**Species distribution model identifies influence of climatic constraints at the leading edge of a native insect outbreak**

Bo Zhang a \*, Shawn J. Leroux a, Joseph J. Bowden b, Kathryn E. Hargan a, Amy Hurford a, Eric R. D. Moise b

a Department of Biology, Faculty of Science, Memorial University of Newfoundland, 45 Arctiv Avenue, St. John’s, NL, A1C 5S7, Canada

b Natural Resources Canada, Canadian Forest Service, Atlantic Forest Centre, Corner Brook, NL, A2H 5G4, Canada

\* Correspondence:

Bo Zhang

Email: bo.zhang@mun.ca

**Abstract**

1. Numerous bioclimatic, vegetation, and spatial-temporal variables can influence the distribution and outbreak patterns of phytophagous insects. In eastern North America, spruce budworm (*Choristoneura fumiferana* Clem.) is one of the most important forest insects with regard to outbreak coverage and impacts to the forest industry. Recent models that identified predictors for distributions and impacts of budworm outbreaks have limited agreement on key drivers. Spatially-explicit, fine-scale predictive models are needed for understanding ongoing and future budworm outbreaks.

2. We used a species distribution model ‒ MaxEnt to examine the relationships between the 1972-1981 budworm defoliation and 13 predictor variables for the island of Newfoundland, Canada ― the leading edge of the current budworm outbreak.

3. The optimal model for predicting the probability of severe budworm defoliation included precipitation in May and June, precipitation in July and August, and maximum temperature in June and July. This probability decreased with the distance to major roads and river channels, suggesting possible corridor effects. We then used this MaxEnt machine learning algorithm to predict the probability of severe defoliation for the forests of the entire island on a 2 km resolution raster. Areas with high probability (> 0.80) were concentrated in the northeast part of the island where black spruce (*Picea mariana* (Mill.) BSP) dominates the forest landscape. The predicted severe defoliation area (probability 0.4-1.0) for black spruce was 1.4 times the area for balsam fir, which is the dominant canopy forming tree species on the island and preferred host species for budworm.

4. The higher predicted probability of severe defoliation in black spruce stands relative to balsam fir stands may arise due to the unique combination of climatic conditions and geographical distributions of two species on the island. Species distribution models for insect outbreaks are relatively rare but may be an important tool in forecasting future outbreaks.

**Key words**: Boreal, Spruce budworm, Newfoundland, MaxEnt, Machine learning, ENMeval 2.0, Precipitation, Temperature

**Introduction**

Phytophagous insects, especially those that exhibit periodic outbreaks over large geographical extents are important disturbance agents in temperate and boreal forest ecosystems (Johnson et al., 2005). Predicting spatial patterns of insect outbreaks is imperative, yet difficult, due to the complexity of underlying ecological processes, such as plant-insect-natural enemy interactions, dispersal, and synchrony (Nenzén et al., 2017). In eastern North America, spruce budworm (*Choristoneura fumiferana* Clem., SBW) is the most important forest insect with regard to outbreak extent and impacts on the forest industry (Régnière et al., 2019). SBW outbreaks occur every 30-40 years, and repeated defoliation typically last up to 10 years (MacLean et al., 2019), causing landscape-scale disturbances on forest production, succession, and ecological stability (MacLean, 2016). The outbreak in 1967-1993 covered more than 50 million ha. in eastern Canada and the northeastern United States and a new outbreak, which began in 2004 in Quebec, Canada, has covered over 8.2 million hectares of spruce-fir forests in that province by 2018 (MacLean et al., 2019). Predicting the spatial distribution of impacted areas from current and future outbreaks is paramount for understanding ecological dynamics of this large-scale ecological disturbance.

There are distinct views on the ecological processes that cause the initiation, termination, and recurrence of the SBW outbreaks (Pureswaran et al., 2016). The “double-equilibrium theory” (Ludwig et al., 1978) presumes that the eruption of budworm populations is triggered by, or coincides with, certain environmental conditions. The “oscillatory theory” asserts that a periodic predator-prey cycling process dominates the outbreak cycles; the budworm population is regulated mainly by its natural enemies (Royama et al., 2017). With limited field budgets, entomologists and natural resource managers usually conduct field observations in a small selection of sites, thus population dynamics are spatially restricted. Making spatially-explicit description and prediction of SBW population dynamics over large geographical areas is difficult and requires accumulating local scale data across landscape extents.

More than 20 spatial and aspatial models have been constructed incorporating different processes to make predictions of SBW outbreaks at various spatial extents (Cooke et al., 2007). Numerous bioclimatic, vegetation, and spatio-temporal predictors are involved in SBW modeling but key predictors of outbreak dynamics and defoliation patterns are not consistent across studies (see summary in Table A.1 in Appendix S1). For example, spring temperature is often expected to be a strong predictor of defoliation patterns as it influences the synchrony between larval emergence from winter diapause and bud flush of host trees. Candau and Fleming (2005) found low defoliation frequencies associated with high spring temperature in Ontario, Canada whereas Li et al. (2020) showed that warmer spring led to higher defoliation in New Brunswick, Canada. Furthermore, spatial and aspatial approaches also generate inconsistent trends; for example, a forest stand characteristic - site quality has negative effects on defoliation in aspatial models (MacKinnon & MacLean, 2003), but spatially-explicit models did not detect significant relationships (Magnussen et al., 2004; Li et al., 2020).

Most SBW models focus on its two main host tree species: balsam fir (*Abies balsamea* (L.) Mill.) and white spruce (*Picea glauca* (Moench) Voss) (Blais, 1981). Balsam fir tends to be more susceptible than spruce spp. to budworm defoliation due to synchronization between spring tree budburst and budworm larval emergence from overwintering (Pureswaran et al., 2019). Black spruce (*Picea mariana* (Mill.) BSP) is more abundant in boreal forests, where balsam fir is often geographically separated from black spruce. Only a few studies, however, have compared the outbreak impacts on balsam fir and black spruce (e.g., Hennigar et al., 2008; Pothier et al., 2012). Published studies on SBW in Newfoundland are sparse where the forests are mainly comprised of balsam fir and black spruce (Arsenault et al., 2016; Leroux et al., 2021). The synchronization of spring tree bud flush and budworm larval emergence can be influenced by the unique cool and wet insular climate on the island of Newfoundland, making predictions for Newfoundland more difficult.

To reconcile the context-dependency in SBW outbreak and environmental predictors, we conduct a case study for the island of Newfoundland – the northeastern tip of SBW distribution and the leading edge of the current SBW outbreak. Our objectives were to use a species distribution model to 1) examine key predictor variables influencing the spatial dynamics of SBW defoliation; 2) predict spatially-explicit severe defoliation areas based on bioclimatic constraints and previous outbreak patterns; and 3) compare outbreak coverage in balsam fir and black spruce forests. We evaluated two predictions: 1) climatic predictors (e.g., temperature and precipitation) will be more influential than forest stand characteristics in describing SBW outbreak across large landscape; 2) susceptibility of black spruce is similar to balsam fir in Newfoundland due to the unique combination of climatic conditions and geographical distribution of the two host species.

**2 Materials and Methods**

**2.1 Study area**

We investigated the spruce budworm outbreak which occurred during the 1970s-1990s on the island of Newfoundland (111390 km2), located in the boreal shield ecozone (Natural Resources Canada, 2020) on the east coast of Canada (Figure 1). The climate on the island of Newfoundland is semi-coastal with an average annual temperature of 5°C and annual precipitation of 1400 mm.

The first written-documentation for a provincial outbreak is from the early 1940s to 1950s. Defoliation on spruce and fir was observed but no tree mortality was reported (Carroll, 1956). A much larger outbreak occurred throughout the 1970s to the early 1990s. SBW defoliation covered about 4 million ha. of forested lands, which is larger than all other disturbances combined (e.g., wildfire, windthrow, and other insects) during the time period (Arsenault et al., 2016).

**2.2 Species distribution model response variable: severe defoliation area**

The Canadian Forestry Service systematically administered annual aerial SBW defoliation assessments for major forested areas across Newfoundland during 1972-1992 (Sterner & Davidson, 1980) and the data was stored in ESRI ArcGIS polygon featureclass format. Defoliation severity was classified into three categories: light, moderate, and severe, corresponding to 5-30%, 31-50%, and 76-100% defoliation, respectively (See Appendix S1 for more details on defoliation assessments).

