Bird Census News

Journal of the European Bird Census Council

www.ebcc.info

2025

Volume 38 n°1–2

Bird Census News

2025, volume 38 n° 1–2 (published January 2026)
ISSN 1381-5261
Free download pdf from www.ebcc.info

Bird Census News is the Journal of the European Bird Census Council or EBCC. The EBCC exists to promote the organisation and development of atlas, census work and population studies in all European countries; it promotes communication and arranges contacts between organisations and individuals interested in census and atlas work, primarily (but not exclusively) in Europe.

Bird Census News reports developments in census and atlas work in Europe, from the local to the continental scale, and provides a forum for discussion on methodological issues.

Cover illustration by Evgeniy Koblik

Bird Census News
Volume 38/1–2, January 2026

EDITORIAL

The technical development in automatic sound identification has been progressing very fast over the last ten years. This has increased the interest in passive acoustic monitoring (PAM) among birds and other taxa. The cost of sound recorder devices have come down and there are currently several online tools available to automatically suggest the identity of the recorded bird sounds. This makes use of PAM possible for a wider range of researchers. There are already many groups using PAM in their research in Europe. What is the current state of the art in PAM? How good are available acoustic tools and how could PAM be used in bird monitoring? To answer for example these questions and to keep on track in the recent development in this research field, EBCC board decided to establish an Acoustic Monitoring Group (AMOG) in 2024.

This Bird Census News volume is dedicated for articles about acoustic monitoring including eight articles. The first article introduces the AMOG, the newest member of the EBCC family. The second article is based on a questionnaire from the EBCC to the European bird monitoring coordinators to ask how they see the role of acoustic monitoring in bird surveys. The following five articles are national examples how PAM could be used to support monitoring. Two of these articles are case studies from the Netherlands (Roodbergen et al.) and the UK (Wilson et al.) that are exploring how acoustic monitoring could be used in common bird monitoring. Two other articles are single species case studies from Switzerland (Kornienko et al.) and Spain (Klaas-Fábregas et al.) exploring the acoustic monitoring of nocturnal owl species. One of the challenges in PAM is to evaluate accurate location of the vocalising bird and thus estimate potential abundances of species. The case study of Bruggemann et al. from Germany using multiple recorders simultaneously is investigating this issue. The last acoustic article by Ovaskainen et al. is about solving large scale problems in acoustic monitoring: improving (i) classifiers and (ii) the workflow from recordings to the analyses as well as (iii) engage citizen scientist to collect audio data on a larger spatial scale. The first and the last audio article also encourages international collaboration among the EBCC network and we can only strongly agree! I hope these papers will stimulate new approaches and initiatives in the field of acoustic monitoring. The final article of the volume is introducing the secretary of the EBCC board: Dawn Balmer.

Last, we would like to inform that the new editor in chief of the BCN will be Dimitrije Radišić. Please, contact if you wish to publish your work in the journal."

Aleksi Lehikoinen & Danae Portolou

CONTENTS

Introducing AMOG — the newest member of the EBCC family!

Thomas Sattler[1] and Maja Roodbergen[2]

[1] Swiss Ornithological Institute, Seerose 1, 6204 Sempach, Switzerland
[2] Sovon Dutch Center for Field Ornithology, Nijmegen 6525, The Netherlands

Acoustic monitoring — using hardware and software to record and identify bird species — has undergone tremendous development over the past 15 years. In particular, Passive Acoustic Monitoring (PAM), which uses autonomous recording units capable of collecting data for weeks or even months, has the potential to revolutionize bird monitoring, as well as the monitoring of other taxa.

Several organisations within the EBCC network have begun to explore — or even actively implement — these innovative acoustic monitoring methods in their work. Alongside the exciting opportunities this brings, acoustic monitoring also presents a range of challenges that may affect current and future bird monitoring practices.

To address these developments and encourage knowledge exchange among its partner organisations, the EBCC established the Acoustic Monitoring Group (AMOG) in June 2024, following an initiative from the EBCC board. AMOG's goal is to coordinate, align, and integrate acoustic bird monitoring initiatives across Europe, strengthening the overall efforts in bird monitoring. If necessary, AMOG will also advise the EBCC board on matters related to acoustic monitoring.

As its first initiative, AMOG distributed a questionnaire on the perception, use, and future potential of acoustic monitoring among EBCC partners, receiving responses from 47 individuals across 32 countries. The results were presented at the EBCC Conference 2025 in Riga and are summarised in this issue of *Bird Census News* (p. 4).

Responding to growing interest and requests from EBCC partners, AMOG also drafted the EBCC Statement on the Use of Auto-ID Tools in Long-Term Monitoring. At present, AMOG recommends refraining from using identification apps such as *Merlin Bird ID* or *BirdNet* in Breeding Bird Surveys (see p. 5 of this BCN). This recommendation is based on concerns that these tools may introduce biases of unknown magnitude into long-term datasets. Before such tools are widely adopted, we must first understand their implications and develop clear guidelines for their use.

At the EBCC Conference in Riga in April 2025, AMOG organised a workshop on different aspects of acoustic monitoring, which was well-attended (approximately 120 participants). This strong turnout reflects the significant interest in the topic within the EBCC network. Participants generally agreed on the great potential of acoustic monitoring for studying elusive species, remote habitats, or generally regions with limited observer coverage. However, there was also consensus that the EBCC and its network are not yet focused on deriving abundance data (e.g. densities) through PAM. While this area is scientifically promising and rapidly evolving, it is not yet ready for large-scale monitoring applications. Workshop participants encouraged AMOG to develop a comprehensive strategy for how the EBCC can contribute to the field of acoustic monitoring, and to prepare practical guidelines for organisations interested in launching their own PAM initiatives.

Currently, AMOG includes 21 members from 15 institutions. If your national bird monitoring organization has some experience with acoustic monitoring and is interested in joining and supporting AMOG, please contact the Co-chairs Aleksi Lehikoinen (aleksi.lehikoinen@helsinki.fi) or Thomas Sattler (thomas.sattler@vogelwarte.ch), including a brief description of your interests and current activities related to acoustic monitoring.

Received 23rd June 2025
Accepted 6th October 2025

Bird Census News 2025, 38/1–2: 4–13

Role of acoustic monitoring in bird surveys and atlases: results of EBCC acoustic monitoring questionnaire

Petr Voříšek[1] and Tatiana Kornienko[2]

[1] European Bird Census Council, Na Bělidle 34, 15000 Prague 5, Czechia
[2] Swiss Ornithological Institute, Seerose 1, CH-6204, Sempach, Switzerland
vorisek@ebcc.info

Abstract. In order to understand the importance and future potential of the currently quickly developing acoustic methods of research, the questionnaire with 36 questions about acoustic monitoring and its role was distributed in Autumn 2024 among the EBCC network. Answers of the respondents from 32 countries showed the interest of the EBCC community in this topic and a lot of possibilities for implementation of acoustic monitoring methods within EBCC. Some countries use it a lot already, some haven't started yet, so EBCC can bring them together for effective and productive collaboration to improve the data sets on bird distribution and abundance across Europe.

Introduction

Acoustic monitoring plays an increasingly important role in ornithological studies. Thousands of recorders are deployed around the world, with high coverage in Europe — for example only for citizen science BirdWeather PUC devices the known number of active stations is more than 1200 items (BirdWeather). The huge amount of AudioMoth, SongMeter Mini, SongMeter Micro is difficult to estimate not only because there are a lot of official projects (some of them can be found here: ecoSound-web), but also due to many private devices. They are especially important to reveal the status of cryptic and difficult to study species either diurnal like Rock Ptarmigan or nocturnal like owls, in remote, effort-consuming areas, regions where there is a lack of ornithologists etc. As the goals of EBCC first of all are "to bring together ornithologists who are interested in the study of distribution, numbers, and demography of European birds and to encourage monitoring of bird distribution, numbers, and demography aimed at improving their conservation and management", it is important to estimate the scale of using acoustic monitoring for all those activities. The questionnaire, which was sent to the EBCC network, was aimed to do it. On the other hand, the important task of EBCC is encouraging and creating specialist working groups to tackle relevant topics, that is why AMOG — Acoustic Monitoring Group — was created (see Bird Census News p. 3, of this volume) to bring together acousticians from all around Europe and to develop acoustic methods which can be integrated into current EBCC monitoring activities.

Materials and methods

The questionnaire was distributed among EBCC national delegates and national coordinators of PECBMS and EBBA2 projects in September 2024. The questionnaire was also advertised in the EBCC Newsletter, which is distributed quarterly among a wider audience interested in bird monitoring work. 47 responses were received from 32 countries, including 43 organisations and four private persons. From most of the countries (65% of all of them) just one answer was received, one quarter of the countries provided two answers, two countries gave three answers, and it was one country which provided four answers from 2 different organisations.

The questionnaire included 36 questions (see Annex). There were different kinds of them:

- multiple choice restricted with some particular options, so the respondent has to choose one (Q. 1, 2, 3, 4, 31).
- score questions restricted with the answer in certain numbers (Q. 5–30, 32) — mostly about importance of acoustic monitoring.
- open-ended, so the respondents could answer in a free manner (Q. 33–36) — comments, suggestions etc.

Results and discussion

The questionnaire results show that there is an interest in developing and using methods of acoustic monitoring that prevails among the survey participants. All the questions and distribution of the replies are shown in the Annex. In open-ended questions some relevant comments have been shown. According to the responses the potential of acoustic monitoring lies in complementing existing monitoring methods and schemes, and the complementarity can also mean that acoustic monitoring brings data on presence/absence of species in remote areas, areas with few birdwatchers, or for species difficult to detect. There are positive expectations from acoustic monitoring playing a role in both atlases and monitoring, but most of the respondents believe that acoustic monitoring gives more valuable data for atlases and inventory than for monitoring as monitoring is about numbers, abundance and trends. Despite the fact that AI Identifiers are developing very fast nowadays, they are still weak in quantity estimation and most probably it will take a few more years to train them to identify not only the presence/absence of the bird species, but also the number of individuals, if at all. In general, caution prevails among the respondents about using ID apps by fieldworkers. The risks are, among others, the data being not compatible with the data obtained by traditional methods. Despite the accuracy of identification for some species being very good, there are species which are almost invisible for classifiers even if the quality of recordings is high, and validation and data processing is time consuming. That is why “blind” use of ID Apps is currently not recommended by EBCC while making surveys for PECBMS or EBP. In some cases it was possibly one of the reasons for the answer “No” on the question about general interest in acoustic monitoring. On the other hand, even if there was no interest in acoustic monitoring from the organisation, acoustic methods were considered as useful in some study areas.

Most of the respondents emphasised that acoustic monitoring was especially important for cryptic, rare and nocturnal species research (including nocturnal migration) and for research with dedicated topics like vocalisation, phenology, individual recognition (for some species individual marking can be done as the peculiarities of the song are easier to detect than catching/ recatching the same individual to check the ring number) etc. and it makes research together with other taxa more efficient as just only one method/device is used for bats, birds, mammals, frogs and insects. The answers have shown that using acoustic devices will probably not significantly reduce the travel costs or costs for volunteer training — despite even one recorder replaces human hours in the field, the amount of the recorders is increasing, and travel costs are needed not only to deploy and bring them back to the office, but also to maintain them like changing batteries or memory cards. Furthermore, the computational capacity to analyse the data will also be higher than analysing traditional survey data.

Nearly half of the respondents think the acoustic monitoring will change the community of the fieldworkers. For instance, by opening monitoring to people with lesser expertise in bird ID (incl. foresters or experts on other taxa). There are examples of this type of work from some countries when recordings for common bird species monitoring are mostly done by protected areas staff, and further identification is run by professional ornithologists. It allows large areas to be covered and significant amounts of data collected. On the other hand, huge amounts of data must be analysed, verified and stored somehow, and it makes it complicated for the countries which either have lack of resources for that or collect that much data by implementing acoustic monitoring a lot.

There were no worries that the attitude of the bird watcher community would be negative, but some doubts regarding engagement of people with birds and nature — no need to spend hours in the field for bird counts, all the fieldwork is about deploying the recorders and maintaining them. The good aspect of it is that acoustic devices can be installed under any weather conditions and at any time of the day, in some cases months in advance if needed, so all the problems with finding the best time for the survey can be skipped — later the researcher can decide which files to use depend on the aim of the study. On the other hand, there is a concern about losing a link between the work and the community of fieldworkers.

