18 February, 2022

# **Reproducible Methodology for Developing Species Distribution Models of Boreal Landbirds in Ontario’s Far North and Adjacent Areas**

# Andrew D. Crosby1,2,4; Peter Sólymos1,2; Lionel Leston1,2; Erin M. Bayne1,2

## 1Boreal Avian Modelling Project, 751 General Services Building, University of Alberta, Edmonton, Alberta T6G 2H1

## 2Department of Biological Sciences, CW405 Biological Sciences Building, University of Alberta, Edmonton, Alberta T6G 2E9

## 4Corresponding author. Email: [crosby@ualberta.ca](mailto:crosby@ualberta.ca)

## **Summary**

The boreal region of North America represents a critical biome for sustaining continental landbird populations (Blancher and Wells 2005). Of the 190 species of landbirds that breed in the boreal, 46 breed only in the Far North, 35 are restricted to the Hudson Bay Lowlands, and nine are on the list of Species at Risk in Ontario (need citation). There are existing species distribution models developed by the Boreal Avian Modelling Project (BAM: <http://www.borealbirds.ca>) for predicting abundance and distribution of boreal bird species across all of Canada. The utility of these national-scale models for predicting bird response to cumulative effects (development, climate change) in the Ring of Fire chromium mining development area may be limited, as they do not include some key predictors available in Ontario. Therefore, it is necessary to develop localized models of breeding bird distributions and responses to disturbance.

The objective of this project is to develop regional species distribution models of breeding landbirds in Ontario’s Far North that can be integrated with outputs from landscape simulation models (e.g. SpaDES R package) to forecast the potential effects of proposed industrial development on bird populations.

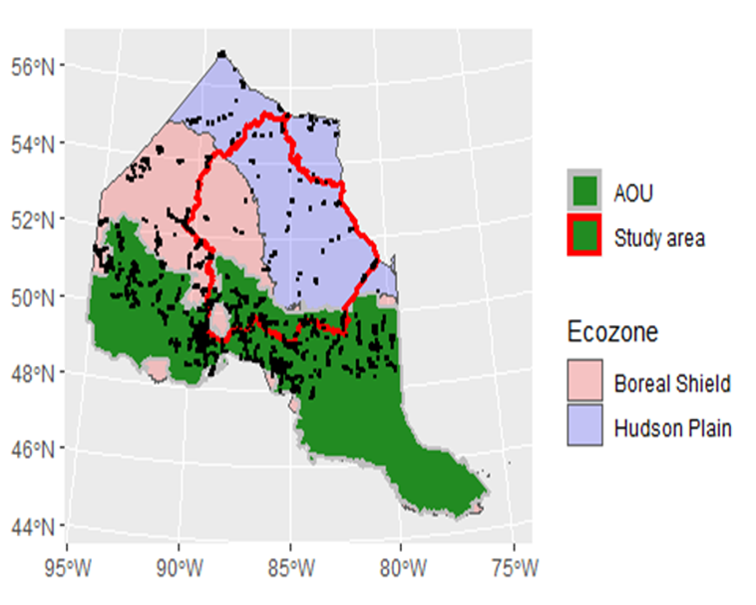
The purpose of this document is to describe reproducible methods for acquiring and processing geospatial data files, and fitting species distribution models, for boreal landbirds in Ontario’s Far North.



**Figure 1.** Map of Ontario, Canada showing the Far North boundary, proposed Ring of Fire development area, and current and proposed transportation infrastructure. Map provided publicly by Northern Policy Institute. Accessed online at <https://www.northernpolicy.ca/upload/documents/publications/maps/research-ring-of-fire-map-15.04.23-revis.pdf>, March 31, 2021.

## **Study Area**

Our area of interest (hereafter study area) was the area encompassed by 5 boreal caribou ranges identified as overlapping with the area of potential disturbance by industrial development for chromium mining in the Far North (Fig. 2). This area was split between the Hudson Plains and Boreal Shield ecozones. The Hudson Plains is a flat, poorly-drained area characterized by extensive wetlands intersected by numerous rivers (Ecological Stratification Working Group 1995). The Boreal Shield ecozone is a rolling landscape of glacial deposits intermixed with bedrock and many water bodies, and is much more heavily forested than the Hudson Plains (Ecological Stratification Working Group 1995). The southern portion of the Boreal Shield in our study area is within the Area of the Undertaking (AOU) in Ontario, which is the area where commercial forest harvesting occurs.



**Figure 2.** Study area within Ontario’s Far North used for assessing cumulative effects of anthropogenic disturbance on wildlife.

