Modeling Marbled Murrelet (Brachyramphus marmoratus) Habitat Using LiDAR-Derived Canopy Data

JOAN C. HAGAR,1 United States Geological Survey, Forest and Rangeland Ecosystem Science Center, Corvallis, OR 97331, USA
BIANCA N. I. ESKELSON, Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, OR 97331, USA
PATRICIA K. HAGGERTY, United States Geological Survey, Forest and Rangeland Ecosystem Science Center, Corvallis, OR 97331, USA
S. KIM NELSON, Oregon Cooperative Wildlife Research Unit, Oregon State University, Corvallis, OR 97331, USA
DAVID G. VESELY, Oregon Wildlife Institute, Corvallis, OR 97339, USA

ABSTRACT LiDAR (Light Detection And Ranging) is an emerging remote-sensing tool that can provide fine-scale data describing vertical complexity of vegetation relevant to species that are responsive to forest structure. We used LiDAR data to estimate occupancy probability for the federally threatened marbled murrelet (Brachyramphus marmoratus) in the Oregon Coast Range of the United States. Our goal was to address the need identified in the Recovery Plan for a more accurate estimate of the availability of nesting habitat by developing occupancy maps based on refined measures of nest-strand structure. We used murrelet occupancy data collected by the Bureau of Land Management Coos Bay District, and canopy metrics calculated from discrete return airborne LiDAR data, to fit a logistic regression model predicting the probability of occupancy. Our final model for stand-level occupancy included distance to coast, and 5 LiDAR-derived variables describing canopy structure. With an area under the curve value (AUC) of 0.74, this model had acceptable discrimination and fair agreement (Cohen’s $k = 0.24$), especially considering that all sites in our sample were regarded by managers as potential habitat. The LiDAR model provided better discrimination between occupied and unoccupied sites than did a model using variables derived from Gradient Nearest Neighbor maps that were previously reported as important predictors of murrelet occupancy (AUC = 0.64, $k = 0.12$). We also evaluated LiDAR metrics at 11 known murrelet nest sites. Two LiDAR-derived variables accurately discriminated nest sites from random sites (average AUC = 0.91). LiDAR provided a means of quantifying 3-dimensional canopy structure with variables that are ecologically relevant to murrelet nesting habitat, and have not been as accurately quantified by other mensuration methods. Published 2014. This article is a U.S. Government work and is in the public domain in the USA.

KEY WORDS Brachyramphus marmoratus, forest structure, habitat model, LiDAR, marbled murrelet, remote sensing.

Habitat models are useful tools in conservation biology when they can be used to accurately predict species distributions over large areas and to identify priority areas for conservation and restoration (Guisan and Zimmermann 2000). Development of accurate, reliable models for species that use old-growth forest canopies can be challenging because of the difficulties of discerning and measuring the relevant features at an appropriate spatial scale in complexly structured, 3-dimensional space. Until recently, acquiring data that characterize vertical structure at a sub-meter scale of resolution to inform habitat models for landscape scale applications has not been feasible. However, recent studies (e.g., Goetz et al. 2007, 2010; Smart et al. 2012; Zhao et al. 2012) have overcome these challenges with successful applications of LiDAR (Light Detection And Ranging) technology in wildlife–habitat models. Because it directly measures both vertical and horizontal vegetation distribution (Lefsky et al. 1999; Hyde et al. 2006), LiDAR is capable of providing information on 3-dimensional habitat structure that influences patterns of biodiversity in general (Bergen et al. 2009), and is particularly relevant for species that respond directly to canopy structure (Vierling et al. 2008, Jung et al. 2012, Palminteri et al. 2012).

The marbled murrelet (hereafter, murrelet; Brachyramphus marmoratus) is a federally threatened seabird that, in Washington, Oregon, and California (USA), nests primarily in the canopy of old-growth and mature forest (Hamer and Nelson 1995, Ralph et al. 1995, Nelson 1997). The conservation of nesting habitat is a critical component of the murrelet recovery plan (USFWS 1997). In particular, the recovery plan identifies the need for refined measures of nest site structure and selection by murrelets, to more accurately estimate the availability of nesting habitat (Recovery Task 4.4.1.2). Raphael et al. (2011) modeled habitat suitability for the murrelet in Washington, Oregon, and California using

Received: 8 March 2013; Accepted: 10 September 2013
Published: 12 February 2014

1E-mail: joan_hagar@usgs.gov
presence only data (Maxent; Phillips et al. 2006, Phillips and Dudik 2008). The models they developed used vegetation covariates derived from integrated field plot data and satellite imagery with Gradient Nearest Neighbor imputation methods (Ohmann and Gregory 2002). Although the performance of the models developed by Raphael et al. (2011) was strong for each of the 3 U.S. states (OR, WA, and CA) modeled (area under the curve [AUC] > 0.85), Gradient Nearest Neighbor–derived data do not perform as well at smaller spatial scales, such as stands and patches (Pierce et al. 2009). Furthermore, variables that discriminate suitable and unsuitable habitat at a local scale are likely to be different from those that perform well at a regional scale (Gaillard et al. 2010). Development of tools to identify habitat at finer spatial scales would assist with district-level planning and management.

Models developed for use at smaller spatial scales have established that variables describing the availability of nesting platforms and vertical canopy complexity are important correlates of suitable nesting habitat for murrelets (Hamer and Nelson 1995, Nelson and Wilson 2002, Hamer et al. 2008, Raphael et al. 2011). However, the nuances of canopy structure that provide suitable nesting platforms are difficult to quantify. Historically, murrelet nesting habitat has been characterized as multistoried or multilayered (Hamer and Nelson 1995, USFWS 1997), but a method for direct quantification of this characteristic has not been developed. Rather, a discrete, categorical variable, the number of tree-canopy layers present, has been derived from Gradient Nearest Neighbor data (Raphael et al. 2011) or estimated in the field by observers on the ground (e.g., Hamer et al. 2008). Although this categorical variable has been useful in classification of habitat suitability for murrelets, continuous, directly measured variables provide a more accurate description of vertical structure and would likely result in superior discriminatory power (Graf et al. 2009). LiDAR offers the potential to provide metrics that directly characterize canopy complexity at a finer spatial resolution than Gradient Nearest Neighbor–derived data, based on empirical rather than imputed data, and without the subjectivity and observer bias inherent in field data. Furthermore, a habitat model based on LiDAR-derived variables would directly address the need for an accurate and repeatable methodology for inventory and monitoring of habitat identified in the recovery plan (USFWS 1997:147). The increasing availability of LiDAR coverage combined with capability of LiDAR to quantify fine-scale, 3-dimensional canopy complexity over broad geographic areas provides a promising application for characterizing murrelet nesting habitat (Lešký et al. 2002). Although LiDAR coverage is not currently complete across the tree-nesting range of the murrelet, data are available for many areas; LiDAR data have been acquired for 54% and 46% of the area within 80 km of the coast for Oregon and Washington, respectively, and both states are actively acquiring additional coverage (P. Haggerty, U.S. Geological Survey, unpublished data).

