National-scale acoustic monitoring of avian biodiversity and migration

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 M. 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

Billions of birds migrate annually, triggered by endogenous behaviors as well as ecoclimatic triggers, which are shifting with climate change. These dynamics play out over large spatiotemporal scales, 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 Norway. We used machine learning to automatically identify bird vocalizations, and with expert validation found we were able to classify 57 species (14 full migrants) with over 80% precision. We show that acoustic surveys can fill data gaps in traditional surveys and facilitate mapping of migratory waves across Norwegian forests. Our study demonstrates how acoustic monitoring can 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 biodiversity and species' responses to stressors and interventions is critical for informing and evaluating conservation and land management strategies (*3*, *4*). Traditionally, ecologists rely on citizen science to track avian phenology, populations, and inform conservation policy (*5*) through either national breeding bird surveys or apps such as eBird, iNaturalist, and Artsobservasjoner.

Breeding bird surveys are typically performed by knowledgeable volunteers (*6*) and capture a snapshot of bird populations in the late-Spring and early-Summer across thousands of sites in the northern hemisphere (*6*). While these surveys are a thorough and crucial means of estimating population trends and ranges for many species (*5*), their lack of temporal resolution makes them ineffective for studying migratory timing and phenology, such as shifts in Spring migratory bird arrival timing with global warming (*1*, *3*). Conversely, datasets which allow the public to record observations opportunistically (e.g., eBird) provide some insight into avian phenology timing (*7*) but suffer from temporal and spatial biases (*8*) and a high variation in data reliability (*9*). Ecologists have also long employed weather radars to track birds and migratory patterns over large distances (*10*). While radar methods effectively track aggregate migrant behavior, such as identifying stopover locations (*10*), they cannot easily distinguish between species.

Automated approaches are increasingly being used to complement manual surveys by improving the scale and resolution of biodiversity monitoring (*11*). Passive acoustic monitoring (PAM) has shown particular promise due to the approach's taxonomic breadth (*12*) and scalability for both data analysis and collection (*11*). Automated analysis of audio is becoming commonplace, supported by the growth of powerful machine learning (ML) models (*13*). ML algorithms such as BirdNET can accurately identify hundreds of species from recordings and anthropogenic noise pollution (*14*, *15*), while others are used to detect biodiversity pressures such as noise pollution (*16*), illegal logging (*17*), or poaching (*18*, *19*). Autonomous and real-time acoustic data collection via cellular network connection and self-powered units (via solar power or connection to a grid) is also growing (*20*, *21*), but few studies (*22*) 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 PAM. 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 (*14*, *15*, *22*). Improved monitoring of avian phenology could help us understand how stressors such as noise pollution (*23*), habitat loss (*24*), and climate change (e.g., "spring mismatch" (*3*, *25*)) impact avian migration routes, behaviors, and breeding outcomes at relevant scales, informing conservation policy in an era of rapid ecological change (*26*).

In this study, we perform one of the first national-scale acoustic surveys (*27*) of avian phenology using autonomous acoustic recorders installed across almost the entire latitudinal extent of Norway. We

demonstrate that PAM can measure avian phenology at high resolution across large scales, that detections of vocalizations from an ML model are reliable for many species through expert validation, and that PAM provides complementary biodiversity data to existing surveys. We also show that by relating PAM detections to ecoclimatic variables (*28*), it is possible to visualize the spread of avian vocalization probability across Spring migration at a national scale.

PAM is reliable and fills data gaps in existing surveys

Our solar-powered, networked PAM stations were installed 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 utilized an expert validation methodology to understand the precision of BirdNET classifications in our study area and focus our analysis on species that the model detections with high precision. Fifty BirdNET detections, and their corresponding audio, were provided to an expert ornithologist for 67 species which had at least 50 BirdNET detections with a model confidence (*29*) over 0.80. After expert labeling of these recordings, we find that BirdNET correctly identified 57 species (including 14 full migrants) with greater than 80% precision and 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 as well as latitudinal differences in biodiversity and migratory arrival (Fig. 1c)

