Biodiversity Informatics, 16, 2021, pp. 1-19

PREDICTING MULTI-SPECIES BARK BEETLE (COLEOPTERA: CURCULIONIDAE: SCOLYTINAE) OCCURRENCE IN ALASKA: OPEN-ACCESS BIG GIS-DATA MINING TO PROVIDE ROBUST INFERENCE

Khodabakhsh Zabihi[a,b,*], Falk Huettmann[a], and Brian D. Young[c]
[a] *EWHALE Lab, Biology and Wildlife Department, Institute of Arctic Biology, University of Alaska Fairbanks (UAF), Fairbanks, AK 99775-7000, USA*
[b] *EXTEMIT-K, Faculty of Forest and Wood Sciences, Czech University of Life Sciences (CULS) Prague, Kamýcká 129, 165 00 Praha 6-Suchdol, Czech Republic*
[c] *STEM Department, Landmark College, Putney, VT 05346, USA*

Abstract. Native bark beetles (Coleoptera: Curculionidae: Scolytinae) are a multi-species complex that ranks among the key disturbances of coniferous forests of western North America. Many landscape-level variables are known to influence beetle outbreaks, such as suitable climatic conditions, spatial arrangement of incipient populations, topography, abundance of mature host trees, and disturbance history that includes former outbreaks and fire. We assembled open-access data for understanding the ecology of bark beetles in Alaska. We used boosted classification and regression trees as a machine-learning data-mining algorithm to predict relationships between 838 occurrence records of 68 bark beetle species and 14 environmental variables, compared to pseudo-absence locations across Alaska. Environmental variables included topography- and climate-related predictors as well as feature proximities and anthropogenic factors. We were able to model, predict, and map multi-species bark beetle occurrences across Alaska at 1-km spatial resolution: about 16% of the mixed forest and 59% of evergreen forest are expected to be occupied by the bark beetles based on current climatic conditions and biophysical landscape attributes. The open-access dataset that we prepared, and the machine learning modeling approach that we used, can provide a foundation for future research not only on scolytines but for other multi-species questions of concern, such as forest defoliators, and wildlife species assemblages worldwide.

Key words: Scolytinaes, Pest insects, Outbreaks, Boosted classification and regression tree, Forest ecology, Spatial modeling, Machine-learning algorithm.

INTRODUCTION

Historic background

Forests are a major terrestrial ecosystem of global relevance encompassing about 30% of the land area on the earth (Schmitt et al. 2009; Liu et al. 2018). Forest ecosystems play a critical role in ecological services and reducing the threat of natural disasters, such as floods, droughts, and landslides (Uy and Shaw 2012). At the global scale, forests can mitigate climate change impacts via carbon sequestration and rainfall infiltration that safeguards drinking water supplies (Uy and Shaw 2012). Leaves and needles play major roles in such equations, but they can be consumed or defoliated by insects, specifically bark beetle species.

Native bark beetles (Coleoptera: Curculionidae: Scolytinae) consist of several species complexes acting together as a community, and constitute one of the key disturbances of coniferous forests of western North America (Bentz et al. 2010; Seidl et al. 2014; Morris et al. 2016). In spite of traditional single-species views and subsequent efforts regarding the ‘bark beetle problem’ (e.g., Bentz et al. 2009), bark beetles can be perceived as a community organism that together plays a wider role as ecosystem engineers (Martikainen et al. 1999; Jonasova and Pracha 2004; Müller et al. 2008). Since 1990, bark beetles have killed billions of coniferous trees across millions of hectares in North American forest ecosystems from Mexico to Alaska (Raffa et al. 2008; Bentz et al. 2009; Bentz et al. 2010).

Although insects such as bark beetles constitute an inherent element of ecosystems, initiating succession and providing food sources for predators such as woodpeckers, bark beetle outbreaks can cause severe direct and indirect impacts on forest ecosystems and species that are dependent on or that interact with forests. Direct effects include tree mortality

---

[*] Corresponding author

and the ensuing changes in forest composition and structure, increased chance of wildfire owing to creation of broad areas with large quantities of dead trees, and a greater chance of windthrow of living residual trees (Schowalter 2012). Indirect impacts can include, for example, reduction of timber value resulting from accelerated salvage harvest activities following outbreaks, decreased carbon sequestration, degradation of fish and wildlife habitats, and reduced recreational capacity (Schowalter 2012). Rapid widespread tree mortality leaves longer-term impacts on structural and functional aspects of ecosystems with ongoing influences on climate, habitats, species, and land use (Kurz et al. 2008; McDowell et al. 2008; Bentz et al. 2010).

From the ecological and biodiversity point of view, insects such as bark beetles represent an inherent part of the wider ecological and global web, but usually are seen as pests to be gotten rid of. An increasing body of evidence suggests that bark beetle outbreaks and post-outbreak conditions form part of the wider successional pattern in a landscape, and can have some positive impacts on ecosystem services. For example, tree loss can increase water yield (Bearup et al. 2014; Morris et al. 2016), improve foraging for livestock and wildlife species, and thus increase the population of some big game species and provide more wildlife viewing and hunting opportunities (Saab et al. 2014; Morris et al. 2016). Bark beetles also serve as major prey species for many insectivores, particularly for woodpeckers (Bonnot et al. 2009; Saab et al. 2014). In addition, bark beetles have fascinating ecological relationships with various fungal species (Munro et al. 2019).

Several studies found that bark beetle outbreaks appear to occur when factors such as drought, aging, and density attenuate trees' resistance abilities against bark beetle attacks (Christiansen and Bakke 1988; Raffa 1988; Fettig et al. 2007; Bentz et al. 2010). Although tree- and stand-level characteristics such as tree vigor and size and stand density are critical for local infestation under endemic conditions (Raffa and Berryman 1983; Simard et al. 2011), landscape-level factors influence the transition of infestation from endemic to epidemic conditions: a local eruption to regional outbreaks (Wallin and Raffa 2004; Raffa et al. 2008; Simard et al. 2011). Strong correlations with environmental parameters provide local patches of suitable habitats that enhance potential for scolytine growth; this growth can lead to outbreaks under certain circumstances (Aukema et al. 2006; Raffa et al. 2008). The spread of outbreaks is spatially and temporally autocorrelated regardless of host tree vigor (Aukema et al. 2006; Aukema et al. 2008; Simard et al. 2011).

Several landscape-level variables facilitate scolytine beetle outbreaks, such as suitable climatic conditions, spatial arrangement of incipient populations, topography, abundance of mature host trees, and disturbance history that includes outbreak and fire history (Aukema et al. 2006). Identifying these variables could aid in understanding the epidemiology and ecology of these species and improve management strategies for outbreak control (Simard et al. 2011).

Rationale

As bark beetle issues are ecologically complex and multidimensional, computational approaches such as data mining, machine learning, and geographic information system (GIS) could be helpful to study and map the current and future spatial distributions of bark beetles. Use of these methods is on the rise (Bhattacharya 2013), which is particularly needed in forest ecology and management. They have been recently applied in quantitative modeling of macroscale ecological niches of tree species (Prasad 2018), mapping aboveground biomass of trees within the Alaskan Boreal Forest (Young et al. 2018), inference questions regarding wildlife (Huettmann et al. 2018b), and variance assessment of predicted climate-scapes based on topographic variation (Huettmann 2018a).

However, in a high-dimensional multivariate (i.e., 10-20 predictors) investigation, tools such as data-mining and machine-learning have not yet been applied to understand current and future presence of scolytines, considered as a community organism. This point is particularly true for broad-scale prediction and mapping for a large region such as Alaska, which contains part of the largest forest area in the world (the boreal forest) and the world's largest temperate rainforest (Tongass). Machine-learning, in contrast to parametric methods such as regression models, can use many algorithms in ensembles (Humphries et al. 2018). Algorithms are developed to comprehend complex, nonlinear relationships in the data without requiring prior model assumptions such as normally distributed model residuals and freedom from spatial autocorrelation (Huettmann 2018c). Natural phenomena do not necessarily follow known distributions and generally have no linear interactions (Ver Hoef et al. 1993). In addition, in regression models including logistic regression, use of residual analysis as a measure of model fitness is unreliable when modeling a high-dimensional data set, with nonlinear combinations of the variables (Breiman 2001). As such, to fulfill model assumptions and have easily interpretable information in parametric models, a widely used but faulty approach is to weed out less-important predictor variables and consider effects of unmeasured variables as "noise" (Breiman 2001). This approach can lead to wrong conclusions and consequently poor management strategies, more a reflection of the model's mechanism rather than a true emulation of nature (Breiman 2001). In contrast, algorithmic models such as boosted classification and regression tree are not affected by limited and biased sets of predictors or by nonlinear relationships among predictors. This characteristic results in more accurate and informative conclusions (Breiman 2001; Elith et al. 2008; Tyralis 2019), very important for biodiversity conservation and forest management.

The commonly-used, single-species approach of species distribution modeling assumes that species respond individually to environmental gradients. However, species distributions may be influenced by biotic interactions, such as competition, with other species within a community (Araújo and Luoto 2007; Heikkinen et al. 2007; Chapman and Purse 2011). In addition, as different species within a community may have similar responses to environmental changes at regional scales (Golicher et al. 2008; Azeria et al. 2009; Chapman and Purse 2011), community-level analyses of spatial patterns of biodiversity may provide beneficial information (Chapman and Purse 2011) for biodiversity conservation and natural resource management.

For example, several bark beetle species, e.g., *Polygraphus rufipennis* and *Pityophthorus nitidulus*, attack the same host trees (e.g., white spruce and black spruce). In this sense, the common, individual-based approach of modeling and mapping spatial distributions of bark beetle species may not provide comprehensive information required for efficient forest management. In addition, the individual-species-based approach may not help in understanding the broader, overall diversity of bark beetle species.

In this study, we set out to map forest landscapes that could favor presence of multi-species bark beetles, with attention to other, more poorly studied landscape reservoirs, such as Arctic tundra shrublands. The resulting map of bark beetle species occurrence will provide useful resources for forest managers and policy makers to prioritize their forest management measures spatially in a time- and cost-effective manner. We assess many environmental variables and proxies, as well as tabulated metadata for 68 species of bark beetles that together comprise a large dataset that we make openly accessible to the scientific community and the broader public. We seek to develop an example of how large-scale environmental data-mining on a big set of species presence and pseudo-absence locations, machine-learning, and geographic information systems can help to predict and map presence of scolytines, as a community organism. We develop this work in Alaskan landscapes, forested and non-forested, without prior model assumptions or consequent data perturbation resulting from violations of assumptions.

