count at site *i* on survey occasion *t* was modeled as:

|  |  |
| --- | --- |
|  | (1) |

is a latent indicator variable (i.e., presence/absence) that is 1 if the site has non-zero density and 0 if the site contributes to the “zero-inflation” component of the model. Accordingly, was modeled as a logit-linear function of covariates:

|  |  |
| --- | --- |
|  | (2) |
|  | (3) |

The non-zero component of the model describes the expected counts at sites that are potentially occupied. This was modeled as a log-linear function of covariates, with random effects for PSU to account for local-scale autocorrelation in residual density among SSUs:

|  |  |
| --- | --- |
|  | (4) |
|  | (5) |

Offsets were calculated for each site, during each survey location, according to standard approaches described in XX.

An estimate of density (males per hectare) at a site, given site-level covariates, is therefore given by:

|  |  |
| --- | --- |
|  | (6) |

Based on previous analyses (Cumming et al. 2014; Stralberg et al. 2015a,b) we derived estimates of mean annual temperature (°C), climate moisture index (mean annual precipitation - potential evapotranspiration), degree days below 0°C, and degree days above 5°C. In addition to climatic covariates, we obtained digital elevation model derived indices including terrain roughness index which indexes changes in slope (Wilson et al. 2007) and topographic wetness index (Sörensen et al. 2006) which quantifies topographic control on hydrological processes and correlates with soil attributes such as horizon depth, silt percentage, and organic matter content. Climatic and terrain based covariates were obtained using the ‘envirem’ package (Title and Bemmels 2018) in the R statistical computing environment (R Core Team 2019).

We also derived a series of local vegetation variables representing forest structure, composition and biomass. Specifically, we obtained estimates percent tree cover (250m resolution), stand age, stand height (m), stand canopy closure, proportion of the the stand biomass comprised of needle-leaf trees, and proportion of the the stand biomass comprised of broadleaf-leaf trees. Percent tree cover was derived from the MODIS vegetation continuous fields version 6 product (DiMiceli et al. 2021). All other forest composition variables were derived from the products of Beaudoin et al (2014), which used a combination of 26 geospatial data layers and national forest inventory data to derive Canada wide estimates of forest attributes.

In order to avoid multicollinearity, we removed all highly correlated variables (r>0.60) prior to analyses; retaining the variables in correlated pairs that we felt would provide the best ecological explanatory power and biological interpretability. As a result, we removed degree days below 0°C and degree days above 5°C as they were highly correlated with mean annual temperature (°C, r >0.7). In addition, stand age, stand height and stand canopy closure were also correlated and we dropped stand height from consideration. In order to derive an orthogonal measure of stand closure, we modeled canopy closure as a function of stand height (using a quadratic fit in OLS regression) and save the residuals of this relationship as our measure of canopy closure. Proportion of the stand biomass composed of needle-leaf trees was inversely related to the proportion of broadleaf trees in the stand, and we retained needleleaf as the sole covariate of forest composition.

*Empirical example*

We modeled the distribution and density Olive-sided Flycatcher (*Contopus cooperi*) as a motivating example. We first began with a saturated model in which included stand age, stand age2, proportion needleleaf, proportion needleleaf2, residual canopy closure, and percent tree cover as fixed effect covariates and a random intercept for primary sampling unit in the count model. The zero-inflation component of the model included mean annual temperature, climate moisture index and topographic wetness index as covariates. A quadratic fit for mean annual temperature was also considered, but models would not converge using frequentist approaches. Using this saturated model, we then varied the inclusion of the variables in the zero-inflation component of the model and used Akaike’s Information Criterion to select the most parsimonious model formulation for zero-inflation. We then used the most parsimonious model formulation in a second stage of model selection where we varied the inclusion of variables in the count component of the model. Prior to analyses, all covariates were centered and standardized to improve model convergence. We only considered main effects and no interactions in the analyses. Models also included offsets to account for detection probability. Offsets were calculated using version 3 of the QPAD models (Sólymos 2016).