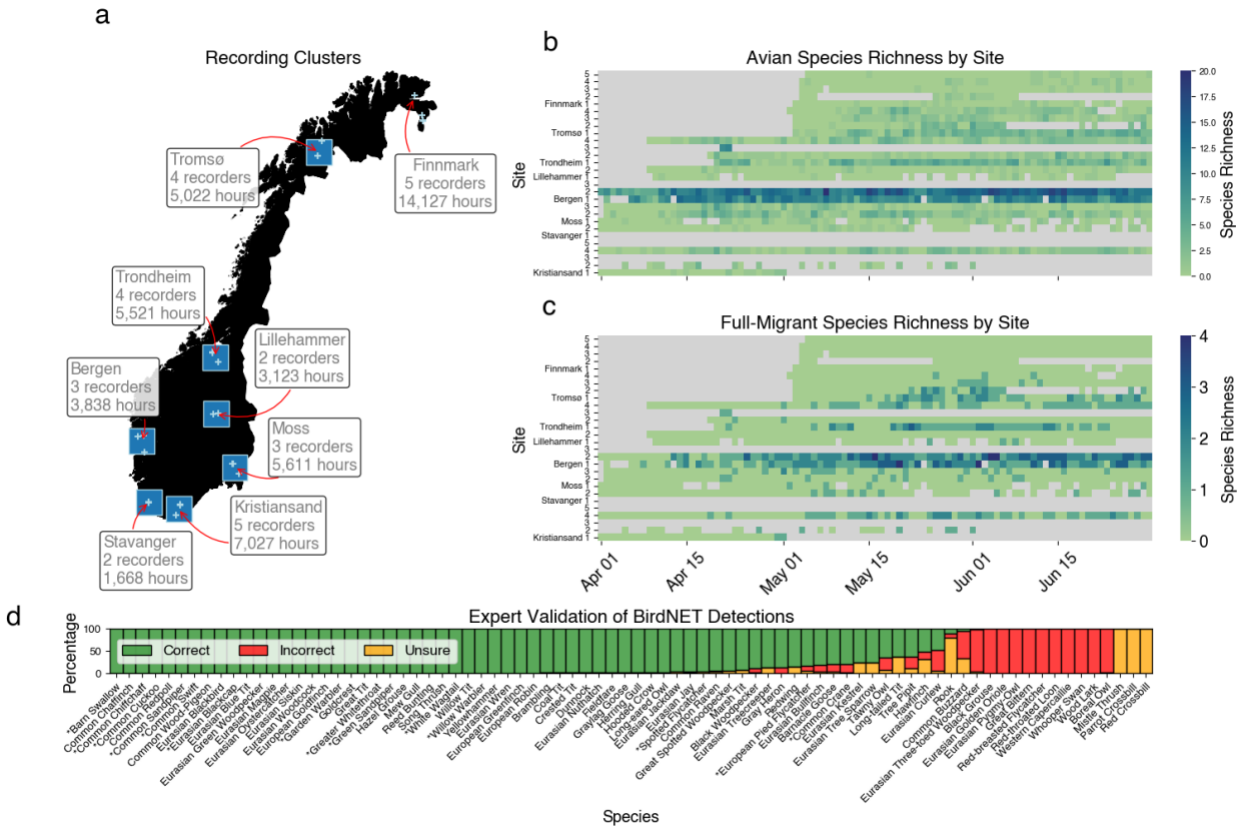

Fig. 1 | Passive acoustic monitoring captures avian species richness during migratory arrival and is highly accurate in detecting dozens of species. a, 28 passive acoustic monitors deployed across Norway (*30*) in 2022 are grouped into eight regions. Blue boxes (90 km x 90 km) at the centroid of most region's recorders are used to filter and count 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 57 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 67 species with ≥50 detections over ≥0.80 model confidence. Of 67 species, 57 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 further on three common migrating passerines detected with 100% precision (Fig. 1d) which had sufficient data for downstream analysis: 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 (*31*), and standardized line transect point counts from Norsk Hekkefuglovervåking, also known as the Norwegian Breeding Bird Survey (BBS) (*6*). We found that PAM delivered a far greater number of total species observations than manual

surveys for WW across the study period within our study regions (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). We note a distinct two-peak pattern of vocalization frequency over time for the three species (Fig. 2a,d,e), that has previously been observed for Willow Warbler (*32*) and other passerines (*33*). The first peak in mid-May represents the pre-breeding season (Fig. 2a), with males establishing territory and performing courtship (*32*). This is followed by a drop in vocalization during pairing, then a second increase during laying and incubation, which may be due to males performing song to attract second females (*32*).

As BBS is only performed between late May to early July, we also compared detections solely on days where both BBS and PAM data were 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 only two to five recorders per 90 km x 90 km region, PAM provides greater sampling effort temporally, giving phenological context for manual surveys which are concentrated during breeding season. In contrast, BBS had a greater spatial sampling effort, being much more likely to have a unique daily detection of these three species than PAM.