Our goal in this study was to explore the utility of LiDAR-derived variables to accurately identify and map suitable nesting habitat for the murrelet at a stand-level spatial scale. Our specific objectives were to 1) identify LiDAR-derived variables that were most strongly associated with stand-level occupancy and nest trees; 2) assess the usefulness of LiDAR in refining understanding of murrelet–habitat relationships by comparing a model incorporating LiDAR-derived variables to a model using Gradient Nearest Neighbor–derived variables; and 3) develop a predictive map of probability of murrelet occupancy based on LiDAR-derived data for the Bureau of Land Management Coos Bay District (hereafter, Coos Bay BLM). Ultimately, we wanted to provide managers with a tool to address recovery goals for this threatened species at the District level, where management actions are implemented.

STUDY AREA

The 1,692 km² study area was located in the Coast Range physiographic province of southwestern Oregon and included the forested lands of Coos Bay BLM and the Elliott State Forest, managed by the Oregon Department of Forestry (Fig. 1). The Coos Bay BLM was managed under the Northwest Forest Plan (USDA/USDI 1994), which led to the establishment of extensive late-successional forest reserves across the study area and other strategies to protect habitat for native forest-associated species. The region was dominated by Pseudotsuga menziesii–mixed conifer forest and stands of Pseudotsuga menziesii–Tsuga heterophylla (Kiilsgaard 1999). Much of the Coos Bay BLM land occurred in 258-ha (1-mi²) sections interspersed among privately owned lands, and consequently included highly fragmented late-successional forests.

METHODS

Murrelet Data

Site-level survey data.—The murrelet data used to develop the habitat model were derived from a BLM database of murrelet survey records collected between 1993 and 2010 on Coos Bay BLM lands. Because of the threatened status of the murrelet (USFWS 1992), the BLM has regularly conducted intensive murrelet surveys since the early 1990s. Surveys were conducted according to protocols established by the Pacific Seabird Group (Evans Mack et al. 2003). Survey sites were not randomly selected because the Pacific Seabird Group protocols require that all potential habitat be surveyed prior to tree harvesting or other forest management projects that are likely to affect murrelets. Thus, all the sites used in our analysis were considered potential murrelet habitat that also were suitable for timber harvest. Potential habitat is defined as 1) old-growth (>200 yr old), 2) mature (>80 yr old), or 3) younger coniferous forest with remnant trees or nesting platforms. A nesting platform is a flat surface in a tree crown formed by a large-diameter branch or a tree deformity (>10 cm in diam, >10 m above forest floor; Evans Mack et al. 2003). Project areas containing extensive potential habitat were divided into survey sites <61 ha in area. Each survey site included multiple stations at which surveys were conducted. To maximize the probability of detecting...
murrelets, stations were placed in openings, along roads, and often up to 50 m outside the survey site if an adjacent location afforded a better position for observations.

We did not use model-based estimation procedures to determine the occupancy status of stands (e.g., MacKenzie et al. 2006). Instead, we classified occupancy status according to categories of observed murrelet behaviors that have been standardized by the Pacific Seabird Group. This approach permitted maps and other products of our analysis to be consistent with the murrelet occupancy classification used by the BLM, and thus fulfill a commitment that our study yield useful habitat assessment tools for the agency.

Murrelet surveys are designed to collect data for classifying survey sites as 1) “occupied” by murrelets, with 2) murrelet “presence,” or 3) “probable absence” (Evans Mack et al. 2003). An “occupied” site is defined as a site where sub-canopy behaviors or signs of nesting were observed during at least one visit to any of the stations within the site. “Presence” is defined as the detection of above-canopy flight or non-stationary aural detections at a site. “Probable absence” is defined as a site with potential habitat where no murrelets were detected after the requisite numbers of surveys. Pacific Seabird Group protocols require that survey sites are visited a minimum of 5 times over a 2-year period; more visits are required when murrelets are detected but occupancy behaviors were not observed (Evans Mack et al. 2003). Although we did not explicitly incorporate estimates of detectability in classifying occupancy status of survey sites, surveys conducted according to Pacific Seabird Group protocols should result in a >95% correct classification of sites as “occupied,” “presence,” or “probable absence” (Evans Mack et al. 2003:14).

We excluded sites classified as “presence” by the BLM because evidence of nesting was insufficient, and we used the survey station records to identify forest stands as “occupied” or “probable absence” according to the following procedure. We delineated 6,522 forest stands on the Coos Bay BLM District based on homogeneity of composition and structure using a BLM GIS (geographic information system) layer of forest inventory. We dissolved boundaries that represent public land survey system section lines, not actual differences in forest vegetation. Because of the ambiguity surrounding the precise locations of murrelet nest trees, the Pacific Seabird Group protocol recommends that all survey sites be classified as occupied even if sub-canopy behaviors were observed at only one station at the site (Evans Mack et al. 2003). We adopted this practice for our stand-level analysis. Because of the long temporal gap between some murrelet surveys and the LiDAR flight, we used a GIS and 2009 satellite imagery to visually inspect every site classified as “occupied” or “probable absence” for evidence of tree harvest since murrelet surveys that would have confounded our analysis. Based on this inspection, we excluded a small number of sites where murrelets were undetected during surveys and the stand had been subsequently harvested before the LiDAR flight. Our classification procedure resulted in 121 forest stands identified as “occupied” and 302 surveyed stands as “probable absence” (henceforth referred to as occupied and unoccupied stands, respectively).