Based on the above analyses, the most parsimonious model was then used to build two competing Bayesian models, and we used posterior predictive checks to select which model formulation was more appropriate. The two competing models varied only in whether or not the zero-inflation component of the model included a quadratic effect for mean annual temperature. Bayesian models were run in the JAGS environment interfaced with the R programming language using the jagsUI library. We ran 35000 iterations with a burnin of 5000 iterations and stored every 50th iteration to estimate the posterior probability densities.

We generated pixel based population size estimates (following Sólymos et al. 2020) by generating spatially explicit predictions of bird densities onto the spatially explicit raster covariate layers and summing the densities across all non-water pixels. When estimates were back-transformed to produced the density estimates, we applied a lognormal bias correction as sum of the log-scale prediction plus 0.5 times the between site (PSU) variance following Sprugel (1983). One raster prediction was made per saved Markov Chain Monte Carlo iteration, and we used percentiles of the posterior distributions to quantify uncertainty in density and population size estimates respectively.

*Simulation*

To confirm the statistical model can accurately estimate provincial abundance, we conducted spatially explicit simulations for seven species of boreal passerines: Olive-sided Flycatcher, Dark-eyed Junco (*Junco hyemalis*), Orange-crowned Warbler (*Oreothlypis celata*), Ovenbird (*Seiurus aurocapilla*), Blackpoll Warbler (*Setophaga striata*), Yellow-rumped Warbler (*Setophaga coronata*), and Canada Warbler (*Cardellina canadensis*) (BAM 2020a-g).

For each species at each survey location, we extracted local density values from BAM national density model output raster maps. For each survey occasion, we generated Poisson-distributed count data while also incorporating empirical QPAD detectability offsets . We then applied the statistical model ( Specifically, each species simulated density at the survey locations was modeled using mean annual temperature and mean annual temperature2 along with climate moisture index as covariates for the zero-inflation submodel, and stand age, proportion needleleaf, residual canopy closure, and percent tree cover in the count submodel. We used the same pixel-based population size estimation approach (above) to generate mean population size (and 95% Credible Intervals) which were then compared to the “True” population size obtained by summing pixel values within the BAM mean species density maps.

*Results*

*Empirical example*

Initial AIC based based selection followed by posterior predictive checks suggested that the model including mean annual temperature and mean annual temperature2 along with climate moisture index in the zero-inflation sub-model, and main effects for each of stand age, proportion of stand in needleleaf trees, residual canopy closure, and percent tree cover in the count submodel provided the best fit to the Olive-sided Flycatcher data. Our analyses suggests the probability of an Olive-sided flycatcher is detected follows a quadratic relationship with mean annual temperature (Table 1), with probability of occurrence being greatest in regions where mean annual temperature as ~ -0.4°C annually and decreased in warmer and colder areas respectively. Probability of occurrence also increased with climate moisture index (Table 1). Of the factors included in the count component of the model, only stand age showed evidence of a meaningful effect (credible interval not overlapping zero) on Olive-sided Flycatcher density, with Flycatcher density decreasing with increasing stand age (Table 1). Using our model, we generated spatially explicit predictions of Olive-sided Flycatcher density across Saskatchewan, which shows densities generally being greatest in the Boreal Shield shield ecozone (BCR8) in the central boreal south of Lake Athabasca (Figure 2). After remove waterbodies from consideration, the posterior median population size estimate for Saskatchewan was 887,588 (95% CI: 594,500 –1,446,650) after applying the lognormal bias correction (Figure 3). Prediction uncertainty is greatest in the boreal fringe near Prince Albert, Saskatchewan (Figure 4), where BOSS sampling does not occur (note that volunteer surveys from the Saskatchewan Breeding Bird Atlas have not been included here). The coefficient of variation was generally higher in the southern boreal forest, but was also higher on the southwestern and northwestern shores of Lake Athabasca (Figure 4).