MATERIAL AND METHODS

Study area

The study was conducted across the state of Alaska, United States, covering a total area of ~152 x 10^6 ha (Figure 1) ranging from approximately 54° to 71° N and 130° W to 173° E (Figure 1). Within

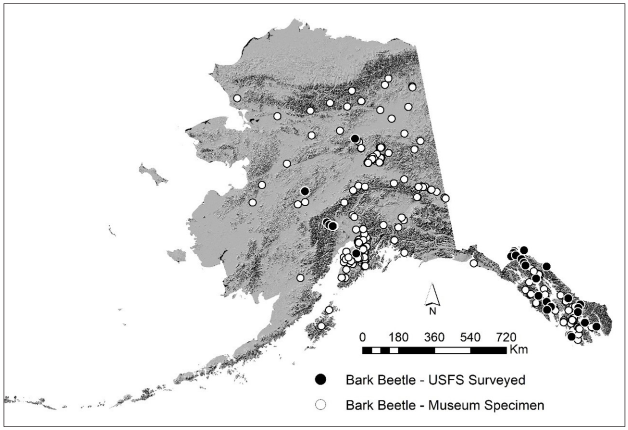


Figure 1. Hillshade map of the study area, the state of Alaska, along with 838 locations of 68 bark beetle species, the presence points for developing and validating the three models, provided by the University of Alaska Museum (UAM; white circles). A different record of 68 locations of 3 bark beetle species, the presence points for additionally assessing/testing the three models, surveyed by U.S. Department of Forest Service (USFS; black circles). See Appendix I for species details.

that area, ~49 x 10^6 ha (~32%) is forested (defined as areas with >10% tree cover; Hutchison 1968, ADFG 2018). Most of the forested area (~43 x 10^6 ha) is in interior Alaska, and is classified as "boreal forest." The remaining forestland, ~5 x 10^6 ha, occurs along the southern coast of Alaska and is classified as "coastal temperate rainforest," which covers the regions of Kodiak, Prince William Sound, and the islands and mainland of South Alaska, including the world's largest temperate rainforest in southeast Alaska (Tongass) and the Chugach national forest in south-central Alaska (ADFG 2018). Coniferous communities of boreal forests are dominated by spruce trees including white spruce (*Picea glauca* (Monech) Voss) and black spruce (*P. mariana* (Miller) Britton, Sterns & Poggenburg) (ADFG 2018). The coastal temperate rainforest includes western hemlock (*Tsuga heterophylla* (Rafinesque) Sargent), Sitka spruce (*P. sitchensis* (Bongard) Carrière), mountain hemlock (*T. mertensiana* (Bongard) Carrière), Alaska yellow cedar (*Chamaecyparis nootkatensis* (D. Don) Farjon & D.K. Harder), western red cedar (*Thuja plicata* (Donn. ex D. Don)), and lodgepole pine (*Pinus contorta* (Douglas ex Loudon)) (ADFG 2018).

Alaska has a glacial history, including many refugia. The glaciated regions of Alaska include 11 mountain ranges: Coast Mountains, Saint Elias Mountains, Chugach Mountains, Kenai Mountains (including Montague Island), Aleutian Range, Wrangell Mountains, Talkeetna Mountains, Alaska Range, Wood River Mountains, Kigluaik Mountains, and the Brooks Range (USGS 2017). The heavily glaciated Alaska range spans an arc ~965 km long that extends from the Alaska-Canada border towards the Alaska Peninsula. Mount McKinley (6195 m), known as Denali, is the highest mountain in the Alaska range and North America (USGS 2017), with extensive alpine and forest cover. Within the study region, large-scale tree mortality is often caused by different insect species and diseases, with bark beetle species being among the important elements (USDA 2008). For example, a massive mortality event that covered >1.3 x 10^6 ha during 1990–1999 was caused by spruce beetles (USDA 2008).

Research design

The year the species were identified varied from 1953 to 2018, with nearly half coming from after 2011 (Figure 2), which justifies our use of 2011 land cover data. Our inferences and conclusions regarding Alaska's potential habitats for the bark beetle organism therefore focus primarily on the recent period after 2010. This study follows a study design that has been applied previously for predictive modeling of the distribution of white spruce [*Picea glauca* (Monech) Voss] and small mammals across Alaska (Ohse et al. 2009, Baltensperger and Huettmann 2015a). We compiled 838 records of 68 bark beetle species, as the occurrence points for the model, provided by the University of Alaska Museum (UAM[1]; Figure 1 and Appendix I) in a separate record for each species. The occurrence data we use are not balanced by species nor do they come from a systematic sampling, but they still serve as a 'presence record' of the bark beetle community in Alaska. In absence of detailed knowledge about many of the bark beetle species of North American landscapes, this compilation helps to shed light on habitat preferences and dynamics of bark beetles in general. Among the pooled bark beetle species, dominant genera were *Dryocoetes* (*N* = 133), *Trypodendron* (*N* = 107), *Ips* (*N* = 104), and *Dendroctonus* (*N* = 83) (Appendix I). The most common species were *Trypodendron lineatum* (*N* = 74), *Dryocoetes affaber* (*N* = 69), *Dendroctonus rufipennis* (*N* = 66), and *Ips perturbatus* (*N* = 52) (Appendix I). The host evergreen trees from which the bark beetle specimens were collected included white spruce (*Picea glauca*), black spruce (*P. mariana*), Sitka spruce (*P. sitchensis*), western hemlock (*Tsuga heterophylla*), lodgepole pine (*Pinus contorta*), mountain hemlock (*T. mertensiana*), Lutz spruce (*P.* x *lutzii*), Tamarack (*Larix laricina*), Yellow cedar (*Cupressus nootkatensis*), and Western redcedar (*Thuja plicata*) (Appendix I). Some specimens were collected within non-evergreen forests. We extracted land cover types underlying 838 beetle locations

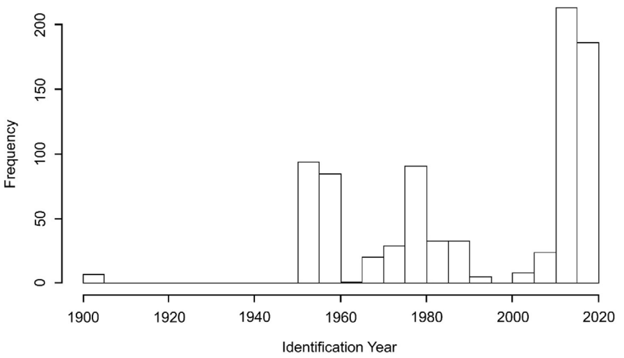


Figure 2. Frequency distributions of the year of observation for the 838 bark beetle presence records.

---

[1] http://arctos.database.museum/SpecimenSearch.cfm.

Table 1. Number and percent of bark beetle specimens collected at different geographic locations calculated by different land cover types derived from NLCD 2011[2]

Nlcd 2011 Designated no.	Land cover Type	Number of bark beetles found	Percent of bark beetles found
11	Open water[3]	47	5.6
12	Perennial ice/snow	5	0.6
21	Developed, open space	7	0.8
22	Developed, low intensity	70	8.4
23	Developed, medium intensity	39	4.7
24	Developed, high intensity	8	1.0
31	Barren land	28	3.3
41	Deciduous forest	68	8.1
42	Evergreen forest	251	30.0
43	Mixed forest	81	9.7
51	Dwarf shrub	62	7.4
52	Shrub/scrub	74	8.8
71	Grassland/herbaceous	16	1.9
90	Woody wetlands	62	7.4
95	Emergent herbaceous wetlands	19	2.3

using the raster map of the National Land Cover Database (NLCD 2011; Table 1).

Following Ohse et al. (2009), Baltensperger and Huettmann (2015a), and Young et al. (2018), we created a background dataset with which to compare bark beetle presences, from across Alaska. We established 5000 random point locations using ArcMap 10.4 (ESRI Inc., Redlands, CA), with a minimum Euclidean distance from each other of 1 km across Alaska. The 838 bark beetle presence locations and 5000 random points were then used to extract underlying pixel values of the environmental variables, as independent variables, for comparison with the binary response variable of presence/absence of bark beetles. We created a lattice point grid with a 1-km Euclidean distance, in QGIS (version 3.4.0), for the entire study area, for a total number of 1,522,655 points. The lattice points were used to extract underlying environmental variable values for mapping and predicting bark beetle occurrences across the study area based on the model.

For additional model assessment, we used 68 independently surveyed locations of 3 bark beetle species—spruce beetle (*Dendroctonus rufipennis*), western balsam bark beetle (*Dryocoetes confusus* Swaine), and northern spruce engraver (*Ips perturbatus*)—with sample sizes of 57, 8, and 3, respectively, from surveys conducted by the U.S. Department of Forest Service in 2016–2017 (Figure 1). The host trees of these observed bark beetle species included white spruce (*Picea glauca* (Monech) Voss), Subalpine fir (*Abies lasiocarpa*), and Sitka spruce (*P. sitchensis* (*Bong.*) Carrière).

Environmental variables

To assess environmental requirements of bark beetles quantitatively, we used 14 independent variables (Table 2). We used aspects of mean monthly maximum and minimum temperature and total precipitation as our climate data. Slope and aspect maps were derived from the available 2 Arc-second digital elevation model (DEM) at 60 m spatial resolution provided by United States Geological Survey (USGS). We used a soil map unit that aggregates different soil components, including multiple soil classes and miscellaneous areas delineated together in spatial polygons. Generally, one to four soil series (or soil taxonomic classes), along with non-soil areas, named as miscellaneous areas, are attached to each map unit (Nauman and Thompson 2013). We also used the land cover raster map of NLCD (2011) that contains the following classes: water, including open water and perennial ice/snow; developed area, including open space, low-, medium-, and high-intensity developed regions; barren land; forest, including deciduous, evergreen, and mixed forest; shrubland, including dwarf shrub and shrub/scrub classes; grassland/herbaceous; emergent herbaceous; and woody and non-woody wetlands (Table 1).

Land status, indicating ownership of the land, was drawn from a vector file provided by the Bureau of Land Management (2013), and had the following categories: private or municipal, state, Bureau of Land Management, native, National Park Service,

---

[2] https://mrlc.gov/data/legends/national-land-cover-database-2011-nlcd2011-legend.

[3] Open water is a pixel and includes coastal and island locations.

Table 2. Description of the environmental variables and additional dataset used to develop and test the models, the three model results as raster maps, and their open-access sources.

