

Considerations for the validation of species-habitat models

Jennifer M. Psyllakis¹ and Michael P. Gillingham²

Abstract

The multitude of approaches to wildlife-habitat modeling reflect the broad objectives and goals of various research, management, and conservation programs. Validating models is an often overlooked component of using models effectively and confidently to achieve the desired objectives. Statistical models that attempt to predict the presence or absence of a species are often developed with logistic regression. In this paper, we review principles of validating logistic regression models, measures of prediction accuracy assessment, and potential sources of prediction errors in a forest-management context. Based on our work in central British Columbia, Canada, we use two species, red squirrel (*Tamiasciurus hudsonicus*) and pileated woodpecker (*Dryocopus pileatus*), to examine the prediction accuracy of logistic regression models. Because different types of detections may increase spatial or temporal uncertainty in empirical models, we use either audio and visual observations or sign (e.g., forage, nests) as the response variable to compare model results. We also compare models using data from Vegetation Resource Inventory (VRI; the regional inventory used by government for timber quality and quantity on the landscape developed from photo interpretation and ground measurements), local plot measurements collected as part of this study, and a combination of the two as explanatory variables in the statistical models. Using detections of sign as the dependent variable resulted in models with higher predictive discrimination for both species, but the difference was not as great for red squirrel, with small home-range sizes, as for pileated woodpeckers, that use landscapes at much larger scales. The final models selected based on low Akaike's Information Criterion (AIC) and predictive discrimination included a combination of locally measured independent variables and VRI data as explanatory variables. Our results suggest that detection type may affect model outcomes and relatively small investments in data acquisition can improve predictive discrimination. We discuss considerations for the development and validation of statistical models intended for use in biodiversity monitoring.

1 Natural Resources and Environmental Studies Graduate Program, University of Northern British Columbia, 3333 University Way, Prince George, BC, V2N 4V9

2 Associate Professor, Natural Resources and Environmental Studies Institute, University of Northern British Columbia, 3333 University Way, Prince George, BC V2N 4V9

Citation: Psyllakis, J.M. and M.P. Gillingham. 2007. Considerations for the validation of species-habitat models. Paper presented at the "Monitoring the Effectiveness of Biological Conservation" conference, 2-4 November 2004, Richmond, BC. Available at: <http://www.forrex.org/events/mebc/papers.html>

Notes: This paper has been published "as presented" at the Conference, and has not been peer reviewed prior to posting. © Copyright 2007 by the authors

Introduction

The worldwide rate of species decline has resulted in international agreements (e.g., Convention on Biological Diversity; United Nations Environment Programme 1992), national strategies (e.g., Canadian Biodiversity Strategy; Minister of Supply and Services Canada 1995), and provincial guidelines (e.g., Landscape Unit Planning Guide; Province of British Columbia 1999) aimed at preserving species diversity in reserves, as well as landscapes managed for industrial purposes. Forests provide habitat necessary for thousands of species, as well as considerable economic and social resources to humans. Activities driven by anthropogenic values have contributed to habitat loss and fragmentation resulting in a global decline of many species from forest environments (Hunter 1990; Harper and Hawksworth 1995). A strategy for conserving, or minimizing impacts on, biodiversity is also required as part of most sustainable forest certifications (e.g., Forest Stewardship Council, Sustainable Forestry Initiative).

Models that build on the relationships between species and their environments provide an important tool for biodiversity monitoring. To be effective, however, species-habitat models need to be explicitly tested (Guisan and Zimmerman 2000; Scott et al. 2002) and model validation is a vital component to confidently implement monitoring objectives (Ottaviani et al. 2004). The process of validation can increase the understanding of species-habitat relationships (Fleishman et al. 2002), the limitations of the statistical model application, and whether or not the model is appropriate for its intended use (Rykiel 1996).

Statistical approaches to species-habitat modeling vary with modeling objectives and available types of data (Guisan and Zimmermann 2000). Logistic regression is often the preferred method to model species presence or absence in relation to habitat variables (Manel et al. 1999; Pearce and Ferrier 2000). The resulting logistic equation predicts the probability of species presence given independent variables and parameters (i.e., the intercept and β coefficients). Validation of logistic regression models usually focuses on the accuracy of predictions (Fielding and Bell 1997; Johnson 2001) and is judged on: 1) reliability - the accuracy of the predicted likelihood of occurrence; and 2) discrimination - the ability of the model to accurately distinguish between occupied and unoccupied sites (Pearce and Ferrier 2000).