## **Methods**

*Avian count data*  
To estimate bird densities and habitat associations across the study area, we used data compiled by the Boreal Avian Modeling Project (BAM, Cumming et al. 2010). BAM is a research consortium including academic, government, and non-governmental organizations that has compiled avian point count surveys from > 100 different research projects, provincial breeding bird atlases and North American Breeding Bird Survey sites across Canada (Barker et al. 2015). BAM has developed quantitative methods for harmonizing datasets with different survey protocols to produce standardized density estimates (QPAD approach, Sólymos et al. 2013), thus facilitating large scale analysis of bird populations at a much higher resolution than previously possible (Crosby et al. 2019). The QPAD approach reformulates distance sampling models of effective detection radius, and removal models of avian singing rates, to control for survey protocol effects on detection probability, so that:

where is expected count at a given point, is point level density, is area sampled, is the probability of an individual singing during the cumulative time interval (availability), and is the probability of an individual being detected, given singing within the point count radius (detectability) (Sólymos et al. 2013). Thus, ‘QPAD’ refers to different elements of the expected count model, which can be modified according to different temproal and environmental covariates present during a survey (Sólymos et al. 2013). We calculated offsets for data wtihin our study area using the QPAD package for program R (R Core Development Team 2017) which can be downloaded directly at <https://github.com/psolymos/QPAD/>. For methods for formatting data and calculating offsets, see Appendix A.

one of the primary issues to be dealt with was a lack of data within the study area, and in the Hudson Plains ecozone in particular (Fig. 2). In order to ensure the most data possible for making inference within the study area, our initial dataset consisted of all BAM points from northern Ontario north of Boreal Conservation Region (BCR) 12, which included BCRs 7 and 8 within the boundaries of Ontario. Although the majority of the study area was within the Hudson Plains, the majority of the data, both within and surrounding the study area, came from the Boreal Shield ecozone, which is distinctly different from the Hudson Plains in terms of geology, soil, topography, and ecology (Ecological Stratification Working Group 1995). Moreover, much of the boreal shield data came from the AOU, which is much more heavily forested than the rest of the region. This means that species habitat relationships from one region are not likely to be representative from those of another region, and so the bulk of the data was not representative of the bulk of the study area. We addressed this situation provisionally through a random point selection system, where we used all points within the Hudson Plains ecozone (both within and outside of the study area), and randomly selected an equal number of points from outside the Hudson Plains. At point count stations that had multiiple surveys, we randomly chose a single survey for modeling purposes.

*Environmental data*  
We derived habitat and other environmental covariates from regional and national scale maps of land-cover, forest attributes (Beaudoin et al. 2014), topography, and some anthropogenic features. These include: kNN estimate forest attributes from Beaudoin et al. (2014); topographic indices such as eleveation, roughness, and slope; Ontario and Far North land cover maps (<https://geohub.lio.gov.on.ca/documents/lio>::ontario-land-cover-compilation-v-2-0/about); fire and harvest information (Hermosilla et al. 2016); forest structure change (Matasci et al. 2018); and Quaternary geology, among others (Appendix A). All data were standardized to 250m resolution raster maps across the point selection region (BCRs 7 and 8 in Ontario). Most of the data processing was done using program R (R Core Development Team 2017), but some was done in ArcGIS becasue of memory limitations or data storage within a GeoDatabase. Descriptions of the process for acquiring and processing geospatial data are found in Appendix A of this document. Further descriptions of data used can be found in Appendix C of this document.

*Species distribution models*  
We used Boosted Regression Trees (BRT, Elith et al. 2008) to model point level bird densities and predict species distributions in relation to environmental covariates across the study region. BRT models work by fitting multiple individual decision trees to progressively improve predictive accuracy, automatically incorporating interactions and easily handling large numbers of variables (Elith et al. 2008). We fit generalized boosted models to raw counts at each survey, assuming a Poisson error distribution, and used the log correction factors calculated by the QPAD method (Sólymos et al. 2013) as an offset to account for detectability (*sensu* Stralberg et al. 2015). We specified the number of trees as 10,000, a learning rate of 0.001, an interaction depth of 3, and a bag fraction of 0.5. To build the BRT models and generate density predictions, we used the ‘dismo’ (Hijmans et al. 2020), ‘gbm’ (Greenwell et al. 2020), and ‘raster’ (Hijmans and Etten 2014) packages within program R (R Core Development Team 2017).

We used 2 stages of model fitting for each species. First, we mapped out the most important predictor variables for each species by fitting a BRT to a single subset of the data (all Hudson Plains points, plus an equal number of randomly chosen points outside the Hudson Plains). We then used the ‘relative.influence’ function in the gbm package to compute the relative influence of each variable in the model. Second, we used 100 bootstrapped subsamples of the data to estimate the distribution of predicted densities of singing males at each raster cell across the study region. Finally, we used the mean predicted values among bootstrap replicates to estimate population size within the study area, assuming a pair adjustment of two, and calculated densities per land cover type (Sólymos et al. 2020).

**Results**  
We fit BRT models and created density distribution maps for 61 species of breeding landbirds in Far North Ontario (Appendix B). The two most prominent predictor variables across species were Ontario Land Cover and vegetation biomass within a 750m buffer of points.

##### Page break

#### Appendix A: R code and narrative supplements for acquiring and processing geospatial data

This document describes the the acquisition and processing steps for geospatial data in the Ontario’s Far North for the purpose of modeling landbird distributions.