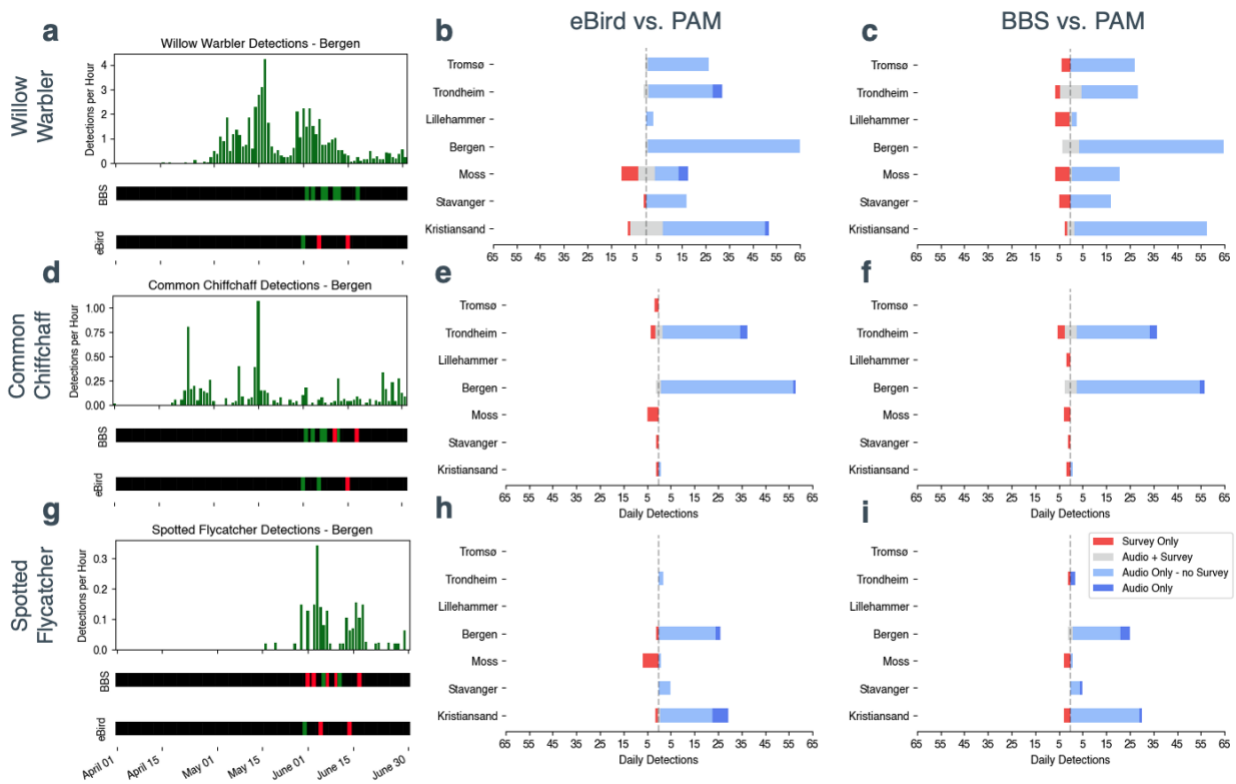

Fig. 2 | For three common migrating passerines, passive acoustic monitoring data can fill gaps and improve temporal resolution of sampling compared to existing biodiversity datasets, but on days where manual surveys and PAM were conducted, manual surveys had more unique detections. a, d, g, Audio recorders 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

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, and thus arriving sooner. The timings also align with these species' respective food sources: 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 and Tromsø, WW traveled northward at approximately 67 km/day, agreeing with previous measurements of WW southward autumn migratory rate of 40.6-80.5 km/day using ringed birds (*34*). Intuitively, we found that at higher latitudes WW arrives at a lower greenness, as measured by satellite-derived Normalized Difference Vegetation Index (NDVI) (*35*) (Spearman ρ=-0.83, p=0.04). As migrating birds must contend with short breeding seasons in northern regions, they often carefully balance their arrival times to match

the timing snowmelt and greening of northern birch forests (*36*), ideally matching peak food abundance to when chicks are present in their nests (*37*). Mistiming migration to these colder, northern areas may lead to mortality (*37*, *38*) and climatic changes in wintering areas in Africa seem to have driven WW to arrive approximately half a day earlier to the Tromsø region each year since 1995 (*37*), potentially threatening this phenological timing. Whilst previously only investigated at regional scales (*26*), our results show that large PAM networks have the potential to reveal long-term changes in avian migratory phenology on national scales and beyond, identifying species or ecosystems at the most risk from climate change.

