Predictive Models for Estimating the Cumulative Effects of Human Development on Migratory Landbirds in the Oil Sands Areas of Alberta

Project Name: Joint Oil Sands Monitoring: Cause-Effects Assessment of Oil Sands Activity on Migratory Landbirds

Péter Sólymos

Alberta Biodiversity Monitoring Institute and the Boreal Avian Modelling Project, Department of Biological Sciences., CW 405, Biological Sciences Bldg., University of Alberta, Edmonton AB, T6G 2E9

C. Lisa Mahon

Environment Canada, Canadian Wildlife Service, Population Assessment Unit, Prairie and Northern Region, 9250-49th Street NW, Edmonton AB, T6B 1K5

Trish Fontaine

Boreal Avian Modelling Project, University of Alberta, Department of Renewable Resources, 751 General Services Building, Edmonton AB T6G 2H1

Erin M. Bayne

University of Alberta, Department of Biological Sciences, Edmonton AB, T6G 2E9

March 31, 2015

Contents

Executive Summary

Project results describe the habitats used by 81 species of forest songbirds and response by birds to human footprint (physical disturbance of vegetation cover by human activity). We used a modelling approach using data from surveys of boreal birds collected by Environment Canada, the Alberta Biodiversity Monitoring Institute, the North American Breeding Bird Survey, and additional data from studies also compiled by the Boreal Avian Modelling Project. The models described the habitat associations and responses to human footprint at two scales: at the station of the data collection (0.07 km2) and at the quarter-section (0.64 km2) around the stations.

Using an estimate of what this landscape would have looked like with no human activity, we predicted how the abundance and distribution of birds would differ from their current abundance and distribution. We also grouped different types of human disturbances according to 'sectors': agriculture, urban/industrial, mines/wells, energy related linear features (transmission lines, pipelines, seismic lines), forestry, and transportation related linear features (roads, rails).

The results across all 81 species showed that the overall impact of the different sectors on bird populations is proportional to the area of disturbed landscape associated with the various sectors. Agriculture and forestry had largest extent in the oil sands region, consequently these sectors contributed the most to the predicted population differences.

After standardizing the population differences by the area of disturbances, the unit area effects of vegetated and non-vegetated linear features were highest, and the impact was often disproportionate relative to the relatively small area of these disturbance types. Such disproportionate effects occur when a particular disturbance is affecting highly suitable habitats of a species, thus a 1% land cover change could result in >1% population difference.

The results can be used to improve management practices for migratory songbirds including several listed species, such as Canada Warbler and Olive-sided Flycatcher (both Threatened, Species at Risk Act), and Black-throated Green Warbler (Special Concern, Alberta Wildlife Act).

Introduction

The Boreal Avian Modelling Project (BAM), Environment Canada (EC), and the Alberta Biodiversity Monitoring Institute (ABMI) collaborated under the Joint Oil Sands Monitoring (JOSM) on using the most extensive standardized point count data collected and compiled by these agencies (BAM, EC, ABMI) and the North American Breeding Bird Survey (BBS) program.

The avian data set was combined with available geospatial information to build predictive models for bird species. In this year's report, we have updated the models summarizes in last year's report (Solymos et al. 2014). Major updates in this 2015 sets of models were:

- additional data from 2014 sampling year (*n*=4238 surveys, targeted mostly on footprint types and wetland areas),
- updated/corrected spatial coordinates for some projects and general updates in the BAM and BBS databases,
- revised modeling strategy in terms of the variables used and the sequence how they enter the model selection procedure (road x habitat interactions, post-harvest, forest regeneration trajectories, climate covariate interactions, description of forest age responses),
- external validation of fitted models,
- aging/regeneration of forestry footprint incorporated into predictions,
- refined geospatial information and the ability to crosswalk current land cover and 'backfilled' land cover types.

Objectives

The updated data and the models were used to

1. describe the habitat associations of the bird species and their responses to human footprint at different spatial scales;
2. determine difference between current and historically assumed ('backfilled') habitat supply for bird populations in the oil sands area;
3. attribute the differences between current and 'backfilled' population estimates to human footprint types categorized into different sectors;
4. evaluate how direct (i.e. habitat loss) and indirect (i.e. proximity effects) effects of the different sectors contribute to the overall difference between current and 'backfilled' population estimates.

Methods

Data sets

Bird surveys

The point count data set included a total of 48242 surveys taken at 25407 locations in the Boreal, Canadian Shield and Foothills natural regions of Alberta between 1997 and 2014. When multiple point count surveys were replicated at the same location more than once within a year, we retained a single randomly chosen from the revisits. Replicates collected in different years were kept in the data set.

The data were contributed by ABMI (6863 surveys from 5990 location), BAM (17429 surveys from 10503 locations), BBS (16962 surveys from 1925 locations), and EC (6989 surveys from 6989 locations). We used counts of 81 species (Appendix 1) that had at least 25 detections and had available estimates for singing rates and effective detection radii (according to Sólymos et al. 2013).

Land cover information

We used a composite wall-to-wall land cover map of Alberta developed by ABMI to characterize vegetation at sampling location and within spatial units (quarter-sections) used for prediction. The wall-to-wall vegetation map was used with footprint classes removed to create the “backfilled” vegetation layer (ABMI 2014). The wall-to-wall human footprint map (ABMI 2013) was merged with the backfilled map in the end. The “backfilled” layer was also combined with other sources of information to better describe habitat conditions, i.e. forest age and wetness, from Alberta AVI, wetness information from various sources (ABMI 2014).

Land cover types in the “backfilled” (reference) and current vegetation + footprint maps are listed in Appendix 3. We refer to these products as *reference* and *current* maps, respectively. The reference vegetation map describes the vegetation that would have been present in the study area if there was no human footprint (updated to the year 2010). This backfilled reference vegetation map incorporates information about fires, describes the ages of natural vegetation for 2010 conditions, and projects ages of the backfilled polygons (areas where human footprint currently exists) for 2010 conditions. The current vegetation map describes the vegetation and human footprint that currently exists within the OSA (updated to the year 2010).

Bioclimatic variables

We have used 7 bioclimatic variables in the modeling:

- MAT: mean annual temperature
- MAP: mean annual precipitation
- PET: potential evapotranspiration
- AHM: annual heat-moisture index
- FFP: frost-free period
- MWMT: mean warm month (Jul) temperature
- MCMT: mean cold month (Jan) temperature

The bioclimatic variables used in the modeling were calculated at a 4-km resolution using monthly climate normals of temperature and precipitation averaged over 1961-1990. The monthly climate normals are based on instrument-measured climate data that were interpolated by PRISM (Daly et al., 2002) and WorldClim (Hijmans et al., 2005). The western North American portion of these data are described by (Wang et al., 2011).

Modeling

Count models

We used Poisson generalized linear models with a log link. The response variable was the total number of individuals counted per survey. We used the QPAD approach (Sólymos et al. 2013)to calculate offsets that account for differences in sampling protocol and nuisance parameters affecting detectability (time of day, time of year, tree cover and composition).

Bootstrap aggregation

We applied ‘branching’ forward stepwise variable selection to minimize bias in predictions. The ‘branching’ process was applied instead of a simple add-one type of variable search in order to minimize model misspecification due to co-linearity is some of the covariates (i.e. spatial variables and footprint). The best supported model was determine based on lowest value of the Akaike’s information criterion (AIC) among competing models.

The `branching' process was combined with bootstrap aggregation (bagging, or bootstrap smoothing) (Breiman 1996) to minimize variance in the predictions. Bootstrap replicates were drawn with replacement for each spatio-temporal blocking units. Spatial blocks were established by dividing the sampling area into 4 quadrants within which the distributions of surveys were even. Four roughly 5-year time periods (1997-2002, 2003-2008, 2009-2014) were used as temporal blocks. This ensured that bootstrap replicates represented the whole sample distribution. The number of bootstrap iterations was B=240. Each draw resulted in a vector of IDs, same IDs were used across species to allow for comparisons across them. With the response data vector this resulted in a total of B independent runs.

Within spatio-temporal units, we sampled survey stations with replacement using the number of survey visits within each selected station as weights with replacement. This ensured that the spatial sampling pattern of the point counts was retained in the bootstrap samples but abundant projects did not dominate the samples. We weighted each survey visit by the inverse square root of the number of observations in 4 km x 4 km grid cells to account for the non-independence (clustered nature) of the data set. Otherwise observations were assumed to be independent conditional on the value of the predictors.

Model stages

At each stage of the branching hierarchy we compared support among several models using AIC. Variables for the top ranked model in a given stage were fixed and added to models in the subsequent level. Model sets at each level also considered a null model, which was the top model from the previous level or, in the case of level 1, a constant density model without covariates (Stage 0). We considered variables describing local scale habitat conditions at the first four stages (150 m buffer, Stages 1-4). At Stage 5 we used spatial smoothing terms (latitude, longitude, climate). Variables describing conditions around the sites at the quarter-section scale were considered at Stages 6. Model terms and subsets corresponding to stages of the branching model hierarchy are in Appendix 2.

Each higher stage modifies coefficients carried forward from previous stages (corresponding to terms that have already entered the model and are already part of the "active set"). For example quarter-section scale effects modify local effects depending on habitat composition and disturbance context. We repeated the multi-stage branching variable selection procedure for each bootstrap iteration, which resulted in *B* sets of parameter estimates (see e.g. Efron 2013). We stored the IDs of best supported models at each level of each iteration to be able to recover the variable selection process.

Lorenz tangent approach

We used the *Lorenz-tangent approach* to determine the threshold for delineating high suitability land cover classes. First, we fitted a model with categorical land cover types as predictors. This model was used to predict expected density (fitted) values for each observation used after accounting for the effect of road. Fitted values were then sorted from the smallest to largest. The cumulative sum of sorted fitted values represented a strictly monotonically increasing convex function (i.e. Lorenz-curve). We determined the value of the sorted (not cumulated) density value (Lorenz threshold) that corresponded to the point in the Lorenz-curve where the slope is 1 (tangent). This point corresponds to the Youden statistic (Youden 1950) which is the maximum difference between the Lorenz-curve and the diagonal. The maximum difference is located at the tangent.

A land cover class was defined as highly suitable when expected density was equal or higher than the Lorenz threshold. The amount of high suitability land cover classes were summarized at the quarter-section scale at each point during the modeling, repeating the thresholding procedure for each bootstrap iteration.

Predictor variables

Predictor variables were chosen based on our current understanding of songbird ecology across Alberta. We used two spatial scales to derive predictor variables for each survey station. Local scale variables were assessed in a 150 m radius of each station. quarter-section scale variables were assessed in a 451 m radius of each survey station, which corresponds to quarter-section area (64 ha) and is the unit for predictive mapping across Alberta.

The predictor variables were grouped to facilitate variables selection (Appendix 2). NULL refers to the selected model at the previous stage. NULL is an intercept only model for stage 1.

- **Stage 1: local habitat** - description of local habitat conditions within a 150 m radius circular buffer, based on current land cover composition. Feature types were collapsed into main vegetation and footprint categories (Appendix 2). Proportional areas within each 150 m radius buffer were calculated for each habitat class. Dominant habitat type was assigned to each survey station based on a simple majority rule.
- **Stage 2: age** - age effects and interactions between forest types and age. Age was calculated as the difference between year of the bird survey and the year of origin. Year

of origin was estimated based on area-weighted age of forest polygons within the 150 m radius circular buffer. Age was modelled as a linear or quadratic relationship using age and square root of age.

