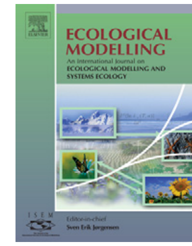
available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ecolmodel



# Sensitivity of species-distribution models to error, bias, and model design: An application to resource selection functions for woodland caribou

*Chris J. Johnson*, Michael P. Gillingham*

*Ecosystem Science and Management Program, University of Northern British Columbia, 3333 University Way, Prince George, British Columbia, V2N 4Z9, Canada*

ARTICLE INFO

*Article history:*
Received 19 June 2007
Received in revised form
14 October 2007
Accepted 27 November 2007
Published on line 15 January 2008

*Keywords:*
GPS
Radiotelemetry
Resource selection function
Sensitivity analysis
Species distribution
Woodland caribou

ABSTRACT

Models that predict distribution are now widely used to understand the patterns and processes of plant and animal occurrence as well as to guide conservation and management of rare or threatened species. Application of these methods has led to corresponding studies evaluating the sensitivity of model performance to requisite data and other factors that may lead to imprecise or false inferences. We expand upon these works by providing a relative measure of the sensitivity of model parameters and prediction to common sources of error, bias, and variability. We used a one-at-a-time sample design and GPS location data for woodland caribou (*Rangifer tarandus caribou*) to assess one common species-distribution model: a resource selection function. Our measures of sensitivity included change in coefficient values, prediction success, and the area of mapped habitats following the systematic introduction of geographic error and bias in occurrence data, thematic misclassification of resource maps, and variation in model design. Results suggested that error, bias and model variation have a large impact on the direct interpretation of coefficients. Prediction success and definition of important habitats were less responsive to the perturbations we introduced to the baseline model. Model coefficients, prediction success, and area of ranked habitats were most sensitive to positional error in species locations followed by sampling bias, misclassification of resources, and variation in model design. We recommend that researchers report, and practitioners consider, levels of error and bias introduced to predictive species-distribution models. Formal sensitivity and uncertainty analyses are the most effective means for evaluating and focusing improvements on input data and considering the range of values possible from imperfect models.



## 1. Introduction

Species-distribution models are becoming an important tool for understanding ecological processes and patterns and for guiding the conservation and management of plants and animals (Raxworthy et al., 2003; Fortin et al., 2005). Once an effective model is identified, results provide a measure of the importance of ecological variables that correlate with species distribution and in some cases abundance (Boyce and McDonald, 1999). Also, model results can be applied to digital spatial data to produce maps representing the likelihood of species occurrence (Carroll et al., 2001). The absolute or rela-

* *Corresponding author*. Tel.: +1 250 960 5357; fax: +1 250 960 5538.
E-mail address: johnsoch@unbc.ca (C.J. Johnson).
0304-3800/$ – see front matter 
doi:10.1016/j.ecolmodel.2007.11.013

tive likelihood of occurrence then serves as a metric to rank habitats for conservation initiatives such as habitat restoration, enhancement or protection (Johnson et al., 2004).

Numerous approaches are available for predicting and mapping species occurrence. Quantitative techniques range from the suite of generalized linear models to rule-based methods (Guisan and Zimmermann, 2000). Although there are many types of distribution models, most are dependent on two sources of data: an unbiased and precise sample of species locations and maps of environmental data that correlate with species distribution. Depending on the species, ecologically plausible variables could represent vegetation, soil parameters, topography, human disturbance, and interspecific interactions (Manly et al., 2002).

Arbitrary decisions during the modeling process, and error and bias in requisite data, can reduce predictive power or lead to incorrect inferences (Elith et al., 2002). A model that poorly reflects actual species–environment relationships will not enlighten our understanding of ecological processes and patterns and might result in misplaced resources or harmful conservation and management actions (Loiselle et al., 2003). Although modellers and practitioners often are aware of potential sources of error, bias, and variation during model construction and use, the impacts are seldom quantified and reported. This is despite the availability of formal methods for conducting sensitivity and uncertainty analyses (Crosetto and Tarantola, 2001). Where case-specific sensitivity and uncertainty analyses are impractical, much guidance can be gleaned from past research. For example, researchers have evaluated and discussed the predictive performance of a number of techniques (Pearce and Ferrier, 2000; Boyce et al., 2002; Loiselle et al., 2003); the sensitivity, uncertainty, and efficacy of expert-based approaches (Dettki et al., 2003; Johnson and Gillingham, 2004); and the lack of ecological theory to support these techniques and their applications (Austin, 2007). Although we have witnessed a recent surge in the use and evaluation of species-distribution models and requisite data, we are unaware of any work that provides a comparison of the relative sensitivity of model predictions to multiple sources of bias and error and variation in model design.