We analyzed the defoliation data for 1972-1981 — the most intense defoliation period in the 1972-1992 aerial survey data (See Appendix S1 for summary of the defoliation data). We overlaid the 1972-1981 aerial defoliation data with a 2 km by 2 km fishnet point layer (ArcGIS: Create Fishnet) covering the entire island of Newfoundland. The fishnet points were assigned with a yes/no dichotomous response: points with at least three consecutive years of moderate to severe defoliation were classified as “severe” defoliation areas and labeled with yes. High tree mortality was likely to occur in these cells, resulting in stand-replacing succession (Arsenault et al., 2016).

**2.3 Species distribution model predictor variables**

**2.3.1 Climate variables**

Climate datasets were obtained from Natural Resource Canada (McKenney et al., 2011). The customized spatial climate data contain monthly climate variables such as maximum temperature, minimum temperature, precipitation, etc. at 60 arcsecond (~ 1.2 km) resolution in ASCII grid format, available for the same time period as our response variable (i.e., 1972-1981).

We chose a common 2 km resolution for all spatial layers to accommodate the lowest resolution climate data; each raster cell is 400 ha. All predictor raster data layers can be found on the online repository (Zhang et al., 2022).

Based on previous budworm models (see Table A.1 in Appendix S1), we chose the following 12 climate variables to reflect conditions that would be expected to influence budworm survival, dispersal, or search rates. Firstly, we calculated the monthly average of the maximum temperature, minimum temperature, and precipitation in May and June, in June and July, in July and August. Secondly, we included the Julian date of the start of growing season and two variables of cumulative degree days above base temperature (5 °C) for two time periods: a) three months prior to the start of growing season, and b) first six weeks after the starting date of the growing season (see Appendix S1 for an ecological justification of each predictor variable).

**2.3.2 Forest species composition**

We extracted forest species composition from the 2015 Newfoundland forest inventory (Department of Fisheries Forestry and Agriculture of Newfoundland and Labrador, 2022). We excluded non-commercial forests (labeled as deciduous shrub and coniferous shrub) from the study. For commercial forests (3.35 million ha.), a maximum of three dominant tree species (by basal area) in each forest stand was recorded in the original spatial data.

Firstly, we reclassified the forest stands into five species groups (see Appendix S1 for more details on justification of grouping) based on current knowledge of susceptibility to budworm defoliation (MacLean & MacKinnon, 1997; Hennigar et al., 2008; Zhang et al., 2018) and defoliation history (Arsenault et al., 2016): balsam fir (bF), black spruce (bS), mixedwood (balsam fir and spruce spp. mixed with broadleaf spp.: main broadleaf/hardwood species are white birch (*Betula papyrifera* Marshall), trembling aspen (*Populus tremuloides* Michx.), and yellow birch (*Betula alleghaniensis* Britton); MW), tamarack (*Larix laricina* (Du Roi) K. Koch; tL), and other (all other stands except the above four groups; OT). We did not classify white spuce into a seperate group as white spruce only accounts for 0.51% of the commercial forests by area. Secondly, we overlaid forest species polygon with a 200 m by 200 m fishnet point layer (ArcGIS: Create Fishnet); each fishnet point was assigned with the tree species code from the polygon that contains the point. We chose this fine 200 m resolution to capture the variation in size and shape of the forest inventory polygons. Then, we converted the fishnet point layer to a 2 km by 2 km resolution raster (ArcGIS: Point to Raster) and populated the cell value using the tree species code from the most frequent fishnet points within each cell (See Table A.2 in Appendix S1 for detailed statistics of each species group).

**2.3.3 Elevation, slope, and aspect**

We extracted elevation data from the digital elevation model SRTM (DEM, Shuttle Radar Topography Mission) at 90 m resolution between -60 and 60 latitude (Farr et al., 2007). The DEM was clipped to the Newfoundland spatial extent and aggregated to 2 km resolution. We computed slope and aspect from the elevation raster (Ritter, 1987). We obtained the DEM using the “getData/alt” function from the “raster” package in R version 3.6.0 (R Core Team, 2021). We used the “resample/bilinear” function from the “raster” package to convert the resolution. Slope and aspect were calculated using “terrain/slope” and “terrain/aspect” functions in “raster”, respectively.

**2.3.4 Adjacency to road network and watercourse**

Plants and animals can use road networks as dispersal corridors (Forman & Alexander, 1998). Insect species have used road networks and roadsides as habitats or travel corridors, especially in fragmented landscapes (Bernes et al., 2017). Similarly, river networks can also provide dispersal corridors for insects (Tonkin et al., 2018). Whether spruce budworm can use road and river networks for dispersal has not been studied empirically.

We obtained the Newfoundland road network from the National Road Network GeoBase Series, Canadian Statistical Geomatics Centre in Shapefile format (Statistics Canada, 2021). We selected only major road classes (road class: ‘freeway’, ‘expressway/highway’, ‘arterial’, and ‘collector’). We created a 2 km-resolution raster layer by computing the nearest distance from each raster cell to the road features using the “gDistance” function in the “rgeos” package in R.

We downloaded riverine data from CanVec Series - Hydrographic Features, Natural Resources Canada in Shapefile format (Natural Resources Canada, 2021). We selected the 5M-scale (1:5000000) dataset so that only major riverine features in Newfoundland were included in this study. We created a 2 km-resolution raster layer using the same method as for the road raster.

**2.4 Species distribution model: analysis**

We fit a species distribution model (SDM) to analyze the relationship between severe budworm defoliation and predictor data layers. We executed the MaxEnt procedure using the “dismo” package (Hijmans et al., 2021) in R version 3.6.0 (R Core Team, 2021). MaxEnt is a deterministic machine learning algorithm SDM based on maximum entropy theory (Phillips et al., 2006).

Our application of SDMs and MaxEnt is different from the classic examples (see reviews in Charney et al., 2021) as we did not model occurrence of a species in question but rather the occurrence of defoliation from an outbreaking insect. SDMs are commonly used, however, to model ecological phenomena beyond species occurrence, for example, Pablo et al. (2019) used a MaxEnt model to describe and predict distribution of tree fungal diseases caused by the pathogenic bacteria *Xylella fastidiosa* (Wells).

For better transparency and reproducibility of the SDM, we followed the ODMAP (Overview, Data, Model, Assessment, and Prediction) protocol (Fitzpatrick et al., 2021) to document the modeling procedure and to report the results.

**2.4.1** **Variable selection**

We used variance inflation factor test (VIF, Naimi et al., 2014) and correlation analysis to identify collinearity in our *a priori* list of 17 non-categorical predictor variables (Table 1). Based on our VIF and correlation analysis (see full details in Appendix S1), we dropped five collinear input variables: maximum temperature in July and August (MaxT0708), minimum temperature in May and June (MinT0506), minimum temperature in June and July (MinT0607), Julian date of the start of growing season (sg\_date), and cumulative degree days above base temperature (5 °C) for the first six weeks after the starting date of the growing season (sg\_gdd\_p2).

Including the forest species composition as a categorical layer, a total of 13 predictor layers (Table 1) were stacked as a RasterStack object as model input. All raster layers were set to a single spatial projection (NAD83, UTM zone 21). The RasterStack contains a total of 23782 raster cells across the island of Newfoundland.

**2.4.2 Presence and background data**

The SDMs uses presence data (also referred as occurrence data) on the species of interest as a response variable (Merow et al., 2013): e.g., presence of an invasive species in new environment. In our 2 km-resolution severe defoliation fishnet point layer, we have a total of 2964 ‘yes’ points denoting the historical presence of severe defoliation. Each MaxEnt model iteration selects a random subset of these presence points (see below). Our presence data were derived from a systematic survey and are therefore less prone to sampling bias (Merow et al., 2013; Baker et al., 2022).

MaxEnt uses background points where presence/absence is unmeasured to contrast present points (Phillips & Dudík, 2008). Since our presence data were systematically sampled from all areas in the forest layer, we also randomly selected background points from the forest layer. We used an “accumulation curve” technique to determine the number of background points (see details in Appendix S1). We selected 14820 random points as background based on a 5:1 background/presence ratio (background/presence = 14820/2964). By design, only one presence or background point can be extracted from a single raster cell. The presence points and background points covered 12.5% and 62.3% of all raster cells of the predictor RasterStack, respectively.