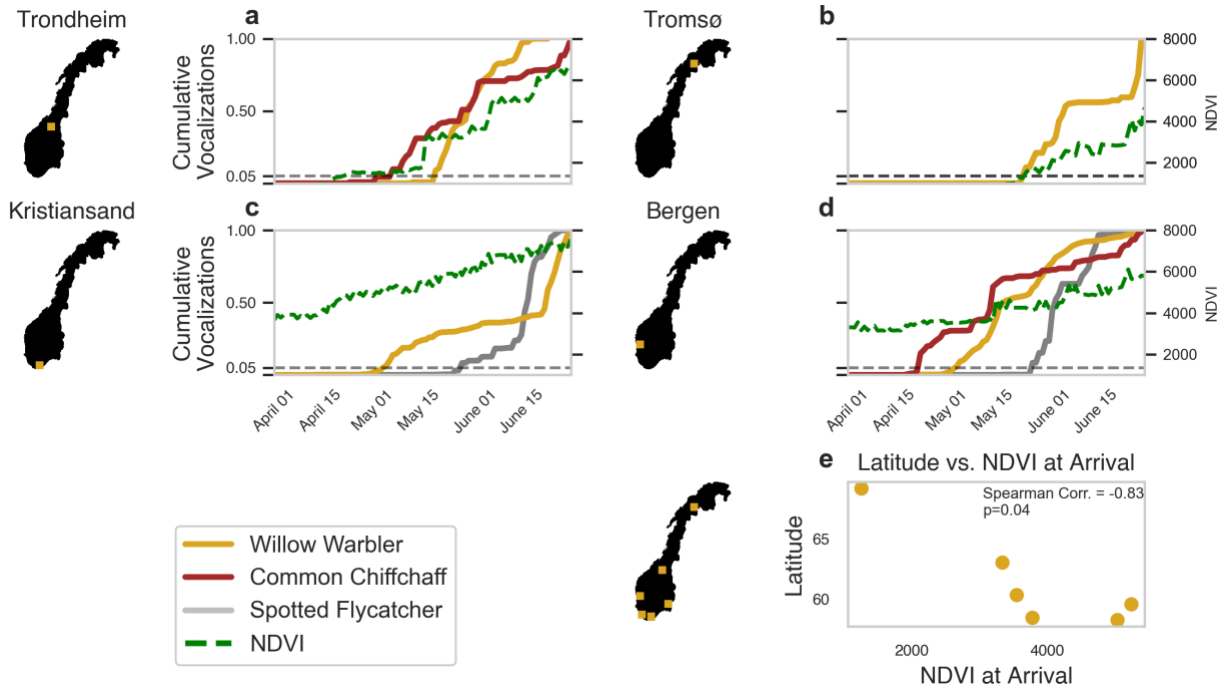

Fig. 3 | Cumulative audio detections of migrating passerines demonstrate interspecies and large-scale variation in migration timing across Norway. a, b, c, d, Passive acoustic monitoring shows ecologically consistent species arrival patterns across Norway (*30*). Spotted Flycatcher, with a diet of flying 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.

We next hyperparameter-tuned and trained random forest regression models (*39*) to act as acoustic species distribution models (aSDMs) (*28*) by predicting the weekly probability of hearing WW and SF vocalizations based on several ecoclimatic features. The aSDMs allowed us to extrapolate spatially beyond our sampling sites to other forested regions (*40*) similar to where we've sampled, mapping migratory waves of birds based on their probability of vocalization (Fig. 4) at a resolution of 0.2 degrees,

approximately 10 km. 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). Between May 8-14, the model predicts that Willow Warbler are largely not yet vocalizing in the northern Tromsø region, but that these areas are populated two weeks later. This aligns with both the raw species detections (Fig. 3b) and the estimate of May 14 as the first arrival date of Willow Warbler in Troms County by Barrett (2002) (*36*).

To test the agreement of the aSDMs with eBird and BBS survey data in forested regions, we sampled the distributions of aSDM values at survey observation points, observing whether higher aSDM vocalization probabilities were associated with positive observations from manual surveys (Fig. 4c,d). The aSDMs agree well with eBird observations by permutation test (Fig. 4c,d) of WW ($p \leq 0.001$, $t=0.102$) and SF ($p=0.014$, $t=0.071$). Similarly, Bota et al 2020 found that avian vocal activity rate of European Bee-eaters, as detected by PAM recorders, correlated positively with their inclusion in complete checklists from citizen scientists (*41*). However, we see that the aSDMs for WW do not significantly agree with the BBS observations (Fig. 4d), perhaps due to the limited temporal range of BBS in late May and June, during which WW are nearly ubiquitous across Norwegian forests (Fig. 4a). In contrast, the BBS observations agree with the aSDM for SF, which have a more heterogenous distribution (Fig. 4b). The general agreement of the aSDMs with citizen-science based datasets illustrates the complementary nature of PAM with survey data. As the aSDM vocalization probabilities capture a metric of species detectability (*9*, *42*), they also offer a means of informing the timing of both breeding bird surveys and more casual citizen scientists. Such aSDM models may also be extended to project how conditions suitable for species vocalization may change over time with climate change (*28*).

