# National-scale acoustic monitoring of avian biodiversity and phenology

I. Avery Bick[1,2*], Vegar Bakkestuen[1], Benjamin Cretois[1], Ben Hillier[4], John A. Kålås[1], Marius Pedersen[2], Kiran Raja[2], Carolyn Rosten[1], Marius Somveille[4], Bård G. Stokke[1], Julia Wiel[1], Sarab S. Sethi[1,3*]

[1]Norwegian Institute for Nature Research, Trondheim, Norway
[2]Norwegian University of Science and Technology, Gjøvik, Norway
[3]Department of Life Sciences, Imperial College London, London, UK
[4]Division of Biosciences, University College London, London, UK
*Corresponding authors

## Abstract

Birds migrate over large spatial scales with complex dynamics which play out over extended time periods, making monitoring of phenology challenging with traditional biodiversity survey approaches. In this study, over a complete spring season, we collected 37,429 hours of audio from 28 networked sensors in forests across the latitudinal extent of Norway to demonstrate how acoustic monitoring can transform avian phenology monitoring. We used machine learning to automatically detect and identify bird vocalizations, and with expert validation found we were able to classify 55 species (14 full migrants) with over 80% precision. We compared audio data to existing avian biodiversity datasets and demonstrated that acoustic surveys could fill large data gaps and improve the temporal resolution at which metrics such as date of arrival for individual species could be estimated. Finally, we combined acoustic data with ecoclimatic variables from satellites and were able to map migratory waves of 10 species across the country at fine spatial resolutions (0.2 degrees). Our study demonstrates how acoustic monitoring can inexpensively and reliably complement existing national-scale biodiversity datasets, delivering high quality data which can support the design and implementation of effective policy and conservation measures.

## Main

Around the world, climate change, habitat loss, and pollution are having severe impacts on biodiversity[1], an integral aspect to each of the 17 United Nations' Sustainable Development Goals (SDGs)[2]. Monitoring species compositions, phenology, and their response to stressors and interventions is critical for informing

and evaluating conservation and land management strategies[3,4]. While the norm is to use manual surveys to acquire detailed and reliable biodiversity data, these tend to be slow and costly, leading to biases or spatiotemporal gaps that can lead to poorly informed decision making[5].

Automated approaches are increasingly being used to complement manual surveys by improving the scale and resolution of biodiversity monitoring possible[6]. Passive acoustic monitoring (PAM) has shown particular promise due to the approach's taxonomic breadth[7] and scalability for both data analysis and collection[6]. Automated analysis of audio is becoming commonplace, supported by the growth of powerful machine learning (ML) models[8]. ML algorithms such as BirdNET[9] can accurately identify hundreds of species from recordings and anthropogenic noise pollution, while others are used to detect biodiversity pressures such as noise pollution[10] or signs of illegal hunting, such as gunshots[11]. Autonomous acoustic data collection via cellular network connection and self-powered units (via solar power or connection to a grid) is also growing[12,13], but few studies have put data collection and analysis together to deliver biodiversity insight at scales and resolutions that would not be possible with traditional methods.

The study of avian species phenology is particularly well-suited for bioacoustic monitoring. Many birds are highly vocal and their dynamics and distributions extend over scales and resolutions that are not easily captured by traditional surveys. Furthermore, due to large existing vocalization libraries, off-the-shelf bird vocalization detection ML models cover thousands of species and are increasingly reliable[9,14]. Improved monitoring of avian phenology could help us understand how stressors such as noise pollution[15], habitat loss[16], and climate change (e.g., "spring mismatch"[3,17]) impact avian migration routes, behaviors, and breeding outcomes at relevant scales, leading to improved design and implementation of conservation and policy actions.

We currently rely on citizen science to track avian phenology; either through coordinated national breeding bird surveys, or through data from apps such as eBird, iNaturalist, and Artsobservasjoner, which allow the public to record observations opportunistically. Breeding bird surveys are typically performed by knowledgeable volunteers[18] and capture a snapshot of bird populations in the late-Spring and early-Summer across thousands of sites in the northern hemisphere. While these surveys are a thorough and crucial means of estimating population trends and ranges for many species, they cannot be effectively used to study migratory timing and phenology, such as shifts in Spring migratory bird arrival timing with global warming[1,3]. Conversely, datasets which contain opportunistic encounters (e.g., eBird) provide some insight into avian phenology timing[19] but suffer from temporal and spatial biases[20] and a high variation in data reliability[21]. PAM can complement both standardized and opportunistic avian phenology

monitoring by plugging both spatial and temporal gaps with data that can be easily compared across taxa and sampling sites over extended and cost-efficient deployments.

In this study, we perform one of the first national-scale acoustic surveys of avian phenology using autonomous acoustic recorders installed across almost the entire latitudinal extent of Norway. We demonstrate that PAM can measure phenology at high resolution across large scales, that detections of vocalizations from a machine learning model are reliable for many species through expert validation, and that PAM provides complementary biodiversity data to existing surveys. We also show that species distribution models based on acoustic data enable generation of national-scale predictions of avian phenology at fine spatial resolution.

## PAM is reliable and fills data gaps in existing surveys

We installed networked recording devices powered by solar panels to survey acoustic data on large temporal and spatial scales across Norway, from Kristiansand on the southern tip of Norway, to Finnmark in the far north, approximately 1,700 km apart (Fig. 1a). Between April 1 and June 30, 2022, we collected 37,429 hours of audio data from 28 sites in eight regional clusters, with an average 75% recorder uptime. Having redundant recorders in each region provided pseudo-replicate data and covered for periods of power loss or recorder failure (see sites Bergen 3 and Kristiansand 1 in Fig. 1b,c).

