BAM Generalized National Models Documentation

Peter Solymos & Diana Stralberg

Version 4.0 (2020-09-08)

Contents

1 Introduction

Reliable information on species' population sizes, trends, habitat associations, and distributions is important for conservation and land-use planning, as well as status assessment and recovery planning for species at risk. However, the development of such estimates at a national scale is challenged by a variety of factors, including sparse data coverage in remote regions (Stralberg et al. 2015), differential habitat selection across large geographies (Crosby et al. 2019), and variation in survey protocols (Sólymos et al. 2013).

With these factors in mind, we developed a generalized analytical approach to model species density in relation to environmental covariates, using the Boreal Avian Modelling Project database of point-count surveys (through 2018) and widely available spatial predictors (Cumming et al. 2010, Barker et al. 2015). We developed separate models for each geographic region (bird conservation regions intersected by jurisdiction boundaries) based on covariates such as tree species biomass (local and landscape scale), forest age, topography, land use, and climate. We used machine learning to allow for variable interactions and non-linear responses while avoiding time-consuming species-by-species parameterization. We applied cross-validation to avoid overfitting and bootstrap resampling to estimate uncertainty associated with our density estimates.

1.1 Contact

Please contact us if you have questions or suggestions via these channels:

- borealbirds.ualberta.ca
- Twitter

- [GitHub](#)

1.2 Citing the models

Please cite this document when using the BAM National Model results as:

Boreal Avian Modelling Project, 2020. *BAM Generalized National Models Documentation, Version 4.0.* Available at https://borealbirds.github.io/. DOI: [10.5281/zenodo.4018336](https://doi.org/10.5281/zenodo.4018336).

2 Methods

2.1 Data

2.1.1 Study area

Models were developed based on data from non-arctic portions of Canada. Sampling effort varied greatly across the study area. In general, northern environmental conditions were underrepresented, and southern boreal conditions were overrepresented in comparison with the rest of the study area.
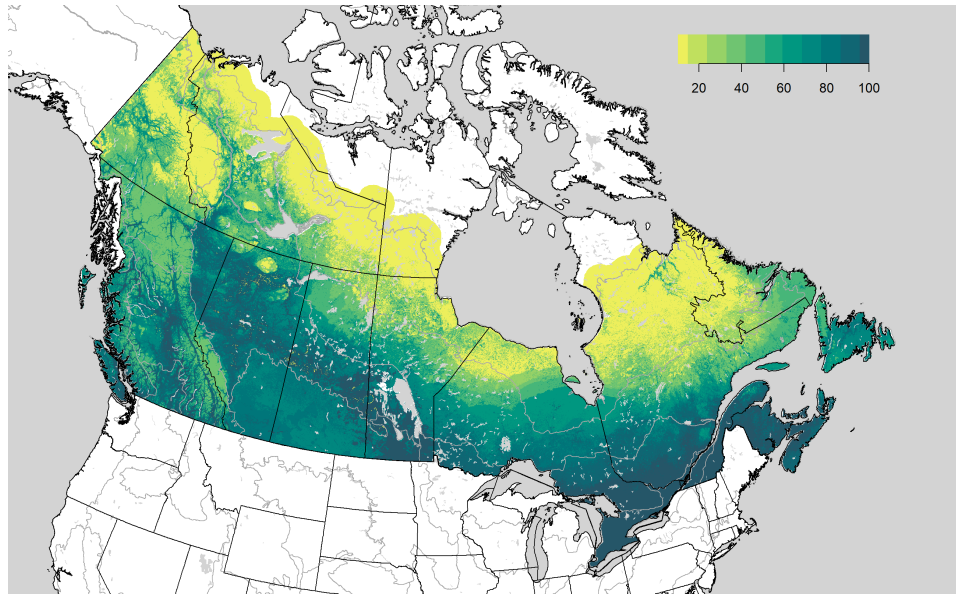
Percentiles of predicted survey effort (number of sites surveyed) based on ~100 environmental covariates and a single boosted regression tree model with a Poisson distribution. Data represent a subsample (1 million 1-km pixels)

of the Canadian study area indicated on the map. Mean number of sites surveyed per 1-km pixel = 0.03.

Separate models were constructed for each of 16 separate spatial subunits consisting of bird conservation regions (BCRs) intersected with Canadian jurisdictional boundaries. Smaller BCR x jurisdiction intersections were merged to maintain adequate sample sizes.
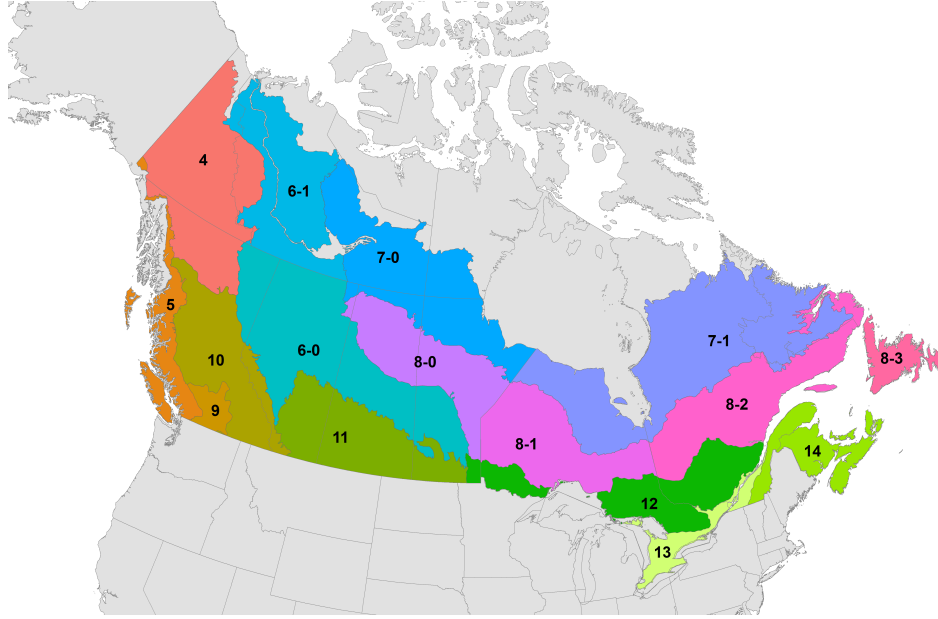
Model-building units based on a combination of bird conservation regions and Canadian jurisdiction boundaries.

2.1.2 Avian data and subsampling

Avian data were extracted from the BAM avian dataset (v. 4) and supplemented with automated recording unit (ARU) data from the WildTrax acoustic database (6,801 surveys between 2012 and 2018). In total, we sampled data from 296,061 point counts across Canada (across a total of 256,316 site locations and 175 distinct projects). North American Breeding Bird Survey and provincial Breeding Bird Atlas data were included in this database, and constitute a significant fraction of available data (30% and 53% respectively).

We present models for 143 landbird species for which density offsets were available and for which data were sufficient to fit cross-validated BRTs in

at least one of the 16 regions. Point-count surveys were conducted between 1991 and 2018 (97% of the point counts were from between 1997 and 2014). We stratified samples by year and geography to produce a more spatially and temporally balanced dataset. We used a 2.5 km x 2.5 km resolution spatial grid to define spatial 'clusters' of data. We resampled the data set in each region so that we had a single data point from each cluster/year combination and fit BRTs to the resampled data set. This subsampling addressed instances where multiple visits to the same location occurred within the same year. The subsampling was repeated 32 times.

The list of species is part of the downloadable Excel file.