**2.4.3 Optimization and tuning of model parameters**

MaxEnt models can be optimized by two main parameters: 1) the combinations of feature classes (Merow et al., 2013) and 2) regularization multiplier values. A combination of linear (L), quadratic (Q), product (P), threshold (T), and hinge (H) feature classes was applied to predictors to determine the shape of the marginal response curves. The regularization multiplier (also known as beta multiplier) is a constant (positive numerical values, MaxEnt default = 1) that is used to tune the regularization coefficient.

To find the optimal model parameters (see full details in Appendix S1), we tested 35 combinations of feature classes and regularization multipliers: seven feature class combinations (L, LQ, LH, LQH, LQP, LQHP, LQHPT) were cross-combined with five regularization multiplier values (0.5, 1.5, 2.5, 3.5, 4.5).

**2.4.4 Model selection and evaluation**

We used three metrics to identify the optimal model: AUC (area under the curve), CBI (continuous Boyce index), and AIC (Akaike information criterion) (see full details in Appendix S1). For the AUC and CBI evaluation, we used two cross-validation methods (see full details in Appendix S1) to partition presence and background points into training and testing datasets: ‘random k-fold’ cross-validation (Hastie et al., 2009) and spatial partitioning ‘block’ method (Radosavljevic & Anderson, 2014).

The random k-fold and block data partitioning methods yielded similar model ranking. Thus, we only presented the results from the block method. We executed all model selection, model evaluation, and data partitioning procedures with the “ENMeval 2.0” package (Kass et al., 2021) in R version 3.6.0 (R Core Team, 2021).

**2.4.5 Variable importance and model output**

We examined multiple model outputs on variable importance comprehensively to gain better insights on budworm biology and its relationship to the host environment. First, we extracted percent contribution and permutation importance from the optimal model. Percent contribution estimates the increase of regularized gain when the corresponding variable is added, whereas permutation importance measures the drop in training AUC when the corresponding variable is randomly permuted. Second, we built relationship curves showing how predicted probability of presence changes as each top-ranked (based on percent contribution and permutation importance) predictor variable is varied, keeping all other predictor variables at their average value. Furthermore, we reran the optimal model with the jackknife test (Shcheglovitova & Anderson, 2013). Jackknife measures model gains by training a series of models with each variable first omitted, then used in isolation.

We plotted and inspected the model prediction with the default cloglog (complementary log-log) output format. Cloglog is a logistic transformation of the estimated probability of presence in each grid cell, with a bounded value between 0 and 1 (Phillips et al., 2017). From the prediction, we further examined the estimated probability of presence, i.e., probability of severe defoliation, in balsam fir and black spruce cells.

**3 Results**

We identified 12.5% (2964 cells, 1.19 million ha.) of the forest layer (23782 cells, 9.51 million ha.) with at least three consecutive years of moderate to severe aerially-surveyed defoliation. See Appendix S1 for a summary of predictor variables in our study region.

**3.1 Strong influence from climatic predictors**

Based on AUC, CBI, AICc, and model complexity (feature class and regularization multiplier), our optimal model was LQHPT-3.5 (a combination of linear, quadratic, product, threshold, and hinge feature classes, regularization multiplier = 3.5, see Table A.3 in appendix for the top 20 of the 35 candidate models). The optimal model produced the smallest decrease of AUC-training to AUC-testing (0.105), yielded the largest CBI value from testing data (0.698), and was ranked 10th by AICc among the top 20 models. Also, LQHPT-3.5 adopted the largest regulation multiplier among the top 10 models, suggesting lower complexity, simpler model prediction, and higher transferability in novel environments.

The percent contribution and permutation importance assessments identified the same six top variables among the 13 predictor variables, but with a slightly different ranking (Table 2): precipitation in May and June (Prcp0506), maximum temperature in June and July (MaxT0607), distance to roads (Dist\_road), elevation, precipitation in July and August (Prcp0708), and distance to rivers (Dist\_river). Both metrics indicated that precipitation in May and June (Prcp0506) was the most influential variable for predicting probability of severe SBW defoliation. The jackknife procedure showed that precipitation in July and August (Prcp0708) generated the highest model gain when each variable was modeled in isolation, indicating that Prcp0708 has the most useful information by itself (Figure 2). On the other hand, Distance to rivers (Dist\_river) caused the largest decrease in gain when it was omitted, suggesting that it carried the most information that was not present in other variables. Forest species did not rank as an important predictor (Table 2, Figure 2).

Probability of predicted severe SBW defoliation was below 0.50 before precipitation in May and June (Prcp0506) reaching 70 mm, and rapidly increased to > 0.70 when Prcp0506 was at the range of 75.0 mm – 91.0 mm, before declining gradually (Figure 3.A). Probability of severe defoliation increased with maximum temperature in June and July (MaxT0607) until temperature reached 18.0 °C (peak probability = 0.88), and then started decreasing. Both distance to road (Dist\_road) and distance to river (Dist\_river) showed negative correlations with severe defoliation probability, suggesting that forests closer to major roads and rivers are more susceptible to budworm outbreak (Figure 3.D and Figure 3.E). Predicted probability dropped to about 0.50 when a location was 29 km and 85 km away from major roads and major rivers, respectively. High severe defoliation probability (> 0.80) was predicted for elevation ranging between 37.4 m - 197.7 m, whereas slope and aspect did not show strong influence. We observed a complex relationship between severe defoliation probability and precipitation in July and August (Prcp0708) (Figure 3.B).

**3.2 Prediction of severe defoliation area**

To predict areas of severe defoliation, we classified the probability from 0-1 into five categories by a 0.2 step. Examining the raster data of probability of severe defoliation, most cells (9485) were predicted to have < 0.20 probability. The 1947 cells with the highest probability (0.80) are concentrated mainly in the northeast where black spruce dominates the forest landscape (Figure 1). The northwest (Northern peninsula) and southern part of Newfoundland (South coast, Burin peninsula, Avalon peninsula) showed low probability of severe defoliation.

**3.3 Black spruce is more susceptible than balsam fir**

The majority of balsam fir forests (61.1%, 2.5 million ha.) showed low probability of 0.20 to severe spruce budworm defoliation, whereas only 30.9% of black spruce forests were in the low probability bin (Figure 4, also see details in Table A.4 in appendix). In 0.40-0.60, 0.60-0.80, and 0.80-1.0 probability bins, the percentage of black spruce forests (19.1%, 21.7%, and 8.3%, respectively) was almost double the percentage of balsam fir forests (10.0%, 11.1%, and 4.1%, respectively) in each bin. Area of predicted severe defoliation for black spruce was larger than for balsam fir in all bins except 0-0.2. Severe defoliation area (0.4-1.0 bins) for black spruce was 1.4 times of the area for balsam fir (Table A.4 in appendix), indicating higher susceptibility of black spruce to spruce budworm than other tree species on the island of Newfoundland.

**4 Discussion**

Predicting spatial patterns of insect outbreaks over large geographical areas is difficult due to complex relationships between insect distributions and their host environment. Indeed, there is little agreement on the top biotic and abiotic predictors of outbreak patterns across study areas and modeling platforms (reviewed in Table A.1 in Appendix S1). We fitted a machine learning species distribution model to identify top predictors of spruce budworm defoliation on the island of Newfoundland - the leading edge of the current budworm outbreak. As expected, we found climatic predictors (especially precipitation and temperature) were more influential than forest stand characteristics in describing SBW outbreak across the island landscape. Black spruce stands were more susceptible than balsam fir stands due to the unique combination of climatic condition and geographical distribution of two species on the island. Surprisingly, all of our variable importance assessments suggest little importance of forest tree species composition to probability of severe defoliation.

**4.1Climate and distance to corridors were strong predictors for severe defoliation**

Precipitation, temperature, bathymetry, distance to water, and habitat characteristics (e.g., vegetation) surfaced as the most important variables among over 400 predictors of species distribution in a review of 2040 MaxEnt SDMs (Bradie & Leung, 2017). In spruce budworm literature, a variety of climatic variables showed strong influence on budworm defoliation in different regions (as summarized in Table A.1 in Appendix S1). For example, spring temperature influenced budworm defoliation positively in New Brunswick, Canada (Li et al., 2020) while summer precipitation showed negative effects in Ontario, Canada (Candau & Fleming, 2005).

