy height. Field-based studies have shown that the Goldcrest prefers well-grown spruce forest stands (Semb-Johansson *et al.* 1991, Cramp & Simmons 1992). For Chaffinch, the only laser-scanning variable included in the final prediction model was the 90th height percentile (H90), showing that the probability of presence increased in the tall

Table 3. Results of logistic regression analyses of presences of eight bird species, as a function of ALS variables.

	Chaffinch	Robin	Goldcrest	Willow Warbler	Great Tit	Tree Pipit	Wren	Siskin
Intercept	-1.8** (0.57)	-3.2*** (0.9)	-1.7 (0.87)	0.21 (0.98)	-3.3** (1.2)	2.7 (1.6)	-12.0*** (2.9)	7.6 (4.3)
Hmean				-0.34*** (0.09)		-5.9** (2.0)		
Hsd							0.77** (0.25)	
Hcv								-14.4* (6.7)
Hskewness			-1.5* (0.69)					
H60			0.08 (0.06)			2.2 (1.3)		
H90	0.09*** (0.03)					1.9* (0.83)		
D0					-19.1* (9.2)	-32.5** (12.4)	7.5* (3.7)	
D1				4.4* (2.2)				
D2					55.4** (18.5)			
D3		4.6** (1.5)						
D4					-48.2*** (14.2)			
D5						108.0** (39.4)		
D6						-72.8* (34.2)		
D7					20.0** (6.8)			
D9								-28.4 (16.4)
AIC	198.6	208.0	191.7	191.3	199.2	149.6	110.8	131.2
Res. Deviance ^a	194.6	204.0	185.7	185.3	189.2	135.6	104.8	125.2
Var. expl. (%) ^b	6.3	5.6	14.3	10.4	7.5	15.1	24.9	4.5
N ₀ ^c	58	88	93	101	102	126	134	134
N ₁ ^d	99	69	64	56	55	31	23	23
N ^e	157	157	157	157	157	157	157	157

Notes: Parameter estimates and standard errors (in parentheses) for ALS variables included in final models after best-subset model reduction. Variables starting with 'H' are variables describing canopy height, while variables starting with 'D' describe canopy density above ten equally spaced layers including lower parts of the canopy (D0) to only higher parts (D9) (for details see methods section).

*** $P < 0.001$.

** $P < 0.01$.

* $P < 0.05$.

^aRes. Deviance = residual deviance.

^bVar. expl (%) = percent variation explained = [(null deviance - residual deviance)/residual deviance] \times 100].

^cN₀ = number of absence observations.

^dN₁ = number of presence observations.

^eN = Total number of observations.

forest. Similarly, based on laser-scanning variables included in predictions models, Müller *et al.* (2009) concluded that tall and dense forests had increased probability of presence of Chaffinch, whereas the ornithological literature generally refers to Chaffinch as a typical habitat generalist (Semb-Johansson *et al.* 1991, Cramp & Perrins 1994). For Willow Warbler, the ALS variables indicated increased probability of presence with lower vegetation height and increased canopy density across the entire vertical column of biological matter (D1). Both field-based studies and studies using ALS variables have shown that the Willow Warbler is less dependent on mature forest than on shrub or secondary-growth forest (Bellamy *et al.* 2009), and Bellamy *et al.* (2009) were able to use laser data to identify suitable Willow Warbler habitats based on mean vegetation height. To sum up: the associations found in this study between bird presences

and ALS variables match fairly well with previous knowledge about the habitat associations of these species. Therefore, we conclude that our results are likely to represent real relationships and not simply being due to the large number of variables used for modelling.

Another possible explanation of the higher predictability for Wren, Goldcrest, and Chaffinch might be that these species are relatively easy to detect. The detection probability differ between bird species (Bibby *et al.* 2000), and the fact that some species can be heard from greater distances than others was to some extent compensated for by using limited-radius counting stations. However, bird species also differ in their behaviour, and both Wren and Goldcrest are very vocal (T.D. Øygard, pers. comm.) and thus less likely to pass undetected. Chaffinch was by far the most abundant species observed in this