2.1.3 Environmental covariates

Model inputs consisted of 219 spatially explicit environmental covariates, as well as survey year (continuous) and survey type (binary).

To capture the influence of changing landscape conditions on avian density, we used vegetation maps from 2001 and 2011 (Beaudoin et al. 2014) and associated our survey data with the layer that represented the closest time period. Surveys conducted in 2005 or earlier were associated with the 2001 dataset, while surveys from 2006 and later were associated with the 2011 dataset. Vegetation variables were derived at a 250-m spatial resolution from k-nearest-neighbor (kNN) models that used forest sample plots from Canada's National Forest Inventory combined with MODIS satellite imagery, as well as climate and terrain data (Beaudoin et al. 2014). Vegetation variables included pixel-level and landscape-level biomass of individual tree species and stand age. Landscape-level covariates were calculated using the focal function in the raster package for R, and were based on a moving-window average using a Gaussian weighting of surrounding pixels (one standard deviation = 750 m).

To capture other sources of landscape variation not represented in vegetation data, we supplemented the biomass and stand age covariates with several terrain, land use, and climate variables. Terrain metrics, calculated using the terrain function in the raster package for R were based on a 100-m digital elevation model for North America. Land-use and landcover variables were based on the 2005 MODIS-based 250-m North American landcover map (Commission for Environmental Cooperation). A binary (0/1) 1-km road variable (Venter et al. 2016) was used to account for the influence of roads at a broad scale. Climate variables were based on a 1-km Climate NA

interpolation of 1981-2010 weather station data (Wang et al. 2016).

We pre-screened the environmental predictor variables to eliminate constant (no variation in a BCR subunit) or highly correlated (Pearson's correlation > 0.9) variables. We also eliminated variables that never entered the cross-validated BRTs to further narrow the variable set for bootstrap to boost computing speed.

The list of covariates is part of the downloadable Excel file.

2.2 Analyses

2.2.1 Density calibration, detectability offsets

We accounted for differences in sampling protocol and covariate effects on detectability using statistical offsets. This included the effects of time of day and day of year on the probability of availability given presence, and the effects of tree cover and land-cover type on the probability of detection given availability (Sólymos et al. 2013). Offsets were calculated based on removal and distance-sampling models (Sólymos 2016, Sólymos et al. 2018). These models were used to predict availability and detectability for each species given survey-specific covariates. The adjustments appeared as offsets in the BRTs so that expected values represented species density.

We assumed that ARU detectability is similar to detectability by human observers (Yip et al. 2017). Nevertheless, we used an indicator variable to account for possible differences in effective area sampled between human counts and ARUs following Van Wilgenburg at el. (2017).

2.2.2 Model building

Separate models were constructed for each BCR subunit plus a 100-km buffer around the Canadian portion of the perimeter. We implemented boosted regression tree (BRT) models using the gbm.step function in the dismo R package with a Poisson distribution and 10-fold cross-validation in a preliminary run to assess the number of boosting iterations required to avoid over-fitting. We capped the number of iterations (trees) at 10,000 maximum. BRT settings were as recommended by Elith et al. (2008) and were consistent with Stralberg et al. (2015). We used the number of trees established based on the cross-validation to run models for each bootstrap sample using the gbm function in the gbm R package. This yielded 32 BRT outputs per species and subregions.

We provide information on variable importance for each species by subunits, and for Canada as an average across the subunits (see in the downloadable Excel file). The top ranking variables with respect to variable importance were year of survey, temperature difference, average summer temperature, summer and annual heat/moisture index, and the proportion of developed areas, black spruce, and aspen.

2.2.3   Model validation

We calculated validation metrics using the training data set by making 32 predictions given the bootstrap based BRT outputs. Scale and location shifts across bootstrap based predictions were evaluated by the overall concordance correlation coefficient (OCCC; Lin 1980, Barnhart et al. 2002). OCCC measures the deviation from 1:1 line through the origin, i.e. perfect agreement between two measures. OCCC is the product of two the overall precision (how far each observation deviated from the best fit line), and the overall accuracy (how far the best line deviates from the 1:1 line).

We used the bootstrap averaged predictions to calculate expected values under the null model [exp(initial intercept estimate of the BRT + offsets)] and the final BRT [estimate from all trees combined x exp(offset)]. These initial and final predictions were used to calculate AUC (initial and final) to assess classification accuracy (counts treated as detection / non-detection) and pseudo R2 to quantify the proportion of variance explained (based on Poisson density based deviance relative to the null and saturated models).

Validation results are part of the downloadable Excel file.

2.3   Predictive mapping

We used the subregional BRT results to make species and subregion specific predictions using 1 km2 resolution raster layers as predictors. Our predictions represent the expected number of male individuals per ha area given off-road habitat and human observers. We generated 32 predictions for each species x subregion combination. When the actual bootstrap sample did not contain any detections of the species we predicted 0. For each species, we mosaiced together the 16 subregion predictions for a bootstrap run (runs were independent across regions). We varied the width of the overlap zone between subregions (0-100 km) to smooth the predictions at the edges of the subregions and to avoid banded patterns. The random buffer was based on the cumulative density of the Beta(2, 2) distribution. We then averaged the

7

32 mosaiced layers to get the bootstrap mean of the predictions.

3   Results

The results from the BAM National Models are available at boreal-birds.github.io for each species.
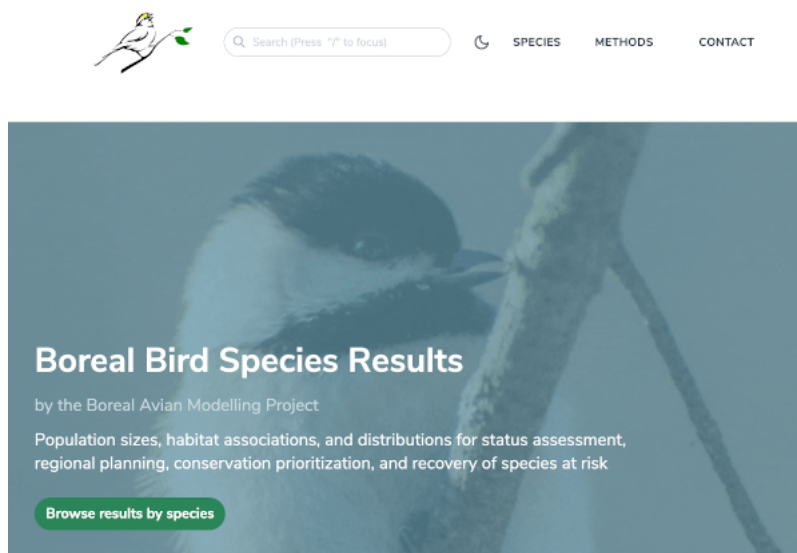

3.1   Species specific results

We walk through the results for Canada Warbler. The same results are available for all 143 species.

3.1.1   Density map

We developed separate models for each BCR subunit to improve local prediction accuracy and avoid out-of-range prediction. However, this resulted in some sharp transitions in predictions across certain boundaries that coincide with large regional differences in density. This variation in density across a large study area presented challenges for mapping, and we had to balance mapping detail with aesthetics to produce meaningful national maps. We emphasize that categorical map legends necessarily introduce subjectivity into the interpretation of species' distribution and abundance patterns, and note that the legend breaks we used may not be the best ones for any particular mapping need. We encourage users to download the raster predictions

and develop their own maps for regional applications.

