**Autonomous recording units can identify spatiotemporal bird activity patterns for better citizen science survey protocols**

# Abstract

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.

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

## Target species

The target species was Olive-sided Flycatcher (OSFL)

* Status in IUCN and COSEWIC, and others
* Breeding behaviour, nesting timing, double peak of the vocal activity
* The contradicted results from eBird, breeding birds survey
* Maybe we could get the surveying raw data from the breeding birds survey in PG area?

## Study area and audio data collection

The study was conducted in the John Prince Research Forest, covering approximately 150 km², located in central British Columbia, Canada (54°27'N, 124°10'W; 700 m a.s.l.) within the dry sub-boreal spruce biogeoclimatic zone. Acoustic data were collected using 66 AudioMoth (cite), each deployed at least 2 km apart to minimize spatial autocorrelation. Acoustic data were collected during the breeding seasons from 2020 to 2022 (May–July), between 4am and 7am. Recordings were scheduled for 1-minute recording intervals followed by 4 minutes of inactivity, resulting in 12 recordings per hour and 36 recordings per day. In total, 67,301 one-minute recordings were obtained. All recordings were standardized to a 48 kHz sampling rate and stored as mono Pulse Code Modulation (PCM) WAV files.

## Audio data processing

Collected acoustic data were analyzed using the BirdNET Analyzer v2.4 model (cite GitHub repository), implemented via the Python module running in a local terminal (parameters detailed in Table 1in Tseng et al. (2026). To retain as many true positive detections as possible in the initial analysis stage, we set the parameter min\_conf, which filters out results below a specified confidence threshold, to 0.1. This low initial threshold allowed us to later apply the species-specific threshold to minimize false positives. Processing the entire dataset, which comprised 1.5 terabytes of audio, required approximately 72 consecutive hours.

The species-specific threshold for Olive-sided Flycatcher (OSFL) was applied to ensure a precision of 0.95. Based on Tseng et al. (2025), who defined BirdNET thresholds for common species in the same dataset and study area, we used a threshold of 0.35 for OSFL. Refer to Tseng et al. (2025) for a detailed framework on how BirdNET species-specific thresholds are defined. Daily OSFL detections were summarized for each site throughout the survey period. At sites with at least one detection, the number of active days varied from 1 to 54, with a mean of 6.84 ± 10.89 days (Fig. full\_aru).

A screenshot of a computer

AI-generated content may be incorrect.

Fig. full\_aru. Daily detections of Olive-sided Flycatcher (OSFL) across sites. Sites are ranked by total OSFL detections, from lowest (top) to highest (bottom). Grid cells with a color gradient indicate days with OSFL detections, while grey cells represent days when the ARU was active but no OSFL detections. Variation in ARU activity periods at each site (grey and colored areas) reflects logistical constraints and field challenges, including battery depletion, firmware issues, and wildlife disturbances.

## Weather and environmental covariates

From ECCC Historical data,

## Modelling for temporal pattern

To identify the vocal activity pattern of Olive-sided Flycatchers (OSFL), we further refined the dataset (after applying the species-specific threshold) by retaining only ARUs (site-year combinations) with at least two consecutive days of OSFL detections (hereafter referred to as “qualified ARUs”). This criterion was applied to exclude ARUs with very few or sporadic detections (Fig. full\_aru) that could bias the modeling of vocal activity, as those with only a single detection in a given site-year may represent opportunistic events rather than true activity patterns. This filtering resulted in 9, 10, and 13 qualified ARUs in 2020, 2021, and 2022, respectively. All vocal activity modeling was conducted using data from these qualified ARUs for the corresponding years, which revealed a higher proportion of active detections near the middle of the breeding season (Fig. aru\_proportion).

A screenshot of a graph

AI-generated content may be incorrect.

Fig. aru\_proportion. Seasonal activity of qualified ARUs, defined as ARUs with at least two consecutive days of Olive-sided Flycatcher (OSFL) detections within a site-year. Qualified ARUs with at least one OSFL detection that day are shown in dark blue, and those without detections are shown in light blue. The proportion of qualified ARUs with OSFL presence increased gradually toward the middle of the breeding season and declined toward the end of the survey period.

We used a Generalized Additive Model (GAM), which is well suited for modeling nonlinear and potentially cyclic patterns, such as seasonal variation across the breeding period. Using the qualified ARUs (Fig. aru\_proportion), daily detection counts were extracted as the response variable, with Julian day as the predictor. To account for repeated measures and site-specific variation, we included site as a random effect, while allowing the seasonal pattern (smooth function of Julian day) to vary by year. This model structure assumes that the expected number of detections changes smoothly over the breeding season, with distinct seasonal patterns for each year, while accounting for random differences in baseline detection rates among sites. The number of detections was modeled as a negative binomial count process to accommodate overdispersion in the data. Models were fitted using the mgcv::gam() function in R (Wood, 2017), with restricted maximum likelihood (REML) estimation. The final model formulations were:

model\_year <- gam(detections ~ s(yday, bs = "cc", by = year) +

s(site, bs = "re"),

family = nb(),

data = data,

method = "REML")

model\_general <- gam(detections ~ s(yday, bs = "cc") +

s(site, bs = "re"),

family = nb(),

data = data,

method = "REML")

These models could generate smooth seasonal activity curves for each year individually (model\_year) and an overall curve representing the general seasonal pattern across all sites (model\_general).

## 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 negative binomial GAM revealed clear seasonal patterns in OSFL vocal activity (Fig. XYZ). Smooth terms for Julian day were all highly significant, showing the strong seasonal patterns of OSFL detections. Random site effects were also significant, indicating considerable variation in baseline detection rates among sites. The model explained 35.2% of the deviance, capturing the overall temporal pattern of OSFL vocal activity across the three years.

## 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

# References