## **Annotated Description of “Ring of Fire” R Project**

This document serves as an annotated checklist of the folder structure and spatial data within the R project “RoF- spatial-data-birds.” The purpose of this R project is to contain raw spatial data (raster layers, shapefiles) for the Ring of Fire region in Ontario, along with analysis scripts for processing that data and extracting spatial information to avian point count locations prior to modelling. By storing spatial data and code in a coherent, well-documented framework, the “RoF-spatial-data-birds” project will: 1) make results from analysis of bird survey data more transparent and reproducible; 2) facilitate updates of analysis results with new bird survey data; 3) facilitate application of analysis techniques to other kinds of survey data (e.g., amphibians, mammals).

Most of the raw spatial data that has been extracted to point counts was processed further spatially within R prior to data extraction. In some cases, however, spatial data was processed outside of R in ArcGIS because processing was easier or due to limitation of memory in R or because the initial raw data was stored in an ArcGIS geodatabase, which is a data structure that cannot be processed within R. In any of these cases, instructions for obtaining those data products will be indicated. This R project does not contain scripts for the actual boosted regression tree models that were run for the Ring of Fire in 2020-2021. Those models are in a separate deliverable along with outputs due to the memory requirements of the outputs.

The R project can be obtained at ***LINK TO BE PROVIDED BY CANADIAN WILDLIFE SERVICE***

### “0\_data” Folder

This folder stores all the point count data used in boosted regression tree models in the Ring of Fire region, along with spatial data extracted to those point counts. There are two subfolders for raw and processed data. Raw data is where initial spatial data downloaded from the web and other sources are stored, along with point count data obtained from the Boreal Avian Modelling Project and other sources. In most cases, spatial data needs to be processed further in a geographic information system (GIS) before it can be extracted to point counts. R itself can be used like a GIS to process spatial data and most spatial data is further processed in R within this R project. There will be notes about cases where raw data had to be processed in ArcGIS first before entry into R, and what to do about getting the data products that are used in R.

### “1\_scripts” Folder

This folder stores all the R code necessary for processing spatial data and extracting that data to point counts in the Ring of Fire region. Individual scripts are numbered according to the order in which they could be done, although there is some flexibility if scripts don’t require an input from another script that hasn’t previously been run.

### Outputs from this project

Beyond the data and scripts folders, R projects may have other folders where intermediate or final outputs and reports are stored. Since the purpose of this R project is to prepare data prior to running models, there are no model outputs. Two of the folders in “0\_data/processed” contain data that will be used in models.

### “0\_data” Subfolders

There are separate folders within this folder for point counts and the raw spatial data that is either extracted to those point counts or used to clip spatial data to smaller areas.

### “0\_data/raw/point counts” Folder

Raw point count data stored in this folder comes from three sources: the Boreal Avian Modelling Project (BAM), North American Breeding Bird Survey (BBS), and the Ontario Forest Bird Monitoring Program (FBMP). BAM already has some but not recent years of FBMP data in its database.

Point count data were initially visualized in maps relative to the Ring of Fire study area to summarize how much data from different sources was available for modelling. Based on these results, FBMP data was not included in the Ring of Fire models because virtually all the data was collected outside of the study area.  
BBS data were downloaded from the web for all years up to 2019 and are completely open source but must be processed further. Processing steps include obtaining latitude/longitude coordinates for individual stops, filtering out poor-quality data collected in marginal conditions, reducing point counts to those in Ontario, summarizing species counts per stop per year, and generating QPAD detection offsets. BBS data is currently stored in long data format in an R data file called **“BBS.91.19.ON.stopsFeb2021”**. A separate R project is used to process raw BBS data to the data needed for spatial data extraction. If that R project is used to obtain additional BBS surveys more recent than 2019, then the name of the BBS R data file will need to be changed within the scripts.

BAM data were extracted from the BAM version 6 Access database, using an ODBC connection. Instructions for setting up such a connection on a laptop or desktop are included in the script. The connection cannot be automated and must be set up manually. BAM data are constantly updated and will be stored on a different platform (WildTrax) soon, at which point steps for obtaining BAM data will need to be updated.

BBS and BAM spatial data locations are combined and used to extract data from all the spatial data layers in the final scripts of the R project. Point count data that have been assigned predictor values from the layers are stored in: “0\_data/processed/point counts with data.”

### “0\_data/raw/spatial data” Folder

There are multiple folders containing either the original raw data downloaded from the web and other sources or links to the original data if the size of the raw files is extremely large. Even where the original data are stored in the R project, links to sources are provided in the scripts and in Table 1 accompanying this R project. Ideally and in most cases, all processing of spatial data is done within R. In those cases where some processing had to be done in ArcGIS, a note is provided in the script and intermediate data products from ArcGIS are stored.

Raw spatial data include the following folders:

**Acidity:** GIS layers of critical loads of acid deposition for soils have been produced by Julian Aherne at Trent University, for either 95% or 80% protection of the most sensitive tree species. The extent to which those critical loads are exceeded (respectively, exceedance 5 % and exceedance 20 %) can be used as one measure of potential acid rain impacts on birds in the Ring of Fire region.

