

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

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

Yolanda F. Wiersma^{1,*}, Randolph Skinner^{1,2}

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

²*Present address:* Atlantic Canada Conservation Data Centre, Corner Brook, Newfoundland A2H 751, Canada

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

*Email: ywiersma@mun.ca

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

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

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

14 INTRODUCTION

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

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

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

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

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

25 **Predictive model for boreal felt lichen (*Erioderma pedicellatum*)**

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

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

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

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

16 METHODS

17 *Erioderma pedicellatum* data

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

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

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

3 **Environmental data**

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

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

15 **Aspect.** We generated a GIS layer for aspect using the ‘Spatial Analysis’ extension in
16 ArcGIS and a provincial digital elevation model (DEM) and extracted aspect values at
17 each data point using the ‘Extract Values to Points’ tool in ArcGIS 9.3 Spatial Analyst.
18 We recalibrated values from a $0-360^\circ$ to a scale of $\pm 180^\circ$ to give ‘northern’ values the
19 same order of magnitude.

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

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

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

5 **Statistical analysis**

6 **Model building**

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

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

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

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

$$34 \quad O = S + D + (T1 \text{ or } T2) + A \quad (1)$$

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

$$39 \quad O = D + A + (T1 \text{ or } T2) \quad (2)$$

40 From this initial model, we developed 18 other *a priori* models (Table 1). These models
41 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

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

9 RESULTS

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

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

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

38 DISCUSSION

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

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

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

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

22 Because aspect was non-parametric, and the GAM does not provide parameter
23 coefficients, we could not statistically determine which direction was most statistically
24 significant for predicting occurrence of *Erioderma pedicellatum*. However, inspection of
25 the distribution of thalli against aspect (Fig. 3) shows an approximate preference for
26 sloped terrain over flat terrain, as well as a slight preference for northerly slopes over
27 southerly slopes. This is consistent with the investigation of 20 sites with confirmed *E.*
28 *pedicellatum* in Nova Scotia, which showed that *E. pedicellatum* preferred aspects of 0 to
29 90° and sites with depressions or low slopes (Cameron & Neily 2008). The preference of
30 *E. pedicellatum* habitat for sloped rather than flat terrain may be in part to limit the
31 amount of direct sunlight reaching inside the tree canopy and the lichen communities
32 therein (Gauslaa et al. 2001). A preference of northern slopes over southern slopes may
33 also be a result of the amount of ambient sunlight filtering through the forest canopy
34 during the day (Campbell & Coxson 2001, Hylander 2005). Alternately, it may be a
35 preference for greater near-surface moisture retention from a northern aspect versus a
36 southern aspect or it may simply be a case of substrate dependence, as north-facing
37 balsam fir trees will have a slower growth life cycle than southern-facing trees. Goudie et
38 al. (2011) suggested that the retardation of balsam fir growth during its full life cycle may
39 give more opportunities for the microhabitat substrate to retain features that are
40 conducive to a more complete boreal felt lichen life cycle. Further research is needed to
41 confirm the relationship between rate of balsam fir growth and *E. pedicellatum* substrate
42 preferences.

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

1 gradients. First, the population density of the preferred substrate (balsam fir) also fits the
2 pattern shown in the boreal felt lichen population densities and distance from coastline.
3 Sudden or abrupt changes in the Damman Forest Types from the coastline to the interior
4 may explain this phenomenon (Meades & Moores 1989). Furthermore, exposure to direct
5 coastal climatic conditions during all 4 seasons (salt water, ice, and higher winds) may be
6 detrimental to boreal felt lichen habitat suitability directly at the coastline (Werth et al.
7 2005). Balsam fir that grow closest to the shoreline may exhibit stunted growth, or be
8 exposed to the oceanic elements year-round.