Our MaxEnt model identified three climatic variables that strongly influenced spruce budworm defoliation on the island of Newfoundland: spring precipitation (Prcp0506), fall precipitation (Prcp0708), and summer maximum temperature (MaxT0607). Few spatial models included or reported significant effects of spring precipitation on defoliation patterns. Li et al. (2020) included fall precipitation in a predictive model for New Brunswick, Canada but did not find it was significantly related to spatial defoliation patterns. The climate on the island of Newfoundland is unique for spruce budworm: cold and wet summer climate on the west coast creates adverse conditions for budworm activities whereas in the inland areas, warm and dry summer is ideal for budworm infestation. This regional variation in climate conditions shaped our model prediction. Specifically, the model predicted a high probability of severe defoliation associated with spring precipitation (Prcp0506) ranging from 75.0 mm – 91.0 mm (Figure 3.A). During the transition period from winter to spring, temperature is critical to the metamorphosis and metabolism for ectothermic insects (Lemoine & Burkepile, 2012). Precipitation is an important co-dominating factor to temperature on insect phenology and survival. Spruce budworm second-instar larvae emerge from winter diapause and subsequently disperse on silk threads in spring, searching for suitable feeding sites. Second-instar dispersal survival is a major mortality event in the budworm life cycle and it is critical to the population size in subsequent feeding stages (Zhang et al., 2020). A lack of water could result in high mortality for newly-emerged insects, particularly if under high temperature (Eskafi & Fernandez, 1990). Low precipitation (< 75.0 mm) may adversely influence spring emergence of second-instar larvae and their synchrony with budburst of the host trees. On the other end, high spring precipitation would inevitably hinder the success of the second-instar larval dispersal via delicate silk threads. Wet conditions may also promote pathogens that can cause mortality of young larvae, as found for other forest insects (Hajek et al., 2015) . The fall precipitation (Prcp0708) generated the highest model gain in the jackknife analysis indicating that it carried the most useful information when modeled alone (Figure 2). Frequent or heavy rainfall events are likely to adversely affect budworm moth flight (Dickison et al., 1983). If the unfavorable weather condition discourages moth dispersal, female moths may have to lay eggs locally. The large amount of egg deposition would in turn result in high larval density and thus more severe defoliation in the subsequent year. Future work could test the influence of spring and fall precipitation on budworm survival and egg deposition along a gradient of precipitation intensities in the natural environment or using artificial rainfall.

We observed a peak probability of severe defoliation at summer temperature (MaxT0607) of 18.0 °C. Generally, higher temperature leads to higher metabolism and thus higher consumption of the host foliage. However, high summer extreme temperatures may be disadvantageous to the natural enemies of budworm if the temperature exceeds the optimum temperatures for their developmental rates (Gray, 2008). The predicted probability of severe defoliation started decreasing above 18.0 °C. This decreasing trend may be caused by correlations between other variables, e.g., summer precipitation (Prcp0607) negatively correlated (-0.74) with maximum summer temperature (MaxT0607). We speculated that beyond 18.0 °C, the low precipitation asscociated with high temperature may decrease foliage consumption and therefore defoliation probability.

Areas close to major road networks and river channels showed high probability of severe defoliation. This result agrees with those of Magnussen et al. (2004), who found budworm defoliation areas are concentrated along rivers in British Columbia, Canada. They attributed the effects to the larger tree sizes found in proximity to rivers. Our model analyzed forest properties in 2 km2 cells and we found that the probability of severe defoliation dropped to about 0.50 at distances of 29 km and 85 km away from major major roads and rivers, respectively. At these distances, tree size was unlikely to contribute to the effects. Alternatively, both natural and anthropogenic corridors can affect the dispersal and movement of avian insects. For example, Mönkkönen and Mutanen (2003) found that noctuid and geometrid moths used riparian corridors in Finnish boreal forests as breeding habitats or dispersal corridors. Similarly, we speculate that budworm moths would use major roads and rivers as traveling routes and the forests nearby as highly accessible habitats. The frequent visits by moths would likely lead to a high amount of egg deposition and thus high defoliation in the forests along these corridors. Forest “edge effects” can also influence insect egg mass distribution. Gypsy moth, (*Lymantria dispar* L.) deposited 2.4 times more egg masses on forest edges than in the interior (Bellinger et al., 1989). It is likely that budworm moths will lay more eggs on trees along the river and road edges and cause higher damage. Future empirical work could test this model prediction by comparing budworm population density and defoliation intensity along river channels and roads networks.

**4.2 Black spruce is more susceptible than balsam fir to spruce budworm under the unique insular climate of Newfoundland**

The model predicted higher susceptibility for black spruce than balsam fir forests have to spruce budworm defoliation. The proportion of black spruce forests with high probability (> 0.6) of severe defoliation is twice as much as the proportion for balsam fir (see Table A.4 in appendix) and this despite balsam fir being the dominant canopy forming tree species on the island. At local scales, when fir and spruce are both available, fir is more susceptible than spruce because its spring budburst is more synchronized to budworm second-instar larval emergence (Pureswaran et al., 2019). When the foliage of primary hosts are depleted, budworm will defoliate less desirable tree species such as tamarack, pine spp., and hemlock (*Tsuga canadensis* (L.) Carr.) (Talerico, 1983; Volney & Fleming, 2007). Hence, spatial models which assess budworm impacts in larger geographical areas often show inconsistent results on defoliation in fir and spruce forests (see summary in Table A.1 in Appendix S1). For example, Bouchard and Auger (2014) found balsam fir were more susceptible than black spruce in Quebec, Canada, whereas Li et al. (2020) did not find forest composition important in predicting defoliation.

We surmise that when modeling insect distribution and disturbance at large spatial extents, the variation in climatic conditions can be more important than the host tree availability and forest conditions (Bradie & Leung, 2017). Under the distinct bioclimatic conditions on the island of Newfoundland (Meades, 2008), defoliation is less severe in balsam fir forests as they dominate the western part of the island where summer is wet and cool. Defoliation is more severe in black spruce forests as they dominate the central and eastern regions where summer is dry and warm (see geographic variation of climatic variables in Figure A.1 in appendix).

Our model analyzed severely defoliated areas over a large landscape, i.e., any balsam fir or black spruce dominant raster cell that experienced three or more consecutive years of moderate to severe aerial-surveyed defoliation. Our goal was to assess susceptibility (defoliation) rather than vulnerability (mortality) of the forests to spruce budworm. It is possible that some fir stands could experience many more years of defoliation than spruce stands and eventually demonstrate higher mortality (Chen et al., 2017). The model did not test stand-level insect-host interaction or host preference per se. Instead, in essence, it analyzed the function of various bioclimatic predictors and host tree species distribution. If a given host species (e.g., black spruce in this study) occurs in a greater proportion in an outbreak-suitable climate zone, it becomes more susceptible due to its realized niche and/or geographic happenstance. Bioclimatic predictors become the main driver of observed insect damage.

**4.3 Species distribution models are a useful tool for predicting insect disturbances**

Although MaxEnt SDMs typically assess and predict distributions of floral or faunal species, we applied it as an innovative tool for predicting the consequences of an native forest insect outbreak. Ecologists have used SDMs and advanced machine learning algorithms extensively to predict the distribution of various pest species around the globe (Gobeyn et al., 2019). Using SDMs to support ecological impact assessments is much needed (Baker et al., 2021) and modeling pest impacts seems to be a natural extension of typical SDM applications, especially given that management of pests is spatially explicit such as our model application. We suggest testing our modeling approach on disturbances caused by other pests in forest or agricultural systems.

Our results contribute to the overall perspective on spatial drivers of budworm defoliation, but also identifies distinctive defoliation patterns that may occur at the leading edge of budworm outbreaks. Our MaxEnt SDM revealed that climatic variables can be more crucial than host condition or forest composition when predicting insect outbreaks at large spatial extents. On the island of Newfoundland, precipitation and temperature were the limiting factors for the distribution of spruce budworm defoliation. The continuously warming climate will further complicate the observation and interpretation of budworm defoliation in the spruce-fir forests in Newfoundland.

ACKNOWLEDGEMENTS

We would like to thank Dan McKenney, Kaitlin DeBoer, Robert LeBlanc, Jeff Motty for their help on data acquisition. This study was a part of an Early Intervention Strategy research project funded by Natural Resources Canada and the Healthy Forest Partnership to SJL, AH, and KEH.

AUTHORS’ CONTRIBUTIONS

All authors contributed to the conception of the article, and B.Z. led the writing with contributions from all authors.

DATA ACCESSIBILITY

The data and R code that support the findings of this study are openly available in Figshare at https://doi.org/10.6084/m9.figshare.c.6147219.v1. Restrictions apply to the availability of spruce budworm defoliation and forest inventory raw data, which were used under data sharing agreements.