**BCR\_Terrestrial\_master:** a shapefile of Bird Conservation regions in North America along with a smaller shapefile containing Bird Conservation Regions 7 and 8 in Ontario.

**Beaudoin:** national-scale raster layers of forest structure at 250-m resolution, including live above-ground biomass (tonnes/ha), forest stand age, and percent coverage by different tree species. Estimates of forest structure predictors were generated for 2001 and 2011 across Canada. Earlier point counts (<2006) were assigned estimates from the layers for 2001 while more recent point counts and prediction raster layers were assigned estimates from the layers for 2011.

**Canada shapefile:** stores a shapefile for Canada used to generate an Ontario raster mask for clipping spatial data.

**Far North Land Cover:** raster layer data at 30-m resolution for different land cover categories across Ontario. More detailed land cover categories at higher resolution in Ontario than those provided by “North American Land Cover 2005 MODIS.” More wetland types than provided by “Ontario Land Cover Compilation” but smaller extent: used for point counts in Ring of Fire study area that are within the Far North Land Cover layer boundaries.

**ForestHeightLidar:** 30-m raster layer with estimates of forest height.

**Global LANDSAT 30TM Tree Cover Layer:** 30-m raster layer with estimates of forest canopy cover.

**Linear Disturbances:** raster layer of distance to nearest road in Ontario, generated from the “Roads On Off Layer Ontario” layer. Used in exploring potential effects of road development on birds in Ring of Fire region.

**Mining Locations:** locations of active and abandoned drill holes used in exploring potential effects of mining (distance to nearest mine) on birds in Ring of Fire region.

**NA\_PoliticalDivisions:** North American shapefile used in maps.

**North America ASCII DEM Elevation:** 1-km elevation data from a digital elevation model (DEM).

**North American Land Cover 2005 MODIS:** Land cover categories from continental raster at 250-m resolution in 2005.

**NTEMS Landsat Layers Derived for 2015:** recent derived Landsat products at 30-m resolution across Canada in extent. Provide measures of recent disturbance (fire mask, fire year, fire severity, harvest mask, harvest year), forest height, forest volume, and forest biomass as of 2015. Due to the size of the original files and difficulties with clipping and resampling the files in R, clipping, and resampling of these files was done in ArcGIS. A link to the original data is provided in the scripts and intermediate products from ArcGIS are stored in this folder.

**Ontario Land Cover Compilation:** raster layer data at 30-m resolution for different land cover categories across Ontario. More detailed land cover categories at higher resolution in Ontario than those provided by “North American Land Cover 2005 MODIS.” Slightly fewer wetland types than provided by “Far North Land Cover Layer” but greater extent: used for point counts in Ring of Fire study area that are outside the Far North Land Cover layer boundaries.

**Ontario Hydro Network:** dam locations in Ontario, used to estimate the number of point counts potentially influenced by hydroelectric development, i.e., point counts within 30 km of a dam. Based on map results, these data were not considered further in Ring of Fire bird models.

**Quaternary Geology:** shapefile data for different landform types, more specific than the landform categories stored in “0\_data/raw/spatial data/Topography.” Includes landforms shaped by glaciation such as eskers and uplands, which may support different habitat and bird communities than lowlands in the Ring of Fire region.

**Ring of Fire Boundaries:** various study area boundaries provided by the Ring of Fire working group.

**Road OnOff Layer Ontario:** whether a 250-m pixel was classified as containing a roadside or not. Used in exploring potential effects of road development on birds in Ring of Fire region.

**Topography:** Slope, topographic position index, terrain ruggedness index, landform (10 categories).

Data from each of these folders was clipped to Ontario and stored in folders with the same or near-same names in “0\_data/processed” and clipped to the study region (Bird Conservation Regions 7 and 8 in Ontario) in “0\_data/processed/prediction rasters.”

### “0\_data/processed” Folder

Besides folders containing the clipped layers for each of the raw data (Fig. 5), there are two folders with “outputs” from this R project that will be transferred to bird models in the Ring of Fire region:

**“0\_data/processed/point counts with data”** contains R data files with point counts compiled from the Boreal Avian Modelling Project and North American Breeding Bird Survey, which occur within Bird Conservation Regions 7 and 8 in Ontario (the area used for modelling). Predictor values have been extracted from the spatial data layers to these point counts.

**“0\_data/processed/prediction rasters”** contains spatial layers for all the predictor variables, which have been clipped to Bird Conservation Regions 7 and 8 in Ontario and resampled to the same resolution (250 m) and reprojected to the same coordinate system (Lambert Conformal Conic). As all the clipped raster layers have the same resolution and projection, they can be made into a raster stack in which each pixel location is assigned values from all the model predictors, enabling R to calculate predicted densities (males/ha) at each pixel location and create spatial distribution maps for each bird species.

### “1\_scripts” Folder

There are currently 16 scripts in the R folder, numbered in the order in which they are done, since some files or data objects in higher-numbered scripts are produced in lower-numbered scripts.

