**Avian biodiversity mapping using acoustic monitoring and occupancy modelling**

**Sunny Tseng, Dexter Hodder, and Ken Otter**

# Abstract

# Introduction

Paragraph 1: Importance of biodiversity monitoring in large scale landscapes for evidence-based management. The challenges we have and the opportunities we have in current days (camera traps, audio recorders, and eDNA)

Paragraph 2: Advances in acoustic monitoring offers advantages of cost and spatialtemporal scale coverage such as Audio Moth, also the developed machine learning tools such as BirdNET.

Paragraph 3: Occupancy modelling is a statistical approach for estimating species distribution and habitat use. It can deal with non-perfect detection probabilities and providing insights into species occurrence.

Paragraph 4: How combining acoustics monitoring with occupancy modelling can enhance biodiversity assessments by integrating large-scale acoustic data with sophisticated statistical models.

Paragraph 5: Objectives are: 1) provide a biodiversity hotspot map based on acoustic monitoring and occupancy modelling to demonstrate how this approach can reveal critical area for conservation management and illustrate the potential of PAM in large-scale landscape management. And 2) provide a framework for combining the processing of AU data with occupancy modelling.

# Methods

## Study area, acoustic monitoring and data processing

The study was conducted in the John Prince Research Forest, located in central British Columbia, Canada, within the dry sub-boreal spruce biogeoclimatic zone. Audio data were collected from 2020 to 2022 during the breeding season (May to July) between 4:00 AM and 7:00 AM, using 66 Audio Moths (cite). Recordings were made for 1 minute every 5 minutes. Each ARU was placed at least two kilometers apart to minimize spatial correlation. Variability in the number of active ARUs at each site occurred due to setup logistics and field challenges such as battery depletion, firmware issues, or disturbances by wildlife (Fig. effort).

Collected acoustic data were analyzed using the BirdNET Analyzer (cite with GitHub repository), utilizing the Windows Setup option to run the Python module in a local environment (parameters detailed in Table parameter). To retain as many detections as possible at the first analysis stage, we set the parameter “min\_conf”, which determines the threshold for ignoring results with confidence below this level, to 0.1. This low threshold was chosen because the optimal confidence threshold varies across species. By keeping as many original detections as possible, we would have the flexibility later to apply species-specific thresholds to filter out false positives. The entire dataset, comprising 1.5 terabytes of audio, required approximately 72 consecutive hours of processing.

Table parameter. BirdNET algorithms arguments, with default value and the values used in this study. The empty cell in the used value column indicates using the default value.

| Argument | Default value | Used value |
| --- | --- | --- |
| i | None | -- |
| o | None | -- |
| lat | -1 |  |
| lon | -1 |  |
| week | -1 |  |
| slist | None |  |
| sensitivity | 1.0 |  |
| min\_conf | 0.1 |  |
| overlap | 0 |  |
| rtype | table | r |
| threads | 1 | 4 |
| batchsize | 1 | 4 |
| locale | en |  |
| sf\_thresh | 0.03 |  |
| classifier | None |  |
| fmin | 0 |  |
| fmax | 15000 |  |
| output\_file | None | -- |
| skip\_existing\_results | FALSE | TRUE |



Fig. effort. Number of active ARUs during the surveying seasons.

## Species list and validated detections

1. How did we get the total number of species based on all detections – use the general high threshold and then manual validation – provide a species list
2. How to set BirdNET species-specific threshold to reach precision higher than 0.95 for all species – provide the protocol of how we do the species-specific threshold setting

The species list was produced in four steps: (1) BirdNET detections were filtered using a confidence threshold of 0.8, and five recording segments with the highest confidence scores for each detected category were manually reviewed. Only categories with at least one confirmed vocalization were retained, resulting in 136 categories. (2) Non-bird categories, such as Car Engine, Red Squirrel, Wood Frog, and Slender Meadow Katydid, were removed, leaving 129 categories. (3) Species not listed in the British Columbia Breeding Bird Atlas (<https://www.birdatlas.bc.ca/>), which includes species recorded in the Prince George area since 2008, were excluded, leaving 123 species. (4) Species not detected from more than one site or more than one day within May to July, were excluded, resulting in a final list of 122 species (Horned Grebe was dropped in this part of filtering as it only got detected in August).

For each species, we followed the recommended methods to define species-specific thresholds for retaining reliable BirdNET detections (cite Wood & Kahl and my own study). Specifically, we employed stratified sampling to select 360 recording segments for each target species. This involved sampling 20 segments from each 0.5 confidence interval class (ranging from 0.1 to 1.0). These segments were manually reviewed through listening or spectrogram analysis to determine whether detections were true positives or false positives. We then used logistic regression with a logit link function, modeling BirdNET confidence as the predictor and detection accuracy (true/false positive) as the response variable. The confidence values were converted to probabilities of true positive detections, which were then used to compute the cumulative number of true positives. For each species, we identified the confidence threshold at which precision reached 0.95, meaning at least 95% of the cumulative detections were true positives. This process was repeated iteratively for all target species.

Table: Species list, including family, and the confidence threshold.

## Response variable: detection-nondetection

Create the weekly observation and make the species detection metrics

## Predictor variables: detection and occurrence covariates

Lidar data for the site-specific covariates, and include other observation specific covariates

## Occupancy modelling

The use of the spOccupancy package – using single or multiple species occupancy modelling

* Guilds such as cavity nesters, warblers, woodpeckers, etc.
* Break up into family group

# Result

# Discussion

# Conclusion

# Literature cited