Based on previous work, we started by developing maps that used mean density (males/ha) within the model-building area (Stralberg et al. 2015) as a presence/absence threshold, with areas of density below this mean density ("absence") represented in light yellow. However, we found that this did not adequately describe the abundance patterns of all species, especially those that are widely distributed. So we adjusted the minimum thresholds according to visual alignment with known range limits. If maps based on these mean density thresholds resulted in a non-trivial number of occurrence locations mapped as absence (light yellow), then we sequentially adjusted these thresholds downward until that was no longer the case (starting with 0.05, then 0.01, then 0.001). 0.001 males/ha was the lowest density that we allowed to be used for this lower density threshold. Equal-interval legends, capped at the 99th percentile of predictions, were used to classify remaining density predictions for mapping.
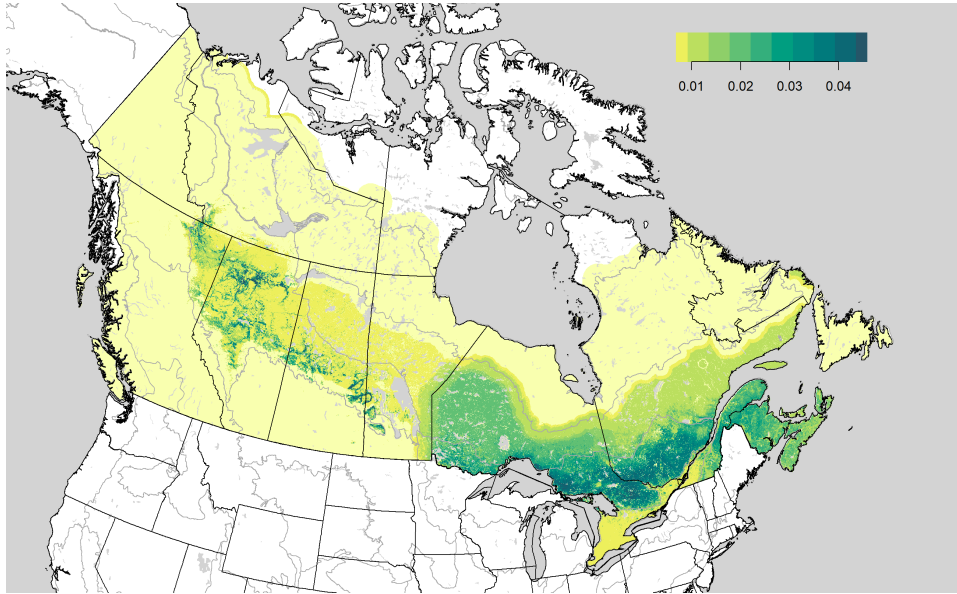
The trade-off to this mapping approach is that there is perceived "over-prediction" in non-range areas (usually in the north, and often only in some BCRs). In some cases this may be related to range map inaccuracies in northern regions, but it also has to do with the sparsity of data in the north and the model's inability to identify covariates that control presence/absence. Additional data are needed to map northern range limits more accurately.

We also present the density maps with species' detections and range map overlaid.
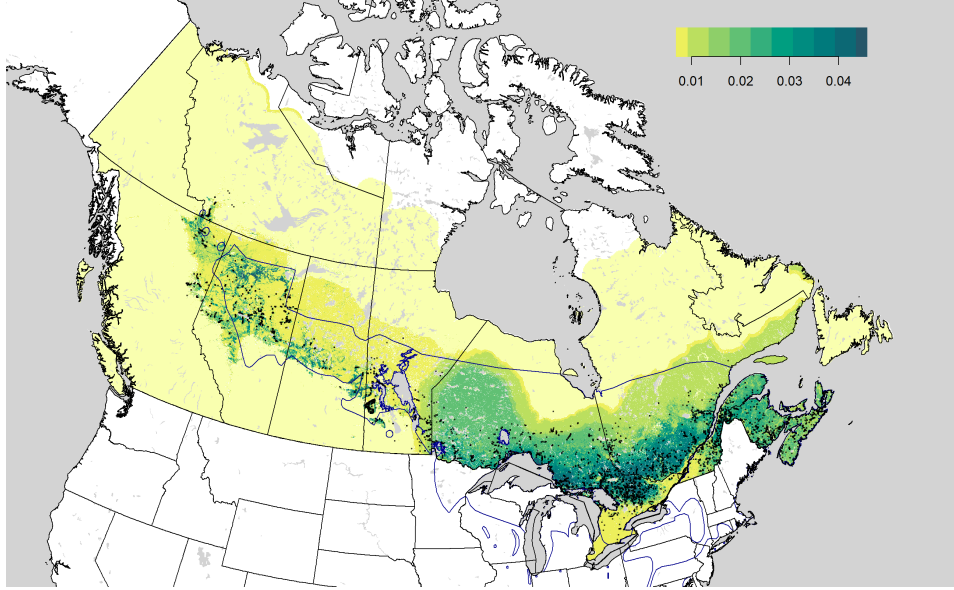
3.1.2 Land cover associations

We used a post-hoc stratification ('post-stratification') approach to estimate land cover based density estimates (males per ha) for each species and regions (Canada and subunits). We classified the predictive maps according to the 2005 MODIS-based North American landcover map into major land cover types (Conifer, Taiga Conifer, Deciduous, Mixedwood, Shrub, Grass, Arctic Shrub, Arctic Grass, Wetland, Cropland) and calculated the mean of the pixel level predicted densities. Uncertainty was based on the 5th and 95th percentiles of the bootstrap distribution.
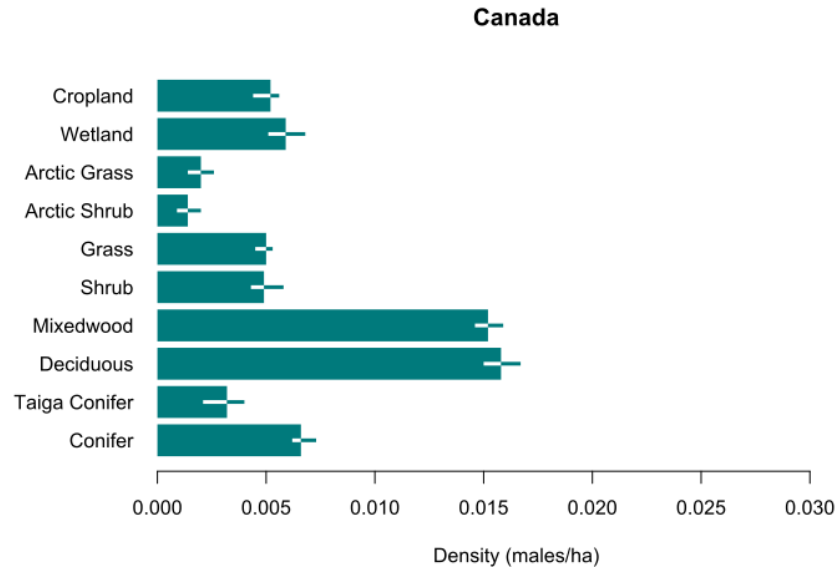
Canada

3.1.3 Population size

Regional population estimates (millions of male individuals) for each species and regions (Canada and subunits) were estimated by summing up the pixel level predictions within the region of interest accounting for the area difference (ha to km2). Uncertainty around the population estimates was based on the 5th and 95th percentiles of the bootstrap distribution.