Variable	**Type**	**Spatial resolution**	**Designated name**	**Original source**	**Secondary source**
Soil map unit	Vector and categorical	N/A	Soilakab	*https://mrlc.gov/data*	*http://hdl.handle.net/11122/10870*
Land status	Vector and categorical	N/A	LandStatakab	*https://sdms.ak.blm.gov/sdms/*	*http://hdl.handle.net/11122/10871*
National land cover data in 2011	Raster and categorical	30 m	LandCo11akab	*https://mrlc.gov/data*	*http://hdl.handle.net/11122/10872*
Mean annual precipitation in 2010	Raster and continuous	60 m	Tempakab	*https://uaf-snap.org/*	*http://hdl.handle.net/11122/10873*
Mean annual temperature in 2010	Raster and continuous	60 m	Precipakab	*https://uaf-snap.org/*	*http://hdl.handle.net/11122/10874*
Digital Elevation Model (DEM)	Raster and continuous	60 m	DEM60makab	*https://ned.usgs.gov*	*http://hdl.handle.net/11122/10875*
Slope	Raster and continuous	60 m	Slope60makab	-	*http://hdl.handle.net/11122/10876*
Aspect	Raster and continuous	60 m	Aspect60makab	-	*http://hdl.handle.net/11122/10877*
Euclidean distance to coastline	Raster and continuous	60 m	DistCoast	*https://gis.data.alaska.gov*	*http://hdl.handle.net/11122/10878*
Euclidean distance to lakes and rivers	Raster and continuous	60 m	DistLakeRiver	*https://usgs.gov/*	*http://hdl.handle.net/11122/10879*
Euclidean distance to drainage network	Raster and continuous	60 m	DistDriNet	*https:// usgs.gov/*	*http://hdl.handle.net/11122/10880*
Euclidean distance to towns	Raster and continuous	100 m	DistTowns	*https://gis.data.alaska.gov/*	*http://hdl.handle.net/11122/10881*
Euclidean distance to main roads	Raster and continuous	100 m	DistRoads	*https://gis.data.alaska.gov/*	*http://hdl.handle.net/11122/10882*
Euclidean distance to infrastructure	Raster and continuous	100 m	DistInfrastruct	*https://gis.data.alaska.gov/*	*http://hdl.handle.net/11122/10883*
Model results	**Type**	**Spatial resolution**	**Designated name**	**Original source**	**Secondary source**
Model 1 (Appendix III)	Raster and continuous	1 km	Model1	-	*http://hdl.handle.net/11122/10921*
Model 2 (Figure 3)	Raster and continuous	1 km	Model2	-	*http://hdl.handle.net/11122/10922*
Model 3 (Appendix IV)	Raster and continuous	1 km	Model3	-	*http://hdl.handle.net/11122/10923*
Binary map (Figure 6)	Raster and categorical	1 km	BinaryMap	-	*http://hdl.handle.net/11122/10924*
Final map (Figure 7)	Raster and categorical	1 km	FinalMap	-	*http://hdl.handle.net/11122/10925*
Additional data	**Type**	**Spatial resolution**	**Designated name**	**Original source**	**Secondary source**
Bark Beetle presence points	Vector and non-categorical	N/A	BBeetleakab	*http://arctos.database.museum/*	*http://hdl.handle.net/11122/10927*
Bark Beetle pseudo-absence points	Vector and non-categorical	N/A	RaPointsakab	-	*http://hdl.handle.net/11122/10928*
1-km space grid points	Vector and non-categorical	N/A	GridPointsakab	-	*http://hdl.handle.net/11122/10929*
Bark Beetle assessment/test points	Vector and non-categorical	N/A	BBeetleValidakab	*https://fs.usda.gov/*	*http://hdl.handle.net/11122/10930*

national wildlife refuge, military, National Forest Service, and state and native together. We also derived raster maps from vector files of coastline, lakes and rivers, drainage network, towns, main roads, and infrastructures, using Euclidean distance toolbox in ArcMap. The drainage network represents more ephemeral water channels, whereas lakes and rivers encompass more permanently standing or flowing water. The most dominant infrastructures in the state, based on the measured length, were pipelines, gas lines, winter trail, transmission line, highway, tractor trail, old railroad, and summer trail, respectively. We used NAD 1983 Alaska Albers geographic projection for all vector and raster layers; raster maps were prepared at spatial resolutions of 60-m and 100-m; open-access sources are provided for all datasets (Table 2).

Model development and assessment

We used a TreeNet gradient boosting model (SalfordSystems, San Diego, CA; Ohse et al. 2009) to infer relationships between environmental variables and presence of the multi-species bark beetle. Bark-beetle presence and random points were set as 1 and 0, respectively, as the binary categorical response variable of the model. TreeNet uses a boosting classification and regression tree approach (Ohse et al. 2009; Baltensperger and Huettmann 2015a; Young et al. 2018; Humphries et al. 2018) to model relationships between model predictors and a response variable. We developed three models based on different sets of predictors.

TreeNet settings were the same for all three models: the well-tested 'default setting' in the *Salford* Predictive Modeler (*SPM*) software suite in which the number of trees and maximum nodes per tree was set 200 and 6, respectively, with10-fold cross-validation. This setting is known to perform very well as a standard on most data (Salford Systems, San Diego, CA; Ohse et al. 2009; Baltensperger and Huettmann 2015a; Humphries et al. 2018). In Model 1, we included all environmental variables as predictors (termed the 'full model'). In Model 2 ('ecological model'), we excluded distance-based variables and included only predictors expressing local conditions: soil map unit, land status, land cover type, mean annual precipitation, mean annual temperature, elevation, slope, and aspect. Finally, in Model 3, we excluded roads (Euclidean distance to roads) to minimize impact of any sampling bias (Kadmon et al. 2004; Ohse et al. 2009; Baltensperger and Huettmann 2015a). In our dataset, almost half of the recorded bark beetle presences were <20 km from roads (Appendix II).

Each model was applied to the 1-km spaced grid points across Alaska. Every point on the grid was assigned a relative index of occurrence (RIO) ranging 0–1. For better visualization of the study area overall, we used inverse distance weighting (IDW) algorithm in ArcMap 10.4 to interpolate the value of RIO across the state. Performance of the three models were assessed by comparing the area under the receiver operating characteristic curve (AUC), and percent of correctly predicted presences (% corr.) derived from the confusion matrix (Fielding and Bell 1997). Variable importance (ranking) in each model was assessed using computed relative importances in SPM; partial dependency plots were used to illustrate relationship between the response and predictor when all other variables were held to average values. However, the variable importance (ranking) score should not be used to conclude the absolute informative value of a variable; rather, the scores indicate the amount of contribution that each variable (either via a primary role in splitting tree branches or in a substitute role to any of the primary splitters) makes in classifying or predicting the target variable (D. Steinberg; Salford Systems, San Diego, CA). As such, the variable rankings are highly related to the performance metric and a chosen tree structure (D. Steinberg; Salford Systems, San Diego, CA). Our focus, rather, is on deriving inference from the predictions (Breiman 2001; Salford Systems, San Diego, CA).

Conventionally, models have been tested using part of the primarily records of the species presence used to build the model; however, this approach may overestimate a model accuracy, especially if the collection of datasets was systematically biased (Newbold et al. 2010). Therefore, it is better to assess a model using an independent dataset (Newbold et al. 2010). We used an independent dataset of 68 locations of 3 bark beetle species surveyed by the U.S. Department of Forest Service during 2016–2017 (USFS 2019) to additionally assess the three models. The model-predicted occurrences of bark beetle species underlying these 68 widely-spaced locations in the study area were compared using descriptive statistics and box plots in R (R Core Team 2018; Humphries et al. 2018).

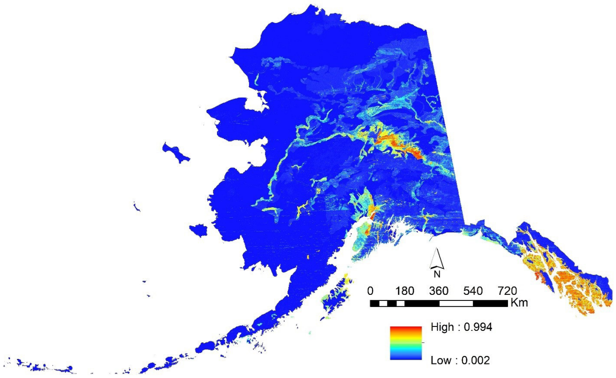


Figure 3. Predicted distribution map of bark beetles in Alaska using model 2 (ecological model). Predicted maps of model 1 (full model) and model 3 (model with excluded roads) are included as appendices III and IV, respectively.

RESULTS

Predicted distribution map

Most occurrence data were in evergreen forest, mixed forest, shrub/scrub, low intensity developed area, and deciduous forest, in decreasing order (Table 1). Our results comprise three maps summarize the RIO of bark beetle suitability patterns across Alaska (Figure 3, Appendices III and IV). After comparing the distribution maps resulting from the full model and the ecological model, the core predicted bark beetle occurrence falls in three hotspot regions: south-central Alaska, southeastern Alaska, and interior Alaska. Model 3 showed a pattern similar to models 1 and 2: hotspots of occurrence corresponded to the vicinity of human settlements and human-built infrastructure such as towns, highways, and railroads (Figure 3, Appendices III and IV). The core hotspots on all three maps (areas close to the urban centers of Fairbanks, Anchorage, and Juneau) likely contributed to the emergence of multi-species bark beetles. In addition to urban areas, other hotspots were along linear features, such as rivers, of central and western Alaska. The overall similarity in the prediction patterns using different combinations of predictors for the three models may signal the generic prediction strength gained from machine-learning algorithms.

Model Performance and Inferences

The AUC values for the three models were very close, all at ~0.99. The percent correctly predicted presences for the three models were 91.9%, 93.2%, and 94.4%, respectively. Relative importances of predictor variables were computed for the three models (Table 3). In Model 1, distance to infrastructure, soil, land status, distance to roads, distance to towns, and land cover type were the most explanatory variables. Model 3 (in which roads were excluded) showed a similar pattern to that observed in model 1. Model 2 revealed variable importance patterns similar to those in the full and ecological models even though the spatially-dependent predictors were excluded from model development. In the ecological model, soil, land status, and land cover type became top-contributing variables. In all three models, land status of "state and native" and "private or municipal" explained the most presence locations whereas "national wildlife refuge" explained pseudo-absence locations. Land cover categories of low and medium development intensities appeared to favor bark beetle presence most. Developed areas with a low or medium intensity most commonly include single-family houses with a mixture of constructed materials and vegetation, and the impervious surfaces from the total cover ranged from 20-49% and 50-79%, respectively.