Recordings were saved locally in 5-minute files, then uploaded to a cloud server via mobile internet, where BirdNET was used to detect and classify bird vocalizations. We provided 50 detections to an expert ornithologist for each species that BirdNET identified with a classification score per species of over 0.80[22], finding that the model correctly identified 55 species (14 full migrants) with greater than 80% precision, 30 of these with 100% precision (Fig. 1d). From the vocalizations of the 14 reliably detectable migrant species, we were able to observe increasing species richness through the spring months (Fig. 1c)

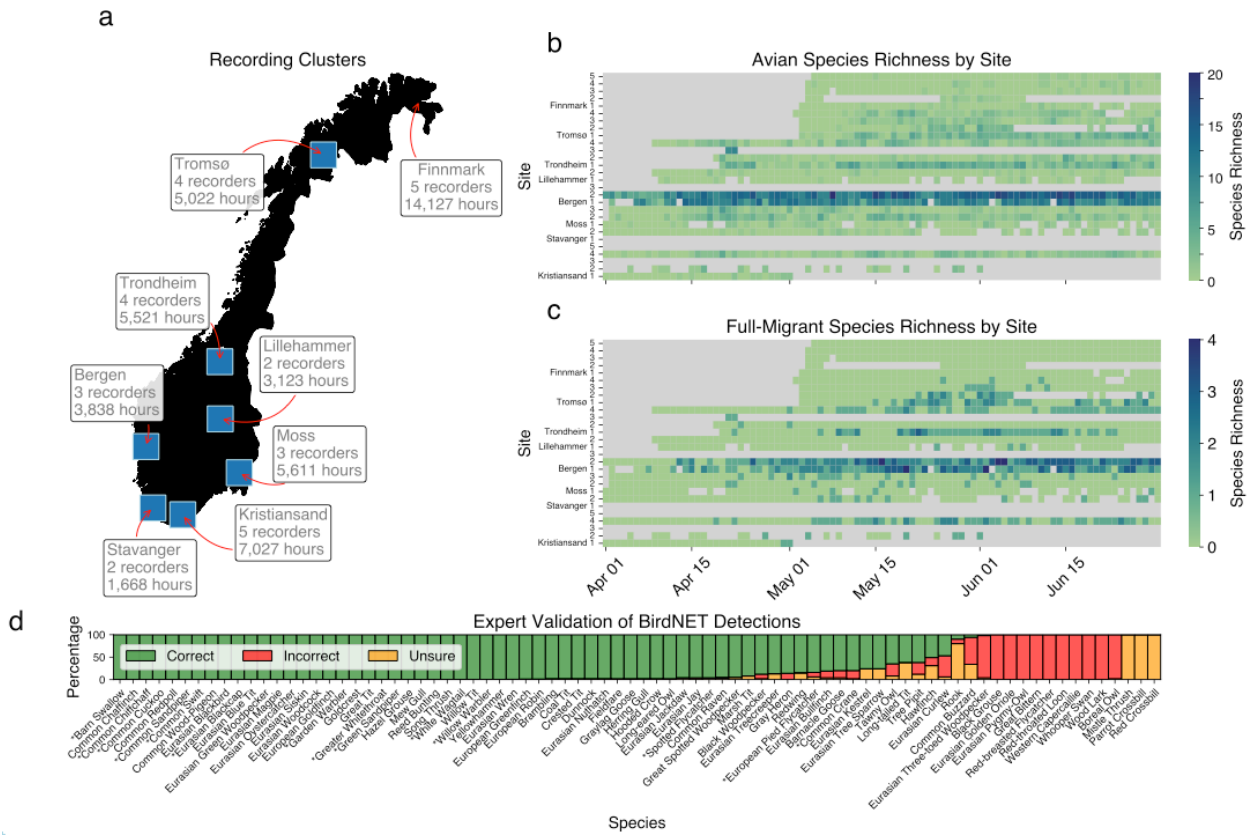


**Fig. 1 | Passive acoustic monitoring captures avian species richness during migratory arrival and is highly accurate in detecting dozens of species. a,** 28 passive acoustic monitors deployed across Norway in 2022 are grouped into eight regions. Blue boxes (90 km x 90 km) at the centroid of most region's recorders are used to mask manual bird survey data for comparison to audio detections, and to spatially average daily NDVI data. Light blue crosses represent coordinates of recorders. **b,** Species richness of 55 avian species and **c,** 14 full-migrant species vary across Norway temporally and latitudinally. Passive acoustic monitoring can provide near-continuous monitoring over long periods (mean 75% recorder uptime across our study period of April 1 to June 30, 2022), but redundancy is key as equipment in the field may fail (sites Bergen 3 and Kristiansand 1) or solar panels may receive insufficient sunlight. **d,** Expert ornithologist validation of BirdNET species identification performance for all species with ≥50 detections ≥0.80 model confidence shows that 41 species are detected with ≥80% precision across 50 randomly sampled detections. Full migrant species are identified by a "*" before their common name. Days where recorder uptime was less than 50% are shown as gray.

We then focused on three common migrating passerines detected with 100% precision which had enough data for downstream analysis (Fig. 1d); Willow Warbler (WW) (*Phylloscopus trochilus*), Common Chiffchaff (CC) (*Phylloscopus collybita*), and Spotted Flycatcher (SF) (*Muscicapa striata*)). We compared the survey effort and sensitivity of PAM with two comprehensive avian occurrence datasets available for Norway: opportunistic encounters recorded on eBird[23], and standardized line transect point counts from Norsk Hekkefugloverváking, also known as the Norwegian Breeding Bird Survey (BBS)[18].

