1

2 Development of a predictive distribution model for Erioderma pedicellatum boreal
3 felt lichen for the island of Newfoundland, Canada

4

5 Yolanda F. Wiersma[1,*], Randolph Skinner[1,2]

6

7 [1]Department of Biology, Memorial University, St. John's, Newfoundland A1B 3X9,
8 Canada

9 [2]Present address: Atlantic Canada Conservation Data Centre, Corner Brook,
10 Newfoundland A2H 751, Canada

11

12 Running Page Head: Wiersma & Skinner: Boreal felt lichen distribution model

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30 *Email: ywiersma@mun.ca

31 ABSTRACT: The worldwide population of the boreal felt lichen Erioderma pedicellatum
32 is currently listed as Critically Endangered by the IUCN, with over 95% of the current
33 population residing on the island of Newfoundland, Canada. Surveys of E. pedicellatum

habitats and populations have primarily been opportunistic, rather than systematic, in nature. We used a geographic information system and compiled occurrence data and pseudo-absence data to develop the first predictive spatial distribution model for E. pedicellatum in Newfoundland. Of the suite of 19 models using 4 different parameters examined, the model with distance from coastline and topographic aspect was the best candidate. The final model had low sensitivity (i.e. a low ability to predict false presence), but high specificity (a strong ability to predict true absence). The final predictive model can contribute to future species status assessments and provincial conservation management decisions that require information on probable species distribution.

KEY WORDS: Habitat · Lichen · Predictive habitat model · Species distribution · Rare species · Generalized additive model · Newfoundland

INTRODUCTION

In conservation biology, determining the spatial distribution of rare species is a challenge. By their nature, rare species offer little natural life history information because of the difficulty of finding, identifying, and maintaining contact with the species long enough to conduct viable studies (Pearson et al. 2007). It can be difficult to find samples of the species within their full range of ecological variability, and thus, predicting their full potential distribution may be difficult. This is the challenge in determining the spatial distribution of the globally rare cyanolichen Erioderma pedicellatum (Hue) P.M. Jørg, known commonly as boreal felt lichen. This species is listed on the International Union for Conservation of Nature (IUCN) Lichen Red List as Critically Endangered worldwide. On the island of Newfoundland, Canada, E. pedicellatum has been listed as a species of special concern by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC), and vulnerable under the Province of Newfoundland and Labrador's Endangered Species Act (Maass & Yetman 2002, Keeping & Hanel 2006).

Erioderma pedicellatum is an epiphytic cyanolichen found in coastal boreal forests with a historical amphi-Atlantic distribution (Keeping & Hanel 2006). Originally discovered in New Brunswick, Canada, in the early 1900s, it has since become extirpated from there (Keeping & Hanel 2006). A detailed survey of Nova Scotia, Canada, has been ongoing for several years, and to date, about 100 individual lichens have been found (Cameron & Neily 2008), although recently a single tree with over 50 individuals was discovered on Cape Breton Island, Nova Scotia (Ayers 2010). In northeastern Europe, it was believed the E. pedicellatum had a range throughout Scandinavia, where it is now believed to have become extirpated (Scheidegger 2003). The exception to this global loss is on the island of Newfoundland, where surveys have discovered in excess of 10000 individual lichens (known as thalli), with more discoveries in each survey year (Keeping & Hanel 2006). Here, E. pedicellatum has been found predominantly on balsam fir Abies balsamea (Keeping & Hanel 2006) and occasionally on yellow birch Betula alleghaniensis and black spruce Picea mariana (Scheidegger 2003), predominantly in wet, coastal forests (Scheidegger 2003). Despite past and ongoing survey efforts, little is

known of the life history of E. pedicellatum (Maass & Yetman 2002); the first demographic model was only recently published (Goudie et al. 2011).

Opportunistic surveys in Newfoundland began over 30 yr ago when Erioderma pedicellatum was first discovered in the province, with concentrated survey efforts beginning in 1998 (Keeping & Hanel 2006). There has been no concerted effort until recently to record absence data. Survey efforts to date have been opportunistic and have focused on gathering abundance data in known hotspots. Furthermore, surveys have been conducted primarily from roadside and trail access routes near known hotspots, in part due to the difficulty of conducting survey work. Newfoundland is largely unpopulated, and accessing areas even a few kilometers from roads and trails can be time consuming and difficult. The terrain is rocky and hilly, with many gulches, ravines, and rivers crisscrossing the landscape. Coastline access is difficult, with fog and wind in the spring, summer, and fall, and ice during the winter. Although the known distribution in Newfoundland is concentrated in 2 locations, on a finer scale, the occurrence of this lichen appears to vary by stand age and in some cases, by species substrate. We thus assume that our samples capture the full range of habitat variability for the species.

During surveys conducted between 1998 and 2008, 2 distinct hyper-populated regions ('hotspots') were mapped in detail (Fig. 1). The Avalon population is centered in the Avalon Peninsula within the Lockyer's Waters region (just northwest of the Avalon Wilderness Reserve in Fig. 1), while another is centered near the Jipujijkuei Kuespem Provincial Park in the Bay d'Espoir region (just west of the Bay du Nord Wilderness Reserve in Fig. 1). The combined populations within these regions comprise nearly 96% of the known individual thalli currently in Newfoundland (24 and 72% of the entire known population, respectively).