- **Stage 3: forestry** – we have used a nonlinear trajectories (Figure 1) to differentiate between stands of natural (fire) vs. cutblock origin. We assumed that the growth trajectory converges to natural disturbance trajectory at year 80 post disturbance. When the dominant habitat type was forestry footprint, the backfilled habitat class was used. This was done so that forestry activities did not form a separate habitat class but instead was treated as young forest with a modified (fCC2 in Appendix 2).
- **Stage 4: contrast** –this stage accounts for 'contamination' so that we can better predict for non-contaminated land cover types by statistically removing effects of linear features. Linear features also had small combined area within the 150 m buffers and as a result did not form a dominant habitat type. Dummy variable indicating that a road intersected the 150 m radius point count buffer. It was either known (BBS) or inferred based on hard linear features being >4% of the 150 m radius buffer (this corresponds to a 9 m wide road). Proportion (0--1) of soft linear features with in the 150 m radius buffer. We also included interactions between these variables and broad land cover categories because edge effects are expected to change with surrounding habitat composition.
- **Stage 5: climate** – spatial smoothing terms using latitude/longitude and climate variables (Appendix 2).
- **Stage 6: surrounding habitat** – proportions of footprint and suitable habitat within 451 m radius buffers around points, land cover types that are suitable were determined by the Lorenz-tangent approach. Footprint types were categorized into the following groups: cultivated, non-cultivated, linear, non-linear, successional, alienating, and total human footprint.

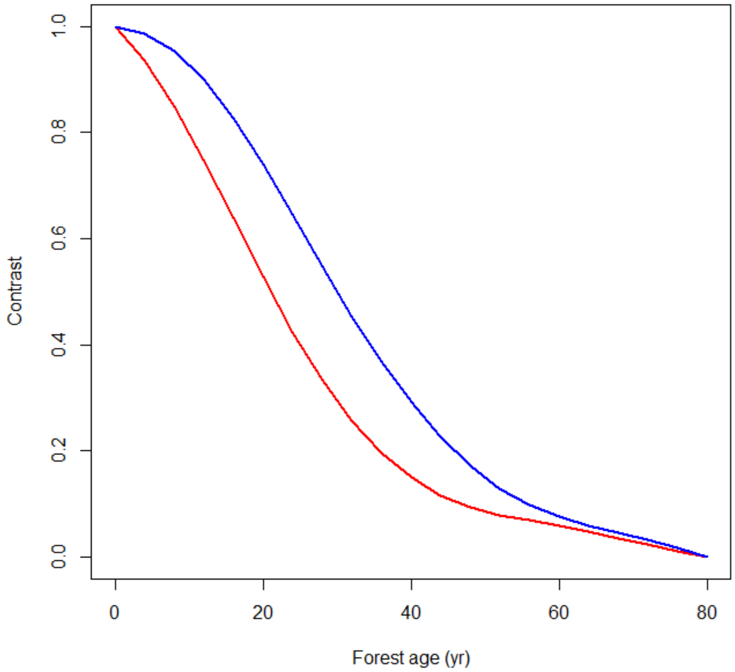

Figure 1. Convergence trajectories for forestry cutblocks based on expert opinion used in modeling. Red = deciduous stands, blue = coniferous stands. Contrast is the complement of recovery.

Model evaluation

10% of the full data set (*n*=4829 surveys) was used as a held-out set for independent/external validation purposes, and this part of the data was not used for model building. When evaluating the model performance, we checked if the models were able to:

1. Rank survey locations according to their relative habitat suitability (this is important for extrapolating the results into un-sampled areas),
2. Estimate the population level proportion of counts accurately (this is important for population size estimation),
3. Discriminates between 0 and >0 observations (this is important for making accurate local predictions).

Prediction

Local scale habitat effects

Local scale habitat associations and forest-age relationships were predicted based on model parameter estimates at stage 4. We used the bootstrap distribution and used the median (50% percentile) as the point estimate and the 5% and 95% percentiles as lower and upper confidence limits, respectively, for a confidence interval with 90% nominal coverage. Values are presented in males/ha units.

Mapping reference and current abundance

We predicted pixel (quarter section, ~800 m x 800 m) level abundance values based on the reference and current vegetation maps. We summarized provincial land cover information (reference and current conditions) for predictive mapping at the quarter-section scale for the entire study area and calculated density for each species for quarter-section units (pixels) within the study area based on model parameter estimates from the first model run (i.e. without the rest of the bootstrap iterations) at stages 4, 5, and 6, so that we can compare the effects of increasing model complexity on the spatial predictions. The attributes (footprint and vegetation) for each quarter-section pixel were calculated based on all polygon attributes found within that pixel. The centroid of the pixel was used to determine latitude/longitude and climate (from 500 m resolution raster layers) variable values. The mean predicted density for a pixel was the weighted average of the polygon-level densities weighted by corresponding polygon areas. We used the median instead of the mean to avoid the distorting effects of outliers. Outliers caused problems for rare predictor combinations (<1%) when predicting outside of the sample space.

Estimating population size and sector specific effects

We aim to quantify the *net effect* of human footprint on the *potential population* of a species in a region of interest. This is can be done by comparing the *current* and *reference* populations (Ncr, Nrf, respectively) in the region, as estimated based on the spatial predictions described in the previous section. If Ncr < Nrf we are talking about a *decreaser* species, if Ncr > Nrf we are talking about an *increaser* species. These qualifiers reflect the net effect. The limitation of this approach is this: Imagine for example, that an early seral forest species show a mixed response to footprint. One type of footprint (e.g. cultivation) has a negative, while another type of footprint (e.g. forestry) has positive effect on potential population. This limitation can be addressed if the effects of footprint types are individually quantified.

For this, we need to be able to attribute land cover change to different footprint types. We want reference and current population estimates for the combined land cover categories of the reference and current maps. For example of a well pad is surrounded by 60-year-old deciduous forest then the backfilled label is likely to be “60yr old deciduous”. Combining the 2 labels (and all possible combinations of vegetation and footprint classes) will be “60yr old deciduous -> well pad”. Each of these labels in a given region is then associated with the corresponding area (A), and the reference (Nrf) and current population size (Ncr) estimates. Labels be collapsed later: for example collapsing all deciduous forest irrespective of age would give the label “Deciduous” under reference, and “Urban/Industrial” as current. The corresponding values of A, Nrf, Ncr become the sum of the categories lumped in higher level classes.

The *unit effect* of a footprint category or sector can be calculated as the percent population size difference, dN=100*(Ncr – Nrf)/sum(Nrf) for the footprint type(s) of interest standardized by percent area (dA) of the footprint type(s). The unit effect index takes a value of 0 when the population effect is 0; the value is 1 when % population gain is equals the habitat conversion %; it is -1 when % population loss equals the habitat conversion %. The index can take any value between minus and plus infinity, thus it is unbounded. We transformed the index into the (-1, 1) range as sign(dN/dA)*plogis(log(abs(dN/dA))), and it is also 0 when dA=0]. As a result of rescaling, the -1 and 1 values will be at -0.5 and 0.5. When the absolute value of the scaled intensity is close to 0.5, it means that a given land conversion has an effect on population that is proportional to the % area converted.

The following sectors were identified for the analyses in this report (see Appendix 3):

- Agriculture,
- Urban/Rural,
- Mines/Wells,
- Forestry,
- Energy related soft linear features (pipelines, seismic lines, transmission lines),
- Transportation (roads, rails, incl. associated vegetated feature types).

The calculations for spatial predictions were coupled together with the ‘sector effects’ calculation. For sector effects, the polygon level estimates were summarized for land cover types within the oil sands study area instead of spatial units (quarter sections). The calculation details for sector effects are given in Appendix 4.

Direct and indirect effects of footprint

The sector effects calculations as outlined above take into account direct effects of land cover conversion from natural vegetation into some form of footprint. By modeling the effects of surrounding habitats and footprint in local abundance of the species, using a stage 6 model a certain footprint type might affect not converted natural vegetation types in its vicinity (quarter section scale). Thus it is possible to quantify the indirect effects of a footprint type by partial backfilling. Partial backfilling was done so that all footprint types except for the footprint type of interest were backfilled, and population size estimates were calculated as before. This was repeated for each group of footprint according to sectors. This allowed us to partition the population difference into sector specific contributions: direct effect is defined as the difference in polygons that are labeled as footprint the current land cover map, while indirect effect is defined as the difference in natural habitats under current conditions. The unit effect of direct and indirect effects were calculated by standardizing the % population difference by the % area of the footprint. This rationale applies only under partial backfilling, because this way it is ensured that the only amount of footprint that is not filled back make the impact on the populations. Thus, the same amount of footprint affects expected populations both directly and indirectly. These results only apply to the region of interest, here the oil sands area.

Results and Discussion

Results for individual species can be found in Appendices 5 and 6.

Example: Canada Warbler

The Canada Warbler (CAWA) model could rank the 0 and >0 counts correctly for both the internal and external validation data set (Figure 2). The fitted vs. observed cumulative distribution of counts improved for >0 counts when using covariates, the proportion of 0s were well approximated by the detectability corrections that were incorporated into the constant model due to differences in sampling effort among the data sets. Increasing model complexity did not result in huge improvement in classification accuracy for the 0 and >0 counts based on AUC values, it was rather incremental. Covariate based AUC values were generally high for the external validation data (>0.79).

CAWA abundance was highest in upland forest stands, especially in old-growth deciduous and mixedwood stands. Abundance was low in pine forests, treed and open wetlands and anthropogenic habitats. Forestry effect was not distinguishable from naturally regenerating young forests (Figure 3).

Using the Lorenz tangent approach, we found that mature and old growth deciduous forest stands were selected more than 95% of the bootstrap iterations as part of the highly suitable habitat classes. Other upland forest stands were also selected with higher frequencies (~50% and above) than other land cover types (Figure 4).

The amount of highly suitable in the surrounding landscape had a positive effect on CAWA abundance, as expected. The amount of surrounding footprint types had positive effect on abundance when replacing low suitability habitats (0% highly suitable habitat). A reason for this can be that some footprint types are also correlated with more suitable habitats (e.g. forestry). The effect of surrounding human footprint became negative when it was replacing highly suitable habitats (100% suitable habitat in the landscape) (Figure 5). Realized combinations of suitable habitats and footprint types in the surrounding area vary greatly, thus combinations between these two extreme situations are likely to drive observed differences between the spatial predictions based on different modeling stages (Figure 6).