PAM delivered a far greater number of total species observations than manual surveys for WW across the study period ($N_{PAM}$ = 687, $N_{eBird}$ = 102, $N_{BBS}$ = 135), CC ($N_{PAM}$ = 300, $N_{eBird}$ = 48, $N_{BBS}$ = 63), and SF ($N_{PAM}$ = 192, $N_{eBird}$ = 60, $N_{BBS}$ = 45), primarily driven by the large number of days when acoustic data was recorded but no eBird or BBS survey were performed (light blue, Fig. 2). However, on days where both BBS and PAM data was available, BBS (red, Fig. 2c, 2f, 2i) generally had a higher probability of uniquely detecting these species than PAM (dark blue, Fig. 2c, 2f, 2i), with 42 unique daily detections across the three species compared to 13 for PAM. Similarly, eBird also exhibited a higher detection sensitivity than PAM on days where both surveys were conducted, with 21 to 14 unique daily detections, respectively (Fig. 2b, 2e, 2h). With two to five recorders per 90 km x 90 km region, PAM captures more overall observations while providing gap-filling and phenological context for manual surveys.

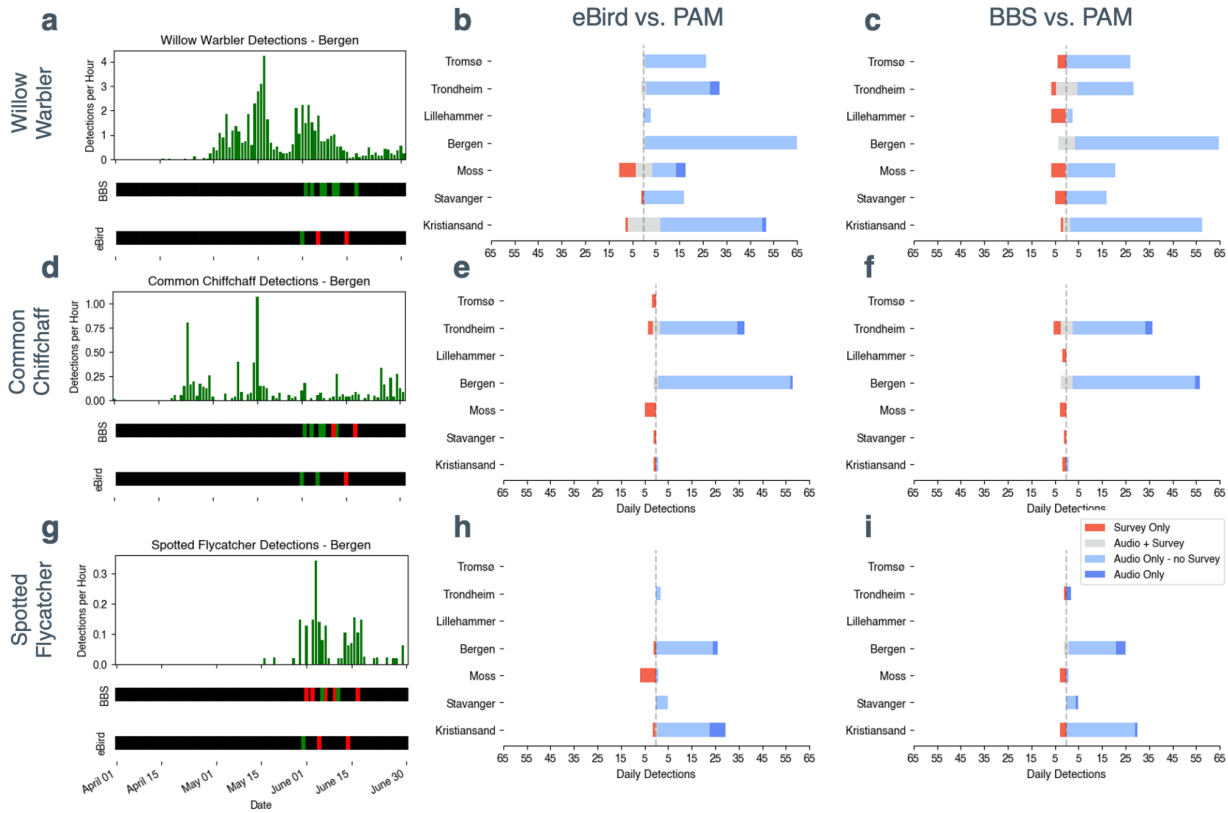


**Fig. 2 | For three common migrating passerines, passive acoustic monitoring data can fill gaps and improve temporal resolution of sampling compared to existing biodiversity datasets, but on days where manual surveys and PAM were conducted, manual surveys had more unique detections. a, d, g,** Audio recorders in Bergen capture temporal patterns in species arrival and vocalization for three migrating passerine species (green bars). Detections (green lines) and non-detections (red lines) from manual surveys in this region fail to capture avian migratory dynamics at the same resolution. PAM captures more total detections of species of interest than **b,e,h,** eBird and **c,f,i,** BBS. However, on days where both audio surveys and manual surveys were performed, both eBird and BBS were more sensitive for detecting each species.

## PAM enables avian phenology monitoring at a national scale

Whilst we saw that existing avian datasets were limited in their temporal coverage and resolution (Fig. 1), with continuous PAM deployments we were able to measure fine-scale migratory dynamics. In Bergen, we estimated that CC, WW, and SF arrived on April 23, May 4, and May 27, respectively (Fig. 3d, Methods). This order of establishment agrees with intuition based on CC wintering further north than WW and SF, which are long-distance migrants, as well as these species' respective food sources for their chicks: primarily crawling insects and larvae for WW and CC, but flying insects for SF, which emerge later in the Spring season. Based on arrival times in Kristiansand to Tromsø, WW traveled northward at approximately 67 km/day, aligning with previous measurements of WW southward autumn migratory rate of 40.6-80.5 km/day using ringed birds[24]. Intuitively, we found that at higher latitudes WW arrives at a lower greenness, as measured by satellite-derived Normalized Difference Vegetation Index (NDVI)[25] (Spearman $\rho=-0.83$, $p=0.04$). As migrating birds must contend with short summers in northern regions, WW must balance arrival times to avoid bad early season weather, but also to match peak food abundance to when chicks are present in their nests[26]. Climatic changes in wintering areas in Africa have driven WW to arrive approximately half a day earlier to the Tromsø region each year since 1995[26], potentially threatening this phenological timing. Comparing multi-year PAM data with complementary datasets (e.g., measuring insect emergence) could shed light on the full effect of seasonal shifts in a way that is not tractable currently.