Calculating the area under the receiver operating characteristics (ROC) curve is a favoured measure used to assess the predictive accuracy of logistic models, when presence and true absence data are available (Fielding and Bell 1997; Pearce and Ferrier 2000). The ROC value is calculated by plotting the number of sites where presence is correctly predicted divided by the total number of positive sites (sensitivity), against the fraction of incorrect cases where presence is predicted (1-specificity; Table 1) across available thresholds. The area under the resulting curve is an estimate of predictive accuracy not biased by threshold probabilities (i.e., $P > 0.5$ designated as presence; Fielding and Bell 1997) or species prevalence (i.e., one outcome greatly outnumbers the other; Manel et al. 2001). A ROC value is interpreted as the percentage of time that a random selection from the positive class will have a higher predictive score than a randomly drawn case from the negative class (Deleo 1993). ROC values of 0.5 indicate the explanatory variables do not improve discrimination beyond random assignment and 1.0 indicates perfect discrimination. A value below 0.5 indicates the model performs more poorly with the explanatory variables than without them. When presence-only data is all that is available (e.g., absence can only be defined as not detected) other techniques for validating models should be applied (see Ottaviani et al. 2004).

Table 1. Matrix of prediction classifications describing the possible outcomes of presence or absence and the associated indices that can be used to describe predictive performance.

| Classification matrix | | Definitions of the four indices of performance | |
|-----------------------|------------------|------------------------------------------------|-----------------------------------------|
| | Recorded Present | Recorded Absent | Sensitivity = $A / (A + C)$ |
| Predicted present | A | B | Specificity = $D / (B + D)$ |
| Predicted Absent | C | D | False positive fraction = $B / (B + D)$ |
| | | | False negative fraction = $C / (A + C)$ |

Statistical models developed to monitor biodiversity are often applied to areas beyond the location where data were collected (Mac Nally 2002). To have the highest confidence in the reliability and discrimination of the model, an external data set (independent from the data used to build the model) should be used in validation (Guisan and Zimmerman 2000; Fleishman et al. 2002). Obtaining an independent data set is often infeasible due to various constraints. Withholding data to test the model or using a resampling technique are options to an independent data set (Fielding and Bell 1997; Boyce et al. 2002) but will still result in optimistic prediction accuracy (Verbyla and Litaitis 1989; Fielding and Bell 1997; Pearce and Ferrier 2000).

Further assessment of the sources of prediction error can lead to improved understanding of the ecological associations among the species and its habitat as well as the utility of the model. Prediction errors can occur due to errors in specifying the model, inappropriate statistical assumptions, measurement errors, and uncertainty related to natural variation (Elith and Burgman 2002; Fielding 2002). Two potential sources of error related to specification error come from bias in detection type and misappropriate inclusion or exclusion of explanatory variables.

Species presence can be established from a variety of detection types including visual detections, auditory detections, tracks, scat, forage sign. Visual detections have little or no spatial or temporal uncertainty associated with them. In contrast, auditory detections are temporally certain, but potentially can have a high degree of spatial uncertainty given that, for some species, calls can travel several hundred meters. Mobile species may call while in flight adding additional uncertainty as they may be in transit between activity areas of their home ranges. Feces, tracks, dens, and nests are all exact in space but vary in their temporal certainty. Certain sign (e.g., cavity nests, dens) are very persistent on the landscape and habitat structure can change significantly around the sign; it may be unclear if the location is still useful to the species it is associated with. Uncertainty in the response data potentially leads to unexplainable variation in the model and reduced reliability and discrimination (Pearce and Ferrier 2000). Determining species absence is more ambiguous, and may require that an alternative modeling and validation approach is adopted (see MacKenzie et al. 2002; Hirzel et al. 2002; Ottaviani et al. 2004).

Misappropriate inclusion or exclusion of explanatory variables also leads to prediction error of species presence or absence. Influence on species distribution often includes variables that are not typically measured in association with studies of species-habitat relationships (e.g., intra- and inter-species interactions). In the context of forest management, habitat data used

as predictor variables can be obtained from sampling designed specifically for species monitoring programs, but also data that is available as part of forest inventories related to harvesting activities. Vegetation Resource Inventory (VRI) is an inventory methodology established by the British Columbia (BC) Provincial government; the primary objective of the inventory is to assess the quantity and quality of timber (Province of British Columbia 2002). When the inventory is complete, data will be available for forested landbase of the BC and may be acquired at little cost. The limits associated with the type of environmental data available or acquisition of appropriate data, both for statistical model construction and its use in monitoring activities, may influence the effectiveness of achieving the desired model objective(s).

Additional statistical model assessment includes an evaluation of the variation explained by the independent variables. In the logistic model, this cannot be calculated in the same way as for other regression models and there are several alternative measures. Pseudo- R^2 , calculated by dividing the model X^2 value by the -2 log-likelihood of the model with only the intercept, results in a proportional reduction in the absolute value of the log-likelihood (Nagelkerke 1991) and is the recommended measure to describe variance explained for logistic regression (Menard 2001). What constitutes a “good” pseudo- R^2 value is unclear, but it can be useful when comparing across models that do not have the same response data (e.g., comparing between detection types). A model with a low pseudo- R^2 may still have high predictive efficiency.