**References:**

Araújo, M.B. & Peterson, A.T. (2012) Uses and misuses of bioclimatic envelope modeling. *Ecology,* **93,** 1527-1539. <https://doi.org/10.1890/11-1930.1>

Arsenault, A., LeBlanc, R., Earle, E., Brooks, D., Clarke, B., Lavigne, D. & Royer, L. (2016) Unravelling the past to manage Newfoundland’s forests for the future. *The Forestry Chronicle,* **92,** 487-502. <https://doi.org/10.5558/tfc2016-085>

Baker, D.J., Maclean, I.M.D., Goodall, M. & Gaston, K.J. (2021) Species distribution modelling is needed to support ecological impact assessments. *Journal of Applied Ecology,* **58,** 21-26. <https://doi.org/10.1111/1365-2664.13782>

Baker, D.J., Maclean, I.M.D., Goodall, M. & Gaston, K.J. (2022) Correlations between spatial sampling biases and environmental niches affect species distribution models. *Global Ecology and Biogeography,* **31,** 1-13. <https://doi.org/10.1111/geb.13491>

Bellinger, R.G., Ravlin, F.W. & Mcmanus, M.L. (1989) Forest edge effects and their influence on gypsy moth (Lepidoptera: Lymantriidae) egg mass distribution. *Environmental Entomology,* **18,** 840-843. <https://doi.org/10.1093/ee/18.5.840>

Bernes, C., Bullock, J.M., Jakobsson, S., Rundlöf, M., Verheyen, K. & Lindborg, R. (2017) How are biodiversity and dispersal of species affected by the management of roadsides? A systematic map. *Environmental Evidence,* **6,** 24. <https://doi.org/10.1186/s13750-017-0103-1>

Bouchard, M. & Auger, I. (2014) Influence of environmental factors and spatio-temporal covariates during the initial development of a spruce budworm outbreak. *Landscape Ecology,* **29,** 111-126. <https://doi.org/10.1007/s10980-013-9966-x>

Bradie, J. & Leung, B. (2017) A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *Journal of Biogeography,* **44,** 1344-1361. <https://doi.org/10.1111/jbi.12894>

Burns, P., Clark, M., Salas, L., Hancock, S., Leland, D., Jantz, P., Dubayah, R. & Goetz, S.J. (2020) Incorporating canopy structure from simulated GEDI lidar into bird species distribution models. *Environmental Research Letters,* **15,** 095002. <https://doi.org/10.1088/1748-9326/ab80ee>

Candau, J.N. & Fleming, R.A. (2005) Landscape-scale spatial distribution of spruce budworm defoliation in relation to bioclimatic conditions. *Canadian Journal of Forest Research,* **35,** 2218-2232. <https://doi.org/10.1139/x05-078>

Carroll, W. (1956) Investigations on the spruce budworm, balsam woolly aphid and hemlock looper conducted by the Forest Insect Survey in Newfoundland, 1955. Interim Rep. 1955, Department of Agriculture.

Charney, N.D., Record, S., Gerstner, B.E., Merow, C., Zarnetske, P.L. & Enquist, B.J. (2021) A test of species distribution model transferability across environmental and geographic space for 108 western North American tree species. *Frontiers in Ecology and Evolution,* **9,** 689295. <https://doi.org/10.3389/fevo.2021.689295>

Chen, C., Weiskittel, A., Bataineh, M. & MacLean, D.A. (2017) Evaluating the influence of varying levels of spruce budworm defoliation on annualized individual tree growth and mortality in Maine, USA and New Brunswick, Canada. *Forest Ecology and Management,* **396,** 184-194. <https://doi.org/10.1016/j.foreco.2017.03.026>

Cooke, B.J., Nealis, V.G. & Régnière, J. (2007) Insect defoliators as periodic disturbances in northern forest ecosystems. in E. Johnson & K.Miyanishi, (Eds.), *Plant disturbance ecology: the process and the response* (pp. 487-525).Elsevier Academic Press.

Department of Fisheries Forestry and Agriculture of Newfoundland and Labrador (2022) Forest inventory program.

Dickison, R.B.B., Haggis, M.J. & Rainey, R.C. (1983) Spruce budworm moth flight and storms: case study of a cold front system. *Journal of Applied Meteorology and Climatology,* **22,** 278-286. [https://doi.org/10.1175/1520-0450(1983)022<0278:sbmfas>2.0.co;2](https://doi.org/10.1175/1520-0450(1983)022%3c0278:sbmfas%3e2.0.co;2)

Dingle, H., Zalucki, M.P., Rochester, W.A. & Armijo-Prewitt, T. (2005) Distribution of the monarch butterfly, *Danaus plexippus* (L.) (Lepidoptera: Nymphalidae), in western North America. *Biological Journal of the Linnean Society,* **85,** 491-500. <https://doi.org/10.1111/j.1095-8312.2005.00512.x>

Eskafi, F.M. & Fernandez, A. (1990) Larval–pupal mortality of Mediterranean fruit fly (Diptera: Tephritidae) from interaction of soil, moisture, and temperature. *Environmental Entomology,* **19,** 1666-1670. <https://doi.org/10.1093/ee/19.6.1666>

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D. & Alsdorf, D. (2007) The shuttle radar topography mission. *Reviews of Geophysics,* **45**. <https://doi.org/10.1029/2005RG000183>

Fitzpatrick, M.C., Lachmuth, S. & Haydt, N.T. (2021) The ODMAP protocol: a new tool for standardized reporting that could revolutionize species distribution modeling. *Ecography,* **44,** 1067-1070. <https://doi.org/10.1111/ecog.05700>

Forman, R.T.T. & Alexander, L.E. (1998) Roads and their major ecological effects. *Annual Review of Ecology and Systematics,* **29,** 207-231. <https://doi.org/10.1146/annurev.ecolsys.29.1.207>

Gobeyn, S., Mouton, A.M., Cord, A.F., Kaim, A., Volk, M. & Goethals, P.L.M. (2019) Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. *Ecological Modelling,* **392,** 179-195. <https://doi.org/10.1016/j.ecolmodel.2018.11.013>

Gray, D.R. (2008) The relationship between climate and outbreak characteristics of the spruce budworm in eastern Canada. *Climatic Change,* **87,** 361-383. <https://doi.org/10.1007/s10584-007-9317-5>

Gray, D.R. (2013) The influence of forest composition and climate on outbreak characteristics of the spruce budworm in eastern Canada. *Canadian Journal of Forest Research,* **43,** 1181-1195. <https://doi.org/10.1139/cjfr-2013-0240>

Grendár, M. & Grendár, M. (2001) Maximum entropy: clearing up mysteries. *Entropy,* **3,** 58-63. <https://doi.org/10.3390/e3020058>

Hajek, A.E., Tobin, P.C. & Haynes, K.J. (2015) Replacement of a dominant viral pathogen by a fungal pathogen does not alter the collapse of a regional forest insect outbreak. *Oecologia,* **177,** 785-797. <https://doi.org/10.1007/s00442-014-3164-7>

Hastie, T., Tibshirani, R., Friedman, J.H. & Friedman, J.H. (2009) The elements of statistical learning: data mining, inference, and prediction. Springer.

Hennigar, C.R., MacLean, D.A., Quiring, D.T. & Kershaw, J.A. (2008) Differences in spruce budworm defoliation among balsam fir and white, red, and black spruce. *Forest Science,* **54,** 158-166. <https://doi.org/10.1093/forestscience/54.2.158>

Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2021) dismo: Species Distribution Modeling. <https://CRAN.R-project.org/package=dismo>

Jardon, Y. & Doyon, F. (2003) Balsam fir stand dynamics after insect outbreak disturbances in western Newfoundland ecoregion (Corner Brook Subregion). Western Newfoundland Model Forest, Inc.

Johnson, D.M., Liebhold, A.M., Bjørnstad, O.N. & Mcmanus, M.L. (2005) Circumpolar variation in periodicity and synchrony among gypsy moth populations. *Journal of Animal Ecology,* **74,** 882-892. <https://doi.org/10.1111/j.1365-2656.2005.00980.x>

Kass, J.M., Muscarella, R., Galante, P.J., Bohl, C.L., Pinilla-Buitrago, G.E., Boria, R.A., Soley-Guardia, M. & Anderson, R.P. (2021) ENMeval 2.0: Redesigned for customizable and reproducible modeling of species’ niches and distributions. *Methods in Ecology and Evolution,* **12,** 1602-1608. <https://doi.org/10.1111/2041-210X.13628>

Kettela, E.G. (1983) A cartographic history of spruce budworm defoliation 1967 to 1981 in eastern North America. Info. Rep. DPC-X-14, Canadian Forestry Service.