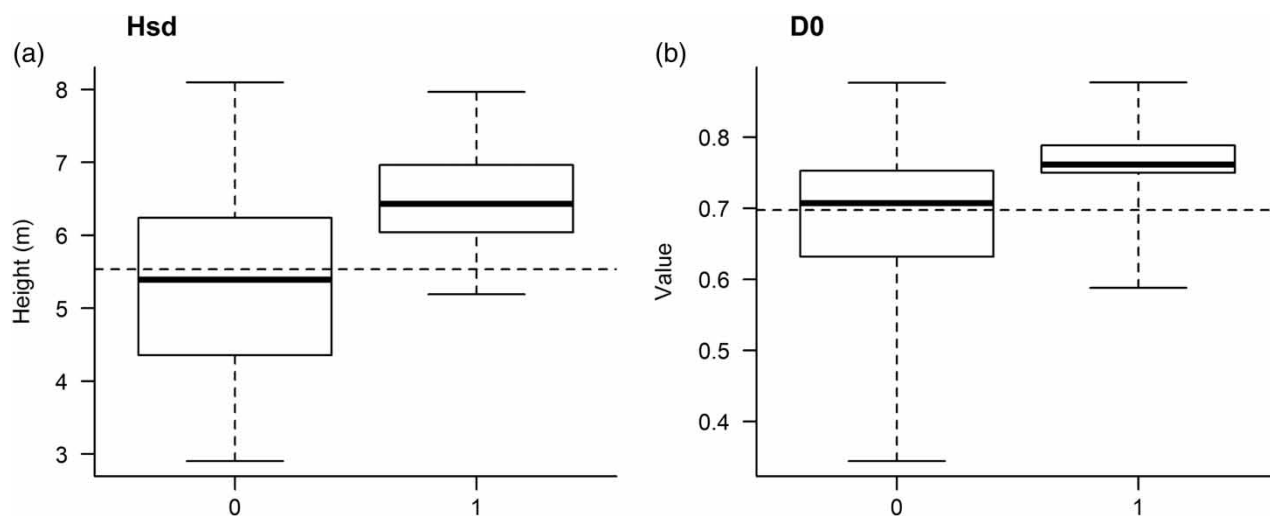


Figure 4. Boxplots for values of the predictor variables (a) standard deviation of the laser height distribution and (b) the lowest canopy density value (D0), i.e. proportion of echoes >0.5 m to total echoes, obtained by ALS of counting stations in which Wren was either observed (1) or not observed (0). The dashed horizontal lines show the average value for the predictor variable (i.e. mean value calculated from $N = 157$ counting stations).

study. At high population densities, the need for territory defence (vocalisations) will be higher, and this could in turn increase detectability.

Prediction models for species richness had poor predictive capability. This could probably be explained by the generally low species richness in our study area, and very low variation among the counting stations. Few studies have used the three-dimensional representation of vegetation offered by ALS to predict bird richness (Goetz *et al.* 2007, Jones *et al.* 2013). Most previous studies exploring the relationship between avian diversity and airborne laser data have concluded that aspects of vegetation height and its variability are important variables (Goetz *et al.* 2007, Clawges *et al.* 2008, Müller *et al.* 2009, Lesak *et al.* 2011), although the models have limited predictability.

Including multispectral data as a complement to the laser-scanning variables substantially improved predictions for Willow Warbler. For the other species and for species richness, no or only slight improvement in predictive performance was found. We included spectral variables in our study mainly because they are known to be useful for identification of tree species, deciduous trees in particular (Hill & Thomson 2005). Deciduous trees, especially birch, are preferred by Willow Warbler to a much larger extent than by the other species in our data set (Semb-Johansson *et al.* 1991, Cramp & Simmons 1992). Our results indicate that spectral variables capture some of the variation in

tree species composition which is of importance for this species. Goetz *et al.* (2007) found only slight differences in performance between models based on ALS variables and models based on both laser and spectral variables, and Müller *et al.* (2009) concluded that ALS data were more useful than spectral data for modelling abundance of single bird species. However, in recent years, simultaneous use of data from ALS and other types of remote sensing data have proven to be a promising approach for describing richness of distinct avian guilds (Jones *et al.* 2013), predicting detections or prevalence of individual bird species (Swatantran *et al.* 2012), and for describing habitat occupation by forest birds (Broughton *et al.* 2012). Thus, the potential of integrating multiple remote sensing data sets to improve prediction models is high (Horning *et al.* 2010).

Our two prediction modelling techniques (i.e. logistic regression and RF) yielded similar results; neither the overall predictive capability nor the species for which we got the best predictions, differed considerably. In contrast, Cutler *et al.* (2007) found moderate superiority of RF to other methods in predicting cavity-nesting bird habitats. Our results accord with the species distribution modelling literature which provides no clear guidance for selecting among relevant methods (Elith & Graham 2009): it is difficult to know in advance which method that will perform the best with a specific data set. Assessing the relative merits of our two modelling techniques is