To validate the stand-level habitat model with an independent set of murrelet observations, we acquired murrelet survey records collected between 1992 and 2012 on the Elliott State Forest. These records were the result of surveys conducted at 2,870 stations according to the same Pacific Seabird Group protocols as were used on BLM lands. We used the same methods to classify the occupancy status of Elliott State Forest stands as previously described for the BLM stand-level analysis. The resulting list contained 130 stands classified as occupied and 510 classified as unoccupied.

Nest tree data.—We also had access to locations for 11 trees in the Elliott State Forest known to be used by murrelets for nesting. These tree records were from a study of murrelet nest sites conducted on Oregon state forest lands between 1995 and 1999 (Nelson and Wilson 2002). Nests were discovered by climbing trees on plots randomly located within stands where previous surveys had revealed sub-canopy murrelet activity.

Figure 1. Location of marbled murrelet surveys (1993–2010) and LiDAR data collection (2008) in the Coast Range of southwestern Oregon, USA, including the Coos Bay District of the Bureau of Land Management (BLM) and the Elliott State Forest, managed by the Oregon Department of Forestry.
LiDAR Data and Other Explanatory Variables

LiDAR data were collected between 3 May and 28 September 2008, using a Leica ALS50 Phase II sensor flown in a Cessna Caravan 208B (Cessna, Wichita, KS). Average pulse density was 8.1/m² and up to 4 returns/pulse were recorded. Vertical root mean square error was 0.05 m. Raw LiDAR point files (LAS format) were processed by the US Forest Service, Pacific Northwest Research Station, using the FUSION software package Gridmetrics routine (McGaughy 2009) to produce a set of metrics describing forest cover, forest height, and topographic characteristics for the entire project area (approx. 652,000 ha) at a pixel size of 22.9 m. The FUSION routine used a height cutoff of 1 m for removing LiDAR returns near the ground, and a canopy threshold of 2 m. The upper limit for outliers also was adjusted to provide for trees of height >91 m. The ground surface was subtracted from the highest hits layer to produce a canopy surface raster at the 0.9-m pixel size. Canopy Surface Ratio (a measure of canopy surface area divided by underlying planimetric area) was calculated with the Digital Elevation Model (DEM) surface toolbar for ArcGIS 10 (Jenness 2004). We used Canopy Surface Ratio to represent canopy surface ruggedness.

To evaluate the performance of LiDAR-derived variables against those from other remote-sensing methods, we derived from Gradient Nearest Neighbor data for the Coos Bay BLM District the same set of variables used by Raphael et al. (2011) to model murrelet habitat suitability for the Oregon Coast (Table 1). Raphael et al. (2011) used 9 vegetation and physiographic attributes correlated with murrelet occupancy derived from the 2006 Gradient Nearest Neighbor imputation model created for the Northwest Forest Plan effectiveness monitoring (Moeur et al. 2011), and LandTrendr data (Kennedy et al. 2010).

For each stand polygon, we calculated summary statistics (min., mean, max., SD) for each FUSION raster, Canopy Surface Ratio, and the Gradient Nearest Neighbor variables. We calculated distance to the Pacific Ocean from stand center using the arcs coded as “coastline” in the National Hydrography Dataset.

Statistical Analysis

Stand-level analysis.—Because the probability of correct classification of sites as occupied or unoccupied following Pacific Seabird Group protocols is considered to be high enough (≥95% according to Evans Mack et al. 2003), occupancy was treated as a binary response variable and logistic regression models were fit to predict the probability of occupancy $\pi(x)$ for a stand given values of explanatory variables $x$, such as stand size, distance to coast, and variables derived from LiDAR metrics:

$$\pi(x) = \frac{\exp(g(x))}{1 + \exp(g(x))}$$

where $g(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$ is the logit of the multiple-regression model with $p$ explanatory variables $x$ and corresponding coefficients $\beta$. The “Gridmetrics” command in FUSION produces >60 metrics that can be used to describe forest structure with LiDAR elevation and intensity data, so it was necessary to go through a process of selecting relevant variables to reduce the set of candidate explanatory variables. We fit a univariable logistic regression model for each variable, and retained only those variables that were significant at least at the 0.05-level for stepwise model selection. Stepwise regression using the step() function in R (R Development Core Team 2011) was performed using a combination of forward and backward selection. The variables in the final model were examined for collinearity using variance inflation factors. If the variance inflation factors values indicated collinearity among the explanatory variables (variance inflation factors >4), the variables that explained most

Table 1. Variables used in a model developed by Raphael et al. (2011) to quantify potential habitat for marbled murrelets in Oregon, USA. Raphael et al. (2011) derived these variables from the 2006 Gradient Nearest Neighbor imputation model created for the Northwest Forest Plan effectiveness monitoring (Moeur et al. 2011), and LandTrendr data (Kennedy et al. 2010). We compared the performance of these variables in a scaled-down model of murrelet habitat in the Coos Bay Bureau of Land Management District, southwestern Oregon, with an alternative model using variables derived from LiDAR data collected between 3 May and 28 September 2008. “GNN” refers to the Gradient Nearest Neighbor model of landscape attributes for the Pacific Northwest (Oohmann and Gregory 2002).