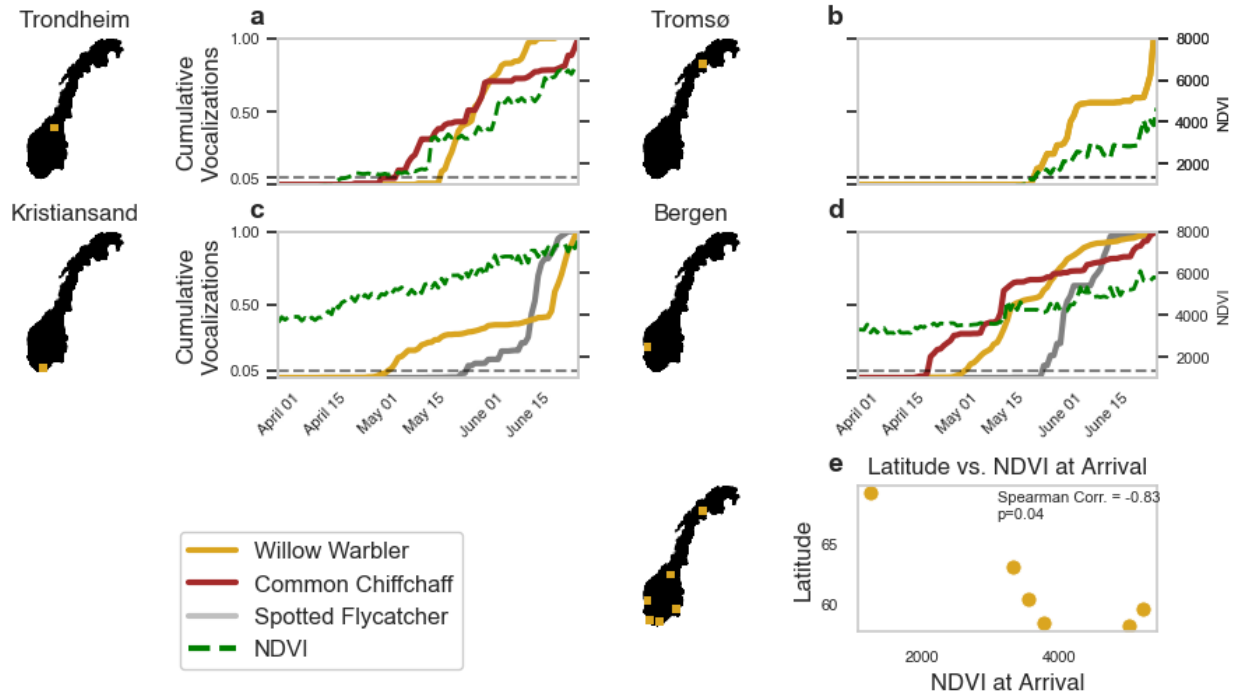


**Fig. 3 | Cumulative audio detections of migrating passerines demonstrate interspecies and large-scale variation in migration timing across Norway. a, b, c, d,** Passive acoustic monitoring shows ecologically consistent species arrival patterns across Norway. Spotted Flycatcher, with a diet of flying insects, arrives later than both Willow Warbler and Common Chiffchaff, whose diets include primarily larvae and crawling insects. **e,** Willow Warbler arriving at higher latitudes arrive at a much lower greenness index than those at lower latitudes (Spearman ρ=-0.83, p=0.04), illustrating a potential tradeoff between lower competition in northern regions, but also lower food availability.

For ten full-migrant species detected reliably and with at least 10 weekly detections across the study period, we hyperparameter-tuned and trained an acoustic species distribution model (aSDM)[27] on weekly audio detection data and ecoclimatic covariates to generate weekly predictions of species distributions across the whole of Norway. The aSDMs allowed us to predict occurrences beyond our sampling sites and observe migratory waves throughout the country's forests[28] at an 0.2 degree resolution (Fig. 4). With the national-scale aSDMs, we observed WW's earlier arrival and higher relative abundance than SF across Norway, as demonstrated by their vocalization probabilities (Fig. 4a,b).

Testing the agreement of the aSDMs with eBird and BBS survey data in forested regions, we generated distributions of aSDM values at presences and absences from each survey (Fig. 4d,e). The aSDMs agree well with eBird observations by permutation test (Fig. 4d) of WW ($p \leq 0.001$, $t=0.105$) and SF ($p=0.024$, $t=0.0634$), and with BBS observations for SF ($p=0.004$, $t=0.0367$), though not with BBS observations for WW (Fig. 4d), perhaps due to the limited temporal range of BBS in late May and June, during which WW are ubiquitous across Norway (Fig. 4a). The general agreement of the aSDMs with citizen-science based

datasets illustrates the complementary nature of PAM with survey data, as do the variable sensitivities and survey efforts of each method (Fig. 2). Additionally, the vocalization probabilities capture a metric of species detectability[21,29], offering a method of informing the timing of both breeding bird surveys and citizen scientists.