*Simulation*

With the exception of Yellow-rumped Warbler, 95% credible intervals for population size estimates encompassed the “true” population size for all the species we simulated (Figure 5). Population size estimates for Yellow-rumped Warbler were biased high, though this may have been driven by convergence issues in the current model runs. Thus, even this relatively simple model structure produced unbiased estimates of species population sizes.

**Discussion**

Our preliminary results suggest that our Bayesian framework building upon previous work by BAM could provide a relatively simple approach to ongoing status reporting. Importantly, even applying a single model parameterization across multiple species appears capable of recovering unbiased estimates of population status with our design based sampling. Further refinements to the model should further improve both model accuracy and precision as well as ease of application to other species. Not only does our approach provide mapped estimates of species densities and population size estimation, but the derivation of relatively interpretable and graphable parameter estimates allow for relatively easy ecological interpretation. In addition, the natural approach to error propagation under a Bayesian approach makes it comparatively easy to publish and map degree of uncertainty in our estimates. These additional products not only improve our scientific credibility, but could also be used to drive further sampling.

*Future refinements*

One current limitation of BAM modeling approaches is that it does not currently leverage data sources aside from point count data. Data sources such as e-Bird and Breeding Bird Atlases have vast archives of data not fitting standardized point count methods that could be leveraged using Bayesian methods because the methods allow for multiple data sources to be combined within the joint likelihoods of models similar to those used here. Specifically, we envision

checklist data (with known survey effort) could be used to improve the estimation of the zero-inflation component of count models which would likely greatly improve precision and prediction accuracy. Developing a modeling framework including these data would have an immediate benefit for our partners as this would allow Birds Canada to incorporate a unified model into mapping for their Atlas products (instead of separate abundance and ‘PObs” models). Furthermore, testing this framework could also point to potential refinements to ECCC field programs and more easily leverage data outside of ECCC and not currently being used by BAM in modeling. For example, having ECCC staff regularly collect checklist data may improve our status and trend reporting if we had a unified model incorporating both checklist and point count data (both human and ARU). This would potentially be a major advantage over BAM machine learning approaches which can only incorporate point count data. However, the ongoing refinement of QPAD offset estimates, and further use of species-specific variable importance values (e.g. see Table C2 and C3 of Stralberg et al. 2015b) from BAMs previous modeling efforts could help drive model parameterization.

Here, we used relatively simple formulations for our preliminary models and limited ourselves to including climate and topographic variables to the zero-inflation sub-model, and habitat covariates to the count sub-model. Our models could be further refined in a number of ways, including creating a series of species-specific ‘indicator’ variables to indicate habitats that are likely structural-zeros. These indicator variables could be included in the zero-inflation sub-model to improve localized prediction of species presence/absence. In addition, it is likely that topographically derived variables such as the topographic wetness index or perhaps estimates of distance to water would be useful for predicting high abundance areas for species such as Yellow Warbler (*Setophaga petechia*).

One chief advantage of the BAM machine learning approach is that there is no need to go through a complicated model selection process for dozens of species. As previously mentioned, using previous BAM variable importance scores could greatly speed model parameterization and likelihood of success. One additional approach to partially automate estimation would be to use GAM smoothers on all linear covariates (while slow) to reduce the reliance on model selection issues and make our approach easier to apply across many species.

Finally, broader application of an approach such as that used here (e.g. the NA Pops project, A. Smith, personal communication) could be applied post-hoc to the BAM data set. Having a suite of Bayesian models at national and regional scales could improve status reporting in a number of ways. First, deriving national or regional models fit to historic data sets would allow us to use parameter estimates from those models as informative priors when creating or updating status reports using data collected over a narrower time-frame. Since the use of informative priors can greatly improve model precision, parameter estimates from national models fit using an approach similar to ours could greatly improve our ability to detect changes in population status. In addition, both map product and changes in parameter estimates could also point to changes in species distribution and/or habitat selection using a single modeling framework and thus allow greater insight into factors influencing avian populations.