Region	Abundance (M males)	Density (males/ha)	Area (M km²)
Canada	4.81 (4.59 — 5.21)	0.0077 (0.0074 — 0.0084)	6.21
4 Northwestern Interior Forest	0.309 (0.207 — 0.401)	0.0057 (0.0038 — 0.0073)	0.546
5 Northern Pacific Rainforest	0.0014 (0.0009 — 0.0023)	0.0001 (0.0001 — 0.0002)	0.136
6 Boreal Taiga Plains	1.06 (0.943 — 1.27)	0.009 (0.008 — 0.0107)	1.18
7 Taiga Shield & Hudson Plains	0.206 (0.156 — 0.287)	0.0014 (0.001 — 0.0019)	1.49
8 Boreal Softwood Shield	1.46 (1.34 — 1.7)	0.0111 (0.0102 — 0.0129)	1.32

3.2 Downloadable material

Average density maps for each species are available for download as raster layers in GeoTIFF format.

Result summaries are also available in Microsoft Excel (xlsx) format. Sheets within the file contain abundance and density estimates and also the list of species, variables, variable importance and validation metrics.

3.3 Programmatic access

Population size and density estimates are available through the Boreal Birds JSON API.

Results include static images and data in JSON (JavaScript Object Notation) format that can be consumed by a wide array of modern programming languages.

4 Applications

The BAM National Model results are meant to be used in various applications. We provide some R scripts here to facilitate the use of the results.

4.1 Prerequisites

We will assume that you use R version $>= 3.6$ with the following packages installed: raster, sf, jsonlite, readxl, ggplot2, and optionally googledrive.

4.2 Working with the JSON API

The JSON API uses JSON as data exchange format. Let's define the URL for the BAM v4 API:

```r
library(jsonlite)
api_root <- "https://borealbirds.github.io/api/v4"
```

4.2.1 Get the list of species

First, we need to get the list of species from the JSON API, so that we know what species codes to use:

```r
tab <- fromJSON(file.path(api_root, "species"))
str(tab)
```

```
## 'data.frame':     143 obs. of   8 variables:
##  $ id        : chr  "ALFL" "AMCR" "AMGO" "AMPI" ...
##  $ idnext    : chr  "AMCR" "AMGO" "AMPI" "AMRE" ...
##  $ idprevious: chr  "YTVI" "ALFL" "AMCR" "AMGO" ...
##  $ scientific: chr  "Empidonax alnorum" "Corvus brachyrhynchos" "Spinus tristis" "
##  $ english   : chr  "Alder Flycatcher" "American Crow" "American Goldfinch" "Ameri
##  $ french    : chr  "Moucherolle des aulnes" "Corneille d'Am&eacute;rique" "Chardo
##  $ family    : chr  "Tyrannidae" "Corvidae" "Fringillidae" "Motacillidae" ...
##  $ show      : logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
```

```r
head(tab[,c("id", "english")])
```

```
##      id            english
## 1 ALFL   Alder Flycatcher
## 2 AMCR       American Crow
## 3 AMGO American Goldfinch
## 4 AMPI       American Pipit
## 5 AMRE   American Redstart
## 6 AMRO       American Robin
```

4.2.2   Get estimates for a species

We can use the `id` column if we need to loop over multiple species. Now
we'll only use one species:

```r
spp <- "CAWA"
results <- fromJSON(file.path(api_root, "species", spp))
results$species$english
```

```
## [1] "Canada Warbler"
```

```r
str(results, max.level=2)
```

```
## List of 3
##  $ species :List of 8
##   ..$ id        : chr "CAWA"
##   ..$ idnext    : chr "CCSP"
##   ..$ idprevious: chr "BWWA"
##   ..$ scientific: chr "Cardellina canadensis"
##   ..$ english   : chr "Canada Warbler"
##   ..$ french    : chr "Paruline du Canada"
##   ..$ family    : chr "Parulidae"
```

```

```
..$ show   : logi TRUE
$ popsize :'data.frame':   12 obs. of 4 variables:
..$ region   : chr [1:12] "Canada" "4 Northwestern Interior Forest" "5 Northern
..$ abundance:'data.frame':   12 obs. of 3 variables:
..$ density :'data.frame':   12 obs. of 3 variables:
..$ areakmsq : num [1:12] 6.21 0.546 0.136 1.18 1.49 1.32 0.0552 0.372 0.444 0.3
$ densplot:'data.frame':   12 obs. of 2 variables:
..$ region: chr [1:12] "Canada" "4" "5" "6" ...
..$ data :'data.frame':   12 obs. of 4 variables:
```

The `species` element in the list contain the species info saw already in the species `table`.

The `popsize` element contains population sizes, densities, and areas for the various regions:

```
(N <- do.call(cbind, results$popsize))
```

```
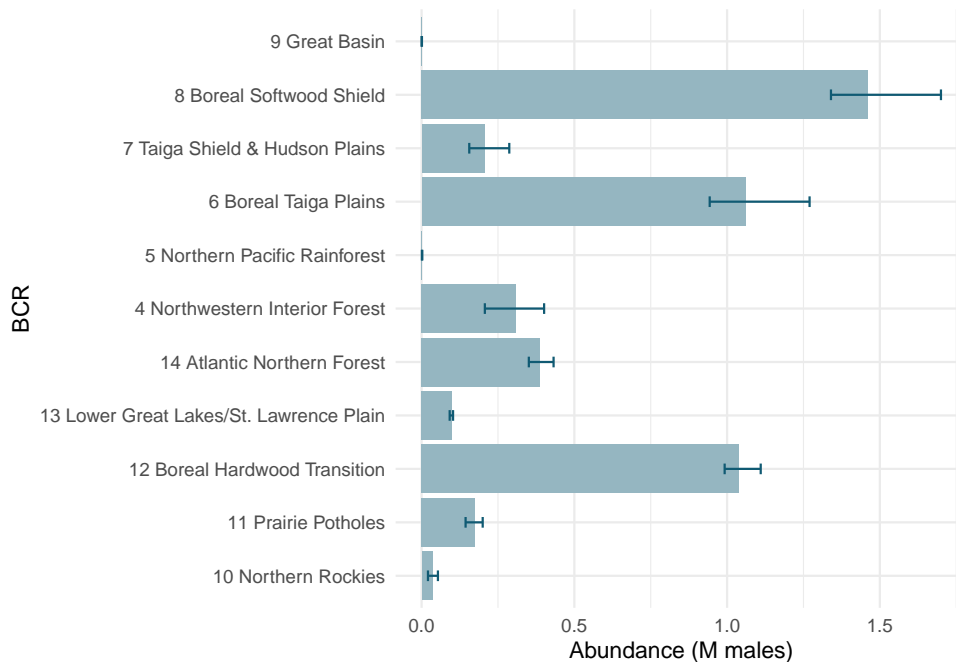##                                   region abundance.estimate abundance.lower
## 1                                 Canada             4.8100          4.5900
## 2         4 Northwestern Interior Forest             0.3090          0.2070
## 3          5 Northern Pacific Rainforest             0.0014          0.0009
## 4                  6 Boreal Taiga Plains             1.0600          0.9430
## 5          7 Taiga Shield & Hudson Plains            0.2060          0.1560
## 6               8 Boreal Softwood Shield             1.4600          1.3400
## 7                          9 Great Basin             0.0002          0.0000
## 8                  10 Northern Rockies               0.0353          0.0208
## 9                  11 Prairie Potholes               0.1750          0.1440
## 10         12 Boreal Hardwood Transition             1.0400          0.9920
## 11 13 Lower Great Lakes/St. Lawrence Plain           0.0986          0.0919
## 12            14 Atlantic Northern Forest             0.3880          0.3510
##    abundance.upper density.estimate density.lower density.upper areakmsq
## 1           5.2100           0.0077        0.0074        0.0084   6.2100
## 2           0.4010           0.0057        0.0038        0.0073   0.5460
## 3           0.0023           0.0001        0.0001        0.0002   0.1360
## 4           1.2700           0.0090        0.0080        0.0107   1.1800
## 5           0.2870           0.0014        0.0010        0.0019   1.4900
## 6           1.7000           0.0111        0.0102        0.0129   1.3200
## 7           0.0011           0.0000        0.0000        0.0002   0.0552
## 8           0.0534           0.0009        0.0006        0.0014   0.3720
## 9           0.2000           0.0040        0.0032        0.0045   0.4440
```