Most of the respondents have expressed no concern about privacy, although in many countries it is forbidden by the legislation to record people’s voices without their permission. In accordance with the answers given, the safety of equipment also is not something that is too concerning —

most modern devices are small, cryptic-coloured and easy-to-hide. As prices significantly decreased and quality increased (high humidity protection, solar panel for long-lasting, live data transfer throughout GSM or satellite like in Eco-Pi (Recording technology — ecoPi) and similar devices) within the last five years, they became more affordable and more widespread.

Nobody of the respondents expects that acoustic devices will completely replace standard human-based monitoring in the longer term. However, some level of complementarity is expected — in remote and difficult to access areas, regions where there is a lack of ornithologists or birdwatchers, within the work with rare/cryptic/nocturnal species and other cases which demand a lot of efforts to collect the data by humans. At the moment there are a lot of projects among EBCC partners, in which acoustic monitoring is already used. These are ones mentioned by the respondents: some separate species monitoring like Rock *Lagopus muta* and Willow Ptarmigan *Lagopus lagopus*, White-winged Snowfinch *Montifringilla nivalis*, nightjars *Caprimulgus* sp., Eurasian Woodcock *Scolopax rusticola*; multiple species monitoring like common bird surveys, farmland birds, nocturnal migration and nocturnal breeding species, including owls. Nearly half of the respondents does not expect any support from EBCC in these activities, but others believe that EBCC can:

- develop guidelines, education and trainings;
- facilitate and implement projects focusing on acoustic monitoring;
- stimulate discussion within EBCC on how to integrate acoustic monitoring in current monitoring schemes;
- become the European reference point for standards on how to mobilise/integrate acoustic data in long term monitoring projects.

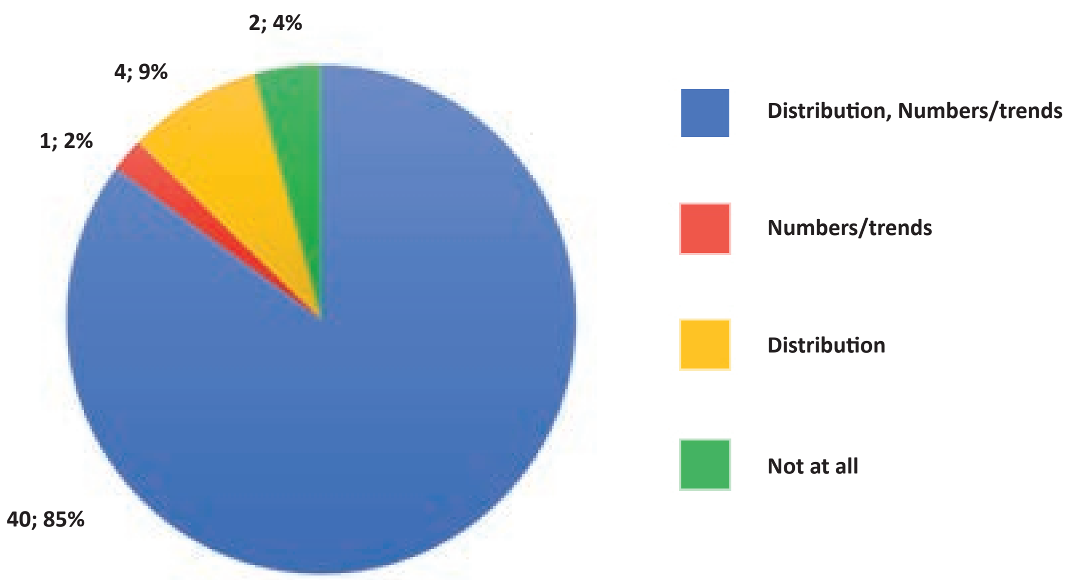
References

ecoSound-web: https://ecosound-web.de/ecosound_web/
BirdWeather: https://www.birdweather.com/
Recording technology — ecoPi: https://www.oekofor.de/en/portfolio/erfassungstechnik_en/

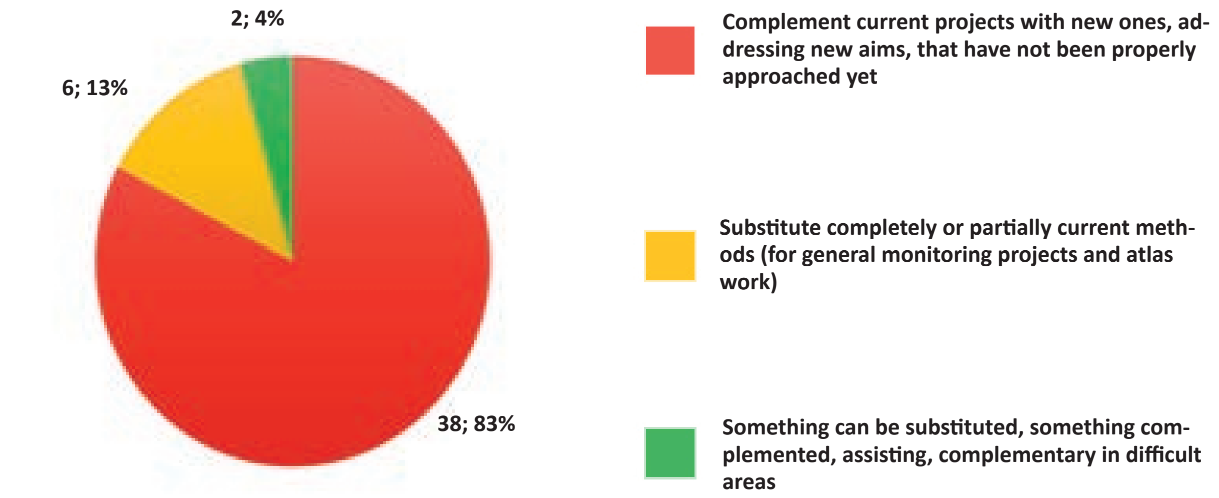
Received: 30th May 2025
Accepted: 17th July 2025

Annex
Questions and results

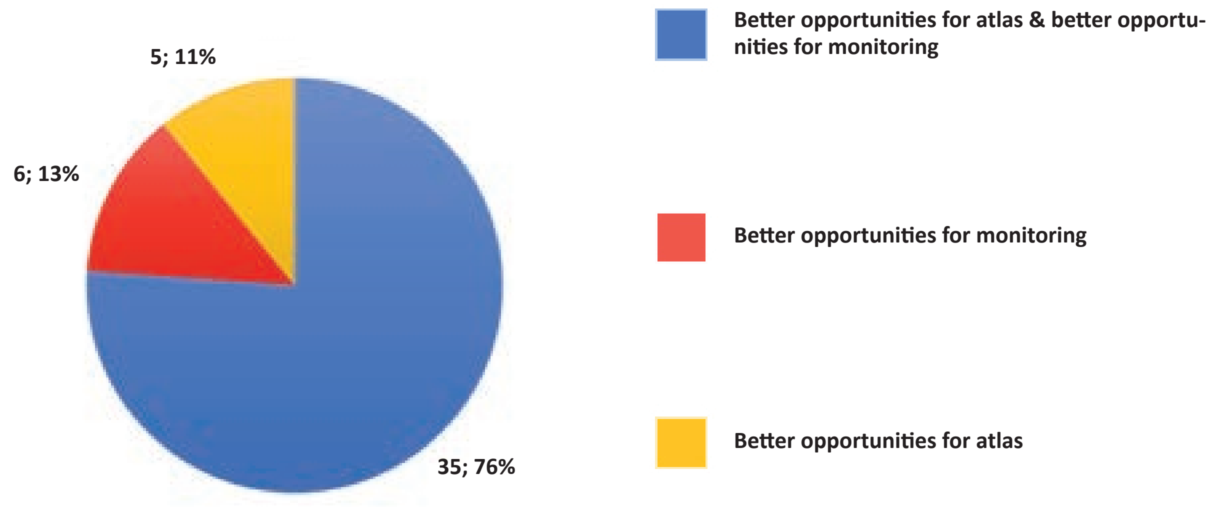
1. Is your organisation interested in acoustic monitoring to monitor bird populations?



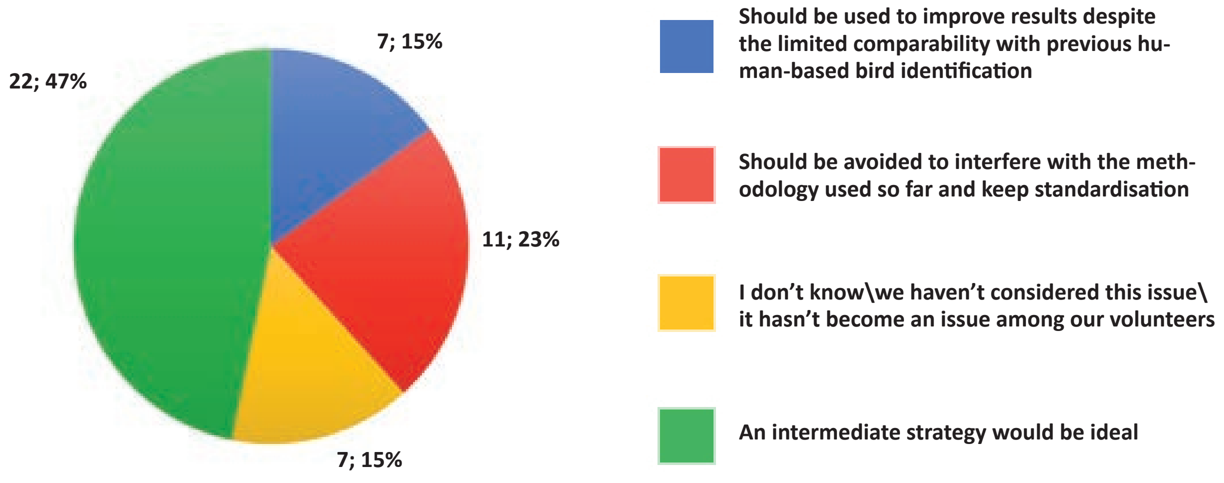
2. In your view, what are the main expected outputs from acoustic monitoring? Will it complement or substitute completely or partially current methods (for general monitoring projects and atlas work)



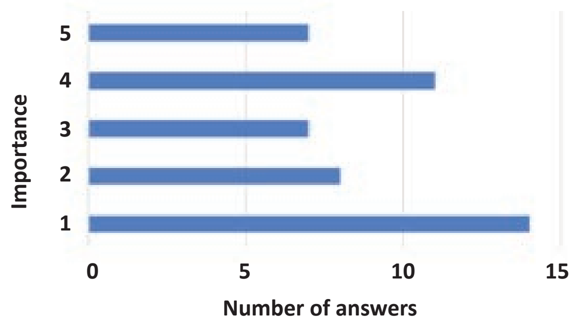
3. What do you think will acoustic monitoring bring to the regular monitoring (temporal dimension) or atlas work (spatial dimension)?



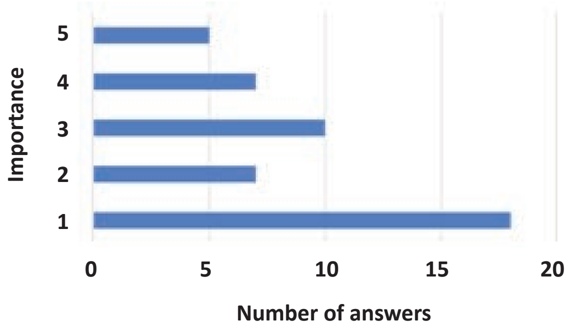
4. What is your view on how volunteers can use auto-identification apps during fieldwork?



