SUSTAINABLE **FOREST** MANAGEMENT NETWORK

RÉSEAU DE GESTION DURABLE DES **FORÊTS**

PROJECT REPORTS
2003/2004

Spatial and temporal validation of habitat models in the boreal mixedwood

P.R. Vernier, F.L. Bunnell, F.K.A. Schmiegelow, S.G. Cumming, and C.J. Walters

May 2003

Published: 20 August 2003

Related SFM Network Project:
bunnellfland6
Landscape issues in sustainable forest management: statistical methods and tools for projecting consequence of management actions

A NETWORK OF CENTRES OF EXCELLENCE
UN RÉSEAU DE CENTRES D'EXCELLENCE

Spatial and Temporal Validation of Habitat Models in the Boreal Mixedwood

Landscape Issues in Sustainable Forest Management: Statistical Methods and Tools for Projecting Consequences of Management Actions

P.R. Vernier[1], F.L. Bunnell[1], F.K.A. Schmiegelow[2], S.G. Cumming[3], and C.J. Walters[4]

[1]Centre for Applied Conservation Research, University of British Columbia
[2]Department of Renewable Resources, University of Alberta
[3]Boreal Ecosystem Research Ltd., Edmonton, Alberta
[4]Fisheries Centre, University of British Columbia

May 2003

ABSTRACT

Sustainable forest management requires knowledge about potential economic and ecological outcomes of policy decisions and management actions. Often this requires the use and evaluation of models that capture and project the relationships between indicators and management policies and practices. This project, which is one of several interrelated projects undertaken by the Boreal Ecology and Synthesis Team (BEEST), had two main objectives related to model development and evaluation. The first objective was to evaluate the accuracy and generality of songbird-habitat models and thus to determine their usefulness as tools for managing forest-dwelling birds and evaluating alternative forest management scenarios. The second objective was to continue to develop and enhance FEENIX, a landscape simulation model, and to apply it to evaluate management scenarios and policy decisions.

We assessed the predictive capability of songbird-habitat models developed in Calling Lake, Alberta using (1) data from the same site but different years (temporal validation), (2) independent data collected at four sites within 125 km of Calling Lake (external validation), and (3) independent data collected in northeastern British Columbia (external validation). All sites were located within the boreal plains ecosection. The models were derived using logistic regression analysis and their accuracy assessed using the area under a receiver operating characteristic curve (AUC; see Model Validation). In the temporal validation component, model coefficients and reliability fluctuated across years but performance was generally satisfactory. Increasing the number of years used to develop the models improved only the predictive ability of the Red-breasted Nuthatch model. Generally, regardless of the number of years of data used, the models were reliable. In Alberta, when we combined all external validation sites, our results indicated reasonable discrimination between occupied and unoccupied sites for 8 species, while for the other 8 species the models showed poor discrimination ability. When the models were tested against each validation site individually, performance depended in part on broad differences in landscape structure and disturbance history. In comparison, when the habitat models were developed and validated using only data from the four validation landscape (internal validation), they generally performed well. Using the BC validation data, none of the four species models tested proved to be adequate – although, using the same set of variables it was possible to develop reliable models using local data. Thus, our analyses indicate that although model coefficients vary geographically, the same set of variables is consistently useful in developing predictive models of site-occupancy within landscapes – especially when developed using local data.

During the past three years, we continued to develop and enhance FEENIX, and apply it to evaluate management scenarios and policy alternatives. Enhancements to FEENIX include improved performance, the addition of several sub-models (i.e. salvage logging, landscape metrics, networks, strata-based harvest scheduling), the ability to import and export common GIS data, and the facility to perform Monte Carlo analyses. A prototype, “stripped-down” version of FEENIX also has been developed to facilitate its use by non-programmers. Recent applications of FEENIX for management purposes are described in a related BEEST project.

ACKNOWLEDGEMENTS

We thank T. Morcos and C. McCallum for database management and GIS assistance. We also thank the numerous field assistants who collected the point count data as part of the Calling Lake Fragmentation Experiment. S. Hannon collaborated on the analysis and provided data for the Alberta validation component of this project. M. Boyce and S. Nielsen collaborated on the temporal validation component. R. Lessard contributed to the development of FEENIX for the past three years while conducting research for his SFM-funded PhD. The BC validation data was provided by Canadian Forest Products Ltd. (thru a contract to Manning, Cooper and Associates). This research was supported by the Sustainable Forest Management Network and by Alberta Pacific Forest Industries Inc. Participating SFM Network Partners/Affiliates: Alberta-Pacific Forest Industries Inc., Government of Alberta, Weyerhaeuser Company.

INTRODUCTION

Sustainable forest management requires knowledge about the potential economic and ecological outcomes of policy decisions and management actions. For example, managers need to understand the consequences of alternative management scenarios and policies on ecological indicators such as forest songbirds. One of the most effective ways to accomplish this is to capture relationships between songbirds and habitat attributes within models, and to integrate these models in tools that are able to simulate management activities and landscape dynamics over large spatial and temporal scales. Such habitat models can be used in two broad ways: (1) by themselves to identify information gaps, assess current landscape conditions, and test specific species-habitat hypotheses and (2) within a landscape simulator to facilitate integrated resource planning, scenario evaluation, and policy analysis. Both uses of habitat models can help managers make better decisions. Over the past several years, the Boreal Ecology and Economics Team (BEEST) has developed two modeling platforms designed to simulate forest harvesting, wildfire, stand dynamics, and wildlife habitat and population dynamics in boreal forests. Each platform comprises a number of sub-models, including habitat suitability models, that are parameterised and/or initialised from the same or similar data sets, but which differ in the scale at which processes are represented, the size of the regions that may be modeled, and in the sorts of questions they are suited to answer. FEENIX, which is described in more detail in this report, operates at a relatively fine resolution (3 ha), and can deal with patch-level questions such as fire ignition and spread, dispersal of seeds or individual vertebrates, tactical harvest scheduling, and cut-block layout. TARDIS, in contrast, works at very broad scales (~ 100 km^2) and is intended primarily as a strategic level, policy analysis tool.

In collaboration with the BEEST, one of our goals over the past few years was to contribute to the development and validation of songbird-habitat models for use on their own and within FEENIX for scenario evaluation and policy analysis. A persistent criticism of habitat models and models used in a management context, is that they are rarely validated. Our ability to generalize predictions from our existing songbird-habitat to other locations in the boreal forest may be restricted by geographic limits in the ranges of bird species, or by undetected variation in links between forest habitat elements and particular bird species or communities within the ecosystem. We explored these issues by attempting to validate our models with data from other locations in the Alberta boreal mixedwood forest and neighbouring British Columbia (BC). By testing the models it may possible to identify sources of uncertainty and guide data collection to reducing the causes of those uncertainties.

We had two main objectives. Our first objective was to evaluate the accuracy and generalizability of songbird-habitat models and thus to determine their usefulness as tools for managing forest-dwelling birds and evaluating alternative forest management scenarios. To do this, we used existing bird survey data that have been collected at independent locations in Alberta and BC in conjunction with other research and management projects. Our focus was on evaluating habitat models for use in areas within the boreal mixedwood forest. Specifically, we assessed the predictive ability of the Calling Lake habitat models using:

1. Multi-year data from the same study area to (1) assess the temporal variability in songbird-habitat relationships (direction, strength, and significance of estimated coefficients), (b) evaluate the temporal variability in the predictive performance of models, and (c) determine if the predictive performance of the models increased with the number of years of data used to fit the model. [Temporal Validation]
2. Independent data collected at four separate locations in Alberta within the same ecological region (boreal plains ecosection). [External Validation]
3. Independent data collected at one location in northeastern BC at the western edge of the same broad ecological region (boreal plains ecosection). [External Validation]

Our second objective was to continue to develop FEENIX and to apply it to evaluate management scenarios. Consequently, in this report, we describe various enhancements that have been made FEENIX over the past 3 years. The application of FEENIX in various management scenarios and policy analyses in Alberta and BC is provided in a companion BEEST project.

HABITAT MODEL EVALUATION

Habitat models are often used to make predictions beyond the sample from which the models were developed (Scott et al. 2002). For example, a model may be used to assess the suitability of a forest management unit for a species of concern or to evaluate and rank alternative future management scenarios. Both examples illustrate the use of habitat models outside of their range of development – both in space and time – raising concerns about the use of models without proper, or external, validation. Validation, however, is rarely done due to limitations in time, funding, and data (Morrison et al. 1999, Guisan and Zimmermann 2000). Recently, in an attempt to address this situation, several papers have reviewed and illustrated methods for evaluating species distribution models and in particular logistic regression models such as the ones described here (Fielding and Bell 1997, Pearce and Ferrier 2000).

A key aspect of prediction, is the consideration of whether a model derived from an analysis of the original data set is transportable to similar forest landscapes in other geographic locations. This concept is sometimes referred to as generalizability or validity, and a model that is found to pass such a test is said to have been validated. Broadly, validation can be carried out both internally and externally (Harrell 2001). Internal validation is used to test the accuracy of a model using the same sample that was used to develop the model; it is restricted to a single geographic site. Methods include data splitting and computer-intensive methods such as ‘leave-one out’ cross-validation, and bootstrapping. External validation addresses the issue of the generalizability of the models by using independent data (e.g., another place) to test a model. A third form of validation, referred to as temporal or prospective validation, provides an intermediate level of assessment whereby models developed in one study area are tested against past or future data collected in the same area.

No matter which validation approach is used (internal, external, or temporal), there are two related aspects of predictive accuracy that can be assessed when evaluating a habitat model: calibration and discrimination (Pearce and Ferrier 2000, Harrell 2001). Calibration describes the agreement between observations and predicted values (e.g., goodness-of-fit) and therefore describes the reliability with which a model predicts the probability of a site being occupied by a species of interest. A reliable model (i.e., well calibrated) should be able to correctly predict the actual proportion of sites occupied by the species of interest. Although revealing, this aspect of model accuracy is not always directly useful to forest managers. Discrimination, on the other hand, is the model's ability to reliably classify locations (e.g., patches or landscape units) into two or more groups (e.g., presence/absence or different habitat suitability classes). Such classifications can be used to provide feedback to influence management activities such as the choice of treatment type to use in a particular location or to evaluate and rank the suitability of different landscape units or management scenarios. Consequently, our focus in this report is on evaluating the discrimination capability of our songbird-habitat models.

Study Areas

All of the songbird-habitat models were initially developed and tested using data collected adjacent to Calling Lake, in north-central Alberta (55° N, 113° W). Subsequently, the models were tested at five independent sites, four near Calling Lake and one in BC (Figure 1). All sites are within the Boreal Plains ecosection (Ecological Stratification Working Group 1996). Three of the validation sites were located between 85 and 125 km away from the development site (Calling Lake). The fourth validation site was located just north of Calling Lake. The fifth site was located near Chetwynd in northeastern BC. Although there are many similarities among the landscapes with respect to merchantable tree species, there are also some important differences relating to the disturbance history of the sites. The Calling Lake site has been logged extensively during the early part of the 1990s but has not experienced any major recent burns. In contrast, the Goodwin site is characterized by a very large burn and few recent clearcuts. The Reference site, as its name implies, has been relatively unlogged and largely undisturbed by recent fires. The North Calling Lake and Owl river sites are more similar to Calling lake in that they have both been logged recently. At the Alberta sites, trembling aspen (*Populus tremuloides*), balsam poplar (*Populus balsamifera*), and white spruce (*Picea glauca*) were the most abundant upland tree species, often occurring together in old, mixed stands, whereas black spruce (*Picea. mariana*) characterized hydric sites. The BC site is located within Block 4 of TFL 48 near Chetwynd and consists of few recent natural disturbance patches but many widely distributed cutblocks, especially in the lower elevations. The leading merchantable tree species were lodgepole pine (*Pinus contorta*), subalpine fir (*Abies lasiocarpa*), white spruce, trembling aspen, and spruce hybrid (*Picea* cross).