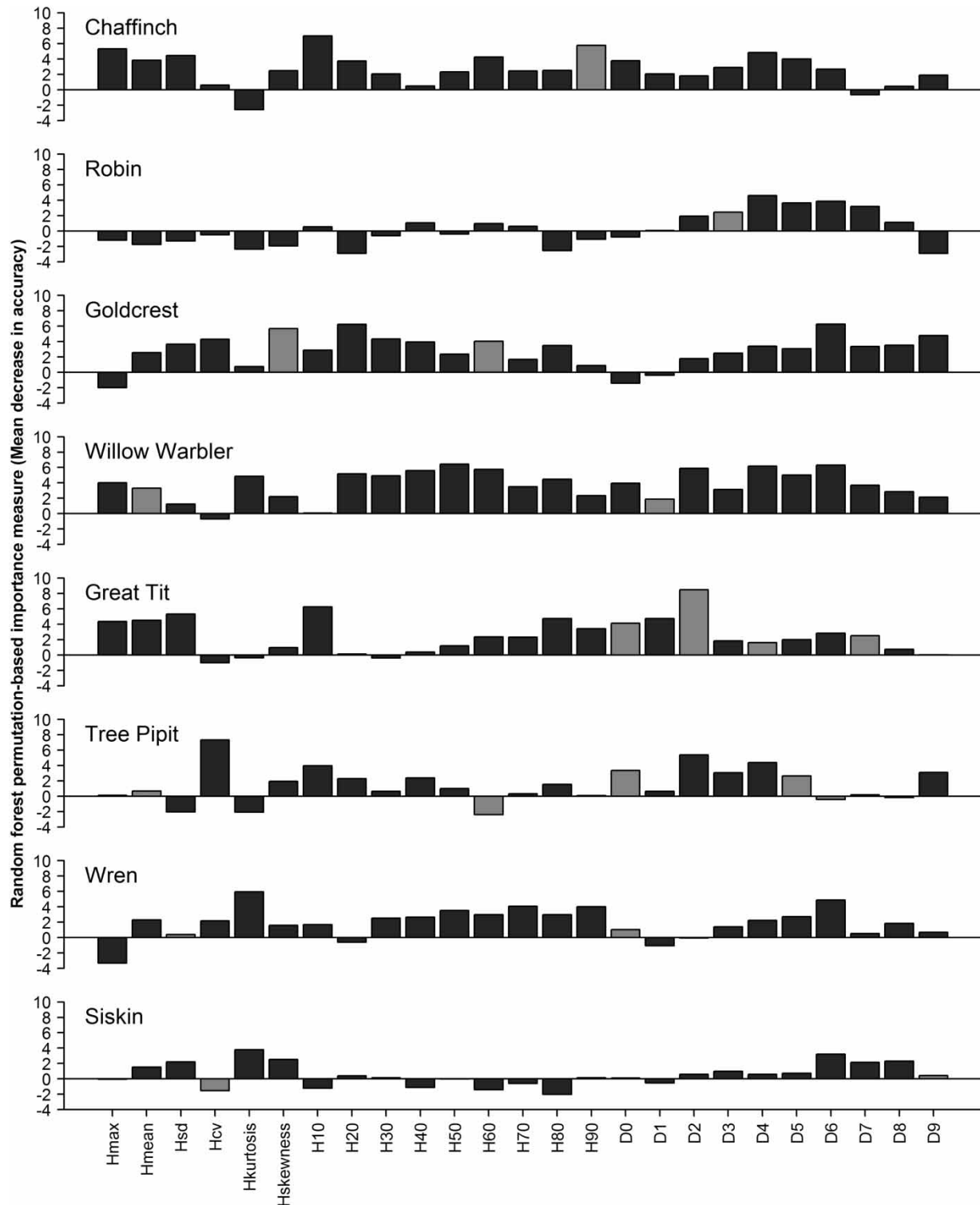


Figure 5. Importance of individual ALS variables (indicated by height of the bars) in RF models predicting presence of the eight most abundant bird species. Higher values indicate a more important variable in the RF models. Grey bars indicate variables which were also included in the final logistic regression prediction models (Table 3). Variables starting with 'H' are variables describing canopy height, while variables starting with 'D' describe canopy density above ten equally spaced layers including echoes from lower parts of the canopy (D0) to only higher parts (D9) (for details see methods section).