Table 3. Relative importance of predictor variables included in model 1 (full model), model 2 (ecological model), and model 3 (model with roads excluded). Top three predictors are shown in bold, and dash indicates non-included predictors

Variable	**Model 1**	**Model 2**	**Model 3**
Distance to infrastructure	**100.00**	-	**100.00**
Map unit soil	**82.50**	**100.00**	**76.94**
Land status	**44.19**	**63.52**	**47.73**
Distance to main roads	36.40	-	-
Distance to towns	34.75	-	37.95
Land cover 2011	30.89	**54.17**	31.56
Mean annual temperature	16.03	51.25	14.60
Distance to drainage network	14.22	-	14.39
Elevation (DEM)	12.70	19.72	11.92
Distance to lakes and rivers	10.25	-	12.37
Distance to coastline	9.30	-	13.99
Mean annual precipitation	8.13	25.26	17.47
Aspect	5.69	14.69	4.81
Slope	4.49	15.42	4.07

In models 1 and 3, distances below 5-6 kilometers from infrastructures explained the most bark beetle presences. A similar pattern was revealed for distance-to-roads in model 1 and distance-to-towns in models 1 and 3, so that distances below 20-25 km from roads and towns favored bark beetle presences. Soil types that with a texture of silt loam, Schrock (usually found on stream terraces with a slope of 0-2%), and typic haplocryands (typically found on 1-8% slopes) (USDA 1998, 2005), were the most important soil types in providing suitable habitats for bark beetle host tree species. In all three models,

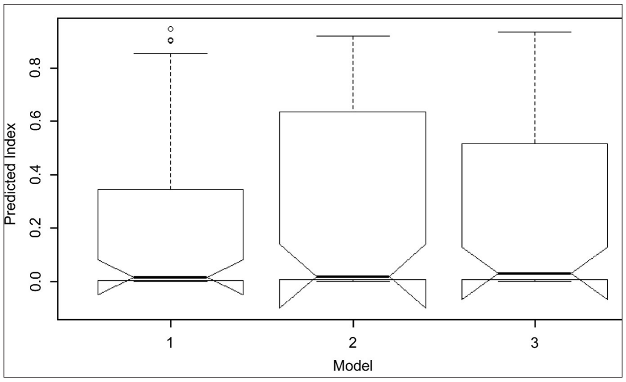


Figure 4. Box plots of models 1-3 used to describe the statistics of the RIO gained by 68 assessment/test points of 3 bark beetle species. The dots in model 1 shows the outliers. The dark thick line within the boxes represents median value within the range of predicted index. 25% of dataset are below the median (1st quantile between the two straight dark lines), and 75% of the dataset are above the median (3rd quantile ends by the upper edge of boxes). The whisker on top of the boxes and the lower triangular shapes below the second dark straight line represent the maximum and minimum values. Note that machine learning models produce a RIO which is not a probability nor symmetrical.

regions with mean annual temperatures >-3.0ºC and mean annual precipitation <350 mm were those that favored the distribution of bark beetles. In addition, areas with slope range <40º and aspect of 100–300º correspond to most bark beetle presences. Elevation showed a different pattern from other predictors: in all three models, areas at elevations <2000 m and >4000 m favored bark beetle presence, whereas elevations of 2000-4000 m was not occupied by bark beetle species. In models 1 and 3, proximity at 4 km from rivers and lakes with standing water, 500 m from the drainage network, and 10 km from the coastline, favored bark beetle presences.

Model assessment

Although the AUC and percent correct prediction statistics revealed that all three models performed well in predicting bark beetle presence and pseudo-absence locations, testing with newly surveyed independent data (2016-2017 records of 68 presence points) revealed differences in their performance. Model 2 was most successful in predicting the independent test points (Figure 4). The median RIOs received by test points using models 1, 2, and 3 were 0.018, 0.021, 0.033, respectively. It should be stated that those RIOs are an index and not probabilities and thus, it includes a range of values that are neither symmetrical nor always reaching 1. That is due to the tree nature of the algorithm used. The model assessment with the alternative test data shows the validity of those concepts (Figure 4; Kandel et al. 2015).

Given that the ecological model predicted the bark beetle occurrences marginally better than two other models, we additionally included the frequency distribution of the predicted RIO for the 68 surveyed points using the ecological model (Figure 5). The frequency plot revealed that about 60% of surveyed bark beetles (assessment/test points) received a predicted index greater than 0.1. Predicted values less than a 0.0049 threshold excluded 5% of the test points within the 95% confidence interval (95% of predicted presence were ≥0.0049; Figure 5). We followed Pearson et al. (2004) and Newbold et al. (2010) to present the prediction map in a binary format but using the 95% confidence interval of the newly surveyed independent presence points, aiming to incorporate current variations in species distribution across the landscape due to temporal-scale changes in the environment (Figure 6). The omission rate was

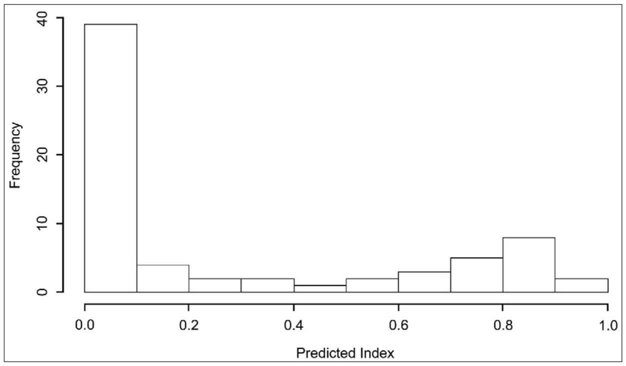


Figure 5. Frequency distribution of gained predicted RIOs for 68 locations of bark beetle presence surveyed by the USFS (assessment/test points) using the ecological model (model 2).

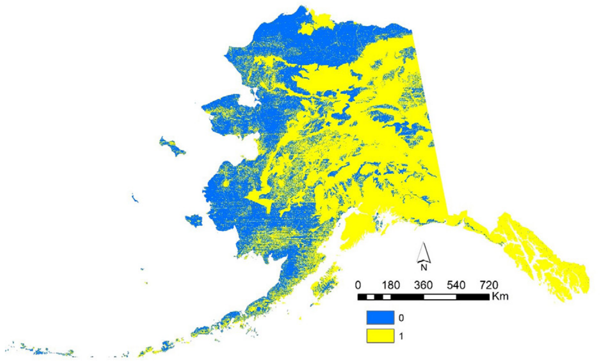


Figure 6. Classified prediction map of multi-species bark beetle occurrence using 95% confidence interval of assessment/test points to differentiate predicted index of relative occurrence (RIO) of the ecological model. Value 1 (presence) represents the favorable habitats and value 0 (absence) represents regions that may not be occupied by scolytines community based on the current climatic conditions and biophysical attributes of the landscape.

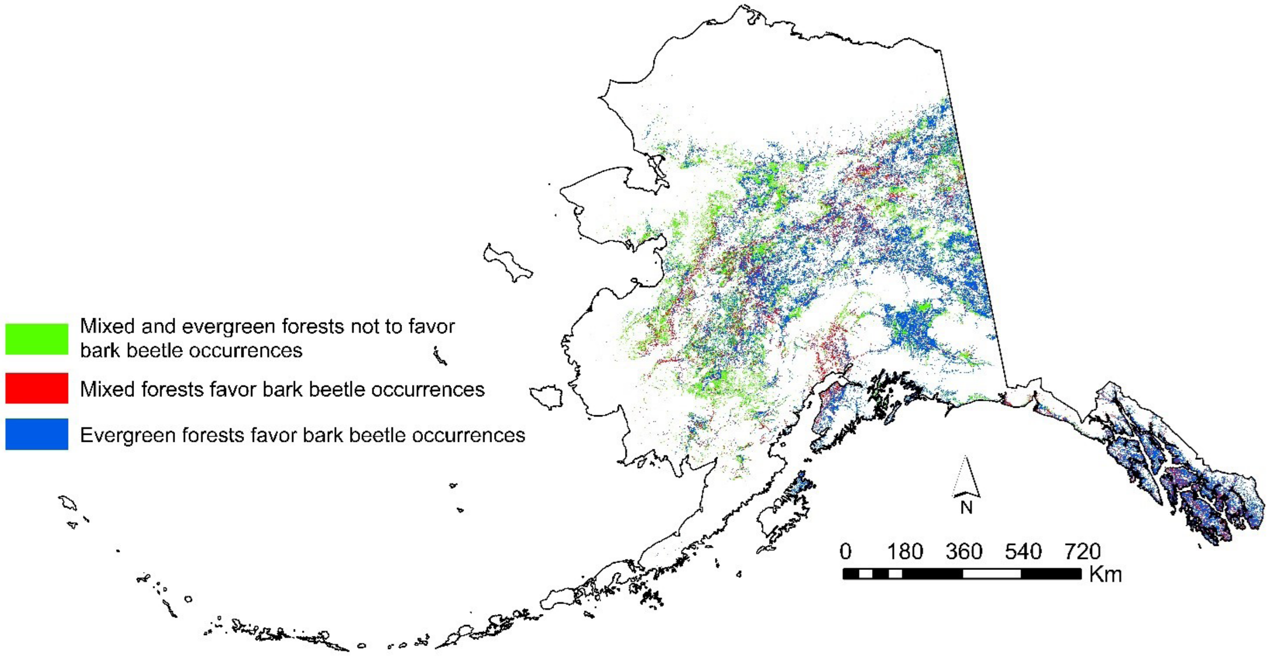


Figure 7. Classified prediction map of multi-species bark beetle occurrences in different forest types: Mixed and evergreen forests that predicted not to favor bark beetle occurrences (green color), mixed forests that expected to favor bark beetle occurrences (red color), and evergreen forests that predicted to be occupied by different bark beetle species (blue color). The 2011 NLCD was the reference map to extract forest type and area across the state of Alaska.

zero after overlaying the 838 presence points on the binary map.

About 60.3% of the surface area of Alaska received a value of 'presence' and is expected to provide favorable habitats for scolytines. Those habitats are not solely forested landscapes but include shrublands as well (Figure 6). We additionally overlaid the mixed and evergreen forest types, extracted from the NLCD 2011 map, on the binary map (Figure 7): ~16% of the mixed forests and 59% of evergreen forests are expected to be suitable for bark beetles, based on current climatic conditions and biophysical attributes of the landscape (Figure 7).

Discussion

Our approach of studying several species of bark beetles, considered as a community organism was new. It should be emphasized that bark beetles live not only on trees but also on shrubs and similar species that together may create poorly studied landscape reservoirs for bark beetles (McDermott et al., 2021). For instance, the Arctic tundra shrubs would allow bark beetle species to live beyond the tree line in northern areas of Alaska. These landscape reservoirs are apparent in our final classified prediction maps (Figures 6 and 7). This finding would be critical for a better understanding of the ecology of bark beetle communities, in addition to designing a better forest pest management strategy.