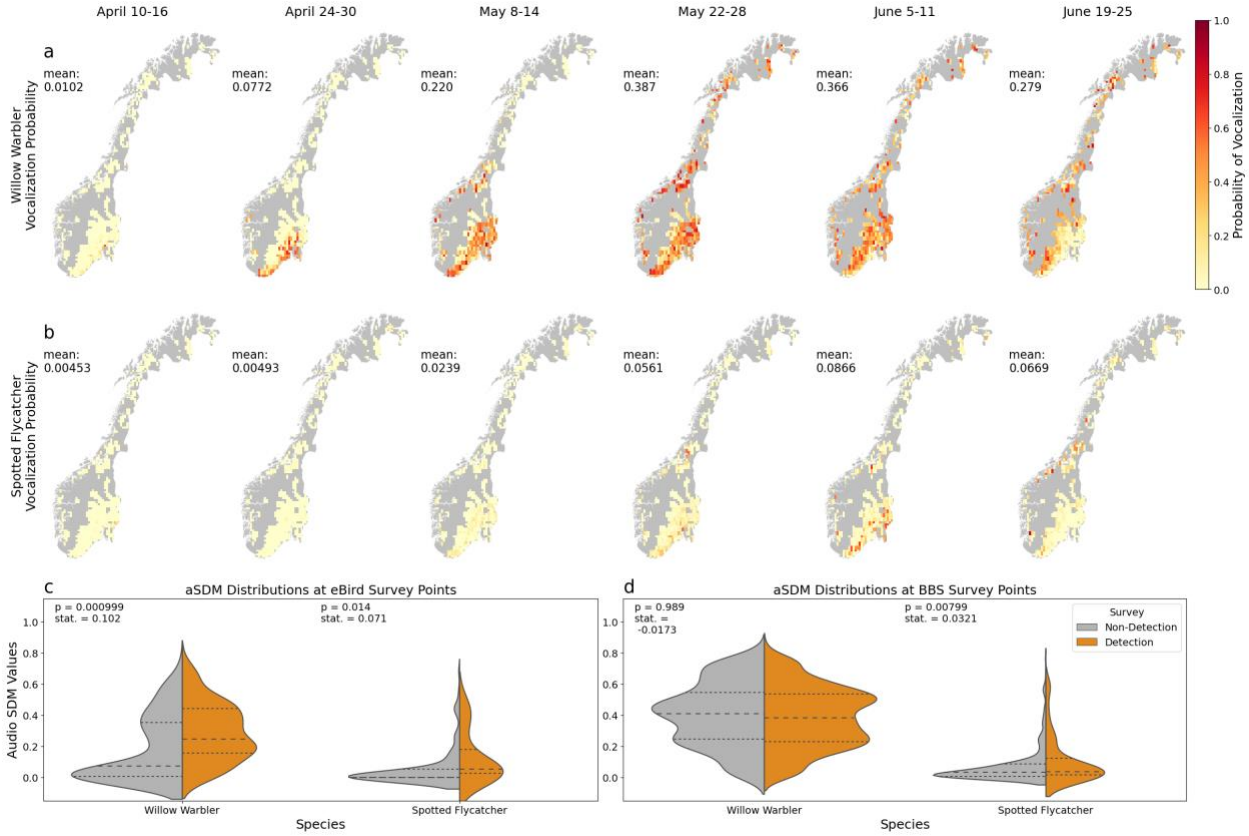

Fig. 4 | Combining audio vocalization detections with ecoclimatic variables allowed us to train audio species distribution models (aSDMs) to predict vocalization probability as an analogue of species occurrence across Norway. Weekly vocalization probability in forested areas across Norway (*30*) for **a,** Willow Warbler and **b,** Spotted Flycatcher shows spatiotemporal species arrival patterns and captures later arrival of Spotted Flycatcher relative to Willow Warbler. **c,** 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. **d,** For BBS survey points, the distribution of aSDM values is significantly greater when the surveyor detected SF, but not for WW.

Discussion and Future Work

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 manual surveys. Pairing networked sensors and automated analyses with one-off expert validation allowed us to track phenology reliably and in near real-time and required minimal cost and effort to deploy and maintain. However, even with improved temporal resolution and lower labor requirements, PAM methods should be viewed as a complement to existing manual observation methods, such as breeding bird surveys. For instance, PAM cannot easily distinguish between vocalizations of resident birds in an area versus those who are passing migrant individuals, or discern whether calls are from individual birds or multiple, without the use of multiple microphones in a single array (*43*). Breeding bird surveys are designed specifically to count

breeding pairs during the period after most migrants have arrived, typically June in Norway. Therefore, PAM observations before this period represent occurrence of a migrating species, but not necessarily their population or residency within a given area.

The expert validation approach employed here enabled estimation of BirdNET model precision in identifying bird vocalizations but is limited by inability to calculate recall. Calculating recall would involve calculating the fraction of actual vocalizations that are correctly detected by BirdNET, requiring each detection to be manually checked for all species. This would be a challenge considering there were 33,643 detections with a model confidence over 0.80. We argue that precision provides a more useful metric, critical for ensuring reliable detections in remote areas with lower bird populations and for facilitating stakeholder trust in model results. However, by thresholding detections at a model confidence over 0.80 for validation, we subsequently filtered out 81% of the total detections from our recorders, possibly significantly reducing the model's recall at the expense of our higher precision (*44*). Depending on goals, future studies may consider lowering thresholds for model confidence, or setting model confidence thresholds on a species-by-species basis (*14*). Financial and practical limitations also affected the number of recorders deployed to each region, making our study sensitive to failure of some recording devices (see Fig. 1) and introducing possible spatial bias in the total number of detections by region, though we partially corrected for this by normalizing daily detections by the daily number of hours recorded.

Future studies may take advantage of the fact that sensors are becoming cheaper, more sensitive, and more reliable (*18*) while machine learning models for identifying species by vocalization are improving in accuracy, taxonomic breadth (e.g., insects, mammals), and computational efficiency (*13*), making national-scale PAM surveys more readily implementable. Efforts to provide large-scale, comprehensive biodiversity mapping, such as the Dutch ARISE program (*45*), may eventually draw upon multiple sources of species observations, including PAM, camera traps (*46*), drones (*47*), citizen science (*48*), and formal surveys (*6*). 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. N. Saino, R. Ambrosini, D. Rubolini, J. Von Hardenberg, A. Provenzale, K. Hüppop, O. Hüppop, A. Lehikoinen, E. Lehikoinen, K. Rainio, M. Romano, L. Sokolov, Climate warming, ecological mismatch at arrival and population decline in migratory birds. *Proc. R. Soc. B Biol. Sci.* **278**, 835–842 (2011).

2. M. Blicharska, R. J. Smithers, G. Mikusiński, P. Rönnbäck, P. A. Harrison, M. Nilsson, W. J. Sutherland, Biodiversity's contributions to sustainable development. *Nat. Sustain.* **2**, 1083–1093 (2019).