Predictive model for boreal felt lichen (Erioderma pedicellatum)

This paper presents the first predictive habitat model for Erioderma pedicellatum on the island of Newfoundland. Based on a review of the existing literature (Maass & Yetman 2002, Scheidegger 2003, Keeping & Hanel 2006, Cameron & Neily 2008, Goudie et al. 2011), and consultation with local experts, we hypothesized that the environmental parameters of aspect, distance from coastline, host tree species, and relative ground/surface moisture content would be important predictors of E. pedicellatum distribution. Two of these factors (tree substrate and distance from coastline) were included in the heuristic model for Nova Scotia (Cameron & Neily 2008). Similar parameters of topography and climate moisture variables were also used as predictors for ecological niche factor analysis (ENFA) models for a range of lichen species in Spain (Martínez et al. 2006).

We hypothesized that Erioderma pedicellatum abundance would decrease from the coastline towards inland areas because of the species' dependence on moist environments (Baldwin & Bradfield 2005). Based on surveys to date which showed that the majority of E. pedicellatum were found on balsam fir, we hypothesized that presence of balsam fir in a stand would be an important predictor of whether the stand was suitable for E. pedicellatum (Lang et al. 1980). We further hypothesized that habitat suitability would be higher within valleys and lower near hills and peaks. Moisture retention and collection at the bottom of valleys and on protected slopes may increase potential habitat

suitability (Rolstad et al. 2001, B. Clarke pers. comm.). Because lichens appear to thrive in damp environments (Brodo et al. 2001), we predicted that suitable E. pedicellatum habitat would more likely be those areas with high topographic convergence and consequently high relative moisture. We also hypothesized that suitable habitat would more likely be on south-facing aspects because these have higher solar insolation for photosynthesis of the photobiont of E. pedicellatum and because higher insolation likely leads to larger trees (and hence more surface area for suitable substrate). Because of Newfoundland's moist climate, we do not feel that south-facing slopes are likely to be too dry to support E. pedicellatum.

In this study, we developed a suite of competing statistical models to explain the spatial distribution of Erioderma pedicellatum on the island of Newfoundland based on the environmental parameters (predictors) described above. We used known locations of lichen thalli together with pseudo-absence data to sample digital map data which were then used in model development. From these, we selected the best model based on statistical analysis and then validated the best model with independent (testing) data.

METHODS

Erioderma pedicellatum data

We used data compiled over 8 yr by the Department of Environment and Conservation Wildlife Division and the Department of Forest Resources and Agrifoods to test hypotheses and build predictive models. We restricted the data to only point count data (which include GPS coordinates and the tree species substrate for each individual lichen thallus for each tree) and discarded data from area counts, which did not supply enough precision on the spatial location of the individual thallus. The finest resolution of the geographic information system (GIS) data layers (described below) was ~80 m, so data were further filtered to excluded any points within 80 m of each other. This left a total of 667 points (mostly clustered in the Bay d'Espoir and Lockyer's Waters regions of the province), from which we randomly chose 10% (n = 67) to reserve as testing data, leaving 600 points for statistical model building (Fielding & Bell 1997).

Because true absence data were lacking, we used pseudo-absence data as a proxy (Wisz & Guisan 2009). Pseudo-absence data have been used in other studies of rare or difficult to find species (e.g.). We created random points across the entire forested part of the island of Newfoundland in ArcGIS 9.3 (ESRI) using Hawth's Tools Random Sampling Toolset (Beyer 2004). We further constrained these random points to forest cover that included balsam fir in the stand (see details on Forest Resource Inventory data below). Mapping of the presence data showed that all known locations of Erioderma pedicellatum were within 20 km of the coast, and so we further restricted the location of pseudo-absence data to forest stands containing balsam fir within 20 km of the coastline. We generated approximately twice as many pseudo-absence points as real presence, since a 2:1 ratio of pseudo-absences to presence data has been suggested to increase the statistical power and to increase the likelihood that the data will give a statistically significant representation of true absence data (Wisz & Guisan 2009). Sample size for pseudo-absence data was n = 1180, from which we randomly selected 10% (n = 118) of these points to reserve for testing data, leaving 1062 pseudo-absence points in the training

data set, for a total presence + pseudo-absence data set of n = 1662 for model building (hereafter referred to as the 'training data').

Environmental data

Forest data. We used the provincial Forest Resource Inventory (FRI; Government of Newfoundland and Labrador, Department of Natural Resources) to determine forest stand structure. FRI data provide a spatial data set for forest stands based on regular surveys and are digitally compiled for use in a GIS (Gillis 2001). The focus of the FRI is on quantifying forest stand productivity for timber harvest, but these types of data have been used in spatial ecology research (Moores et al. 1996, Gillis 2001). The identity of stands was based on an overlay of the training data and the FRI layer in ArcGIS (v. 9.3).

Distance from coastline. We measured proximity of sample points to coastline using the 'Near' tool from the 'Proximity Analysis' toolbox in ArcGIS. We measured distance from each training data point to the nearest section of the coastline, using a coastline layer of the island of Newfoundland.