```
## 10          1.1100          0.0280          0.0266          0.0297   0.3730
## 11          0.1030          0.0097          0.0090          0.0101   0.1020
## 12          0.4320          0.0199          0.0180          0.0222   0.1950
```

```
library(ggplot2)
ggplot(N[-1,], aes(x=region, y=abundance.estimate)) +
    geom_bar(stat="identity", fill="#95B6C1") +
    coord_flip() +
    geom_errorbar(aes(ymin=abundance.lower, ymax=abundance.upper),
                  width=0.2, color="#105A73") +
    ylab("Abundance (M males)") +
    xlab("BCR") +
    theme_minimal()
```


The `densplot` element contains the regional landcover specific densities:

```
## pick a region
results$densplot$region
```

```
##  [1] "Canada" "4"       "5"       "6"       "7"       "8"       "9"       "10"
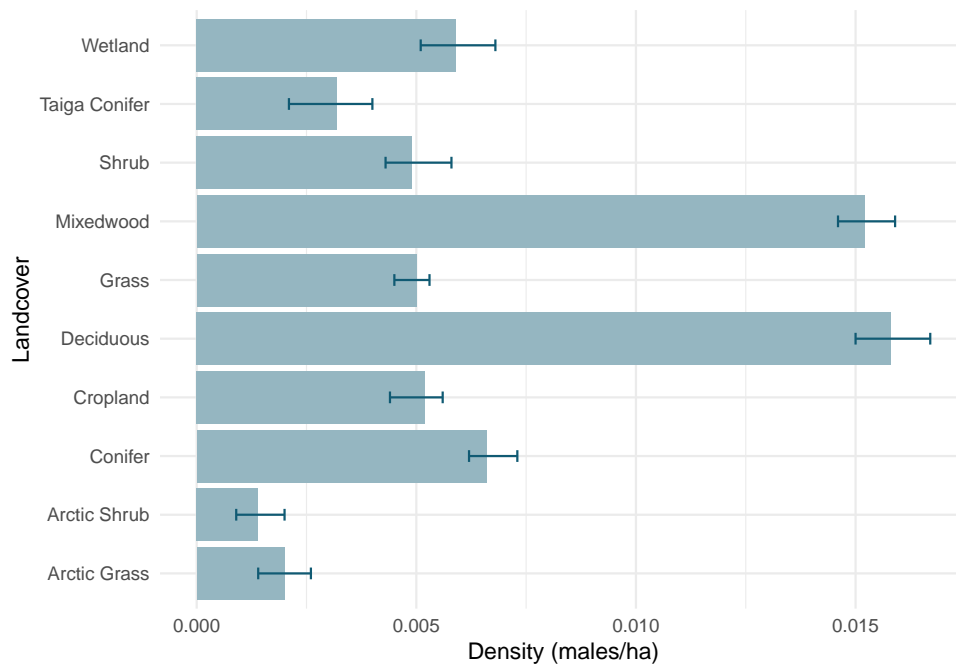##  [9] "11"      "12"      "13"      "14"
```

15

```
region <- 1
results$densplot$region[region]
```

```
## [1] "Canada"
```

```
## densities for this region
(D <- as.data.frame(lapply(results$densplot$data, "[[", 1)))
```

```
##        landcover estimate  lower  upper
## 1        Conifer   0.0066 0.0062 0.0073
## 2  Taiga Conifer   0.0032 0.0021 0.0040
## 3      Deciduous   0.0158 0.0150 0.0167
## 4      Mixedwood   0.0152 0.0146 0.0159
## 5          Shrub   0.0049 0.0043 0.0058
## 6          Grass   0.0050 0.0045 0.0053
## 7   Arctic Shrub   0.0014 0.0009 0.0020
## 8   Arctic Grass   0.0020 0.0014 0.0026
## 9        Wetland   0.0059 0.0051 0.0068
## 10      Cropland   0.0052 0.0044 0.0056
```

```
ggplot(D, aes(x=landcover, y=estimate)) +
    geom_bar(stat="identity", fill="#95B6C1") +
    coord_flip() +
    geom_errorbar(aes(ymin=lower, ymax=upper),
                  width=0.2, color="#105A73") +
    ylab("Density (males/ha)") +
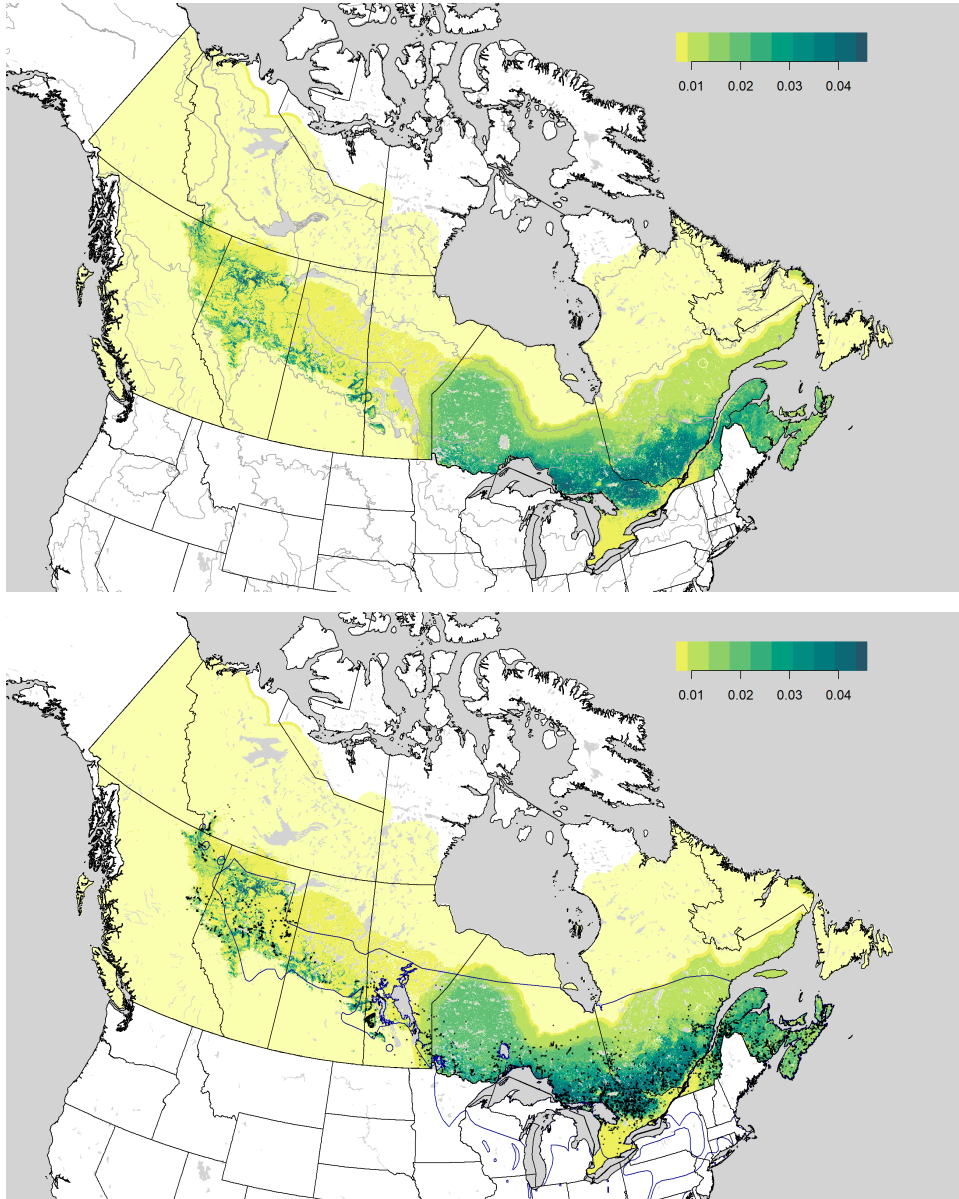    xlab("Landcover") +
    theme_minimal()
```

These results reflect population and density as males only (million males, males/ha, respectively). To apply a pair adjustment, the numbers have to be multiplied by 2.

4.2.3 Density map images

The density map images can be accessed as `https://borealbirds.github.io/api/v4/species/CAWA/:` and `https://borealbirds.github.io/api/v4/species/CAWA/images/mean-det.png`.

4.3 Assessing validation results

Accessing the `BAMv4-results-2020-02-20.xlsx` file gives us the following tables (sheet names in parenthesis):

- metadata explaining sheets and columns (metadata)

- list species of species (species)
- list of variables (variables)
- variable importance (importance)
- model validation results (validation)
- population estimates (abundances)
- lend cover based density estimates (densities)

Download the Excel file in a temporary file:

```
library(readxl)
tmp <- tempfile()
download.file(file.path(api_root, "BAMv4-results-2020-02-20.xlsx"), tmp)
```

Now we can read in different sheets:

```
vars <- read_xlsx(tmp, sheet="variables")
str(vars)
```

```
## tibble [219 x 4] (S3: tbl_df/tbl/data.frame)
##  $ variable  : chr [1:219] "YEAR" "ARU" "AHM" "bFFP" ...
##  $ definition: chr [1:219] "Year of survey" "ARU (1) or human point count (0)" "An
##  $ resolution: chr [1:219] NA NA "1 km" "1 km" ...
##  $ source    : chr [1:219] NA NA "Wang T., Hamann A., Spittlehouse D., & Carroll C
```

Let's read in all the tables and delete the temp file:

```
sheets <- c(
    "metadata",
    "species",
    "variables",
    "importance",
    "validation",
    "abundances",
    "densities")
tabs <- list()
for (sheet in sheets)
    tabs[[sheet]] <- read_xlsx(tmp, sheet)
unlink(tmp)
```

Here are variable importance results:

```
i <- tabs$importance
i <- i[i$id == spp & i$region != "Canada",]
```

```r
i$BCR <- sapply(strsplit(i$region, " "), "[[", 1)
ii <- stats::xtabs(importance ~ variable + BCR, i)
heatmap(ii, margins = c(5, 10), col=hcl.colors(100, "Teal", rev=TRUE))
```
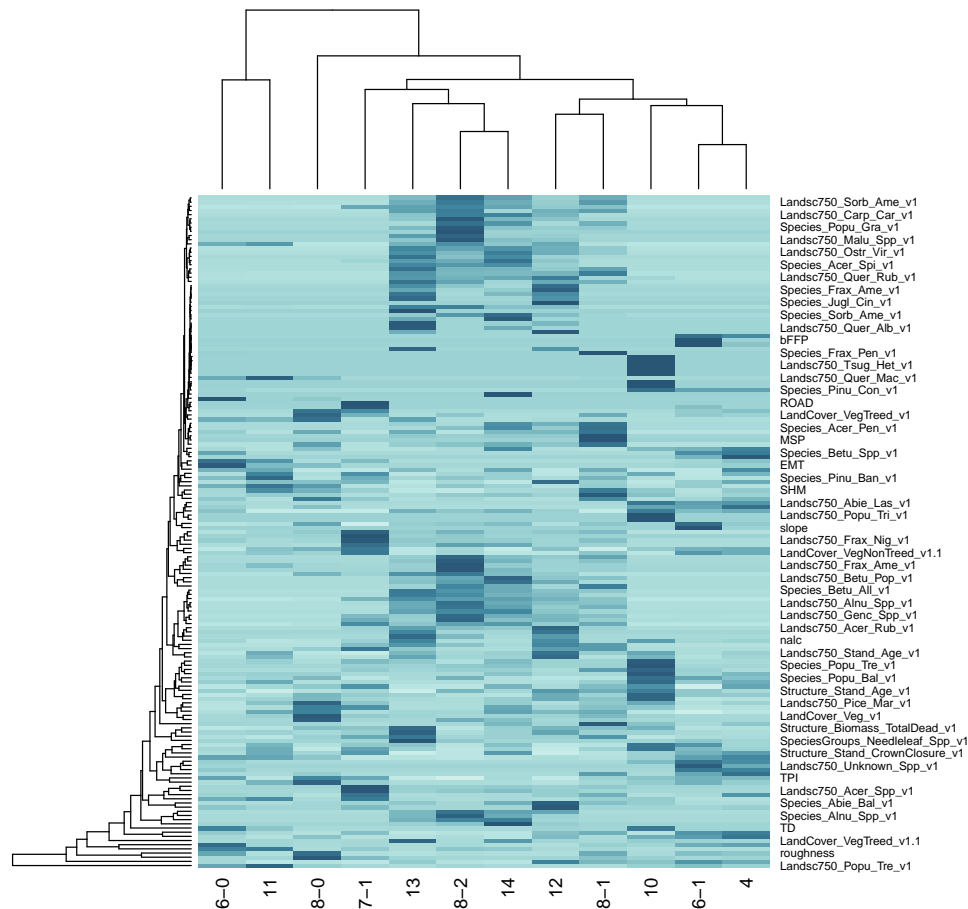

These are validation results for the Canada Warbler:

```r
v <- tabs$validation
v <- v[v$id == spp,]
v[order(v$prevalence, decreasing=TRUE), c(4,5,8:12)]
```

```
## # A tibble: 14 x 7
##    region              prevalence AUC_final pseudo_R2   occc  oprec  oaccu
##    <chr>                    <dbl>     <dbl>     <dbl>  <dbl>  <dbl>  <dbl>
##  1 6-1 Boreal Taiga Plains,~  0.0215     0.934     0.265  0.596  0.664  0.898
##  2 12 Boreal Hardwood Trans~  0.0214     0.807     0.0923 0.924  0.934  0.990
```