The forest and non-forested landscapes (e.g., Arctic shrublands) that are predicted to favor bark beetle communities represent potential habitats that may support intra- and inter-specific competition within and among bark beetle species. From a biodiversity standpoint, these various favorable habitats for multi-species bark beetles may help to preserve or even promote biodiversity, as well as co-evolution within the bark beetle community. On the other hand, forested areas that are supposed not to be occupied by bark-beetle species (green shading, Figure 7) would also be important from a forest management perspective to be protected against future anthropogenic disturbances that may promote the infestations.

We were able to assemble open-access data for understanding the ecology of bark beetle communities on a broad scale within the immense geographic area of Alaska. This dataset, together with the machine-learning modeling approach that we used, can provide a foundation for future research use. The methodology applies not only to scolytines, but also to other multi-species questions of concern, such as forest defoliators and small and big game wildlife species worldwide (see, e.g., Huettmann and Schmid 2014, Humphries et al. 2018). The boosted classification and regression tree approaches that we used are particularly useful in dealing with such ecological and environmental datasets, with the common characteristics of being big, complex, and spatially autocorrelated (Breiman 2001; Elith et al. 2006; Humphries et al. 2018). Furthermore, our results will allow us, in the future, to focus on predictions based on climate change scenarios for future time periods (see Baltensperger and Huettmann 2015b).

The higher predictive power of model 3, relative to the full model (model 1), may highlight potential effects of sampling bias in collecting bark beetle occurrence data that results from opportunistic sampling along roads or inconsistent sampling effort over time and space (Yost et al. 2008; Zabihi et al. 2017). Often, species occurrence data have such characteristics owing to lack of awareness or sampling bias in the geographic space (Stockwell and Peterson 2002; Graham et al. 2004; Elith et al. 2006; Yost et al. 2008; Ohse et al. 2009; Baltensperger and Huettmann 2015a; Zabihi et al. 2017). However, the predictive performance of the three models, using the machine-learning algorithm, is sufficiently high that it likely reveals signals in the environmental landscape that can be captured even with sampling bias of species presence. This strength of algorithmic models in finding associations between environmental variables and species occurrences are evident in all three models.

For example, even though we lowered the number of predictors from 14 to 13 to 8 in models 1, 2, and 3, respectively, the relative importance of predictor variables did not change (Table 3). In addition, the predictive strength of the three models is evident in the three resulting species distribution maps, in which the hotspots of bark beetle suitability are in the southeast, south, and interior of Alaska (Figure 3, Appendices III and IV). In contrast to the algorithmic model that we used, more traditional parametric models can become unstable by removing less important predictor variables from the model and consequently lead to wrong conclusions (Breiman 2001).

Although one advantage of algorithmic models is in including more predictors to make more information available for prediction (Breiman 2001), we selected Model 2 for producing a binary predictive map in view of its slightly higher predictive ability using additional independent test points. This model may have had higher predictive ability, thanks to removal of multicollinearity effects between distance-based predictors (e.g., roads, drainage networks, towns, infrastructures, and coastlines) and those predictors considered in the model. For example, drainage network is a function of elevation, slope, and aspect (Ohse et al. 2009), which were included in all three models. Also, soils and landcover can be a function of drainage networks; for example, rich soils are usually found in well-drained locations such as valleys, whereas mountaintops and alpine zones usually do not have fertile soils, and vegetation classes reflect those correlations and interactions indeed.

As is evident from the top predictor variables and visually from the maps in all three models, human settlements and infrastructures are important factors in shaping the distribution of bark beetle species. The hotspot of bark beetle occurrences in the north corresponds well with the periphery of established pipelines; those in the south and southeast are around the cities of Anchorage and Juneau, respectively. For example, Tongass National Forest has one of the highest densities of road networks in southeastern Alaska, where roads have been used for logging and deer hunting since the mid-1950s (Brinkman 2009). The interior hotspots correspond to the vicinity of Fairbanks and along highways.

We further found that land ownership and management, such as state lands, native lands, and private lands, are closely associated with bark beetle occurrence, likely as a consequence of land use practices that may disturb forest landscapes. For example, state lands, managed primarily by various divisions of the Alaska Department of Natural Resources (AKDNR), have been influenced by designated land use in the form of sale or lease to the public; lease for commercial, industrial, and recreational use; selling minerals; and temporary permits for use and access (Alaska Department of Transportation and Public Facilities, Northern Region, 2018). Native lands are aboriginal lands that are owned by individual village corporations having regional rights of exploiting minerals. Private lands are owned by individual entities, municipalities, and boroughs, and are generally concentrated close to cities, villages, and populated regions along highways and roads (Alaska Department of Transportation and Public Facilities, Northern Region, 2018). The emergence and spreads of bark beetle attacks in the vicinity of human settlements and recreation sites, with public use and infrastructure developments such as pipelines, roads, and hiking trails, could be related to the associated disturbances and compaction damages in soil structure. These soil disturbances consequently compromise tree roots and may eventually lead to higher chance of infections (FS-R10-FHP 2019). Urbanization in Arctic and sub-Arctic regions increases impervious surfaces, creating urban heat islands (Chandler 1960; Oke 1988) that may impact the climate and associated ecosystem components, such as the spread of bark beetle infestations.

In sum, anthropogenic factors seem to be closely associated with, and perhaps accelerate, beetle outbreaks in different ways. For example, human-induced climate change results in warmer and dryer summers that reduce tree resiliency, and milder winters that decrease beetle mortality (Müller et al. 2008; Müller 2011). In addition, untreated spruce slash-and-debris, due to, for instance, highway and powerline constructions, can elevate spruce beetle populations (Schmid 1977; Werner et al. 2006). Logs, slash, or dead and dying trees favor several bark beetle species, such as *Ips* spp., because of little or no host resistance against beetle attacks (Fettig et al. 2007).

The soil texture of silt loam was closely associated with bark beetle presences, likely due to providing suitable habitats for host trees. For example, Schrock and Haplocryands soils provide habitats for white spruce (USDA 1998, 2005). Across the study area, regions with a higher mean annual temperature and lower annual precipitation, relative to other areas in Alaska, were more closely associated with bark beetle occurrence. This finding mirrors that of Økland et al. (2019), in which high summer temperatures and low precipitation favored the flight period and reproduction rate of most bark beetle species, even those at high latitudes with cooler climates. The aspect range (100–300º), including south- and west-facing slopes, is likely to provide a warmer and more favorable microclimate for bark beetle activities in addition to providing favorable habitats for host tree species. For example, white spruce occurrence was concentrated on south-facing slopes in previous studies (Viereck and Little 2007; Ohse et al. 2009).

Our community-based modeling approach could be debated based on variations in species' interactions with local environments at fine scales of individual host trees or stands, mostly considered as issues of spatial autocorrelation. However, our modeling approach of using a non-parametric model of boosted classification and regression tree that uses many algorithms, ensembles, and responses (Humphries et al. 2018) aimed to learn and model these complex, nonlinear relationships in the data without prior assumptions such as being free of spatial autocorrelation (Huettmann 2018c). In addition, different species within a community may have similar responses to changes in the environment at regional scales (Golicher et al. 2008; Azeria et al. 2009; Chapman and Purse 2011), so community-level analyses of spatial patterns of biodiversity may be beneficial (Chapman and Purse 2011) for biodiversity conservations and natural resource management purposes.

We used a historical collection of bark beetle specimens from UAM without consideration of sampling design strategies or assumptions such as balanced sample sizes for different species. Although unbalanced samples may represent true populations of species across landscapes, future work might test these ideas by removing different species from the model. However, our approach treating the species presences across all bark beetles represents a way of dealing with numerous small sample-size species in our dataset. A model with high sensitivity, even if it results in some overpredictions, will minimize omission of sites that are actually suitable, which is particularly meaningful for rare species (Engler et al. 2004; Barbet-Massin et al. 2012). The 2011 NLCD map that we used does not provide detailed information about different types of conifer species; preparing and using such a map in our models could provide additional detail as regards host tree communities and their effects on bark beetle assemblages.

Acknowledgments

This study was supported by grant No. CZ.02.1.01/0.0/0.0/15_003/0000433, "EXTEMIT – K," financed by the Operational Program Research, Development and Education (OP RDE). We greatly appreciate the constructive comments and suggestions from the reviewers. We are grateful to all data contributors, namely the Alaska Division of Forestry and the United States Department of Forest Service, and Derek Sikes, Curator of Insects at the University of Alaska Museum, for helping with accessing the data used in this study, as well as for providing comments on an earlier version of the manuscript. We further thank John Morton for discussions, and IAB-UAF, namely Brian Barnes and his team, for support for Khoda Zabihi to carry out this research in the EWHALE lab. Huettmann thanks Hazel Berrios, Thor Chrome, Maya Hera, and 'Cub' Sparks Poa. This is EWHALE lab publication # 210.

LITERATURE CITED

ACIA. 2004. Impacts of a Warming Arctic Climate Impact Assessment. Cambridge University Press, UK. Accessed 5 April 2019.[4]

ADFG (Alaska Department of Fish and Game). 2018. Key Habitats of Featured Species. Accessed 5 April 2019.[5]

DOT&PF (Alaska Department of Transportation and Public Facilities), Northern Region. Land Ownership and Management (Appendix C). 2018. Accessed 5 April 2019.[6]

Araújo, M. B., and M. Luoto. 2007. The importance of biotic interactions for modelling species distributions under climate change. Glob. Ecol. Biogeogr. 16:743-753.

Azeria, E. T., D. Fortin, C. Hébert, P. Peres-Neto, D. Pothier, and J. C. Ruel. 2009. Using null model analysis of species co-occurrences to deconstruct biodiversity patterns and select indicator species. Divers. Distrib. 15:958-971.

Baltensperger, A., P., and F. Huettmann. 2015a. Predictive spatial niche and biodiversity hotspot models for small mammal communities in Alaska: applying machine-learning to conservation planning. Landsc. Ecol. 30:681-679.

Baltensperger, A. P., and F. Huettmann. 2015b. Predicted shifts in small mammal distributions and biodiversity in the altered future environment of Alaska: an open access data and machine learning. PLoS ONE 13:e0194377.

Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol. 3*:327-338.

Bearup, L. A., R. M. Maxwell, D. W. Clow, and J. E. McCray. 2014. Hydrological effects of forest transpiration loss in bark beetle-impacted watersheds. Nat. Clim. Change 4:481-486.