Literature Cited:

Barker, R. J., M. R. Schofield, W. A. Link, and J. R. Sauer. 2018. On the reliability of N‐mixture models for count data. Biometrics 74: 369-377.

Boreal Avian Modelling Project, 2020a. BAM Generalized National Models Documentation, Version 4.0. Available at https://borealbirds.github.io/. DOI: 10.5281/zenodo.4018335. [Results for Olive-sided Flycatcher (*Contopus cooperi*)]

Boreal Avian Modelling Project, 2020b. BAM Generalized National Models Documentation, Version 4.0. Available at https://borealbirds.github.io/. DOI: 10.5281/zenodo.4018335. [Results for Dark-eyed Junco (*Junco hyemalis*)]

Boreal Avian Modelling Project, 2020c. BAM Generalized National Models Documentation, Version 4.0. Available at https://borealbirds.github.io/. DOI: 10.5281/zenodo.4018335. [Results for Orange-crowned Warbler (*Oreothlypis celata*)]

Boreal Avian Modelling Project, 2020d. BAM Generalized National Models Documentation, Version 4.0. Available at https://borealbirds.github.io/. DOI: 10.5281/zenodo.4018335. [Results for Ovenbird (*Seiurus aurocapilla*)]

Boreal Avian Modelling Project, 2020e. BAM Generalized National Models Documentation, Version 4.0. Available at https://borealbirds.github.io/. DOI: 10.5281/zenodo.4018335. [Results for Blackpoll Warbler (*Setophaga striata*)]

Boreal Avian Modelling Project, 2020f. BAM Generalized National Models Documentation, Version 4.0. Available at https://borealbirds.github.io/. DOI: 10.5281/zenodo.4018335. [Results for Yellow-rumped Warbler (*Setophaga coronata*)]

Boreal Avian Modelling Project, 2020g. BAM Generalized National Models Documentation, Version 4.0. Available at https://borealbirds.github.io/. DOI: 10.5281/zenodo.4018335. [Results for Canada Warbler (*Cardellina canadensis*)]

Bayne, E. M., M. Knaggs, and P. Slymos. 2017. How to most effectively use autonomous recording units when data are processed by human listeners. Unpublished report prepared for Environment and Climate Change Canada. 78pp. <http://www.abmi.ca/home/publications/451-500/489>

Beaudoin, A., P. Y. Bernier, L. Guindon, P. Villemaire, X. J. Guo, G. Stinson, T. Bergeron, S. Magnussen, and R. J. Hall. 2014. Mapping attributes of Canada’s forests at moderate resolution through kNN and MODIS imagery. Canadian Journal of Forest Research 44: 521-532.

Cumming, S.G., Stralberg, D., Lefevre, K.L., Sólymos, P., Bayne, E.M., Fang, S., Fontaine, T., Mazerolle, D., Schmiegelow, F.K.A., Song, S.J., 2014. Climate and vegetation hierarchically structure patterns of songbird distribution in the Canadian boreal region. Ecography 37: 137–151. <https://doi.org/10.1111/j.1600-0587.2013.00299.x>

DiMiceli, C., J. Townshend, M. Carroll, and R. Sohlberg. 2021. Evolution of the representation of global vegetation by vegetation continuous fields. Remote Sensing of Environment 254: 112271.

Environment and Climate Change Canada. 2020a. ECCC Boreal Avian Monitoring Strategy.. Unpublished report. Ottawa, ON, Canada. 14pp.

Environment and Climate Change Canada. 2020b. Environment and Climate Change Canada Guidance for Collecting Data to Support Assessment of Project Effects on Birds in Canada. Unpublished report. Ottawa, ON, Canada. 92pp.