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description</th>
<th>Unit</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>CANCOV_CON</td>
<td>Percent canopy cover of all conifers</td>
<td>Percentage</td>
<td>GNN</td>
</tr>
<tr>
<td>CANCOV_HDW</td>
<td>Percent canopy cover of all hardwoods</td>
<td>Percentage</td>
<td>GNN</td>
</tr>
<tr>
<td>MNDBHB_CON</td>
<td>Basal-area weighted mean diam of all live conifers</td>
<td>Inches</td>
<td>GNN</td>
</tr>
<tr>
<td>PLATFORMS</td>
<td>Platforms per acre derived from GNN. See Raphael et al. (2011)</td>
<td>No. per acre</td>
<td>Derived from GNN</td>
</tr>
<tr>
<td>MULTISTORY_50</td>
<td>Percentage of 50-ha circular area classified as GNN IMAP_LAYERS (no. of tree-canopy layers present) equal 3. See Raphael et al. (2011).</td>
<td>Percentage</td>
<td>Derived from GNN</td>
</tr>
<tr>
<td>PCTMATURE_50</td>
<td>Percentage of 50-ha circular area classified as GNN VEGCLASS 10 (large conifer, moderate to closed canopy) or 11 (giant conifer, moderate to closed canopy). See Raphael et al. (2011).</td>
<td>Percentage</td>
<td>Derived from GNN</td>
</tr>
<tr>
<td>SLOPE_PCT</td>
<td>Slope</td>
<td>Percentage</td>
<td>USGS NED 30-m DEM</td>
</tr>
<tr>
<td>BRIGHTNESS</td>
<td>Tasseled cap transformation of Landsat TM differentiating dry from wet soils</td>
<td>Index</td>
<td>LandTrendr</td>
</tr>
<tr>
<td>GREENNESS</td>
<td>Tasseled cap transformation of Landsat TM measuring presence and density of green vegetation</td>
<td>Index</td>
<td>LandTrendr</td>
</tr>
</tbody>
</table>
variability were retained in the model, while correlated variables were dropped. Variables that had not been included in the stepwise regression model were added one at a time to see if they were significant in combination with the other variables. We checked for significant interactions among the variables in the final model.

To compare the explanatory power of LiDAR-derived variables against other remote-sensing data, we fit a logistic regression model using variables derived from Gradient Nearest Neighbor and LandTrendr that were found to be the most important predictors of murrelet occupancy in Oregon by Raphael et al. (2011; Table 1). The original model in which these variables were used was developed to quantify potential habitat for marbled murrelets at the regional spatial scale of western Oregon (Raphael et al. 2011). Our goal was to test the performance of the same variables in a scaled-down model of murrelet habitat in the Coos Bay BLM District, for comparison with an alternative model using LiDAR-derived variables. However, when all variables identified by Raphael et al. (2011) as significant regional correlates of murrelet habitat were employed in our logistic regression model, collinearity issues arose. Although the Maxent model used by Raphael et al. (2011) was “more stable in the face of correlated variables […] so there is less need to remove correlated variables” (Elith et al. 2011:50), logistic regression models are sensitive to collinearities among explanatory variables (Hosmer and Lemeshow 2000:140). Some of the Gradient Nearest Neighbor variables were inherently correlated with each other; for example, the inverse relationship between percent canopy cover of conifers and percent canopy cover of hardwoods. Dependencies among LandTrendr, Gradient Nearest Neighbor, and Gradient Nearest Neighbor–derived variables also were expected because LandTrendr variables were used to derive the Gradient Nearest Neighbor variables, which were then used in the development of the Gradient Nearest Neighbor–derived variables. Therefore, we fit a reduced Gradient Nearest Neighbor model by dropping insignificant and collinear variables from the full model one by one and retaining only the significant variables. We used this model, henceforth referred to as the Gradient Nearest Neighbor model, for comparisons with the LiDAR model.

We assessed model performance in several ways. First, we calculated the AUC statistic (Hosmer and Lemeshow 2000:160). Following the interpretation of AUC values by Hosmer and Lemeshow (2000:162), AUC \( \leq 0.5 \) suggests no ability of the model to discriminate between occupied and unoccupied stands. Area-under-curve values between 0.7 and 0.8 are considered to provide acceptable discrimination, while 0.8 \( \leq \) AUC \( < 0.9 \) and AUC \( \geq 0.9 \) provide excellent and outstanding discrimination, respectively.

Second, we used leave-one-out cross-validation to evaluate prediction accuracy. We fitted the final model for \( n - 1 \) stands at a time and predicted the occupancy probability \( \pi(x) \) for the stand that was omitted in the model fit. A stand was classified as occupied for \( \pi(x) \geq 0.286 \), a threshold value that reflected differential class sizes in the modeling data set (Chen et al. 2006, 121/423 = 0.286). We calculated the overall accuracy, sensitivity, specificity, and Cohen’s \( \kappa \) as indicators of model performance. Overall accuracy is the proportion of correctly predicted stands. Sensitivity is the proportion of observed occupied stands that were predicted to be occupied. Specificity is the proportion of observed unoccupied stands that were predicted to be unoccupied. Cohen’s \( \kappa \) is another measure of agreement between observed and predicted occupancy, which reaches a maximum of 1 when agreement is perfect. Following the interpretation of Cohen’s \( \kappa \) in Altman (1991:404), \( \kappa \leq 0.2 \) shows poor agreement and \( \kappa \)-values between 0.21 and 0.4 show fair agreement; 0.41 \( \leq \kappa < 0.6 \), 0.61 \( \leq \kappa < 0.8 \), and \( \kappa \geq 0.8 \) stand for moderate, good, and very good agreement, respectively.

Third, we provided a visual representation of stand classification by plotting histograms of the predicted occupancy probabilities.

Finally, we used 640 stands from the Elliott State Forest that were surveyed for murrelets (130 occupied, 510 unoccupied) as an independent validation data set. Occupancy was predicted with the LiDAR and Gradient Nearest Neighbor models to assess the ability of the LiDAR model to discriminate between occupied and unoccupied sites of an independent data set, and to compare performance of the 2 models. A stand was classified as occupied for \( \pi(x) \geq 0.203 \), a threshold value that reflected differential class sizes in the Elliott State Forest data set (Chen et al. 2006; 130/640 = 0.203).

Analysis of nest tree data.—We employed a use–availability design (see Keating and Cherry 2004) to compare canopy characteristics at the 11 nest trees with those at a random sample of points from the 7 stands that had nest trees. We used a \( 3 \times 3 \) grouping of 22.9-m pixels to summarize LiDAR-derived variables at each sample point, a spatial scale assumed to approximate a nest grove (i.e., included the nest tree and immediately neighboring trees). We sampled approximately one point (either nest tree or random) for every 15 acres (6 ha) in each stand. We used 500 iterations of random point selection, and analyses were performed on the 500 resulting data sets.