```
##  3 8-1 Boreal Softwood Shie~    0.0204    0.770    0.0530 0.632  0.694  0.910
##  4 14 Atlantic Northern For~     0.0163    0.741    0.0640 0.826  0.862  0.958
##  5 13 Lower Great Lakes/St.~     0.0137    0.839    0.138  0.915  0.925  0.989
##  6 8-2 Boreal Softwood Shie~     0.0136    0.823    0.0879 0.598  0.666  0.897
##  7 4 Northwestern Interior ~     0.0132    0.936    0.253  0.638  0.718  0.890
##  8 6-0 Boreal Taiga Plains,~     0.0122    0.857    0.165  0.861  0.890  0.967
##  9 Canada                        0.0120    0.852    0.136  0.579  0.709  0.779
## 10 8-0 Boreal Softwood Shie~     0.0104    0.835    0.241  0.152  0.340  0.448
## 11 7-1 Taiga Shield & Hudso~     0.00505   0.997   -0.0723 0.0289 0.0877 0.330
## 12 11 Prairie Potholes           0.00418   0.872    0.250  0.767  0.801  0.958
## 13 10 Northern Rockies           0.00180   0.982    0.276  0.515  0.635  0.810
## 14 7-0 Taiga Shield & Hudso~     0.000426  0.0703   0      0.0754 1      0.0754
```

AUC was used to assess classification accuracy and pseudo R2 to quantify the proportion of variance explained. OCCC measures correspondence across bootstrap based predictions. OCCC is the product of two the overall precision (how far each observation deviated from the best fit line), and the overall accuracy (how far the best line deviates from the 1:1 line).

4.4 Working with maps

The 1 km2 resolution GeoTIFF raster files are in this shared Google Drive folder, anyone can view.

We can access the list of available files and download the files using the googledrive package:

```
library(googledrive)

## this should let you authenticate
drive_find(n_max = 30)

## now list files in the shared folder
f <- "https://drive.google.com/drive/folders/1exWa6vfhGo1DNUL4ei2baDz77as7jYzY?usp=sh
l <- drive_ls(as_id(f), recursive=TRUE)

## add species codes
l$species_id <- sapply(strsplit(l$name, "-"), "[[", 2)

## download
tmp <- tempfile(fileext = ".tif")
```

```
file_id <- l$id[l$species_id == spp]
tif_file <- drive_download(file_id, tmp)

tif_file <- drive_download(file_id,
    path="~/Downloads/pred-CAWA-CAN-Mean.tif",
    overwrite=TRUE)

tif_path <- tif_file$local_path
```

Alternatively, download the file for a species and point to its path:

```
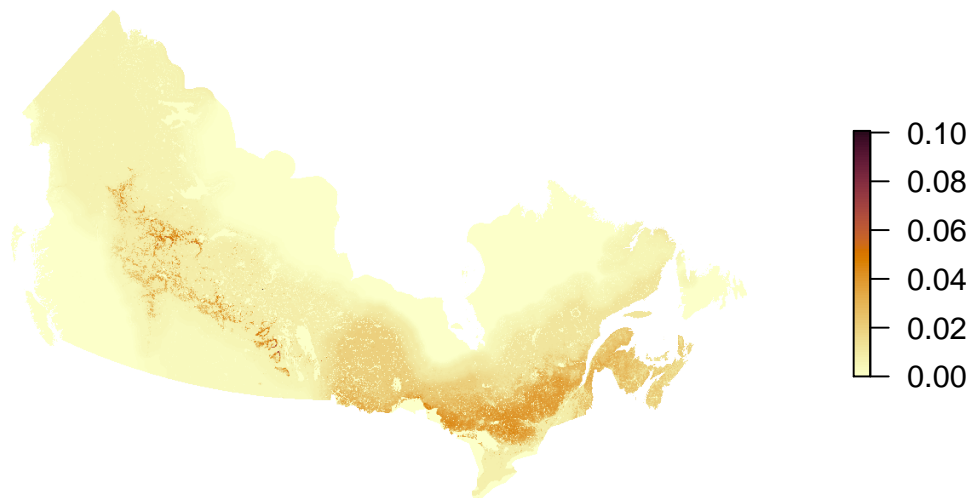tif_path <- "~/Downloads/pred-CAWA-CAN-Mean.tif"
```

Now we can work with the raster:

```
library(raster)
r <- raster(tif_path)

plot(r, axes=FALSE, box=FALSE, col=hcl.colors(100, "Lajolla"))
```


4.4.1   Population size for custom boundary

Next, we read in a custom boundary file. Let's use the provincial boundary of Alberta now (stored as a GeoJSON file). We transform the polygon to match the projection of our raster layer and plot the two together:

```r
library(sf)

#bound <- st_read("https://raw.githubusercontent.com/ABbiodiversity/cure4insect/maste
bound <- st_read("https://raw.githubusercontent.com/ABbiodiversity/cure4insect/master
```

```
## Reading layer `Alberta' from data source `https://raw.githubusercontent.com/ABbiod
## Simple feature collection with 1 feature and 1 field
## geometry type:  POLYGON
## dimension:      XY
## bbox:           xmin: -120.0016 ymin: 48.99666 xmax: -110.0048 ymax: 60.00046
## epsg (SRID):    4326
## proj4string:    +proj=longlat +datum=WGS84 +no_defs
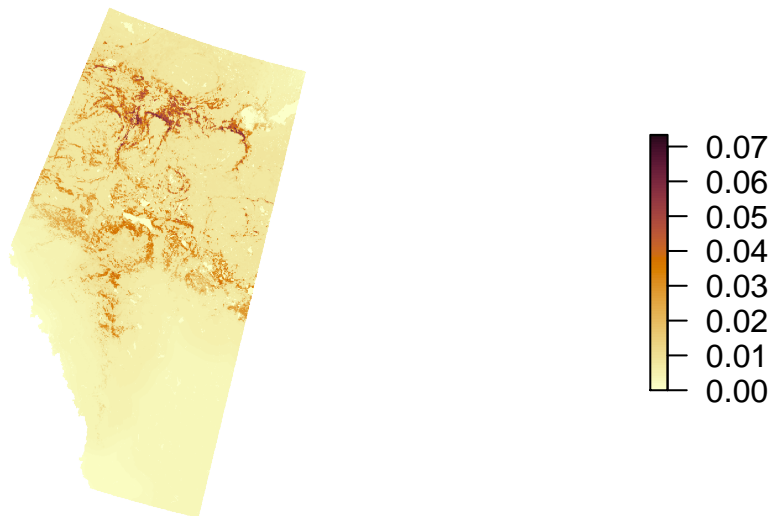```

```r
bound <- st_transform(bound, st_crs(r))

plot(r, axes=FALSE, box=FALSE, col=hcl.colors(100, "Lajolla"))
plot(bound$geometry, add=TRUE, border="red", col=NA)
```
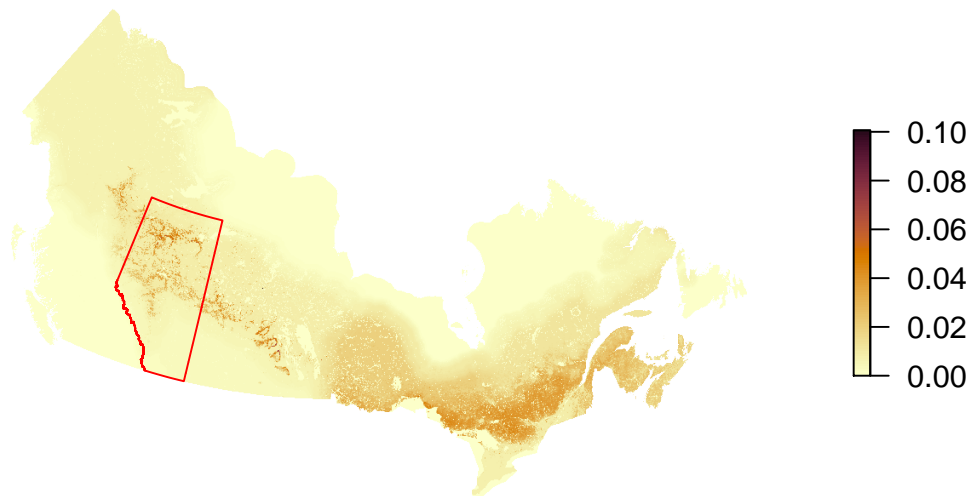

Let's crop the density map to the extent of Alberta and mask areas outside of the boundary:

```r
r2 <- crop(r, bound)
r2 <- mask(r2, bound)
plot(r2, axes=FALSE, box=FALSE, col=hcl.colors(100, "Lajolla"))
```

23

We can now sum up the raster cells to get population size (million individuals) with pair adjustment:

```
(N <- sum(values(r2), na.rm=TRUE) * 100 * 2) / 10^6
```

```
## [1] 1.082979
```

4.4.2 Post-stratified density estimates

Post-hoc stratification ('post-stratification') is an approach to estimate land cover based density estimates (males per ha) for a species based on the density map and a classification layer.