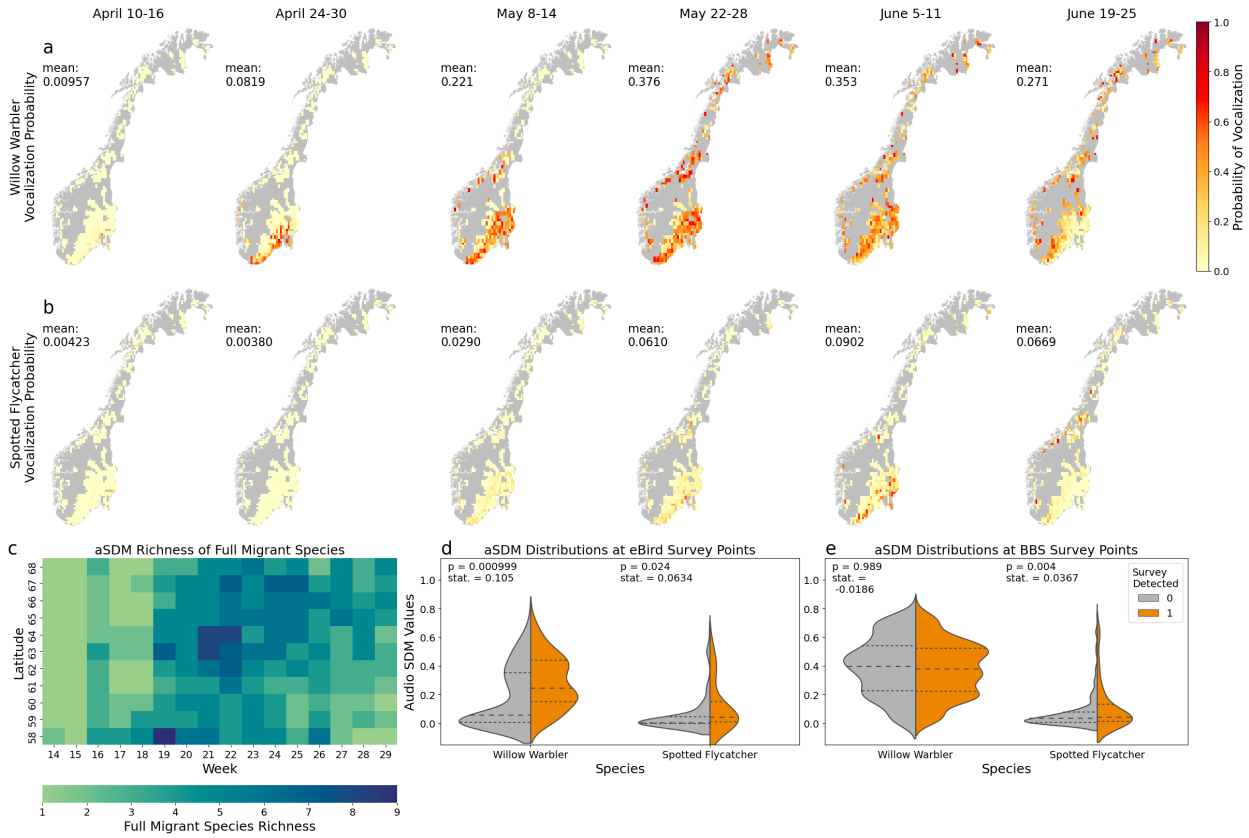


**Fig. 4 | Combining audio vocalization detections with ecoclimatic variables allowed us to train audio species distribution models (aSDMs) to predict vocalization probability as an analogue of species occurrence across Norway.** Weekly vocalization probability in forested areas across Norway for **a,** Willow Warbler and **b,** Spotted Flycatcher shows spatiotemporal species arrival patterns and captures later arrival of Spotted Flycatcher relative to Willow Warbler. **c,** Weekly species richness of 10 migrating species across latitudinal bands**.** Species are considered present in each band if mean aSDM values are ≥ 0.05. **d,** At eBird survey points, the distribution of aSDM values are significantly greater when the surveyor detected WW and SF versus when the eBird surveyor did not detect the species. **e,** For BBS survey points, the distribution of aSDM values is significantly greater when the surveyor detected SF, but not for WW.

# Discussion

Our results show that national-scale PAM deployments can deliver a step change in both the resolution and scale at which we can track avian phenology when compared to existing surveys at comparable scales. Pairing networked sensors and automated analyses with one-off expert validation allowed us to track phenology reliably and in near real-time, whilst surveys required minimal cost and effort to deploy

and maintain. As sensors become cheaper, more sensitive, and more reliable[11] and machine learning models improve in accuracy, taxonomic breadth, and computational efficiency[8], national-scale PAM surveys will become more widely accessible and will be extended for monitoring phenology of non-avian species (e.g., insects, mammals). Ultimately, PAM has the potential to transform our understanding of species phenology and can pave the way for effective data-driven conservation and policy decision making, leading to improved outcomes for biodiversity globally.

# References


1. Saino, N. *et al.* Climate warming, ecological mismatch at arrival and population decline in migratory birds. *Proc. R. Soc. B Biol. Sci.* **278**, 835–842 (2011).
2. Blicharska, M. *et al.* Biodiversity's contributions to sustainable development. *Nat. Sustain.* **2**, 1083–1093 (2019).
3. Robertson, E. P. *et al.* Decoupling of bird migration from the changing phenology of spring green-up. *Proc. Natl. Acad. Sci.* **121**, e2308433121 (2024).
4. Zwerts, J. A. *et al.* FSC-certified forest management benefits large mammals compared to non-FSC. *Nature* (2024) doi:10.1038/s41586-024-07257-8.
5. Hudson, M.-A. R. *et al.* The role of the North American Breeding Bird Survey in conservation. *The Condor* **119**, 526–545 (2017).
6. Besson, M. *et al.* Towards the fully automated monitoring of ecological communities. *Ecol. Lett.* **25**, 2753–2775 (2022).
7. Gibb, R., Browning, E., Glover-Kapfer, P. & Jones, K. E. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.* **10**, 169–185 (2019).
8. Stowell, D. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ* **10**, e13152 (2022).
9. Kahl, S. BirdNET: A deep learning solution for avian diversity monitoring. *Ecol. Inform.* 10 (2021).
10. Cretois, B., Rosten, C. M. & Sethi, S. S. Voice activity detection in eco-acoustic data enables privacy protection and is a proxy for human disturbance. *Methods Ecol. Evol.* **13**, 2865–2874 (2022).
11. Prince, P. *et al.* Deploying Acoustic Detection Algorithms on Low-Cost, Open-Source Acoustic Sensors for Environmental Monitoring. *Sensors* **19**, 553 (2019).
12. Aniceto, A. S., Ferguson, E. L., Pedersen, G., Tarroux, A. & Primicerio, R. Temporal patterns in the soundscape of a Norwegian gateway to the Arctic. *Sci. Rep.* **12**, 7655 (2022).

