**Audio monitoring for temporal pattern of migratory birds: using Olive-sided Flycatcher as an example**

# Outline

During breeding season, most birds show a daily period of high singing activity, which is known as “dawn chorus” (cite). It is widely known that various environmental factors influence the dawn chorus start time, such as ambient temperature, precipitation, cloud cover, lunar phase, and existence of other species (cite). The rapid development of autonomous recording units and machine learning algorithms had notably reduced the difficulty in monitoring dawn chorus. Studies had been done to investigate the relationships between environmental factors and the dawn chorus start time; however, bird species in North American have received little attention (cite). In this study, the relationships between dawn chorus start time and ambient temperature, precipitation, cloud cover, lunar phase, and site biodiversity will be investigated for Olive-sided Flycatcher (*Contopus cooperi*), whose status is under special concern in Canada (cite). This research will inform the effects of environmental factors on the dawn chorus start time of Olive-sided Flycatcher, not only providing a baseline information for the species but also setting up a standard framework for future dawn chorus studies.

# Objectives

* Find the monthly pattern of Olive-sided Flycatcher vocal density by cumulative detections
* Determine the factors that related to the start time of dawn chorus from OSFL

Related papers:

* A global assessment of BirdNET performance: differences among continents, biomes, and species
* Using data from camera traps and autonomous recording units to evaluate and improve species-habitat inferences
* Diel and seasonal vocal activity patterns revealed by passive acoustic monitoring suggest expert recommendations for breeding bird surveys need adjustment
* Phenological mismatch between breeding birds and their surveyors and implications for estimating population trends

Potential source to compare the OSFL trends:

Idea – get the Canada wide trends from these three sources and make an comparison?

* Trend from eBird data across Canada, and BC: <https://science.ebird.org/en/status-and-trends/species/olsfly/trends-map?week=1>
* Trend from various resources in Canada, produced by Nature Counts: <https://naturecounts.ca/nc/socb-epoc/species.jsp?sp=olsfly#status-and-trends>
* Trends from breeding birds survey across Canada, and BC: <https://bbsbayes.github.io/bbsBayes2/articles/bbsBayes2.html>

Something else:

BBS protocol: <https://www.canada.ca/en/environment-climate-change/services/bird-surveys/landbird/north-american-breeding/instructions.html#toc1>

# Method and materials

## Study area

The study was conducted in the John Prince Research Forest (~150 km2 in area), located in central British Columbia, Canada, within the dry sub-boreal spruce biogeoclimatic zone (Fig.1A). Audio data were collected from 2020 to 2022 during the breeding season (May to July) between 4am and 7am, 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/date occurred due to setup logistics and field challenges such as battery depletion, firmware issues, or disturbances by wildlife (Fig.1B).

The audio data was collected in John Prince Research Forest (54° 27'N, 124° 10'W, 700 m a.s.l) in 2020 to 2022 breeding season. A total of 44 recorders (AudioMoth; Open Acoustic Devices, 2020) were evenly distributed across the region (fig. of a map). Adjacent recorders were placed at least 2 km apart to ensue independent sampling. All recorders were under an identical recording schedule, repeating daily from four am to seven am, one minute on, followed by four minutes off. Recorders were deployed in the field beginning on X April 2021 (mean deployment date X April 2021; range X April 2021 – X May 2021; X±Y recorded/recorder). This resulted in 67,301 one-minute recordings collected. All recordings were formatted into a 48 kHz sampling rate and the mono pulse code modulation WAV.

## Audio data

Process through BirdNET, filter the detections by the developed threshold, based on previous publication.

Collected acoustic data were analyzed using the BirdNET Analyzer v2.4 model (cite with GitHub repository), utilizing the Windows Setup option to run the Python module in a local environment (parameters detailed in Table 1). 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.