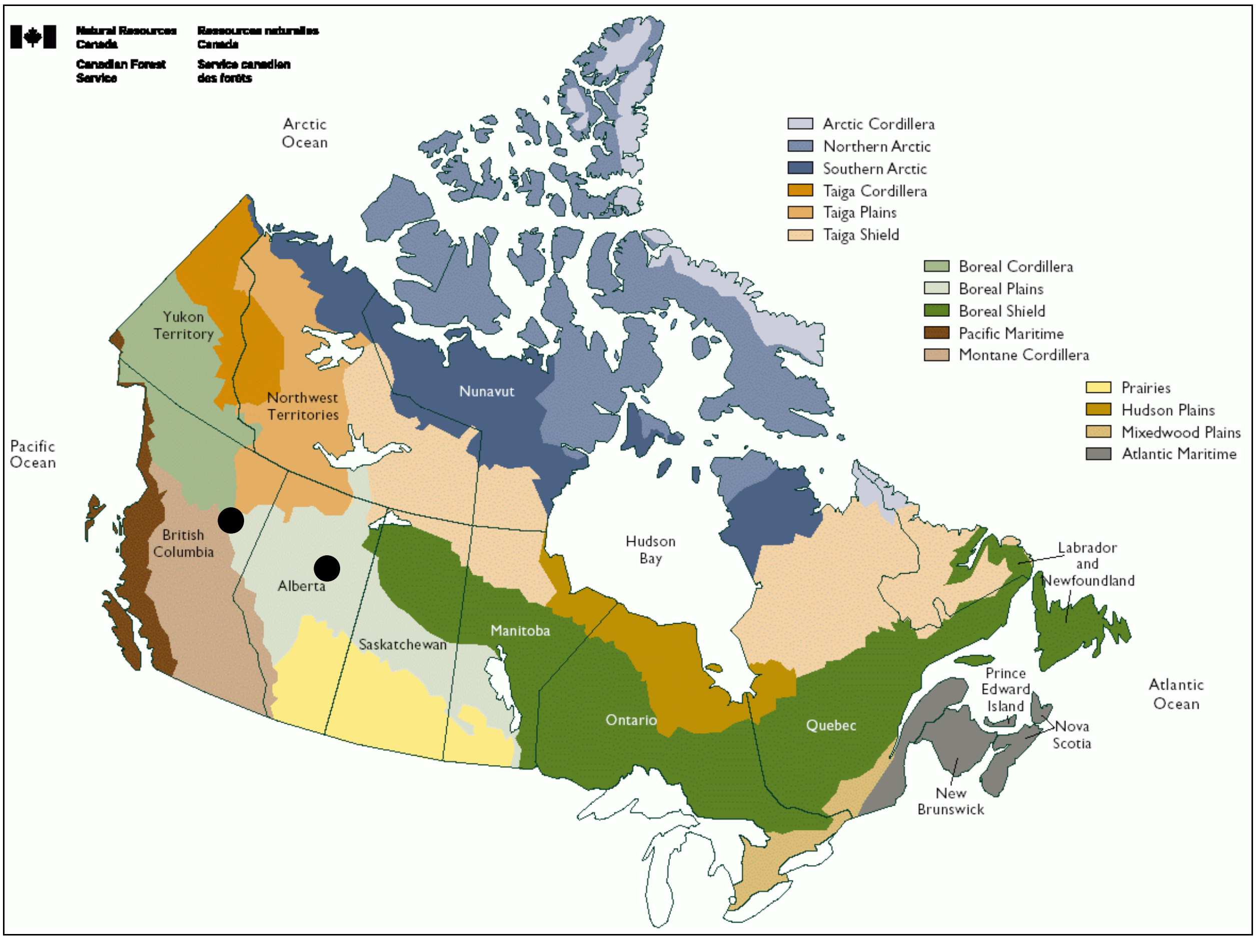
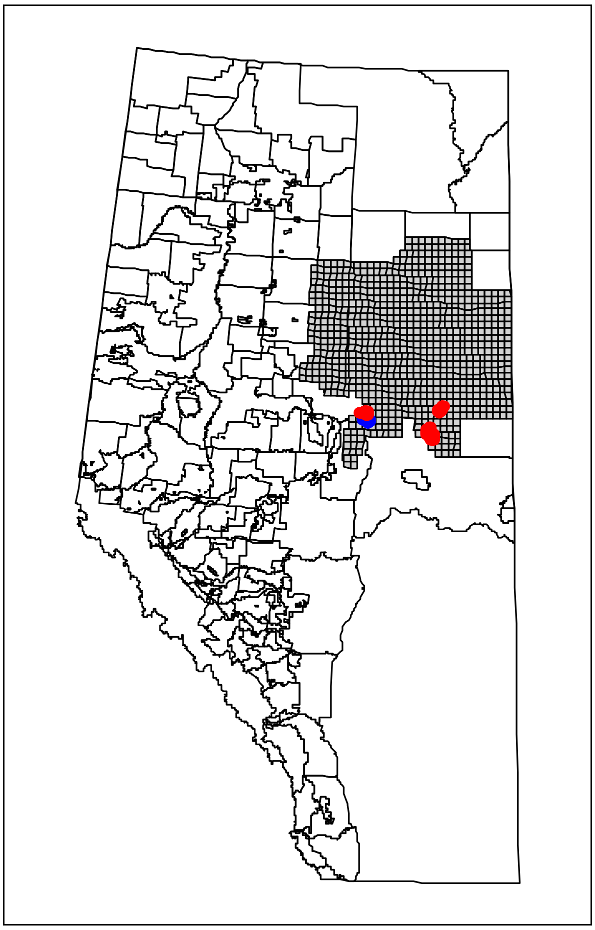
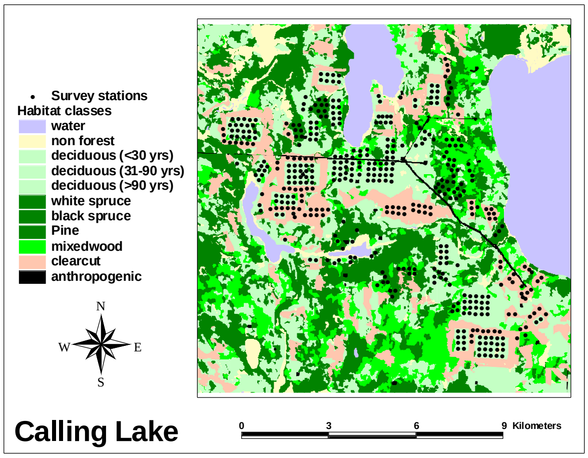


Figure 1. General locations of the Alberta and BC bird survey sites (black circles). The map of the Terrestrial Ecozones of Canada is from the 2002 State of Canada's Forests website located at http://www.nrcan.gc.ca/cfs-scf/national/what-quoi/sof/latest_e.html.

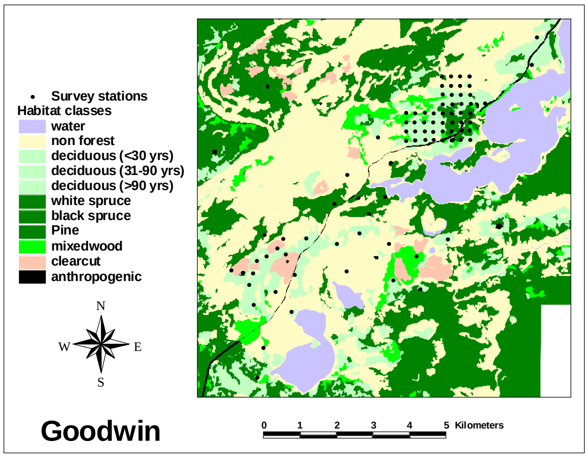
Bird Survey Data

Calling Lake site. We developed habitat models for selected songbird species using bird abundance data collected by point-count surveys conducted between 1993-2001 as part of the Calling Lake Fragmentation Experiment study in north-central Alberta (Schmiegelow et al. 1997). A total of 436 permanent sampling stations were located within 65 sites, which were defined as contiguous areas of similar forest type and age. Site types included areas clearcut in 1993 as part of the experimental design, young and old deciduous forests, mature coniferous forests, and mixedwood forests. There was at least 200 m between each sampling station. In each year that a station was sampled, point counts were conducted four times during the breeding season, from the third week in May through early July. The point count stations were geo-referenced and linked to the Calling Lake GIS database. We converted all bird abundance data to presence/absence data prior to analysis. Schmiegelow et al. (1997) provide more details on the survey design.

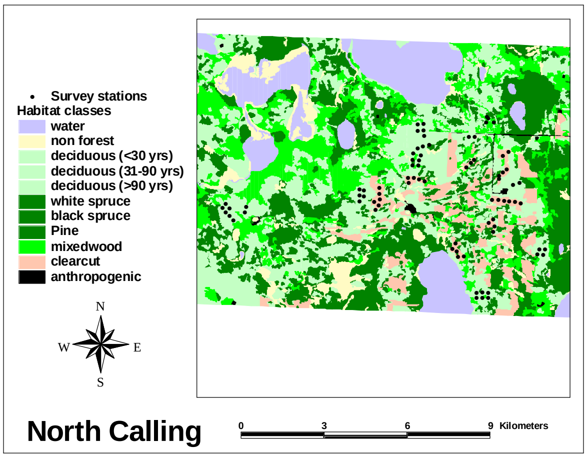




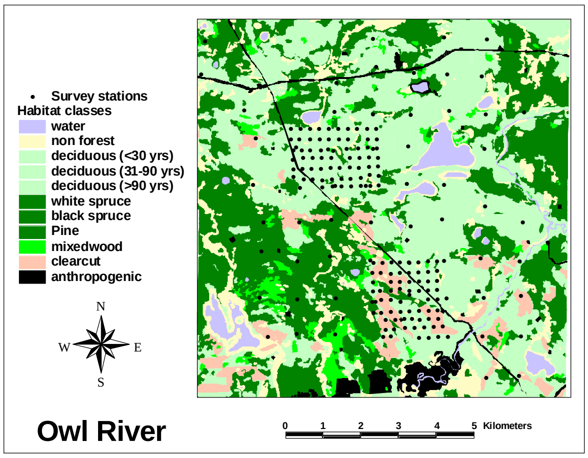
Survey stations
Habitat classes
water
non forest
deciduous (<30 yrs)
deciduous (31-90 yrs)
deciduous (>90 yrs)
white spruce
black spruce
Pine
mixedwood
clearcut
anthropogenic
N
W
E
S
Calling Lake
0
3
6
9 Kilometers



Survey stations
Habitat classes
water
non forest
deciduous (<30 yrs)
deciduous (31-90 yrs)
deciduous (>90 yrs)
white spruce
black spruce
Pine
mixedwood
clearcut
anthropogenic
N
W
E
S
Goodwin
0
1
2
3
4
5 Kilometers



Survey stations
Habitat classes
water
non forest
deciduous (<30 yrs)
deciduous (31-90 yrs)
deciduous (>90 yrs)
white spruce
black spruce
Pine
mixedwood
clearcut
anthropogenic
N
W
E
S
North Calling
0
3
6
9 Kilometers



Survey stations
Habitat classes
water
non forest
deciduous (<30 yrs)
deciduous (31-90 yrs)
deciduous (>90 yrs)
white spruce
black spruce
Pine
mixedwood
clearcut
anthropogenic
N
W
E
S
Owl River
0
1
2
3
4
5 Kilometers

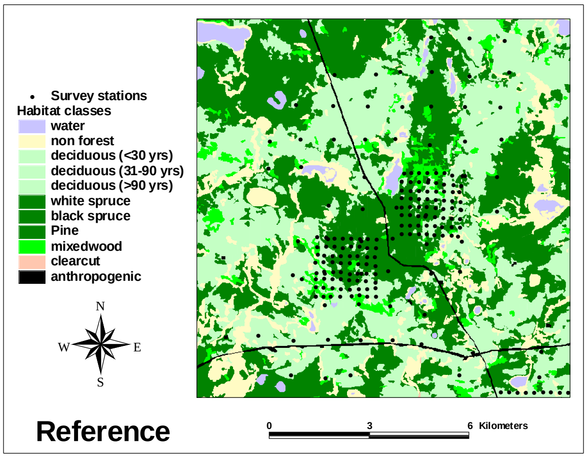


Figure 2. Top left map indicates the location of development (blue stations) and validation (red stations) sites in north-central Alberta. All Alberta sites are located within the ALPAC FMU (grid). The 5 other maps show the distribution of survey stations and habitat classes within each of the study sites. The BC site used a modified habitat classification system and is not shown here.