We performed a comprehensive sensitivity analysis for one type of species-distribution model, a resource selection function (RSF) formulated using logistic regression (Manly et al., 2002). Sensitivity analyses provide support for model predictions and highlight areas where assumptions need to be addressed and source data improved or augmented (Crosetto and Tarantola, 2001). With the objective of maintaining realistic ecological relationships we performed the analysis using previously published location data for woodland caribou (*Rangifer tarandus caribou*) and a map of vegetation generated from a classified Landsat Thematic Mapper image (Johnson et al., 2002a, 2003). We measured the sensitivity of model coefficients, prediction success, and maps of selected habitats to four factors: alternate model structures, various levels of bias and error in animal locations, and thematic misclassification of a vegetation map. Resource selection and resource selection probability functions are now ubiquitous in the conservation and ecological literature and, thus, are an excellent case to demonstrate methods for sensitivity analyses of species-distribution models (Arthur et al., 1996; Boyce and McDonald, 1999; Compton et al., 2002; Boyce et al., 2003; Johnson et al., 2004; Fortin et al., 2005).

## 2. Methods

### 2.1. Study area

We developed and assessed RSF models for a population of woodland caribou known as the Wolverine herd located approximately 250 km northwest of Prince George, British Columbia, Canada (Fig. 1, Heard and Vagt, 1998). The study area varies in elevation from valley bottoms at ~900 m to alpine summits at ~2050 m and is characterized by numerous vegetation associations. Forest types below 1100 m elevation are dominated by lodgepole pine (*Pinus contorta*), white spruce (*Picea glauca*), hybrid white spruce (*P. glauca* × *P. engelmannii*), and subalpine fir (*Abies lasiocarpa*). Between 1100 and 1600 m elevation, a moist cold climate prevails with forest types consisting primarily of Engelmann spruce (*P. engelmannii*) and subalpine fir. Areas at elevations >1600 m are alpine tundra and are distinguished by gentle to steep windswept slopes vegetated by shrubs, herbs, bryophytes, and lichens, with occasional trees in krummholz form (Meidinger and Pojar, 1991).

### 2.2. Animal locations

For the sensitivity analyses, we used animal location data collected from 16 individual female caribou of the Wolverine herd monitored between March 1996 and March 1999. Caribou were located with differentially correctable Global Positioning System (GPS) collars scheduled to record one location every third or fourth hour (GPS 1000, Lotek Engineering, Newmarket, Ontario, Canada; Johnson et al., 2002b). For these analyses, we used only GPS locations collected during winter (December 1 to March 31). During this period, monitored caribou were known to demonstrate three coarse-scale selection strategies: foraging across forested habitats, foraging across alpine habitats, or foraging for some period of time within both forested and alpine habitats (Johnson et al., 2002a). Because we wanted to reduce model complexity due to behavioural variation, we restricted our analyses to only those monitored caribou that exclusively occurred across forested habitats. Following screening and the exclusion of out-of-season locations, we retained 2178 GPS fixes for the sensitivity analysis. All locations for the 16 caribou were pooled. We did not control for inconsistent sample sizes across animals; thus, model inference to the population is likely biased to caribou with the greatest relocation frequency.

### 2.3. Modelling approach

A RSF can take many mathematical forms, but is defined as any function that provides predictions of resource use that are proportional to the true probability of use (Manly et al., 2002). We used logistic regression to formulate RSFs that described the selection patterns and predicted the occurrence of female woodland caribou from the Wolverine herd. Logistic regression is commonly used to model species–environment