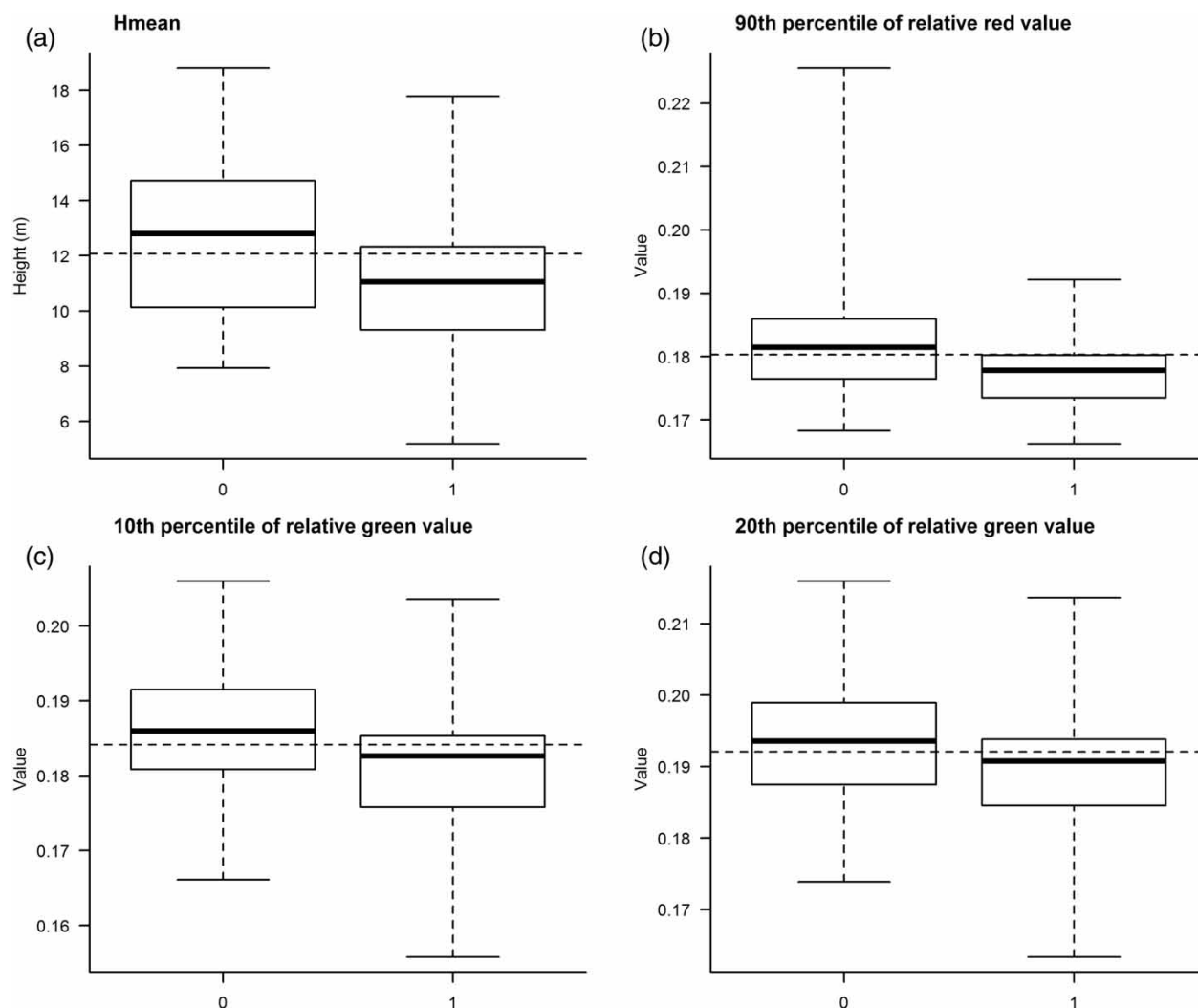


Figure 6. Boxplots for values of the predictor variables (a) mean height of the laser height distribution (Hmean), obtained by ALS of counting stations in which Willow Warbler were either observed (1) or not observed (0). Boxplots for values of the predictor variables (b) 90th percentile of relative red value, (c) 10th percentile of relative green value, and (d) 20th percentile of relative green value for the same counting stations, obtained from multispectral data (aerial images). The dashed horizontal lines show the average value for the predictor variable (i.e. mean value calculated from $N = 157$ counting stations).

probably better done with simulated data where the population is known, or by use of independently collected data for model evaluation (Guisan & Zimmermann 2000, Austin 2007, Halvorsen 2012). However, our conclusions are strengthened by the fact that logistic regression and RF models yielded similar results in our evaluation by cross-validation.

Presence observations of the eight bird species were unevenly distributed on the five different sampling strata. The stratified sampling design was only used to ensure that important habitat variation was captured by the counting stations, and the five strata were not used as explanatory variables in any of the models.

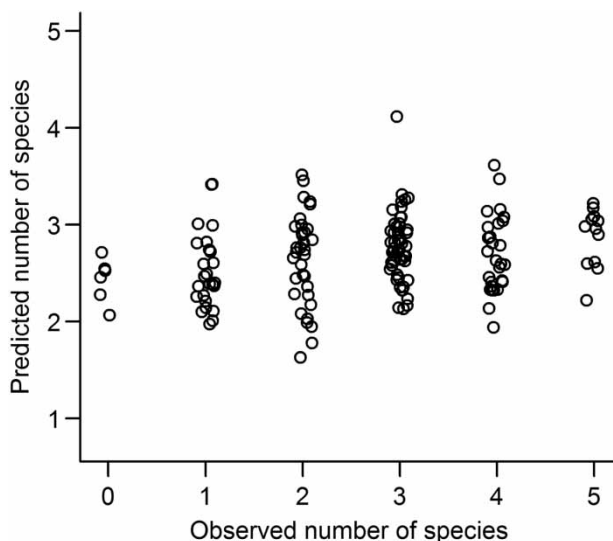
Thus, any 'cross-contamination' between these coarse habitat 'classes' (e.g. to what extent pine distribution corresponds to 'ridges' as mapped in the terrain models, and spruce occurs in the swamp forest areas) should not influence our conclusions. However, the fact that the ALS data were used to define four of the five strata indicates that such data convey relevant information about bird-habitat associations. The uneven distribution of species on the different strata also suggests that the stratification actually yielded better predictions compared to what would have been achieved with, for example, systematic sampling or simple random sampling.