In this paper we investigate the effects of two types of potential model error on the predictive accuracy and evaluation of species-habitat models: 1) the type of detection used to determine species presence and, 2) different sources data for explanatory variables. We use two species as examples of how differences in the response variable and habitat data source may affect the model selection and discrimination, red squirrel (*Tamiasciurus hudsonicus*) and pileated woodpecker (*Dryocopus pileatus*). We selected these species for examples because they are relatively common in our data study area, their sign is distinctive (i.e., there is little uncertainty in species identification from sign), and the red squirrel uses the landscape at small-spatial scales and is not highly mobile in short-time periods, whereas the pileated woodpecker uses the landscape at relatively large scales and is highly mobile in short time periods.

Methods

Study Area

The data were collected near the northern extent of the Interior Douglas Fir (IDF) Biogeoclimatic zone in central British Columbia (51°51'N, 121° 50'W; Meidinger and Pojar 1991; Figure 1) from May 2001 through January 2004. The IDF is characterised by stands of closed- and open-canopy Douglas fir (*Pseudotsuga menziesii*). At higher elevations, or where crown fires have occurred in the past, lodgepole pine (*Pinus contorta*) is common and there are localised stands of hybrid white spruce (*Picea engelmannii* x *glauca*) and trembling aspen (*Populus tremuloides*). At lower elevations some areas have large grassland communities; non-forested wetlands are common throughout the study area.

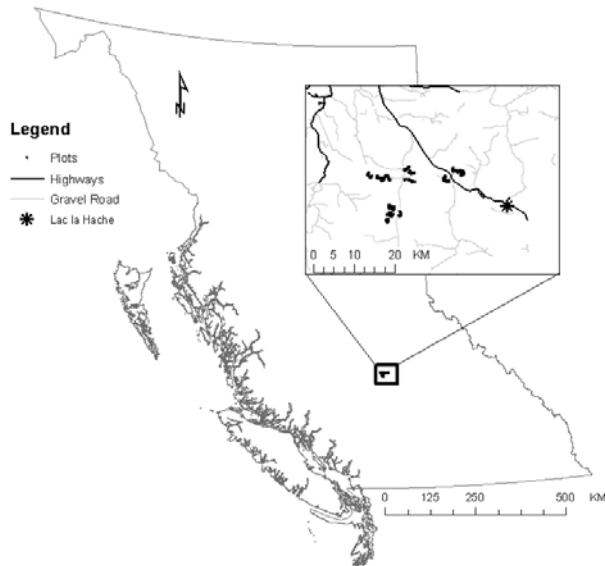


Figure 1. The location of the study area (51°51'N, 121° 50'W) in British Columbia, Canada. We established 243 plots (inset) and surveyed them May 2001 – January 2004. The town of Lac la Hache is mapped for reference.

Fire was an important disturbance process in our study area historically, but is now actively suppressed. Insect outbreaks continue to influence stand dynamics. Forest harvesting and grazing are the predominant anthropogenic disturbances. The area is particularly diverse for vertebrate species as several are at the northern limit of their range (Stevens 1995).

Habitat Data

We established 243 plots to encompass a wide range of structural characteristics at a range of spatial scales (Figure 1). Each plot is connected by a 150-m or 300-m transect for a total of nearly 42 km of transects. Each transect was flagged to ensure that the same route was followed on successive visits. We collected extensive vegetation data for all plots using standardised methodologies modified from several sources (summarized in Gillingham and Parker 2001). Modifications were made so that data are relevant from both the perspectives of wildlife and forest inventory, as well as for the examination of spatial variation.

We used surveyor tapes, laid out on perpendicular axes through plot centre, to measure shrub species and cover, canopy gaps, and coarse-woody debris along the intercept. For coarse woody debris, we recorded the diameter of the piece perpendicular to where it crossed the axis, the tree species, decay class, and any sign of use by wildlife. At five, 2-m radius plots located at plot centre and at 11.28 m away from plot centre on each axis, we measured the coverage for litter, coarse-woody debris, herb species, moss and lichens, shrub species, sapling species, bare ground, and rock. Within a 5.64-m radius plot around plot centre we tallied all trees and stumps and all trees >30 cm diameter at breast height (DBH) and snags within an 11.28-m radius. For these trees we recorded DBH, height, species, health, evidence of wildlife use, and whether or not the tree was standing or fallen. Each tree and sapling location within the plot was also stem mapped. We recorded general information for each plot including canopy closure around plot centre (average of 4 measurements taken on each axes), aspect, slope, canopy stratification and complexity, disturbance history (evidence of

fire, grazing, logging), any wildlife species detected while taking vegetation measures or their sign, and the elevation above sea level. We were supplied VRI data by the forest company that operates in our study area.