**“1\_lookatpointcountsminesanddamsinRoFregion”** is a script for taking the point count data BAM already has and running GIS intersections of the point count data with the Ring of Fire study region. GIS operations are performed in R. Output from this script consists of annotated maps showing: 1) the number of stations and survey visits, mines, and dams in different areas, along with relative contributions of different data sources to point counts; 2) number of stations near (within a specified distance of) mines and dams. Based on these maps, “distance to nearest mine” was considered for Ring of Fire bird models but “distance to nearest dam” was not.

**“2\_BeaudoinData\_cliptoOntariostudyarea\_RingOfFire”** is a script for taking remotely sensed raster data of forest age and structure at a Canadian extent, clipping it to the study region, and summarizing the data at two additional spatial scales larger than “local” scale (the original 250-m resolution of the raster layers). GIS operations are performed in R. Output from this script consists of processed raster layers stored in “0\_data/processed/Beaudoin,” in one of two subfolders, “2001” or “2011,” based on the year of the data.

**“3\_ExceedanceData\_cliptoOntariostudyarea\_RingOfFire”** is a script for taking shapefile layers containing acid rain exceedance values at a Canadian extent, rasterizing the exceedance values for different variables, and clipping those raster layers to the study region. GIS operations are performed in R. Output from this script consists of raster layers stored in “0\_data/processed/Acidity/.”

**“4\_NTEMS\_HarvestData\_cliptoWOTHstudyarea”** was to be a script for taking remotely sensed raster data of forest harvest and fire age and structure (forest height, volume, and biomass) at a Canadian extent, clipping it to the study region, then resampling the layers to 250 m from 30-m resolution to permit raster stacking. Data are more recent (2016) and at finer resolution (30 m) than the Beaudoin satellite data from 2001-2011. GIS operations are performed in R. We ended up finding it easier to do the clipping and resampling in ArcGIS due to excessive memory and time requirements in R, combined with concerns about errors and poor data quality of the clip in R.

**“5\_Elevation\_cliptoOntariostudyarea”** is a script for taking remotely sensed raster data of elevation at a continental extent and clipping it to the study region. GIS operations are performed in R. Output from this script consists of processed raster layers stored in “0\_data/processed/Elevation.”

**“6\_Topography\_cliptoOntariostudyarea”** is a script for taking remotely sensed raster data of slope, topographic position index, and terrain ruggedness index at a continental extent, and clipping it to the study region. GIS operations are performed in R. Output from this script consists of processed raster layers stored in “0\_data/processed/Topography.”

**“7\_TreeCover30m\_cliptoOntariostudyarea”** is a script for taking remotely sensed raster data of tree cover (30 m resolution) at a global extent and clipping it to the study region. GIS operations are performed in R. Output from this script consists of processed raster layers stored in “0\_data/processed/Global LANDSAT 30TM Tree Cover Layer.”

**“8\_NALC2005Ontariostudyarea”** is a script for taking remotely sensed raster data of land cover classes (250 m resolution) at a continental extent and clipping it to the study region. GIS operations are performed in R. Output from this script consists of processed raster layers stored in “0\_data/processed/North American Land Cover 2005 MODIS.”

**“9\_roadsonoffOntariostudyarea”** is a script for taking remotely sensed raster data of road data at a continental extent, clipping it to the study region, and assigning pixels as on or off-road. GIS operations are performed in R. Output from this script consists of a processed raster layer stored in “0\_data/processed/RoadsOnOff Layer Ontario.”

**“10\_roaddistancecliptoOntariostudyarea”** is a script for taking a distance raster to the nearest road created in ArcGIS (250 m resolution) and clipping it to the study region. Output from this script consists of a processed raster layer stored in “0\_data/processed/Linear Disturbances.”

**“11\_minedistancecliptoOntariostudyarea”** is a script for taking a distance raster to the nearest mine created in ArcGIS (250 m resolution) and clipping it to the study region. Output from this script consists of a processed raster layer stored in “0\_data/processed/Mining Locations.”

**“12\_FarNorthLandCoverstudyarea”** is a script for taking remotely sensed raster data of land cover classes from the Ontario Far North Land Cover product (stitched together in ArcGIS from three separate layers), resampled from 30 m to 250 m resolution, clipping it to the study region, and summarizing the data at two additional spatial scales larger than “local” scale (the resampled 250-m resolution of the raster layers). Except for stitching the three original layers together in ArcGIS, GIS operations are performed in R. Output from this script consists of processed raster layers stored in “0\_data/processed/Far North Land Cover.”

**“13\_OntarioLandCoverstudyarea”** is a script for taking remotely sensed raster data of land cover classes from the Ontario Land Cover Compilation product (resampled from 30 m to 250 m resolution in ArcGIS), clipping it to the study region, creating binary files of each cover class, and summarizing the data at two additional spatial scales larger than “local” scale (the resampled 250-m resolution of the raster layers). Except for stitching the three original layers together in ArcGIS, GIS operations are performed in R. Output from this script consists of processed raster layers stored in “0\_data/processed/Ontario Land Cover.”