Table 4. Results of logistic regression analyses of presence of Willow Warbler, as a function of ALS and spectral variables.

	β	se	z	P
Intercept	19.2	5.2	3.7	0.00019
Mean canopy height (Hmean)	-0.32	0.067	-4.7	<0.0001
90th percentile of relative red value	-111.0	43.6	-2.5	0.011
10th percentile of relative green value	-413.1	104.9	-3.9	<0.0001
20th percentile of relative green value	414.3	106.7	3.9	0.00010

Notes: Parameter estimates, standard errors, z-statistic and P-values for variables included in the final logistic regression model after best-subset model reduction. Percent variation explained = [(null deviance - residual deviance)/residual deviance] \times 100] was 31%.

In conclusion, ALS data were useful for predicting presence of some bird species in a Scandinavian natural boreal forest. Including information from multispectral aerial images substantially improved prediction for the species most strongly associated with deciduous trees. Yet, for most species and for species richness, predictability was limited. We suggest that predictability of bird species may be improved by better matching of recording scales for birds and predictor variables, and by improving measures of bird habitat-use, for example, by use of telemetry data, territory mapping, and longer term data from the same study area.

**Figure 7.** Observed values of bird species richness per counting station, against the values predicted in the Poisson regression with ALS variables as predictors. A small random noise is added to the observed value for increased visibility.

ACKNOWLEDGEMENTS

We thank T.D. Øygard for the bird data sampling, and Svein Dale for providing the species list for Østmarka and general advice about field sampling of bird data.

REFERENCES

- Aplin, P.** 2005. Remote sensing: ecology. *Prog. Phys. Geogr.* **29**: 104–113.
- Austin, M.** 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecol. Model.* **200**: 1–19.
- Axelsson, P.** 2000. DEM generation from laser scanner data using adaptive TIN-models. *Int. Arch. Photogramm. Remote. Sens.* **33**: 110–117.
- Baltsavias, E.P.** 1999. Airborne laser scanning: basic relations and formulas. *ISPRS J. Photogramm.* **54**: 199–214.
- Bässler, C., Stadler, J., Müller, J., Förster, B., Göttele, A. & Brandl, R.** 2011. LiDAR as a rapid tool to predict forest habitat types in Natura 2000 networks. *Biodivers. Conserv.* **20**: 465–481.
- Bater, C.W., Coops, N.C., Gergel, S.E., LeMay, V. & Collins, D.** 2009. Estimation of standing dead tree class distributions in northwest coastal forests using lidar remote sensing. *Can. J. For. Res.* **39**: 1080–1091.
- Bellamy, P.E., Hill, R.A., Rothery, P., Hinsley, S.A., Fuller, R.J. & Broughton, R.** 2009. Willow Warbler *Phylloscopus trochilus* habitat in woods with different structure and management in southern England. *Bird Study* **56**: 338–348.
- 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. *J. Geophys. Res. Biogeosci.* **114**: G00E06.
- Bibby, C.J., Burgess, N.D., Hill, D.A. & Mustoe, S.H.** 2000. *Bird Census Techniques*. 2nd edn. Academic Press, London.
- Breiman, L.** 2001. Random forests. *Mach. Learn.* **45**: 5–32.
- Brokaw, N.V.L. & Lent, R.A.** 1999. Vertical structure. In Hunter, M.L. (ed.) *Maintaining Biodiversity in Forest Ecosystems*, 373–395. Cambridge University Press, Cambridge.
- Broughton, R.K., Hill, R.A., Freeman, S.N., Bellamy, P.E. & Hinsley, S.A.** 2012. Describing habitat occupation by woodland birds with territory mapping and remotely sensed data: an example using the Marsh Tit (*Parus palustris*). *Condor* **114**: 812–822.
- Broughton, R.K., Hinsley, S.A., Bellamy, P.E., Hill, R.A. & Rothery, P.** 2006. Marsh Tit *Parus palustris* territories in a British broad-leaved wood. *Ibis* **148**: 744–752.
- Calcagno, V.** 2013. *glmulti: model selection and multimodel inference made easy*. R package version 1.0.7. <http://CRAN.R-project.org/package=glmulti>
- Chen, C., Liaw, L. & Breiman, L.** 2004. *Using Random Forest to Learn Imbalanced Data*, 1–12. Tech. Rep. Dept. Statistics, Univ., Berkeley, CA.
- 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. Sens. Environ.* **112**: 2064–2073.
- Cohen, J.** 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**: 37–46.
- Cramp, S. & Perrins, C.M.** 1994. *Handbook of the Birds of Europe, the Middle East and North Africa: the Birds of the Western Palearctic. Volume 8: Crows to Finches*. Oxford University Press, Oxford.
- Cramp, S. & Simmons, K.E.L.** 1988. *Handbook of the Birds of Europe, the Middle East and North Africa: the Birds of the Western Palearctic*.