We started the variable selection process with a univariable analysis of each variable; variables that were significant at the 0.05-level were retained for step-wise model selection. Only the most significant of the 4 summary statistics (min., max., mean, SD) of each LiDAR variable was retained if more than one was significant at the 0.05-level. A combination of forward and backward selection was performed using step() in R (R Development Core Team 2011). The variables that were selected in the majority of the 500 data sets were retained for further modeling.

We assumed that the data from our use–availability design were approximately equivalent to a case-control design (Keating and Cherry 2004), and therefore analyzed the data with logistic regression and interpreted the results in terms of odds ratios. We used the same variable selection process as for the stand occupancy analysis, above. The variance of a random stand effect, included to account for similarities of points within stands, was so small (<0.00001) as to be negligible. Therefore, we assumed all stands had a similar
logistic curve. We used PROC GLIMMIX in SAS (SAS Institute, Inc. 2012) to determine odds ratios and their corresponding confidence intervals.

RESULTS

Five LiDAR-derived variables and distance-to-coast were selected in our final model (Table 2; Fig. 2). This model will henceforth be referred to as the LiDAR model. Maximum percentage of all returns above the mean height (ALLCVABVMN_\text{max}), maximum of height distribution of 99th percentile (EL_{p99\_max}) and 10th percentile of first returns (EL_{p10\_max}), and standard deviation of percentage of first returns above the modal height (FRSTCVABVMD_std) tended to be larger in occupied stands (Fig. 3). Minimum kurtosis of height distribution (EL_{kurt\_min}) was smaller in occupied stands than in unoccupied stands. Occupied stands tended to be slightly closer to the coast.

The LiDAR model had an AUC value of 0.74 (which corresponds to acceptable discrimination) and a $\kappa$ value of 0.24 (which corresponds to fair agreement between observed and predicted occupancy). Overall accuracy, sensitivity, and specificity based on the confusion matrix (Table 3) equaled 64, 63, and 65, respectively. The LiDAR model predicted more unoccupied stands as occupied (Fig. 4A gray) than occupied stands as unoccupied (Fig. 4A white).

In spite of significant interactions of FRSTCVABVMD_std with both EL_{p99\_max} and EL_{p10\_max}, improvement of a model including interaction terms over the simple LiDAR model was marginal (AUC = 0.74, overall accuracy = 65, sensitivity = 66, specificity = 65, $\kappa$ = 0.26). Hence, we preferred the simpler LiDAR model without interactions.

Three Gradient Nearest Neighbor variables (SLOPE_PCT, PCTMATURE_50, and CANCOV_CON) were significant in the Gradient Nearest Neighbor model (Table 4). Predicted probabilities for the Gradient Nearest Neighbor were lower than for the LiDAR model.

### Table 2

Variable descriptions and coefficients, with standard errors and significance levels, for a model of marbled murrelet occupancy developed with LiDAR-derived variables for the Coos Bay Bureau of Land Management District, Oregon, USA. LiDAR data were collected between 3 May and 28 September 2008, and metrics were produced using FUSION software (McGaughey 2009) at a pixel size of 22.9 m.

| Coeff. | Variable description and ecological interpretation | Estimated coeff. | SE | Signif. levels
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>-15.416</td>
<td>3.609</td>
<td>***</td>
</tr>
<tr>
<td>ALLCVABVMN_\text{max}</td>
<td>LiDAR canopy-cover metric: percentage of all returns above mean canopy ht (stand max.). Interpretation: cover in upper portion of canopy.</td>
<td>0.078</td>
<td>0.037</td>
<td>*</td>
</tr>
<tr>
<td>EL_{p99_max}</td>
<td>LiDAR canopy ht distribution: 99th percentile of first returns (stand max.). Interpretation: max. ht of tallest trees.</td>
<td>0.048</td>
<td>0.016</td>
<td>**</td>
</tr>
<tr>
<td>EL_{p10_max}</td>
<td>Canopy ht distribution: 10th percentile of first returns (stand max.). Interpretation: max. ht of bottom of canopy.</td>
<td>0.068</td>
<td>0.024</td>
<td>**</td>
</tr>
<tr>
<td>FRSTCVABVMD_std</td>
<td>LiDAR canopy-cover metric: percentage of first returns above the modal ht (SD). Interpretation: variation in cover in upper canopy.</td>
<td>0.142</td>
<td>0.065</td>
<td>*</td>
</tr>
<tr>
<td>EL_{kurt_min}</td>
<td>Canopy ht distribution: kurtosis of ht distribution (stand min.). Interpretation: distribution of vegetation across canopy ht intervals.</td>
<td>1.940</td>
<td>0.976</td>
<td>*</td>
</tr>
<tr>
<td>Distance-to-coast (km)</td>
<td>Distance to the Pacific Ocean from stand center (source: National Hydrography Dataset)</td>
<td>-0.059</td>
<td>0.012</td>
<td>***</td>
</tr>
</tbody>
</table>

*0.01 < $P \leq 0.05$; **0.001 < $P \leq 0.01$; ***0 < $P \leq 0.001$.

Figure 2. Comparison of structural profiles derived from 2008 LiDAR data of example forest stands in the Coos Bay BLM District, Oregon, USA, representing 3 marbled murrelet occupancy classes. The table below the diagram lists the variables selected for a model of marbled murrelet occupancy and the variable values associated with each occupancy class. EL_{p99\_max} and EL_{p10\_max} reflect height at the top and bottom of the canopy, respectively; EL_{kurt\_min} describes kurtosis of canopy height distribution; ALLCVABVMN_\text{max} and FRSTCVABVMD_std reflect cover in the upper canopy. See Table 2 for detailed variable descriptions.
Neighbor model ranged from 0.06 to 0.71 for unoccupied stands and from 0.11 to 0.54 for occupied stands (Fig. 4B). For all explanatory variables in the LiDAR model except distance-to-coast, the same trends between occupied and unoccupied stands were observed for the Elliott State Forest as for the Coos Bay District. The surveyed stands in the Elliott State Forest tended to be slightly closer to the coast than those in the Coos Bay district (Fig. 3), but distance to