Aspect. We generated a GIS layer for aspect using the 'Spatial Analysis' extension in ArcGIS and a provincial digital elevation model (DEM) and extracted aspect values at each data point using the 'Extract Values to Points' tool in ArcGIS 9.3 Spatial Analyst. We recalibrated values from a 0–360° to a scale of ±180° to give 'northern' values the same order of magnitude.

Topographic convergence and topographic relative moisture. More nuanced measures of topography that have been shown to correlate with surface moisture include the topographic convergence index (TCI) and topographic relative moisture index (TRMI). TCI calculates the speed of movement and collection of moisture based on terrain (i.e. where water runoff collects), and provides an index of low-to-high moisture collection at the surface. The TRMI model uses the TCI data, but then further refines the TCI parameters to provide relative moisture or wetness levels near the surface beyond the collection areas (Parker 1982). TCI and TRMI were initially developed to determine surface moisture for the Appalachian Mountains, the US Midwest, and the Rocky Mountains (Wilds & van Manen 1995, Wolock & McCabe 1995). These indices have since been used worldwide and are accepted as robust proxies for surface moisture values when 'real' moisture data are unavailable or unreliable (Wolock & McCabe 1995).

We built TCI and TRMI layers for the entire island of Newfoundland in ArcInfo with Arc Macro Language (AML) scripts that were previously developed to calculate TCI and TRMI for the Appalachian mountain range. To determine whether the TCI and TRMI parameters in the original model of Wilds & van Manen (1995) were suited for Newfoundland's varied terrain, or whether reparameterization was necessary, we conducted a sensitivity analysis. Parameters from both models were changed by a magnitude of ±5 and then re-run for the province. For both the TCI and TRMI, all possible parameters were tested for sensitivity. TCI parameters included flow area and flat surface determination. The TRMI parameters included a combined planiform and profile curvature, relative slope potential (RSP), slope, and aspect. Only 1 parameter was changed at a time, and then each model was separately run in ArcInfo. The resulting TCI or TRMI values did not vary considerably from the original values, leading us to

conclude that reparameterization was not necessary and that the model parameters for calculating TCI and TRMI for the Appalachians was appropriate for Newfoundland's terrain. We sampled the TCI/TRMI values with the training data using the 'Extract Values to Points' tool in ArcGIS 9.3 Spatial Analyst.

Statistical analysis

Model building

Data were compiled from the GIS, and then each model predictor was tested for normality. Of the 3 predictors chosen, moisture (both the TCI and TRMI data sets) showed a Poisson distribution. The TCI and TRMI data sets were transformed using the square root of each datum. Aspect and distance from coastline had bimodal distributions. Bimodal distributions are generally dealt with using non-parametric statistical analysis, and thus we chose to use the generalized additive model (GAM), a non-parametric counterpart to the parametric generalized linear model (GLM).

GAM is widely used in biological and ecological model development when data are non-normal and data transformation is not practical (Guisan et al. 2002). The GAM differs from the traditional GLM by replacing the weighted linear regression in the adjusted dependent variable by a weighted backfitting (one that is repetitively replaced) algorithm (Hastie & Tibshirani 1986). This iterative fitting smooths partial residuals by separating the parametric portion of the fit from the non-parametric portion, fitting the parametric portion using weighted least squares within a Gauss-Seidel backfitting algorithm (Hastie & Tibshirani 1986).

The power of the GAM is the ability to choose which predictors use the non-parametric smoother and which ones do not. This allows for a more robust model, as non-parametric (smoothed) data are dealt with differently than the parametric (non-smoothed) data, all within the same model. We used the default parameters for the GAM as outlined in the R Statistical package, version 12.9.0-12.12 .0 (R Development Core Team 2010) with the mgcv library (Hastie 2010). It is recommended that unless warnings are generated in the R program, or if the model fitting fails to converge, the default parameters be used (Hastie 2010).

We had 4 environmental variables (aspect, distance from coastline, species substrate, and moisture), with 5 measurements: species substrate (S), distance from coastline (D), TCI (T1), TRMI (T2), and aspect (A). These were the physical parameters from our original hypothesis. The full model was therefore

$$O = S + D + (T1 \text{ or } T2) + A \qquad (1)$$

where O represents presence–absence occurrences, with 1 = presence or 0 = absence. Because all occurrences in the training data had balsam fir as the tree substrate, the parameter species substrate (S) was omitted from the model. Thus, the new base model was

$$O = D + A + (T1 \text{ or } T2) \qquad (2)$$

From this initial model, we developed 18 other a priori models (Table 1). These models were then subjected to a series of statistical significance tests based on the GAM.

Model selection

We selected the best predictive model as follows. Each of the 19 models was analyzed using the GAM library (with the mgcv library). For each model, the predictors were individually evaluated based on their p values and ranked based on their respective adjusted R^2 values, deviance explained, and the UBRE scores (which is a modified Akaike's Information Criterion value that measures model fitness). We set the final p value for significance testing at p = 0.05. Those models that had all significant predictors were then evaluated on their adjusted R^2 values and their deviance explained values. The adjusted R^2 value provides insight on how much of the model statistically explains the real-life biological and ecological factors, while the deviance explained is a rough guide that contrasts the model with the 'full' model, or a model with all parameters fully fitted. Both a high R^2 value and deviance explained value would fare well in the final model selection; as R^2 and deviance explained values increase, the statistical fit of the model improves.