- Volume 5: *Tyrant Flycatchers to Thrushes*. Oxford University Press, Oxford.
- Cramp, S. & Simmons, K.E.L.** 1992. *Handbook of the Birds of Europe, the Middle East and North Africa: the Birds of the Western Palearctic*. Volume 6: *Warblers*. Oxford University Press, Oxford.
- Crawley, M.J.** 2013. *The R Book*. 2nd edn. Wiley, Chichester.
- Cutler, D.R., Edwards, T.C. Jr., 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., Ørka, H.O., Gobakken, T., Gianelle, D. & Næsset, E.** 2013. Tree species classification in boreal forests with hyperspectral data. *IEEE T. Geosci. Remote* **51**: 2632–2645.
- Dirksen, J.W.** 2013. Modelling presence of swamp forest and forest dwelling birds in a boreal forest reserve using airborne laser scanning. PhD Thesis. Norwegian University of Life Sciences.
- Elith, J. & Graham, C.H.** 2009. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography* **32**: 66–77.
- Elith, J. & Leathwick, J.R.** 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Syst.* **40**: 677–697.
- 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 Sens. Environ.* **113**: 946–956.
- Franklin, J.** 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge.
- Goetz, S., Steinberg, D., Dubayah, R. & Blair, B.** 2007. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sens. Environ.* **108**: 254–263.
- Goetz, S.J., Steinberg, D., Betts, M.G., Holmes, R.T., Doran, P.J., Dubayah, R. & Hofton, M.** 2010. Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird. *Ecology* **91**: 1569–1576.
- Gogol-Prokurat, M.** 2011. Predicting habitat suitability for rare plants at local spatial scales using a species distribution model. *Ecol. Appl.* **21**: 33–47.
- Graf, R.F., Mathys, L. & Bollmann, K.** 2009. Habitat assessment for forest dwelling species using LiDAR remote sensing: Capercaillie in the Alps. *Forest Ecol. Manag.* **257**: 160–167.
- Guisan, A. & Zimmermann, N.E.** 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* **135**: 147–186.
- Halvorsen, R.** 2012. A gradient analytic perspective on distribution modelling. *Sommerfeltia* **35**: 1–165.
- Halvorsen, R.** 2013. A strict maximum likelihood explanation of MaxEnt, and some implications for distribution modelling. *Sommerfeltia* **36**: 1–132.
- Hastie, T., Tibshirani, R. & Friedman, J.** 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd edn. Springer, New York.
- Hill, R.A. & Broughton, R.K.** 2009. Mapping the understorey of deciduous woodland from leaf-on and leaf-off airborne LiDAR data: a case study in lowland Britain. *ISPRS J. Photogramm.* **64**: 223–233.
- Hill, R.A., Hinsley, S.A., Gaveau, D.L.A. & Bellamy, P.E.** 2004. Predicting habitat quality for Great Tits (*Parus major*) with airborne laser scanning data. *Int. J. Remote Sens.* **25**: 4851–4855.
- Hill, R.A. & Thomson, A.G.** 2005. Mapping woodland species composition and structure using airborne spectral and LiDAR data. *Int. J. Remote Sens.* **26**: 3763–3779.
- Hinsley, S.A., Hill, R.A., Bellamy, P.E. & Baltzer, H.** 2006. The application of lidar in woodland bird ecology: climate, canopy structure, and habitat quality. *Photogramm. Eng. Rem. S.* **72**: 1399–1406.