13. Sethi, S. S. *et al.* SAFE Acoustics: An open-source, real-time eco-acoustic monitoring network in the tropical rainforests of Borneo. *Methods Ecol. Evol.* **11**, 1182–1185 (2020).

14. Sethi, S. S. *et al.* Automatic vocalisation detection delivers reliable, multi-faceted, and global avian biodiversity monitoring. Preprint at https://doi.org/10.1101/2023.09.14.557670 (2023).

15. Ware, H. E., McClure, C. J. W., Carlisle, J. D. & Barber, J. R. A phantom road experiment reveals traffic noise is an invisible source of habitat degradation. *Proc. Natl. Acad. Sci.* **112**, 12105–12109 (2015).

16. Halfwerk, W., Lohr, B. & Slabbekoorn, H. Impact of Man-Made Sound on Birds and Their Songs. in *Effects of Anthropogenic Noise on Animals* (eds. Slabbekoorn, H., Dooling, R. J., Popper, A. N. & Fay, R. R.) vol. 66 209–242 (Springer New York, New York, NY, 2018).

17. Butler, C. J. The disproportionate effect of global warming on the arrival dates of short-distance migratory birds in North America. *Ibis* **145**, 484–495 (2003).

18. J. A. Kålås, I. J. Øien, & B. Stokke. Norwegian breeding bird monitoring scheme - Norsk hekkefugloverváking. GBIF https://doi.org/10.15468/6jmw2e (2022).

19. La Sorte, F. A. *et al.* Global change and the distributional dynamics of migratory bird populations wintering in Central America. *Glob. Change Biol.* **23**, 5284–5296 (2017).

20. Zhang, G. Spatial and Temporal Patterns in Volunteer Data Contribution Activities: A Case Study of eBird. *ISPRS Int. J. Geo-Inf.* **9**, 597 (2020).

21. Johnston, A. *et al.* Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. *Divers. Distrib.* **27**, 1265–1277 (2021).

22. Wood, C. M. & Kahl, S. Guidelines for appropriate use of BirdNET scores and other detector outputs. *J. Ornithol.* (2024) doi:10.1007/s10336-024-02144-5.

23. Fink, D. *et al.* eBird Status and Trends, Data Version: 2022. Cornell Lab of Ornithology https://doi.org/10.2173/ebirdst.2022 (2023).

24. Hedenström, A. & Pettersson, J. Migration routes and wintering areas of Willow Warblers Phylloscopus trochilus (L.) ringed in Fennoscandia.

25. Lyapustin, A. MODIS/Terra+Aqua Vegetation Index from MAIAC, Daily L3 Global 0.05Deg CMG V061. NASA EOSDIS Land Processes Distributed Active Archive Center https://doi.org/doi.org/10.5067/MODIS/MCD19A3CMG.061 (2023).

26. Barrett, R. T. The dependence of long-distance migration to North Norway on environmental conditions in the wintering area and en route. *Ornis Nor.* **40**, 14 (2017).

27. Desjonquères, C. *et al.* Acoustic species distribution models ( ASDMs ): A framework to forecast shifts in calling behaviour under climate change. *Methods Ecol. Evol.* **13**, 2275–2288 (2022).

28. Copernicus Climate Change Service, Climate Data Store. Land cover classification gridded maps from 1992 to present derived from satellite observations. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) https://doi.org/10.24381/cds.006f2c9a (2019).

29. Zamora-Marín, J. M., Zamora-López, A., Calvo, J. F. & Oliva-Paterna, F. J. Comparing detectability patterns of bird species using multi-method occupancy modelling. *Sci. Rep.* **11**, 2558 (2021).

30. OpenStreetMaps Contributors. OpenStreetMaps Norway Dataset. Geofabrik (2022).

31. Heath, B. E., Sethi, S. S., Orme, C. D. L., Ewers, R. M. & Picinali, L. How index selection, compression, and recording schedule impact the description of ecological soundscapes. *Ecol. Evol.* **11**, 13206–13217 (2021).

32. Beery, S., Cole, E., Parker, J., Perona, P. & Winner, K. Species Distribution Modeling for Machine Learning Practitioners: A Review. in *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS)* 329–348 (ACM, Virtual Event Australia, 2021). doi:10.1145/3460112.3471966.

33. Cristian Lussana. seNorge/TX: daily maximum temperature over Norway (22.09) [Data Set]. Zenodo https://doi.org/10.5281/zenodo.6965960 (2022).

34. Cristian Lussana. seNorge/TN: daily minimum temperature over Norway (22.09) [Data set]. Zenodo https://doi.org/10.5281/zenodo.6965929 (2022).

35. Cristian Lussana. seNorge/TG: daily mean temperature over Norway (22.09) [Data set]. Zenodo https://doi.org/10.5281/zenodo.6965866 (2022).

36. Cristian Lussana. seNorge/RR: daily total precipitation amounts over Norway (22.09) [Data set. Zenodo https://doi.org/10.5281/zenodo.6965906 (2022).