**“14\_ExtractDataToOntarioPointCounts”** is a script for converting all point counts (BAM, BBS) to a spatial points data frame, then extracting data from the processed raster layers created in scripts 2 to 13 to those point counts. “15\_QuaternaryGeology\_2021\_04\_23” is a script for extracting some additional geological data related to landforms (identification of upland vs. lowland point counts) to the point counts mentioned in script 14.

**“16\_ClipSpatialLayersToBCR7and8”** is a script for making sure that all raster layers (from which spatial data are extracted to point counts) are clipped to the same extent and projection to enable raster stacking (the use of these raster layers in predicting abundance of bird species based on model prediction).

##### Page break

#### Appendix B: R code and narrative supplements for fitting BRT models

This document describes the processing and modeling steps for bird-habitat data in the Ontario’s Far North for the purpose of creating species distribution models and making predictive maps of species distributions, estimates of population size, and estimates of mean density by land cover type.

## **Purpose**

This document serves as a guideline for creating Species Distribution Models (SDM) for bird species in the Ring of Fire study area using Boosted Regression Trees. R code and data is stored in an R project titled “ring-of-fire-brt” that can be accessed at:

***LINK TO BE PROVIDED BY CANADIAN WILDLIFE SERVICE***

The google drive project contains all data for running models, non-bootstrapped outputs with variable importance and associated plots, and model outputs, predictions, and maps for 2 species (ALFL and AMCR) using to bootsrap resamples.

Code for this project, without data or outputs, can be found at:

<https://github.com/crosbya1/ring-of-fire-brt>

Models can be reproduced using code in the “1\_scripts” folder by running the .R files in the order in which they are numbered.  
Data used for generating the models is accessed from the “RoF-spatial-data-birds” R project described in Appendix A of this report.1

1Note that BAMv6\_ONBBS.RData in the project folder is not the same as the one from the RoF-spatial-data-birds project on G-Drive. There is is an error on the G-Drive version so that it doesn’t calculate offsets for all surveys, so the code doesn’t quite work with that one. The version in the ring-of-fire-brt project is an earlier version that works with this code.

Resolution of all raster maps is 250m2

## **Folder and file structure**

This project uses the same folder structure as the project described in **Appendix A**.

### 0\_data

* **processed**  
  – **archive:** older files no longer in use – **prediction-rasters:** raster files accessed from the same folder in RoF-spatial-data-birds  
  – **predictor-layers:** files generated by running code from this project  
  – **shapefiles:**  
  – other files generated from code in this project
* **raw**  
  – **archive:** older files no longer in use  
  – **point\_counts\_with\_data:** files accessed from the “point counts with data” folder in RoF-spatial-data-birds  
  – **shapefiles:**  
  – BAMv6\_ONBBS : point counts and offsets for Northern Ontario

### 1\_scripts

* **archive:** older scripts no longer in use
* **functions:** any saved functions would be here (should be empty)
* **model-scripts:** any specialized model scripts would be here (should be empty)
* R scripts for creating data and running models

### 2\_pipeline

* **store:** stored objects from running scripts  
  – **brt2-xv:** non-bootsrtapped BRT model objects for each species  
  – **brt2-xv-pred-mosaic:** Plots of improtant variables from brt2-xv model objects  
  – **brt-boot-1:** bootstrapped BRT model objects for each species – **brt-boot-pred:** mean density predictions for northern Ontario derived from bootstrapped BRT models  
  – **brt-boot-pred-mosaic:** raster maps of mean density predictions for each species.
* **temp:** temporary objects generated from running scripts (should be empty)