Lemoine, N.P. & Burkepile, D.E. (2012) Temperature-induced mismatches between consumption and metabolism reduce consumer fitness. *Ecology,* **93,** 2483-2489. <https://doi.org/10.1890/12-0375.1>

Leroux, S.J., Charron, L., Hermanutz, L. & Feltham, J. (2021) Cumulative effects of spruce budworm and moose herbivory on boreal forest ecosystems. *Functional Ecology,* **35,** 1448-1459. <https://doi.org/10.1111/1365-2435.13805>

Li, M., MacLean, D.A., Hennigar, C.R. & Ogilvie, J. (2020) Previous year outbreak conditions and spring climate predict spruce budworm population changes in the following year. *Forest Ecology and Management,* **458,** 117737. <https://doi.org/10.1016/j.foreco.2019.117737>

Ludwig, D., Jones, D.D. & Holling, C.S. (1978) Qualitative analysis of insect outbreak systems: the spruce budworm and forest. *The Journal of Animal Ecology,* **47,** 315-332. <https://doi.org/10.2307/3939>

MacKinnon, W.E. & MacLean, D.A. (2003) The influence of forest and stand conditions on spruce budworm defoliation in New Brunswick, Canada. *Forest Science,* **49,** 657-667. <https://doi.org/10.1093/forestscience/49.5.657>

MacLean, D.A. (2016) Impacts of insect outbreaks on tree mortality, productivity, and stand development. *The Canadian Entomologist,* **148,** S138-S159. <https://doi.org/10.4039/tce.2015.24>

MacLean, D.A., Amirault, P., Amos-Binks, L., Carleton, D., Hennigar, C., Johns, R. & Régnière, J. (2019) Positive results of an early intervention strategy to suppress a spruce budworm outbreak after five years of trials. *Forests,* **10,** 448. <https://doi.org/10.3390/f10050448>

MacLean, D.A. & MacKinnon, W.E. (1997) Effects of stand and site characteristics on susceptibility and vulnerability of balsam fir and spruce to spruce budworm in New Brunswick. *Canadian Journal of Forest Research,* **27,** 1859-1871. <https://doi.org/10.1139/x97-146>

Magnussen, S., Boudewyn, P. & Alfaro, R. (2004) Spatial prediction of the onset of spruce budworm defoliation. *The Forestry Chronicle,* **80,** 485-494. <https://doi.org/10.5558/tfc80485-4>

McKenney, D.W., Hutchinson, M.F., Papadopol, P., Lawrence, K., Pedlar, J., Campbell, K., Milewska, E., Hopkinson, R.F., Price, D. & Owen, T. (2011) Customized spatial climate models for North America. *Bulletin of the American Meteorological Society,* **92,** 1611-1622. <https://doi.org/10.1175/2011bams3132.1>

Meades, W. (2008) Vegetation of Newfoundland. Tech. Rep. No. 15, Conservation of Arctic Flora and Fauna.

Merow, C., Smith, M.J. & Silander Jr, J.A. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography,* **36,** 1058-1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>

Mönkkönen, M. & Mutanen, M. (2003) Occurrence of moths in boreal forest corridors. *Conservation Biology,* **17,** 468-475. <https://doi.org/10.1046/j.1523-1739.2003.01414.x>

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2014) Where is positional uncertainty a problem for species distribution modelling? *Ecography,* **37,** 191-203. <https://doi.org/10.1111/j.1600-0587.2013.00205.x>

Natural Resources Canada (2020) Forest classification. <https://www.nrcan.gc.ca/our-natural-resources/forests/sustainable-forest-management/measuring-and-reporting/forest-classification/13179>

Natural Resources Canada (2021) Lakes, Rivers and Glaciers in Canada - CanVec Series - Hydrographic Features. <https://ftp.maps.canada.ca/pub/nrcan_rncan/vector/canvec/shp/Hydro/>

Nealis, V.G. & Régnière, J. (2009) Risk of dispersal in western spruce budworm. *Agricultural and Forest Entomology,* **11,** 213-223. <https://doi.org/10.1111/j.1461-9563.2008.00414.x>

Nenzén, H.K., Filotas, E., Peres-Neto, P. & Gravel, D. (2017) Epidemiological landscape models reproduce cyclic insect outbreaks. *Ecological Complexity,* **31,** 78-87. <https://doi.org/10.1016/j.ecocom.2017.04.004>

Pablo, M., Giovanni, L., Santilli, G., Huang, W. & Zappacosta, D. (2019) Maxent model application for tree pests monitoring. in IEEE international geoscience and remote sensing symposium (pp. 6664-6667). <https://doi.org/10.1109/IGARSS.2019.8898056>

Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E. & Blair, M.E. (2017) Opening the black box: an open-source release of Maxent. *Ecography,* **40,** 887-893. <https://doi.org/10.1111/ecog.03049>

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling,* **190,** 231-259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography,* **31,** 161-175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>

Pothier, D., Elie, J.-G., Auger, I., Mailly, D. & Gaudreault, M. (2012) Spruce budworm-caused mortality to balsam fir and black spruce in pure and mixed conifer stands. *Forest Science,* **58,** 24-33. <https://doi.org/10.5849/forsci.10-110>

Pureswaran, D.S., Johns, R., Heard, S.B. & Quiring, D. (2016) Paradigms in eastern spruce budworm (Lepidoptera: Tortricidae) population ecology: a century of debate. *Environmental Entomology,* **45,** 1333-1342. <https://doi.org/10.1093/ee/nvw103>

Pureswaran, D.S., Neau, M., Marchand, M., De Grandpré, L. & Kneeshaw, D. (2019) Phenological synchrony between eastern spruce budworm and its host trees increases with warmer temperatures in the boreal forest. *Ecology and evolution,* **9,** 576-586. <https://doi.org/10.1002/ece3.4779>

R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>

Radosavljevic, A. & Anderson, R.P. (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography,* **41,** 629-643. <https://doi.org/10.1111/jbi.12227>

Régnière, J., Cooke, B.J., Béchard, A., Dupont, A. & Therrien, P. (2019) Dynamics and management of rising outbreak spruce budworm populations. *Forests,* **10,** 748. <https://doi.org/10.3390/f10090748>

Ritter, P. (1987) A vector-based slope and aspect generation algorithm. *Photogrammetric Engineering and Remote Sensing,* **53,** 1109-1111.

Riva, F., Acorn, J.H. & Nielsen, S.E. (2018) Narrow anthropogenic corridors direct the movement of a generalist boreal butterfly. *Biology Letters,* **14,** 20170770. <https://doi.org/10.1098/rsbl.2017.0770>

Royama, T., Eveleigh, E.S., Morin, J.R.B., Pollock, S.J., M Carthy, P.C., M Dougall, G.A. & Lucarotti, C.J. (2017) Mechanisms underlying spruce budworm outbreak processes as elucidated by a 14-year study in New Brunswick, Canada. *Ecological Monographs,* **84,** 600-631. <https://doi.org/10.1002/ecm.1270>

Shcheglovitova, M. & Anderson, R.P. (2013) Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes. *Ecological Modelling,* **269,** 9-17. <https://doi.org/10.1016/j.ecolmodel.2013.08.011>

Statistics Canada (2021) National Road Network - NRN - GeoBase Series. <https://geo.statcan.gc.ca/nrn_rrn/nl/nrn_rrn_nl_SHAPE.zip>

Sterner, T.E. & Davidson, A.G. (1980) Forest insect and disease conditions in Canada. Catal. No. Fo21-1/1980E, Canadian Forestry Service.

Talerico, R.L. (1983) Summary of life history and hosts of the spruce budworm. Gen. Tech. Rep. NE-85, US Forest Service, Northeastern Forest Experimental Station.