Bentz, B. J., J. Logan, J. MacMahon, C. Allen, et al. 2009. Bark beetle outbreaks in Western North America: Causes and consequences. Proceedings: Bark Beetle Symposium, November 15-17, 2005, Snowbird, Utah. University of Utah Press, Chicago, IL. p. 42.

Bentz, B. J., J. Régnière, C. J. Fettig, et al. 2010. Climate change and bark beetles of the western United States and Canada: direct and indirect effects. BioScience 60:602-13.

Bhattacharya, M. 2013. Machine Learning for Bioclimatic Modelling. International Journal of Advanced Computer Science and Applications, 4:1-8.

Bonnot, T. W., J. J. Millspaugh, and M. A. Rumble. 2009. Multi-scale nest-site selection by black-backed woodpeckers in outbreaks of mountain pine beetles. Forest Ecol. Manag. 259:220-228.

Breiman, L. 2001. Statistical modeling: the two cultures. Stat. Sci. 16:199-215.

Brinkman, T. J., F. S. III. Chapin, G. P. Kofinas, and D. K. Person. 2009. Linking hunter knowledge with forest change to understand changing deer harvest opportunities in intensively logged landscapes. Ecol. and Soc. 14:36.

Chandler, T. J. 1960. Wind as a factor of urban temperatures: A survey in north-east London, Weather, 15:204-213.

Chapman, D. S. and B. V. Purse. 2011. Community versus single-species distribution models for British plants. J. Biogeogr. 38:1524-1535.

Christiansen, E., and A. Bakke. 1988. The spruce bark beetle of Eurasia. In: Berryman A. A., (Eds), Dynamics of forest insect populations: patterns, causes, and implications. Plenum, New York, pp 480-504.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. Overton, A. T. Peterson, S. J. Phillips, K. S. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129-151.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. J. of Anim. Ecol. 77:802-813.

Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J. of Appl. Ecol. 41:263-274.

Fettig, C. J., K. D. Klepzig, R. F. Billings, A. S. Munson, T. E. Nebeker, J. F. Negrón, and J. T. Nowak. 2007. The effectiveness of vegetation management practices for prevention and control of bark beetle infestations in coniferous forests of the western and southern United States. Forest Ecol. and Manag. 238:24-53.

Fielding, A.H., and J.F. Bell. 1997. A review methods for the assessment of prediction errors in conservation

---

[4] http://www.library.arcticportal.org/1299.

[5] https://www.adfg.alaska.gov/static/species/wildlife_action_plan/appendix5_forest_habitats.pdf.

[6] http://www.dot.state.ak.us/nreg/westernaccess/documents/corridor_planning_report_appx_c.pdf.

presence/absence models. Env. Cons. 24:38-49.

Flynn, M., J. D. Ford, T. Pearce, and S. Harper. IHACC Research Team. 2018. Participatory scenario planning and climate change impacts, adaptation and vulnerability research in the Arctic. Environ. Sci. Policy 79:45-53.

FS-R10-FHP. 2019. Forest Health Conditions in Alaska - 2019. Anchorage, Alaska. U.S. Forest Service, Alaska Region. Publication R10--PR-45. 68 pp. Accessed February 19, 2021.[7]

Golicher, D. J., L. Cayuela, J. R. M. Alkemade, M. González-Espinosa, and N. Ramírez-Marcial. 2008. Applying climatically associated species pools to the modelling of compositional change in tropical montane forests. Glob. Ecol. Biogeogr. 17:262-273.

Graham, C. H., S. Ferrier, F. Huettmann, C. Moritz, and A. T. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol. Evol. 19:497-503.

Heikkinen, R. K., M. Luoto, R. Virkkala, R. G. Pearson, and J. Körber. 2007. Biotic interactions improve prediction of boreal bird distributions at macroscales. Glob. Ecol. Biogeogr. 16:754-763.

Huettmann, F. 2018a. Advanced data mining (cloning) of predicted climate-scapes and their variances with machine learning: an example from southern Alaska shows topographical biases and strong differences. In: G. R. W., Humphries, D. R., Magness, and F. Huettmann (Eds.), Machine Learning for Ecology and Sustainable Natural Resource Management (1st edition, pp. 227-241). Cham, Switzerland: Springer Nature Switzerland AG, 441 pp.

Huettmann, F., and M. Schmid. 2014. Publicly available open access data and machine learning model-predictions applied with open source GIS for the entire Antarctic Ocean: A first meta-analysis and synthesis from 53 charismatic species. In: B. Veress and J. Szigethy (Eds.), Horizons in Earth Science Research. Volume 11 (Chapter 3): pp. 24-34.

Huettmann, F., E. H. Craig, K. A. Herrick, A. P. Baltensperger, G. R. W. Humphries, D. J. Lieske, K. Miller, T. C. Mullet, S. Oppel, C. Resendiz, I. Rutzen, M. S. Schmid, M. K. Suwal, and B. D. Young. 2018c. Use of machine learning (ML) for predicting and analyzing ecological and 'presence only' data: an overview of applications and a good outlook. In: G. R. W., Humphries, D. R., Magness, and F. Huettmann (Eds.), Machine learning for ecology and sustainable natural resource management (1st edition, pp. 315-332). Cham, Switzerland: Springer Nature Switzerland AG, 441 pp.

Huettmann, F., Mi. Chunrong, and Y. Guo. 2018b. 'Batteries' in machine learning: A fFirst experimental assessment of inference for Siberian Crane breeding grounds in the Russian high arctic based on 'Shaving' 74 predictors. In: G.R.W., Humphries, D.R., Magness, & F. Huettmann (Eds.), Machine learning for ecology and sustainable natural resource management (1st edition, pp. 315-332). Cham, Switzerland: Springer Nature Switzerland AG, 441 pp.

Humphries, G. R. W., D. R. Magness, and F. Huettmann (eds.). 2018. Machine Learning for Ecology and Sustainable Natural Resource Management (1st edition). Cham, Switzerland: Springer Nature Switzerland AG, 441 pp.

Hutchison, O. K. 1968. Alaska's forest resource. Institute of Northern Forestry, USFS, Pacific Northwest Forest and Range Experiment Station. 74 p.

Kadmon, R., O. Farber, and A. Danin. 2004. Effect of road-side bias on the accuracy of predictive maps produced by bioclimatic models. Ecol. Appl. 14:401-413.

Kandel, K., F. Huettmann, M. K. Suwal, G. R. Regmi, V. Nijman, K. A. I. Nekaris, S. T. Lama, A. Thapa, H. P. Sharma, and T. R. Subedi. 2015. Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: Red panda (*Ailurus fulgens*) in the Hindu-Kush Himalaya region. Biodivers. Conserv. 181:150-161.

Kurz, W. A., C. C. Dymond, G. Stinson, G. J. Rampley, E. T. Neilson, A. L. Carroll, T. Ebata, and L. Safranyik. 2008. Mountain pine beetle and forest carbon feedback to climate change. Nature 452:987-990.

Liu, Z. L., C. H. Peng, T. Work, J. N. Candau, A. DesRochers, and D. Kneeshaw. 2018. Application of machine-learning methods in forest ecology: recent progress and future challenges. Environ. Rev. 26:339-350.

Martikainen, P., J. Siitonen, L. Kaila, P. Punttila, and J. Rauh. 1999. Bark beetles (Coleoptera, Scolytidae) and associated beetle species in mature managed and old-growth boreal forests in southern Finland. For. Ecol. Manage. 116:1-3.

McDermott, M.T., P. Doak, C.M. Handel, G.A. Breed, C.P.H. Mulder. 2021. Willow drives changes in arthropod communities of northwestern Alaska: ecological implications of shrub expansion. Ecosphere 12:e03514.

McDowell, N., et al. 2008. Mechanisms of plant survival and mortality during drought: Why do some plants survive while others succumb to drought? New Phy-

---

[7] https://www.fs.usda.gov/Internet/FSE_DOCUMENTS/fseprd712413.pdf

tologist 178:719-739.

Morris, J. L., S. Cottrell, C. J. Fettig, W. D. Hansen, R. L. Sherriff, V. A. Carter, J. L. Clear, J. Clement, R. J. DeRose, J. A. Hicke, P. E. Higuera, K. M. Mattor, A. W. R. Seddon, H. T. Seppä, J. D. Stednick, and S. J. Seybold. 2016. Managing bark beetle impacts on ecosystems and society: priority questions to motivate future research. J. Appl. Ecol. 54:750-760.

Müller, J., H. Busler, M., T. Rettelbach, and P. Duelli. 2008. The European spruce bark beetle Ips typographus in a national park: from pest to keystone species. Biodivers. Conserv. 17:2979-3001.

Müller, M. 2011. How natural disturbance triggers political conflict: bark beetles and the meaning of landscape in the Bavarian Forest. Glob. Environ. Change 21:935-946.

Munro, H. L., B. T. Sullivan, C. Villari, and K. J. K. Gandhi. 2019. A review of the ecology and management of black turpentine beetle (Coleoptera: Curculionidae), Environ. Entomol. 48:765-783.

Nauman, T. W., and J. A. Thompson. 2013. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. Geoderma 2014:385-399.

Newbold, T., T. Reader, A. El-Gabbas, W. Berg, W.M. Shohdi, and S. Zalat, 2010. Testing the accuracy of species distribution models using species records from a new field survey. Oikos, 119:1326-1334.

Ohse, B., F. Huettmann, S. Ickert-Bond, and G. Juday. 2009. Modeling the distribution of white spruce (*Picea glauca*) for Alaska with high accuracy: an open access role-model for predicting tree species in last remaining wilderness areas. Polar Biol. 32:1717-29.

Oke, T. R. 1988. The Urban Energy Balance, Prog. Phys. Geogr. 12:471-508.

Pearson, R.G., T.P. Dawson, and C. Liu. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27:285-298.

Prasad, A. M. 2018. Machine learning for macroscale ecological niche modeling - a multi-model, multi-response ensemble technique for tree species management under climate change. In: G. R. W., Humphries, D. R., Magness, and F. Huettmann (Eds.), Machine learning for ecology and sustainable natural resource management (1st edition, pp. 315-332). Cham, Switzerland: Springer Nature Switzerland AG, 441 pp.

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [8]

Raffa, K. F. 1988. The mountain pine beetle in western North America. In: Berryman, A.A. (ed.), Dynamics of forest insect populations: patterns, causes, and implications. Plenum, New York, pp. 556-576.