Models that had all significant predictors and had relatively high R^2 and deviance explained values were then further evaluated using a modified chi-squared analysis of variance (ANOVA) test using the GAM (mgcv) library package (anova.gam). Models that had 2 significant predictors, with a third predictor that was nearly significant (i.e. near, but just over the 0.05 p-value significance), and with relatively high deviance explained and R^2 values were included, just to be sure that all significant values (p values, R^2 values, and deviance explained values) were taken into account. The ANOVA chi-squared test determines the most significant model from the suite of models (see Table 1).

Model validation

We validated the final model using the testing (reserved) data (Fielding & Bell 1997). We predicted the probability of a location being suitable for Erioderma pedicellatum using the best model from the model selection procedure and the environmental data at the testing data locations. The predictive model was generated in R using the GAM (mgcv) library and the 'predict.gam' command and a predictive probability of occurrence calculated based on the exponential link with a logistical transformation to weight each value between 0 and 1. Finally, values were then assigned a binary category (0 or 1) if they were below or above a value of 0.5, respectively. These values were then assigned to a confusion matrix (Fielding & Bell 1997) to assess model fit.

From the confusion matrix, we calculated several indices of model fit. Sensitivity is the measurement of the amount of positive (presence) measurements taken that are actually correct, while specificity relates to the number of negatives (absences) that are correctly identified (Fielding & Bell 1997).

Generating a predictive surface

We developed a predictive surface of Erioderma pedicellatum occurrence across the island (within 20 km of the coastline) in ArcGIS 9.3. Because the GAM does not yield parameter coefficients, we could not simply use map algebra to create a predictive surface. Rather, we created a lattice of data points separated by 1 km using Hawth's

Tool's Sampling Tools and 'Generate Sampling Points' (Beyer 2004). We then sampled these data points for their respective environmental data and used the data to calculate the predicted value with 'prediction.gam' in R, using the exponential link with a logistical transformation. We then imported the predicted values (ranging from 0 to 1) back to the ArcGIS 9.3 layer in the same data set as the 1 km grid data points. We used the 'Spatial Analyst' toolbox to create an inverse distance-weighted interpolation based upon the imported logistic values. The final output was a predictive surface of the island of Newfoundland based upon the best statistical model.

RESULTS

Based on the ANOVA chi-squared test, the final model selected was Model 3 (Table 1), which was: O = sD + sA, where O is occurrence, s represents the smooth (GAM) function, D is distance from coastline, and A is the aspect. The next closest model (Model 12: O = sD + sA + sTCI) also had all significant predictors, and had the best R^2 values and deviance explained values (Table 1). However, the chi-squared ANOVA test suggested that Model 3 was more significant than Model 12, consistent with the principle of parsimony. Erioderma pedicellatum occurrence showed a non-linear relationship with distance from coastline (Fig. 2), with highly positive correlation at a distance of approximately 2 and 18 km from the coast, and a low correlation at a distance of approximately 5 km. E. pedicellatum was most abundant at sites with northern or southern aspects and least abundant in flat areas (Fig. 3). E. pedicellatum occurrence was negatively correlated with intermediate values of TCI and positively correlated with extreme values (i.e. tops of hills and valley bottoms; Fig. 4) and did not show any pattern with TRMI (Fig. 5).

Model validation resulted in a confusion matrix (Table 2) and a range of accuracy measures (Table 3). Sensitivity and specificity outlines the accuracy of the presence and absence data, respectively. Model sensitivity was 14.93%, and specificity was 71.79%. Overall diagnostic power was at 63.59%, the correct classification rate was at 51.09%, and Kappa was very low (9.24%).

The predictive surface for the entire island is shown in Fig. 6. Detailed maps for the prediction surface from each of the 2 areas of the province where intensive survey work for Erioderma pedicellatum have occurred are shown in Fig. 7 (Avalon) and Fig. 8 (Bay d'Espoir). Fig. 6 represents the predictive model surface as applied to the island based on the abiotic parameters, and does not restrict the suitable habitats to only those that overlap with balsam fir stands, because at the scale of the whole island, the resolution is insufficient to distinguish balsam fir stands. However, within the pixels with high probability values for habitat suitability in Fig. 6, only those that overlap with stands of balsam fir would be considered probable E. pedicellatum habitat.

DISCUSSION

The final predictive model was the best of 18 possible a priori models. The model parameters for all candidate models were based on field research spanning more than a decade, and were picked because they were consistent both in the database and from expert opinion. Because this is an initial predictive model for Erioderma pedicellatum, the initial parameters met 2 criteria: they could (1) be easily obtained from existing data

sets and (2) be translated to a GIS software program. The best of all candidate models, given the data, contained only the parameters of aspect and distance from coastline.

For comparison, the heuristic model developed for Erioderma pedicellatum in Nova Scotia (Cameron & Neily 2008) used distance from coastline (less than 30 km), and balsam fir within 80 m of sphagnum wetlands as their primary biological and geographical predictors using a filtering query developed in a GIS. In this paper, we investigated a suite of possible predictors statistically, and then applied the statistically best fitted model via GIS. In both our model and the Nova Scotia model, distance from coastline was a strong predictor. This convergence of an environmental parameter between the 2 types of models in 2 different provinces suggests that this is an important factor in predicting E. pedicellatum occurrence. It is most likely a proxy value for moisture or micro-climate; further field work should investigate atmospheric moisture and microclimate conditions along a gradient from the coast inland to confirm this hypothesis and further refine our understanding of E. pedicellatum habitat requirements.