### 3\_outputs

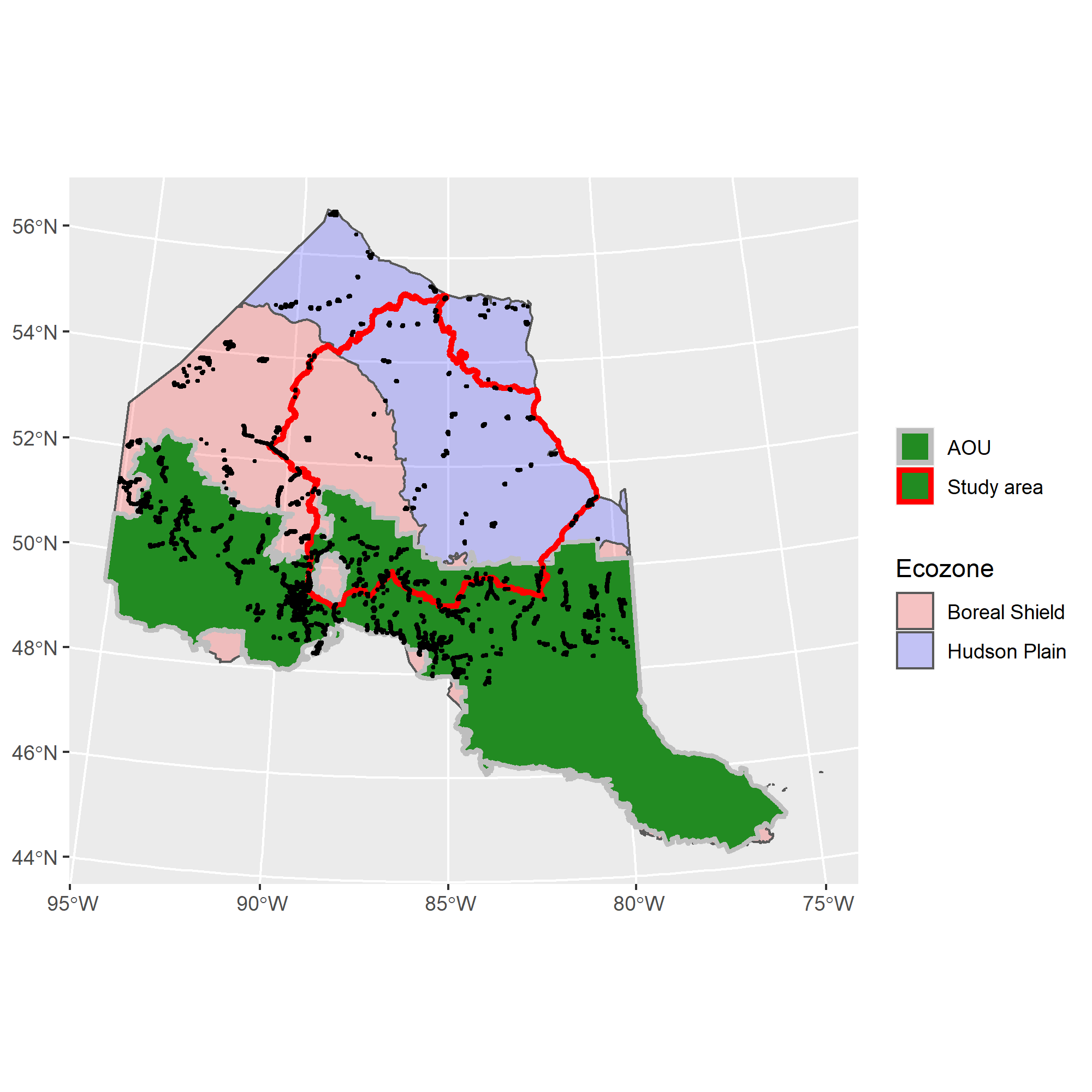
* **figures:** any figures generated by running the code
* **maps:**  
  – **brt-boot-pred-mosaic:** .png maps of predicted mean density across northern Ontario
* **tables:**  
  – SppBRTVarImp\_v2.csv : variable importance for each species  
  – SppDensityByOLCC\_v2.csv : predicted density for each land cover type for each species

### 4\_reports

* **common:** files used for creating reports with R markdown
* R markdown code and output for Appendix C in this report

##### Page break

#### Appendix C: Point count and environmental data used for RoF analysis



**Figure 1.** Spatial distribution of BAM point count stations (black dots) north of Bird Conservation Region 12 in Ontario.

## The breakdown of point locations is as follows:

Total: 11566  
Hudson Plain Ecozone: 1460  
Boreal Shield Ecozone: 10106

In AOU: 8487

Study Area total: 2387  
Study Area Hudson Plain: 860  
Study Area Boreal Shield: 1527  
Study Area AOU: 1309  
Study Area Far North: 1078

# Variables extracted for BAM point count stations within the larger region. This consistst of points in Ontario north of BCR 12.

All raster variables were standardized to 250m resolution raster values. If original raster was <250m resolution, the mean value among cells was used.

## Variables (and name in database)

**Elevation** (elevation): height above sea level in meters  
**Tree cover** (treecover): percentage of area covered by tree crowns  
**On road?** (road\_yesno): point on road (1) or off road (0)  
**Distance from road** (road\_dist\_m): distance of pixel containing point from nearest road in meters  
**Adjusted distance from road** (road\_dist\_m.adj): distance adjusted for width of pixel  
**Distance to mine site** (mine\_dist\_m): distance of pixel to nearest mine site in meters  
**Slope** (slope): average slope within pixel (degrees)  
**TRI** (TRI): Topographic roughness index – The topographic ruggedness index (TRI) was developed by Riley, et al. (1999) to express the amount of elevation difference between adjacent cells of a DEM. Indicates the ruggedness of the terrain surrounding the cell. Scale-dependent.  
**TPI** (TPI): Topographic position index – Calculates whether a raster cell is on a hill, valley, or in between. Positive TPI values represent locations that are higher than the average of their surroundings, as defined by the neighborhood (ridges). Negative TPI values represent locations that are lower than their surroundings (valleys). TPI values near zero are either flat areas (where the slope is near zero) or areas of constant slope (where the slope of the point is significantly greater than zero). Scale-dependent.

### Landcover classes: These were calculated at the local level [local; pixel is in that class (1) or not (0)] and with a Gaussian filter at 250m (\_G250) and 750m (\_G750), where the value is the distance-weighted proportion of pixels in that land-cover class surrounding the point.