Raffa, K. F., and A. A. Berryman. 1983. The role of host plant resistance in the colonization behavior and ecology of bark beetles (Coleoptera: Scolytidae). Ecol. Monogr. 53:27-49.

Raffa, K. F., B. H. Aukema, B. J. Bentz, A. L. Carroll, J. A. Hicke, M. G. Turner, and W. H. Romme. 2008. Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of bark beetle eruptions. Bioscience 58:501-517.

Jonasova, M., and K. Pracha. 2004. Central-European mountain spruce (Picea abies (L.) Karst.) forests: regeneration of tree species after a bark beetle outbreak. Ecol. Eng. 23:15-27.

Saab, V. A., Q. S. Latif, M. M. Rowland, T. N. Johnson, A. D. Chalfoun, S. W. Buskirk, et al. 2014. Ecological consequences of mountain pine beetle outbreaks for wildlife in western North American forests. Forest Sci. 60:539-559.

Schmid, J. M. 1977. Guidelines for minimizing spruce beetle populations in logging residuals. U.S. For. Ser. Res. Pap. RM-185, Fort Collins, Colorado.

Schmitt, C. B., N. D. Burgess, L. Coad, A. Belokurov, C. Besancon, L. Boisrobert, et al. 2009. Global analysis of the protection status of the world's forests. Biol. Conserv. 142:2122-2130.

Schowalter, T. D. 2012. Ecology and management of bark beetles (Coleoptera: Curculionidae: Scolytinae) in southern pine forests. J. Integr. Pest Manage. 3:A1-A7.

Seidl, R., M. J. Schelhaas, W. Rammer, and P. J. Verkerk. 2014. Increasing forest disturbances in Europe and their impact on carbon storage. Nature Climate Change 4:806-810.

Steinberg, D. What is the Variable Importance Measure? [Blog post]. Salford Systems, San Diego, CA. Accessed August 2, 2019.[9]

Stockwell, D. R. B., and A. T. Peterson. 2002. Controlling bias in biodiversity data. In: J. M., Scott, P. J., Heglund, M. L., Morrison, J. B., Haufler, M. G., Raphael, W. A., Wall, and F. B., Samson (Eds.), Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, pp 537-546.

Stucky, B., J. Balhoff, N. Barve, V. Barve, L. Brenskelle, M. Brush, G. Dahlem, J. Gilbert, A. Kawahara, O. Keller, A. Lucky, P. Mayhew, D. Plotkin, K. Selt-

---

[8] http://www.R-project.org/.

[9] *https://www.salford-systems.com/blog/dan-steinberg/item/37-regression-tree-ensembles.*

mann, E. Talamas, G. Vaidya, R. Walls, M. Yoder, G. Zhang, and R. Guralnick. 2019. Developing a vocabulary and ontology for modeling insect natural history data: example data, use cases, and competency questions. Biodivers. Data J. 7:e33303.

Tyralis, H., G. Papacharalampous, and A. Langousis. 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water 11:910.

USDA (United States Department of Agriculture), Forest Service, Pacific Northwest Research Station, 2009. The Western Bark Beetle Research Group: a unique collaboration with forest health protection: proceedings of a symposium at the Society of American Foresters Conference, October 23-28, Portland, Oregon. Accessed February 17, 2021.[10]

USDA (United States Department of Agriculture). 1998. Soil Survey of Yentna Area, Alaska. Accessed April 2, 2019.[11]

USDA (United States Department of Agriculture). 2005. Soil Survey of Yentna Area, Alaska. Accessed April 2, 2019.[12]

USDA (United States Department of Agriculture). 2008. Insects and Diseases of Alaskan Forests. Accessed August 29, 2019.[13]

USFS (United States Forest Service). 2019. Forest Health Conditions in Alaska Reports & ADS Damage Maps (2002-2019). Accessed September 12, 2019.

USGS (United States Geological Survey). 2017. Alaska Range; Wood River Mountains; Kigluaik Mountains; Brooks Range. Accessed August 4, 2019.[14]

Ver Hoef, J. M., N. Cressie, R. N. Fisher, and T. J. Case. 2001. Uncertainty and spatial linear models for ecological data. Spatial uncertainty in ecology: implications for remote sensing and gis applications. C. T., Hunsaker, M. F., Goodchild, M. A., Friedl, and T. J., Case (eds.), pp. 214-237. Springer, New York.

Viereck, L. A., and E. L., Little. 2007. Alaska Trees and Shrubs. Snowy Owl Books, Fairbanks.

Wallin, K. F., and K. F., Raffa. 2004. Feedback between individual host selection behavior and population dynamics in an eruptive herbivore. Ecol. Monogr. 74:101-116.

Werner, R. A., E. H. Holsten, S. M. Matsuoka, and R. E. Burnside. 2006. Spruce beetles and forest ecosystems in south-central Alaska: a review of 30 years of research. Forest Ecol. Manage. 227:195-206.

Yost, A. C., et al., 2008. Predictive modeling and mapping sage-grouse (*Centrocercus urophasianus*) nesting habitat using maximum entropy and a long-term dataset from southern Oregon. Ecol. Informatics 3:375-386.

Young, B. D., J. Yarie, D. Verbyla, F. Huettmann, and S. Chapin III. 2018. Mapping aboveground biomass of trees using forest inventory data and public environmental variables within the Alaskan boreal forest. In: G. R. W., Humphries, D. R., Magness, and F. Huettmann (eds.), Machine Learning for Ecology and Sustainable Natural Resource Management (1st edition, pp. 315-332). Cham, Switzerland: Springer Nature Switzerland AG, 441 pp.

Zabihi, K., G. B. Paige, A. L. Hild, S. N. Miller, A. Wuenschel, and M. J. Holloran. 2017. A fuzzy logic approach to analyse the suitability of nesting habitat for greater sage-grouse in western Wyoming. J. Spat. Sci. 62:215-234.

---

[10] https://www.fs.fed.us/pnw/pubs/pnw_gtr784.pdf.

[11] https://www.nrcs.usda.gov/Internet/FSE_MANUSCRIPTS/alaska/AK631/0/yentna.pdf.

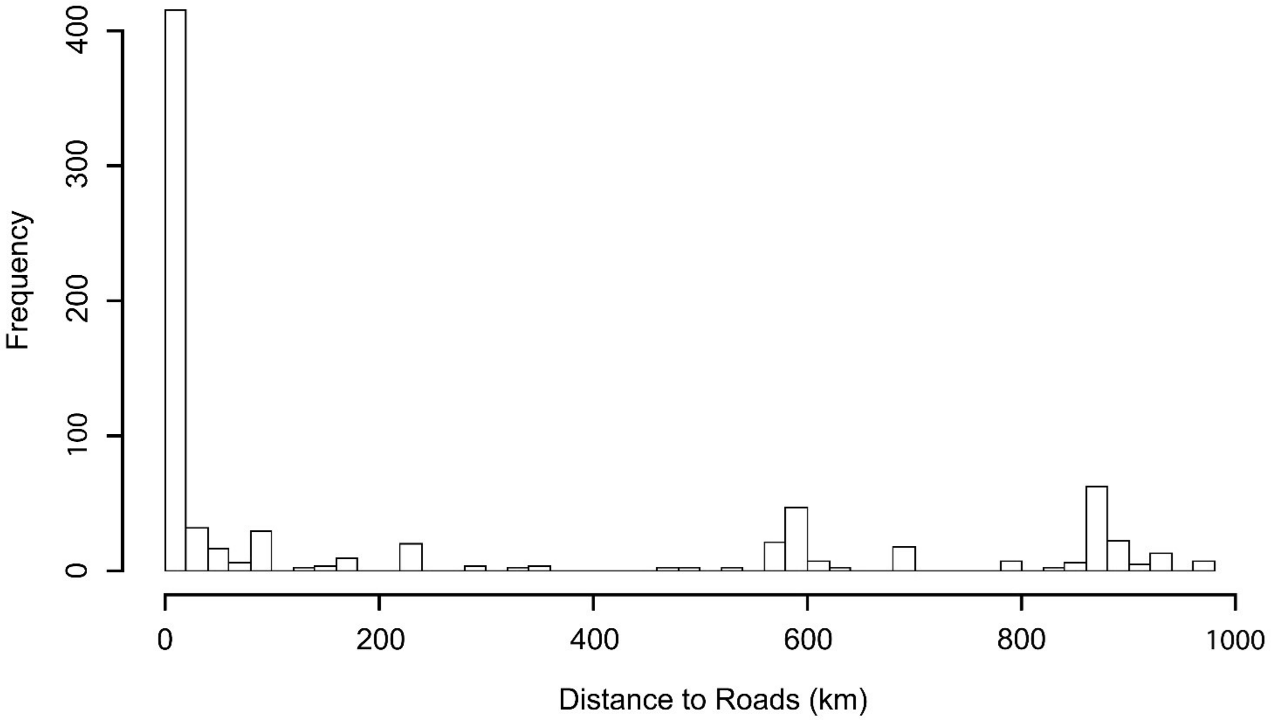
[12] https://www.nrcs.usda.gov/Internet/FSE_MANUSCRIPTS/alaska/AK653/0/SaintPaul.pdf

[13] https://www.fs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb5315942.pdf.