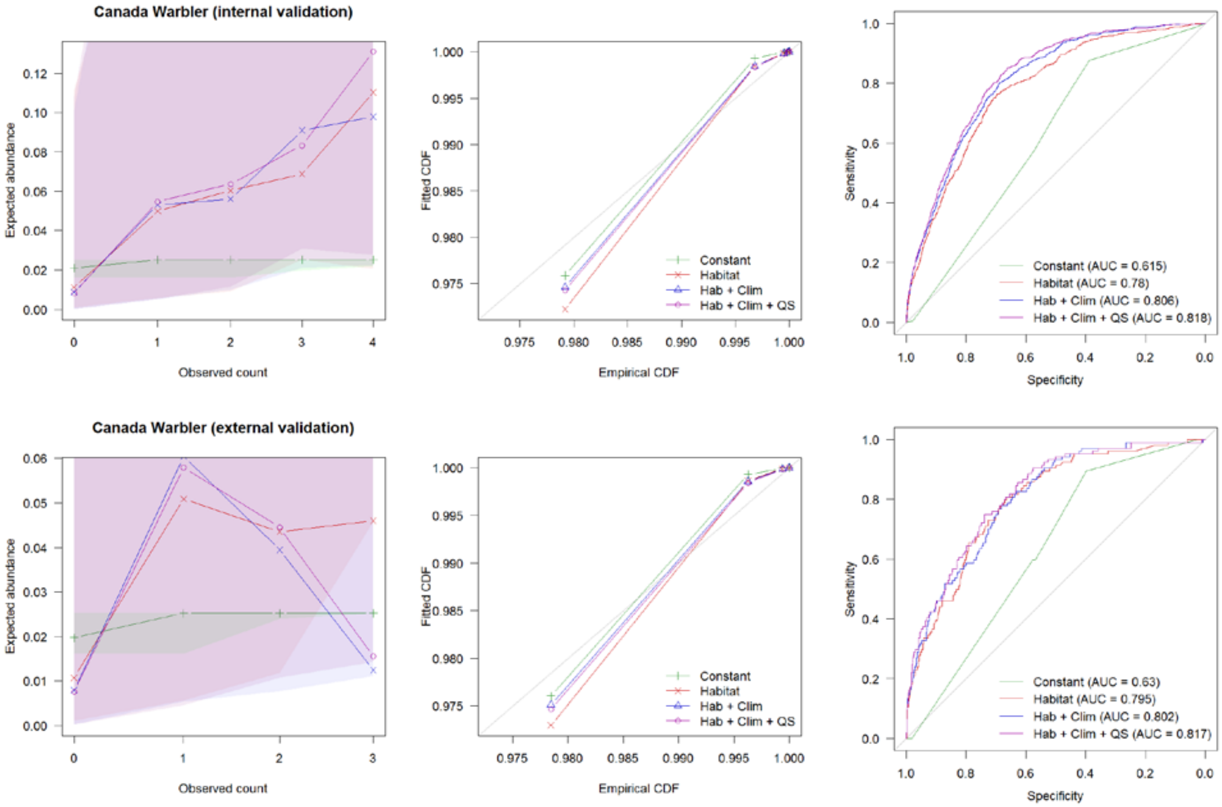

Figure 2. Goodness of fit based on bootstrap smoothed predictions using the data set used for model building (internal validation; top row), and using the held-out part of the data set (not used for model building, external validation; bottom row). First graph in each row show the model's ability to accurately rank survey points based on expected bird density given the predictor combinations at that location. An increasing expected abundance value with increasing observed counts indicates good ranking compared to the null (constant) model. The middle graphs show the cumulative density function (CDF) based on the observed counts (Observed CDF) vs. the fitted model (Fitted CDF). Quantile points closer to the diagonal line indicate better fit. Right hand side plots are receiver operator characteristic functions (ROC) with area under the curve (AUC) values for each modeling stage. Higher AUC indicate better ability to discriminate between 0 and >0 observations. Predicted values incorporate detectability offsets besides bootstrap smoothed predicted values.

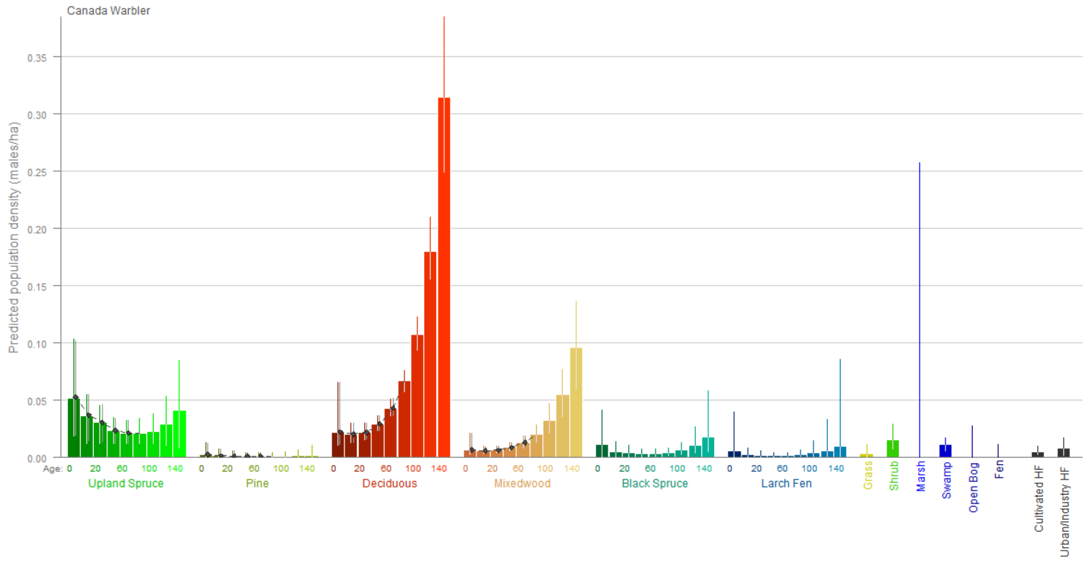

Figure 3. Local (point count, 150 m radius buffer) scale density estimates. Forested habitats (green) are plotted by 20-year age classes, non-forested habitat classes including urban-industrial and cultivation footprint types are plotted in the right hand side of the graph. Forestry effect is shown in grey lines for harvestable stand types. Effects of linear features were removed statistically. Vertical bars represent 90% confidence intervals based on bootstrap.

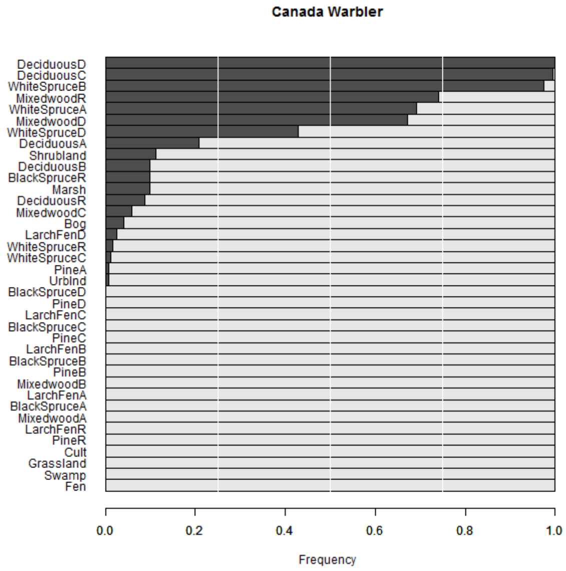

Figure 4. Habitat suitability was estimated independent of the variable selection process to allow for the determination of the amount of surrounding suitable habitat in the 451 m radius buffer (64 ha = area of a quarter section). High suitability habitat classes were determined based on the Lorenz threshold after ranking habitats based on their estimated Poisson means. Selection frequencies are based on bootstrap runs. Age categories (A--D) correspond to a modified version of the avian habitat classification system by Environment Canada: R = recent burn; A = herb/shrub stage; B = pole/sapling stage, C = young forest; D = mature and old forest.

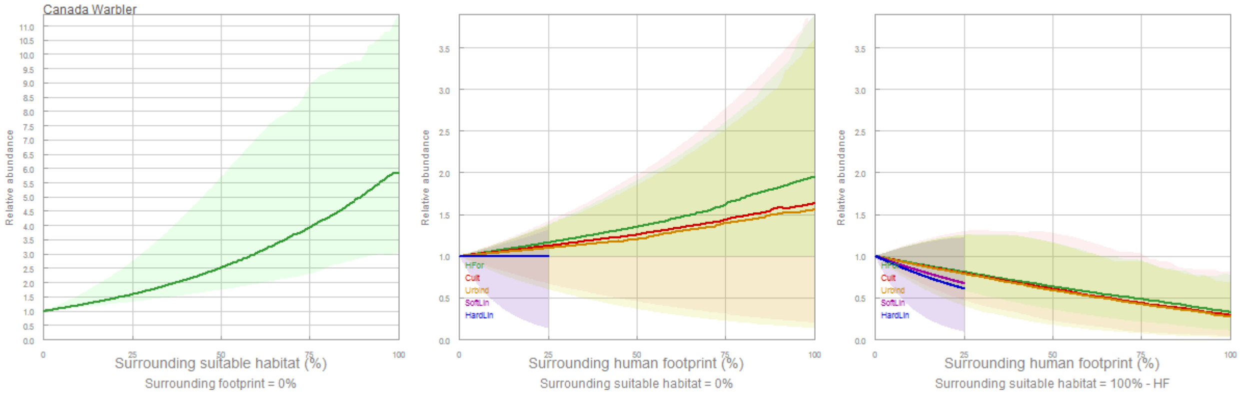

Figure 5. Quarter-section level effects of surrounding land cover type and amount. The plot in the left shows the effect of surrounding suitable are based on Lorenz threshold (amount of footprint other than what is included in suitable habitats was set to 0%). The middle and right plot shows the effects of different groups of human footprint types under two extreme scenarios. The middle figure shows the situation when footprint increases in a 100% unsuitable landscape, the right figure shows when 100% suitable landscape is increasingly converted into human footprint. Shaded polygons represent 90% confidence intervals based on bootstrap.