Species Detections

Vocal and visual detections of red squirrels were recorded during point-count surveys and encounter transects. We also recorded foraging sign, middens, and nests during encounter transects and intensive plot searches. Vocal and visual detections were recorded for pileated woodpeckers during playback surveys, point counts, and encounter transects. Foraging sign and nest cavities were recorded during encounter transects and intensive plot searches.

Point counts began no earlier than 30 min before sunrise and were ceased no later than 0900 h from late-May until early-July. The centre of each plot was used as a station for point counts to detect calling and singing birds (Wilson et al. 2000). Red squirrels frequently vocalise and are often recorded during point-count surveys. One minute after arriving at the centre of a plot, an observer facing north began a 6-min recording period. Distance and direction to detection from plot centre were estimated and recorded as either < 50 m, 50-75 m, or > 75 m. Each detection was assigned a species and activity code (song, call, visual, flyover, or drumming). Observers were rotated between plots (one visit per observer per plot) and direction of travel along the transect was changed between surveys to reduce bias associated with the observer and time of day.

We conducted playback surveys, broadcasting recordings of calls and drumming, for the seven species of woodpeckers expected to occur in the study area, including pileated woodpecker. The call playback technique attempts to solicit woodpecker responses to broadcasted recordings (Johnson et al. 1981). Woodpecker surveys began no earlier than 30 min before sunrise and ended by 1100 h from mid-May until mid-June at every other plot (minimum distance of 300 m between playback stations). When a woodpecker was detected, distance and direction to bird from plot centre were estimated and recorded as for point counts. When point counts were conducted in conjunction with playbacks, playbacks were always conducted after the passive listening period of the point count was over. Poor weather such as high winds, rain, and fog can inhibit both bird behaviour and observer ability; therefore, point-count and playback surveys were only conducted in appropriate weather conditions. Each plot was surveyed by point count and playback three times per year for a total of nine visits.

Along the established transects connecting plots, we also conducted encounter transect surveys with unlimited width. If a species was detected along the transect, a Global Positioning System (GPS) waypoint was recorded with species ID, distance and bearing from waypoint, as well as a sign code (singing, call, visual, den/nest, track, remains, browse, or feces). In addition, extensive searches were made within all plots for animal sign. Each plot was searched intensively (to a 50-m radius) for all sign of vertebrates (i.e., visuals, nests, dens, feeding, feces, remains). Detections from all methods were entered into a database and georeferenced.

Model Construction and Predictive Evaluation

We used logistic regression (Hosmer and Lemshow 1989) to examine the relationship of species occurrence to vegetative structure and composition. Candidate variables were selected from potential variables measured and supplied (i.e., VRI data). We considered a

variable a candidate if it was, or was related to, an aspect of the species' habitat requirements (Table 2). If variables existed that were measures of the same characteristic, but from different a data source (e.g., plot percent gap and VRI sum of crown closure), only one was included in a candidate model. We only considered plots in which structure was not altered over the duration of the study (e.g., harvested; $n = 228$).

To test for collinearity among independent variables, we calculated variance inflation factor scores for all variables in the model after linear regression (Neter et al. 1985). Variation inflation increases with increasing collinearity among variables and results in overestimates of variance explained. Although there is no set rule for a variance inflation factor indicating a collinearity problem we adopted a value of 5 or above, which corresponds with a tolerance score of 0.2, a recommended threshold (Menard 2001). If collinearity was indicated, we reran our model using only one of the indicated problem variables and compared outcomes. We retained the variable that resulted in the highest pseudo- R^2 and predictive accuracy.

Competing models were ranked using the Akaike's Information Criteria (AIC; Akaike 1973; Burnham and Anderson 1998). AIC model selection estimates the information loss when the probability distribution with the true model is approximated by the probability distribution associated with the model that is to be evaluated. Choosing the model with the lowest expected information loss between the true model and the approximating model is asymptotically equivalent to choosing a model that has the lowest AIC value (Burnham and Anderson 1998). We applied a correction to the AIC value to account for small sample sizes (AIC_c) and determined the Akaike weight (w), the likelihood of the model given the data (Burnham and Anderson 1998).