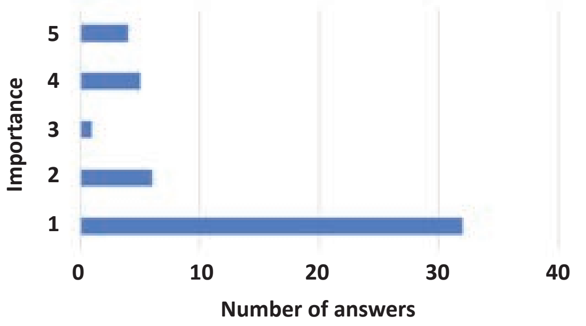
5. Please score the following monitoring activities for how much you think acoustic monitoring devices will be complementary to current projects (1 = important to 5 = unimportant) [Breeding bird monitoring in areas with a limited number of ornithologists and birdwatchers (a person non-trained in bird identification could contribute to monitoring)]



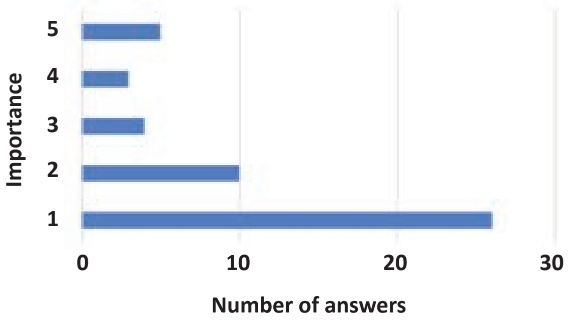
6. Please score the following monitoring activities for how much you think acoustic monitoring devices will be complementary to current projects (1 = important to 5 = unimportant) [Atlas/inventory work in areas with a limited number of ornithologists and birdwatchers (a person non-trained in bird identification could contribute to monitoring)]



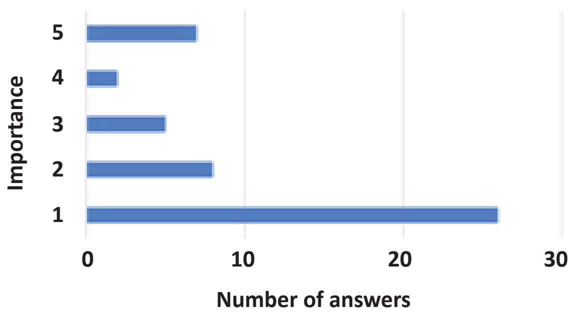
7. Please score the following monitoring activities for how much you think acoustic monitoring devices will be complementary to current projects (1 = important to 5 = unimportant) [Dedicated research projects e.g. focussed on particular sites or species]



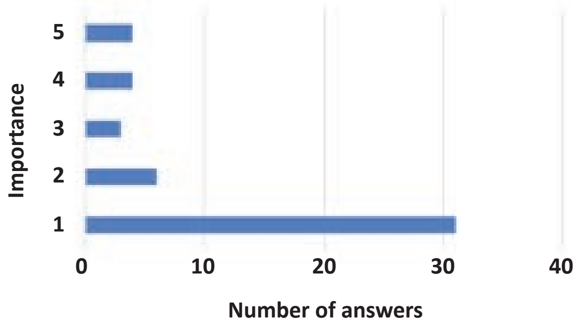
8. Please score the following monitoring activities for how much you think acoustic monitoring devices will be complementary to current projects (1 = important to 5 = unimportant) [Nocturnal breeding bird species]



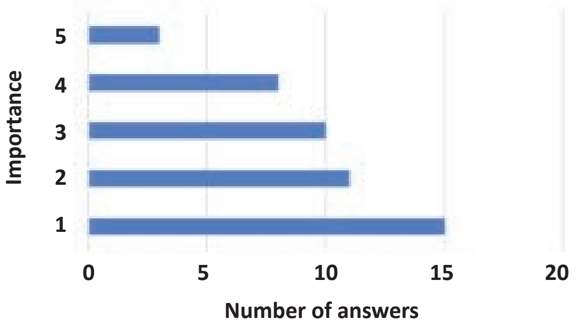
9. Please score the following monitoring activities for how much you think acoustic monitoring devices will be complementary to current projects (1 = important to 5 = unimportant) [Nocturnal migration]



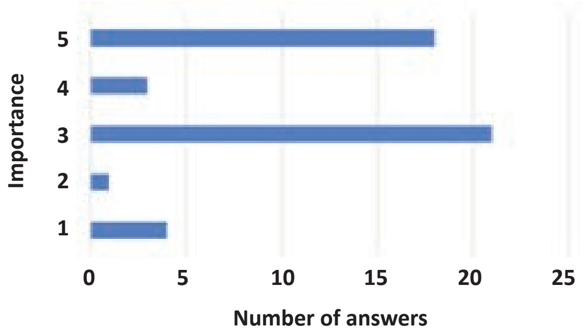
10. Please score the following monitoring activities for how much you think acoustic monitoring devices will be complementary to current projects (1 = important to 5 = unimportant) [Monitoring of some rare or cryptic species]



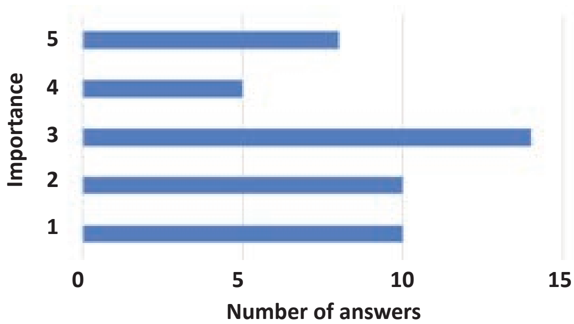
11. Please score the following monitoring activities for how much you think acoustic monitoring devices will be complementary to current projects (1 = important to 5 = unimportant) [Complementarity with the monitoring of other taxa]



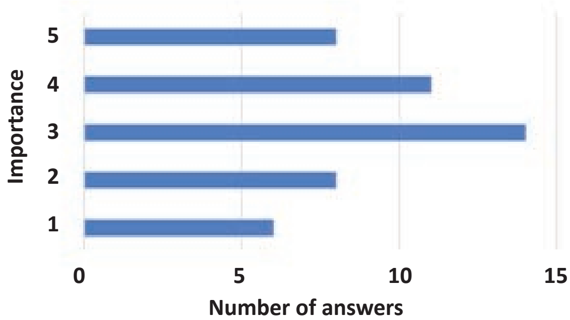
12. Please score the following monitoring activities for how much you think acoustic monitoring devices will be complementary to current projects (1 = important to 5 = unimportant) [Other (please specify in comments)]



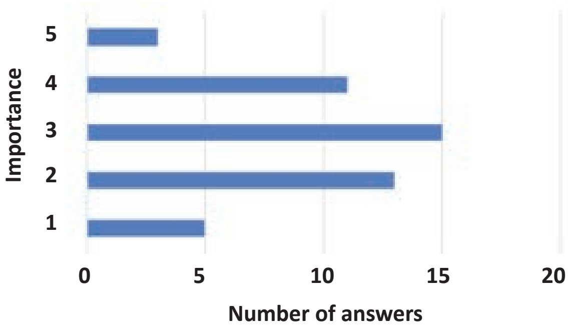
13. Please score the following for how important they are in shaping the value of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Reduced travel costs]



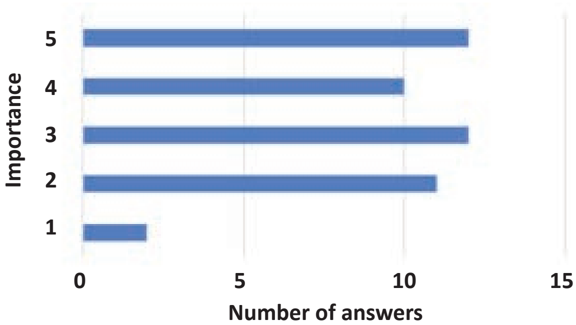
14. Please score the following for how important they are in shaping the value of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Reduced costs of volunteer training]



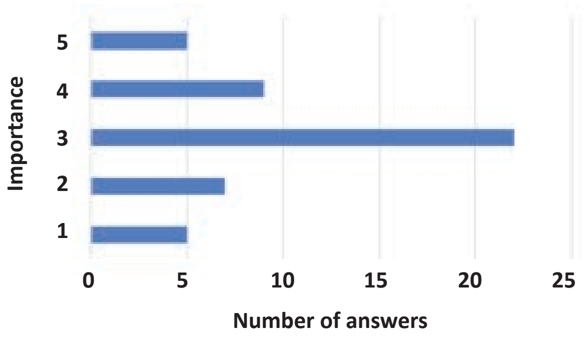
15. Please score the following for how important they are in shaping the value of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Increased accessibility allowing greater diversity of volunteers]



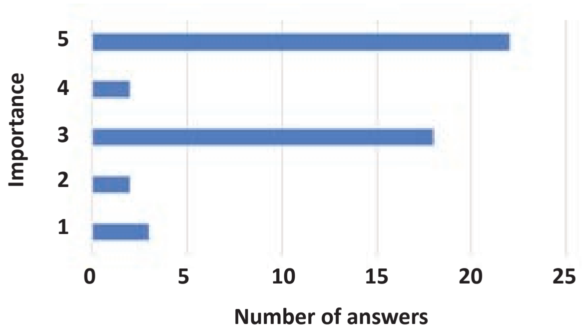
16. Please score the following for how important they are in shaping the value of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Other]



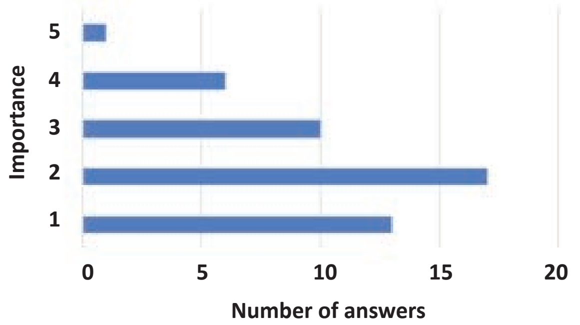
17. Please score the following for how important they are in shaping the value of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Positive attitude of the bird watcher community]



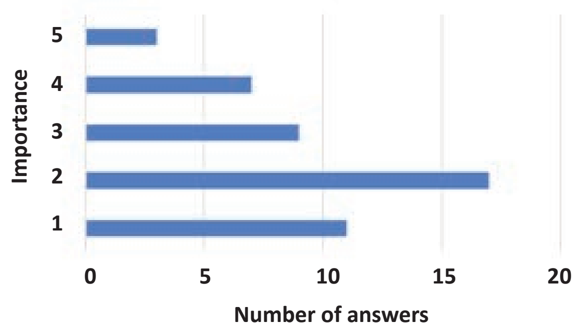
18. Please score the following for how important they are in shaping the value of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Other]



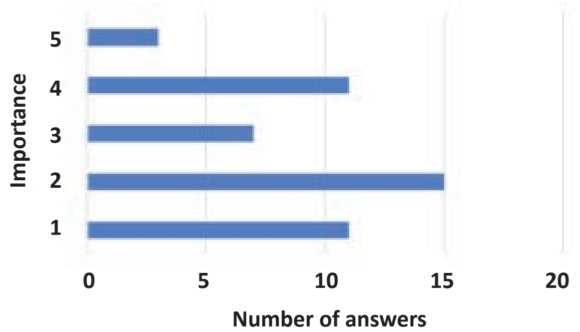
19. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Cost of devices]



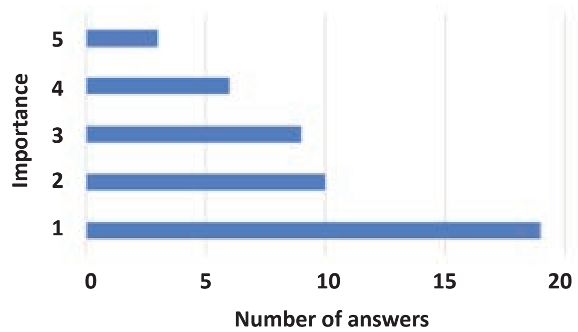
20. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Cost of IT infrastructure]



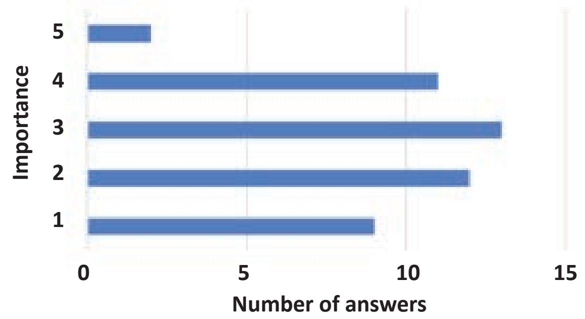
21. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Cost of storing data]



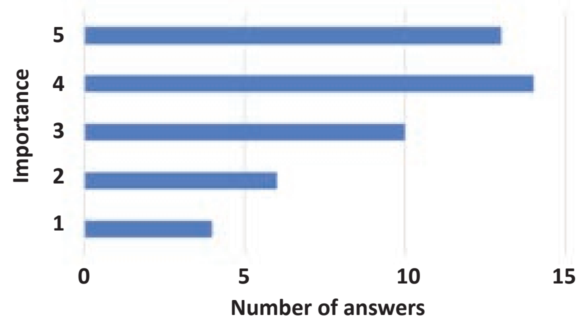
22. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Cost of processing and verifying data]



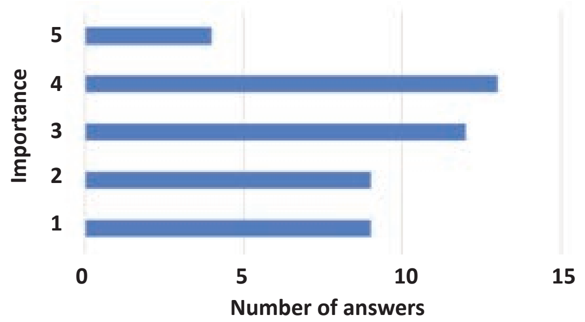
23. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Lack of in-house expertise for acoustic projects and data management]