Matsuoka S.M., C. L. Mahon, C. M. Handel, P. Sólymos, E. M Bayne., P. C. Fontaine, and C. J. Ralph. 2014. Reviving common standards in point-count surveys for broad inference across studies. The Condor: Ornithological Applications 116(4): 599-608.

Royle, J. A. (2004). N‐mixture models for estimating population size from spatially replicated counts. Biometrics 60(1):108-115.

Sólymos, P. 2016. QPAD version 3 documentation. Technical Report, Boreal Avian Modelling Project, URL<http://www.borealbirds.ca/library/index.php/technical_reports>

Sólymos, P., S. M. Matsuoka, E. M. Bayne, S. R. Lele, T. Fontaine, S. G. Cumming, D. Stralberg, F. K. A. Schmiegelow, and S. J. Song. 2013. Calibrating indices of avian density from non-standardized survey data: making the most of a messy situation. Methods in Ecology and Evolution 4: 1047–1058.<https://doi.org/10.1111/2041-210X.12106>

Sólymos, P., J. D. Toms, S. M. Matsuoka, S. G. Cumming, N.K.S. Barker, W.E. Thogmartin, D. Stralberg, A. D. Crosby, F. V. Dénes, S. Haché, C. L. Mahon, F. K. A. Schmiegelow, and E. M. Bayne. 2020. Lessons learned from comparing spatially explicit models and the Partners in Flight approach to estimate population sizes of boreal birds in Alberta, Canada. The Condor: Ornithological Applications 122. <https://doi.org/10.1093/condor/duaa007>

Sörensen, R., U. Zinko, and J. Seibert. 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrology and Earth System Sciences 10: 101-112.

Sprugel, D. G. 1983. Correcting for bias in log‐transformed allometric equations. Ecology 64: 209-210.

Stralberg, D., E. M. Bayne, S. G. Cumming, P. Sólymos, S. J. Song, and F. K. A. Schmiegelow. 2015a. Conservation of future boreal forest bird communities considering lags in vegetation response to climate change: a modified refugia approach. Diversity and Distributions 21: 1112-1128.

Stralberg, D., S. M. Matsuoka, A. Hamann, E. M. Bayne, P. Sólymos, F. K. A. Schmiegelow, X. Wang, S. G. Cumming, and S. J. Song. 2015b. Projecting boreal bird responses to climate change: The signal exceeds the noise. Ecological Applications 25, 52–69. <https://doi.org/10.1890/13-2289.1>

Title, P.O., and J. B. Bemmels. 2018. ENVIREM: An expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. Ecography 41:291–307.

Van Wilgenburg, S. L., C. L. Mahon, G. Campbell, L. McLeod, M. Campbell, D. Evans, W. Easton, C. M. Francis, S. Haché, C. S. Machtans, C. Mader, R. F. Pankratz, R. Russell, A. C. Smith, P. Thomas, J. D. Toms, and J. A. Tremblay. (2020). A cost efficient spatially balanced hierarchical sampling design for monitoring boreal birds incorporating access costs and habitat stratification. PLoS ONE 15(6):e0234494. <https://doi.org/10.1371/journal.pone.0234494>

Wilson, M. F. J, B. O’Connell, C. Brown, J. C. Guinan, and A. J. Grehan. 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. Marine Geodesy 30: 3-35.

Table 1. Parameter estimates from a preliminary zero-inflated Poisson count model fit to Olive-sided Flycatcher (*Contopus cooperi*) counts from Saskatchewan and Manitoba.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | 95% Credible Interval | |
| Parameter | *β* | LCL | UCL |
| *Zero-inflation sub-model* |  |  |  |
| Intercept | 2.25 | 1.26 | 3.86 |
| Mean Annual Temperature | -1.22 | -1.79 | -0.71 |
| Mean Annual Temperature2 | -1.84 | -2.91 | -1.04 |
| Climate moisture index | 0.46 | -0.39 | 0.11 |
|  |  |  |  |
| *Count sub-model* |  |  |  |
| Intercept | -4.78 | -5.27 | -4.35 |
| Percent Needleleaf | -7.5x10-3 | -0.24 | 0.22 |
| Stand Age | -0.28 | -0.58 | 2.0x10-3 |
| Percent tree cover | 0.01 | -0.17 | 0.19 |
| Residual stand closure | -0.02 | -0.15 | 0.11 |