Figure 3. Distribution of values of explanatory variables selected in a logistic regression model of marbled murrelet occupancy for occupied (1) and unoccupied (0) stands. The model was developed using murrelet occupancy data collected between 1993 and 2010 and LiDAR data collected between 3 May and 28 September 2008 on the Coos Bay BLM District, and tested using murrelet occupancy data collected between 1992 and 2012 from the Elliott State Forest, southwestern Oregon, USA.
coast did not differ between occupied and unoccupied stands. Occupied stands in the Coos Bay District tended to have larger values for variables related to tree height and cover (ALLCVABVMN_max, EL_p99_max, EL_p10_max, and FRSTCVABVMD_std), and smaller values for the variable related to kurtosis of canopy height distribution (EL_kurt_min). The ranges of the LiDAR variables for Elliott State Forest occupied stands were similar to those observed for the Coos Bay District. However, the LiDAR variables for unoccupied sites in the Elliott State Forest tended to have a larger range of values than those observed for unoccupied sites in the Coos Bay district (Fig. 3).

When applied to the independent data set of the Elliott State Forest, the LiDAR model resulted in a Cohen’s $k$ of 0.18, thus showing poor agreement between predicted and observed occupancy for the Elliott State Forest. Based on the confusion matrix, the LiDAR model had an overall accuracy of 51 and sensitivity and specificity values of 91 and 41, respectively (Table 5; Fig. 5A). The Gradient Nearest Neighbor model, which already resulted in poor agreement for the Coos Bay data, performed even worse when applied to the Elliott State Forest data, with Cohen’s $k$ of 0.04. The Gradient Nearest Neighbor model predicted the majority of unoccupied stands to be occupied, resulting in low overall accuracy (31), with high sensitivity values (94) and low specificity values (15 [Table 5; Fig. 5B]).

Table 3. Comparison of results of leave-one-out cross-validation for models of marbled murrelet occupancy between 1993 and 2010 on the Bureau of Land Management Coos Bay District (OR, USA). The LiDAR model uses variables derived from data collected in 2008; the GNN Model uses variables identified by Raphael et al. (2013) as relevant to murrelet occupancy and derived from the 2006 Gradient Nearest Neighbor imputation model (Moeur et al. 2011).

<table>
<thead>
<tr>
<th>Status</th>
<th>LiDAR model</th>
<th>GNN model</th>
<th>Observed Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Occupied</td>
<td>Unoccupied</td>
<td>Occupied</td>
</tr>
<tr>
<td>Occupied</td>
<td>76</td>
<td>45</td>
<td>68</td>
</tr>
<tr>
<td>Unoccupied</td>
<td>107</td>
<td>195</td>
<td>130</td>
</tr>
</tbody>
</table>

Analysis of Nest Tree Data
Two variables related to tree height and cover (EL_p99_max and ALL1CVABVMD_mean) were significantly larger for nest trees than for random points for 500 and 499 of the 500 data sets, respectively (Fig. 6; example of one of the 500 data sets). EL_p99_max represents maximum canopy height; ALL1CVABVMD is the ratio of all returns above modal canopy height to the total number of LiDAR pulses (i.e., a measure of cover in the upper canopy). These 2 variables entered the final model in the majority of the data sets when the variable selection process was performed. Although distance to stand edge tended to be larger for nest trees than for random points (Fig. 6), the difference was not significant and it did not enter into the logistic regression model. On average across the 500 data sets, EL_p99_max and ALL1CVABVMD_mean achieved outstanding discrimination between nest sites and random points (average AUC = 0.91). Area-under-curve values ranged from 0.82 to 0.98, thus achieving at least excellent discrimination for all 500 data sets.

With ALL1CVABVMD_mean held constant, a 10-unit increase in EL_p99_max increased the odds of a point being a nest tree 2 times, on average, across the 500 data sets, ranging between 1.1 and 3.8. The lower and upper 95% confidence limits of the odds ratios, on average, across the 500 data sets were 1.4 and 6.2, respectively. A 10-unit
Table 4. Model coefficients with standard errors, significance levels, and variance inflation factors (VIF) for a model of marbled murrelet occupancy between 1993 and 2010 on the Coos Bay, Oregon Bureau of Land Management District (OR, USA) using variables identified by Raphael et al. (2011) as relevant to murrelet occupancy and derived from the 2006 Gradient Nearest Neighbor imputation model (Moeur et al. 2011).

<table>
<thead>
<tr>
<th>Model coeff.</th>
<th>Estimated coeff.</th>
<th>SE</th>
<th>Significance levels</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.159</td>
<td>0.826</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLOPE_PCT</td>
<td>0.021</td>
<td>0.009</td>
<td>*</td>
<td>1.02</td>
</tr>
<tr>
<td>PCTMATURE_50</td>
<td>0.016</td>
<td>0.006</td>
<td>**</td>
<td>1.59</td>
</tr>
<tr>
<td>CANCOV_CON</td>
<td>-0.038</td>
<td>0.015</td>
<td></td>
<td>1.59</td>
</tr>
</tbody>
</table>

*0.01 < P ≤ 0.05; **0.001 < P ≤ 0.01.

Table 5. Classification of occupancy by marbled murrelets between 1993 and 2010 for the Elliott State Forest, Oregon, USA, based on a model using LiDAR-derived independent variables and a model using variables identified by Raphael et al. (2011) as relevant to murrelet occupancy and derived from the 2006 Gradient Nearest Neighbor imputation model (Moeur et al. 2011).

<table>
<thead>
<tr>
<th>Status</th>
<th>Predicted</th>
<th>GNN model</th>
<th>Observed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LiDAR model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupied</td>
<td>Occupied</td>
<td>118</td>
<td>121</td>
<td>130</td>
</tr>
<tr>
<td>Unoccupied</td>
<td>Unoccupied</td>
<td>299</td>
<td>431</td>
<td>510</td>
</tr>
</tbody>
</table>

Figure 5. Distribution of predicted marbled murrelet occupancy probabilities in the Elliott State Forest, southwestern Oregon, USA, from (A) variables derived from LiDAR data collected between 3 May and 28 September 2008 data and (B) variables derived from the 2006 Gradient Nearest Neighbor imputation model (GNN; Moeur et al. 2011) for unoccupied (dark gray) and occupied stands (white). Overlap in distributions shown in light gray. The vertical line represents the cut-off value (=0.203) used to determine occupancy.