For each species, we defined species-specific thresholds for retaining reliable BirdNET detections following the methods recommended by Wood and Kahl (2024) and Tseng et al. (2024). Stratified sampling was used to select 360 recording segments per species, with 20 segments sampled from each 0.5 confidence interval class (ranging from 0.1 to 1.0). Each segment was manually reviewed via listening or spectrogram analysis to classify detections as true or false positives. Logistic regression with a logit link function was applied, modeling BirdNET confidence scores as the predictor and detection accuracy as the response. A threshold achieving a precision of 0.95, indicating that at least 95% of remaining detections were true positives, was then identified for each species. See Tseng et al. 2024 for detailed method. This process was repeated iteratively across all target species.

Each site with at least one detection of OSFL has varied days of OSFL presence, ranging from 1 to 54 days, mean days is 6.84 +- 10.89 days (Fig.).

A screenshot of a computer

AI-generated content may be incorrect.

Fig. XYZ the occupancy of OSFL in sites and date. Sites were ranked by number of total OSFL detections from low (top) to high (bottom).

## Weather and environmental covariates

From ECCC Historical data,

***Modelling for temporal pattern***

Generalized additive model with polynomial terms. Try to see whether the temporal variation change due to the weather.

Qualified ARU site - at least 2 consecutive days of detection” as evidence of recurring presence during that period. This filtering was done to only use sites with at least two consective days of detection OSFL. This results in 9, 10, and 13 sites in 2020, 2021, and 2022, respectively. All the modelling and exploratory were done using data from these sites in according years. The seasonal activity can be captured roughly by the proportion of these qualified ARUs that with OSFL detections (Fig.XYZ).

“Model the number of detections as a negative binomial count process.  
We used the GAM model given the expected pattern of natural cycle of breeding activity. Using the qualified ARUs (Fig XYZ), we extracted the daily detection as the response variable (Fig XYZ) and the Julian day as response variable. We further added year and site to account for the random effect. We use the XYZ::XYZ() function in R, to fit the GAM model, The final model in R looks like this:

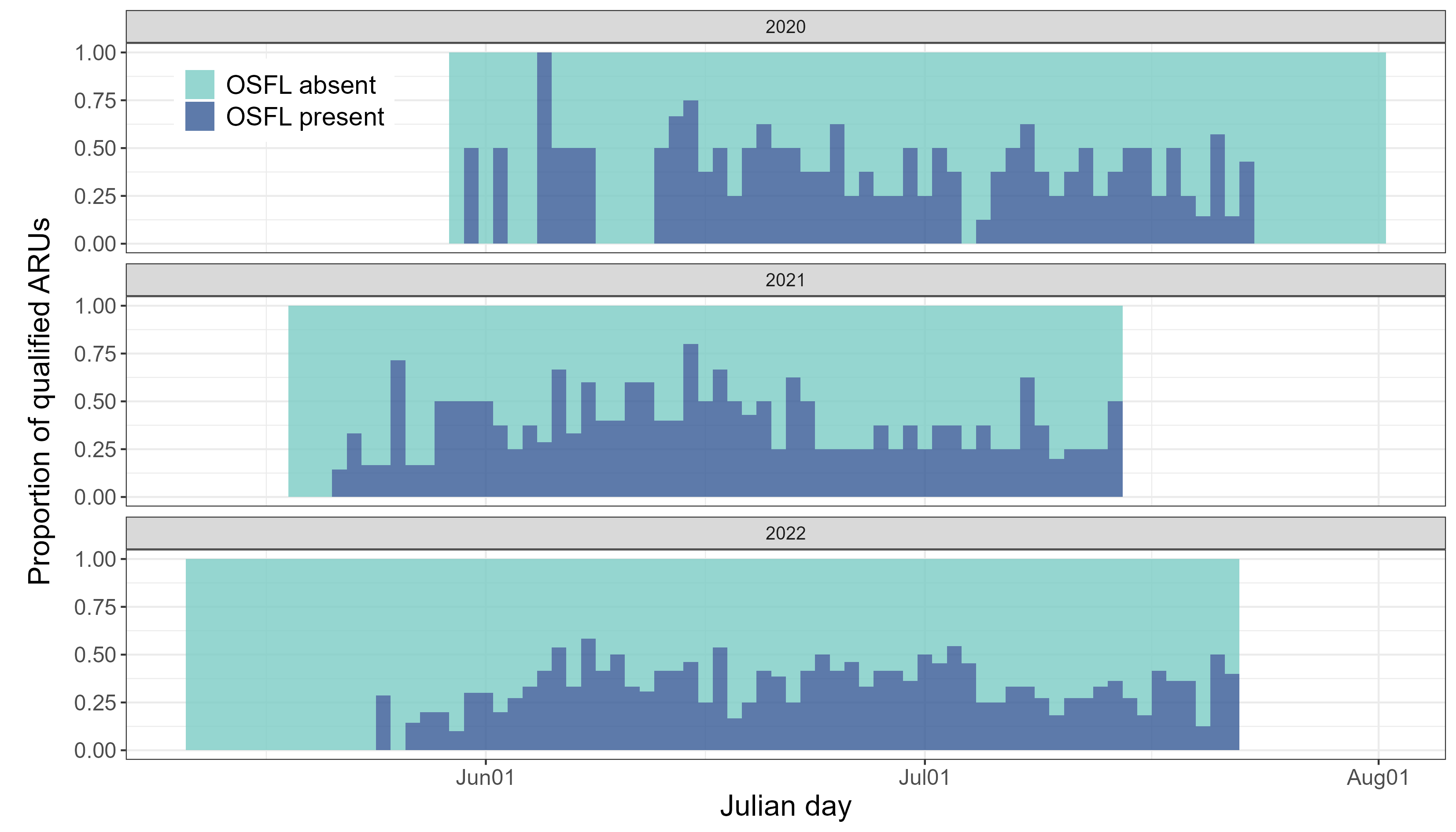


Fig.XYZ The activity of qualified ARUs.

The expected number of detections changes smoothly over the day of year,  
with a separate seasonal pattern for each year,  
while accounting for random differences in baseline detection rates across sites.

***Modelling for spatial pattern***

We used all data available from our study site (Fig. XYZ) without filtering. We selected the use of occupancy modelling given their assumptions: XYZ (?)

Occupancy modelling with LiDAR covariates. Try to identify whether the spatial variation change due to the environmental variation.

Use data from all sites (no filtering out low detection data), but need to use the result from the temporal pattern to identify the breeding season (?), or use temporal covariate to account for the temporal variation.Use grouping to get the detection matrix.

# Results

## Temporal pattern

The GAM explained 35.2% of the deviance in daily detection counts, indicating that the model effectively captured broad seasonal patterns in vocal activity, though substantial day-to-day variability remained.”

The generalized additive model (GAM) explained 35.2% of the deviance in daily detection counts, indicating that it captured substantial seasonal variation in vocal activity across years and sites. The smooth terms for day of year were highly significant in all years (p < 0.001), suggesting clear temporal patterns in detections during the breeding season, while the random effect for site accounted for additional spatial variability.

## Spatial pattern

Discussion

For temporal pattern

 Because OSFL typically raise only **one** brood, a single clear breeding peak is expected. However, if many first attempts fail and pairs renest, you can get a **secondary peak** in vocal/activity rates ~3–6 weeks after the first peak (incubation ≈15–19 d + nestling ≈15–19 d; renesting and detection timing add more lag). [All About Birds+1](https://www.allaboutbirds.org/guide/Olive-sided_Flycatcher/lifehistory?utm_source=chatgpt.com)

 Renesting is explicitly noted in multiple regional reports (e.g., COSEWIC / SARA) — they emphasize one brood raised per season but frequent re-nesting after failure.

 Check **timing**: compute the lag between the two peaks. If it’s roughly 30–45 days, renesting or the fledging period is plausible (incubation + nestling).

 Check **site-level patterns**: plot seasonal curves per site (or a heatmap of detections by site × yday). If different sites peak at different times, pooled data will look bimodal.

 Check **per-year patterns**: are both peaks present in each year or only in some years?

 Look at **call types** (if your BirdNET labels call/song types) — are peaks driven by the same vocalization type?

 Compare to environmental covariates — insect emergence indices, temperature, or heavy rain windows that might suppress or shift calling.

For spatial pattern