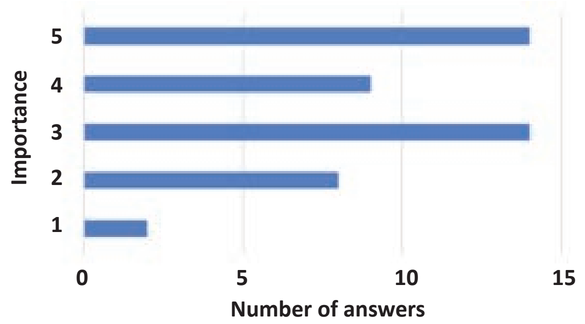
24. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Concerns about privacy]



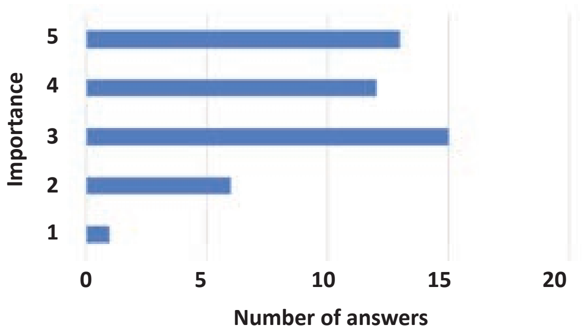
25. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Safety of equipment]



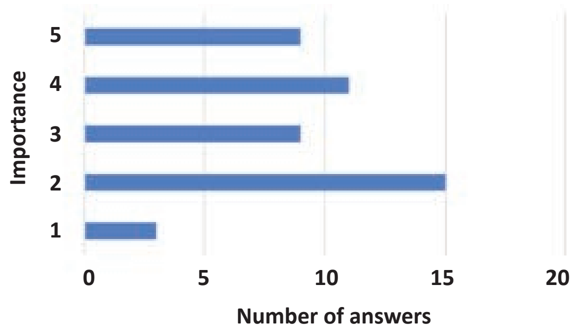
26. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Diminished support by sponsors]



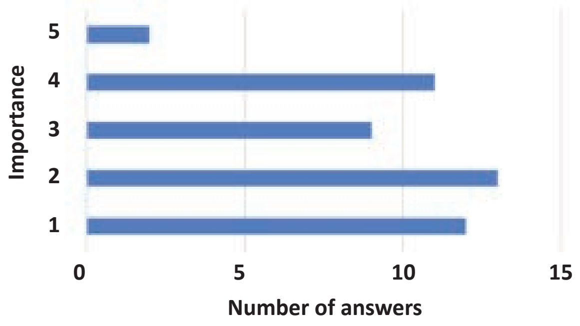
27. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Negative attitude of the bird watcher community]



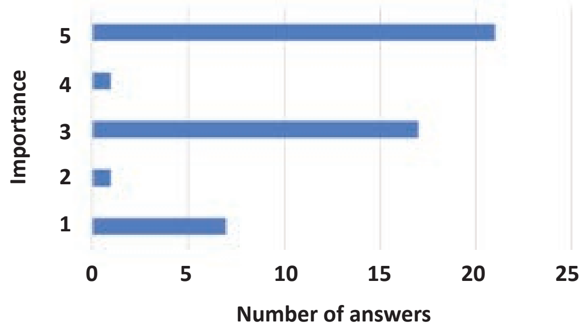
28. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Reduce engagement of people with birds/nature]



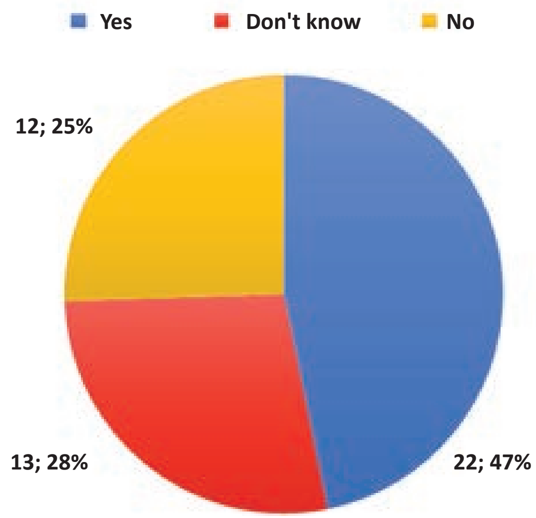
29. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Limited comparability with the data collected so far by humans]



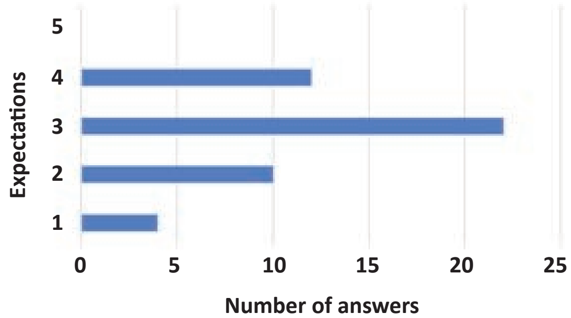
30. Please score the following for how important they are in constraining the use of acoustic monitoring in bird monitoring (1 = important to 5 = unimportant) [Other (please specify in the Comments)]



31. Do you think acoustic monitoring development would change the type of collaborators in fieldwork? If yes, please elaborate in the Comments.



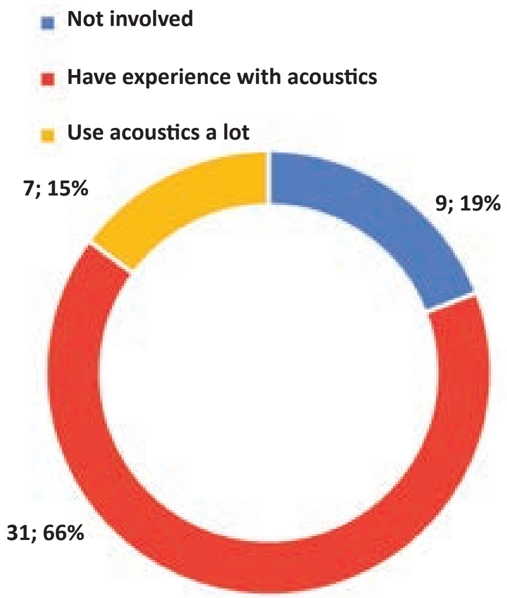
32. Evaluate from 1 (no expectation at all) to 5 (complete replacement of current monitoring human-based approach) your 20-year future vision on acoustic monitoring in bird monitoring and atlas work. Please comment on it in the Comments.



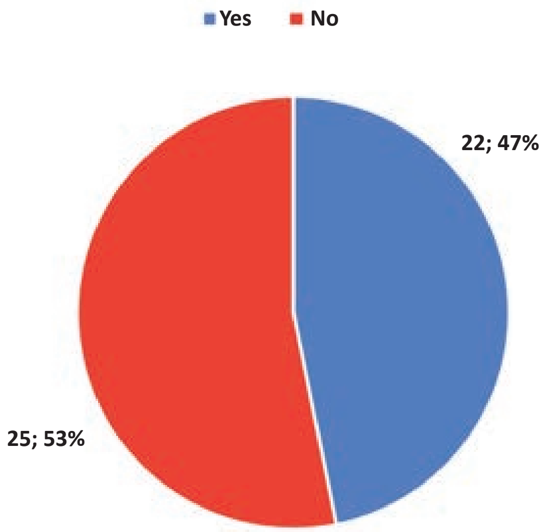
33. Is there anything that EBCC could do to support you/your organisation/scheme in the implementation of acoustic monitoring?
Some remarkable comments are given below:

- develop guidelines, education and trainings;
- facilitate and implement projects focusing on acoustic monitoring;
- stimulate discussion within EBCC on how to integrate PAM [Passive acoustic monitoring] in current monitoring schemes;
- become the European reference point for standards on how to mobilise/integrate acoustic data in long term monitoring projects.

34. Is your organisation already involved in monitoring work which uses acoustic monitoring? If yes, please provide answers according to the following structure directly in the text. If there are more projects, please add info on all of them.



35. Are you aware of any consultants undertaking acoustic monitoring in your country? If yes, please provide comments, if possible.



36. Comments

Some remarkable comments are given below:

- *'…We are very positive to using acoustic techniques for very specific and local surveys. But for long-term large-scale monitoring — DON'T GO THERE! '*
- *'My biggest fear with the development of passive acoustic monitoring is the loss of the link with the work of volunteers and the grassroots community. The main challenge is to manage to combine the two in harmony.'*
- *'…Consistent data, which is not dependent on the observer's abilities and skills, gathered for longer periods of time, fully comparable between different sampling places. It feels like a right way to the general aim of any science — finding out the truth.'*

Bird Census News 2025, 38/1–2: 14–25

Applying Passive Acoustic Monitoring for Dutch common bird species: evaluating detection and classifier performance

Maja Roodbergen, Chiel Boom and Niels van Harten

Sovon Dutch Centre for Field Ornithology, Postbus 6521, 6503 GA Nijmegen, The Netherlands
maja.roodbergen@sovon.nl

Abstract. To evaluate the usability of audio-recordings for monitoring, we recorded audio data during 232 ten-minute point counts in agricultural habitat. All recordings were automatically classified, using four classifiers, and 70 recordings were annotated manually. Finally, two new classifiers were developed by training an existing classifier using annotated recordings of 326 species occurring in the Netherlands, with and without use of secondary species labels. Here we show the results of comparisons between 1) the field data of the point counts vs the manually annotated audio data, and 2) the automatic classifications by the four existing and the two new classifiers.

Introduction

Being a densely populated country, with many interested and skilled birders, monitoring of bird numbers and distribution is extensive in the Netherlands. Monitoring schemes are coordinated by Sovon Dutch Centre for Field Ornithology and carried out by c. 10,000 volunteers, which go out in the field to count birds. Though extensive, some species, habitats and/or periods are underrepresented, due to (e.g.) detectability, accessibility and attractiveness issues. Technological advances in low-cost automatic audio recording devices in combination with developments in AI facilitating automatic sound recognition, have opened the potential for acoustic monitoring (e.g. Browning et al. 2017, Shonfield & Bayne 2017).

However, being a relatively new method, exploring the comparability of acoustic and regular existing monitoring methods is pivotal to assess the applicability of acoustic monitoring and to avoid methodological trend breaks. In addition, classifying audio recordings manually is time consuming. Automatic classification using classifiers, developed and trained for this purpose (e.g. Kahl et al. 2021), could greatly improve efficiency. However, thus far, classifiers have been trained using species-specific recordings and may perform inadequately for soundscapes with multiple species vocalising simultaneously as happens during spring choruses. Analyses using classifiers which miss and/ or misidentify vocalising local species produce incorrect presence-absence data. Moreover, species differ in their detectability and identifiability, making between species comparisons difficult. Validating present, freely available classifiers is therefore a prerequisite for their practical implementation.

A habitat type that is typically underrepresented in Dutch monitoring schemes is intensively used farmland, as it contains few and usually only common species and therefore is of little interest to birders. As it is dynamic and constitutes a large part of the land surface of the Netherlands, improving coverage of this habitat type is desirable. Breeding bird monitoring in agricultural sites in the Netherlands to a large degree consists of point counts (Teunissen et al. 2019). Data from audio recordings from a static location can best be compared to point counts (e.g. Klingbeil & Willig 2015, Van Wilgenburg et al. 2017). Therefore, while performing point counts in the agricultural landscape of the province of Noord-Brabant, field observers simultaneously recorded bird sounds using a recorder. We used these data to answer the questions: 1) 'How do presence-absence data collected in agricultural landscapes using audio recorders differ from data collected using point counts?'; 2) 'What is the performance of four freely available bird sound classifiers when it comes to confirming the occurrences of farmland bird species?' and

3) 'Can the performance of the best of these four classifiers be improved for Dutch bird species?'

Methods

Data collection

In 2022, four different field workers performed a total of 232 point counts in 13 agricultural sites with Agri-Environment Schemes in the province of Noord-Brabant, with 5 points per site, repeated 4 times during the breeding season. Counts were performed between sunrise and 5 hours after sunrise and lasted 10 minutes, during which all individuals tied to the location (i.e. excluding fly-overs) within a radius of 300 m around the point were registered on a map, noting species and breeding code (Teunissen et al. 2019).
During the point counts, bird sounds were recorded using an AudioMoth located at observer height and set at a sample rate of 48 kHz and medium gain.