3. E. P. Robertson, F. A. La Sorte, J. D. Mays, P. J. Taillie, O. J. Robinson, R. J. Ansley, T. J. O'Connell, C. A. Davis, S. R. Loss, Decoupling of bird migration from the changing phenology of spring green-up. *Proc. Natl. Acad. Sci.* **121**, e2308433121 (2024).

4. J. A. Zwerts, E. H. M. Sterck, P. A. Verweij, F. Maisels, J. Van Der Waarde, E. A. M. Geelen, G. B. Tchoumba, H. F. Donfouet Zebaze, M. Van Kuijk, FSC-certified forest management benefits large mammals compared to non-FSC. *Nature*, doi: 10.1038/s41586-024-07257-8 (2024).

5. M.-A. R. Hudson, C. M. Francis, K. J. Campbell, C. M. Downes, A. C. Smith, K. L. Pardieck, The role of the North American Breeding Bird Survey in conservation. *The Condor* **119**, 526–545 (2017).

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

7. F. A. La Sorte, D. Fink, P. J. Blancher, A. D. Rodewald, V. Ruiz-Gutierrez, K. V. Rosenberg, W. M. Hochachka, P. H. Verburg, S. Kelling, Global change and the distributional dynamics of migratory bird populations wintering in Central America. *Glob. Change Biol.* **23**, 5284–5296 (2017).

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

9. A. Johnston, W. M. Hochachka, M. E. Strimas-Mackey, V. Ruiz Gutierrez, O. J. Robinson, E. T. Miller, T. Auer, S. T. Kelling, D. Fink, 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).

10. K. G. Horton, J. J. Buler, S. J. Anderson, C. S. Burt, A. C. Collins, A. M. Dokter, F. Guo, D. Sheldon, M. A. Tomaszewska, G. M. Henebry, Artificial light at night is a top predictor of bird migration stopover density. *Nat. Commun.* **14**, 7446 (2023).

11. M. Besson, J. Alison, K. Bjerge, T. E. Gorochowski, T. T. Høye, T. Jucker, H. M. R. Mann, C. F. Clements, Towards the fully automated monitoring of ecological communities. *Ecol. Lett.* **25**, 2753–2775 (2022).

12. R. Gibb, E. Browning, P. Glover-Kapfer, K. E. Jones, Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.* **10**, 169–185 (2019).

13. D. Stowell, Computational bioacoustics with deep learning: a review and roadmap. *PeerJ* **10**, e13152 (2022).

14. C. Pérez-Granados, BirdNET: applications, performance, pitfalls and future opportunities. *Ibis* **165**, 1068–1075 (2023).

15. S. Kahl, BirdNET: A deep learning solution for avian diversity monitoring. *Ecol. Inform.*, 10 (2021).

16. B. Cretois, C. M. Rosten, S. S. Sethi, Voice activity detection in eco-acoustic data enables privacy protection and is a proxy for human disturbance. *Methods Ecol. Evol.* **13**, 2865–2874 (2022).

17. C. Astaras, J. M. Linder, P. Wrege, R. D. Orume, D. W. Macdonald, Passive acoustic monitoring as a law enforcement tool for Afrotropical rainforests. *Front. Ecol. Environ.* **15**, 233–234 (2017).

18. P. Prince, A. Hill, E. Piña Covarrubias, P. Doncaster, J. Snaddon, A. Rogers, Deploying Acoustic Detection Algorithms on Low-Cost, Open-Source Acoustic Sensors for Environmental Monitoring. *Sensors* **19**, 553 (2019).

19. P. H. Wrege, E. D. Rowland, S. Keen, Y. Shiu, Acoustic monitoring for conservation in tropical forests: examples from forest elephants. *Methods Ecol. Evol.* **8**, 1292–1301 (2017).

20. A. S. Aniceto, E. L. Ferguson, G. Pedersen, A. Tarroux, R. Primicerio, Temporal patterns in the soundscape of a Norwegian gateway to the Arctic. *Sci. Rep.* **12**, 7655 (2022).

21. S. S. Sethi, R. M. Ewers, N. S. Jones, A. Signorelli, L. Picinali, C. D. L. Orme, SAFE Acoustics: An open-source, real-time eco-acoustic monitoring network in the tropical rainforests of Borneo. *Methods Ecol. Evol.* **11**, 1182–1185 (2020).

22. S. S. Sethi, A. Bick, M.-Y. Chen, R. Crouzeilles, B. V. Hillier, J. Lawson, C.-Y. Lee, S.-H. Liu, C. H. D. Freitas Parruco, C. Rosten, M. Somveille, M.-N. Tuanmu, C. Banks-Leite, Automatic vocalisation detection delivers reliable, multi-faceted, and global avian biodiversity monitoring. [Preprint] (2023). https://doi.org/10.1101/2023.09.14.557670.

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

24. W. Halfwerk, B. Lohr, H. Slabbekoorn, “Impact of Man-Made Sound on Birds and Their Songs” in *Effects of Anthropogenic Noise on Animals*, H. Slabbekoorn, R. J. Dooling, A. N. Popper, R. R. Fay, Eds. (Springer New York, New York, NY, 2018; http://link.springer.com/10.1007/978-1-4939-8574-6_8)vol. 66 of *Springer Handbook of Auditory Research*, pp. 209–242.