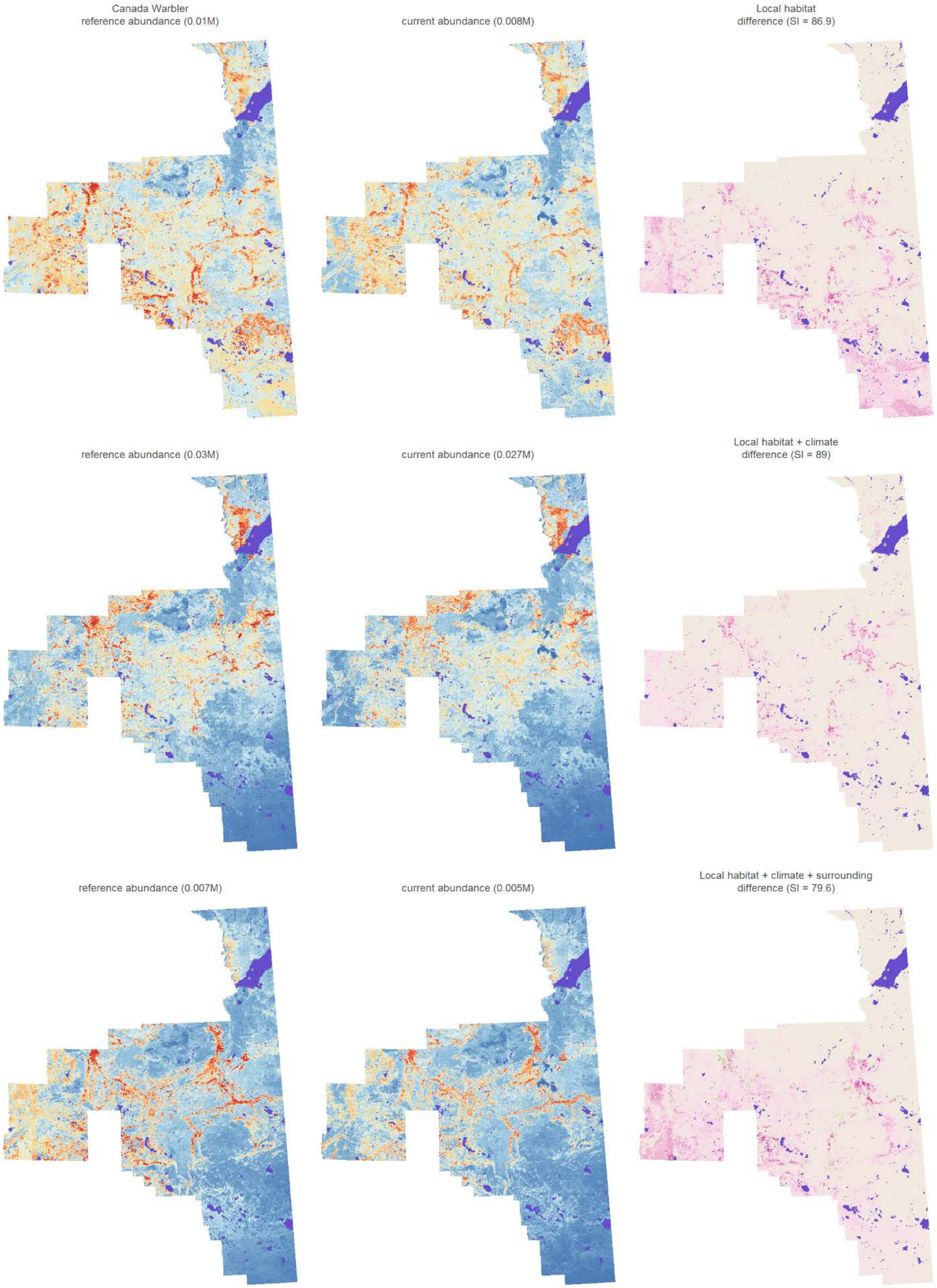
Canada Warbler
reference abundance (0.01M)
current abundance (0.008M)
Local habitat
difference (SI = 86.9)
reference abundance (0.03M)
current abundance (0.027M)
Local habitat + climate
difference (SI = 89)
reference abundance (0.007M)
current abundance (0.005M)
Local habitat + climate + surrounding
difference (SI = 79.6)

Figure 6. Predictions of reference abundance (left) in each quarter section were made after all human footprint in the quarter section had been `backfilled' based on native vegetation in the surrounding area. Predictions of current abundance (middle) in each quarter section were made based on the vegetation and human footprint present in the quarter section in 2010. Quarter sections depicted in red were predicted to have the highest abundance for the species, grading through light tan to dark blue where the species was predicted to be less abundant or absent. For each quarter section the difference between predicted current and reference conditions was determined (right). In quarter sections depicted in green the species was predicted to have higher abundance under present conditions than under reference conditions, with the opposite true for quarter sections depicted in red. The intensity of green and red depict the relative magnitude of increase or decrease for the species between reference and current conditions. The first row is based on local habitat effects only (model stage 4), second row illustrates local effects and climate (stage 5), while the third row incorporates the surrounding habitat effects as well (stage 6). Estimated population sizes (in million males) and two sided intactness (SI) values corresponding to a given sub-plot are indicated in the figure legends.

Sector effects revealed that forestry had greatest effect on CAWA habitats and expected populations, because the unit effect was >0.5 resulting in a 5.71% loss of the expected population. Mines/Wells also had >0.5 unit effect, but resulted only in 1.1% loss due to the smaller extend of that footprint type. Unit effects of these footprint types were high because these tend to affect best CAWA habitats where abundance is concentrated – thus leading to disproportional loss of suitable habitats. Although unit effect of agriculture was smaller, its overall population effect was higher than that of mines and wells due to the large extent of cultivated land (Figure 7).

When looking at direct vs. indirect effects of the different sectors on CAWA habitats, the indirect negative effect of soft linear features was highest (-6.28%), followed by the direct effect of forestry (-5.71%) and agriculture (-2.83%), and the indirect effect of transportation (-2.13%). Altogether, indirect effect led by energy related soft linear features contributed to third of the total footprint effect (Figure 8).

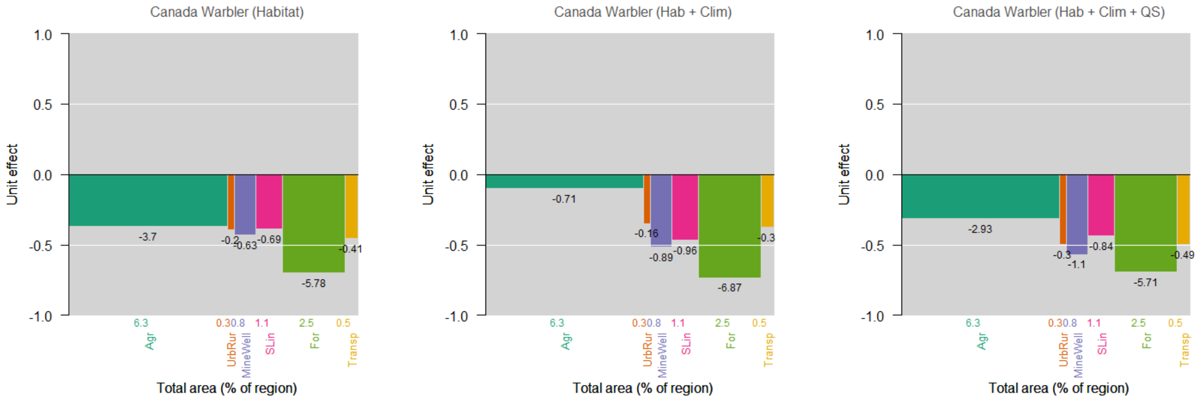

Figure 7. Effects of different sectors (x axis) on expected abundance (numbers on the bars) are proportional to the areas of the rectangles (unit effect times area of footprint). The \% population difference is calculated relative to the total reference population abundance in the region. X axis shows the percent area of the footprint types in the region. Y axis is the rescaled unit effect of the footprint type. Absolute values close to 0.5 indicate that the population effect is closely linear. The left, middle and right plots show the sector effects based on model stages 4, 5, and 6, respectively. Surrounding habitat effects (indirect footprint effects) in case of stage 6 are not incorporated into the right hand side figure.

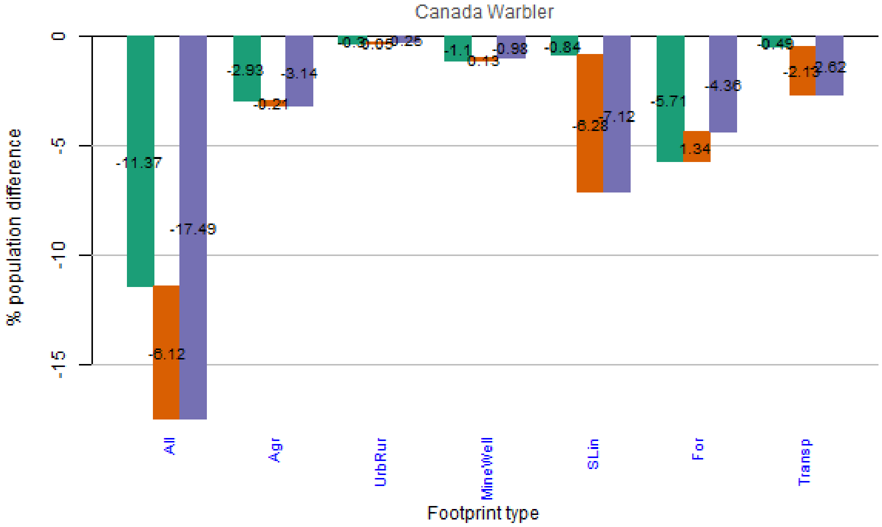

Figure 8. The direct effects (green bars) of different footprint types grouped according to sectors are based on the known habitat conversion rates from the `backfilled' and current landscape. Surrounding habitat effects in case of model stage 6 were used to estimate indirect effects (orange bars) of the sectors on the species populations based on partial `backfilling' of the landscape during prediction to represent the effect of a single sector. The total effect on the species population is the sum of the direct and indirect effects (blue bars). Numbers represent percent population difference within the region relative to the total reference population size.

Community level overview

We have used the Lorenz tangent based selection frequencies

Habitat associations based on the Lorenz tangent approach that were used to delineate suitable habitats at the quarter section scale provided a simple means for summarizing habitat associations across species. We compiled bootstrap based selection frequencies for habitat classes and performed a redundancy analysis (RDA). RDA is a multivariate ordination technique similar to linear models, and is used when responses to environmental gradients are assumed to be linear. The ‘triplot’ in Figure 9 shows the summary of 3 attributes: species, habitat classes, and sector specific % population change.

The figure shows 3 gradients. The 1st axis spans across a wetness gradient from wet/lowland habitats in the left to dry/upland habitat in the right. The second gradient along the 2nd axis is related to disturbance: from undisturbed (bottom) to disturbed (top). A 3rd gradient runs diagonally from the upper left corner to the bottom right, indicating a young-to-old age gradient with open and early seral habitats at one end and mature and old-growth forests at the other end. Increaser species were mostly associated with anthropogenic habitats, while many of the decreaser species were associated with undisturbed old forests. Effects of forestry were correlated with the diagonal age gradient. Urban/industrial, agriculture and mines/wells correlated with the disturbance gradient.