Evaluation data set

For the comparison of audio data with point counts and the performance tests of the classifiers, 70 audio recordings were selected (Table 1). We aimed at including all sites and all four rounds, but prioritised two sites where species diversity was relatively high (Maasheggen and Schijndel). The selected audio-recordings were annotated manually by three expert field workers from Sovon Dutch Centre for Field Ornithology, by composing a list of all species heard during the 10-minute recordings.
The 70 recordings used in the analyses contain 759 minutes of audio with 864 annotations for 73 species, after removal of all uncertain annotations. The Common Chaffinch *Fringilla coelebs* was the most frequently recorded species appearing in 63 out of 70 recordings, while 13 species only occurred in a single recording. The soundscapes included between 4 and 22 annotations, averaging 12 annotations per soundscape. All audio had a bit rate of 768 kbps, though the quality of the recordings varied. Some soundscapes suffered from disturbances from wind, traffic or field observers.
The evaluation dataset used to answer the first question (comparing audio with point counts) contained 1332 records (unique species-point-visit combinations, Table 1) from 91 species that were present on the audio-recording and/or in the point count dataset.

Table 1. Number of records (unique species-point-visit-combinations) in the evaluation dataset per point and visit (records from both point counts and audio-recordings).

		Visit			
Site	Point	1	2	3	4
De Bleken	dblk1	0	0	16	17
Gastelse Heide	gsth1	0	12	20	13
	gsth2	0	17	0	0
Keent	kent1	32	31	26	0
	kent2	35	22	0	0
	kent4	26	0	0	0
	kent5	22	0	0	0
Lage Zwaluwe	lgzw1	0	0	22	14
Made Noord	mdnd1	0	0	11	18
Maasheggen	mshg1	10	22	16	17
	mshg2	21	22	22	21
	mshg3	20	21	19	20
	mshg4	20	21	21	19
	mshg5	21	19	16	14
Rucphen Heikant	rcph1	0	0	25	19
Rielsche heide	rlhd1	0	0	0	8
Reek-Schaijk	rsch1	20	19	18	0
	rsch2	0	19	0	0
Schijndel	schd1	25	22	20	15
	schd3	18	17	18	9
	schd4	18	20	23	18
	schd5	16	20	15	12
	schd7	18	20	25	16
Steenbergen Noord	stbn1	0	0	16	10
Zeeland	zlnd1	20	20	17	0
	zlnd2	0	20	0	0

Model descriptions

We compared four bird sound classification models: BirdNET, the Google Bird Vocalization Classifier (GBVC), AvesEcho and Aquila. BirdNET and GBVC are global models, whereas AvesEcho focuses on Europe and Aquila on the Netherlands.
BirdNET (v2.4, Kahl et al. 2021) is a neural network classifying over 6500 bird species. It operates on three-second-long audio segments. The model is trained using recordings from Xeno-Canto, the Macaulay Library of Natural Sounds and most likely additional unknown sources as the model has been updated after publishing. We

used an overlap of two seconds resulting in one prediction for every second of audio. We did not use the available post-processing mask to filter results based on location and time of the year in our evaluation as it might exclude species which could be present. Instead, we used a selection of 326 bird species occurring in the Netherlands.

The Google Bird Vocalization Classifier (GBVC, v4) is trained on Xeno-Canto recordings and classifies over 10,000 bird species based on 5-second-long audio segments. We inputted segments with an overlap of 4 seconds, resulting in one prediction for every second of the recording. In our evaluation, we again limited the model's output to the same 326 species.

Researchers at the Naturalis Biodiversity Center, led by Burooj Ghani, developed a classification model targeted at European bird vocalisations: AvesEcho (Ghani et al. 2023). The model used in this study was still in its development phase (v0) and was a single-label multi-species classification model, meaning it produced a single species prediction given an input segment of three seconds.

Aquila was developed by Aquila Ecology, a small Dutch company developing technological solutions for ecological research. Their classification model performed well for bats (pers. comm A. Krediet) and also classifies 332 bird species that occur in the Netherlands. This species list excludes 28 species which we distinguish for BirdNET, GBVC and AvesEcho. Most of these are rare species which were not present in the evaluation dataset, apart from five species which were present in the evaluation set (45 annotations) but not included in the Aquila classifier: Red Junglefowl *Gallus gallus,* Mandarin Duck *Aix galericulata,* Pheasant *Phasianus colchicus,* Egyptian Goose *Alopochen aegyptiaca* and Greater Canada Goose *Branta canadensis.*

Model evaluation

We evaluated the four models using our evaluation dataset consisting of 70 soundscapes. The models predicted scores for 3 to 10-second-long segments of a soundscape while we required a single score for each species per soundscape. Therefore, we took the maximum score to obtain a single score for each species given a soundscape. To get a single score for the performance of each model on the soundscapes we used the following approach: 1) We calculated the area under the ROC curve (AUC, Hanley & McNeil 1982) for each species present in the evaluation set. The ROC curve is produced by plotting the true positive rate (sensitivity) on the y-axis against the false positive rate (false alarm rate) on the x-axis at varying threshold settings (Hoo et al. 2017). The larger the area under this curve, the better its performance. An AUC score of 0.5 is as good as random guessing while a score of 1.0 describes a perfect classifier for the data. 2) We averaged the AUC scores calculated in step 1 to obtain a single score per model. We call this metric AUC-mean as it is the mean value of the species' AUC scores. This method ignores species absent from the evaluation set as those have no true positive rate. To not ignore the cost of false positives for other species, we used a second metric we call AUC-binary. Here, we first binarized the problem by accumulating the scores for the bird species in the evaluation set into one set, and the scores for all other species into a second. This representation can be used to plot a ROC curve using all scores and obtain AUC-binary by calculating the area under this ROC curve. We also created a detection error trade-off (DET, Martin et al. 1997) plot, which is similar to a ROC curve but uses the false negative rate (miss rate) instead of the true positive rate and the cumulative normal distribution to scale the curve into a more linear representation. The AUC-binary measure is affected by calibration errors as all species share the same threshold setting. Therefore, it measures a combination of the performance for all species and how well the species are calibrated. To get a more detailed picture, we also plot the ROC curve for a few species. The code used for evaluation can be found in our GitHub repository (van Harten 2023).

Model training

Besides comparing existing models using our evaluation set, we trained the Google Bird Vocalization Classifier, the best performing classifier for our data, for Dutch usage. We replaced the default classifier of GBVC, predicting 10932 classes, with our own classifier predicting 326 species and called this model NLC. First, we obtained training data for this classifier from Xeno-Canto and pre-processed it. Subsequently, we trained the classifier, and in the end, we evaluated the model's performance. The code needed to train and evaluate the classifier can be found in our repository (van Harten 2023). We optimized our model trying to maximize the AUC-mean score for the validation split while also preferring sim-

pler approaches and approaches known to help generalization (like drop-out, label smoothing and up-sampling).
We trained a third model called NLC-NoSec. This model is similar to NLC but ignores the secondary labels of Xeno-Canto. We were interested to see how the model performs being ignorant of these background vocalisations during training.

Training data

To train a model fine-tuned on Dutch data, we used both recordings containing bird sounds for all 326 species and recordings without bird sounds. We obtained bird sound recordings from Xeno-Canto, with a maximum of 100 recordings per species. As non-event data, we used the Warblrb10k development dataset for DCASE 2018 Bird Audio Detection task 3, which consists of 8,000 short smartphone recordings from around the UK and includes weather noise, traffic noise, human speech and even human bird imitations. We used nearly 2,000 recordings that do not contain bird vocalisations.
The methods used for data preprocessing, the model architecture and the training procedure are described in the Appendix. The resulting models were again evaluated using AUC-mean, AUC-binary and a DET-plot (see above).

Results

Comparison audio vs point counts

Of the 91 species present on the audio-recordings (manual annotation) and/or in the point count dataset, 65 were recorded with both methods (though not always at the same point-visit combination), 19 were only observed during point counts and 7 only in audio. Point counts resulted in 1112 records, while audio recordings resulted in 857 records (species-point-visit combinations). Of these, 637 records overlapped, being present in both point count and audio data, while 475 records occurred exclusively in the first, and 220 exclusively in the latter.
Excluding species with less than 10 records resulted in 44 species, of which 18 overlapped in at least 50% of the records in which they were observed, with Common Chaffinch (95%), Blackcap *Sylvia atricapilla* and Common Chiffchaff *Phylloscopus collybita* (both 81%) showing most overlap (Fig. 1); Greater Canada Goose (0%), Grey Heron *Ardea cinerea* (0%) and Eurasian Jay *Garrulus glandarius* (7%) showed least overlap, though all three were observed in both point counts and audio-recordings, if not simultaneously. Sixteen species were most often recorded in point counts only, while only 4 species were most often recorded in audio only (Greater Canada Goose, Oystercatcher *Haematopus ostralegus*, Greylag Goose *Anser anser* and Jackdaw *Corvus monedula*).

Table 2. Mean performances (AUC-mean and AUC-binary) of the four evaluated classifiers BirdNET, GBVC, AvesdEcho and Aquila.

Model	AUC-mean	AUC-binary
BirdNET	0.834	0.911
GBVC	0.836	0.919
AvesEcho	0.832	0.825
Aquila	0.741	0.774

Evaluating classifiers

Comparing existing classifiers

The mean value of all present species' AUC-mean scores was highest for the Google Bird Vocalization Classifier (Table 2), however, the values for both BirdNET and AvesEcho were close. The performance of the Aquila model stayed behind. When looking at the AUC-binary, the performance gap between AvesEcho and both GBVC and BirdNET increased. Looking at the DET plot (Fig. 2), we again see BirdNET and GBVC close together. AvesEcho performs worse at lower thresholds compared to the other models and the performance of Aquila lags.
Figure 4 shows the AUC-mean scores (for examples see Fig. 3) for the 37 most frequent species. AvesEcho seemed unable to classify the Common Starling *Sturnus vulgaris* for the evaluation set, and also seemed to have problems identifying Egyptian Goose, Song Thrush *Turdus philomelos*, Eurasian Coot *Fulica atra* and Common Blackbird *Turdus merula* (AUC < 0.7). The two latter species were also challenging for GBVC and BirdNET, which however performed better for Egyptian Goose, Song Thrush and Eurasian Coot. Aquila could not classify Common Pheasant, Egyptian Goose and Greater Canada Goose, as these species were not included in their species list. However, it performed even worse than random for Eurasian Coot and Common Blackbird, and had problems identifying Willow Warbler *Phylloscopus trochilus*, Greylag Goose, Common Linnet *Linaria cannabina* and European Robin *Erithacus*