We did not include distance from sphagnum wetlands as a possible predictor to compare to the Nova Scotia model (Cameron & Neily 2008) because the geospatial data to describe wetlands in the Newfoundland FRI were not detailed enough to distinguish sphagnum wetlands from other wetland types. Should high quality spatial data on wetlands become available, it would be interesting to see if the statistical model corroborates the heuristic one for that parameter. The Nova Scotia model did not include aspect, which was included in our best model.

Because aspect was non-parametric, and the GAM does not provide parameter coefficients, we could not statistically determine which direction was most statistically significant for predicting occurrence of Erioderma pedicellatum. However, inspection of the distribution of thalli against aspect (Fig. 3) shows an approximate preference for sloped terrain over flat terrain, as well as a slight preference for northerly slopes over southerly slopes. This is consistent with the investigation of 20 sites with confirmed E. pedicellatum in Nova Scotia, which showed that E. pedicellatum preferred aspects of 0 to 90° and sites with depressions or low slopes (Cameron & Neily 2008). The preference of E. pedicellatum habitat for sloped rather than flat terrain may be in part to limit the amount of direct sunlight reaching inside the tree canopy and the lichen communities therein (Gauslaa et al. 2001). A preference of northern slopes over southern slopes may also be a result of the amount of ambient sunlight filtering through the forest canopy during the day (Campbell & Coxson 2001, Hylander 2005). Alternately, it may be a preference for greater near-surface moisture retention from a northern aspect versus a southern aspect or it may simply be a case of substrate dependence, as north-facing balsam fir trees will have a slower growth life cycle than southern-facing trees. Goudie et al. (2011) suggested that the retardation of balsam fir growth during its full life cycle may give more opportunities for the microhabitat substrate to retain features that are conducive to a more complete boreal felt lichen life cycle. Further research is needed to confirm the relationship between rate of balsam fir growth and E. pedicellatum substrate preferences.

The predictor 'distance from coastline' showed a bimodal pattern, with occurrence predicted to be highest near the coastline (approximately 1 to 3 km) and at a distance of approximately 16 to 18 km. This may be due to several biotic and abiotic

gradients. First, the population density of the preferred substrate (balsam fir) also fits the pattern shown in the boreal felt lichen population densities and distance from coastline. Sudden or abrupt changes in the Damman Forest Types from the coastline to the interior may explain this phenomenon (Meades & Moores 1989). Furthermore, exposure to direct coastal climatic conditions during all 4 seasons (salt water, ice, and higher winds) may be detrimental to boreal felt lichen habitat suitability directly at the coastline (Werth et al. 2005). Balsam fir that grow closest to the shoreline may exhibit stunted growth, or be exposed to the oceanic elements year-round.

While the model containing aspect and distance from coastline was the most statistically significant model, model validation showed that the final model was not particularly robust. It had a sensitivity of only 14.9% and specificity of 71.8% and a Kappa reading of only 9.2% (Table 3). The low sensitivity indicates that the model is not predicting the presence of Erioderma pedicellatum without the risk of a Type I error (Fielding & Bell 1997). Cameron & Neily's (2008) heuristic model also had low predictive power; they found E. pedicellatum in 7% of the stands predicted to be suitable. Thus the low model sensitivity here may be due to the lichen's overall rarity, and not to the fact that suitable habitat is being poorly predicted. It may also be possible that E. pedicellatum is a metapopulation (Cameron & Neily 2008) and thus not all suitable habitats may contain lichens at all times. Despite the low power to predict presence, the relatively high specificity means that the model is giving a good percentage of correct predictions for the absence data (that is, the model is relatively good at predicting what areas are not suitable E. pedicellatum habitat). This is very useful from a forest management perspective in that it indicates areas where forest harvest or other anthropogenic disturbance may be approved without the need for extensive survey work for E. pedicellatum first, and with minimal risk of destroying lichen habitat. Kappa is very low (9.24%), which indicates that the model has overall poor agreement with the criteria set out by Fielding & Bell (1997) for model fitness. The very low Kappa reading can be explained by the disparity of data used for the presence and pseudo-absence. Fielding & Bell (1997) indicated that when 1 category of data outnumbers another by a significant amount, the Kappa reading becomes less reliable. In our study, pseudo-absence data outnumbered presence data by a ratio of 2:1, but we felt this was necessary to ensure better overall model fitness (Wisz & Guisan 2009). The underlying poor value for sensitivity and overall diagnostic power (63.9%) and correct classification rate (51.9%) are indicators that the model needs more refinement to improve overall predictive fitness.