Temporal validation. We developed models using presence-absence data for 5 boreal forest songbird species: Black-throated Green Warbler, Red-breasted Nuthatch, White-throated Sparrow, Yellow-rumped Warbler, and Yellow Warbler (Table 1). This is the same suite of species for which we recently developed abundance models (Vernier et al., 2002). The songbird models in this analysis component were developed and tested using data from 1993-99, which, at the time were the most current available to us (Boyce et al. 2002).

Alberta validation sites. We validated our habitat models using bird abundance data collected by point-count surveys conducted in 1996 and 1997 in four separate locations. The sampling protocol was similar to that described for the Calling Lake bird survey. As with the model development data, we converted all bird abundance data to presence/absence data prior to analysis. Details of the sampling protocols and study area can be found in Hannon (1999). In our analyses of model performance, we only considered songbird species that occurred in at least 10% of the stations but no more than 90% (Table 1).

BC validation site. We used bird abundance data collected by point-count surveys conducted in 2000 and 2001 as part of a Northern Goshawk inventory of Block 4, TFL 48 near Chetwynd in northeastern BC (Manning, Cooper and Associates 2001). We only included survey stations that were located in the Boreal White and Black Spruce (BWBS) and Sub-boreal Spruce (SBS) biogeoclimatic zones. As with the model development data, we converted all bird abundance data to presence/absence data prior to analysis. In contrast to the Alberta bird surveys, the songbird surveys in BC were biased toward older forest types where Goshawks were more likely to occur. Manning, Cooper and Associates (2001) provide more details on the bird survey work. We evaluated habitat models for Black-throated Green Warbler, Red-eyed Vireo, Swainson's Thrush, and Yellow Warbler (Table 1).

Table 1. Code, common name, and Latin name for species modeled in this project. AB = Alberta external validation; BC = BC external validation; Temporal = temporal validation.

Code	***Common name***	***Latin name***	***Study area***
AMRE	American Redstart	*Setophaga ruticilla*	AB
BGNW	Black-throated Green Warbler	*Dendroica virens*	AB, BC, Temporal
CHSP	Chipping Sparrow	*Spizella passerine*	AB
COWA	Connecticut Warbler	*Oporornis agilis*	AB
LEFL	Least Flycatcher	*Empidonax minimus*	AB
MOWA	Mourning Warbler	*Oporornis philadelphia*	AB
OVEN	Ovenbird	*Seiurus aurocapillus*	AB
RBGR	Rose-breasted Grosbeak	*Pheucticus ludovicianus*	AB
RBNU	Red-breasted Nuthatch	*Sitta canadensis*	AB, BC, Temporal
REVI	Red-eyed Vireo	*Vireo olivaceus*	AB, BC
SWTH	Swainson's Thrush	*Catharus ustulatus*	AB, BC
TEWA	Tennessee Warbler	*Vermivora peregrina*	AB
WIWR	Winter Wren	*Troglodytes troglodytes*	AB
WTSP	White-throated Sparrow	*Troglodytes troglodytes*	AB, Temporal
YRWA	Yellow-rumped Warbler	*Dendroica coronata*	AB, Temporal
YWAR	Yellow Warbler	*Dendroica petechia*	AB, BC, Temporal

Habitat Data

Forest inventory data and habitat classification. For the Alberta bird survey sites, we measured habitat patterns around each bird sampling station using 1:20,000 digital Alberta Vegetation Inventory (AVI) maps for both model development and model validation sites from Alberta. In BC, we did the same thing using Canfor's Vegetation Resource Inventory (VRI) data. Although the Alberta and BC forest inventory databases differ in certain respects (e.g., database structure, stand attributes, etc.), both contained the attributes required to develop a common, albeit not fully equivalent, habitat classification system. The forest cover layers of the AVI and VRI data contain several attributes useful for modeling wildlife habitat relationships such as species composition, crown closure, height, estimated stand age, and the location of non-forest cover types such as permanent clearings, lakes, and wetlands. Two additional map layers described the location of streams and logging roads. We developed a habitat classification system based on overstory tree species, stand age, and management and disturbance history (Table 2). The classification system was used to create a generalized map of forest and non-forest habitat classes for each site.

Table 2. Habitat classification system used to calculate several local and neighborhood-level habitat variables. Italicized comments relate to the habitat classification system developed for the BC study area.

Habitat class	***Description***
Water	River, lake, ice, river, and reservoirs
Non-forested	Vegetated – non forested upland and wetland areas

Early deciduous	> 70% deciduous with burn but not cut modifier or origin ≥ 1970 (fire-origin stands ≤ 30 years)
Young deciduous	> 70% deciduous and 31-90 years (1910 ≤ origin < 1970); no cut modifier
Old deciduous	> 70% deciduous and origin < 1910
White spruce	Conifer stands with > 70% white spruce as leading species
Black spruce	Conifer stands with black spruce as leading species
Pine	Conifer stands with pine as leading species
Mixedwood	Mixed deciduous/white spruce; include O_DECID if understory Crown > A and understory Sw > 30%; mixed deciduous/conifer stands
Recent cut	Clearcuts; any stand with a cut modifier including burns that have been salvaged; *Cutblocks where stand age ≤ 30 years*
Non-vegetated	Anthropogenic (wellsites, large cutlines, etc.); *non-vegetated – natural or anthropogenic areas*

Habitat-based predictor variables. For all Alberta and BC sites, we used original and derived map layers to measure habitat characteristics around each bird sampling station at two spatial scales: the local-scale, which matched the size and shape of the circular bird sampling stations (100 m radius, or 3.14 ha buffer), and the neighbourhood scale, which extended from 100-500 m beyond the sampling stations (75.4 ha annular buffer). At the local scale, we measured habitat class, stand height, crown closure, deciduous proportion, distance to nearest cutblock, and distance to nearest lake or river. At the neighbourhood scale (74 ha annular buffer) we measured the proportion of early (<15 years) and late (>90 years) seral forest, the proportion of deciduous and mixedwood forest, the presence of black spruce patches, and the variety of habitat classes. We selected predictor variables that included a similar range of variation in both model building and model testing datasets (Harrell 2001). These 12 habitat characteristics comprised our candidate set of predictor variables (Table 3). The process of selecting, generating, and evaluating the variables for inclusion in statistical models is described in Vernier et al. (2002).

Table 3. Habitat variables were derived from AVI and VRI data. Local habitat variables(L) were measured within a 100 m radius while neighborhood variables (N) were measured in a 400 m radius beyond each local (inner) buffer.

Variable	*Description*	*Study area*
L_CCUT	Station located in recent cutblock (<15 yrs)	AB, BC, Temporal
L_YDEC	Station located in young deciduous stand (<=90 yrs)	AB, Temporal
L_ODEC	Station located in old deciduous stand (>90 yrs)	AB, BC, Temporal
L_MIXED	Station is located within a mixedwood stand	AB, BC
L_PINE	Station is located in pine stand	Temporal
L_SIZE	Size of patch the station lies within (see habitat classes)	Temporal
L_CUTDIST	Distance of station centre to nearest anthropogenic edge (e.g. cutblock)	AB, BC
L_WATERDIST	Distance of station centre to nearest water body (river, lake, or reservoir)	AB, BC
L_CROWN	Mean crown closure of forested polygons x forested area (percent)	AB, BC, Temporal
L_DEC	Mean deciduous proportion of forested polygons x forested area	AB, BC, Temporal
L_HT	Mean stand height of forested polygons x forested area	AB, BC, Temporal

N_CUT	Proportion of neighbourhood in a recent cutblock	AB, BC, Temporal
N_MID	Proportion of neighbourhood in mid seral forest (15-90 years)	Temporal
N_LATE	Proportion of neighbourhood in late seral forest (origin < 1910)	AB, BC, Temporal
N_DEC	Proportion of neighbourhood in deciduous forest	AB, BC, Temporal
N_MIXED	Proportion of neighbourhood in mixedwood forest	AB, BC
N_SB	Presence of black spruce forest	AB, BC
N_SW	Presence of white spruce forest	Temporal
N_RICH	Number of habitat classes in neighbourhood	AB
N_SIMP	Diversity of habitat classes in neighbourhood (Simpson's index)	BC, Temporal
N_EDGEN*	Density of natural edges	AB, Temporal

* calculated using habitat classification system and edge contrast matrix

Model Development

All presence/absence models (GLM, logistic regression) were developed using the same general approach that we recently used to develop abundance models (Vernier et al., 2002; GLM, Poisson regression). For each species at each station we calculated the value of the Logit (P) using the equation

$$\text{Logit (P)} = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

from which we calculated the probability of occurrence using

$$p_i = e^{\eta}{}_i / 1 + e^{\eta}{}_i$$

where p_i is the detection probability (probability of occurrence at a given site or patch occupancy) of a species in the i^{th} station, and $\eta_i = \mathbf{x}'_i\boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$ is the linear predictor. $\mathbf{x}_i$ is the vector of independent variables for the i^{th} station and $\boldsymbol{\beta}$ is the vector of parameters to be estimated. Up to three different models were developed for each species: one that only included local-level habitat variables, another that included only neighbourhood-level habitat variables, and a third that included a combination of both types of variables. Variables in each model were selected by backward stepwise regression (p-to-enter < 0.05, p-to-remove <0.10). The best model among the competing models was selected using Akaike's Information Criterion (AIC). Where necessary, we used STATA's cluster option to calculate variance estimates that are robust to influential observations, within site correlations, and undetected over-dispersion (StataCorp 2001).

To build the models we used 1-7 years of data for the temporal validation component (Calling Lake 1993-1999) and 9 years of data for the external validation components (Calling Lake 1993-2001). The discrepancy reflects the fact that we performed the temporal evaluation of the models two years prior to the external validation. In addition, we compare model performance among the three different model structures only for the Alberta external validation component. In the other cases, we only evaluated the "best" model.

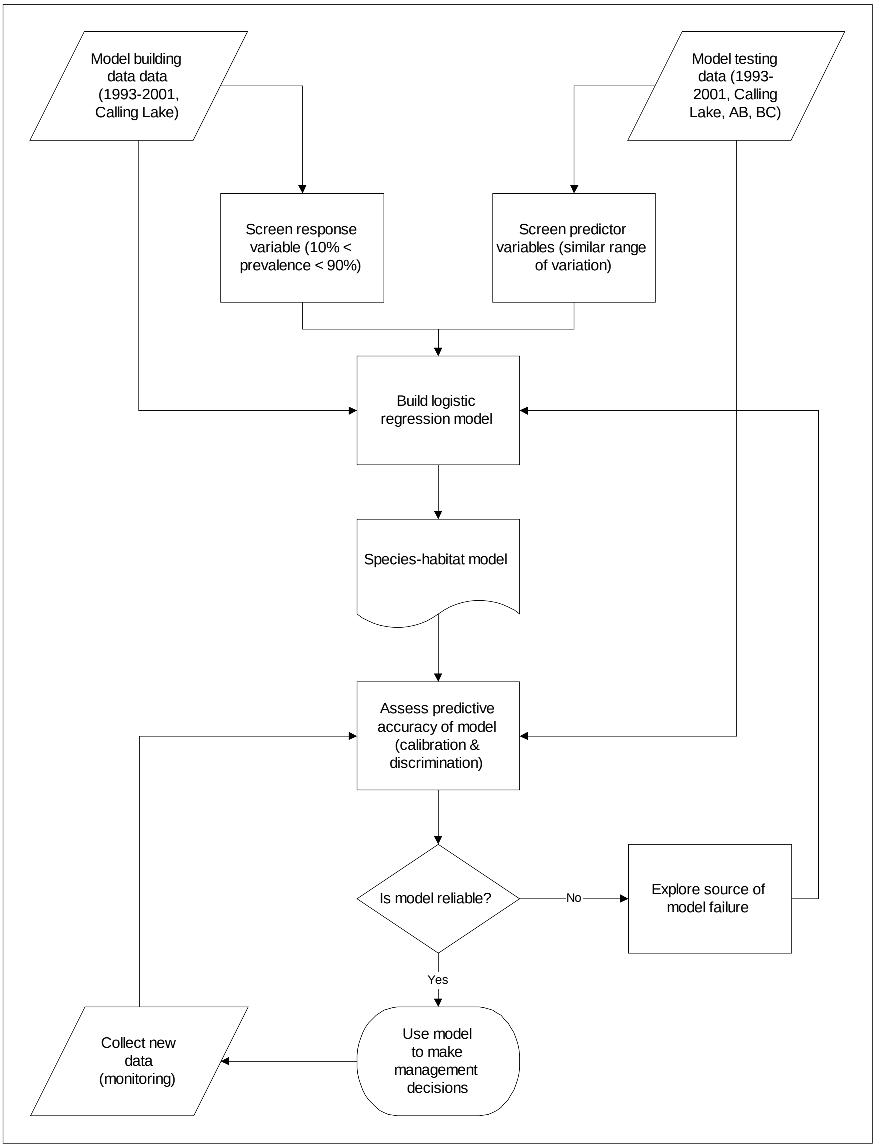


Figure 3. Simplified flowchart of the habitat model validation process.