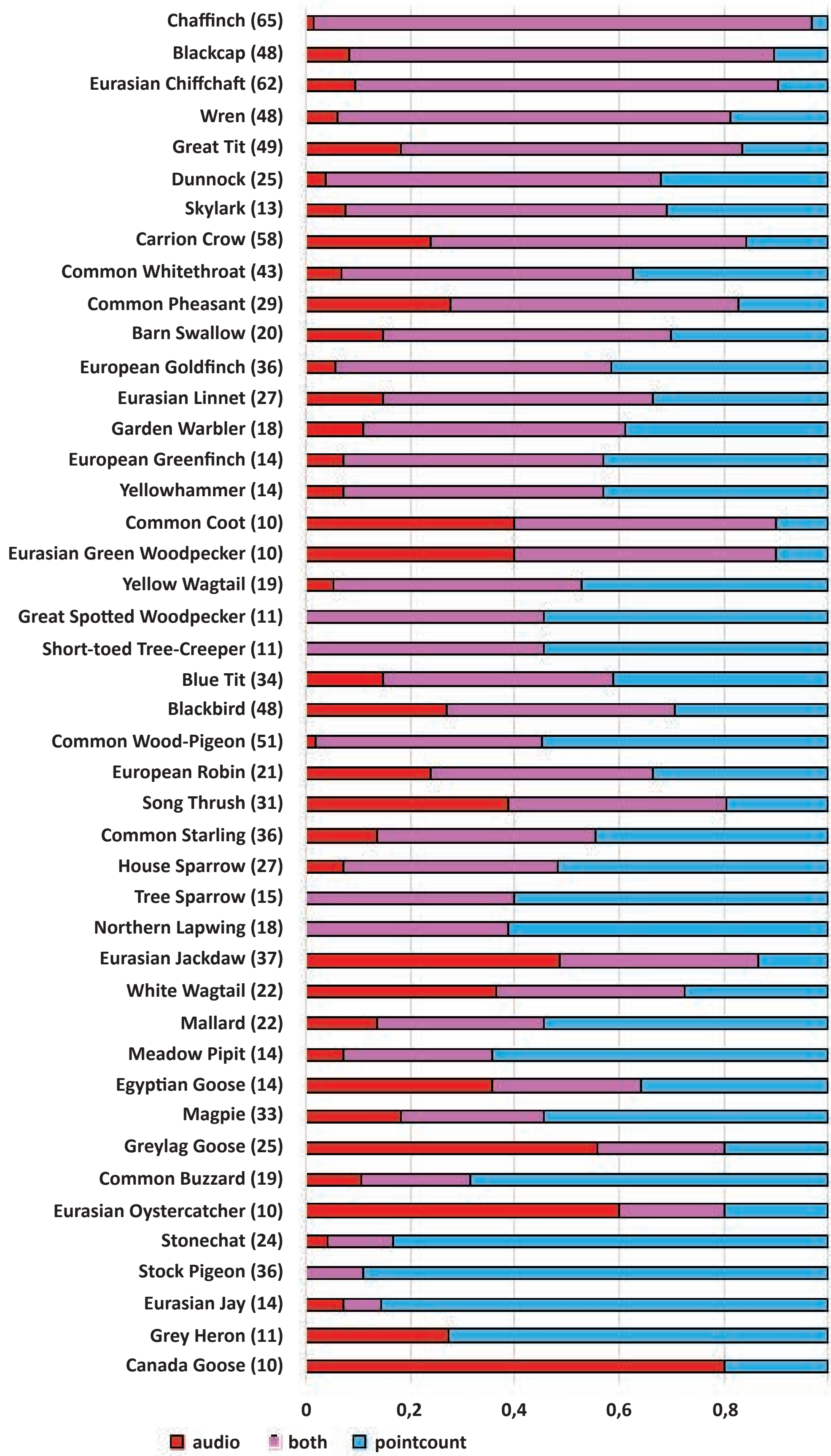


Fig. 1. Proportions of species-point-visit combinations in which a species was observed from the audio-recording only (red), during the point count only (blue) or during both (purple) for 44 species with 10 or more records (in audio-recordings and point counts combined). Number of records given between brackets. The species are ordered by the degree of overlap.

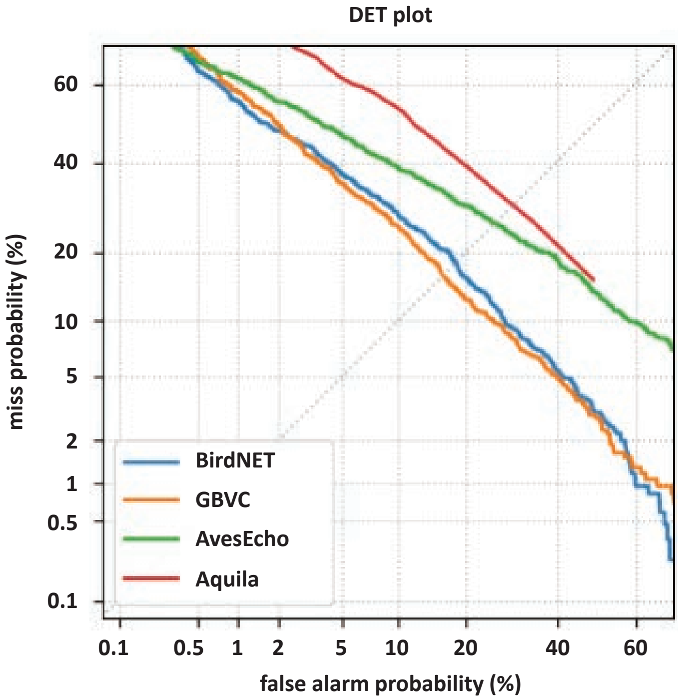


Fig. 2. Detection Error Trade-off curve for BirdNET, GBVC, AvesEcho and Aquila, accumulating targets (bird vocalisations) and non-targets (other sounds).

Table 3. Mean performances (AUC-mean and AUC-binary) of the two new classifiers NLC and NLC-NoSec and GBVC.

Model	AUC-mean	AUC-binary
GBVC	0.836	0.919
NLC	0.867	0.944
NLC-NoSec	0.838	0.834

rubecula, species for which the other models performed relatively well.

Training and evaluating new classifiers

Our approach using a custom classifier for Dutch application (NLC) improves results on both AUC metrics (Table 3). NLC-NoSec scores close to GBVC for AUC-mean but worse for AUC-binary. Looking at the AUC scores for the 37 most frequent species individually (Fig. 6), results vary for different species. Outliers seem to be the European Green Woodpecker, Greater Canada Goose, Song Thrush and Common Blackbird. For the first species, both NLC and NLC-NoSec perform better than GBVC while for the other species NLC performs best but NLC-NoSec worst. In general, NLC performs better than GBVC over the entire range (Table 3) while NLC-NoSec performs competitively for high thresholds but falls behind when lowering the threshold (Fig. 5).

Discussion

Audio recordings vs point counts

An important difference between point count data and data collected using (simple) audio recorders is that the first contains information on presence, abundance and density, while the latter mainly on presence, though analyses techniques are available which enable estimation of abundance from audio-data, using Time To Detection (Strebel et al. 2020), or removal models (van Wilgenburg et al. 2017). However, for these analysis techniques to work properly, the audio-recorders used need to be calibrated for the species present and location/circumstances, to be able to determine detection probabilities.

In 48% of species-point-visit combinations, species were both observed during the point count and heard on the audio-recording. Species showing good overlap between the two methods were all locally occurring songbirds of closed habitats, such as Common Chaffinch, Blackcap, Eurasian Wren *Troglodytes troglodytes* and Great Tit *Parus major*. These species are usually observed by their vocalisations when performing point counts.

Most species were recorded more often during the point counts than on the audio-recordings. This can largely be attributed to visual observations of birds that did not vocalise during the 10-minute count. These were often large and/or conspicuous and relatively silent species of open habitats, such as Stock Dove *Columba oenas*, Eurasian Jay, Stonechat *Saxicola rubicola*, Grey Heron, Common Buzzard *Buteo buteo* or Lapwing *Vanellus vanellus*.

Some species (e.g. Greater Canada and Greylag Goose, Eurasian Oystercatcher and Jackdaw) were observed more often from the audio-recordings than during point counts. These were often species with loud vocalisations and a large action radius and were probably individuals occurring outside the 300 m radius or flying over, and therefore deliberately omitted during the point count. However, it is likely that some individuals/species were truly missed during the point count, e.g. while the observer was focussing on other species; an advantage of using audio-recordings

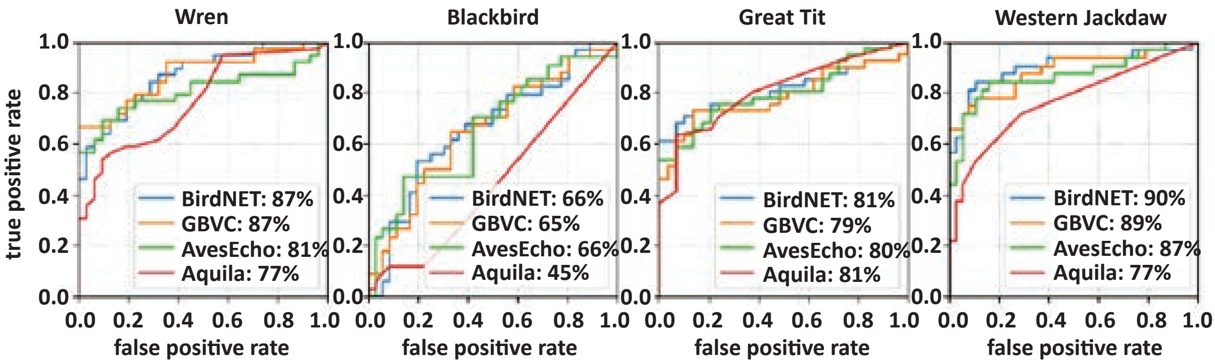


Fig. 3. Plots showing the ROC curves and AUC scores for BirdNET, GBVC, AvesEcho and Aquila for four common species. The area under the ROC curve (AUC) for each species, shown in Fig. 4, was calculated using these curves.

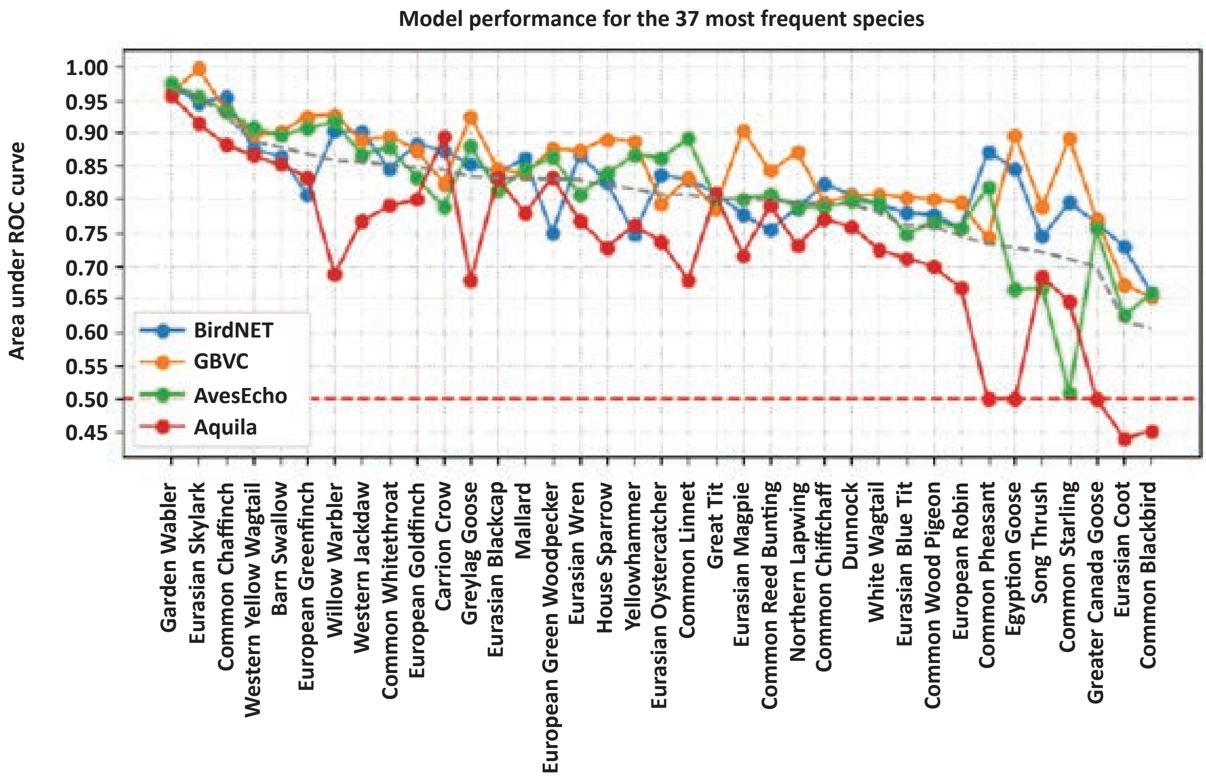


Fig. 4. Comparison of AUC scores for BirdNET, GBVC, AvesEcho and Aquila for 37 species most frequent in the evaluation set.

is that one can replay the recording when many species are vocalising simultaneously, to be able to identify and register every species.

Using acoustic observations only, (at least for now) it is not possible to distinguish between individuals inside or outside a specific detection radius, nor between individuals tied to a location or flying over. It is important to keep in mind such methodological differences, which come on top of the discrepancies in the mode of observation (sight and audio vs audio-only).

Differences in observation probabilities due to the lack of visual cues can be partly overcome by 1) applying ARU's in habitats/situations where visual detection is generally less used (forest, nighttime) and 2) increasing recording effort (duration). For the latter to be time and cost-effective, classification should be done automatically (Venier et al. 2017). However, automatic classification is associated with other issues, as discussed in the next section.