To further assess the subset of models we calculated the pseudo- R^2 (Nagelkerke 1991) as our measure of variation explained and ROC values to estimate predictive accuracy. We classified ROC values between 0.5-0.7 as low, 0.7-0.9 as good and >0.9 as high model prediction accuracy (Manel et al. 2001). For comparing models with different response variables (i.e., audio or visual detections versus sign detection) we consider the highest pseudo- R^2 and highest ROC value as the best model. Using the detection type of the best model we then assessed competing models (i.e., different combinations of independent variables). For those competing models, the lowest AIC_c value and highest Akaike weight defined the best model. After the final model was selected, we reviewed cases for high leverage and studentised residual values to determine if any cases were disproportionately driving the model relationship (Menard 2001). We used the program Stats (version 8.2 StataCorp 2003) for all statistical analyses and employed the DESMAT procedure (Hendrix 2001) for design matrices.

Table 2. List of candidate variables used to construct competing models for red squirrel (TAHU) and pileated woodpecker (PIWO). A structure variable was considered a candidate if it related directly to known associations between the species and its habitat.

| Data Source | Variable | Species |
|-------------------------------|---------------------------------|------------|
| Plot Measures | Main canopy height | TAHU, PIWO |
| | Percent gap | TAHU, PIWO |
| | Percent shrub cover | TAHU |
| | Percent herb cover | TAHU |
| | CWD volume(m ³) | TAHU, PIWO |
| | Percent suspended CWD | TAHU |
| | Live tree basal area | TAHU, PIWO |
| | Dead tree basal area | TAHU, PIWO |
| | DBH ≥30cm basal area | TAHU, PIWO |
| | Deciduous stems per ha | TAHU, PIWO |
| | Frequency Douglas Fir | TAHU, PIWO |
| | Frequency Spruce | TAHU, PIWO |
| | Coniferous stems per ha | TAHU, PIWO |
| Vegetation Resource Inventory | Shrub crown closure | TAHU |
| | Adjusted live basal area | TAHU, PIWO |
| | Herb cover percent | TAHU |
| | Adjusted canopy closure | TAHU, PIWO |
| | Douglas Fir cover | TAHU, PIWO |
| | Aspen cover | TAHU, PIWO |
| | Spruce cover | TAHU, PIWO |
| | Adjusted leading species height | TAHU, PIWO |
| Structure class (categorical) | TAHU, PIWO | |

Results

Red squirrel

We recorded audio and visual detections of red squirrel at 178 plots and detected sign at 205 plots (219 plots in total). Audio and visual detections were primarily made during point count surveys. Sign detections included foraging sign, middens, remains, and nests. The best model was constructed with sign as the response variable (Table 3). This model explained approximately 16% more variation and the prediction accuracy improved by 8%, but both models had good predictive accuracy (sign ROC = 0.88, audio and visual ROC = 0.80; Table 3). Several of the same variables were included in both models, specifically, structure class, percent gap, and percent spruce. Comparing the results of competing models for only sign as the response variable, the best model was constructed with a combination of plot-level data and VRI data (AIC_c = 181.55; Table 4). The multi-source model was highly favoured as the best model with an Akaike weight of 0.93, or a 93% likelihood of being the best model; however, prediction accuracy did not improve dramatically (VRI ROC = 0.86, multi-source ROC = 0.88; Table 4).

Table 3. Model results for red squirrel comparing audio and visual detections to sign detections.

| Response variable | Independent variables | Pseudo-R ² | ROC |
|-------------------|-----------------------|-----------------------|------|
| Audio Visual | Structure Class | 0.20 | 0.80 |
| Sign | Percent Gap | 0.36 | 0.88 |
| | Frequency Spruce | | |
| | Structure Class | | |
| | Percent Gap | | |
| | Frequency Spruce | | |
| | Dead trees basal area | | |
| | Shrub crown closure | | |

Table 4. Competing models for red squirrel sign detections using different sources for independent data variables, plot measures and Vegetative Resource Inventory database.

| Data Source | Independent variables | AIC _c | ΔAIC | w | ROC |
|--------------|----------------------------------------------------------------------------------------------------|------------------|-------|-------|------|
| Multi-Source | Structure class Percent Gap Spruce stems / ha Dead tree basal area Shrub crown closure | 181.55 | 0 | 0.93 | 0.88 |
| VRI | Adjusted canopy closure Structure class Spruce cover | 187.31 | 5.28 | 0.067 | 0.86 |
| Plot | Percent Gap Spruce stem / ha | 212.36 | 29.28 | <0.01 | 0.76 |

Pileated woodpecker

We recorded audio and visual detections for the pileated woodpecker at 63 plots and sign at 33 plots (85 plots in total). Sign detections included foraging and nest cavities. The best model was constructed using sign as the response variable (Table 5). There was little overlap among explanatory variables included in the two models. Sign detection as the response variable dramatically improved model performance. Variation explained improved by 14% and predictive accuracy improved from poor (ROC = 0.66) to good (ROC = 0.82; Table 5). Comparing the results of competing models using only sign as the response variable, the best model was constructed with a combination of plot and VRI data (AIC_c = 179.55; Table 6). The likelihood of the multi-source model being the best approximating model was 66% and predictive accuracy improved by 10%, but remained good (Table 6).