9 While the model containing aspect and distance from coastline was the most
10 statistically significant model, model validation showed that the final model was not
11 particularly robust. It had a sensitivity of only 14.9% and specificity of 71.8% and a
12 Kappa reading of only 9.2% (Table 3). The low sensitivity indicates that the model is not
13 predicting the presence of *Erioderma pedicellatum* without the risk of a Type I error
14 (Fielding & Bell 1997). Cameron & Neily's (2008) heuristic model also had low
15 predictive power; they found *E. pedicellatum* in 7% of the stands predicted to be suitable.
16 Thus the low model sensitivity here may be due to the lichen's overall rarity, and not to
17 the fact that suitable habitat is being poorly predicted. It may also be possible that *E.*
18 *pedicellatum* is a metapopulation (Cameron & Neily 2008) and thus not all suitable
19 habitats may contain lichens at all times. Despite the low power to predict presence, the
20 relatively high specificity means that the model is giving a good percentage of correct
21 predictions for the absence data (that is, the model is relatively good at predicting what
22 areas are not suitable *E. pedicellatum* habitat). This is very useful from a forest
23 management perspective in that it indicates areas where forest harvest or other
24 anthropogenic disturbance may be approved without the need for extensive survey work
25 for *E. pedicellatum* first, and with minimal risk of destroying lichen habitat. Kappa is
26 very low (9.24%), which indicates that the model has overall poor agreement with the
27 criteria set out by Fielding & Bell (1997) for model fitness. The very low Kappa reading
28 can be explained by the disparity of data used for the presence and pseudo-absence.
29 Fielding & Bell (1997) indicated that when 1 category of data outnumbers another by a
30 significant amount, the Kappa reading becomes less reliable. In our study, pseudo-
31 absence data outnumbered presence data by a ratio of 2:1, but we felt this was necessary
32 to ensure better overall model fitness (Wisz & Guisan 2009). The underlying poor value
33 for sensitivity and overall diagnostic power (63.9%) and correct classification rate
34 (51.9%) are indicators that the model needs more refinement to improve overall
35 predictive fitness.

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

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

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

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

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

43 CONCLUSION

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

18
19 *Acknowledgements.* We appreciate the time and conversations with B. Clarke, J.
20 McCarthy, B. Roberts, and C. Hanel, who shared their expert opinions and insights on
21 boreal felt lichen. Thanks to C. Marks and J. Luther for their technical expertise and
22 advice to R.S. regarding GIS, and to C. Doucet for providing R.S. with office space. C.
23 Doucet, M. Krishnapillai, L. Hermanutz, M. Piercey-Normore, and 2 anonymous
24 reviewers provided helpful comments on an earlier version of the manuscript. R.S. was
25 funded by Memorial University's School of Graduate Studies, the Newfoundland and
26 Labrador Department of the Environment Wildlife Division, and Y.F.W. by a GEOIDE
27 grant. Infrastructure to complete the research was provided through Canadian Foundation
28 for Innovation grant to Y.F.W. and through in-kind support from the Wildlife Division
29 and the Grenfell Campus of Memorial University.

30 LITERATURE CITED

- 31 <jrn>Ayers T (2010) Rare, endangered lichens found on super tree in Richmond County.
32 Cape Breton Post, 14 January. Available at
33 [www.capebretonpost.com/Business/Natural-resources/2010-01-14/article-](http://www.capebretonpost.com/Business/Natural-resources/2010-01-14/article-771762/Rare,-endangered-lichens-found-on-super-tree-in-Richmond-County/1)
34 [771762/Rare,-endangered-lichens-found-on-super-tree-in-Richmond-County/1](http://www.capebretonpost.com/Business/Natural-resources/2010-01-14/article-771762/Rare,-endangered-lichens-found-on-super-tree-in-Richmond-County/1) </jrn>
- 35 <jrn>Baldwin LK, Bradfield GE (2005) Bryophyte community differences between edge
36 and interior environments in temperate rain-forest fragments of coastal British
37 Columbia. *Can J For Res* 35:580–592 [doi:10.1139/x04-209](https://doi.org/10.1139/x04-209)</jrn>
- 38 <eref>Beyer H (2004) Hawth's analysis tools for ArcGIS. Available at
39 www.spatial ecology.com/htools (accessed June 2010)</eref>
- 40 <bok>Brodo IM, Sharnoff SD, Sharnoff S (2001) Lichens of North America. Yale
41 University Press, New Haven, CT</bok>