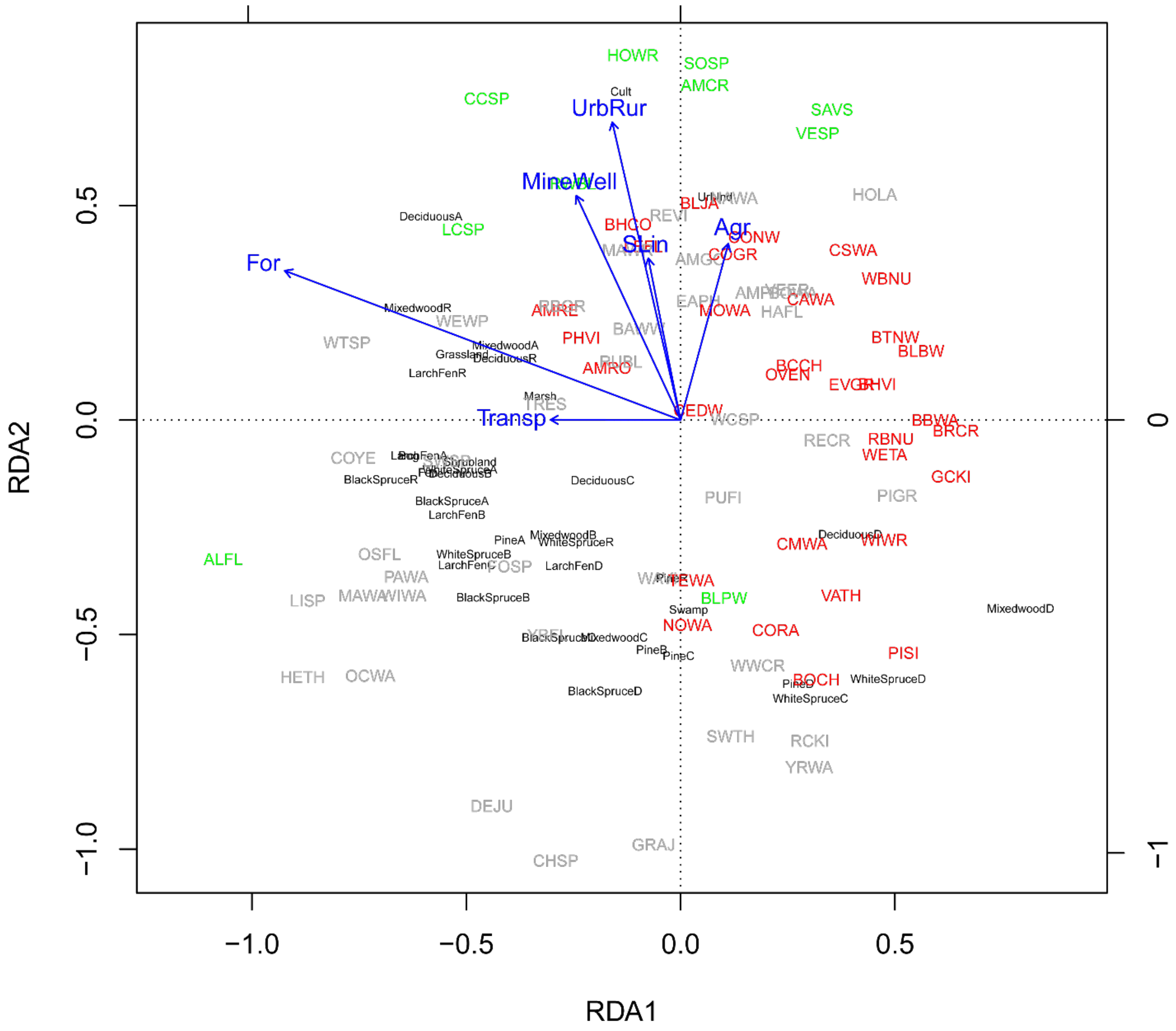

Figure 9. Ordination diagram showing the relationship between species, habitat classes (black), and sectors (blue). Species in green are 'increasers' (Ncr > Nrf), red indicated 'decreaser' species (Ncr < Nrf). Grey species IDs represent neutral species (similar Ncr and Nrf). The relative position of species and habitats indicate correlations among the labels, blue arrows represent the direction and strength of correlation with human footprint sector effects.

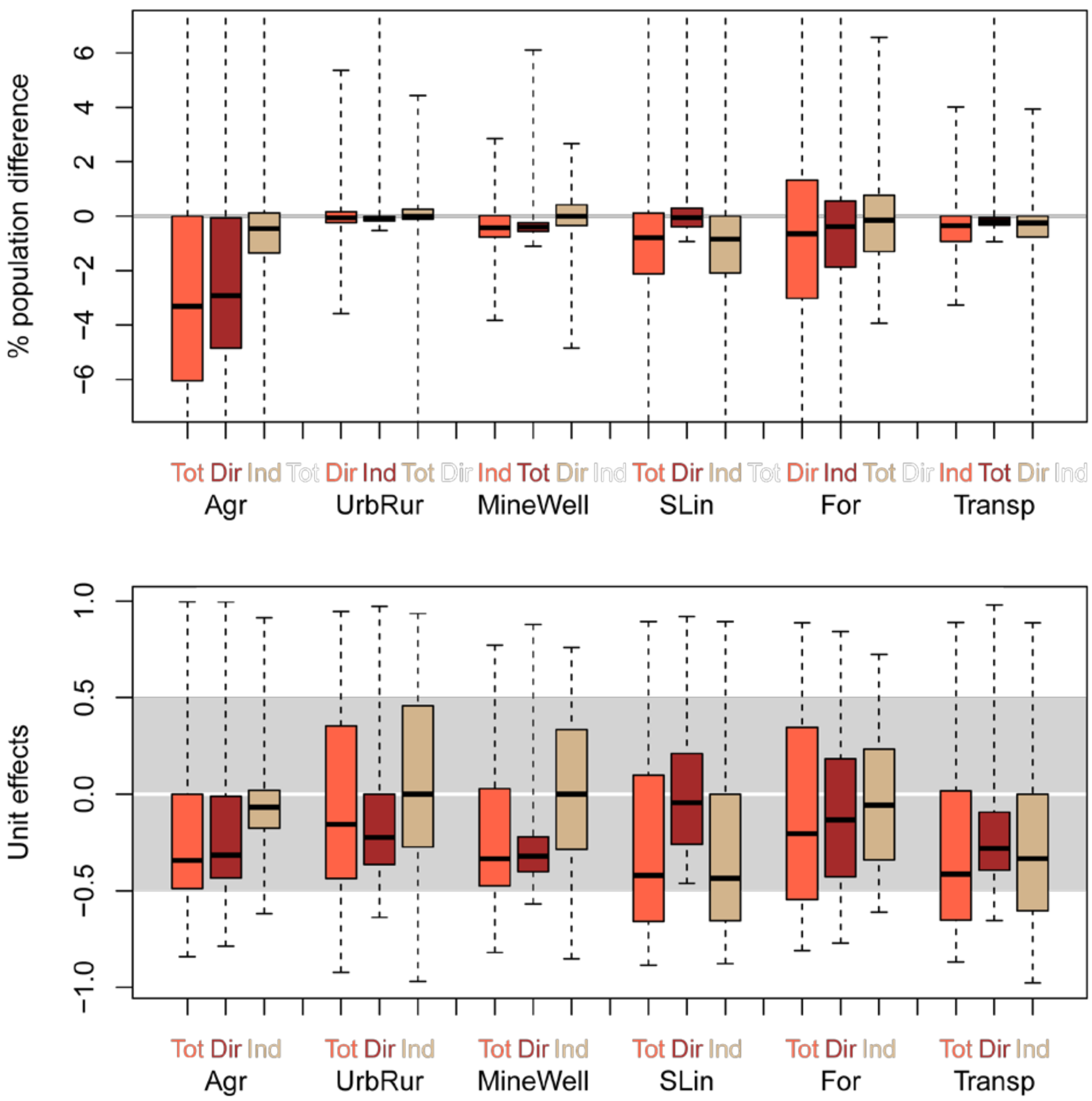

Figure 10. Sector specific effects on % population difference between current and reference estimated population sizes within the oil sands area (top) and unit effects for each sector after standardizing % population difference by % area of the footprint associated with a sector (bottom). Whiskers represent min and max, boxes are inter quartile ranges, lines are medians; 81 species were used.

Figure 10 show the sector effects across all 81 species analyzed. Agriculture had largest effect on species populations, and that effect came mostly from direct effects. The population effect from urban/industrial sites and mines/wells was relatively minor compared to agriculture. The second larges contributor to population differences was forestry, here the effects were mostly due to direct effects and to a lesser extent due to indirect effects. Energy sector related linear features also had big effect mostly due to indirect effects. Transportation related effects were similar to energy related linear features but smaller in magnitude.

The absolute magnitude of population differences by sectors reflect the % area of the footprint associated with each sector within the oil sand area. Area of agriculture was the highest (6.3%), followed by forestry (2.5%), soft linear features (1.1%), mines/wells (0.84%), transportation (0.5%), and urban/industrial (0.3%).

After standardizing the direct and indirect % population effects by the % area of the footprint we get unit effects that show the potential of a given sector to impact species populations. Bottom half of Figure 10 show unit effects where we can see that indirect effects of linear features have the larges effect per unit area, the unit effect being less than -0.5 for many species. A unit effect less than -0.5 indicates that the population loss is disproportionate, e.g. convering 1% of the landscape leads to the potentially loss of >1% of the population. To some extent this is also true for the total (direct+indirect) unit effect of forestry. Sectors with extreme unit effect values have powerful effects on populations because those selectively target higly suitable habitats for many species. For example forestry targeting old forests. In the case of linear features, direct effects are less powerful than indirect effects, which means that linear features affect nearby suitable habitats most likely through edge effects.

Limitations

- The current abundance estimate has no reference to a particular year, it represents average habitat conditions over the time span of the data set with inter annual variation present.
- Reference abundances were estimated based on the backfilled vegetation map and using the same statistical models as current abundance. Reference abundance represent the predictions conditional on the layer where footprint was removed and the original habitat ‘restored’ based on vegetation in the neighborhood. We did not estimate the actual abundance of the species prior to footprint. The reference abundance estimates is best interpreted as a measure of change in habitat suitability of species as a result of footprint. The scale of this change is relative to the current abundance estimate, and measured in same units.
- Some predictors used in the modeling might be measured with error (e.g. forest age) or misclassified (e.g. habitat classes). These errors might have biased our results and increased the variance of our estimates. These errors are also present in the predictions.
- The spatial and habitat specific estimates in this report did not utilize all the bootstrap runs, thus the bootstrap smoothing, but only the 1st run out of *B* replicates. This was due to computing limitations, and we are working on updating the estimates using all bootstrap replicates.

Acknowledgements

We would like to thank to Daiyuan Pan and Jim Schieck for help in accessing the geospatial summaries required for the modeling; Dave Huggard, Ermias Azeria for feedback on model covariates and sector effects calculations; Diana Stralberg for providing the climate data layers. This research was enabled in part by support provided by WestGrid (www.westgrid.ca) and Compute Canada Calcul Canada (www.computecanada.ca). This report is a contribution of the Boreal Avian Modelling (BAM) Project, an international research collaboration on the ecology, management, and conservation of boreal birds. We acknowledge the BAM Project's members, data partners, and funding agencies (including Environment Canada, the U.S. Fish & Wildlife Service and Joint Oil Sands Monitoring Implementation Plan), listed in full at www.borealbirds.ca/index.php/acknowledgements.

References

Alberta Biodiversity Monitoring Institute. 2013. ABMI Wall-to-wall Human Footprint Map 2010 Version 4.3. URL http://www.abmi.ca. Alberta Biodiversity Monitoring Institute, Alberta, Canada.

Alberta Biodiversity Monitoring Institute. 2014. 2000 Alberta Backfilled Wall-to-Wall Land Cover. Version 2.5 - Metadata. URL http://www.abmi.ca. Alberta Biodiversity Monitoring Institute, Alberta, Canada.

Barnhart, H. X., M. Haber, and J. Song. 2002. Overall concordance correlation coe_cient for evaluating agreement among multiple observers. Biometrics, 58:1020-1027.

Breiman, L. Bagging predictors. Machine Learning, 24:123-140, 1996.

Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris. 2002. A knowledge-based approach to the statistical mapping of climate. Climate Research, 22:99-113.

Efron, B. 2014. Estimation and accuracy after model selection. Journal of the American Statistical Association, 109:991-1007.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 1965-1978.

Schisterman, E. F., N. J. Perkins, A. Liu, and H. Bondell. Optimal cut-point and its corresponding Youden index to discriminate individuals using pooled blood samples. Epidemiology, 16:73-81, 2005.

Sólymos, P., Matsuoka, S. M., Bayne, E. M., Lele, S. R., Fontaine, P., Cumming, S. G., Stralberg, D., Schmiegelow, F. K. A., and Song, S. J., 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.

Wang, T., A. Hamann, D. L. Spittlehouse, and T. Q. Murdock. 2011. ClimateWNA - high-resolution spatial climate data for western North America. Meteorology and Climatology, 51, 16-29.

Youden, W. J. Index for rating diagnostic tests. Cancer, 3:32-35, 1950.