Table 5. Model results for pileated woodpecker comparing audio and visual detections to sign detections.

| Response variable | Independent variables | Pseudo-R ² | ROC |
|-------------------|-------------------------------------------------------------------------------------------------------|-----------------------|------|
| Audio Visual | CWD volume Percent gap Douglas fir stems /ha DBH ≥30cm basal area | 0.06 | 0.66 |
| Sign | Percent Gap Structure class Douglas fir cover Main canopy height Adjusted live basal area | 0.21 | 0.82 |

Table 6. Competing models for pileated woodpecker sign detections using different sources for independent data variables - plot measures and Vegetative Resource Inventory database.

| Data Source | Independent variables | AIC _c | ΔAIC | w | ROC |
|--------------|-------------------------------------------------------------------------------------------------------|------------------|-------|------|------|
| Multi-Source | Percent Gap Structure class Douglas fir cover Main canopy height Adjusted live basal area | 179.55 | 0 | 0.66 | 0.82 |
| VRI | Adjusted canopy closure Douglas fir cover Adjusted live basal area | 181.16 | 1.611 | 0.30 | 0.72 |
| Plot | Main canopy height | 185.21 | 5.66 | 0.04 | 0.65 |

Discussion

For both the red squirrel and pileated woodpecker, logistic models using sign detections outperformed those with audio and visual detections. Measurement uncertainty from several sources may explain this result. Firstly, it is sometimes difficult to associate vocalisations to specific locations due to distance sound travels and measurement error assigning the detection to a spatial location. Secondly, vocalisations may be made while the individual is in transit between parts of its home range, as is the case for the pileated woodpecker. Finally, the observer must be at the location at a moment in time when the individual is near the same location. For species with large home ranges, the probability of an observer being in the same location, in space and time, is lower than that for a species with a small home range. It is possible, therefore, that we may have not detected presence. In contrast, given our survey methods for sign, it was highly unlikely that we would not detect presence if it was there.

Sign detections are often associated with a specific component of a species' habitat requirements (e.g., foraging or nesting sites). Logistic models with sign data as the response variable likely had higher predictive efficiency because foraging and nesting substrates are often selected for a smaller scale, within the context of a home range. For example, red

squirrels have home ranges of 1-3 ha (Obbard 1987), easily within a single stand of similar habitat characteristics. Audio and visual detections are likely recorded within the same stand as foraging and other life history activities take place; therefore, it is not surprising that model results are similar for the different detection types for red squirrel. The additional variables included in the sign model for red squirrel, shrub cover and dead wood basal area, may be more related to the associated effects of middens on ground vegetation and the availability of cones. Potential temporal uncertainty associated with changes to vegetative structure and composition from disturbance (e.g., alteration due to harvesting practices) around sign detections may prove a greater issue in long-term monitoring studies.

The pileated woodpecker has a large home range; pairs in the Pacific Northwest use between 300 and 600 ha, while unpaired birds used up to 1400 ha (Bull and Holthausen 1993). Pileated woodpeckers select large snags and logs to forage on and large diameter trees for nesting (Bull and Holthausen 1993; Carey et al. 1991). These structural components may be localised or clustered within a highly variable home range, resulting in better predictive accuracy of models for sign. Further, because we employed a survey methodology that was designed to solicit responses from individuals if they were in the area (i.e., to minimise false absence detections) we may have inadvertently attracted an individual from its original location and the associated characteristics of that original location. It is unclear if false absence or spatial uncertainty related to an individual's movement is more detrimental to the predictive efficiency of logistic species-habitat models and their use in a biodiversity monitoring program.

Overall, differences in logistic-model results, and the potential difference in a monitoring program designed around them, emphasize the importance of assessing sources of potential model error and the predictive efficiency. For a species that uses landscapes at relatively small scales (e.g., red squirrel), sign and visual or audio detections are likely to be within the same area and not have a large effect on a monitoring program. In contrast, there was little overlap among variables included in the different statistical models for pileated woodpecker and the predictive efficiency varied greatly. It may be more appropriate to use complementary models over a general model that has weak predictive accuracy. Ultimately, the decision of which is the most appropriate approach will depend on what poses the highest risk to the species and the goals and objectives the statistical model is intended to help achieve.

In this paper we examined the effects of two kinds of potential model error: detection type and availability of structure data. We showed that: 1) different types of detection can potentially affect model results and assessment of accuracy; and, 2) that adding a small amount of locally measured data can improve predictive efficiency dramatically. Measuring these data may have relatively low costs for increased confidence in the model. We conducted our analyses using our original data set; independent data is the preferred approach to address all aspects of validation. It is, therefore, likely that we have reported optimistic predictive accuracy (Chatfield 1995). Additionally, although the area under the ROC curves is a robust measure of prediction accuracy, its use in the biological sciences is relatively recent. Some caution should be taken when using the method outside of its original development (Kraemer 1988) although the approach is highly recommended for presence and absence data (Ottaviani et al. 2004).