Figure 1. Geographic distribution of sampling using the “Boreal Optimal Sampling Strategy” design in Manitoba, Saskatchewan and Northwest Territories (2014-2019). Red dots indicate the location of 3251 secondary sampling units (point-count stations), distributed amongst 325 primary sampling units.

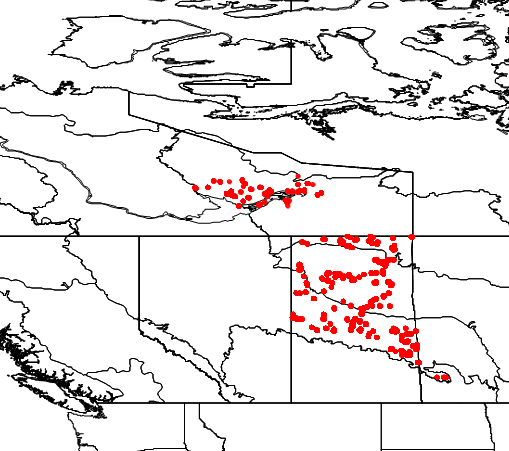


Figure 2. Mean posterior predicted density of Olive-sided Flycatcher (*Contopus cooperi*) in Saskatchewan, Canada (2014-2019).

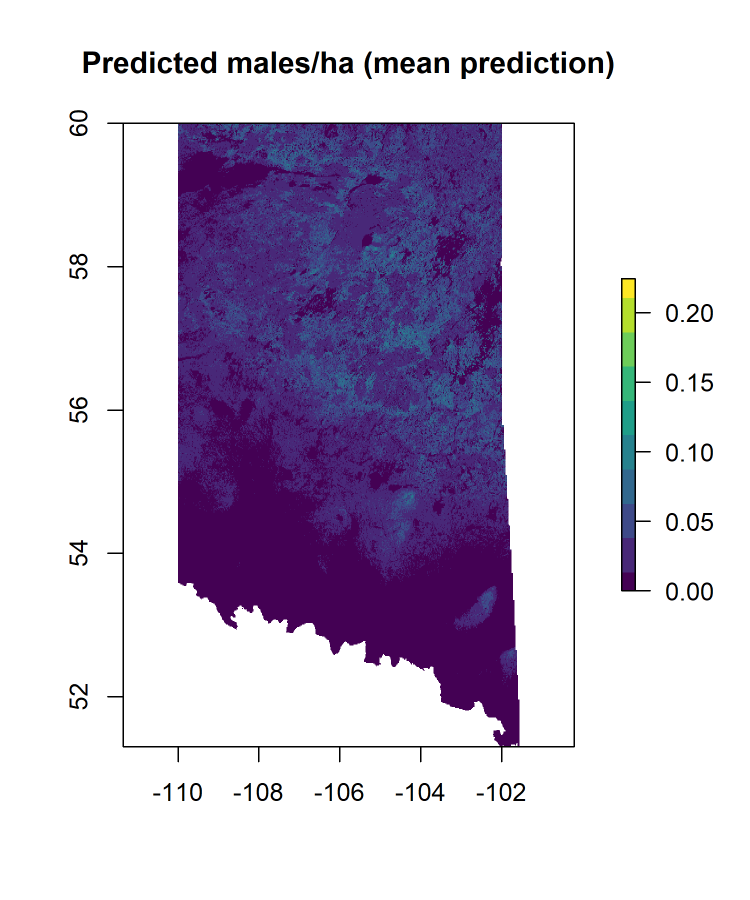


Figure 3. Distribution of posterior population size estimates for Olive-sided Flycatcher (*Contopus cooperi*) in Saskatchewan from a Bayesian species distribution model.