Appendices

Appendix 1. List of species mentioned in this report

Species ID	English name	Scientific name
ALFL	Alder Flycatcher	*Empidonax alnorum*
AMCR	American Crow	*Corvus brachyrhynchos*
AMGO	American Goldfinch	*Spinus tristis*
AMPI	American Pipit	*Anthus rubescens*
AMRE	American Redstart	*Setophaga ruticilla*
AMRO	American Robin	*Turdus migratorius*
BBWA	Bay-breasted Warbler	*Setophaga castanea*
BAWW	Black-and-white Warbler	*Mniotilta varia*
BCCH	Black-capped Chickadee	*Poecile atricapillus*
BTNW	Black-throated Green Warbler	*Setophaga virens*
BLBW	Blackburnian Warbler	*Setophaga fusca*
BLPW	Blackpoll Warbler	*Setophaga striata*
BHVI	Blue-headed Vireo	*Vireo solitarius*
BLJA	Blue Jay	*Cyanocitta cristata*
BOWA	Bohemian Waxwing	*Bombycilla garrulus*
BOCH	Boreal Chickadee	*Poecile hudsonicus*
BHCO	Brown-headed Cowbird	*Molothrus ater*
BRCR	Brown Creeper	*Certhia americana*
CAWA	Canada Warbler	*Cardellina canadensis*
CMWA	Cape May Warbler	*Setophaga tigrina*
CEDW	Cedar Waxwing	*Bombycilla cedrorum*
CSWA	Chestnut-sided Warbler	*Setophaga pensylvanica*
CHSP	Chipping Sparrow	*Spizella passerina*
CCSP	Clay-colored Sparrow	*Spizella pallida*
COGR	Common Grackle	*Quiscalus quiscula*
CORA	Common Raven	*Corvus corax*
COYE	Common Yellowthroat	*Geothlypis trichas*
CONW	Connecticut Warbler	*Oporornis agilis*
DEJU	Dark-eyed Junco	*Junco hyemalis*
EAPH	Eastern Phoebe	*Sayornis phoebe*
EVGR	Evening Grosbeak	*Coccothraustes vespertinus*
FOSP	Fox Sparrow	*Passerella iliaca*
GCKI	Golden-crowned Kinglet	*Regulus satrapa*
GRAJ	Gray Jay	*Perisoreus canadensis*
HAFL	Hammond's Flycatcher	*Empidonax hammondii*
HETH	Hermit Thrush	*Catharus guttatus*
HOLA	Horned Lark	*Eremophila alpestris*

HOWR	House Wren	*Troglodytes aedon*
LCSP	Le Conte's Sparrow	*Ammodramus leconteii*
LEFL	Least Flycatcher	*Empidonax minimus*
LISP	Lincoln's Sparrow	*Melospiza lincolnii*
MAWA	Magnolia Warbler	*Setophaga magnolia*
MAWR	Marsh Wren	*Cistothorus palustris*
MOWA	Mourning Warbler	*Geothlypis philadelphia*
NAWA	Nashville Warbler	*Oreothlypis ruficapilla*
NOWA	Northern Waterthrush	*Parkesia noveboracensis*
OSFL	Olive-sided Flycatcher	*Contopus cooperi*
OCWA	Orange-crowned Warbler	*Oreothlypis celata*
OVEN	Ovenbird	*Seiurus aurocapilla*
PAWA	Palm Warbler	*Setophaga palmarum*
PHVI	Philadelphia Vireo	*Vireo philadelphicus*
PIGR	Pine Grosbeak	*Pinicola enucleator*
PISI	Pine Siskin	*Spinus pinus*
PUFI	Purple Finch	*Haemorhous purpureus*
RBNU	Red-breasted Nuthatch	*Sitta canadensis*
REVI	Red-eyed Vireo	*Vireo olivaceus*
RWBL	Red-winged Blackbird	*Agelaius phoeniceus*
RECR	Red Crossbill	*Loxia curvirostra*
RBGR	Rose-breasted Grosbeak	*Pheucticus ludovicianus*
RCKI	Ruby-crowned Kinglet	*Regulus calendula*
RUBL	Rusty Blackbird	*Euphagus carolinus*
SAVS	Savannah Sparrow	*Passerculus sandwichensis*
SOSP	Song Sparrow	*Melospiza melodia*
SWTH	Swainson's Thrush	*Catharus ustulatus*
SWSP	Swamp Sparrow	*Melospiza georgiana*
TEWA	Tennessee Warbler	*Oreothlypis peregrina*
TRES	Tree Swallow	*Tachycineta bicolor*
VATH	Varied Thrush	*Ixoreus naevius*
VEER	Veery	*Catharus fuscescens*
VESP	Vesper Sparrow	*Pooecetes gramineus*
WAVI	Warbling Vireo	*Vireo gilvus*
WETA	Western Tanager	*Piranga ludoviciana*
WEWP	Western Wood-Pewee	*Contopus sordidulus*
WBNU	White-breasted Nuthatch	*Sitta carolinensis*
WCSP	White-crowned Sparrow	*Zonotrichia leucophrys*
WTSP	White-throated Sparrow	*Zonotrichia albicollis*
WWCR	White-winged Crossbill	*Loxia leucoptera*
WIWA	Wilson's Warbler	*Cardellina pusilla*

WIWR	Winter Wren	*Troglodytes hiemalis*
YBFL	Yellow-bellied Flycatcher	*Empidonax flaviventris*
YRWA	Yellow-rumped Warbler	*Setophaga coronata*

Appendix 2. Models terms and stages of the hierarchical model selection procedure

Stage	Model ID	Stage/model ID	Terms
Hab	0	1.0	NULL
Hab	1	1.1	. + hab1
Hab	2	1.2	. + hab1b
Age	0	2.0	NULL
Age	1	2.1	. + wtAge
Age	2	2.2	. + wtAge + wtAge2
Age	3	2.3	. + wtAge + wtAge2 + wtAge:isCon + wtAge2:isCon
Age	4	2.4	. + wtAge + wtAge2 + wtAge:isUpCon + wtAge:isBSLarch + wtAge2:isUpCon + wtAge2:isBSLarch
Age	5	2.5	. + wtAge + wtAge2 + wtAge:isMix + wtAge:isPine + wtAge:isWSpruce + wtAge:isBSLarch + wtAge2:isMix + wtAge2:isPine + wtAge2:isWSpruce + wtAge2:isBSLarch
Age	6	2.6	. + wtAge05
Age	7	2.7	. + wtAge05 + wtAge05:isCon
Age	8	2.8	. + wtAge05 + wtAge05:isUpCon + wtAge05:isBSLarch
Age	9	2.9	. + wtAge05 + wtAge05:isMix + wtAge05:isPine + wtAge05:isWSpruce + wtAge05:isBSLarch
Age	10	2.1	. + wtAge05 + wtAge
Age	11	2.1	. + wtAge05 + wtAge + wtAge05:isCon + wtAge:isCon
Age	12	2.1	. + wtAge05 + wtAge + wtAge05:isUpCon + wtAge05:isBSLarch + wtAge:isUpCon + wtAge:isBSLarch
Age	13	2.1	. + wtAge05 + wtAge + wtAge05:isMix + wtAge05:isPine + wtAge05:isWSpruce + wtAge05:isBSLarch + wtAge:isMix + wtAge:isPine + wtAge:isWSpruce + wtAge:isBSLarch
CC	0	3.0	NULL
CC	1	3.1	. + fCC2
Contrast	0	4.0	NULL
Contrast	1	4.1	. + ROAD01
Contrast	2	4.2	. + SoftLin_PC
Contrast	3	4.3	. + ROAD01 + SoftLin_PC
Contrast	4	4.4	. + ROAD01 + ROAD01:LCC2
Contrast	5	4.5	. + ROAD01 + SoftLin_PC + ROAD01:LCC2
Contrast	6	4.6	. + ROAD01 + ROAD01:LCC3
Contrast	7	4.7	. + ROAD01 + SoftLin_PC + ROAD01:LCC3
Contrast	8	4.8	. + ROAD01 + ROAD01:LCC5
Contrast	9	4.9	. + ROAD01 + SoftLin_PC + ROAD01:LCC5
Space	0	5.0	NULL
Space	1	5.1	. + xPET
Space	2	5.2	. + xMAT
Space	3	5.3	. + xAHM

Space	4	5.4	. + xFFP
Space	5	5.5	. + xMAP + xFFP
Space	6	5.6	. + xMAP + xFFP + xMAP:xFFP
Space	7	5.7	. + xMAT + xMAP + xPET + xAHM
Space	8	5.8	. + xMAT + xMAP + xPET + xAHM + xPET:xMAP + xMAT:xAHM
Space	9	5.9	. + xMAT + xMAP
Space	10	5.1	. + xMWMT + xMCMT
Space	11	5.1	. + xAHM + xPET
Space	12	5.1	. + xlat + xlong + xlat:xlong + xPET
Space	13	5.1	. + xlat + xlong + xlat:xlong + xMAT
Space	14	5.1	. + xlat + xlong + xlat:xlong + xAHM
Space	15	5.2	. + xlat + xlong + xlat:xlong + xFFP
Space	16	5.2	. + xlat + xlong + xlat:xlong + xMAP + xFFP
Space	17	5.2	. + xlat + xlong + xlat:xlong + xMAP + xFFP + xMAP:xFFP
Space	18	5.2	. + xlat + xlong + xlat:xlong + xMAT + xMAP + xPET + xAHM
Space	19	5.2	. + xlat + xlong + xlat:xlong + xMAT + xMAP + xPET + xAHM + xPET:xMAP + xMAT:xAHM
Space	20	5.2	. + xlat + xlong + xlat:xlong + xMAT + xMAP
Space	21	5.2	. + xlat + xlong + xlat:xlong + xMWMT + xMCMT
Space	22	5.2	. + xlat + xlong + xlat:xlong + xAHM + xPET
Space	23	5.2	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xPET
Space	24	5.2	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xMAT
Space	25	5.3	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xAHM
Space	26	5.3	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xFFP
Space	27	5.3	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xMAP + xFFP
Space	28	5.3	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xMAP + xFFP + xMAP:xFFP
Space	29	5.3	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xMAT + xMAP + xPET + xAHM
Space	30	5.3	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xMAT + xMAP + xPET + xAHM + xPET:xMAP + xMAT:xAHM
Space	31	5.3	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xMAT + xMAP
Space	32	5.3	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xMWMT + xMCMT
Space	33	5.3	. + xlat + xlong + xlat:xlong + xlat2 + xlong2 + xAHM + xPET
QSHF	0	6.0	NULL
QSHF	1	6.1	. + Remn_QS
QSHF	2	6.2	. + Remn_QS + Remn2_QS
QSHF	3	6.3	. + THF_QS
QSHF	4	6.4	. + Lin_QS + Nonlin_QS
QSHF	5	6.5	. + Succ_QS + Alien_QS
QSHF	6	6.6	. + Succ_QS + Noncult_QS + Cult_QS
QSHF	7	6.7	. + THF_QS + THF2_QS