- Hinsley, S.A., Hill, R.A., Bellamy, P.E., Harrison, N.M., Speakman, J.R., Wilson, A.K. & Ferns, P.N.** 2008. Effects of structural gradients and functional habitat gaps on breeding woodland birds: working harder for less. *Landscape Ecol.* **23**: 615–626.
- Hinsley, S.A., Hill, R.A., Fuller, R.J., Bellamy, P.E. & Rothery, P.** 2009. Bird species distributions across woodland canopy structure gradients. *Community Ecol.* **10**: 99–110.
- Hinsley, S.A., Hill, R.A., Gaveau, D.L.A. & Bellamy, P.E.** 2002. Quantifying woodland structure and habitat quality for birds using airborne laser scanning. *Funct. Ecol.* **16**: 851–857.
- Holmes, R.T. & Robinson, S.K.** 1981. Tree species preferences of foraging insectivorous birds in a northern hardwood forest. *Oecologia* **48**: 31–35.
- Horning, N., Robinson, J.A., Sterling, E.J., Turner, W. & Spector, S.** 2010. *Remote Sensing for Ecology and Conservation. A Handbook of Techniques*. Oxford University Press, Oxford.
- Johnson, C.J. & Gillingham, M.P.** 2008. Sensitivity of species-distribution models to error, bias, and model design: an application to resource selection functions for woodland caribou. *Ecol. Model.* **213**: 143–155.
- Jones, T.G., Arcese, P., Sharma, T. & Coops, N.C.** 2013. Describing avifaunal richness with functional and structural bioindicators derived from advanced airborne remotely sensed data. *Int. J. Remote Sens.* **34**: 2689–2713.
- Karr, J.R. & Roth, R.R.** 1971. Vegetation structure and avian diversity in several new world areas. *Am. Nat.* **105**: 423–435.
- Kim, Y., Yang, Z., Cohen, W.B., Plugmacher, D., Lauver, C.L. & Vankat, J.L.** 2009. Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint lidar data. *Remote Sens. Environ.* **113**: 2499–2510.
- Koukoulas, S. & Blackburn, G.A.** 2004. Quantifying the spatial properties of forest canopy gaps using LiDAR imagery and GIS. *Int. J. Remote Sens.* **25**: 3049–3071.
- Kraus, K. & Pfeifer, N.** 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS J. Photogramm.* **53**: 193–203.
- Landis, J.R. & Koch, G.G.** 1977. The measurement of observer agreement for categorical data. *Biometrics* **33**: 159–174.
- Lefsky, M.A., Cohen, W.B., Parker, G.G. & Harding, D.J.** 2002. Lidar remote sensing for ecosystem studies. *Bioscience* **52**: 19–30.
- Lesak, A.A., Radeloff, V.C., Hawbaker, T.J., Pidgeon, A.M., Gobakken, T. & Contrucci, K.** 2011. Modeling forest songbird species richness using LiDAR-derived measures of forest structure. *Remote Sens. Environ.* **115**: 2823–2835.
- Liaw, A. & Wiener, M.** 2002. Classification and regression by randomforest. *R News* **2**: 18–22.
- Lim, K., Treitz, P., Wulder, M., St-Onge, B. & Flood, M.** 2003. LiDAR remote sensing of forest structure. *Prog. Phys. Geog.* **27**: 88–106.
- Lobo, J.M., Jiménez-Valverde, A. & Hortal, J.** 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* **33**: 103–114.
- MacArthur, R. & MacArthur, J.W.** 1961. On bird species diversity. *Ecology* **42**: 594–598.
- Martinuzzi, S., Vierling, L.A., Gould, W.A., Falkowski, M.J., Evans, J.S., Hudak, A.T. & Vierling, K.** 2009. Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sens. Environ.* **113**: 2533–2546.
- McDermid, G.J., Franklin, S.E. & LeDrew, E.F.** 2005. Remote sensing for large-area habitat mapping. *Prog. Phys. Geogr.* **29**: 449–474.