#### Far North Land-cover and Ontario Land-cover

Bedrock:  
Communities:  
Conifer Swamp:  
Conifer Treed:  
Deciduous Swamp:  
Deciduous Treed:  
Disturbed non-woody:  
Disturbed treed:  
Freshwater marsh:  
Heath:  
Intertidal marsh:  
Mixed treed:  
Mudflat:  
Open bog:  
Open fen:  
Open water:  
Sand-gravel mines:  
Sparse treed:  
Supertidal marsh:  
Thicket swamp:  
Treed bog:  
Treed fen:  
Turbid water:

#### LCC 2005

Barren:  
Cropland:  
Polar barren:  
Polar grass:  
Polar shrub:  
Taiga:  
Temperate or sub-polar broadleaf:  
Temperate or sub-polar grassland:  
Temperate or sub-polar mixed:  
Temperate or sub-polar needleleaf:  
Temperate or sub-polar shrub:  
Urban:  
Water:  
Wetland:

### Forest attributes from *k*NN MODIS imagery [the Beaudoin layer; Beaudoin et al. (2014)]. These were also extracted at the local level and with 250m and 750m Gaussian filters.

#### Species composition

Non-vegetated %: percent area in non-vegetated land-cover  
Vegetated %: percent of area in vegetated land-cover  
Treed %: percent of area in treed land-cover  
Vegetated non-treed %: percent of area in open (grass and shrub) vegetated land-cover  
Species-specific canopy cover %: pretty much every species you can think of  
Broadleaf canopy cover %:  
Needleaf canopy cover %:  
Unknown canopy cover %:

#### Structure

Total live above ground biomass (AGB): in tonnes/ha  
Total dead above ground biomass: dead standing biomass in tonnes/ha

##### As percentage of total live AGB

Branch biomass:  
Foliage biomass:  
Bark biomass:  
Wood biomass:

Age: years  
Crown closure %:  
Height: meters  
Merchantable volume: m3/ha  
Total volume: m3/ha

### Other attributes derived from new national layers. Exceedence is ian index of the ability of the area to buffer acid rain.

Fire mask: burned or not  
Fire year:  
Fire severity:  
Harvest mask: harvested or not  
Harvest year:  
Biomass in 2015:  
Volume in 2015:  
Height in 2015:  
Exceedence 05:  
Exceedence 20:

### Added esker data re-sampled to 250m (esker Y/N)

##### Page break

## **References**

Barker NKS, Fontaine PC, Cumming SG, et al (2015) Ecological monitoring through harmonizing existing data: Lessons from the boreal avian modelling project. Wildlife Society Bulletin 39:480–487. <https://doi.org/10.1002/wsb.567>

Beaudoin A, Bernier PY, Guindon L, et al (2014) Mapping attributes of Canada’s forests at moderate resolution through k NN and MODIS imagery. Canadian Journal of Forest Research 44:521–532. <https://doi.org/10.1139/cjfr-2013-0401>

Crosby AD, Bayne EM, Cumming SG, et al (2019) Differential habitat selection in boreal songbirds influences estimates of population size and distribution. Diversity and Distributions 25:1941–1953. <https://doi.org/10.1111/ddi.12991>

Cumming SG, Lefevre KL, Bayne E, et al (2010) Toward Conservation of Canada’s Boreal Forest Avifauna: Design and Application of Ecological Models at Continental Extents. Avian Conservation and Ecology 5:art8. <https://doi.org/10.5751/ACE-00406-050208>

Ecological Stratification Working Group (1995) A National Ecological Framework for Canada

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. Journal of Animal Ecology 77:802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>

Greenwell B, Boehmke B, Cunningham J (2020) Package ’gbm’

Hermosilla T, Wulder MA, White JC, et al (2016) Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring. International Journal of Digital Earth 9:1035–1054. <https://doi.org/10.1080/17538947.2016.1187673>

Hijmans RJ, Etten J van (2014) raster: Geographic data analysis and modeling

Hijmans RJ, Phillips S, Leathwick J, Elith J (2020) Package ’dismo’

Matasci G, Hermosilla T, Wulder MA, et al (2018) Three decades of forest structural dynamics over Canada’s forested ecosystems using Landsat time-series and lidar plots. Remote Sensing of Environment 216:697–714. <https://doi.org/10.1016/j.rse.2018.07.024>

R Core Development Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

Sólymos P, Matsuoka SM, Bayne EM, et al (2013) Calibrating indices of avian density from non-standardized survey data: making the most of a messy situation. Methods in Ecology and Evolution 4:1047–1058. <https://doi.org/10.1111/2041-210X.12106>

Sólymos P, Toms JD, Matsuoka SM, et al (2020) Lessons learned from comparing spatially explicit models and the Partners in Flight approach to estimate population sizes of boreal birds in Alberta, Canada. The Condor 122:1–22. <https://doi.org/10.1093/condor/duaa007>

Stralberg D, Bayne EM, Cumming SG, et al (2015) Conservation of future boreal forest bird communities considering lags in vegetation response to climate change: a modified refugia approach. Diversity and Distributions 21:1112–1128. <https://doi.org/10.1111/ddi.12356>