A key reason that our model had such poor overall predictive power is because the majority of Erioderma pedicellatum data used to build the models came from 2 hyper-densely populated regions, and therefore the models will be biased to the environmental parameters present at those sites. Thus it seems possible that these 2 areas have a unique combination of features that make them suitable for E. pedicellatum. Although this gives the impression that these are the only areas where E. pedicellatum occurs on the entire island, the species has been found in other parts of the island (but to date at lower densities), including in Terra Nova National Park (K. Tulk pers. comm.) and in Bay du Nord, just north of the Bay d'Espoir area, the northern Peninsula, along the Burin Peninsula and Burgeo highways, and in other parts of the Avalon Peninsula (Fig. 1). Therefore, it is more likely that the uneven spatial distribution of training data was the

cause of the weak predictive power of the model. Future sampling work should attempt to more systematically sample lichen abundance across the island to generate a more even spatial distribution of data.

A further challenge in model development was defining surface moisture at a province-wide extent and yet a fine-enough resolution to correspond to stand size. Field researchers widely believe that airborne humidity and moisture levels could be among the most important predictors for Erioderma pedicellatum occurrence. The challenge will be to find some form of measuring or estimating moisture at a high enough resolution to correspond to E. pedicellatum distribution. Weather station data are too coarse, and while the TCI and TRMI have been shown to correlate well with moisture in other studies (Parker 1982, Wolock & McCabe 1995), they were not significant predictors in our best model (although TCI was included in the second-best model). TCI and TRMI measure surface moisture based on predicted patterns of runoff in response to terrain, but in Newfoundland, atmospheric moisture in the form of fog may be more significant to E. pedicellatum occurrence. Thus, data loggers that measure surface and atmospheric moisture at microclimate scales might be a useful tool. It would be logistically challenging to deploy these across the island, but perhaps a future study could measure moisture (and other environmental parameters) at a smaller spatial extent. Findings from a field experiment to compare these parameters in forest stands with and without E. pedicellatum could be used to further refine the predictive model.

There may be other statistically significant predictors of Erioderma pedicellatum population distribution which were not included in our study. The predictors we chose were based upon expert advice from years of field work in the province of Newfoundland and Labrador. ENFA models for lichens in Spain suggest that environmental factors such as precipitation, winter precipitation, altitudinal range, and maximum and mean altitude are important predictors of lichen distribution (Martínez et al. 2006), and these may be worth examining for E. pedicellatum in Newfoundland. There is the possibility that other environmental or habitat predictors have been overlooked because they are difficult to collect, or data are lacking, or they are too complex to articulate as a single predictor in a GIS environment. Several potential predictors such as temperature, tree age, adjacency to bogs and fens, and the inclusion of other lichens in the nearby ecological community might be likely predictors of E. pedicellatum occurrence. The demographic model of Goudie et al. (2011) suggests that stand characteristics are important contributors to population viability, and thus future modeling that investigates stand age, crown closure, and tree diameter may improve the predictive power of distribution models.

Finally, the lack of true absence data may have had an impact on modeling success. Pseudo-absence data have been used when true absence data have been lacking in several case studies (Wisz & Guisan 2009). Pseudo-absence data have also been used for rare and hard to find species, but they are not always ideal. The provincial government is now including a protocol for recording absence data, so when sample size is sufficiently large (and hopefully more evenly spatially distributed across the island), we will be able to build predictive models using true absence data.

CONCLUSION

Developing a predictive model based on limited life history information for a species that is hard to find and rare is always challenging. We developed the first predictive habitat model for boreal felt lichen Erioderma pedicellatum on the island of Newfoundland, Canada. While our model yielded some insights into abiotic parameters that appear to be important for this species (aspect, distance from coastline), our final model was only good for predicting habitat that was unlikely to contain E. pedicellatum, and not very reliable at predicting sites where it might occur (although given the species' overall rarity, this does not discount that these might be suitable habitat areas for future colonization). This may be due to the limitations of either the survey data or the input data, or both. Rare species are challenging to model because of limited data (but see Glavich et al. 2005 for an example of a logistic model for epiphytic lichen in the Pacific Northwest). Despite the limitations of our models, our work does provide the first predictive statistical model ever for this rare species, and is a refinement of (and consistent with aspects of) the previous heuristic model. Through the process of model development and statistical testing, we have been able to identify areas for future research that further our understanding of the rare boreal felt lichen E. pedicellatum and that will hopefully contribute to management of its habitat on the island of Newfoundland.

LITERATURE CITED

<jrn>Ayers T (2010) Rare, endangered lichens found on super tree in Richmond County. Cape Breton Post, 14 January. Available at www.capebretonpost.com/Business/Natural-resources/2010-01-14/article-771762/Rare,-endangered-lichens-found-on-super-tree-in-Richmond-County/1 </jrn>

<jrn>Baldwin LK, Bradfield GE (2005) Bryophyte community differences between edge and interior environments in temperate rain-forest fragments of coastal British Columbia. Can J For Res 35:580–592 doi:10.1139/x04-209</jrn>

<eref>Beyer H (2004) Hawth's analysis tools for ArcGIS. Available at www.spatialecology.com/htools (accessed June 2010)</eref>

<bok>Brodo IM, Sharnoff SD, Sharnoff S (2001) Lichens of North America. Yale University Press, New Haven, CT</bok>

<jrn>Cameron RP, Neily T (2008) Heuristic model for identifying the habitats of Erioderma pedicellatum and other rare cyanolichens in Nova Scotia, Canada. Bryologist 111:650–658 doi:10.1639/0007-2745-111.4.650</jrn>