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

26. R. Y. Oliver, D. P. W. Ellis, H. E. Chmura, J. S. Krause, J. H. Pérez, S. K. Sweet, L. Gough, J. C. Wingfield, N. T. Boelman, Eavesdropping on the Arctic: Automated bioacoustics reveal dynamics in songbird breeding phenology. *Sci. Adv.* **4**, eaaq1084 (2018).

27. P. Roe, P. Eichinski, R. A. Fuller, P. G. McDonald, L. Schwarzkopf, M. Towsey, A. Truskinger, D. Tucker, D. M. Watson, The Australian Acoustic Observatory. *Methods Ecol. Evol.* **12**, 1802–1808 (2021).

28. C. Desjonquères, S. Villén-Pérez, P. De Marco, R. Márquez, J. F. Beltrán, D. Llusia, Acoustic species distribution models (aSDMs): A framework to forecast shifts in calling behaviour under climate change. *Methods Ecol. Evol.* **13**, 2275–2288 (2022).

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

30. R. Hijmans, University of California, Berkeley, Museum of Vertebrate Zoology, First-level Administrative Divisions, Norway, 2015, UC Berkeley, Museum of Vertebrate Zoology (2015).

31. D. Fink, T. Auer, A. Johnston, M. Strimas-Mackey, S. Ligocki, O. Robinson, W. Hochachka, L. Jaromczyk, C. Crowley, K. Dunham, A. Stillman, I. Davies, V. Rodewald, V. Ruiz-Gutierrez, C. Wood, eBird Status and Trends, Data Version: 2022, Cornell Lab of Ornithology (2023); https://doi.org/10.2173/ebirdst.2022.

32. D. Gil, J. A. Graves, P. J. B. Slater, Seasonal patterns of singing in the willow warbler: evidence against the fertility announcement hypothesis. *Anim. Behav.* **58**, 995–1000 (1999).

33. I. G. Moran, K. C. Lukianchuk, Stéphanie. M. Doucet, A. E. M. Newman, H. Williams, D. R. Norris, D. J. Mennill, Diel and seasonal patterns of variation in the singing behaviour of Savannah Sparrows (Passerculus sandwichensis). *Avian Res.* **10**, 26 (2019).

34. A. Hedenström, J. Pettersson, Migration routes and wintering areas of Willow Warblers Phylloscopus trochilus (L.) ringed in Fennoscandia. *Ornis Fenn.* **64** (1987).

35. A. Lyapustin, MODIS/Terra+Aqua Vegetation Index from MAIAC, Daily L3 Global 0.05Deg CMG V061, version 6.1, NASA EOSDIS Land Processes Distributed Active Archive Center (2023); https://doi.org/doi.org/10.5067/MODIS/MCD19A3CMG.061.

36. R. T. Barrett, The phenology of spring bird migration to north Norway. *Bird Study* **49**, 270–277 (2002).

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

38. I. Newton, Weather-related mass-mortality events in migrants. *Ibis* **149**, 453–467 (2007).

39. L. Breiman, Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

40. Copernicus Climate Change Service, Climate Data Store, Land cover classification gridded maps from 1992 to present derived from satellite observations, version 2020, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (2019); https://doi.org/10.24381/cds.006f2c9a.

41. G. Bota, J. Traba, F. Sardà-Palomera, D. Giralt, C. Pérez-Granados, Acoustic Monitoring of Diurnally Migrating European Bee-Eaters Agrees with Data Derived from Citizen Science. *Ardea* **108** (2020).

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

43. B. E. Heath, R. Suzuki, N. P. Le Penru, J. Skinner, C. D. L. Orme, R. M. Ewers, S. S. Sethi, L. Picinali, Spatial ecosystem monitoring with a Multichannel Acoustic Autonomous Recording Unit (MAARU). *Methods Ecol. Evol.* **15**, 1568–1579 (2024).

44. E. J. Malamut, "Using Autonomous Recording Units and Image Processing to Investigate Patterns in Avian Singing Activity and Nesting Phenology," thesis, UCLA (2022).

45. E. Van Ommen Kloeke, C. Huijbers, K. Beentjes, J. Kamminga, P. Bakker, W. Kissling, ARISE: Building an infrastructure for species recognition and biodiversity monitoring in the Netherlands. *Biodivers. Inf. Sci. Stand.* **6**, e93613 (2022).

46. L. N. Rich, B. J. Furnas, D. S. Newton, J. S. Brashares, Acoustic and camera surveys inform models of current and future vertebrate distributions in a changing desert ecosystem. *Divers. Distrib.* **25**, 1441–1456 (2019).

47. S. S. Sethi, M. Kovac, F. Wiesemüller, A. Miriyev, C. M. Boutry, Biodegradable sensors are ready to transform autonomous ecological monitoring. *Nat. Ecol. Evol.* **6**, 1245–1247 (2022).

48. J. W. Fitzpatrick, F. Gill, M. Powers, J. V. Wells, K. V. Rosenberg, Introducing eBird: The Union of Passion and Purpose.