Model Validation

We assessed the predictive accuracy of the songbird habitat models using Receiver Operating Characteristic (ROC) analysis. ROC analysis is a method of measuring and comparing the accuracy of a model at predicting whether each observation is a member of one of two groups (e.g. presence / absence). The ROC curve plots the Sensitivity (true positive rate) against 1-Specificity (false positive rate). The larger the area under the ROC curve (AUC), the better the model is at predicting group membership. As way of guidance for managers using such models, Swets (1988) considers models with AUC values between 0.5 and 0.7 to indicate poor discrimination capacity, values between 0.7 and 0.9 to indicate reasonable discrimination ability appropriate for many uses, and rates higher than 0.9 to indicate very good discrimination. A value of 0.5 indicates a model with no predictive power. Compared to other measures of model accuracy (e.g., sensitivity, specificity, correct classification rate, Kappa), AUC is not sensitive to prevalence (i.e. proportion of sites occupied by a species).

Table 4. Number of stations by year for developmental ("Original Models") and validation datasets.

Study Area	***1993***	***1994***	***1995***	***1996***	***1997***	***1998***	***1999***	***2000***	***2001***
Original Models									
Calling Lake, AB	174	264	341	234	367	385	385	282[1]	258[1]
Temporal Validation									
Calling Lake, AB		264	341	234	367	385	385		
External Validation									
Goodwin, AB					86				
North Calling, AB			53		80				
Owl River, AB				186	122				
Reference, AB					191				
Chetwynd, BC								560	560

[1] 2000 and 2001 data were not used for developing the models used in the temporal validation component.

Temporal validation (Calling Lake). To assess the temporal variability in songbird-habitat relationships, we developed separate models for each year (1993-1999) and recorded the direction, strength, and significance of the estimated coefficients. To evaluate the temporal variability in the predictive performance of the models, we used the models developed in objective 1 to calculate AUC_{in} for each year (e.g., 1995) and calculated AUC_{out} using data from the following year (e.g., 1996) excluding, the last year of sampling. To determine if the predictive performance of the models increased with the number of years used to fit the model, we developed models using 1 year of data, 2 years of data, 3 years of data, and so on, and validated each of these models using data from the following year. For example, a model using 5 years of data would be developed 3 times (i.e., 1993-97, 1994-98, 1995-99) and tested 3 times using AUC_{in} (i.e., 1993-97, 1994-98, 1995-99) and 2 times using AUC_{out} (i.e., 1998 and 1999). Out-of-sample tests for models that included 1999 data were not possible. We summarized our results graphically using the average AUC_{in} and AUC_{out} value for each "number of years" group (e.g. the mean of the 3 models developed with 5 years of data). We made no attempt to interpret inter-model variability because the number of possible models decreased linearly as the number of years included in the model increased. For instance, a model based on 1 year of data could be developed 7 times, while one base on 7 years of data could only be developed once.

External validation (Alberta). We evaluated songbird habitat models using 2 years of data (1996-1997) from 4 independent validation sites. The number of stations in each year for each of the model development and validation sites is shown in Table 4. We calculated the AUC_{in} and AUC_{out} for each of the three models (i.e. models using local, neighbourhood, and local + neighbourhood variables) for each species using the same data that was used to develop the models (internal validation) and using data from four geographically independent sites (external validation). In all, we validated the predictive performance of 16 songbird-habitat models.

External validation (BC). We evaluated modified versions of our "best" habitat models for Black-throated Green Warbler, Red-eyed Vireo, Swainson's Thrush, and Yellow Warbler using AUC_{in} and AUC_{out}. The models were modified because of minor differences between the forest inventory data of the two regions (see Table 2). In addition, we only included survey stations that were located in the BWBS and SBS biogeoclimatic zones – the two zones that are most similar to the Alberta sites.

Results of Temporal Validation (Alberta – Calling Lake)

There was moderate variability among years in the strength and significance of songbird-habitat model coefficients, and the overall model (all years) was generally not a good indicator for individual-year models (Table 5). Exploration of changes in abundance between years could provide further insight into why this might occur. Conversely, the direction (sign) of the coefficients, was largely consistent across years (16/21 over all species). For each species except Yellow-rumped Warbler, only one variable was consistently significant across years, and only for Red-breasted Nuthatch and Yellow Warbler was this variable also significant for the overall (all years) model.

The predictive performance of songbird models was more variable across years when assessed using out-of-sample data (ROC_{out}) than when using in-sample data (ROC_{in}) (Figure 4). Generally, all bird species had good model accuracy with ROC_{in} and ROC_{out} values > 0.7 for all years; the exceptions being Yellow-rumped Warbler in 1993 and Red-breasted Nuthatch in 1993-1995. For 3 of the bird species (Black-throated Green Warbler, White-throated Sparrow, and Yellow Warbler), the values of ROC_{in} and ROC_{out} are very similar and show little variability over time (Figure 4). In contrast, but only for the first 3 years, the other 2 species (Red-breasted Nuthatch and Yellow-rumped Warbler) have very different values and exhibit high variability. Differences in patterns for the first 3 years may be accounted for by landscape-level adjustments to forest harvesting in the area (Schmiegelow and Hannon, 1999; Norton et al. 2000). Nevertheless, although the strength and significance of model coefficients are quite variable, the models themselves are consistently in the "useful applications" and "high accuracy" categories, with the exceptions noted above. In other words, when prediction is the objective, the models appear to be robust, even though their reliability may vary across years. This suggests that care be taken when developing models shortly after disturbances.

Table 5. Estimated coefficients for species presence/absence models for the years 1993-1999. Emboldened coefficients are significant at the 5% level; italicized coefficients are significant at the 10% level.

Species / Variable	***all years***	***1993***	***1994***	***1995***	***1996***	***1997***	***1998***	***1999***
BGNW								
l_ht	-0.018	**0.251**	**0.279**	**0.272**	**0.195**	**0.136**	**0.179**	**0.122**
n_cut	2.138	4.609	**3.569**	**3.514**	0.023	0.164	-0.813	2.704
n_late	**5.032**	0.929	**2.671**	**4.409**	**2.233**	**2.133**	**2.929**	**2.523**
n_dec	**5.575**	**4.999**	**3.098**	**1.953**	1.142	**3.184**	**1.975**	**4.875**
n_sw	0.566	**1.624**	0.679	*1.239*	0.479	**0.891**	0.127	**0.997**
n_simp	**7.942**	2.577	**4.447**	**4.174**	**4.122**	**3.789**	**4.865**	**4.834**
RBNU								
l_ht	**0.127**	**0.164**	**0.149**	**0.19**	**0.16**	**0.232**	**0.295**	**0.144**
l_dec	**-4.053**	0.023	**-4.088**	-1.477	**-3.602**	**-2.986**	**-2.994**	**-1.463**
WTSP								
n_dec	**7.317**	**5.659**	**7.083**	0.587	**3.309**	**4.234**	**3.114**	**6.006**
n_simp	*4.233*	5.009	**6.993**	-0.013	2.812	**5.227**	**4.515**	**4.832**
l_dec	2.184	**2.839**	**2.207**	**4.921**	**4.247**	**3.129**	**2.442**	**2.657**
l_ccut	19.297	-0.933	**1.533**	22.229	**5.753**	20.489	**3.967**	-0.389
n_mid	-2.85	*-1.646*	**-5.632**	**-3.856**	-1.99	**-3.618**	**-4.532**	**-2.168**
l_pine	1.082	0.446	0.44	**-3.728**	-1.439	-0.946	0.307	0.736
n_edgen	-0.014	**-0.085**	*-0.029*	-0.03	-0.036	**-0.036**	-0.024	**-0.06**
YRWA								
n_late	**5.48**	3.197	**3.14**	0.505	**1.864**	*1.712*	**2.361**	**4.72**
l_odec	-0.362	*-1.959*	**-3.203**	**-16.922**	-0.403	-0.46	0.325	-1.092
l_ccut	17.191	**-6.52**	**-4.28**	**-20.862**	-25.478	**-4.466**	**-3.831**	**-3.491**
n_mid	**3.044**	**4.32**	**3.275**	0.908	**7.121**	**3.141**	**2.58**	**2.974**
l_ydec	0.939	**-3.122**	-2.128	-17.01	**-3.037**	*-1.774*	-0.94	-0.577
l_size	**-0.006**	**-0.005**	**-0.004**	-0.001	**-0.003**	**-0.003**	-0.001	**-0.005**
YWAR								
l_odec	**2.458**	**2.635**	**3.333**	**2.188**	**2.82**	**1.496**	**2.441**	**2.603**
l_crown	**-0.026**	**-0.005**	*-0.014*	-0.01	**-0.038**	**-0.038**	**-0.039**	-0.012

The relationship between mean model performance (i.e., the average of the models with the same number of years of data) and the number of years used to develop the model is summarized in Figure 5. With the exception of Black-throated Green Warbler, out-of-sample tests (ROC_{out}) were more variable than in-sample tests (ROC_{in}). In fact, in-sample tests appeared to be little affected by the number of years used to develop the models. Out-of-sample evaluations were more variable, but only in the case of Red-breasted Nuthatch did performance increase consistently with number of years. This result is not surprising, given that among those species we analyze here, the Red-breasted Nuthatch exhibits the highest spatial and temporal variance in distribution and abundance (Carlson and Schmiegelow, 2002). For two species, White-throated

Sparrow and Yellow-rumped Warbler, out-of-sample tests actually indicated a loss in predictive performance, albeit minor, with number of years. We make no attempt to assign significance, as differences in the number of possible models as a function of the number of years included in the model made interpretation of variance problematic. Nevertheless, both in-sample and out-of-sample model performance was always greater than 0.7 indicating "useful applications" and "high accuracy" models.