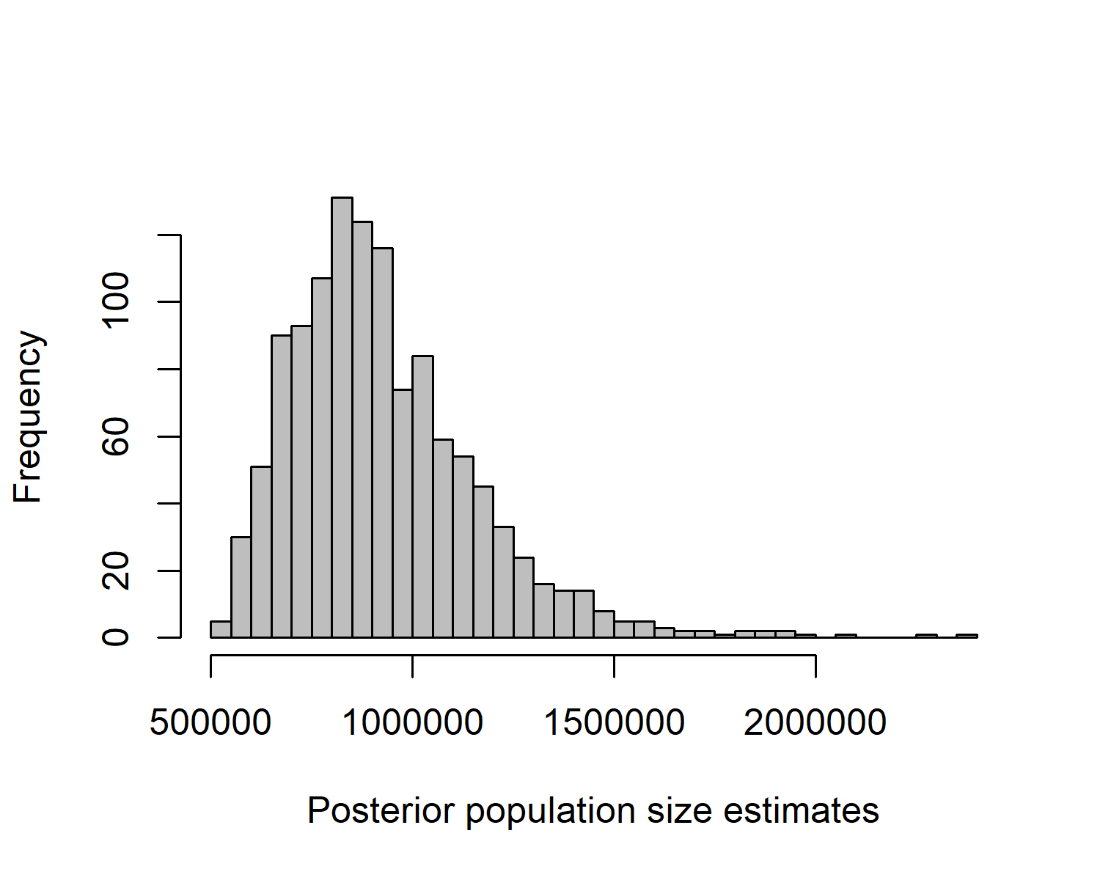


Figure 4. Coefficient of variation in predicted densities of Olive-sided Flycatcher (*Contopus cooperi*) in Saskatchewan, Canada (2014-2019).