Figure 6. Distributions of explanatory variables characterizing marbled murrelet nest trees (Nest tree = 1) and random points (Nest tree = 0) in the Elliott State Forest, southwestern Oregon, USA, for an example data set resulting from one of 500 iterations of random point selection. EL_p99_max and ALL1CVABVMD_mean are derived from 2008 LiDAR data and represent maximum canopy height (99th percentile of first returns) and the ratio of all returns above modal canopy height to the total number of LiDAR pulses (i.e., a measure of cover in the upper canopy), respectively.
increase in ALLCVABVMD_mean for fixed values of EL_p99_max increased the odds of a point being a nest tree 2.1 times on average, ranging from 1.26 to 16.41. Lower and upper 95% confidence limits averaged 1.1 and 5.0, respectively.

DISCUSSION

Our results illustrate how LiDAR-derived variables can be used to improve model prediction accuracy for murrelet habitat at local spatial scales over models using variables derived from traditional remote sensing data (GNN and LandTrendr). Whereas the model using GNN and GNN-derived variables developed by Raphael et al. (2011) is useful for estimating region-wide availability of suitable habitat for murrelets (Raphael et al. 2011:29), our model based on LiDAR-derived variables provides a refined estimate of the availability of murrelet nesting habitat, useful for discriminating among mature forest stands on smaller, sub-regional spatial scales (Fig. 7). Because the LiDAR model outperformed the Gradient Nearest Neighbor model in every aspect of model evaluation that we used, we conclude that the LiDAR explanatory variables, in combination with distance-to-coast, have more explanatory power than the available Gradient Nearest Neighbor and Gradient Nearest Neighbor–derived variables for modeling habitat at stand to watershed spatial scales. Our work shows that incorporating LiDAR-derived variables into models can more accurately represent murrelet nesting habitat. Models incorporating LiDAR-derived variables may provide more refined estimates of the availability of murrelet nesting habitat at local spatial scales (Fig. 7) than do models that are based on Gradient Nearest Neighbor and Gradient Nearest Neighbor–derived variables for 2 main reasons. First, whereas Gradient Nearest Neighbor data are limited to 2-dimensional habitat features, LiDAR variables provide empirical, 3-dimensional measurements of forest structure (Lefsky et al. 1999, 2002) that can be directly related to functional habitat use (Palminteri et al. 2012). Furthermore, the ability of LiDAR to provide continuous, rather than categorical, metrics describing vegetation structure allows for more realistic representation of the ecological gradients that influence species’ distributions (McGarigal and Cushman 2005). In the case of the murrelet, a species that is highly influenced by canopy structure, LiDAR variables are likely to have better discriminatory power than 2-dimensional, categorical Gradient Nearest Neighbor variables because of their ability to quantify 3-dimensional gradients in canopy complexity with continuous metrics.

Secondly, Raphael et al. (2011) developed their Maxent model using presence-only and background data. Their model provided excellent predictive power at a broad regional scale because the background data represented a full range of stand age classes, including stands that were too young to be considered potential murrelet nesting habitat. However, the contribution of variables to models of species distribution has been shown to change with the size of the area from which pseudo-absence points are drawn (VanDerWal et al. 2009). The variables in the Raphael et al. (2011) model were selected to discriminate among the broader range of variability represented across the region, whereas the LiDAR-derived variables selected from our spatially constrained data used forest structural attributes to discriminate among stands that were relatively similar in age.

The reduction in model performance that we observed when applying the LiDAR model to the Elliott State Forest was expected because of differences between Elliott State Forest and the Coos Bay District in 1) the ranges of some explanatory variables (the Elliott State Forest included younger stands than Coos Bay), and 2) the degree of difference between unoccupied and occupied stands for some explanatory variables (Fig. 3). Although, the LiDAR model performance decreased when applied to the independent data set of the Elliott State Forest, it still outperformed the Gradient Nearest Neighbor model, which performed very poorly for the Elliott Forest in terms of overall accuracy, specificity, and Cohen’s $k$. The improved performance of the LiDAR model over the Gradient Nearest Neighbor model when applied to the independent Elliott State Forest data provides evidence that, at stand to watershed spatial scales, LiDAR variables have greater explanatory power than do Gradient Nearest Neighbor–derived variables.

The LiDAR-derived variables selected in our model described dimensions of canopy structure that reflect age-related stand characteristics and are consistent with murrelet nesting ecology. The ability of LiDAR to discriminate among age and structure classes of forest stands has previously been demonstrated (e.g., Lefsky et al. 1999, Falkowski et al. 2009). Large values of the maximum height of the canopy (EL_p99_max) combined with high values of

Figure 7. Example output of probability of occupancy by marbled murrelets in the Coos Bay BLM District, Oregon, USA, from a model using LiDAR-derived variables to quantify habitat suitability at a local spatial scale. Gray areas are private lands that were not modeled. Marbled murrelet survey data were used to develop model were collected between 1993 and 2010; LiDAR data collected in 2008.
cover in the upper portion of the canopy (ALLC-VABVMN_max, ALL1CVABVMD_mean), which we found to be positively associated with murrelet occupancy and nest site locations, indicate large tree crowns that have developed with age (Parker and Russ 2004). Kurtosis of height distribution, a LiDAR-derived variable describing vertical stratification of canopy vegetation, also indicates differences among forest age and structure classes. Jones et al. (2012) found that variables describing kurtosis and standard deviation of height were significant in differentiating between young and mature forests in heavily managed coastal forests in British Columbia. Lower values of kurtosis for height distribution were indicative of structurally complex, multilayered canopies in natural stands, whereas elevated kurtosis values indicated structurally simple canopies in riparian hardwood forest types (Antonarakis et al. 2008). Our finding that stands occupied by murrelets had lower minimum kurtosis values than unoccupied stands indicates greater vertical stratification of canopies in occupied stands, and is consistent with the established association of murrelets with multi-storied canopies (e.g., Hamer and Nelson 1995, Nelson and Wilson 2002, Raphael et al. 2011). The LiDAR-derived kurtosis metric provides a directly measured index of canopy layering as an alternative to a variable (no. of canopy layers) that had previously been defined categorically and estimated subjectively.