Model evaluation

The main limitations of classification model evaluation are generally the size and characteristics of the evaluation set. Only 73 out of 326 target

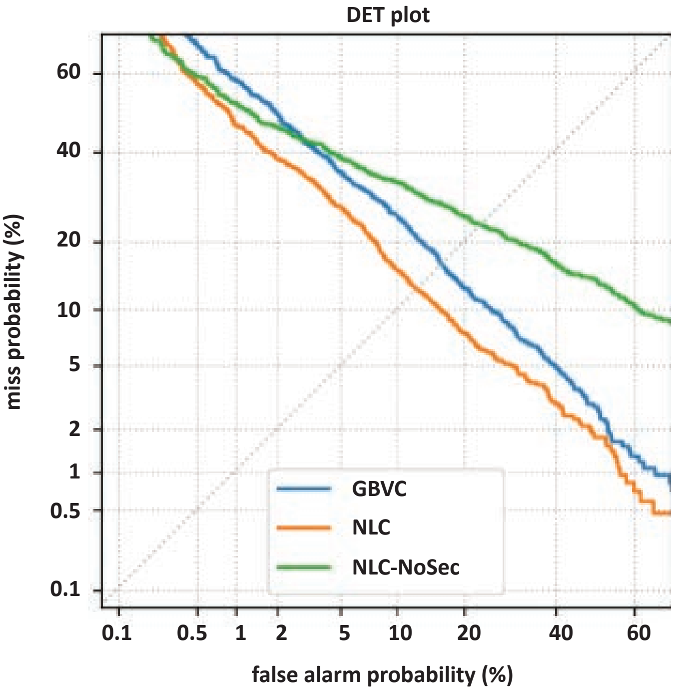


Fig. 5. Detection Error Trade-off curve for GBVC and the new classifiers NLC and NLC-NoSec, accumulating targets (bird vocalisations) and non-targets (other sounds).

species are present in the used evaluation dataset, so model performance could not be assessed for species not present in the evaluation dataset. Any conclusions on model performance are therefore limited to this set of 73, mainly agricultural and common species.

Moreover, for the species for which there were annotations, there are only twelve on average. Ideally, one would make annotations for every five seconds of the recording instead of per 10-minute recording. In this way, there would have been 120 times as many targets (bird sounds) and non-targets (other sounds) combined, which could be used to get a more accurate performance measure. Despite these limitations, the evaluation dataset provided an impression of how the different models perform.

When using a model in practice, one has to pick a confidence threshold, discarding all predictions with a score lower than the chosen threshold. A major difficulty is that the optimal threshold differs per species. So, one could best vary the threshold setting per species. Ideally, these thresholds are set using the results from a large and representative evaluation set. One could also use a manual approach in which the system first uses a very low confidence threshold for all species. Over time, the user could increase the threshold setting for a species if the system returns many false alarms for that species. A downside of this method is that it is not systematic and is mostly based on the number of false alarms.

Performance of classifiers

All evaluated models performed far from perfect on the evaluation set. When allowing for a false alarm probability of 1%, the miss probability is around 60% (Fig. 2). As there are 864 targets (vocalisations) and 21,956 non-targets in the dataset, this results in 220 false alarms while missing around 6 out of every 10 targets. Though classifiers are improving rapidly (GBVC at the point of writing has evolved to version 8 and AvesEcho to version 1), for most purposes it is pivotal to perform a manual validation of results.

In general, the performances of the classifiers, excluding Aquila, showed comparable patterns for the most frequent species, suggesting that species-specific vocalisation characteristics to a large degree determine their identifiability, regardless of the classifier used. Alternatively, if the classifiers used similar training sets for these species (e.g. all of the models evaluated use Xeno-Canto as (one of) their primary training data source), the correlation may reflect species-specific differences in the quality of the training data.

The overall performance of BirdNET and GBVC on our evaluation dataset is comparable, but differs for some species and at different thresholds. Choosing one model over the other should therefore ideally be done by evaluating the candidate models on a dataset similar to the intended application.

AvesEcho seems to be competitive at higher thresholds, but performs worse at lower thresholds, resulting in more false positives, while yielding fewer additional true positives. It could be that calibration is worse for AvesEcho than for BirdNET and GBVC, as unlike AUC-binary, the AUC-mean score for the three models is similar. Moreover, the prototype of the model we evaluated is single-label. The performance for the newer version of the model (Aves Echo v1) will probably have improved, as it is now multi-label. It would be interesting to see whether this newer version of the model can match or even out-

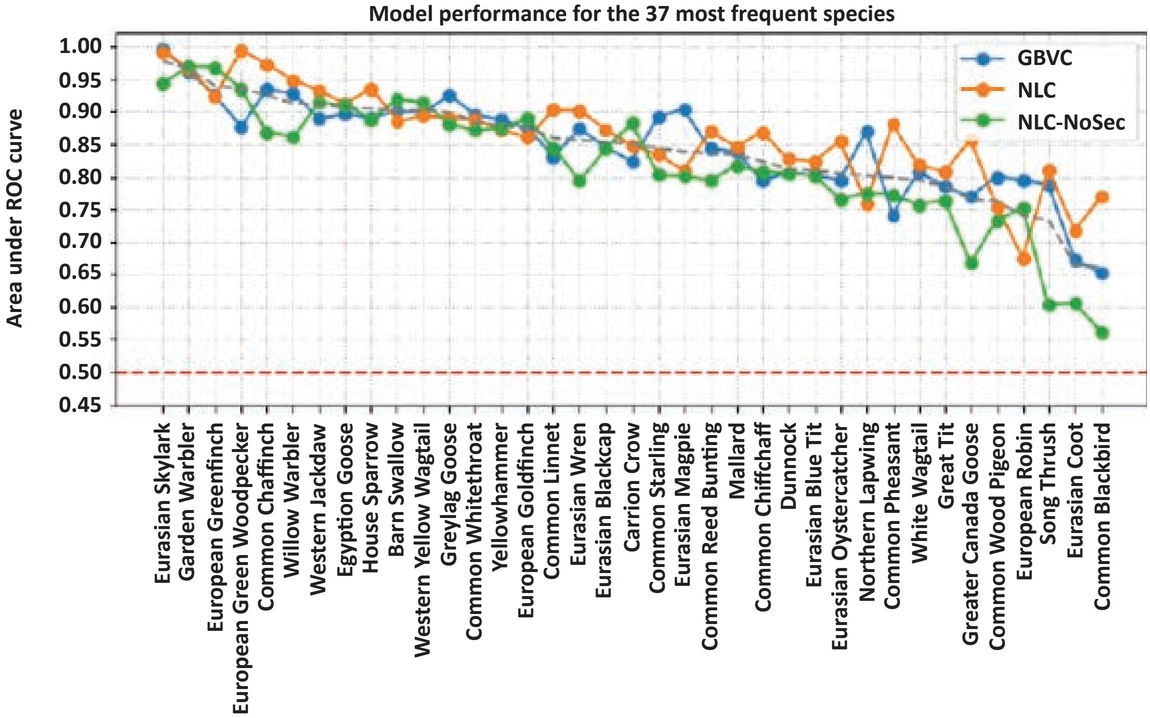


Fig. 6. Comparison of AUC scores for GBVC, and the new classifiers NLC and NLC-NoSec for 37 species most frequent in the evaluation set.

compete state-of-the-art models when applied to soundscapes.

The classifier of Aquila had the lowest performance on our evaluation dataset. The dataset on which the model was trained did not include five of the species present in our evaluation dataset. This can be seen from Figure 4, where Common Pheasant and Egyptian and Greater Canada Goose, three of those species, had an AUC = 0.5, exactly equalling performance when classified randomly. Including these species in the comparison will thus result in a lower overall performance score. Removing these species on which the model was not trained, only slightly increased the AUC value from 0.741 to 0.757. The Aquila classifier was mainly developed for insect and bat sounds, for which it performs markedly better (personal communication A. Krediet). Possibly the model is more suited for sounds of these species groups rather than for bird sounds (since the sounds of these species groups are more distinctive than bird sounds), or the classifier was trained with a smaller set of bird data than the other classifiers.

All of the models evaluated use Xeno-Canto as (one of) their primary data source to train the model on bird vocalisations. However, the recordings from Xeno-Canto have multiple characteristics that make models trained on this data likely less suitable for analysing soundscapes: weak labels (we know which bird is vocalising but not when in the recording), noisy labels (not all species vocalising in the background are labelled (Vellinga and Planqué 2015) which is particularly harmful in the context of soundscapes where we want to be able to detect background vocalisations) and focal recordings (most recordings are focal, meant to capture the labelled bird as clearly as possible). Another drawback is that not all species are equally well-represented.

The four models we evaluated predict scores for 3 (BirdNET, AvesEcho), 5 (GBVC) or 10-second-long (Aquila) segments of a soundscape. As a consequence, species with longer, more elaborate vocalisations may be harder to classify. An example would be the Blackbird, which proved difficult to classify for all of the models assessed. Another possible consequence is that one long (multi-segment) vocalisation could be attributed to multiple species. We have seen this occur with, again, Blackbird song, of which segments were misclassified as Mistle Thrush. This could be solved if classifiers would use longer segments, or if they would use information from adjacent segments. Remarkably, Aquila, the model using the longest segments (10 s) performed worst for the Blackbird, even worse than random.

Performance of new classifiers

Our results indicate the usefulness of creating a custom classifier to improve performance for a specific use case. Based on try-outs with the Google Bird Vocalization Classifier, we suspect to be able to improve in two ways:

• As we transfer to a Dutch setting, we only have to predict 3% of the species of the global setting. This matters most if the vocalisation of a target species is very similar to that of another species not in our target set. By removing the other species, the model will be more confident of the target species. It looks like an example of this can be seen in the case of the European Green Woodpecker *Picus viridis*. Whereas GBVC tries to distinguish the Iberian Green Woodpecker *Picus sharpie*, our model does not, resulting in a higher AUC score for the species, even for NLC-NoSec.

• As we switch from focal recordings to soundscapes, we need to increase our sensitivity to background vocalisations. We did not dispose of annotated multispecies recordings for training the model, but to check for the effect of background vocalisations, we decided to ignore the secondary labels present in some Xeno-Canto recordings. Ignoring secondary labels, NLC-NoSec performs worse on the evaluation set than NLC at lower thresholds. Two possible reasons for this drop are that the model is calibrated worse or that it has more difficulty with fainter/overlapping vocalisations. Given that the AUC-mean score is also lower than for NLC, it is not only a calibration issue. Therefore, we suspect that using secondary labels when training helps the model to predict fainter/overlapping vocalisations. This would also mean that an important step to improve classifiers would be to train them using annotated multispecies recordings.

For some species, the GBVC still performed better than the NLC, possibly because GBVC uses a larger training dataset than NLC, for which a maximum of only 100 recordings were selected. The performance of NLC could therefore probably be further improved by increasing the training dataset per species.

Conclusions

Passive Acoustic Monitoring at present is only/mainly useful for collecting presence data, not for trends in numbers or densities. Applicability for absence information is likely limited to vocal species. That said, it can be a useful addition for distribution mapping, including atlases, especially for nocturnal, vocalising species, and underrepresented habitats or sites. In the future, new analyses techniques, in combination with improved (and calibrated!) ARUs may enable abundance estimation for a subset of species on a local scale, though not without a considerable effort in the field set-up and validation/calibration of results. Furthermore, it is important to keep in mind that data collected using PAM differs systematically from data collected by point counts or other field methods, especially in open landscapes and for less vocal species. Therefore, it is important to always beware of a potential methodological trend break and always label data collected by PAM as such.

The quality of recordings greatly determines their usefulness for species classification. Though effects have not been analysed here, some recordings were excluded as they contained too much background noise and were deemed useless for the evaluation. Minimizing background noise is therefore a first step in improving audio monitoring results. This also highlights a potential risk of using low-cost ARU's, which might use low quality microphones with considerable within device variation, which can result in highly variable performance among devices in the field.

To improve a model's performance on soundscapes, it is pivotal to use annotated soundscapes as (additional) training data. Annotations should contain the species name and start and end time for each vocalisation. In addition, including recordings without bird sounds in the training dataset could improve the distinction between bird vocalisations and background noise. We expect models could also be greatly improved if they would also use the information from adjacent sound fragments. Furthermore, the species-specific performance of models can relatively easily be improved by varying and optimizing the threshold setting per species.

Finally, we suggest using a hybrid approach in which a model's predictions are validated by an expert. The validated data can then be used for training, which can further improve the model (human-in-the-loop).

Acknowledgements

We would like to thank Vincent Kalkman and Burooj Ghani (AvesEcho) and Anne Krediet (Aquila) for kindly offering us their classifiers for evaluation; Fred de Boer for providing the AudioMoths and Eveline Schaar for deploying them in the field; Frank Majoor, Jesse Keijzer and Bas Hissel for performing the point counts; Willem van Manen, Joost van Bruggen and Bas Hissel for performing the manual annotations of the audio-recordings.

References

Browning, E., R. Gibb, P. Glover-kapfer & K. E. Jones. 2017. Passive acoustic monitoring in ecology and conservation. WWF Conservation Technology Series 1 (2). WWF-UK, Woking, United Kingdom.