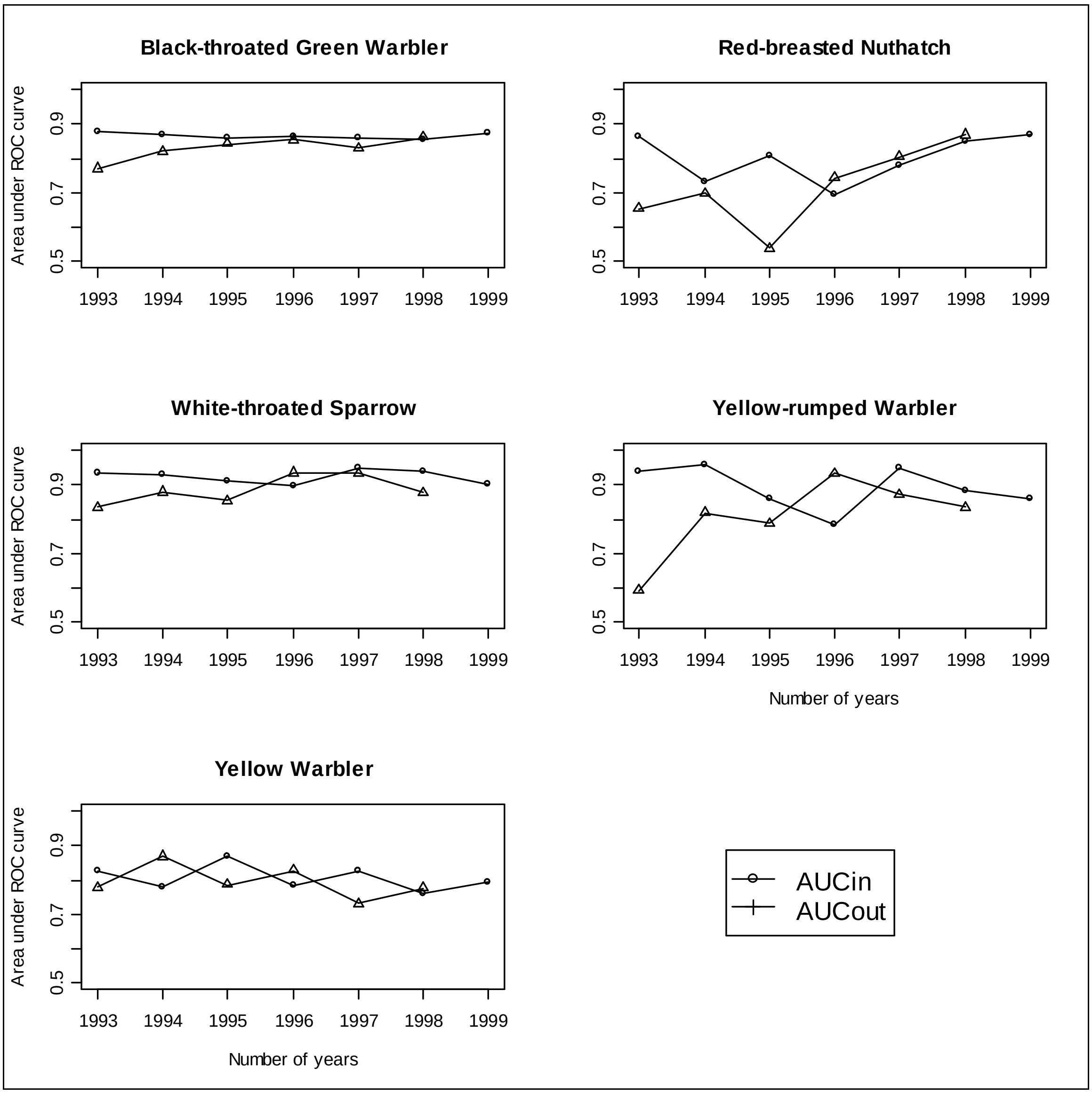


Figure 4. Temporal variability in ROC (AUC_{in} and AUC_{out}) for 5 species of boreal songbirds in Alberta. All models were developed using 1 year of data and validated using the following year of data. No validation was possible for the 1999 models because there is no data for the following year.

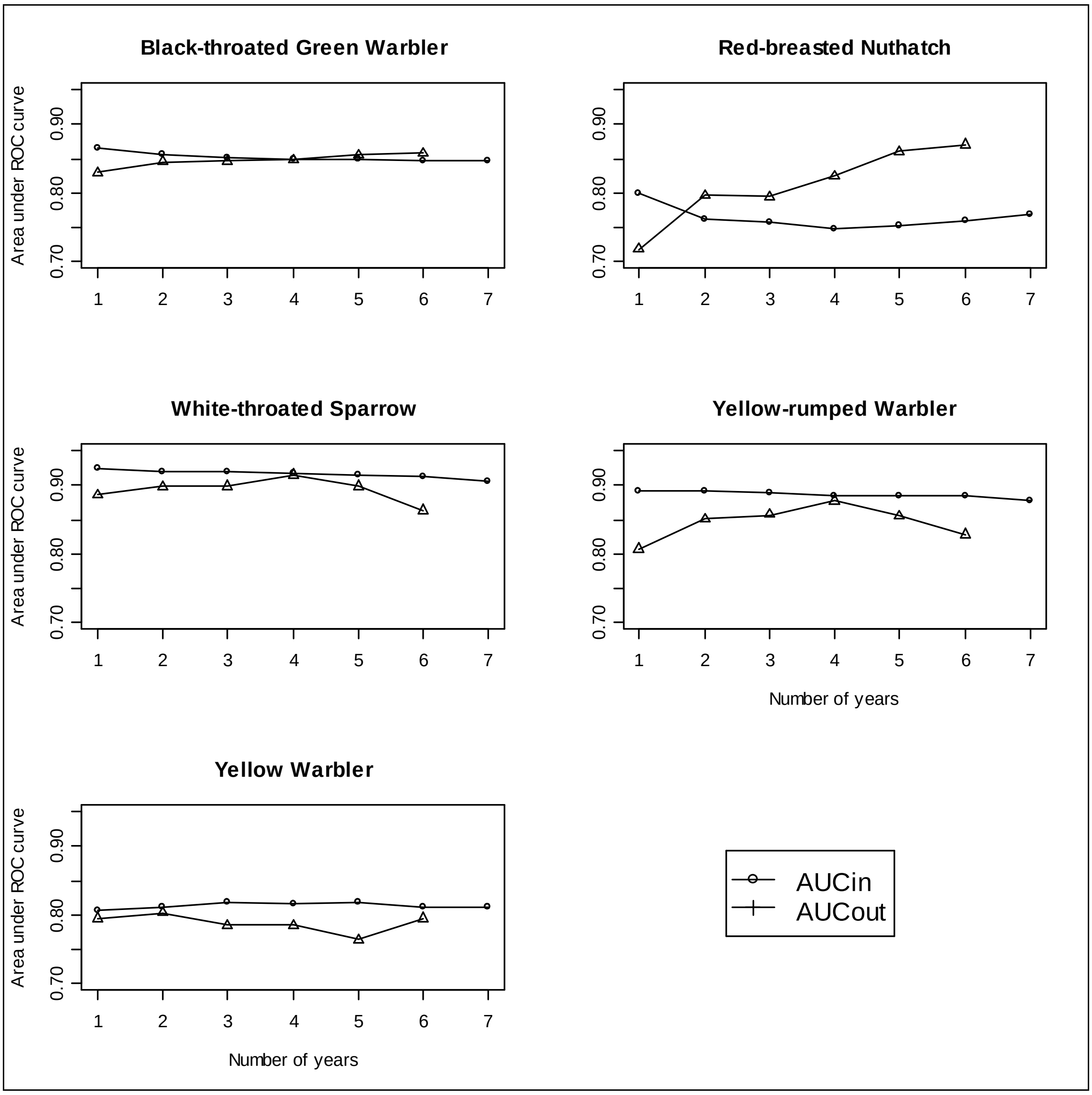


Figure 5. Relationship between mean model performance (AUC_{in} and AUC_{out}) and the number of years used to build the models. Data were validated using data from the following year. Note that out-of-sample predictions were not possible for models developed using all 7 years of data.

Results of External Validation (Alberta – Mixedwood)

We developed logistic regression models for 16 species of forest songbirds that occur in the boreal mixedwood forest using data from the Calling Lake study area (Table 6). The number of significant predictor variables in each model ranged from 4 for the Yellow-rumped Warbler to 10 for the Chipping Sparrow and Least Flycatcher. Local habitat variables were included 67 times in the models compared to 49 times for neighbourhood habitat variables. All local habitat variables except for L_WATERDIST occurred in at least 7 models and as many as 11 models. All neighbourhood habitat variables except for N_MIXED occurred in at least 6 and as many as 12 of the models. The internal predictive accuracy of models, as measured by AUC, ranged from 0.677 for Chipping Sparrow to 0.842 for American Redstart. Only two species' models, those for Chipping Sparrow and Tennessee Warbler, were below the 0.7 cutoff indicating a reasonable model – and in both cases they were just below with AUC values of 0.677 and 0.691, respectively. In contrast, the other 14 models are considered to have reasonable discrimination ability appropriate for management purposes, especially in the area in which they were developed i.e. Calling Lake. For all species, the best model, in terms of predictive accuracy, was the model comprising a combination of both local and neighbourhood variables.

Table 6. Summary of logistic regression models for selected songbird species using the Calling Lake bird survey data. All predictor variables are significant (p < 0.10). LL = log likelihood; AUCin = in-sample predictive ability.

Species	***Habitat variables***	***LL***	***AUC n=3031***
AMRE	-4.401 - 1.85·H_CUT + 1.276·H_ODEC -0.017·L_CC + 5.351·N_CUT + 1.623·N_LATE + 3.119·N_DECID - 0.503·N_MIXED -0.358·N_SB	-1028	0.842
BGNW	-3.623 - 2.589·H_CUT + 1.322·H_ODEC + 0.689·H_MIXED -1.701·L_DEC + 0.102·L_HT - 0.018·L_CC + 0.784·N_LATE + 1.374·N_DECID + 0.363·N_SB + 0.16·N_RICH	-1275	0.815
CHSP	-0.79 - 0.618·H_YDEC - 1.303·L_DEC + 0.033·L_HT - 1.728·N_CUT+ 0.378·N_MIXED + 0.442·N_SB	-1560	0.677
COWA	-1.364 - 1.797·H_CUT + 0.633·H_YDEC + 2.55·L_DEC - 0.06·L_HT - 0.011·L_CC - 1.05·N_LATE + 0.774·N_MIXED	-1232	0.758
LEFL	-2.565 + 1.479·H_ODEC + 1.1·H_YDEC + 2.443·L_DEC - 0.051·L_CC - 0.001·DIST + 2.923·N_CUT + 1.139·N_LATE + 2.346·N_DECID - 0.421·N_MIXED - 0.311·N_SB	-1263	0.796
MOWA	-3.265 + 1.332·H_ODEC + 1.181·H_YDEC + 1.098·L_DEC - 0.026·L_CC + 1.259·N_CUT + 1.032·N_LATE + 1.719·N_DECID - 0.503·N_SB + 0.189·N_RICH	-1541	0.756
OVEN	0.589 - 2.148·H_CUT + 1.274·H_YDEC + 0.961·H_MIXED + 0.909·L_DEC + ·L_HT - 1.255·N_CUT - 0.856·N_LATE	-1365	0.816
RBGR	-2.703 + 1.538·H_ODEC + 0.935·H_YDEC + 0.031·L_HT - 0.015·L_CC + 2.112·N_CUT + 1.008·N_DECID - 0.103·N_RICH	-1231	0.730
RBNU	-1.6 - 2.784·H_CUT - 0.496·H_YDEC + 0.551·H_MIXED - 1.13·L_DEC + 0.068·L_HT - 0.008·L_CC - 0.935·N_CUT + 0.175·N_RICH	-1357	0.751
REVI	-0.208 + 1.14·H_ODEC + 1.685·H_YDEC + 0.604·H_MIXED + 0·DIST + 2.553·N_CUT + 1.08·N_DECID - 0.206·N_RICH	-1660	0.720
SWTH	-0.944 - 1.767·H_CUT - 1.563·L_DEC + 0.054·L_HT + 1.444·N_LATE - 0.77·N_DECID	-1424	0.752
TEWA	-0.988 + 0.831·H_ODEC + 0.413·H_MIXED - 0.006·L_CC + 2.574·N_CUT + 1.418·N_DECID	-1658	0.691
WIWR	-1.938 - 2.485·H_CUT - 1.253·L_DEC + 0.046·L_HT + 0·DIST + 1.29·N_CUT+ 2.339·N_LATE	-1315	0.760
WTSP	-2.196 + 1.982·H_CUT + 1.809·H_ODEC + 1.334·H_YDEC + 0.678·H_MIXED + 0.069·L_HT - 0.035·L_CC + 3.307·N_CUT + 1.068·N_LATE + 2.243·N_DECID	-1149	0.826
YRWA	1.909 - 4.056·H_CUT - 2.315·L_DEC + 0.097·L_HT - 0.928·N_DECID	-1083	0.834
YWAR	-3.764 + 0.617·H_ODEC - 1.105·H_MIXED + 2.329·L_DEC - 0.046·L_CC + 2.248·N_CUT + 1.201·N_LATE + 1.739·N_DECID - 0.579·N_SB + 0.144·N_RICH	-1172	0.791

We tested all 16 songbird habitat models using external validation data from 4 independent sites individually and as a whole. The tests were only performed if a species' prevalence at a given test site was between 10-90%. No clear pattern emerged from the external validation of the Calling Lake models. Fifteen of the 16 habitat models performed reasonably well in at least one of the validation sites or when all validation sites were combined (Table 7). The Red-breasted Nuthatch models were the exception but they were only tested at the Owl River site and with all the validation data. Two models, those for Chipping Sparrow and Tennessee Warbler performed reasonably well even though the original model was considered to be poor. Eight of the 16 habitat models had reasonable predictive accuracy when tested against the validation data as a whole. Among the individual validation sites there were no clear patterns – some species' models appeared to perform better at some sites than at others. Models that were successfully validated using all the data did not necessarily validate reasonably well at individual sites and vice-versa – models that did not validate using all the data did validate at some individual sites. The North Calling Lake and Reference sites had the most number of models validated, 8 and 7, respectively. Five models were validated at the Owl River site and only 2 at the Goodwin site that had the smallest sample size. Only one model (Ovenbird; local variables only) performed reasonably well with all the validation data and at each validation site. The Ovenbird model (local + neighbourhood variables) and the Red-eye Vireo model (local + neighbourhood variables) both performed reasonably well using all validation data and at 3 of the 4 validation sites. Among the species whose model(s) performed reasonably well when tested against external validation data, individually or as a whole, 13 models comprised local and neighbourhood level habitat variables, 8 consisted of local habitat variables only, and 6 included neighbourhood habitat variables only.