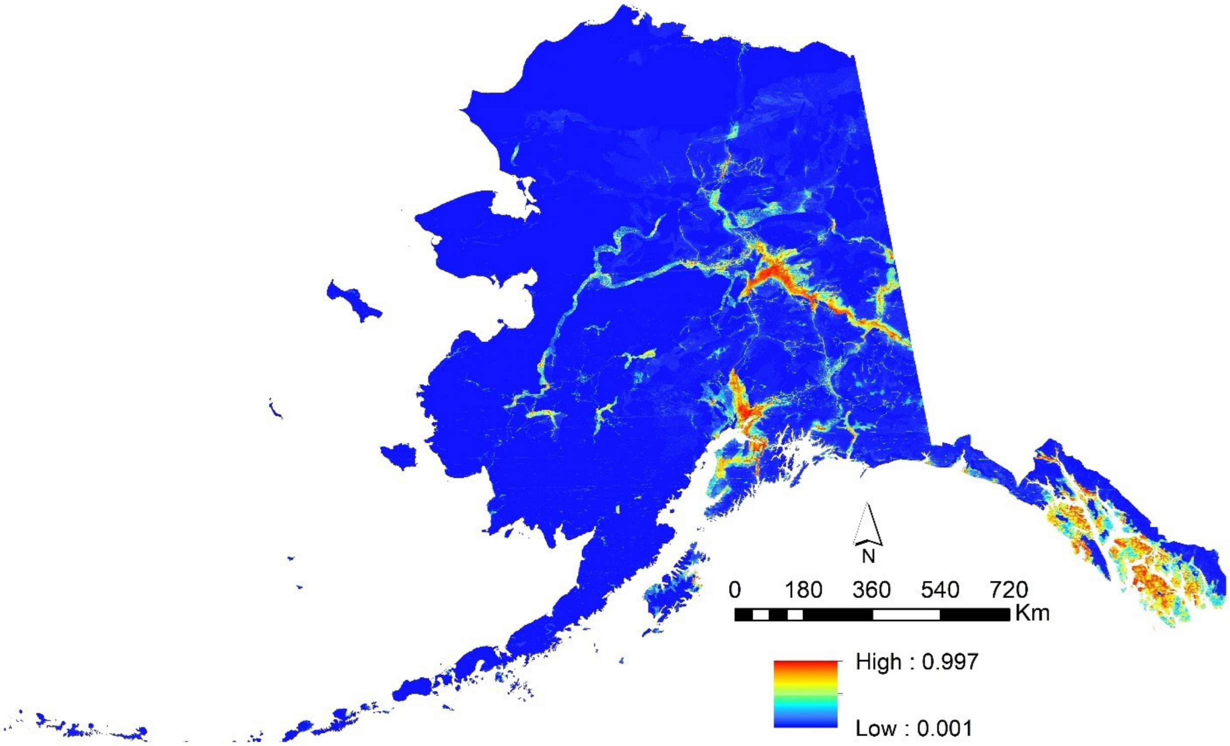
[14] https://pubs.usgs.gov/pp/p1386k/pdf/10_1386K_alaskarange.pdf.

Appendix I. Bark beetle species, host trees species, and beetle sample size used as presence points to extract model inputs/predictors. N/A represents collected bark beetle specimens with no host tree species included.

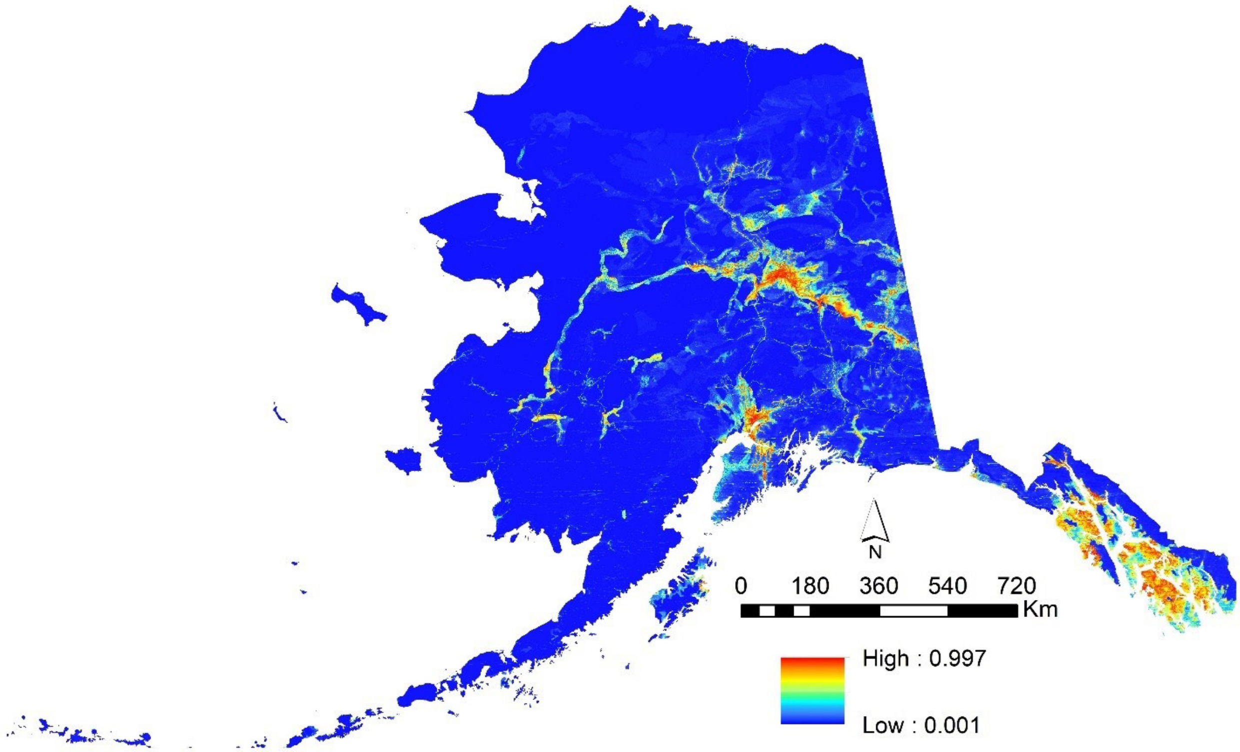
Species	Host tree	Species sample size	Genus	Genus sample size
Scierus annectans	White spruce (*Picea glauca*)	1	*Scierus*	1
Alniphagus aspericollis	N/A	1	*Alniphagus*	1
Carphoborus andersoni	Black spruce (*Picea mariana*) and white spruce	5	*Carphoborus*	12
Carphoborus carri	Black spruce and white spruce	5		
Carphoborus intermedius	White spruce	1		
Carphoborus sp.	White spruce	1		
Cryphalus ruficollis	White spruce	7	*Cryphalus*	7
Crypturgus borealis	Black spruce and white spruce	18	*Crypturgus*	18
Dendroctonus punctatus	White spruce	11	*Dendroctonus*	83
Dendroctonus rufipennis	White spruce	66		
Dendroctonus simplex	Tamarack (*Larix laricina*)	4		
Dendroctonus sp.	White spruce	2		
Dolurgus pumilus	Sitka spruce (*Picea sitchensis*)	22	*Dolurgus*	23
Dolurgus sp.	N/A	1		
Dryocoetes affaber	Black spruce, white spruce, Sitka spruce, western hemlock (*Tsuga heterophylla*), and lodgepole pine (*Pinus contorta*)	69	*Dryocoetes*	133
Dryocoetes autographus	Black spruce, white spruce, Sitka spruce, western hemlock, mountain Hemlock (*Tsuga mertensiana*)	51		
Dryocoetes caryi	Lutz spruce (*Picea x lutzii*)	10		
Dryocoetes sp.	N/A	3		
Gnathotrichus retusus	N/A	2	*Gnathotrichus*	3
Gnathotrichus sp.	N/A	1		
Hylastes nigrinus	N/A	2	*Hylastes*	2
Hylurgops rugipennis	White spruce, Sitka spruce, lodgepole pine, and western hemlock	36	*Hylurgops*	52
Hylurgops sp.	Western hemlock	14		
Hylurgops subcostulatus	N/A	2		
Ips borealis	White spruce	9	*Ips*	104
Ips perroti	Black spruce and white spruce	2		
Ips perturbatus	White spruce	52		
Ips pini	Lodgepole pine	7		
Ips sp.	White spruce	3		
Ips perturbatus	White spruce	3		
Ips tridens	Sitka spruce, white spruce, and Lutz spruce	28		
Lymantor alaskanus	N/A	1	*Lymantor*	1
Orthotomicus caelatus	White spruce, Sitka spruce, and lodgepole pine	13	*Orthotomicus*	14
Orthotomicus sp.	White spruce	1		
Phloeosinus cupressi	Yellow cedar (*Cupressus nootkatensis*)	3	*Phloeosinus*	17
Phloeosinus pini	White spruce	7		
Phloeosinus punctatus	Western redcedar (*Thuja plicata*)	3		
Phloeosinus sequoiae	Yellow cedar and western redcedar	2		
Phloeosinus sp.	Yellow cedar and western redcedar	2		
Phloeotribus lecontei	Black spruce and white spruce	3	*Phloeotribus*	12
Phloeotribus piceae	Black spruce and white spruce	9		
Pityophthorus bassetti	White spruce	2	*Pityophthorus*	74
Pityophthorus murrayanae	White spruce	2		
Pityophthorus nitidulus	Balck spruce, white spruce, Lutz spruce, Sitka spruce, and lodgepole pine	21		
Pityophthorus nitidus	Black spruce and white spruce	5		
Pityophthorus opaculus	White spruce	4		
Pityophthorus pulchellus	Lodgepole pine	1		
Pityophthorus recens	Lutz spruce	1		
Pityophthorus sp.	White spruce, black spruce, Sitka spruce, and Lodgepole pine	34		
Pityophthorus tuberculatus	Lodgepole pine	2		
Pityophthorus Borealis	White spruce	1		
Pityophthorus venustus	White spruce	1		
Polygraphus convexifrons	White spruce and Lutz spruce	6	*Polygraphus*	59
Polygraphus rufipennis	White spruce, black spruce, and Sitka spruce	52		
Polygraphus sp.	N/A	1		
Procryphalus mucronatus	N/A	1	*Procryphalus*	3
Procryphalus utahensis	N/A	2		
Pseudips concinnus	Sitka spruce and Lutz spruce	16	*Pseudips*	21
Pseudips mexicanus	Lodgepole pine	3		
Pseudips sp.	N/A	2		
Pseudohylesinus granulatus	N/A	1	*Pseudohylesinus*	52
Pseudohylesinus sericeus	Lodgepole pine	5		
Pseudohylesinus sitchensis	Sitka spruce	2		
Pseudohylesinus sp.	Western hemlock	27		
Pseudohylesinus tsugae	Western hemlock and mountain hemlock	17		
Scierus annectans	N/A	10	*Scierus*	14
Scierus pubescens	White spruce	4		
Scolytinae sp.	White spruce and black spruce	8	Scolytinae	8
Scolytus piceae	White spruce, black spruce, and tamarack	12	*Scolytus*	12
Trypodendron betulae	White spruce and black spruce	4	*Trypodendron*	107
Trypodendron lineatum	White spruce, black spruce, Sika spruce, and western hemlock	74		
Trypodendron retusum	White spruce	9		
Trypodendron rufitarsis	Mountain hemlock and white spruce	5		
Trypodendron sp.	White spruce and Sitka spruce	15		
Trypophloeus populi	White spruce	1		
Trypophloeus striatulus	N/A	2	*Trypophloeus*	3
Xylechinus montanus	White spruce	2		
Total Sample Size		838	*Xylechinus*	2838



Appendix II. Frequency distribution of Euclidean distance to roads (km) for 838 bark beetle presence locations. The peak distance close to the roads could indicate an ecological corridor for the spread of bark beetles.



Appendix III. Predicted distribution map of bark beetle using model 1 (full model)



Appendix IV. Predicted distribution map of bark beetle using model 3 (model with excluded roads)