- 1 <jrn>Cameron RP, Neily T (2008) Heuristic model for identifying the habitats of
2 *Erioderma pedicellatum* and other rare cyanolichens in Nova Scotia, Canada.
3 *Bryologist* 111:650–658 [doi:10.1639/0007-2745-111.4.650](https://doi.org/10.1639/0007-2745-111.4.650)</jrn>
- 4 <jrn>Campbell J, Coxson D (2001) Canopy microclimate and arboreal lichen loading in
5 subalpine spruce–fir forest. *Can J Bot* 79:537–555</jrn>
- 6 <jrn>Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction
7 errors in conservation presence/absence models. *Environ Conserv* 24:38–49
8 [doi:10.1017/S0376892997000088](https://doi.org/10.1017/S0376892997000088)</jrn>
- 9 <jrn>Gauslaa Y, Ohlson M, Solhaug KA, Bilger W, Nybakken L (2001) Aspect-
10 dependent high-irradiance damage in two transplanted foliose forest lichens, *Lobaria*
11 *pulmonaria* and *Parmelia sulcata*. *Can J For Res* 31:1639–1649</jrn>
- 12 <jrn>Gillis MD (2001) Canada’s national forest inventory (responding to current
13 information needs). *Environ Monit Assess* 67:121–129 [PubMed](#)
14 [doi:10.1023/A:1006405820244](https://doi.org/10.1023/A:1006405820244)</jrn>
- 15 <jrn>Glavich DA, Geiser LH, Mikulin AG (2005) Rare epiphytic coastal lichen habitats,
16 modeling, and management in the Pacific Northwest. *Bryologist* 108:377–390
17 [doi:10.1639/0007-2745\(2005\)108\[0377:RECLHM\]2.0.CO;2](https://doi.org/10.1639/0007-2745(2005)108[0377:RECLHM]2.0.CO;2)</jrn>
- 18 <jrn>Goudie RI, Scheidegger C, Hanel C, Munier A, Conway E (2011) New population
19 models help explain declines in the globally rare boreal felt lichen *Erioderma*
20 *pedicellatum* in Newfoundland. *Endang Species Res* 13:181–189
21 [doi:10.3354/esr00326](https://doi.org/10.3354/esr00326)</jrn>
- 22 <eref>Government of Newfoundland and Labrador, Department of Environment and
23 Conservation (2010) Species at risk. Available at
24 www.env.gov.nl.ca/env/wildlife/endangeredspecies/index.html (accessed on 29 May
25 2010) </eref>
- 26 <jrn>Guisan A, Edwards T, Hastie T (2002) Generalized linear and generalized additive
27 models in studies of species distributions: setting the scene. *Ecol Model* 157:89–100
28 [doi:10.1016/S0304-3800\(02\)00204-1](https://doi.org/10.1016/S0304-3800(02)00204-1)</jrn>
- 29 <eref>Hastie T (2010) Package ‘gam’: generalized additive models. Available at
30 <http://cran.r-project.org/web/packages/gam/gam.pdf></eref>
- 31 <jrn>Hastie T, Tibshirani R (1986) Generalized additive models. *Stat Sci* 1:297–318
32 [doi:10.1214/ss/1177013604](https://doi.org/10.1214/ss/1177013604)</jrn>
- 33 <jrn>Hylander C (2005) Aspect modifies the magnitude of edge effect on bryophyte
34 growth in boreal forests. *J Appl Ecol* 42:518–525 [doi:10.1111/j.1365-](https://doi.org/10.1111/j.1365-2664.2005.01033.x)
35 [2664.2005.01033.x](https://doi.org/10.1111/j.1365-2664.2005.01033.x)</jrn>
- 36 <bok>Keeping B, Hanel C (2006) A five year (2005–2011) management plan for the
37 boreal felt lichen (*Erioderma pedicellatum*) in Newfoundland and Labrador.
38 Government of Newfoundland and Labrador, </bok>
- 39 <jrn>Lang W, Reiners A, Pike LH (1980) Structure and biomass dynamics of epiphytic
40 lichen communities of balsam fir forests in New Hampshire. *Ecology* 61:541–550
41 [doi:10.2307/1937420](https://doi.org/10.2307/1937420)</jrn>