<jrn>Campbell J, Coxson D (2001) Canopy microclimate and arboreal lichen loading in subalpine spruce–fir forest. Can J Bot 79:537–555</jrn>

<jrn>Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24:38–49 doi:10.1017/S0376892997000088</jrn>

<jrn>Gauslaa Y, Ohlson M, Solhaug KA, Bilger W, Nybakken L (2001) Aspect-dependent high-irradiance damage in two transplanted foliose forest lichens, Lobaria pulmonaria and Parmelia sulcata. Can J For Res 31:1639–1649</jrn>

<jrn>Gillis MD (2001) Canada's national forest inventory (responding to current information needs). Environ Monit Assess 67:121–129 PubMed doi:10.1023/A:1006405820244</jrn>

<jrn>Glavich DA, Geiser LH, Mikulin AG (2005) Rare epiphytic coastal lichen habitats, modeling, and management in the Pacific Northwest. Bryologist 108:377–390 doi:10.1639/0007-2745(2005)108[0377:RECLHM]2.0.CO;2</jrn>

<jrn>Goudie RI, Scheidegger C, Hanel C, Munier A, Conway E (2011) New population models help explain declines in the globally rare boreal felt lichen Erioderma pedicellatum in Newfoundland. Endang Species Res 13:181–189 doi:10.3354/esr00326</jrn>

<eref>Government of Newfoundland and Labrador, Department of Environment and Conservation (2010) Species at risk. Available at www.env.gov.nl.ca/env/wildlife/endangeredspecies/index.html (accessed on 29 May 2010) </eref>

<jrn>Guisan A, Edwards T, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol Model 157:89–100 doi:10.1016/S0304-3800(02)00204-1</jrn>

<eref>Hastie T (2010) Package 'gam': generalized additive models. Available at http://cran.r-project.org/web/packages/gam/gam.pdf</eref>

<jrn>Hastie T, Tibshirani R (1986) Generalized additive models. Stat Sci 1:297–318 doi:10.1214/ss/1177013604</jrn>

<jrn>Hylander C (2005) Aspect modifies the magnitude of edge effect on bryophyte growth in boreal forests. J Appl Ecol 42:518–525 doi:10.1111/j.1365-2664.2005.01033.x</jrn>

<bok>Keeping B, Hanel C (2006) A five year (2005–2011) management plan for the boreal felt lichen (Erioderma pedicellatum) in Newfoundland and Labrador. Government of Newfoundland and Labrador, </bok>

<jrn>Lang W, Reiners A, Pike LH (1980) Structure and biomass dynamics of epiphytic lichen communities of balsam fir forests in New Hampshire. Ecology 61:541–550 doi:10.2307/1937420</jrn>

<bok>Maass WSG, Yetman D (2002) COSEWIC assessment and status report on the boreal felt lichen, Erioderma pedicellatum, in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa, ON</bok>

<jrn>Martínez I, Carreño F, Escudero A, Rubio A (2006) Are threatened lichen species well-protected in Spain? Effectiveness of a protected areas network. Biol Conserv 133:500–511 doi:10.1016/j.biocon.2006.08.003</jrn>

<bok>Meades WJ, Moores L (1989) Forest site classification manual. A field guide to the Damman Forest Types of Newfoundland, 1st edn. Newfoundland Department of Forestry and Agriculture, Forestry Canada, </bok>

<jrn>Moores L, Pittman B, Kitchen G (1996) Forest ecological classification and mapping: their application in ecosystem management in Newfoundland. Environ Monit Assess 39:571–577 doi:10.1007/BF00396169</jrn>

<jrn>Parker AJ (1982) The topographic relative moisture index: an approach to soil-moisture assessment in mountain terrain. Phys Geogr 3:160–168</jrn>

<jrn>Pearson RG, Raxworhty CJ, Nakamura M, Peterson AT (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J Biogeogr 34:102–117 doi:10.1111/j.1365-2699.2006.01594.x</jrn>

<eref>R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available at www.R-project.org/</eref>

<jrn>Rolstad J, Gjerde I, Storaunet KO, Rolstad E (2001) Epiphytic lichens in Norwegian coastal spruce forests: historic logging and present forest structure. Ecol Appl 11:421–436 doi:10.1890/1051-0761(2001)011[0421:ELINCS]2.0.CO;2</jrn>

<eref>Scheidegger C (2003) Erioderma pedicellatum. In: 2008 IUCN red list of threatened species. Available at www.iucnredlist.org (accessed on 22 November 2008) </eref>

<jrn>Werth S, Tommervik H, Elvebakk A (2005) Epiphetic macrolichen communities along regional gradients in northern Norway. J Veg Sci 16:199–208 doi:10.1111/j.1654-1103.2005.tb02356.x</jrn>

<bok>Wilds S, van Manen F (1995) Arc markup language script for the topographic convergence index. US Geological Survey, </bok>

<jrn>Wisz MS, Guisan A (2009) Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecol 9:8 PubMed doi:10.1186/1472-6785-9-8 doi:10.1186/1472-6785-9-8</jrn>

<jrn>Wolock DM, McCabe GJ (1995) Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. Water Resour Res 31:1315–1324 doi:10.1029/95WR00471</jrn>