- Mikhail, E.M., Bethel, J.S. & McGlone, J.C.** 2001. *Introduction to Modern Photogrammetry*. Wiley, New York.
- Miller, J.R., Turner, M.G., Smithwick, E.A.H., Dent, C.L. & Stanley, E.H.** 2004. Spatial extrapolation: the science of predicting ecological patterns and processes. *Bioscience* **54**: 310–320.
- Moen, A.** 1999. *National Atlas of Norway: Vegetation*. Norwegian Mapping Authority, Hønefoss.
- Montgomery, D.C. & Peck, E.** 1992. *Introduction to Linear Regression Analysis*. 2nd edn. Wiley, New York.
- Müller, J., Moning, C., Bäessler, C., Heurich, M. & Brandl, R.** 2009. Using airborne laser scanning to model potential abundance and assemblages of forest passerines. *Basic Appl. Ecol.* **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 Sens. Environ.* **114**: 490–495.
- Næsset, E.** 1997a. Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS J. Photogramm.* **52**: 49–56.
- Næsset, E.** 1997b. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sens. Environ.* **61**: 246–253.
- Næsset, E.** 2004. Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic full-scale operational project. *Scand. J. For. Res.* **19**: 554–557.
- Næsset, E. & Bjerknes, K.O.** 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. *Remote Sens. Environ.* **78**: 328–340.
- Næsset, E., Holmgren, T., Hyypä, H., Hyypä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å. & Söderman, U.** 2004. Laser scanning of forest resources: the Nordic experience. *Scand. J. For. Res.* **19**: 482–499.
- Økland, R.H. & Eilertsen, O.** 1993. Vegetation-environment relationships of boreal coniferous forests in the Solhomfjell area, Gjerstad, S Norway. *Sommerfeltia* **16**: 1–254.
- Ørka, H.O., Gobakken, T., Næsset, E., Ene, L. & Lien, V.** 2012. Simultaneously acquired airborne laser scanning and multispectral imagery for individual tree species identification. *Canadian J. Remote Sens.* **38**: 125–138.
- Packalén, P., Suvanto, A. & Maltamo, M.** 2009. A two stage method to estimate species-specific growing stock. *Photogramm. Eng. Rem. S.* **75**: 1451–1460.
- Pal, M.** 2005. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **26**: 217–222.
- Pearce, J. & Ferrier, S.** 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* **133**: 225–245.
- Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Peterson, A.T.** 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J. Biogeogr.* **34**: 102–117.
- Peck, K.M.** 1989. Tree species preferences shown by foraging birds in forest plantations in northern England. *Biol. Conserv.* **48**: 41–57.
- Reutebuch, S.E., McGaughey, R.J., Andersen, H.E. & Carson, W.W.** 2003. Accuracy of a high-resolution lidar terrain model under a conifer forest canopy. *Can. J. Forest Res.* **29**: 527–535.
- Seavy, N.E., Viers, J.H. & Wood, J.K.** 2009. Riparian bird response to vegetation structure: a multiscale analysis using LIDAR measurements of canopy height. *Ecol. Appl.* **19**: 1848–1857.
- Semb-Johansson, A., Frislid, R. & Hogstad, O.** (eds). 1991. *Norges Dyr. Fuglene*. Cappelen, Oslo. (In Norwegian).
- Sithole, G. & Vosselman, G.** 2004. Experimental comparison of filter algorithms for bare-earth extraction from airborne laser scanning point clouds. *ISPRS J. Photogramm.* **59**: 85–101.
- Smart, L.S., Swenson, J.J., Christensen, N.L. & Sexton, J.O.** 2012. Three-dimensional characterization of pine forest type and red-cockaded woodpecker habitat by small-footprint, discrete-return lidar. *Forest Ecol. Manag.* **281**: 100–110.
- Sutherland, W.J., Newton, I. & Green, R.E.** 2004. *Bird Ecology and Conservation. A Handbook of Techniques*. Oxford University Press, Oxford.
- Swatantran, A., Dubayah, R., Goetz, S., Hofton, M., Betts, M.G., Sun, M., Simard, M. & Holmes, R.** 2012. Mapping migratory bird prevalence using remote sensing data fusion. *PLoS ONE* **7**: e28922.
- Tattoni, C., Rizzolli, F. & Pedrini, P.** 2012. Can LiDAR data improve bird habitat suitability models? *Ecol. Model.* **245**: 103–110.
- Trainor, A.M., Walters, J.R., Morris, W.F., Sexton, J., & Moody, A.** 2013. Empirical estimation of dispersal resistance surfaces: a case study with red-cockaded woodpeckers. *Landscape Ecol.* **28**: 755–767.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. & Steininger, M.** 2003. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* **18**: 306–314.
- Venables, W.N. & Ripley, B.D.** 2002. *Modern Applied Statistics with S*. 4th edn. Springer, New York.
- Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S. & Clawges, R.M.** 2008. Lidar: shedding new light on habitat characterization and modeling. *Front. Ecol. Environ.* **6**: 90–98.
- Waser, L.T., Ginzler, C., Kuechler, M., Baltsavias, E. & Hurni, L.** 2011. Semi-automatic classification of tree species in different forest ecosystems by spectral and geometric variables derived from airborne digital sensor (ADS40) and RC30 data. *Remote Sens. Environ.* **115**: 76–85.
- Wesołowski, T.** 1983. The breeding ecology and behaviour of Wrens *Troglodytes troglodytes* under primeval and secondary conditions. *Ibis* **125**: 499–515.
- Wesołowski, T., Mitrus, C., Czeszczewik, D. & Rowiński, P.** 2010. Breeding bird dynamics in a primeval temperate forest over thirty-five years: variation and stability in the changing world. *Acta Ornithol.* **45**: 209–232.
- Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E. M. & Willis, K.J.** 2005. Conservation biogeography: assessment and prospect. *Diversity Distrib.* **11**: 3–23.
- Willson, M.F.** 1974. Avian community organization and habitat structure. *Ecology* **55**: 1017–1029.
- 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 Sens. Environ.* **124**: 730–741.
- Wulder, M.** 1998. Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. *Prog. Phys. Geog.* **22**: 449–476.
- Zuur, A.F., Hilbe, J.M. & Ieno, E.N.** 2013. *A Beginner's Guide to GLM and GLMM with R*. Highland Statistics Ltd., Newburgh.