As a final step, we developed new logistic regression models for each species using all of the validation data combined. The results (not included in this report; Vernier et al. *in prep*) clearly show that these new models performed better at the validation sites in terms of discrimination capability than the original models developed in the Calling Lake study area – although the test only included internal data and thus was not as stringent as the validation results described in the previous section. Fourteen of the 16 models had reasonable predictive capability. Models for Least Flycatcher, Rose-breasted Grosbeak, and Red-breasted Nuthatch were the exceptions. In general the models were more parsimonious i.e., fewer predictor variables entered the models. On average each species included 3.7 variables compared to 6.1 for the original models. As with the original models the same five predictor variables (N_CUT, N_DECID, L_DEC, N_LATE, L_CC) were most often selected among the 14 candidate variables.

Table 7. Results of Alberta validation analysis. AUC local, AUC nbr, AUC local+nbr = area under the ROC curve for model using local, neighbourhood, or both sets of variables. AUC values are not provided where prevalence is less than 0.10 or greater than 0.90.

			Validation Sites				
Species	**Statistic**	**Calling Lake**	**All valid. sites**	**Goodwin**	**North CL**	**Owl River**	**Reference**
AMRE	Prevalence	0.213	0.152	0.011	0.203	0.227	0.063
	AUC local	**0.801**	0.558		**0.787**	0.504	
	AUC nbr	**0.774**	0.633		**0.797**	0.519	
	AUC local+nbr	**0.842**	0.652		**0.891**	0.539	
BGNW	Prevalence	0.32	0.105	0	0.241	0.11	0.052

	AUC local	**0.79**	0.633		**0.751**	0.553	
	AUC nbr	**0.746**	**0.721**		**0.752**	0.574	
	AUC local+nbr	**0.815**	**0.703**		**0.753**	0.593	
CHSP	Prevalence	0.313	0.317	0.222	0.474	0.175	0.482
	AUC local	0.627	**0.732**	0.648	0.622	**0.792**	0.67
	AUC nbr	0.652	0.678	0.615	0.631	**0.712**	0.568
	AUC local+nbr	0.677	**0.723**	0.673	0.585	**0.769**	0.664
COWA	Prevalence	0.232	0.306	0.189	0.173	0.39	0.319
	AUC local	**0.745**	0.561	0.47	0.351	0.398	**0.775**
	AUC nbr	0.694	0.599	0.572	0.366	0.368	**0.809**
	AUC local+nbr	**0.758**	0.584	0.41	0.393	0.422	**0.783**
LEFL	Prevalence	0.274	0.22	0.189	0.263	0.237	0.178
	AUC local	**0.771**	0.614	0.472	0.581	0.568	**0.742**
	AUC nbr	**0.708**	0.575	0.289	0.533	0.567	0.687
	AUC local+nbr	**0.796**	0.622	0.379	0.577	0.593	**0.755**
MOWA	Prevalence	0.407	0.378	0.278	0.759	0.406	0.115
	AUC local	**0.736**	0.67	0.527	0.668	0.616	**0.735**
	AUC nbr	0.693	**0.716**	0.467	0.594	0.52	**0.756**
	AUC local+nbr	**0.756**	**0.72**	0.496	0.646	0.609	**0.752**
OVEN	Prevalence	0.561	0.582	0.456	0.158	0.805	0.576
	AUC local	**0.814**	**0.861**	**0.81**	**0.771**	**0.862**	**0.799**
	AUC nbr	**0.709**	**0.806**	0.619	**0.766**	**0.803**	**0.714**
	AUC local+nbr	**0.816**	**0.864**	**0.753**	0.684	**0.869**	**0.801**
RBGR	Prevalence	0.209	0.248	0.1	0.331	0.334	0.12
	AUC local	0.695	0.597	0.333	0.626	0.576	0.679
	AUC nbr	0.665	0.67	0.694	0.618	0.564	**0.704**
	AUC local+nbr	**0.73**	0.651	0.527	0.641	0.606	0.652
RBNU	Prevalence	0.287	0.126	0.078	0.083	0.188	0.079
	AUC local	**0.738**	0.538			0.606	
	AUC nbr	0.695	0.465			0.514	
	AUC local+nbr	**0.751**	0.528			0.591	
REVI	Prevalence	0.511	0.528	0.333	0.451	0.685	0.419
	AUC local	0.675	0.685	0.647	**0.719**	0.608	0.681
	AUC nbr	0.656	**0.741**	**0.743**	0.549	**0.746**	0.679
	AUC local+nbr	**0.72**	**0.764**	**0.765**	**0.759**	0.74	0.679
SWTH	Prevalence	0.31	0.073	0.078	0.18	0.045	0.042
	AUC local	**0.743**			0.818		
	AUC nbr	**0.715**			0.807		
	AUC local+nbr	**0.752**			0.827		
TEWA	Prevalence	0.603	0.863	0.889	0.805	0.88	0.864
	AUC local	0.664	0.601	0.516	0.527	**0.706**	**0.864**
	AUC nbr	0.663	0.52	0.483	0.534	0.477	**0.877**
	AUC local+nbr	0.691	0.573	0.576	0.592	0.595	**0.873**

WIWR	Prevalence	0.266	0.109	0.267	0.301	0.045	0.005
	AUC local	**0.749**	0.652	0.606	**0.707**		
	AUC nbr	**0.713**	0.833	0.682	0.689		
	AUC local+nbr	**0.76**	0.734	0.671	0.669		
WTSP	Prevalence	0.747	0.699	0.778	0.917	0.666	0.565
	AUC local	**0.797**	**0.777**	0.563		**0.723**	**0.781**
	AUC nbr	**0.77**	**0.704**	0.46		0.645	**0.717**
	AUC local+nbr	**0.826**	**0.753**	0.411		**0.767**	**0.742**
YRWA	Prevalence	0.675	0.481	0.511	0.414	0.344	0.733
	AUC local	**0.833**	**0.727**	0.63	0.696	0.641	**0.724**
	AUC nbr	**0.735**	**0.707**	0.616	0.695	0.618	**0.705**
	AUC local+nbr	**0.834**	**0.731**	0.639	0.668	0.633	**0.755**
YWAR	Prevalence	0.221	0.086	0.056	0.271	0.065	0.005
	AUC local	**0.781**			0.612		
	AUC nbr	**0.736**			0.699		
	AUC local+nbr	**0.791**			0.649		
No. of stations:		2690	722	90	133	308	191

Results of External Validation (BC – Northeast)

We calculated the AUC for both the original (in-sample) data from Calling Lake Alberta (AUC_{in}, Table 8) as well as independent (out-of-sample) data collected in Block 4, TFL 48 (AUC_{out}). For all four songbird species, the AUC_{in} values were greater than 0.75, indicating that the models are reliable – at least in Alberta. However, when we tested our models in BC, AUC values were quite poor (AUC_{out}; Table 8), indicating that our Alberta models had poor predictive ability when transferred to BC. We then used the same set of candidate predictor variables (Table 3) and the approach described earlier to re-develop (refine) our songbird habitat models, this time using the BC validation data. Table 9 summarizes the structure of the refined models as well as their predictive ability using the BC data. In this case, 3 of the 4 songbird models performed well (AUC > 0.8), the exception being the habitat model for Swainson's Thrush – a species whose habitat associations may not be adequately captured with data emphasizing overstory structure.

Table 8. Summary of logistic regression models for BGNW, REVI, SWTH, and YWAR developed using Alberta survey data. LL = log likelihood; AUC_{in} = in-sample predictive ability; AUC_{out} = out-of-sample predictive ability.

Species	*Habitat variables*	*LL*	*AUC_{in} N=3031*	*AUC_{out} N=431*
BGNW	-10.425 + 3.888·N_SB + 0.494·H_ODEC + 0.548·H_MIXED + 2.974·N_MIXED - 1.369·LX_DEC + 0.207·LX_HT - 0.01·LX_CROWN + 4.53·N_DEC + 3.682·N_SIMP + 2.891·N_CUT + 1.324·N_LATE	-1336.6	0.839	0.571
REVI	0.52 - 2.897·N_SB - 0.506·H_MIXED + 2.079·LX_DEC - 1.442·N_SIMP - 0.845·N_LATE - 0.363·L_WATERDIST	-1767.8	0.760	0.602

SWTH	-0.7 + 1.199·N_MIXED - 1.341·LX_DEC + 1.79·N_LATE - 3.282·H_CUT	-1529.8	0.763	0.474
YWAR	-0.332 - 4.001·N_SB + 0.94·H_ODEC - 4.038·N_MIXED + 2.399·LX_DEC - 0.052·LX_CROWN + 1.034·N_LATE - 0.341·L_WATERDIST – 1.182·L_CUTDIST	-1224.3	0.819	0.660

Table 9. Summary of logistic regression models for BTGW, REVI, SWTH, and YWAR developed using bird survey data from Block 4, TFL 48. LL = log likelihood; AUC_{in} = in-sample predictive ability.

Species	*Habitat variables*	*LL*	***AUC*$_{in}$ *N=560***
BGNW	-20.582 - 1.346·H_ODEC + 0.173·LX_HT + 11.455·N_SIMP + 8.329·N_LATE	-75.5	0.873
REVI	-4.398 + 3.477·N_DEC	-87.4	0.802
SWTH	-1.801 + 2.603·N_MIXED + 1.102·N_DEC	-294.3	0.663
YWAR	-4.827 + 4.242·N_DEC + 1.128·N_MIXED	-112.7	0.848

Discussion and Management Implications

A number of in-sample and out-of-sample model evaluation techniques are available for presence/absence modeling (Fielding and Bell, 1997). Without such evaluations it is difficult to interpret the predictive ability of habitat models, and therefore, the reliability of these models as resource management tools. Although in-sample resubstitution techniques are frequently used, they have a tendency to produce over-fitted models, optimistic estimates of model performance and loss of generality (Harrell 2001). Consequently, out-of-sample or external validation approaches provide a more realistic assessment of model reliability. A compromise, but less stringent approach when multiple years of data are available from the same site, is to evaluate the performance of models developed using one or more years of data with data from subsequent year(s) of sampling. In this project, we assessed the performance of a number of songbird-habitat models using all three approaches. However, we only used the internal validation results as a means of comparing the results from the temporal and external validation analyses.

The temporal validation analyses identified that the habitat model coefficients we estimated were quite variable between years, making the development of general models (multi-year response) difficult for most species. Such behavior by some bird species complicates the possible application of such models in natural resource management and conservation planning. Likewise, year-to-year variability in model performance was evident for most species. However, most songbird models performed adequately across years with minor exceptions. The relationship between model performance and the number of years of data used to develop a model was most pronounced for Red-breasted Nuthatch where there was a clear increase in reliability as more data were used in the modeling process. Surprisingly, this was not the case for the other four species investigated.