Fig. 1. Newfoundland, Canada, showing boreal felt lichen Erioderma pedicellatum sites in red (source: Government of Newfoundland and Labrador 2006, Keeping & Hanel 2006)

Fig. 2. Erioderma pedicellatum. Distance from coastline versus partial residuals of modeled presence/absence of boreal felt lichen on the y-axis, using the generalized additive model. Dotted lines indicate the 95% confidence interval; solid line represents the distance from coastline partial residual

Fig. 3. Erioderma pedicellatum. Occurrences (from filtered data) in relation to topographic aspect

Fig. 4. Erioderma pedicellatum. Actual topographic convergence index (TCI) values with partial residuals of modeled presence/absence of boreal felt lichen plotted on the y-axis, using the generalized additive model. The solid lines are the partial residuals; the dashed lines represent the 95% confidence interval

Fig. 5. Erioderma pedicellatum. Actual topographic relative moisture index (TRMI) values with partial residuals of modeled presence/absence of boreal felt lichen (solid lines) plotted on the y-axis and the 95% confidence intervals (dashed lines), using the generalized additive model

Fig. 6. Erioderma pedicellatum. Final island-wide predictive surface interpolation for Newfoundland, Canada. Dark red (bright green) represents the lowest (highest) probability of suitable habitat. Numeric values represent the statistical predictor values for the model. Blue data points represent currently known populations that were used for this study. Grey areas denote regions where the probability of occurrence was not interpolated. Balsam fir stands are not represented, as the resolution was too low to be shown effectively. For maps that include the balsam fir stands, see Figs. 7 & 8

Fig. 7. Erioderma pedicellatum. Final predictive surface interpolation, focusing on the Avalon region shown only within balsam fir stands. The brighter green (darker red) regions and higher (lower) numbers represent more (less) suitable predicted habitat. Grid cells are at 500 m due to computer memory limitations

Fig. 8. Erioderma pedicellatum. Final predictive surface interpolation, focusing on the Bay d'Espoir region shown only within balsam fir stands. Other details as in Fig. 7

Table 1. Erioderma pedicellatum. Potential models for Newfoundland, Canada, and their respective statistical analyses, including when applicable, chi squared, p values, Z values R^2 adjusted, and deviance explained. O: occupancy, D: distance from coastline, A: topographic aspect, T1: topographic convergence index, T2: topographic relative moisture index, s: smoothing function used in the generalized additive model, na: not applicable

Model	Equation	χ^2	p	Z	Deviance explained	R^2 adjusted
1	O = sD + sA + T1	380.06	2.00×10^{-16}	na		
		16.79	0.36	na		
		na	0.092	1.68	32.30	0.37

#	Model					
2	O = sD + T1	385.40	2.00 × 10⁻¹⁶	na		
		na	0.070	1.81	31.30	0.36
3	O = sA + sD	380.85	2.00 × 10⁻¹⁶	na		
		16.93	0.033	na	32.20	0.37
4	O = sA + T1	22.10	0.0065	na		
		na	0.20	1.29	1.23	0.010
5	O = sD	386.10	2.00 × 10⁻¹⁶	na	31.20	0.37
6	O = sA	20.15	0.017	na	1.13	0.001
7	O = T1	na	0.74	0.33	0.01	−0.00050
8	O = sD + sA + T2	380.91	2.00 × 10⁻¹⁶	na		
		19.93	0.012	na		
		na	0.072	1.80	32.3	0.37
9	O = sD + T2	386.80	2.00 × 10⁻¹⁶	na		
		na	0.25	1.15	31.2	0.37
10	O = sA + T2	20.22	0.012	na		
		na	0.53	-0.30	1.14	0.0095
11	O = T2	na	0.76	-0.30	0.01	−0.0005
12	O = sD + sA + sT1	330.30	2.00 × 10⁻¹⁶	na		
		18.31	0.021	na		
		29.06	0.0061	na	33.70	0.39
13	O = sD + sT1	332.66	2.00 × 10⁻¹⁶	na		
		28.68	0.00071	na	32.70	0.3770
14	O = sA + sT1	17.12	0.030	na		
		111.76	2.00 × 10⁻¹⁶	na	6.84	0.078
15	O = sT1	115.50	2.00 × 10⁻¹⁶	na	5.87	0.070
16	O = sD + sA + sT2	381.10	2.00 × 10⁻¹⁶	na		
		20.05	0.012	na		
		3.37	0.066	na	32.40	0.37
17	O = sD + sT2	386.82	2.00 × 10⁻¹⁶	na		
		1.30	0.26	na	31.2	0.37
18	O = sA + sT2	20.20	0.012	na		
		0.37	0.55	na	1.14	0.0095
19	O = sT2	0.10	0.75	na	0.01	−0.00050

1

2 Table 2. Confusion matrix outlining the predicted and actual positive and negative values
3 for Model 3

		Actual	
		+	−
Predicted	+	10	33
	−	57	84
	SUM	67	117

4 Table 3. Accuracy measures based on the confusion matrix in Table 2

Accuracy measure	%
Prevalence	36.41
Overall diagnostic power	63.59
Correct classification rate	51.09

Sensitivity	14.93
Specificity	71.79
Kappa	9.24

1