Overall, assessing prediction accuracy is only one component of model validation. Validation needs to be an iterative process so that confidence is maintained in the model's usefulness through continued monitoring (Johnson 2001). Other factors, outside of those used in

predictive-habitat models, may ultimately have greater impact on the presence of species (e.g., invasive species, climate change). A better understanding of the relations among species distribution and forest characteristics, both in the short- and long-term, will provide knowledge to mitigate the loss of species from forest environments. Ultimately, assessing model performance will allow for informed trade-offs and lead to improved effectiveness of biodiversity monitoring and should be included in any program that is using models as biodiversity monitoring tools. In conclusion, every biodiversity monitoring program that uses species-habitat models should make every effort to ensure that models are valid for their intended use. Clear articulation of model objectives and a thorough consideration to appropriate types of data collection, a standard of acceptable error, and independent data for evaluation of error will improve the effectiveness of biodiversity monitoring programs.

Acknowledgements

This work was partially supported by Forest Renewal BC, the Forest Science Program and Riverside Forest Products Ltd. Jennifer Psyllakis received support from the National Sciences and Engineering Research Council of Canada and the University of Northern British Columbia. We are grateful for assistance in the field from K. Bachmann, S. Barker, D. Baxter, D. Gummeson, M. Hagedorn, E. Hebert, C. Hotson, H. Jack, E. Jones, H. Knight, L. Miller, E. Schuet, and N. Thomas.

References

- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. *In* B. N. Petrov and F. Caski (editors). *Proceeding of the Second International Symposium on Information Theory*. Budapest, Akademiai Kiado, pp. 267-281.
- Boyce, M.S., P.R. Vernier, S.E. Nielsen, and F.K.A. Schmiegelow. 2002. Evaluating resource selection functions. *Ecol. Modell.* 157: 281-300.
- Burnham, K.P. and D.R. Anderson. 1998. *Model Selection and Inference. A Practical Information-Theoretic Approach*. Springer-Verlag, NY.
- Bull, E. L. and R. S. Holthausen. 1993. Habitat use and management of pileated woodpeckers in northeastern Oregon. *J. Wildl. Manage.* 57: 335-345.
- Carey, A.B., M. M. Hardt, S. P. Horton, and B. L. Biswell. 1991. Spring Bird Communities in the Oregon Coast Range. *In* L.F. Ruggiero, K.B. Aubry, A.B. Carey, and M.H. Huff (editors). *Wildlife and vegetation of unmanaged Douglas-fir forests*. U.S. Department of Agriculture, Gen. Tech. Rep. PNW-GTR-285., Portland, OR, pp 123-140.
- Chatfield, C. 1995. Model uncertainty, data mining and statistical inference. *J. R. Stat. Soc.* 158: 419-466.
- Deleo, J.M. 1993. Receiver operating characteristic laboratory (ROCLAB): software for developing decision strategies that account for uncertainty. *In* *Proceedings of the*

Second International Symposium on Uncertainty Modelling and Analysis, IEEE Computer Society Press, College Park, MD, pp. 318–325.

- Elith, J., and M. Burgman. 2002. Predictions and their validation: rare plants in Central Highlands, Victoria, Australia. *In* J.M. Scott, P.J. Heglund, and M.L. Morrison (editors). *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, DC, pp. 303-313.
- Fielding, A.H. 2002. What are appropriate characteristics of an accuracy measure? *In* Scott, J.M., P.J. Heglund, and M.L. Morrison (editors). *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, DC, pp. 271-280.
- Fielding, A.H. and J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence-absence models. *Environ. Conserv.* 24: 38-49.
- Fleishman, E., R. Mac Nally, and J.P. Fay. 2002. Validation tests of predictive models of butterfly occurrence based on environmental variables. *Conserv. Biol.* 17: 806-817.
- Gillingham, M.P., and K.L. Parker. 2001. Summary report for the first year of the Project on Lifeform Classification. Contribution agreement report to Lignum, Ltd. 10 Dec. 21 pp. + 337 pp. appendices. Available at <http://web.unbc.ca/biodiversity/>
- Guisan, A. and N.E. Zimmerman. 2000. Predictive habitat distribution models in ecology. *Ecol. Modell.* 135: 147-186.
- Harper, J.L. and D.L. Hawksworth. 1995. Preface. *In* Biodiversity Measurement and Estimation. D.L. Hawksworth (editor). The Royal Society. Chapman and Hall. London., pp. 5-12.
- Hendrickx, J. 2001. Contrasts for categorical variables: update. *Stata Technical Bulletin* 59: 2-5.
- Hirzel, A.H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83: 2027-2036.
- Hosmer, D.W., and S. Lemeshow. 1989. *Applied Logistic Regression*. J. Wiley and Sons, Toronto.
- Hunter, M.L. Jr. 1990. *Wildlife, Forests, and Forestry: Principles of Managing Forests for Biologic Diversity*. Prentice-Hall. Englewood Cliffs, NJ.