Tonkin, J.D., Altermatt, F., Finn, D.S., Heino, J., Olden, J.D., Pauls, S.U. & Lytle, D.A. (2018) The role of dispersal in river network metacommunities: Patterns, processes, and pathways. *Freshwater Biology,* **63,** 141-163. <https://doi.org/10.1111/fwb.13037>

Volney, W.J.A. & Fleming, R.A. (2007) Spruce budworm (Choristoneura spp.) biotype reactions to forest and climate characteristics. *Global Change Biology,* **13,** 1630-1643. <https://doi.org/10.1111/j.1365-2486.2007.01402.x>

Zhang, B., Leroux, S.J., Bowden, J.J., Hargan, K.E., Hurford, A. & Moise, E.R.D. (2022) Datasets and R codes: species distribution model identifies influence of climatic constraints at the leading edge of a native insect outbreak. *figshare*. <https://doi.org/10.6084/m9.figshare.c.6147219.v1>

Zhang, B., MacLean, D.A., Johns, R.C. & Eveleigh, E.S. (2018) Effects of hardwood content on balsam fir defoliation during the building phase of a spruce budworm outbreak. *Forests,* **9,** 530. <https://doi.org/10.3390/f9090530>

Zhang, B., MacLean, D.A., Johns, R.C., Eveleigh, E.S. & Edwards, S. (2020) Hardwood-softwood composition influences early-instar larval dispersal mortality during a spruce budworm outbreak. *Forest Ecology and Management,* **463,** 118035. <https://doi.org/10.1016/j.foreco.2020.118035>

**Table 1.** Predictor variables included in MaxEnt model to determine probability of severe defoliation from spruce budworm.

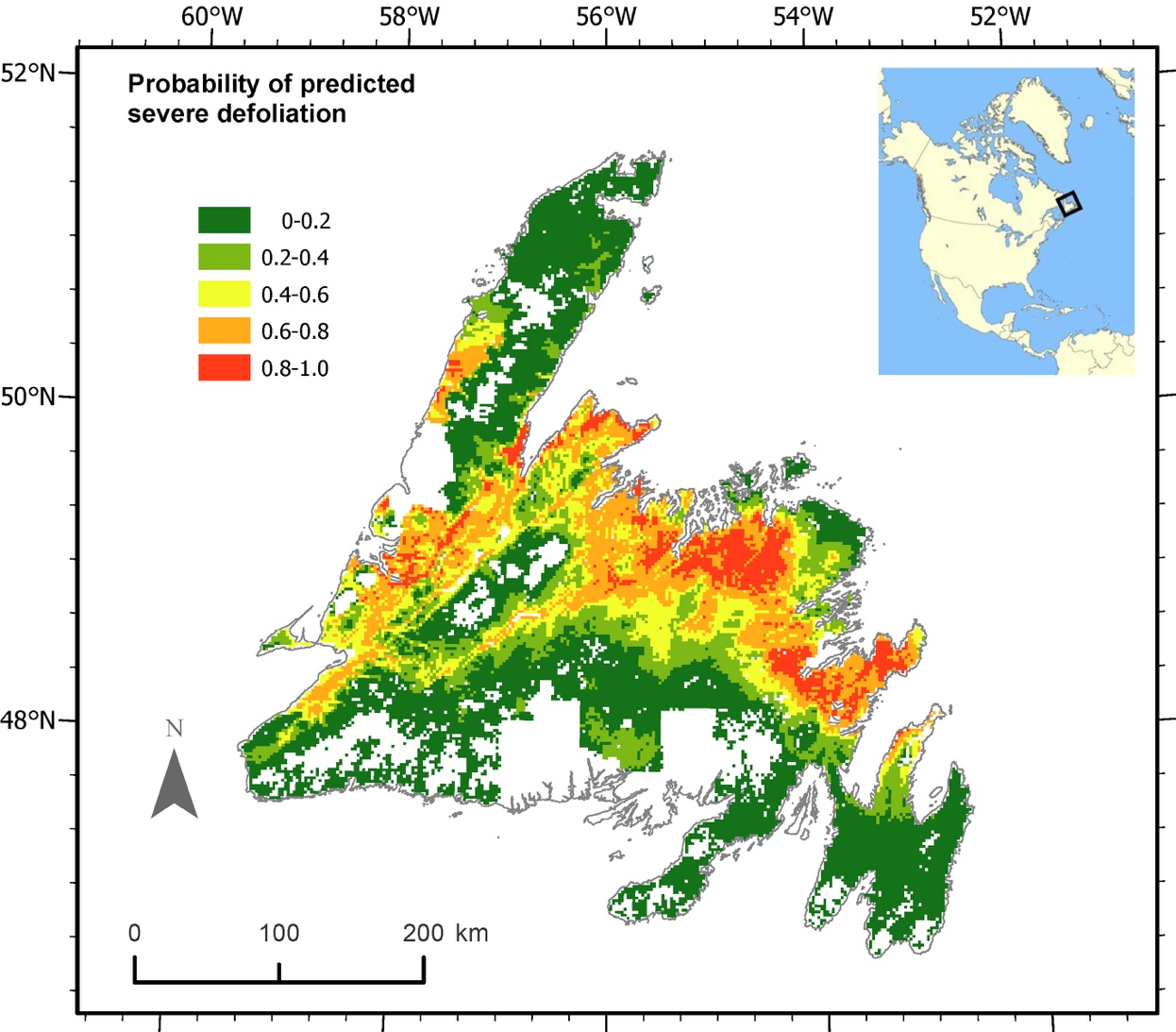
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Predictor variables** | **Description** | **Unit** | **Original data resolution** | **Data source** |
| MaxT0506 | Maximum temperature in May and Jun. | °C | 60 arcsecond | Natural Resources Canada: Historical monthly climate grids |
| MaxT0607 | Maximum temperature in Jun. and Jul. | °C |
| MaxT0708 | Maximum temperature in Jul. and Aug. | °C |
| MinT0506 | Minimum temperature in May and Jun. | °C |
| MinT0607 | Minimum temperature in Jun. and Jul. | °C |
| MinT0708 | Minimum temperature in Jul. and Aug. | °C |
| Prcp0506 | precipitation in May and Jun. | mm |
| Prcp0607 | precipitation in Jun. and Jul. | mm |
| Prcp0708 | precipitation in Jul. and Aug. | mm |
| Sg\_date | Julian date of start of growing season | date |
| Sg\_gdd\_p1 | Cumulative degree days above 5 °C for the three months prior to the start of growing season | °C•d |
| Sg\_gdd\_p2 | Cumulative degree days above 5 °C for the six weeks after the start of growing season | °C•d |
| Elevation | Elevation | m | 90 m | DEM, Shuttle Radar Topography Mission |
| Slope | Slope | ° |
| Aspect | Aspect | ° |
| Dist\_road | Nearest distance to major road features | m | 2 km (derived) | Statistics Canada: National Road Network GeoBase Series |
| Dist\_river | Nearest distance to major riverine features | m | 2 km (derived) | Natural Resources Canada: CanVec Series - Hydrographic Features |
| Forest\_ species\* | Forest tree species |  |  | Government of Newfoundland and Labrador |

\* Categorical predictor, spatial polygon layer

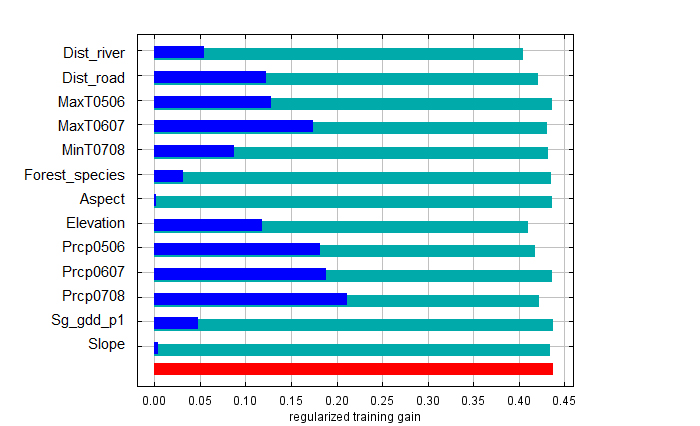
**Table 2.** Variable importance for the optimal MaxEnt model predicting probability of severe defoliation from spruce budworm.

|  |  |  |
| --- | --- | --- |
| **Variable\*** | **Percent contribution (%)** | **Permutation importance**  **(%)** |
| Prcp0506 | 25.7 | 24.7 |
| MaxT0607 | 23.1 | 8.0 |
| Dist\_road | 14.2 | 7.4 |
| Elevation | 13.3 | 21.1 |
| Prcp0708 | 7.7 | 10.2 |
| Dist\_river | 6.0 | 18.5 |
| MaxT0506 | 3.6 | 0.9 |
| Prcp0607 | 2.4 | 1.7 |
| Slope | 1.9 | 1.3 |
| MinT0708 | 1.1 | 4.8 |
| Forest\_species | 0.6 | 0.9 |
| Aspect | 0.3 | 0.2 |
| Sg\_gdd\_p1 | 0.1 | 0.1 |