The selection of EL_p10 max in the occupancy model was unexpected because this variable reflects the height of the bottom of the canopy, which is not typically used for nesting. Although this variable may be an indicator of flight space, murrelets do not necessarily fly below the level of the live crown to access their nests. Alternatively, maximum values of EL_p10 likely indicate tall trees with high live crowns, which is characteristic of nest trees that occur in older-aged forests.

Murrelets have extremely fast flight, and typically make use of openings in the canopy near or adjacent to nests as a means for stalled landings on, and taking off from the nest limb (Nelson and Peck 1995, Singer et al. 1995). The higher scores of FRSTCVABVMD_std, with which probability of murrelet occupancy was positively associated, indicate potential selection for high variability of cover (i.e., gappiness) in the upper canopy (above modal ht). This variable may be useful as an indicator of gaps among overstory trees in late-seral forest that form 3-dimensional corridors that could be used as flight routes through the forest by murrelets (Singer et al. 1995). In addition, it could indicate remnant old-growth trees that might provide suitable nesting habitat in patches of younger forest. Direct quantification of the dimensions of canopy gaps on a spatial scale relevant to murrelets has not previously been feasible using other remote-sensing or field methods. LiDAR data have been used to derive metrics describing the horizontal and vertical properties of gaps, including gap connectivity (Koukoulas and Blackburn 2004). Such metrics were not among the FUSION metrics available for our analysis, but may be useful for identifying flight paths used by murrelets. Evaluated in total, the variables identified by our model as important correlates of occupancy describe ecologically meaningful, 3-dimensional features of canopy structure that have not been easily or accurately quantified by other mensuration methods.

None of the variables that entered our model allowed perfect discrimination of occupied and unoccupied stands (Fig. 3). However, it is important to consider that all stands in our sample were considered potential habitat when evaluating the performance of the model. The difficulty of discriminating between occupied and unoccupied stands is reflected in the distribution of the predicted probabilities of occupancy (Fig. 4). The resulting inflated rate of error of commission (assigning high probability of occupancy to unoccupied stands) can be attributed to the fact that the population used to develop the model was limited to stands that were all considered to be potential murrelet nesting habitat based on age (mostly >80 yr old) and structural characteristics such as the presence of residual trees (K. Palermo, Bureau of Land Management, personal communication). The estimates of occupancy are conservative because they erred on the side of predicting unoccupied stands as occupied (false positives). For models that are used to assess habitat for threatened or endangered species, erring conservatively on the side of false positives as ours did is in most cases more acceptable than false negatives (Morrison et al. 2006). Furthermore, inter-annual variability in occupancy status of suitable stands by murrelets can occur for many reasons, including mortality of territorial birds and changes in off-shore foraging conditions (Evans Mack et al. 2003, Peery et al. 2004), providing ecological justification for the bias of our model toward false positive prediction errors (Fielding and Bell 1997).

In spite of the difficulty in discriminating among occupancy classes related to a limited range of variation in stand characteristics (Aberg et al. 2000), our model nonetheless was able to differentiate habitat based on fine-scale structural attributes. These results indicate that the detailed descriptions of habitat features at relevant spatial scales available through LiDAR may provide superior discriminatory power among habitats that may appear similar from a human perspective (Tattoni et al. 2012). By providing new ways to quantify three-dimensional forest structure, LiDAR expands the lexicon that can be used to describe and understand the physical features that influence habitat selection and use, thereby increasing our capacity to understand habitat relationships. In this case, variables representing maxima in tree height and cover provided insight into the fine scale structural attributes that characterized stands occupied by murrelets. These variables highlight the importance of individual, old-growth components that are most likely to provide nesting platforms in mature and late-seral forest stands (Nelson 1997, Baker et al. 2006).

**MANAGEMENT IMPLICATIONS**

Our model of murrelet occupancy directly addresses needs identified in the Recovery Plan for refined measures of nest site structure and selection by murrelets. It provides a tool to improve estimates of nesting habitat availability specifically on Coos Bay BLM lands. Because LiDAR metrics are
obtainable and repeatable over large areas, our model can be used to monitor fine-scale changes in murrelet habitat availability in the Coos Bay BLM District over time. As no model should be used to extrapolate beyond the data used in its construction, we do not recommend application of our model beyond the Coos Bay BLM District without additional development and testing. However, our work demonstrates the advantages of using LiDAR over other remote sensing data for estimating habitat availability at local spatial scales.

ACKNOWLEDGMENTS

Funding for this study was provided by the Oregon State Office of the U.S. Department of the Interior Bureau of Land Management. We thank J. Guetterman, J. Heaney, and C. Schumacher of the Coos Bay Bureau of Land Management (BLM) for their patient explanations of BLM survey and data management protocols. R. Smith of the Oregon Department of Forestry assisted with interpretation of Elliott State Forest databases. G. McFadden of the Oregon State Office of the BLM was instrumental in coordinating our study within a broader framework of federal and state LiDAR research applied to natural resource management. We thank J. P. Hollenbeck and 2 anonymous reviewers for helpful comments that improved the quality of this manuscript. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

LITERATURE CITED


U.S. Department of Agriculture [USDA], and U.S. Department of Interior [USDI]. 1994. Record of decision for amendments to Forest Service and Bureau of Land Management planning documents within the range of the northern spotted owl: standards and guidelines for management of habitat for late-successional and old-growth forest related species within the range of the northern spotted owl. U.S. Department of Agriculture [USDA], and U.S. Department of Interior [USDI], Washington, D.C., USA.


U.S. Department of Agriculture [USDA], and U.S. Department of Interior [USDI]. 1994. Record of decision for amendments to Forest Service and Bureau of Land Management planning documents within the range of the northern spotted owl: standards and guidelines for management of habitat for late-successional and old-growth forest related species within the range of the northern spotted owl. U.S. Department of Agriculture [USDA], and U.S. Department of Interior [USDI], Washington, D.C., USA.


Associate Editor: Koper.