49. OpenStreetMaps Contributors, OpenStreetMaps Norway Dataset, Geofabrik (2022); https://download.geofabrik.de/europe/norway.html.

50. S. S. Sethi, N. S. Jones, B. D. Fulcher, L. Picinali, D. J. Clink, H. Klinck, C. D. L. Orme, P. H. Wrege, R. M. Ewers, Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. *Proc. Natl. Acad. Sci.* **117**, 17049–17055 (2020).

51. S. Sethi, Bugg, version 1 (2022); https://github.com/bugg-resources.

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

53. K. Lisa Yang Center for Conservation Bioacoustics, BirdNET-lite, (2022); https://github.com/kahst/BirdNET-Lite.

54. J. S. U. Hedlund, S. Jakobsson, C. Kullberg, T. Fransson, Long-term phenological shifts and intra-specific differences in migratory change in the willow warbler *Phylloscopus trochilus*. *J. Avian Biol.* **46**, 97–106 (2015).

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

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

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

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

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

60. Kartverket (Norwegian Geographical Survey), Norway National Detailed Elevation Model, Kartverket; https://hoydedata.no/LaserInnsyn2/.

Materials and 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 (*40*). When possible, sites were chosen within 200m of accessible roads (*49*) 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). The different numbers of recorders per region were due to physical and logistical difficulties installing records in certain locations, including land access or unavailable sunlight. While uncorrected spatial biases could exist in our analyses, we do correct for total number of hours recorded per region to partially alleviative this, as described below in the description of methods for estimating species arrival.

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, see Sethi et al 2020(*50*)).

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. The Bugg recorders used a built-in Knowles SPH0641LU4H-1 omnidirectional PDM MEMS microphone, with a 100Hz-80KHz dynamic range, which was weatherproofed using an acoustic ePTFE membrane (*51*).

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 (*52*). 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 (*53*) to extract and identify vocalizations to species level in near real-time (*15*). In total, 176 unique species were identified in this dataset by BirdNET-lite. Of these, 67 species had at least 50 detections with a model confidence over 0.8. For this subset, we provided an expert ornithologist with 50 random detections for each species, sampled across all sites without stratification, with each species taking approximately 20-30 minutes to validate. Of these, BirdNET-lite correctly identified 57 species with 80% or greater precision. 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. In the Northern regions of Finnmark and Tromsø, recorders were installed in early May.

With snow still covering the area, this temporal discrepancy is unlikely to miss observations–the first Willow Warbler in Tromsø was detected in mid-May, agreeing with previous manual estimates of first arrival (*36*).

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 (*54*). 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. We then plot arrival curves along with normalized difference vegetation index (NDVI), sourced from the daily 0.05 degree resolution gap-filled MODIS Vegetation Index dataset (*35*). 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 (*31*) were aggregated within each 90x90 km region, then filtered by several criteria to standardize data (*9*). 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 (*9*). Duplicate surveys that appear in group surveys were dropped. Lastly, surveys in non-forested areas, as identified by the Copernicus data set (*40*), 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 (*6*) 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, visiting each point only once (*6*). 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 (*55*) to predict the probability of vocalization (*28*) for Willow Warbler and Spotted Flycatcher. 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 (*35*), maximum (*56*), minimum (*57*), and mean (*58*) temperature, and precipitation (*59*), as well as site elevation (*60*) 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 (*35*) 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 2.

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) (approximately 10 km) and used to predict probability of vocalization across Norway. To restrict our vocalization probability maps to ecosystems similar to our study sites, we employ a masking step to filter out non-majority-forested cells from the aSDM based on the 300m resolution Copernicus land cover dataset (*40*). 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. Total detections of Willow Warbler (WW), Spotted Flycatcher (SF), and Common Chiffchaff (CC) by study region and monitoring method.

Region	Species	BBS Detections	eBird Detections	Audio Detections
Kristiansand	WW	95	34	425
Bergen	WW	65	1	2431
Trondheim	WW	88	5	158
Tromsø	WW	36	1	124
Stavanger	WW	48	2	29
Moss	WW	88	18	41
Lillehammer	WW	58	0	3
Kristiansand	SF	4	4	97
Bergen	SF	2	2	88
Trondheim	SF	3	0	4
Tromsø	SF	0	0	0

Stavanger	SF	0	0	8
Moss	SF	13	10	1
Lillehammer	SF	0	0	0
Kristiansand	CC	9	2	1
Bergen	CC	8	2	339
Trondheim	CC	51	9	151
Tromsø	CC	0	5	0
Stavanger	CC	2	2	0
Moss	CC	6	7	0
Lillehammer	CC	3	0	0

SI Table 2. Audio Species Distribution Model (aSDM) model hyperparameters, as well as test and train set mean absolute error (MAE) for Willow Warbler and Spotted Flycatcher. 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	Random Forest Parameters
Willow Warbler (*Phylloscopus trochilus*)	0.0761	0.226	{'n_estimators': 386, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'log2', 'max_depth': 80, 'bootstrap': True}
Spotted Flycatcher (*Muscicapa striata*)	0.0301	0.0704	{'n_estimators': 176, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': None, 'max_depth': 100, 'bootstrap': True}