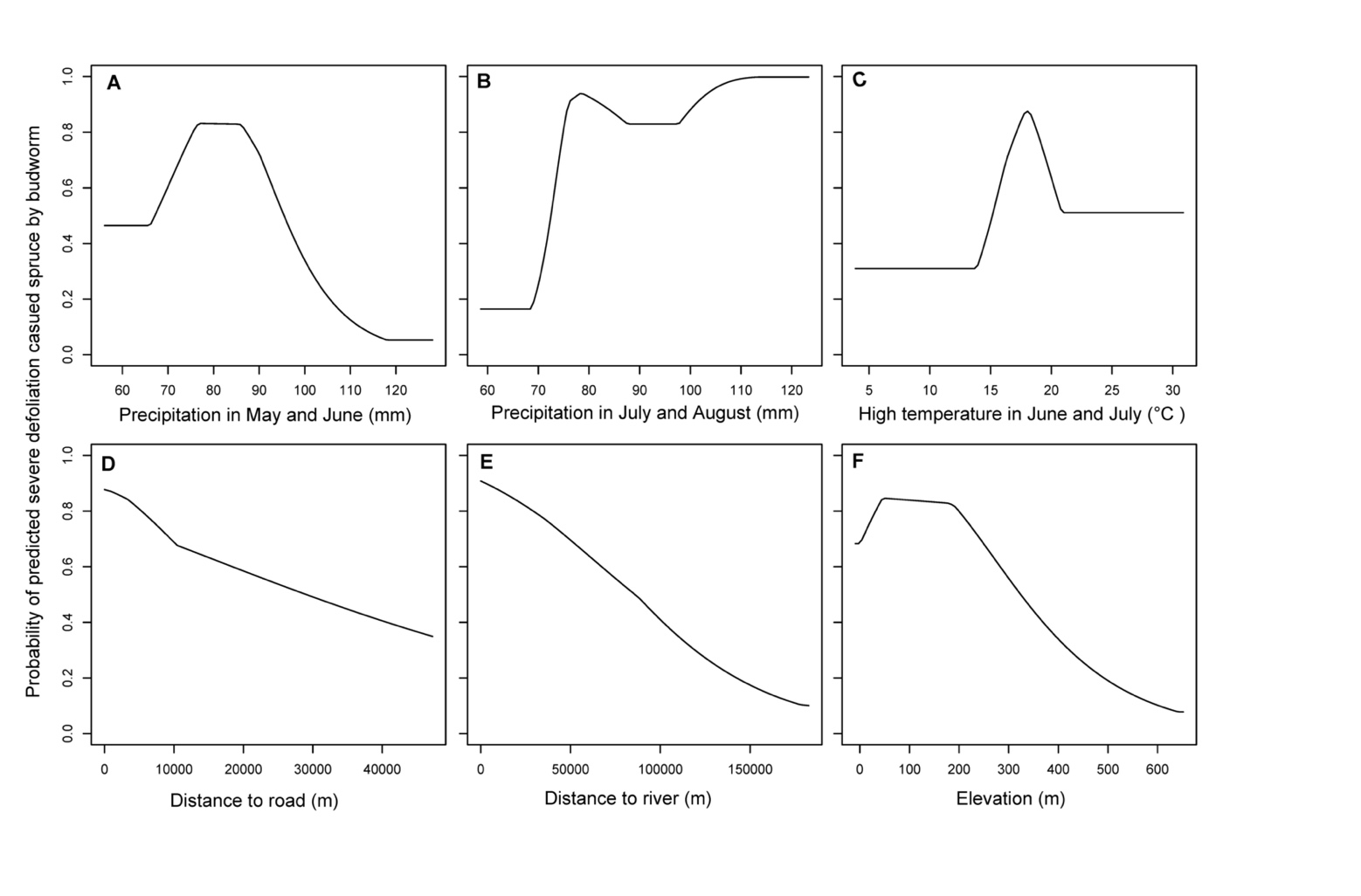
\* See Table 1 for variable definition



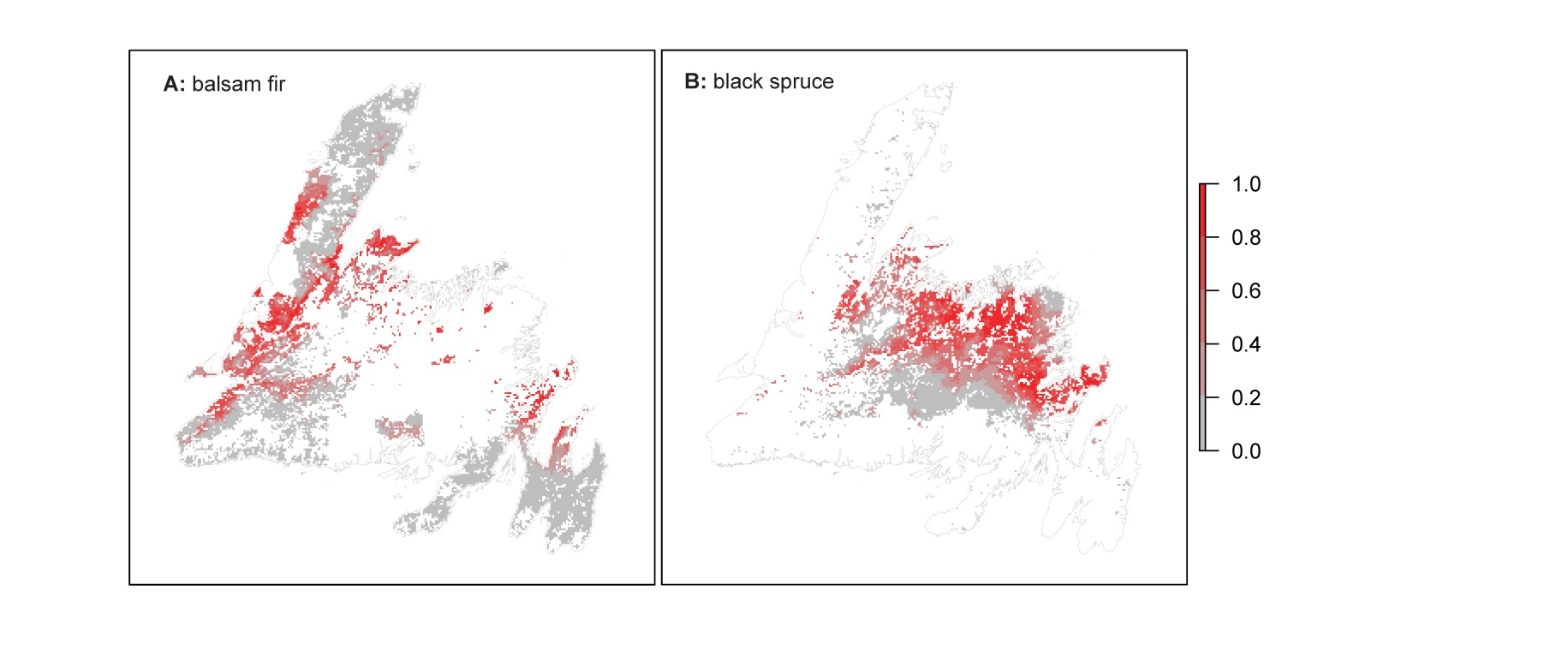
**Figure 1.** Probability of predicted severe defoliation (in five bins of 0.2 step, defined as three consecutive years of moderate to severe aerial-surveyed defoliation) from spruce budworm on the island of Newfoundland. White areas on the land mass are areas of non-commercial forest as determined by the Department of Fisheries Forestry and Agriculture of Newfoundland and Labrador. Inset map shows the general location of the island of Newfoundland in North America.



**Figure 2.** Jackknife test of variable importance for the 13 predictor variables on severe defoliation from spruce budworm. The change in regularized model gain is illustrated when each variable was omitted (cyan) or was used in isolation (blue). The gain when all variables were included is shown in red. See Table 1 for variable definitions.



**Figure 3.** Relationship between probability of severe defoliation caused by spruce budworm and six selected predictor variables. Each curve depicts the relationship between the response and the predictor when all other predictors were hold at their average values.



**Figure 4.** Probability of predicted severe defoliation (in five bins of 0.2 step, defined as three consecutive years of moderate to severe aerial-surveyed defoliation) from spruce budworm in balsam fir (A) and black spruce (B) forests on the island of Newfoundland.