- 1 <bok>Maass WSG, Yetman D (2002) COSEWIC assessment and status report on the
2 boreal felt lichen, *Erioderma pedicellatum*, in Canada. Committee on the Status of
3 Endangered Wildlife in Canada, Ottawa, ON</bok>
- 4 <jrn>Martínez I, Carreño F, Escudero A, Rubio A (2006) Are threatened lichen species
5 well-protected in Spain? Effectiveness of a protected areas network. *Biol Conserv*
6 133:500–511 [doi:10.1016/j.biocon.2006.08.003](https://doi.org/10.1016/j.biocon.2006.08.003)</jrn>
- 7 <bok>Meades WJ, Moores L (1989) Forest site classification manual. A field guide to
8 the Damman Forest Types of Newfoundland, 1st edn. Newfoundland Department of
9 Forestry and Agriculture, Forestry Canada, </bok>
- 10 <jrn>Moores L, Pittman B, Kitchen G (1996) Forest ecological classification and
11 mapping: their application in ecosystem management in Newfoundland. *Environ*
12 *Monit Assess* 39:571–577 [doi:10.1007/BF00396169](https://doi.org/10.1007/BF00396169)</jrn>
- 13 <jrn>Parker AJ (1982) The topographic relative moisture index: an approach to soil-
14 moisture assessment in mountain terrain. *Phys Geogr* 3:160–168</jrn>
- 15 <jrn>Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT (2007) Predicting species
16 distributions from small numbers of occurrence records: a test case using cryptic
17 geckos in Madagascar. *J Biogeogr* 34:102–117 [doi:10.1111/j.1365-2699.2006.01594.x](https://doi.org/10.1111/j.1365-2699.2006.01594.x)</jrn>
- 19 <eref>R Development Core Team (2010) R: a language and environment for statistical
20 computing. R Foundation for Statistical Computing, Vienna. Available at [www.R-](http://www.R-project.org/)
21 [project.org/](http://www.R-project.org/)</eref>
- 22 <jrn>Rolstad J, Gjerde I, Storaunet KO, Rolstad E (2001) Epiphytic lichens in
23 Norwegian coastal spruce forests: historic logging and present forest structure. *Ecol*
24 *Appl* 11:421–436 [doi:10.1890/1051-0761\(2001\)011\[0421:ELINCS\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2001)011[0421:ELINCS]2.0.CO;2)</jrn>
- 25 <eref>Scheidegger C (2003) *Erioderma pedicellatum*. In: 2008 IUCN red list of
26 threatened species. Available at www.iucnredlist.org (accessed on 22 November
27 2008) </eref>
- 28 <jrn>Werth S, Tommervik H, Elvebakk A (2005) Epiphetic macrolichen communities
29 along regional gradients in northern Norway. *J Veg Sci* 16:199–208
30 [doi:10.1111/j.1654-1103.2005.tb02356.x](https://doi.org/10.1111/j.1654-1103.2005.tb02356.x)</jrn>
- 31 <bok>Wilds S, van Manen F (1995) Arc markup language script for the topographic
32 convergence index. *US Geological Survey*, </bok>
- 33 <jrn>Wisz MS, Guisan A (2009) Do pseudo-absence selection strategies influence
34 species distribution models and their predictions? An information-theoretic approach
35 based on simulated data. *BMC Ecol* 9:8 [PubMed](https://pubmed.ncbi.nlm.nih.gov/1472-6785-9-8/) [doi:10.1186/1472-6785-9-8](https://doi.org/10.1186/1472-6785-9-8)
36 [doi:10.1186/1472-6785-9-8](https://doi.org/10.1186/1472-6785-9-8)</jrn>
- 37 <jrn>Wolock DM, McCabe GJ (1995) Comparison of single and multiple flow direction
38 algorithms for computing topographic parameters in TOPMODEL. *Water Resour Res*
39 31:1315–1324 [doi:10.1029/95WR00471](https://doi.org/10.1029/95WR00471)</jrn>

1 **Fig. 1.** Newfoundland, Canada, showing boreal felt lichen *Erioderma pedicellatum* sites
 2 in red (source: Government of Newfoundland and Labrador 2006, Keeping & Hanel
 3 2006)

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

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

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

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

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

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

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

31

32 **Table 1.** *Erioderma pedicellatum*. Potential models for Newfoundland, Canada, and their
 33 respective statistical analyses, including when applicable, chi squared, p values, Z values
 34 R^2 adjusted, and deviance explained. O: occupancy, D: distance from coastline, A:
 35 topographic aspect, T1: topographic convergence index, T2: topographic relative
 36 moisture index, s: smoothing function used in the generalized additive model, na: not
 37 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

2	O = sD + T1	385.40 na	2.00×10^{-16} 0.070	na 1.81	31.30	0.36
3	O = sA + sD	380.85 16.93	2.00×10^{-16} 0.033	na na	32.20	0.37
4	O = sA + T1	22.10 na	0.0065 0.20	na 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 19.93 na	2.00×10^{-16} 0.012 0.072	na na 1.80	32.3	0.37
9	O = sD + T2	386.80 na	2.00×10^{-16} 0.25	na 1.15	31.2	0.37
10	O = sA + T2	20.22 na	0.012 0.53	na -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 18.31 29.06	2.00×10^{-16} 0.021 0.0061	na na na	33.70	0.39
13	O = sD + sT1	332.66 28.68	2.00×10^{-16} 0.00071	na na	32.70	0.3770
14	O = sA + sT1	17.12 111.76	0.030 2.00×10^{-16}	na 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 20.05 3.37	2.00×10^{-16} 0.012 0.066	na na na	32.40	0.37
17	O = sD + sT2	386.82 1.30	2.00×10^{-16} 0.26	na na	31.2	0.37
18	O = sA + sT2	20.20 0.37	0.012 0.55	na 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