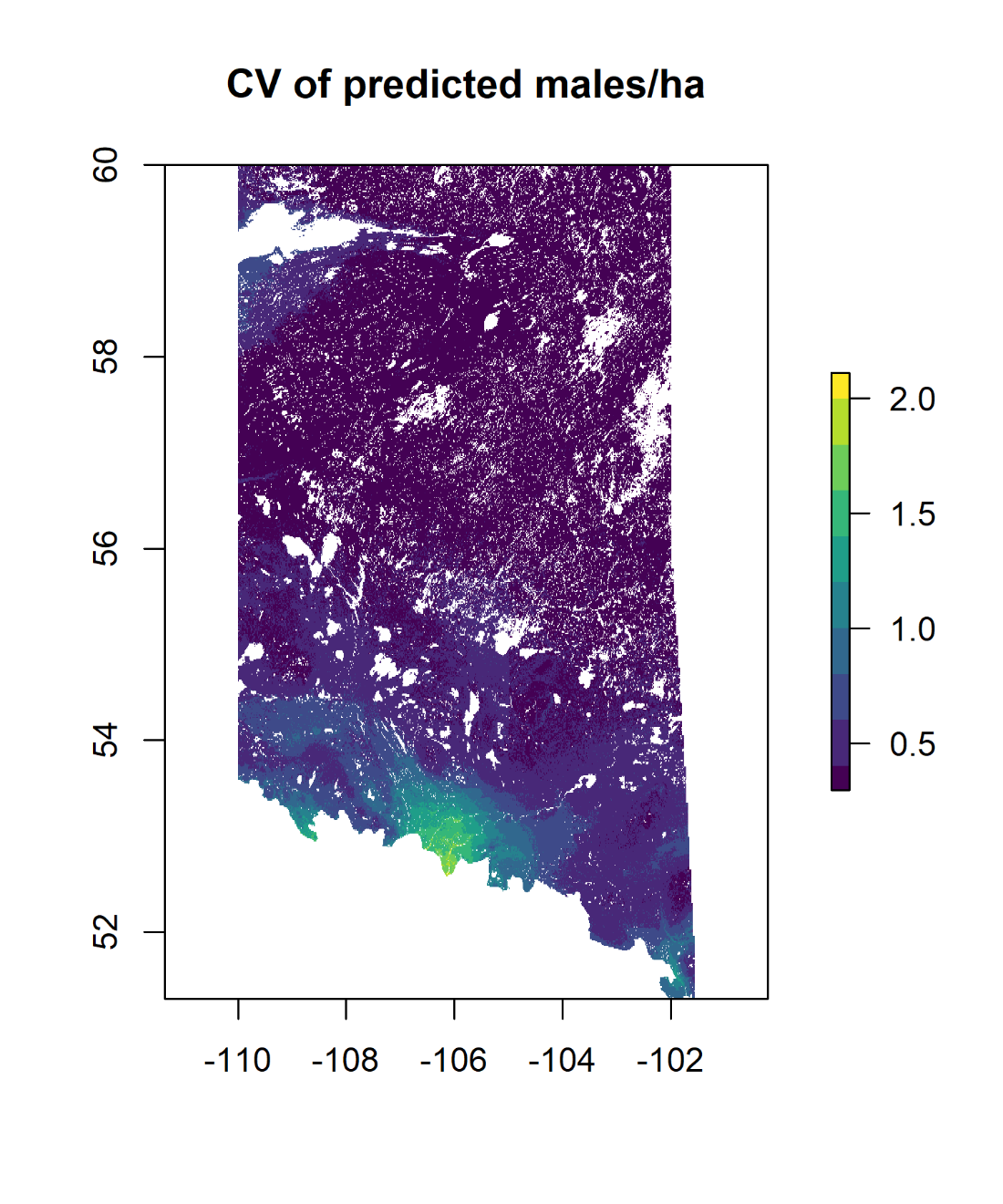


Figure 5. Performance of a pixel based population size estimation for seven species of passerine using a Bayesian zero-inflated Poisson regression model incorporating climatic, topographic and local habitat covariates (see Methods). Black line indicates 1:1 correspondence, dots and error bars represent mean and 95% Credible Intervals for population size estimates respectively.