QSHF	8	6.8	. + Lin_QS + Nonlin_QS + Nonlin2_QS
QSHF	9	6.9	. + Succ_QS + Alien_QS + Succ2_QS
QSHF	10	6.1	. + Succ_QS + Noncult_QS + Cult_QS + Succ2_QS
QSHF	11	6.1	. + Succ_QS + Alien_QS + Alien2_QS
QSHF	12	6.1	. + Succ_QS + Noncult_QS + Cult_QS + Noncult2_QS
QSHF	13	6.1	. + Succ_QS + Alien_QS + Succ2_QS + Alien2_QS
QSHF	14	6.1	. + Succ_QS + Noncult_QS + Cult_QS + Succ2_QS + Noncult2_QS
QSHF	15	6.2	. + Remn_QS + THF_QS
QSHF	16	6.2	. + Remn_QS + Lin_QS + Nonlin_QS
QSHF	17	6.2	. + Remn_QS + Succ_QS + Alien_QS
QSHF	18	6.2	. + Remn_QS + Succ_QS + Noncult_QS + Cult_QS
QSHF	19	6.2	. + Remn_QS + THF_QS + THF2_QS
QSHF	20	6.2	. + Remn_QS + Lin_QS + Nonlin_QS + Nonlin2_QS
QSHF	21	6.2	. + Remn_QS + Succ_QS + Alien_QS + Succ2_QS
QSHF	22	6.2	. + Remn_QS + Succ_QS + Noncult_QS + Cult_QS + Succ2_QS
QSHF	23	6.2	. + Remn_QS + Succ_QS + Alien_QS + Alien2_QS
QSHF	24	6.2	. + Remn_QS + Succ_QS + Noncult_QS + Cult_QS + Noncult2_QS
QSHF	25	6.3	. + Remn_QS + Succ_QS + Alien_QS + Succ2_QS + Alien2_QS
QSHF	26	6.3	. + Remn_QS + Succ_QS + Noncult_QS + Cult_QS + Succ2_QS + Noncult2_QS
QSHF	27	6.3	. + Remn_QS + Remn2_QS + THF_QS
QSHF	28	6.3	. + Remn_QS + Remn2_QS + Lin_QS + Nonlin_QS
QSHF	29	6.3	. + Remn_QS + Remn2_QS + Succ_QS + Alien_QS
QSHF	30	6.3	. + Remn_QS + Remn2_QS + Succ_QS + Noncult_QS + Cult_QS
QSHF	31	6.3	. + Remn_QS + Remn2_QS + THF_QS + THF2_QS
QSHF	32	6.3	. + Remn_QS + Remn2_QS + Lin_QS + Nonlin_QS + Nonlin2_QS
QSHF	33	6.3	. + Remn_QS + Remn2_QS + Succ_QS + Alien_QS + Succ2_QS
QSHF	34	6.3	. + Remn_QS + Remn2_QS + Succ_QS + Noncult_QS + Cult_QS + Succ2_QS
QSHF	35	6.4	. + Remn_QS + Remn2_QS + Succ_QS + Alien_QS + Alien2_QS
QSHF	36	6.4	. + Remn_QS + Remn2_QS + Succ_QS + Noncult_QS + Cult_QS + Noncult2_QS
QSHF	37	6.4	. + Remn_QS + Remn2_QS + Succ_QS + Alien_QS + Succ2_QS + Alien2_QS
QSHF	38	6.4	. + Remn_QS + Remn2_QS + Succ_QS + Noncult_QS + Cult_QS + Succ2_QS + Noncult2_QS

Appendix 3. A list of land cover categories used in modeling

Veg & HF & Age	Age description	Veg & HF	EC	Sector
Deciduous0	Unknown	Deciduous	DeciduousD	Deciduous
DeciduousR	Recent (0-9 yr old)	Deciduous	DeciduousR	Deciduous
Deciduous1	10-19 yr old	Deciduous	DeciduousA	Deciduous
Deciduous2	20-39 yr old	Deciduous	DeciduousB	Deciduous
Deciduous3	40-59 yr old	Deciduous	DeciduousC	Deciduous
Deciduous4	60-79 yr old	Deciduous	DeciduousD	Deciduous
Deciduous5	80-99 yr old	Deciduous	DeciduousD	Deciduous
Deciduous6	100-119 yr old	Deciduous	DeciduousD	Deciduous
Deciduous7	120-139 yr old	Deciduous	DeciduousD	Deciduous
Deciduous8	140-159 yr old	Deciduous	DeciduousD	Deciduous
Deciduous9	160 yr and older	Deciduous	DeciduousD	Deciduous
Mixedwood0	Unknown	Mixedwood	MixedwoodD	Mixedwood
MixedwoodR	Recent (0-9 yr old)	Mixedwood	MixedwoodR	Mixedwood
Mixedwood1	10-19 yr old	Mixedwood	MixedwoodA	Mixedwood
Mixedwood2	20-39 yr old	Mixedwood	MixedwoodB	Mixedwood
Mixedwood3	40-59 yr old	Mixedwood	MixedwoodC	Mixedwood
Mixedwood4	60-79 yr old	Mixedwood	MixedwoodD	Mixedwood
Mixedwood5	80-99 yr old	Mixedwood	MixedwoodD	Mixedwood
Mixedwood6	100-119 yr old	Mixedwood	MixedwoodD	Mixedwood
Mixedwood7	120-139 yr old	Mixedwood	MixedwoodD	Mixedwood
Mixedwood8	140-159 yr old	Mixedwood	MixedwoodD	Mixedwood
Mixedwood9	160 yr and older	Mixedwood	MixedwoodD	Mixedwood
WhiteSpruce0	Unknown	WhiteSpruce	WhiteSpruceD	WhiteSpruce
WhiteSpruceR	Recent (0-9 yr old)	WhiteSpruce	WhiteSpruceR	WhiteSpruce
WhiteSpruce1	10-19 yr old	WhiteSpruce	WhiteSpruceA	WhiteSpruce
WhiteSpruce2	20-39 yr old	WhiteSpruce	WhiteSpruceB	WhiteSpruce
WhiteSpruce3	40-59 yr old	WhiteSpruce	WhiteSpruceC	WhiteSpruce
WhiteSpruce4	60-79 yr old	WhiteSpruce	WhiteSpruceC	WhiteSpruce
WhiteSpruce5	80-99 yr old	WhiteSpruce	WhiteSpruceD	WhiteSpruce
WhiteSpruce6	100-119 yr old	WhiteSpruce	WhiteSpruceD	WhiteSpruce
WhiteSpruce7	120-139 yr old	WhiteSpruce	WhiteSpruceD	WhiteSpruce
WhiteSpruce8	140-159 yr old	WhiteSpruce	WhiteSpruceD	WhiteSpruce
WhiteSpruce9	160 yr and older	WhiteSpruce	WhiteSpruceD	WhiteSpruce
Pine0	Unknown	Pine	PineD	Pine
PineR	Recent (0-9 yr old)	Pine	PineR	Pine
Pine1	10-19 yr old	Pine	PineA	Pine

Pine2	20-39 yr old	Pine	PineB	Pine
Pine3	40-59 yr old	Pine	PineC	Pine
Pine4	60-79 yr old	Pine	PineC	Pine
Pine5	80-99 yr old	Pine	PineD	Pine
Pine6	100-119 yr old	Pine	PineD	Pine
Pine7	120-139 yr old	Pine	PineD	Pine
Pine8	140-159 yr old	Pine	PineD	Pine
Pine9	160 yr and older	Pine	PineD	Pine
BlackSpruce0	Unknown	BlackSpruce	BlackSpruceD	BlackSpruce
BlackSpruceR	Recent (0-9 yr old)	BlackSpruce	BlackSpruceR	BlackSpruce
BlackSpruce1	10-19 yr old	BlackSpruce	BlackSpruceA	BlackSpruce
BlackSpruce2	20-39 yr old	BlackSpruce	BlackSpruceB	BlackSpruce
BlackSpruce3	40-59 yr old	BlackSpruce	BlackSpruceC	BlackSpruce
BlackSpruce4	60-79 yr old	BlackSpruce	BlackSpruceC	BlackSpruce
BlackSpruce5	80-99 yr old	BlackSpruce	BlackSpruceD	BlackSpruce
BlackSpruce6	100-119 yr old	BlackSpruce	BlackSpruceD	BlackSpruce
BlackSpruce7	120-139 yr old	BlackSpruce	BlackSpruceD	BlackSpruce
BlackSpruce8	140-159 yr old	BlackSpruce	BlackSpruceD	BlackSpruce
BlackSpruce9	160 yr and older	BlackSpruce	BlackSpruceD	BlackSpruce
LarchFen0	Unknown	LarchFen	LarchFenD	LarchFen
LarchFenR	Recent (0-9 yr old)	LarchFen	LarchFenR	LarchFen
LarchFen1	10-19 yr old	LarchFen	LarchFenA	LarchFen
LarchFen2	20-39 yr old	LarchFen	LarchFenB	LarchFen
LarchFen3	40-59 yr old	LarchFen	LarchFenC	LarchFen
LarchFen4	60-79 yr old	LarchFen	LarchFenC	LarchFen
LarchFen5	80-99 yr old	LarchFen	LarchFenD	LarchFen
LarchFen6	100-119 yr old	LarchFen	LarchFenD	LarchFen
LarchFen7	120-139 yr old	LarchFen	LarchFenD	LarchFen
LarchFen8	140-159 yr old	LarchFen	LarchFenD	LarchFen
LarchFen9	160 yr and older	LarchFen	LarchFenD	LarchFen
Bog		Bog	Bog	Bog
Fen		Fen	Fen	Fen
Swamp		Swamp	Swamp	Swamp
Marsh		Marsh	Marsh	Marsh
Shrubland		Shrubland	Shrubland	Shrubland
Grassland		Grassland	Grassland	Grassland
Bare		Bare	EXCLUDE	Bare
Water		Water	EXCLUDE	Water
BorrowpitsDugoutsSumps		Water	EXCLUDE	HFWater
Canals		Water	EXCLUDE	HFWater

CultivationCropPastureBareground		Cult	Cult	Agr
HighDensityLivestockOperation		Cult	Cult	Agr
IndustrialSiteRural		UrbInd	UrbInd	UrbRur
MineSite		UrbInd	UrbInd	MineWell
MunicipalWaterSewage		Water	EXCLUDE	HFWater
OtherDisturbedVegetation		UrbInd	UrbInd	UrbRur
PeatMine		UrbInd	UrbInd	MineWell
Pipeline		SoftLin	EXCLUDE	SLin
RailHardSurface		HardLin	EXCLUDE	Transp
RailVegetatedVerge		SoftLin	EXCLUDE	Transp
Reservoirs		Water	EXCLUDE	HFWater
RoadHardSurface		HardLin	EXCLUDE	Transp
RoadTrailVegetated		SoftLin	EXCLUDE	Transp
RoadVegetatedVerge		SoftLin	EXCLUDE	Transp
RuralResidentialIndustrial		UrbInd	UrbInd	UrbRur
SeismicLine		SoftLin	EXCLUDE	SLin
TransmissionLine		SoftLin	EXCLUDE	SLin
Urban		UrbInd	UrbInd	UrbRur
WellSite		UrbInd	UrbInd	MineWell
WindGenerationFacility		UrbInd	UrbInd	MineWell
CCDeciduous0	Unknown	CCDeciduous	DeciduousD	For
CCDeciduousR	Recent (0-9 yr old)	CCDeciduous	DeciduousR	For
CCDeciduous1	10-19 yr old	CCDeciduous	DeciduousA	For
CCDeciduous2	20-39 yr old	CCDeciduous	DeciduousB	For
CCDeciduous3	40-59 yr old	CCDeciduous	DeciduousC	For
CCDeciduous4	60-79 yr old	CCDeciduous	DeciduousD	For
CCMixedwood0	Unknown	CCMixedwood	MixedwoodD	For
CCMixedwoodR	Recent (0-9 yr old)	CCMixedwood	MixedwoodR	For
CCMixedwood1	10-19 yr old	CCMixedwood	MixedwoodA	For
CCMixedwood2	20-39 yr old	CCMixedwood	MixedwoodB	For
CCMixedwood3	40-59 yr old	CCMixedwood	MixedwoodC	For
CCMixedwood4	60-79 yr old	CCMixedwood	MixedwoodD	For
CCWhiteSpruce0	Unknown	CCWhiteSpruce	WhiteSpruceD	For
CCWhiteSpruceR	Recent (0-9 yr old)	CCWhiteSpruce	WhiteSpruceR	For
CCWhiteSpruce1	10-19 yr old	CCWhiteSpruce	WhiteSpruceA	For
CCWhiteSpruce2	20-39 yr old	CCWhiteSpruce	WhiteSpruceB	For
CCWhiteSpruce3	40-59 yr old	CCWhiteSpruce	WhiteSpruceC	For
CCWhiteSpruce4	60-79 yr old	CCWhiteSpruce	WhiteSpruceC	For
CCPine0	Unknown	CCPine	PineD	For