The songbird-habitat models performed reasonably to poorly when coefficients estimated in

Alberta were used to make predictions at other locations in Alberta and BC. In Alberta, the performance of the songbird-habitat models differed substantially between the original Calling Lake data and the validation data. Using the original data, 14 of the 16 models had reasonable predictive ability. The other two were just below the cutoff defining a suitable model (AUC $\geq$ 0.7). In contrast, when we tested the models using all of the validation data, only 8 models were considered to have been validated. Similarly, when we tested the models against the individual validation sites, between 2 and 8 models proved to be reliable. Only the Ovenbird habitat model performed reasonably well across all validation sites. The White-throated Sparrow model performed reasonably well with all datasets except for the Goodwin site. When we tested four of the Calling Lake songbird models against independent data collected in northeastern BC all four models performed poorly. However, the predictive performance of 3 of the 4 models improved when local data was used to develop the models. The same was true for most of the models tested in Alberta and refined using local data (Vernier et al. *in prep*).

There are a number of possible reasons why several songbird-habitat models fared poorly when coefficients estimated in Calling Lake were used to make predictions in other geographic locations. These include:

- Non random / un-representative survey design;
- Regional differences in landscape-level habitat composition and configuration;
- Differences in survey protocols including observer variability;
- Natural spatial and temporal variability in songbird abundance;
- Differences in the prevalence of songbird species and avian community structure in the different sites and between BC and Alberta;
- Important habitat characteristics not identified or measured;
- Differences in detectability of species in different sites; and
- Spatial-temporal variation in habitat use.

Most of these reasons are likely to be more pronounced in the BC validation site where landscape characteristics and survey protocols were most divergent from those of Calling Lake and the other Alberta validation sites. In fact, the songbird survey data from northeast BC were collected as part of a Northern Goshawk study and was not designed to be random or representative of all forest types in the region.

Although the validation analyses pointed to some weaknesses in the current suite of songbird-habitat models, they also demonstrated that it is possible to use the same set of habitat attributes in the model development and model validation sites, within Alberta, to develop models that had equally good predictive ability, at least for the majority of the species, when tested using in-sample data (i.e. the validation data in this case). In fact, the top 5 predictor variables in Alberta in terms of the number of times they were included in the habitat models, were the same for both sets of models. The same pattern was also evident in BC where we used the same set of habitat attributes in both Alberta and BC to develop models that had equally good predictive ability for at 3 of our 4 focal species, when tested using in-sample data. Moreover, the predictor variables that we used in the BC and Alberta habitat models are easy to measure and have the advantage that they can be manipulated thru management activities to achieve desired goals. Thus, even though most of our models are not generalizable to all locations, they can be adapted (re-estimated) when or where necessary using the same set of habitat variables along with local survey data. This last point underlines the importance collecting local data to evaluate and refine models to local conditions and to feedback into the adaptive management process.

In some circumstances, the development of reliable models may be quite straightforward (e.g. habitat specialists like Northern Spotted Owls). However, given the spatially and temporally dynamic nature of habitat selection common to many species, models that can be used across geographic areas are not necessarily expected. An alternative is to develop expert-based models (e.g., HSI) or to strive for general models by understanding functional responses and the influence of environmental variation on the availability or quality of habitat resources (Boyce and McDonald 1999). Understanding such relationships is of crucial importance in natural resource management and conservation, because managers and conservationists are asked to provide habitat-based models describing the influence of changing land-use activities on sensitive or rare species (cumulative effects assessments, population viability analyses, climate change models, etc.). Nevertheless, theoretical models also need to be validated to assess their reliability in a management context. A less satisfying but possibly necessary approach is simply to develop different models for different seasons, years or localities – as indicated by the results of our analyses. Generalist species, like some forest songbirds, likely will require such an approach, because substantial differences in selection are apparent across years, between geographic locations, and over regional scales. Given the spatial and temporal variability in model performance, it may be best to use such habitat models in a relative sense when evaluating management scenarios i.e. using models to compare alternative management scenarios rather than to make site/time specific predictions.

LANDSCAPE SIMULATION MODELING (FEENIX)

During the past three years, and in collaboration with other BEEST projects, we continued to develop and enhance FEENIX, and apply it to evaluate management scenarios and policy alternatives. In the following sections we describe the technological and functional developments of the FEENIX software. A detailed description of an earlier version of FEENIX can be found in Cumming et al. (1998). The application of FEENIX for research and management purposes is described in a companion BEEST project.

FEENIX Enhancements

Profiling (FEENIX internals): A simple run-time profiling facility has been added to the model. This feature helps developers attempting to improve model performance by providing performance benchmarks and identifying bottlenecks where the greatest gains may be made.

Efficiency (FEENIX internals): A redesign of model internals has decreased model execution times by approximately 80% and also markedly reduced memory requirements. A typical simulation experiment (e.g. 100 model runs of 300 years each on a 300,000ha landscape) can now be executed in 8hr on a midrange contemporary PC (e.g. a 2.4 MHz Pentium IV machine), instead of 1.5 to 2 days as previously. As a result, systematic simulation experiments are now much more feasible.

Spruce-dynamics model (FEENIX internals): The original implementation of the mixed stand dynamics model (Cumming et al. 1998) contained several algorithmic, design and coding errors. These have been corrected. Although preliminary assessments indicate that simulated landscape dynamics (Cumming et al. 1999) are not markedly affected by these errors, their correction does increase confidence in interpretation of model results.

Monte Carlo simulation: FEENIX has successively produced full Monte Carlo scenario analyses. It can now be used more efficiently and effectively as a policy analysis tool. Users can define key files and parameter ranges that define a management scenario in a single project definition file. Output can consist of maps or tracking variable files and can be performed for any number of simulations, length of simulation run, or range of management options. All output files are prefixed with the title of the project definition file to provide the user with a reference of which output corresponds to which scenario. The operation is completely automatic once the definition files have been produced.

FEENIX and GIS: FEENIX now reads and operates on raster-based ASCII map files, such as those that can be generated from ArcView and GRASS geographic information systems (GIS). The resolution of input layers defines the resolution of a FEENIX model, although currently the model only has the option of running at 9ha or 3ha resolutions. The flexibility of having the resolution defined by the input data source is intended to support development of new spatial models using the "stripped-down" FEENIX shell (see next paragraph). FEENIX also has the capability to output files in ESRI ASCII grid format. The choice of output layers and the interval of layer production are set by the user. This choice can be predefined in a project definition file or executed by the user during a simulation run. Map production can be automatically defined for scenario analysis in Monte Carlo simulations or within the graphic user interface at any point in time.

FEENIX shell: A prototype, "stripped-down" version has been developed and contains only basic spatial data management and graphics. It is intended to simplify model production and use, and to make it more accessible to non-programmers. It currently reads base-maps as raster grids and converts to FEENIX format. Functionality is basic but provides the foundation for creating spatial ecological models. The tool is ideal for bottom-up ecological model development, so that researchers can build models without the overhead of complex detail and management options currently in FEENIX. Because all internals are compatible with FEENIX, the functionality can then be transferred to FEENIX and integrated by an experienced programmer. The shell version has all the same input, output and scenario evaluation capabilities as the full version.

Salvage logging: In scenarios where both harvesting and fire are enabled, merchantable burnt areas may be salvaged. This prototype feature is enabled by the DoSalvage option. The implementation simulates key features of regulations in force in the Province of Alberta as of 2002. In particular, contiguous patches of burnt area within each salvaged fire are excluded from salvage. The proportion and size of these leave areas depends on the size of the fire. A reduction factor (0.85 by default) is applied to salvage volumes, and unsalvaged standing burnt wood decays at an annual rate (0.5 by default). Salvage volumes are counted towards per-strata or per-landbase Annual Allowable Cuts (AACs) except that salvage volumes may exceed calculated AACs by up to 50% to allow for limited surge cutting after large fires.

Patch configuration metrics: Support for the computation of landscape pattern metrics has been added at the application level. Most of the pattern metrics identified by Cumming and Vernier (2002) are generated, including the FRAGSTATS metrics MPI, AWMPFD, TCA and MSI which are based on relatively efficient implementations of the published definitions (McGarrigal and Marks 1995). These metrics are computationally intensive and should only be generated at intervals within a simulation run. In typical applications, we generate pattern metrics every 5 years, which accounts for 20-30% of total execution time. In principle, these metrics could be generated externally by exporting data from FEENIX and running FRAGSTATS or some similar utility. In practice, however, this process would be both cumbersome and error-prone, especially in Monte-Carlo simulation experiments.

Networks: Various approaches for sub-resolution features such as stream and road-networks have been evaluated in prototype. At a 3ha resolution, FEENIX cells are approximately 170m square, but most roads and streams are much narrower. Thus, roads and streams should not be modelled as distinct vegetation or land-cover types, but rather as a cell attribute or modifier. Our approaches use the lengths of streams, roads or similar features intersecting a cell as a cell attribute. These lengths can be easily generated from GIS coverages. Some capabilities of vector-based network representations, such as used in ARC/INFO, have been implemented, including a join operator to connect disjoint but adjacent network segments. These capabilities add substantially to the dynamic road network construction and routing algorithms already present in the model, and have been critical to several extension or knowledge transfer exercises as described elsewhere. Further development of these capabilities over the next two years will be contributed through the SFMN project of Kurz, Stelfox and Cumming.

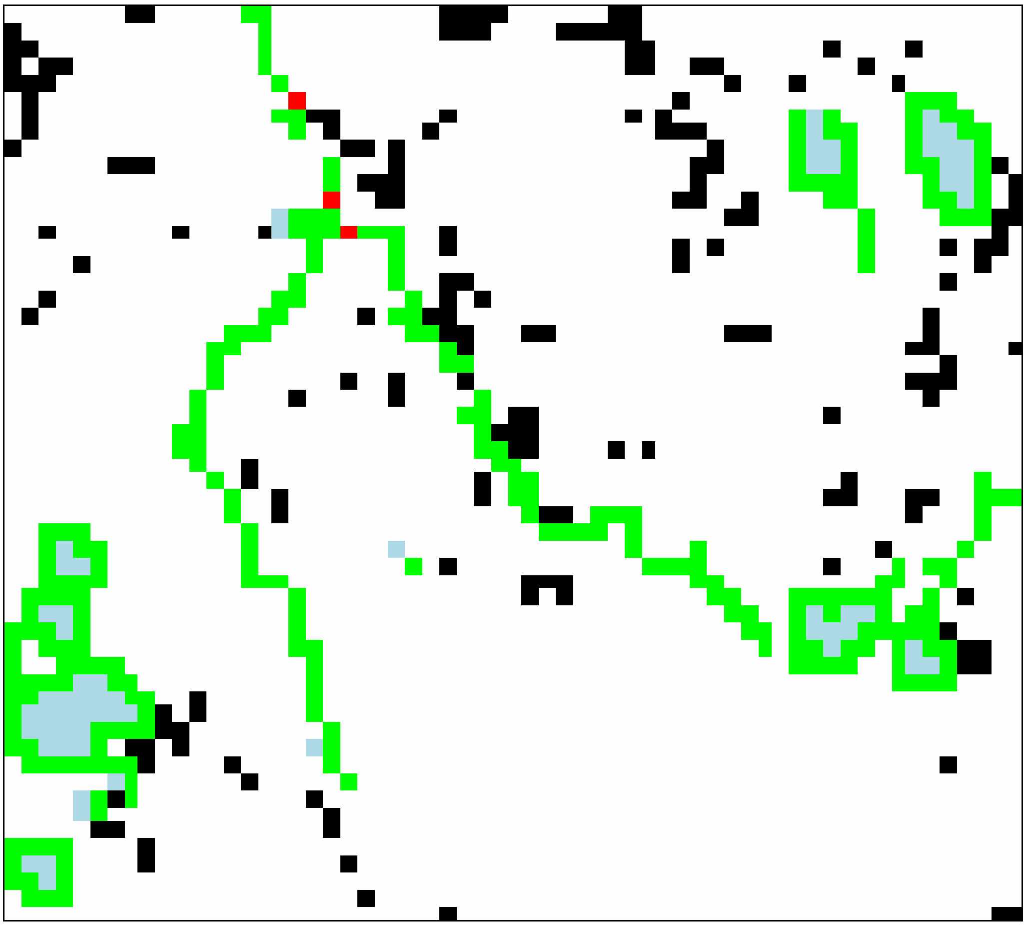


Figure 7. An example of FEENIX's capability at generating networks within a 10,000 ha landscape north of Lac La Biche, Alberta. Some disjoint pieces of a stream network (in green) were joined up (red cells). The pale blue areas are lakes, and the black areas are forested areas that were netted out of the operable landbase. Also in green are buffer areas around the lakes.