- Johnson, R.R., B.T. Brown, L.T. Haight and J.M. Simpson. 1981. Playback recording as a special avian censusing technique. *In* C.J. Ralph and J.M. Scott (editors). Estimating numbers of terrestrial birds. Studies in Avian Biology 6. Allen Press, Lawrence, KS, pp. 68-75.
- Johnson, D.H. 2001. Validating and evaluating models. *In* T.M. Shenk and T.B. Franklin (editors). Modeling in Natural Resource Management: development, interpretation, and application. Island Press, Washington, DC, pp. 105-119.
- Kraemer, H.C. 1988. Assessment of 2 x 2 associations: Generalisation of signal detection methodology. *Am. Stat.* 42: 37-49.
- MacKenzie, D.I., J.D. Nichols, G.B. Lachman, S. Droege, J.A. Royle, and C.A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83: 2248-2255.
- Mac Nally, R. 2002. Improving inference in ecological research: issues of scope, scale, and model validation. *Comments Theor. Biol.* 7: 237-256.
- Manel, S., J. Dias, and S.J. Ormerod. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with Himalayan river bird. *Ecol. Modell.* 120: 337-347.
- Manel, S., H.C. Williams, and S.J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.* 38: 921-931.
- Meidinger, D., and J. Pojar (editors). 1991. Ecosystems of British Columbia. B.C. Ministry of Forests, Spec. Rep. Ser. 6. Victoria, BC.
- Menard, S. W. 2001. Applied Logistic Regression Analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-106. Thousand Oaks, CA.
- Minister of Supply and Services Canada. 1995. Canadian Biodiversity Strategy: Canada's Response to the Convention on Biological Diversity. Catalogue No. En21-134/1995E.
- Nagelkerke, N.J.D. 1991. A note on general definition of the coefficient of determination. *Biometrika.* 78: 691-692.
- Neter, J., W. Wasserman, and M.H. Kutner. 1985. Applied linear statistical models: Regression, analysis of variance, and experimental designs. Irwin Pub., Homewood, IL.
- Obbard, M. E. 1987. Red squirrel. *In* M. Novak, J.A. Baker, M.E. Obbard, and A.M. Mallock (editors). Wild furbearer management and conservation in North America. Ontario Trappers Association, Ontario Ministry of Natural Resources, Queen's Printer, Toronto, ON, pp. 264-281.
- Ottaviani, D., G.J. Lasinio, and L. Boitani. 2004. Two statistical methods to validate habitat suitability models using presence-only data. *Ecol. Modell.* 179: 417-443.

- Pearce, J. and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Modell.* 133: 225-245.
- Province of British Columbia. 1999. Landscape Unit Planning Guide. Ministry of Forests, Forest Practices Branch Operations Division and Ministry of Environment, Lands and Parks Resource Stewardship Branch. Victoria, BC.
- Province of British Columbia. 2002. Vegetation Resource Inventory (VRI) Update. Ministry of Sustainable Resources Management. Victoria, BC.
- Rykiel, E.J., Jr. 1996. Testing ecological models: the meaning of validation. *Ecol. Modell.* 90: 229-244.
- Scott, J.M., P.J. Heglund, and M.L. Morrison (editors). 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale.* Island Press, Washington, DC.
- StataCorp. 2003. *Stata Statistical Software: Release 8.* College Station, TX: StataCorp LP.
- Stevens, V. 1995. *Wildlife Diversity in British Columbia: Distribution and Habitat Use of Amphibians, Reptiles, Birds, and Mammals in Biogeoclimatic Zones.* Res. Br., B.C. Min. For., Wildl. Br., B.C. Min. Environ, Lands and Parks., Victoria, BC. Work Pap. 04/1995.
- United Nations Environment Programme. 1992. *Convention on Biological Diversity.* United Nations.
- Verbyla, D.L. and J.A. Litaitis. 1989. Resampling methods for evaluation of classification accuracy of wildlife habitat models. *Environ. Manage.* 13: 783-787.
- Wilson, R.R., D.J. Twedt, and A.B. Elliott. 2000. Comparison of line transects and point counts for monitoring spring migration in forested wetlands. *J. Field. Ornith.* 71: 345-355.