37. Kartverket (Norwegian Geographical Survey). Norway National Detailed Elevation Model. Kartverket.

# Methods

## Recording Site Selection

Recording sites were located at BBS survey points (within 100m) to maximize compatibility between the two datasets. The sites were also restricted to forested areas, as determined by the "Forest" IPCC Class in 300m resolution land cover data from Copernicus[28]. When possible, sites were chosen within 200m of accessible roads[30] or trails for ease of deployment and within areas with 4G data coverage. To maintain confidentiality of BBS survey points, random noise of up to 1 km was added to the location of each recorder in the publicly available version of this dataset.

Across Spring 2022 in Norway, 28 recording stations were installed across eight regions (from south to north): Kristiansand (5), Stavanger (2), Moss (3), Bergen (3), Lillehammer (2), Trondheim (4), Tromsø (4), and Finnmark (5). These recorders were then grouped into regions, defined by a 90 x 90 km box sharing a centroid with the recorders, the largest size possible without overlap between regions. The Finnmark cluster of recorders was dropped from species arrival analysis, as only one valid detection of a species of interest for this study was found. However, these sites are included in Figure 1 and for training and evaluating the audio species distribution models (Fig. 4). Sites within clusters were spread such that each could be considered acoustically independent (i.e., >100m apart).

## Audio Collection

In total, 73,516 hours of audio was collected from March to October 2022. Audio at each site was collected using a Bugg device 1-2m from ground level. Devices were powered through using 50-100W solar panels and 20-30 Ah 12V batteries, and thus were sensitive to shade and cloud cover. Still, 19 of 28 recorders had uptimes over 75% across the study period. Files were saved locally in five-minute increments at a 44.1 kHz sampling rate in compressed MP3 format (VBR0 using FFMPEG's LAME codec), then uploaded to a Google Cloud server via the 4G network. Previous work has found MP3 compression at such high quality has little effect on downstream eco-acoustic analyses[31]. Where network service was unavailable, audio data was manually uploaded upon device retrieval in Fall 2022.

## Species Detection

Audio data was processed on a custom Google Cloud web app using BirdNET-lite to extract and identify vocalizations to species level in near real-time[9]. We provided an expert ornithologist with 50 audio clips

of randomly sampled detections for each species with 50 or more detections with a BirdNET model confidence over 0.80. In total, 176 unique species were identified in this dataset by BirdNET. Of these, BirdNet correctly identified 55 species with 80% or greater accuracy. Of these species, 14 are full migrants to Norway each Spring and were considered when calculating daily avian species richness at each site.

## Species Arrival

We calculated species arrival for a subset of three fully migrant passerines that were commonly detected in our dataset and had high detection accuracy from BirdNET: Willow Warbler (*Phylloscopus trochilus*), Spotted Flycatcher (*Muscicapa striata*), and Common Chiffchaff (*Phylloscopus collybita*). We defined the study period as April 1 through June 30, 2022, by which most migrating bird species have arrived across Norway.

We estimated the date of migratory arrival for each species at a given region by setting a threshold of 5% total detections across our study period, a common method for setting arrival dates in migratory studies (Hedlund 2014, etc.). Thus, when calculating arrival for a given species, we only consider regions with more than 20 valid detections. To correct for varying spatiotemporal sampling effort, the daily detections of each species were normalized by the daily sum of recorder uptime within each region. To calculate the percentage of total vocalizations over time, we first take a cumulative sum of these normalized uptimes across the study period, then normalized these from 0 to 1. We estimate the date of species arrival as the first day with ≥ 5% of total vocalizations.

NDVI is sourced from the daily 0.05 degree resolution gap-filled MODIS Vegetation Index dataset[25]. Daily NDVI per region is calculated by averaging all NDVI cells within each 90x90 km region.

## Comparison of Audio to Manual Survey Methods

### eBird and BBS Surveys

Raw eBird surveys[23] were aggregated within each 90x90 km region, then filtered by several criteria to standardize data[21]. First, to ensure that pseudo-absences could be assumed from eBird surveys, only complete checklists were included, those where observers marked that all observed species were included. Surveys with high effort, those traveling over 5km or lasting longer than 90 minutes, were dropped to ensure that observations occurred close to the provided survey start coordinates. Duplicate surveys that appear in group surveys were dropped. Lastly, surveys in non-forested areas, as identified by the

Copernicus data set[28], were dropped to ensure comparability against the audio dataset. Lastly, surveys were filtered to the elevation range of our recording stations, 73m to 764m.

Typically, Norwegian Breeding Bird Survey (BBS) data is spatially aggregated to protect the specific coordinates of survey points from increased traffic which could affect avian nesting habits. Non-public BBS data was obtained from the survey organizers[18] for the three migrating passerines of interest. The dataset consists of observations of the number of breeding pairs of each species at 9,257 unique points across Norway by experienced volunteers. Many volunteers are trained by attending the BIRD ID course at Nord University, but a prerequisite for all is good knowledge of bird sounds and appearance. Regional coordinators communicate with observers before they perform surveys, and observations are skipped in rare cases where poor knowledge or reduced hearing abilities is suspected by the survey organizers. The BBS is carried out from 20 May through 10 July each year after all species have arrived at their breeding sites, to be able to exclude individuals that are traveling. Volunteers travel along transects, making observations at several points. Between 4AM and 10AM, they count the number of breeding pairs they observe in 5 minutes at each point[18]. These observations were also filtered to forested regions and to the elevation range (73m to 764m) of our recording stations.

## Audio Species Distribution Models

We trained and cross-validated a random forest regression model on several weekly ecoclimatic covariates[32] (Beery 2021) to predict the probability of vocalization[27] for each of ten full migrant species (i.e. for a total of 10 random forest models) that were both precisely identified by BirdNET (≥ 80%) and had at least 10 weekly detections across all sites: Barn Swallow (*Hirundo rustica*), Common Chiffchaff (*Phylloscopus collybita*), Common Cuckoo (*Cuculus canorus*), Common Sandpiper (*Actitis hypoleucos*), Eurasian Blackcap (*Sylvia atricapilla*), Green Sandpiper (*Tringa ochropus*), White Wagtail (*Motacilla alba*), Willow Warbler (*Phylloscopus trochilus*), Spotted Flycatcher (*Muscicapa striata*), and European Pied Flycatcher (*Ficedula hypoleuca*).