Let's use the 2005 MODIS-based North American landcover map as an example within the Alberta boundary. We calculate the mean of the pixel level predicted densities (PS).

```
## read in raster
lc <- raster("https://raw.githubusercontent.com/ABbiodiversity/recurring/master/offse

## crop and mask to boundary
lc <- mask(crop(lc, bound), bound)

## extract cell values
LC <- data.frame(
    lc=values(lc),      # land cover classes, integer
    density=values(r2), # males / ha
```

```
    area=1)              # km^2
## remove NA values (cells outside of boundary but inside bounding box)
LC <- LC[!is.na(LC$density),]

## land cover classes
labs <- c(
    "Conifer"=1,
    "Taiga Conifer"=2,
    "Deciduous"=5,
    "Mixedwood"=6,
    "Shrub"=8,
    "Grass"=10,
    "Arctic Shrub"=11,
    "Arctic Grass"=12,
    "Wetland"=14,
    "Cropland"=15)
LC$label <- names(labs)[match(LC$lc, labs)]

head(LC)
```

```
##       lc    density area    label
## 1045   6 0.006929385    1 Mixedwood
## 1046   6 0.006855813    1 Mixedwood
## 1047   6 0.006849375    1 Mixedwood
## 1825   6 0.007014129    1 Mixedwood
## 1826   6 0.006815174    1 Mixedwood
## 1827   1 0.006703767    1   Conifer
```

```
## aggregate density by land cover
(PS <- aggregate(list(density=LC$density), list(landcover=LC$label), mean))
```

```
##        landcover     density
## 1  Arctic Grass 0.003469213
## 2       Conifer 0.006983323
## 3      Cropland 0.004930945
## 4     Deciduous 0.019037611
## 5         Grass 0.004553917
## 6     Mixedwood 0.013687571
## 7         Shrub 0.006586476
## 8 Taiga Conifer 0.007123922
```

```
## 9       Wetland 0.008748280
```

Here are the post stratified density values:

```
ggplot(PS, aes(x=landcover, y=density)) +
    geom_bar(stat="identity", fill="#95B6C1") +
    coord_flip() +
    ylab("Density (males/ha)") +
    xlab("Landcover") +
    labs(title=paste(results$species$english, "in Alberta"),
         caption="Based on post-stratification") +
    theme_minimal()
```
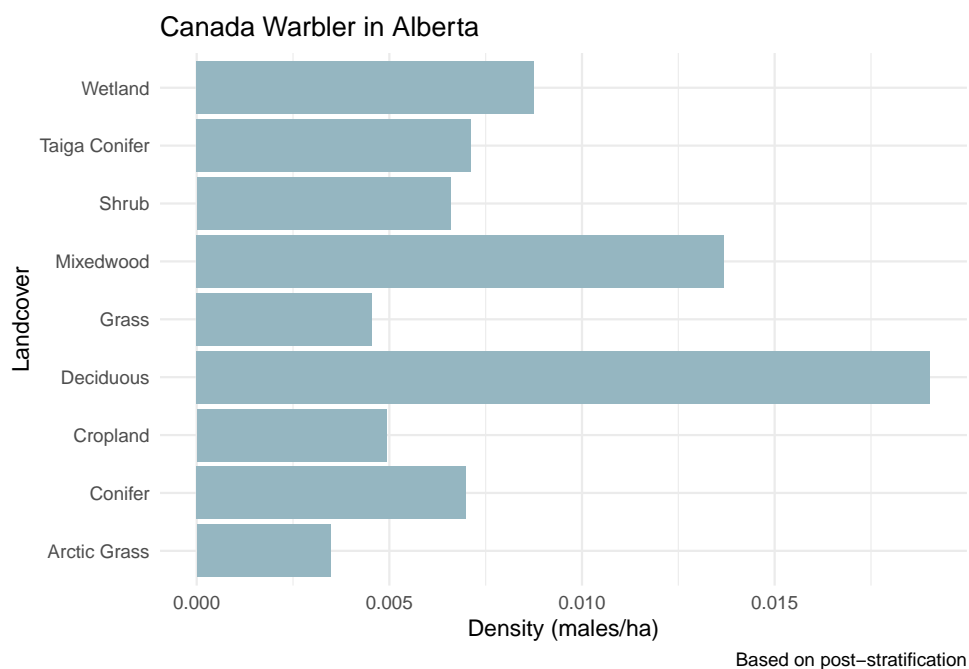

These results can be used in region specific analyses that require density values as inputs, for example landcover based scenario analyses.

5   Source Code

5.1   Generating results

The code for the BAM Generalized National Models (GNM) is hosted in the borealbirds/GNM GitHub (git) repository.

The code in the `/R` folder of the repository includes R scripts for processing the data (observations, offsets, and predictors) and mosaicing together the regional predictive map pieces.

The code in the `/graham` folder contains code to run on Compute Canada's Graham cluster. The code includes regional boosted regression models and predictions.

The `/www` folder of the repository includes scripts to summarize the outputs and organize the results into a presentable format hosted as part of the API repository.

5.2   Storing the results

The results from the BAM Generalized National Models (GNM) are hosted in the borealbirds/api GitHub (git) repository.

Assets are served from the `/docs` folder of the git master branch via GitHub pages.

Benefits of hosting these results via GitHub pages include security (serving assets over https, protection against certain cyber attacks) and speed (via the use of content delivery network). The data can be consumed by other computers because GitHub pages responses have cross-origin resource sharing headers.

The species mean density raster layers are available in GeoTIFF format from this Google Drive folder.

5.3   Presenting the results

The results from the BAM Generalized National Models (GNM) are presented via GitHub pages based on the borealbirds/borealbirds.github.io GitHub (git) repository.

The website's features include:

- Uses Vue and Gridsome
- Tailwind CSS v1 (with PurgeCSS)
- Based on this Gridsome template with light/dark theme.
- Search among species with Fuse.js and vue-fuse
- 404 Page
- RSS Feed
- Sitemap in XML

- Google search crawling is allowed
- Comments via Disqus
- Bird images from Unsplash for the 404 page

To install the tools to build the website, follow these steps:

```
# Install Gridsome CLI tool
npm install --global @gridsome/cli

# Clone the repo
git clone -b dev https://github.com/borealbirds/borealbirds.github.io
cd borealbirds.github.io

# Install dependencies
npm install

# Run development server with hot reloading
gridsome develop

## now look at http://localhost:8080
```

To locally build and deploy, use the `_build.sh` script (you will need write access to the GitHub repository).

Automatic build & deployment enabled via GitHub Actions, see setup in the `.github/workflows/build.yml` file.