Strata-based harvest scheduling: The FEENIX forest management model has been substantially revised to be flexible, general and robust. Each merchantable cell has a yield class defined by a two species-group yield table or volume/age curve (hardwood and softwood species-groups are sufficient for our applications in the boreal mixedwood region). FEENIX interpolates yield curves to 1yr resolution and derives mean annual increments, culmination or rotation ages and minimum and maximum harvest ages. Cells are assigned to strata based on their yield class and other attributes, such as location or tenure. Within strata, yield classes are assigned a management objective. For example, the same mixed yield class could be managed for coniferous production in one stratum, and for deciduous or combined production in another. Periodic AACs are calculated for each stratum at 5yr intervals using a variation of the Hanzlik formula (Davis and Johnson, 1987. p.560). This allows for dynamic responses to fires and to changes in forest structure caused by management actions such as stand conversion and natural processes such as succession. It also allows for consistent evaluation of the AAC implications of alternate management strategies, which would not be possible if AACs were pre-determined.

Elaborate *blocking rules* are used to assemble harvestable blocks for each stratum, using strata-specific block size targets and age-ranges. Blocking occurs periodically (by default at 5yr intervals) to simulate the process of Detailed Forest Management Plans. Small blocks are merged with adjacent or nearby large blocks of the same type. This significantly reduces the frequency of small blocks that result from limitations of the blocking algorithm and/or the initial gridding process used to create the input maps. Finally, some versions of FEENIX allow for the creation of un-harvested retention areas within large cutblocks.

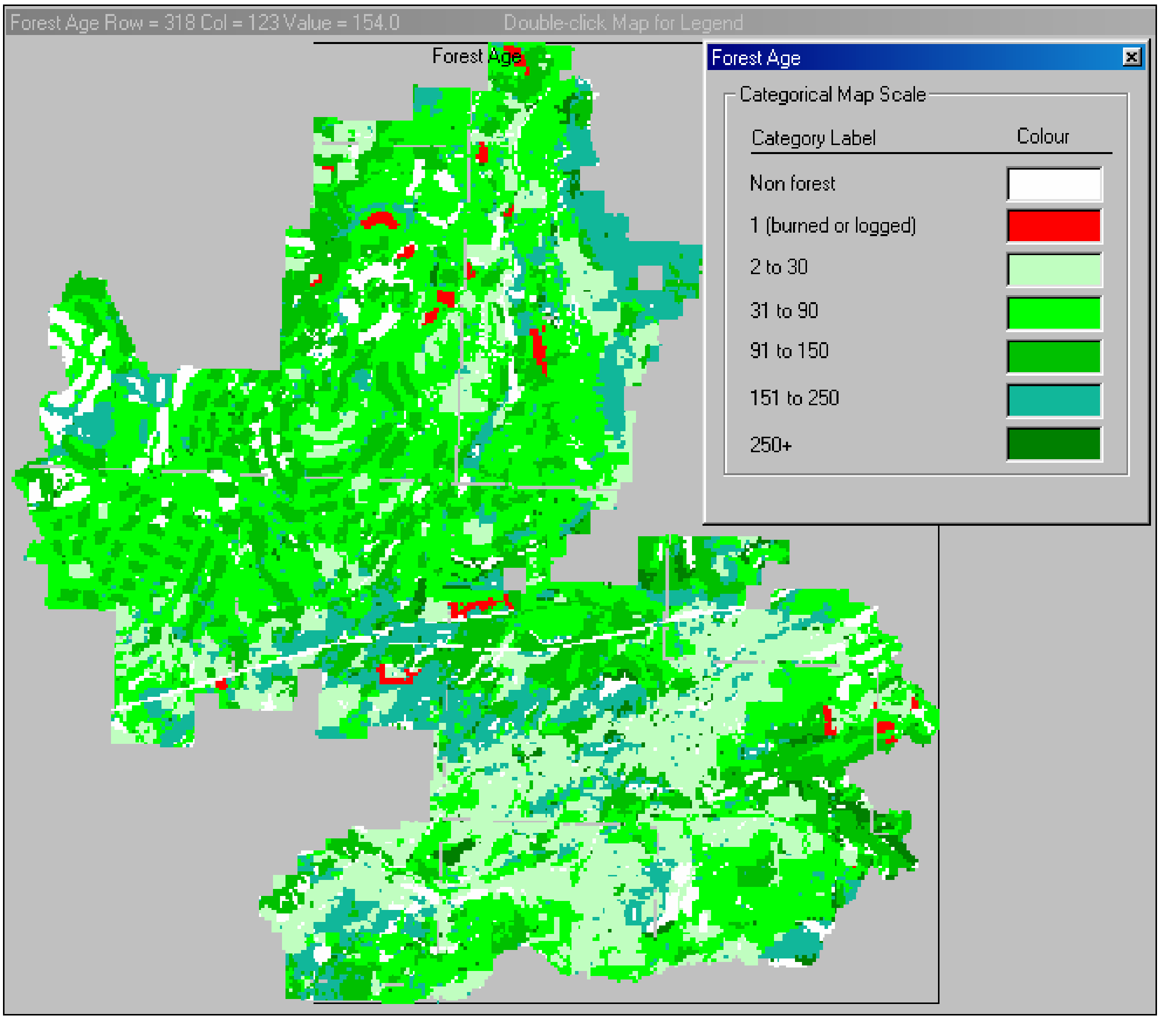


Figure 6. Screen capture of FEENIX blocker and harvest scheduler in action on a small forest estate in the Okanagan.

Blocks are *sequenced* for harvest according to strata-specific rules. Available rules are oldest first, highest volume first, weighted combinations of age and volume, and highest loss. Age, volume and loss are block-level means computed from the age, yield class and net harvestable areas of each cell in the block. Volumes and losses are calculated as m^3 of target volumes / ha. Losses are calculated as the difference between the projected block volume in the next planning period and the present block volume. FEENIX simulates a *reinventory* of the forest at userspecified intervals (20yrs by default). Un-harvested cells that have grown older than their maximum operable age may be assigned a new yield class, landbase or ownership and apparent age, based on the species composition and other characteristics maintained by the stand dynamics sub-model.

The present version of the harvesting model performs well compared to previous versions, and resolves many persistent difficulties related to the stability and sustainability of harvesting during the 2^{nd} or later rotation. Our use of the Hanzlik formula may result in significant fluctuations in AAC during the first rotation. However, experience with simulations conducted on several very different forest estates shows that simulated AACs stabilize during the 2^{nd} rotation. We conjecture that these stable AAC levels closely approximate the long-run non-declining yields that would be calculated from the starting inventory.

REFERENCES

Boyce, M.S. and McDonald, L.L., 1999. Relating populations to habitats using resource selection functions. Trends Ecol. Evol., 14: 268-272.

Boyce, M., P.R. Vernier, S. Nielsen, and F.K.A. Schmiegelow. 2002. Evaluation of Resource Selection Functions. Ecological Modelling 157: 281-300.

Bunnell, F.L., R.W. Wells, J.D. Nelson, and L.L. Kremsater. 1999. Effects of harvest policy on landscape pattern, timber supply and vertebrates in an East Kootenay watershed. Pp. 271-293 *in* J.A. Rochelle, L.A. Lehman, and J. Wisniewski (eds.). Forest fragmentation: wildlife and management implications. Brill, Leiden, Netherlands.

Burnham, K.P., and Anderson, D.R. 1998. Model Selection and Inference. A Practical Information-Theoretic Approach. Springer-Verlag, New York.

Carlson, M.J. and Schmiegelow, F.K.A. 2002. Cost-effective sampling design for broad-scale avian monitoring. Conservation Ecology (*submitted*).

Cumming, S. G., D. A. Demarchi and C. Walters. 1998. A grid-based spatial model of forest dynamics applied to the boreal mixedwood region. SFMN Working Paper 1998-8.

Cumming, S.G. and Vernier, P.R. 2002. Statistical models of landscape pattern metrics, with applications to regional scale dynamic forest simulations. Landscape Ecology 17: 433-444

Cumming, S. 2003. Pilot modelling of tradeoffs between habitat retention, forest management practices and AAC on TFL 49. Project report for Riverside Forest Products Ltd, Armstrong B.C.

Cumming, S. and Wong, C. 2002. Pilot modelling of past fire dynamics in the IDF. Prepared for Lignum Ltd., Williams Lake, BC. http://www.lignum.com/publications-research_papers.asp

McGarigal, K. and Marks, B.J. 1995. Fragstats: Spatial pattern analysis program for quantifying landscape structure. General Technical Report PNW-GTR-351. US Forest Service Pacific Northwest Research Station, Oregon, Portland, USA.

Ecological Stratification Working Group. 1996. A National Ecological Framework for Canada. Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources Research and Environment Canada, State of Environment Directorate, Ottawa/Hull. 125pp. And Map at scale 1:7.5 million. http://www.ec.gc.ca/soer-ree/English/Framework/framework.cfm.

ESRI 2002. ArcView GIS Version 3.3. Environmental Systems Research Institute, Inc. Redlands, California, USA.

Fielding, A.H., Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24, 38-49.

Guisan, A. and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecological Modelling. 135, 147-186.

Hannon, S.J. 1999. Avian response to stand and landscape structure in burned and logged landscapes in Alberta. In T.S. Veeman, D.W. Smith, B.G. Purdy, F.J. Salkie, and G.A. Larkin (eds.) Science and Practice: Sustaining the Boreal Forest. The Sustainable Forest Management Network Conference, Edmonton, Alberta, February 14-17, 1999.

Harrell, F.E. Jr. 2001. Regression Modelling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer-Verlag, NewYork, Inc. New York, NY.

Manel, S., Williams, H.C., Ormerod, S.J. 2001 Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38(5): 921-930.

Manel, S., Dias, J.-M., and Ormerod, S.J. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with Himalayan river bird. Ecological Modelling 120: 337-347.

Manning, Cooper and Associates. 2001. 2001 Northern Goshawk Inventory of Block 4, Canfor TFL 48, Chetwynd, BC. Report prepared for Canadian Forest Products Ltd., Chetwynd Operation, Chetwynd, BC.

Morrison, M.L., Marcot, B.G., and Mannan, R.W. 1999. Wildlife-Habitat Relationships: Concepts and Applications. 2nd Edition. The University of Wisconsin Press. Madison, Wisconsin.

Norton, M.R., Hannon, S.J. and Schmiegelow. F.K.A., 2000. Fragments are not islands: patch vs landscape perspectives on songbird presence and abundance in a harvested boreal forest. Ecography 23:209-223.

Pearce, J. and Ferrier, S. 2000. Evaluating the predictive performance of habitat models using logistic regression. Ecological Modelling 133: 225-245.

Schmiegelow, F.K.A. and Hannon, S.J., 1999. Forest-level effects of fragmentation on boreal songbirds: the Calling Lake Fragmentation Studies. Pages 201-221 in J.A. Rochelle, L.A. Lehmann and J. Wisniewski (eds), Forest Fragmentation: Wildlife and Management Implications. Brill, Leiden.

Schmiegelow, F.K.A., C.S. Machtans, and S.J. Hannon. 1997. Are boreal birds resilient to forest fragmentation? An experimental study of short-term community responses. Ecology 78: 1914-1932.

Scott, J.M., Heglund, P.J., and Morrison, M.L., eds. 2002. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington D.C.

Snowling, S.D., Kramer, J.R., 2001. Evaluating modelling uncertainty for model selection. Ecological Modelling. 138, 17-30.

StataCorp. 2001. Stata Statistical Software: Release 7.0 College Station, Texas: Stata Corporation.

Swets. J.A. 1988. Measuring the accuracy of diagnostic systems. Science 240: 1285-1293.

Vernier, P., Schmiegelow, F.K.A., and Cumming, S.G. 2002. Modelling bird abundance from forest inventory data in the boreal mixedwood forests of Canada. In: Predicting Species Occurrences: Issues of Scale and Accuracy. Island Press, Washington D.C.