The weekly binary detection (0,1) of a species at each of the 28 sites served as the target variable, with weekly averages of NDVI[25], maximum[33], minimum[34], and mean[35] temperature, and precipitation[36], as well as site elevation[37] as the predictor variables. Predictor variables were extracted for each site's coordinates. In Finnmark and Tromsø, where snow cover continues late into the spring, several weeks of NDVI coverage were not available from the MODIS dataset[25] and were zero-filled. The random forest regression models were hyperparameter-tuned using a random grid search approach over 1,000 iterations. The model

parameters with the best mean squared error performance on the test set were chosen via 4-fold cross-validation. The tuned model parameters and train and test set errors are available in SI Table 1.

We then generated weekly aSDMs using the trained models for each species. Covariate rasters were resampled to a common resolution of 0.2 degrees (epsg:4326) and used to predict probability of vocalization across Norway. Based on the aSDM cell size, a masking step then filtered out non-majority-forested cells from the aSDM, using the 300m resolution Copernicus land cover dataset[28].

Lastly, aSDMs were then compared against the filtered complete eBird and BBS lists (i.e., presence and absence included). For each month, we compared the distributions of aSDM values with positive survey detections (1) and for the survey absences (0) using a permutation test with 1,000 resamples and alternative "greater". This test evaluates whether the means of the distributions of aSDM values were greater when eBird or BBS detects a species, versus when they do not.

## Acknowledgements

We thank Tom Roger Østerås of Norsk hekkefugloverváking for his expert opinion in validating BirdNET species detections, NINA field staff for assistance with audio recorder setup, and the Norwegian Environment Agency (Miljødirektoratet) for funding of data collection. IAB was funded by a doctoral research grant 323294 from the Research Council of Norway.

## Author Contributions

CR, SSS, and BC organized the Sound of Norway (thesoundofnorway.com) project which facilitated this study. IAB identified suitable locations for acoustic recorders and worked with CR to receive landowner permissions to deploy recorders. CR, SSS, IAB, and JW deployed and collected acoustic recorders. IAB, SSS, MS, and BH developed methodologies and research objectives, with advisement from VB, MP, and KR. IAB performed all data processing and modeling and generated all figures. IAB drafted the manuscript with significant contributions from SSS. All coauthors reviewed and approved the final manuscript.

## Data Availability

All code, manual survey data, acoustic detections, and ecoclimatic covariates required to generate each figure in this study are available at doi.org/10.5281/zenodo.11238950. Random

noise of up to 0.01 degrees has been added to all coordinates of recording stations and Norwegian Breeding Bird survey points to maintain the confidentiality and data integrity of the survey.

# Supporting Information

**SI Table 1.** Audio Species Distribution Model (aSDM) test and train set mean absolute error (MAE) for ten full-migrant species. Predicted aSDM values range from 0.0 to 1.0. One Random Forest model was trained for each species and hyperparameters were chosen over 1000 iterations of randomized parameters to optimize the test set MAEs.

| **Species** | **Train MAE** | **Test MAE** | **Train MSE** | **Test MSE** | **Random Forest Parameters** |
|---|---|---|---|---|---|
| Barn Swallow (*Hirundo rustica*) | 0.0222 | 0.0575 | 0.00462 | 0.0264 | {'n_estimators': 176, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': None, 'max_depth': 10, 'bootstrap': True} |
| Common Chiffchaff (*Phylloscopus collybita*) | 0.0514 | 0.109 | 0.0121 | 0.0486 | {'n_estimators': 147, 'min_samples_split': 5, 'min_samples_leaf': 1, 'max_features': None, 'max_depth': 70, 'bootstrap': True} |
| Common Cuckoo (*Cuculus canorus*) | 0.0122 | 0.0284 | 0.00278 | 0.0109 | {'n_estimators': 128, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'log2', 'max_depth': None, 'bootstrap': True} |
| Common Sandpiper (*Actitis hypoleucos*) | 0.0295 | 0.0948 | 0.00628 | 0.0581 | {'n_estimators': 434, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'log2', 'max_depth': None, 'bootstrap': True} |
| Eurasian Blackcap (*Sylvia atricapilla*) | 0.0166 | 0.0555 | 0.00444 | 0.0374 | {'n_estimators': 119, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'sqrt', 'max_depth': 30, 'bootstrap': True} |
| Green Sandpiper (*Tringa ochropus*) | 0.023 | 0.0458 | 0.00671 | 0.0306 | {'n_estimators': 386, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'log2', 'max_depth': 80, 'bootstrap': True} |
| White Wagtail (*Motacilla alba*) | 0.0181 | 0.0356 | 0.00446 | 0.0193 | {'n_estimators': 462, 'min_samples_split': 5, 'min_samples_leaf': 2, 'max_features': None, 'max_depth': 50, 'bootstrap': True} |
| Willow Warbler (*Phylloscopus trochilus*) | 0.0766 | 0.226 | 0.0184 | 0.132 | {'n_estimators': 1083, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'log2', 'max_depth': 60, 'bootstrap': True} |
| Spotted Flycatcher (*Muscicapa striata*) | 0.0291 | 0.0701 | 0.00701 | 0.0388 | {'n_estimators': 606, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': None, 'max_depth': None, 'bootstrap': True} |
| European Pied Flycatcher (*Ficedula hypoleuca*) | 0.0418 | 0.0965 | 0.00968 | 0.0543 | {'n_estimators': 128, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'log2', 'max_depth': None, 'bootstrap': True} |