CCPineR	Recent (0-9 yr old)	CCPine	PineR	For
CCPine1	10-19 yr old	CCPine	PineA	For
CCPine2	20-39 yr old	CCPine	PineB	For
CCPine3	40-59 yr old	CCPine	PineC	For
CCPine4	60-79 yr old	CCPine	PineC	For
CCBlackSpruce0	Unknown	CCBlackSpruce	BlackSpruceD	For
CCBlackSpruceR	Recent (0-9 yr old)	CCBlackSpruce	BlackSpruceR	For
CCBlackSpruce1	10-19 yr old	CCBlackSpruce	BlackSpruceA	For
CCBlackSpruce2	20-39 yr old	CCBlackSpruce	BlackSpruceB	For
CCBlackSpruce3	40-59 yr old	CCBlackSpruce	BlackSpruceC	For
CCBlackSpruce4	60-79 yr old	CCBlackSpruce	BlackSpruceC	For
CCLarchFen0	Unknown	CCLarchFen	LarchFenD	For
CCLarchFenR	Recent (0-9 yr old)	CCLarchFen	LarchFenR	For
CCLarchFen1	10-19 yr old	CCLarchFen	LarchFenA	For
CCLarchFen2	20-39 yr old	CCLarchFen	LarchFenB	For
CCLarchFen3	40-59 yr old	CCLarchFen	LarchFenC	For
CCLarchFen4	60-79 yr old	CCLarchFen	LarchFenC	For
CCOpenTypes0	Unknown	CCOpenTypes	EXCLUDE	For
CCOpenTypesR	Recent (0-9 yr old)	CCOpenTypes	EXCLUDE	For
CCOpenTypes1	10-19 yr old	CCOpenTypes	EXCLUDE	For
CCOpenTypes2	20-39 yr old	CCOpenTypes	EXCLUDE	For
CCOpenTypes3	40-59 yr old	CCOpenTypes	EXCLUDE	For
CCOpenTypes4	60-79 yr old	CCOpenTypes	EXCLUDE	For

Appendix 4. Algorithm for calculating regional predictions for sector effects

Terminology

- **Habitat coefficient (c[h]):** relative abundance measure of a species in habitat class h. Habitat classes include natural vegetation and footprint as well. Relative abundance is expressed on the linear predictor scale (c[h] = intercept + contrast[h]) for the sake of generality (applies to Poisson and Binomial irrespective of the link function used, but see some comments later), and for additivity among the terms to hold. This coefficient is independent of spatial/climatic effect estimated at later steps in the all-in-one coefficient approach.
- **QS level spatial/climatic and HF effects (q[i]):** the sum of spatial terms (linear predictor scale) estimated after the habitat related step in the estimation process. In that step, c[h] was used as an offset. q[i] is the spatial term without the c[h] offset. QS level HF effect creates a reference (q_rf[i]) and current (q_cr[i]) version, where reference simply means that QS level HF effect is 0, and take observed values of HF and apply corresponding coefficients for current. All these terms (climate, space, HF) are additive on the linear predictor scale, and are added up to get q[i] (q_rf[i]) and q_cr[i]).
- **Polygon level predicted relative abundance (p[i,h]):** polygon level predicted value is the combination of the habitat and spatial terms weighted by the area of the polygon representing habitat h in QS i. c[h]+q[i] is on the linear predictor scale. A measure on the probability scale is calculated as p[i,h]=inverse_link_function(c[h]+q[i]).
- **Polygon level predicted population size (N[i,h]):** N[i,h]=p[i,h]* A[i,h].
- **QS level mean relative abundance (p_mean[i]):** this is the area weighted mean of polygon level predicted values on the probability scale. p_mean[i]=sum(p[i,h]*A[i,h]) / sum(A[i,h]). A[i,h] can be expressed as proportion: a[i,h]= A[i,h]/sum(A[i,h]) [summation here goes for habitats]. In this case, the calculation of QS level relative abundance reduces to: p_mean[i]=sum(p[i,h]*a[i,h]). Note, that calculations for current and relative abundances are based on the same polygons within QS. The only difference is the h label of polygons.
- **Habitat level mean relative abundance (p_mean[h]):** in any given study region, we can calculate the mean relative abundance for habitat type h as: p_mean[h]=sum(p[i,h]*A[i,h]) / sum(A[i,h]), where summation goes for QSs within the study region. Caution is needed when converting area into proportion, because there the sum needs to be the area of the region (we want to compare ‘total’ abundance contributions across habitat types). Thus it is advised to keep track of area instead of proportions to avoid confusions.
- **Total population size:** we calculate sums of N[i,h] for any combination of habitats and QSs.

Tracking population changes

There are two options: the label of a given polygon is the same for current (cr) and reference (rf) conditions (h_rf = h_cr), or they differ (h_rf != h_cr). The QS level summaries currently do not track this transition, but instead collapse polygons into reference (without HF types) and current (with HF types) classes. A straightforward modification of the summary is that instead create separate tables for current and reference, create only one with labels ‘h_rf->h_cr’. For example ‘Decid4->UrbInd’ would mean that a 60-80 yr old deciduous stand became urban-industrial polygon. ‘Decid4’ would indicate no change. Currently there is no need to use ‘Decid4->Decid4’ because it cannot change label (backcasting and forecasting would be a different story, as age would change as a result, but that wouldn't directly count towards sector effect except for through total population size). As a result, we end up with ~4800 potential polygon types (note: we are tracking all CC types and age classes as well – this needs to be known for prediction of current abundance).

Calculating predictions

Suppose now, that we have a matrix A of dimension n x m, where n is the number of QSs in the subregion, and m is the number of potential 'h_rf->h_cr' habitat combinations. Labels are then LinkIDs for rows and habitat transition labels for columns. Cells correspond to area of the polygon in km^2.

We also have a lookup table M where 'h_rf->h_cr' transition labels are mapped to the reference (h_rf) and current (h_cr) labels. Here is a toy example:

h_rf->h_cr	h_rf	h_cr	zero	exclude
Decid4->UrbInd	Decid4	UrbInd	FALSE	FALSE
Decid4	Decid4	Decid4	FALSE	FALSE
Water	Water	Water	FALSE	TRUE
Decid4	Decid4	Road	TRUE	FALSE

The lookup table can also contain optional weights/filters for:

- strata that should not be used when calculating mean abundances (e.g. water): in this case the corresponding polygon area needs to be set to 0, A[i,h]=0. This way N[i,h]=0=p[i,h]*0, but mean abundance is calculated without taking into account this strata. The toy table than read: exclude Water from summation of area. We get this by replacing corresponding values in A by 0:
  A[,exclude] <- 0
- Strata that should be calculated as 0 abundance (road surface, mine): we set relative abundance to 0, thus N[i,h]=0=0*A[i,h]. This way the area still counts when calculating mean and total population. We get this by replacing habitat specific predicted abundance values by 0:
  p[i,zero] <- 0, or equivalently: inverse_link_function(c)[zero] <- 0.

Calculation of reference and current relative abundance is no different than what is described before, except that the same label can appear in the sum multiple times. So the original coefficients, c[h] need to be mapped to the 2 columns: h_rf and h_cr.

Computational efficiencies

Using sparse matrices makes this process a bit more efficient, but we still need to do the prediction in subregion level areas to keep things within memory limits. This means also that we have to patch together the results later.

We create two vectors for habitat effects, one for reference abundance (c_rf[1:m]), using the mapping onto h_rf; and another one for current abundance (c_cr[1:m]), using the mapping onto h_cr.

Eventually, we are interested in summarizing these two matrices:

- D_rf[1:n,1:m] = t(c_rf[1:m] * t(A[1:n,1:m])) * q_rf[1:n]
- D_cr[1:n,1:m] = t(c_cr[1:m] * t(A[1:n,1:m])) * q_cr[1:n]

Weighted row means of D_rf and D_cr are QS level predictions for reference and current maps, respectively.

Column sums of D_rf and D_cr divided by column sums of A give the mean relative abundance in a habitat class (as indexed in the lookup table) within the region considered. This integrates over all the polygons in the study area: d_rf[h] = sum(D_rf[,h]) / sum(A[,h]), d_cr[h] = sum(D_cr[,h]) / sum(A[,h]); and these are the two vectors (including all classes from the lookup table) we are interested in for sector effects along with the vector from A_total[h]=sum(A[,h]) for keeping track of areas.

Calculating sector effects

We start from d_rf, d_cr and A_total (each of these vectors are of length m, and each element refels to a line in the lookup table for the 'transition' labels.

We can now create an arbitrary classification, e.g. lumping together the main HF for example. The sum of Nrf=sum(d_rf*A_total) is the total reference population. We treat this as 100%. Than we calculate the difference: dN = Ncr – Nrf = (d_cr*A_total) – (d_rf*A_total) = A_total * (d_cr - d_rf) to get 'population change'. This is 0 where the label hasn't changed. Then the sum of dN for different HF types will give the total population change for that HF type (sectors, of lumping refers to sectors).

meanDrf is mean density in the whole region: Nrf / sum(A_total). Calculation then:

- n[h] = dN[h] / Nrf, this is the potential pop change relative to total reference population
- a[h] = A_total[h] / sum(A_total [h]), proportional area in habitat h
- n[h] / a[h] = {A_total[h] * (d_cr[h] – d_rf[h]) / (meanDrf * sum(A_total[h])) } * (sum(A_total [h]) / A_total [h]) = (d_cr[h] – d_rf[h]) / meanDrf = dd[h] / meanDrf = dd[h]

dd[h] is density in habitat h relative to regional reference mean density. The sector effect plots showed dd[h] as a function of a[h], so that relative pop change, n[h]=a[h]*d[h] is, the area of the rectangle for habitat h.

Appendix 5. Habitat coefficients, population size estimates, and sector effects

Separate file: JOSM-birds-2015-appendices.xlsx

Appendix 6. Model outputs

Habitat associations, goodness-of-fit measures, predictive maps, sector effects plots are provided for each of the 81 bird species in a separate file: JOSM_report_2015_appendix-small.pdf