Ghani, B., T. Denton, S. Kahl & H. Klinck. 2023. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. *Scientific Reports,* 13: 1–14.

Google Bird Vocalization Classifier https://www.kaggle.com/models/google/bird-vocalization-classifier. Accessed: 2023-11-20.

Hanley, J. A. & B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology,* 143: 29–36.

van Harten, N. 2023. NL-Bird-PAM. https://github.com/nielsvharten/NL- Bird- PAM/.

Hoo, Z. H., J. Candlish & D. Teare. 2017. What is an ROC curve? *Emergency Medicine Journal,* 34: 357–359.

Kahl, S., C. M. Wood, M. Eibl & H. Klinck. 2021. BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics,* 61: 101236.

Klingbeil, B. T. & M. R. Willig. 2015. Bird biodiversity assessments in temperate forest: The value of point count versus acoustic monitoring protocols. *PeerJ,* 3: e973.

Martin, A., G. Doddington, T. Kamm, M. Ordowski & M. Przybocki. 1997. the Det Curve in Assessment of Detection Task Performance. Pages 1895–1898 5th European Conference on Speech Communication and Technology, EUROSPEECH 1997.

Shonfield, J. & E. M. Bayne. 2017. Autonomous recording units in avian ecological research: current use and future applications. *Avian Conservation and Ecology,* 12: 14.

Strebel, N., C. J. Fiss, K. F. Kellner, J. L. Larkin, M. Kéry & J. Cohen. 2021. Estimating abundance based on time-to-detection data. *Methods in Ecology and Evolution,* 12: 909–920.

Teunissen, W. A., P. Wiersma, A. de Jong, E. Kleyheeg & J. Vergeer. 2019. Handleiding Meetnet Agrarische Soorten (MAS).

Vellinga, W. P. & R. Planqué. 2015. The Xeno-canto collection and its relation to sound recognition and classification. Page CEUR Workshop Proceedings.

Venier, L. A., M. J. Mazerolle, A. Rodgers, K. A. McIlwrick, S. Holmes & D. Thompson. 2017. Comparison of semiautomated bird song recognition with manual detection of recorded bird song samples. *Avian Conservation and Ecology,* 12: 1.

Van Wilgenburg, S. L., P. Sólymos, K. J. Kardynal & M. D. Frey. 2017. Paired sampling standardizes point count data from humans and acoustic recorders. *Avian Conservation and Ecology,* 12: 13.

Xeno-Canto https://xeno-canto. org/about/xeno-canto. Accessed: 2023-11-17.

Appendix

The methods used for data preprocessing, the model architecture and the training procedure for the new classifiers NLC and NLC-NoSec

Data preprocessing

We trained a classifier that uses GBVC embeddings of 1280 features as input. To generate these embeddings, all audio was resampled to 32 kHz as GBVC is trained using this sample rate. To increase the diversity of our training data, we used an overlap of 4 seconds, ignoring the last three embeddings, unless the recording was shorter than 4 seconds, in which case we only used the first embedding. As the input length for GBVC is 5 seconds, this resulted in 30 embeddings for a 30-second-long recording. We also used the model's predictions for all training data to filter which embeddings of a recording to use. We discarded embeddings for which the GBVC prediction (after softmax) was less than 0.02 for the species labelled in the recording. If no segments had a prediction of at least 0.02, we kept the first embedding hoping that the target species vocalises within the first five seconds of the recording.

Model architecture

Our classifier is a simple, fully connected neural network having one hidden layer of size 512. First, we applied dropout on our input. After the first fully connected layer, we used batch normalization and ReLU activation followed by dropout. After the second fully connected layer, we only used sigmoid activation. Given the 1280 input and 326 output dimensions, the classifier has 823k trainable parameters.

Training procedure

We split our training data in a train and validation split of 80% and 20% respectively. We upsampled species with fewer recordings by repeating samples using the number of embeddings as their probability. This way, recordings for which there are more embeddings are more likely to be sampled multiple times. To diversify inputs and improve the model's ability to learn overlapping vocalisations, we employed E-stitchup. This approach combines two embeddings into one by randomly selecting the value at each index of the combined embedding from the value at that index for one of the two original embeddings (Cameron & Lundgaard 2019). The probability to sample from one embedding is determined by a value lambda, for each combination drawn from a beta distribution with $\alpha = \beta = 1$. Given that Xeno-Canto allows specifying secondary labels of species also vocalising in a recording, we give these classes a non-zero probability p. After experimentation, we found $p = 0.9$ to work best. Besides, we use label softening, subtracting 0.1 from all classes with non-zero probability and adding 0.1 to all other classes.
As our optimizer, we use ADAM (Kingma & Ba 2014) with an initial learning rate of 0.001 and a cosine annealing schedule gradually decreasing the learning rate to zero over 30 epochs. We train in batches of 32 samples where each sample is constructed out of two embeddings randomly selected from two different recordings using E-stitchup. We use a dropout probability of 0.1.

Cameron R.W. and K.T. Lundgaard. 2019. "E-stitchup: Data augmentation for pretrained embeddings". In: *arXiv preprint arXiv*: 1912.00772.
Kingma P.D. and J. Ba. 2014. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv*: 1412.6980.

Received: 14th July 2025
Accepted: 12th December 2025

Bird Census News 2025, 38/1–2: 26–30

The potential of acoustic monitoring to inform and expand common bird monitoring

Mark W. Wilson[1], James J.N. Heywood[2], Adham Ashton-Butt[1], Anthony Wetherhill[1] and Simon Gillings[2]

[1] British Trust for Ornithology — Scotland. Unit 15, Beta Centre, Stirling University Innovation Park, Stirling, FK9 4NF
[2] British Trust for Ornithology, The Nunnery, Thetford, Norfolk IP24 2PU, UK

mark.wilson@bto.org, james.heywood@bto.org, adham.ashton-butt@bto.org, anthony.wetherhill@bto.org, simon.gillings@bto.org

Abstract. Acoustic monitoring has the potential to be a very effective method of surveying birds, particularly in remote areas where traditional survey coverage is limited by availability of appropriately skilled surveyors. In order to assess its suitability — particularly for augmenting long-term abundance monitoring — this study compares data collected by acoustic recorders and by surveyors on existing sites (1-km squares) of the UK Breeding Bird Survey (BBS).
Twenty-eight volunteers in Scotland deployed an acoustic recorder on their BBS survey square during spring 2023. An automated classifier (BirdNET Analyzer) was run on all recordings to detect birds and identify them to species. A random subset of recordings was checked manually in order to evaluate classifier performance. We discuss the performance of BirdNET in the context of its suitability for acoustic monitoring in Scotland and compare BirdNET outputs with recent BBS data collected from the same sites. We also consider ongoing and potential future work in this area.

Introduction

The UK Breeding Bird Survey (BBS) is one of over 30 bird monitoring schemes operating in Europe and contributes to the PECBMS (Massimino et al. 2024). Nearly 3,000 skilled volunteers take part in BBS, each monitoring one or more 1-km squares. Each square receives two survey visits per year — an early visit (usually done between mid-April and the end of May) and a late visit (usually done between mid-May and the end of June). During each visit, BBS surveyors record all the birds they see and hear while walking through the square along two parallel, 1-km long transect lines. Surveyors also note the species, number of individuals, distance from the transect and identity of the closest 200m transect section for each detection. The survey and the data collected is used for monitoring of 119 species in the UK during the breeding season, with key outputs including population trends (Heywood et al. 2025) and indicators (Defra 2024). However, there are still species and habitats for which we would like more information. As is the case in many other countries, biological monitoring is more difficult in parts of the UK where there are fewer people and/or where the landscape and terrain is physically challenging. This challenge is especially keen in the north and west of Scotland, which supports several species of breeding bird with narrow distributions at the southern edge of an arctic/northern European range. UK breeding species with populations largely concentrated in Scottish peatlands — e.g. Greenshank *Tringa nebularia* and Dunlin *Calidris alpina* — and montane habitats — e.g. Dotterel *Eudromias morinellus* and Ptarmigan *Lagopus muta* — are challenging to monitor using traditional methods and may require bespoke surveys relying heavily on expensive, professional coverage.

Whilst skilled volunteer bird surveyors are in short supply in upland and montane areas, many of these landscapes are popular with hikers and other outdoors enthusiasts. Most of the people in these groups do not currently have sufficient knowledge or interest to undertake bird surveys, but they might nevertheless be able to make valuable contributions to bird monitoring by deploy-

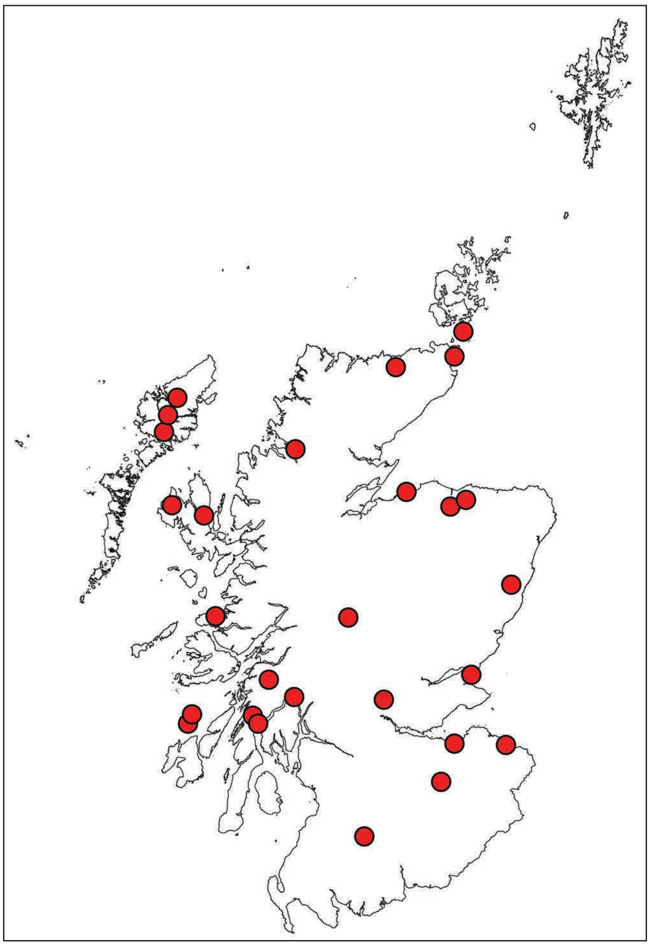


Figure 1. The location of the recorders deployed on BBS squares in 2023.

ing acoustic recording devices in remote areas. To understand the kind of data that could be collected using acoustic recorders in this way, we invited BBS surveyors in Scotland to deploy recorders on their monitoring sites. This allows us to compare data from acoustic recordings with standardised count data collected by human observers at the same locations.

Methods

Recorders

One-hundred and twenty BBS volunteers in Scotland whose squares had been surveyed for at least three of the previous five years (2018 to 2022) were invited to deploy a recorder. Thirty Wildlife Acoustic Song-meter Micro recorders were distributed between the 60 volunteers who responded to this invitation, allowing a wide range of locations and habitat types to be sampled (Fig. 1). Recorders were deployed on 29 squares, with 28 of these successfully generating a large amount of acoustic data. The duration of individual recorders deployments ranged from 28–69 days, during which time they recorded audio for one minute in every 15, yielding a total of 2,077 hours of recordings. The recorders were pre-programmed prior to deployment, and each volunteer was asked to deploy their recorder within 50m of their transect route, between one and two metres from the ground, with sound able to access it from all directions, and with the microphone pointing towards the transect. The deployment was designed to fit around the two survey visits made by BBS volunteers each year — recorders were deployed on early visits and collected during late visits.

Analysis

The automated classifier BirdNET Analyzer v2.3 (Kahl et al. 2021) was used to detect birds and identify them to species in every 3-second segment of each recording. Identifications were constrained by location and date and had a minimum confidence score (a metric that reflects BirdNET's assessment that a classification is correct) of 0.1.

A large sample (10,080) of randomly selected 3-second clips taken from these recordings was manually classified by an expert in bird sound identification (AW) to assess BirdNET's performance.

The BirdNet classifier records were compared with the BBS records in three ways: 1) For each combination of species and BBS square, call rate (defined as the mean number of call detections by BirdNET per hour during the week with the most calls) was compared with counts from BBS surveyors. 2) At the level of the whole BBS square, BirdNET classified call rates were compared with the five-year mean counts (with the count for each year being the highest count from either visit). 3) At a finer spatial scale, call rates were compared with counts from the 200-m section on which the recorder was placed.

Results

Classification performance

BirdNET generated 750,945 detections of 195 species. Classifier precision was generally high, with 79% of BirdNET detections verified by a hu-

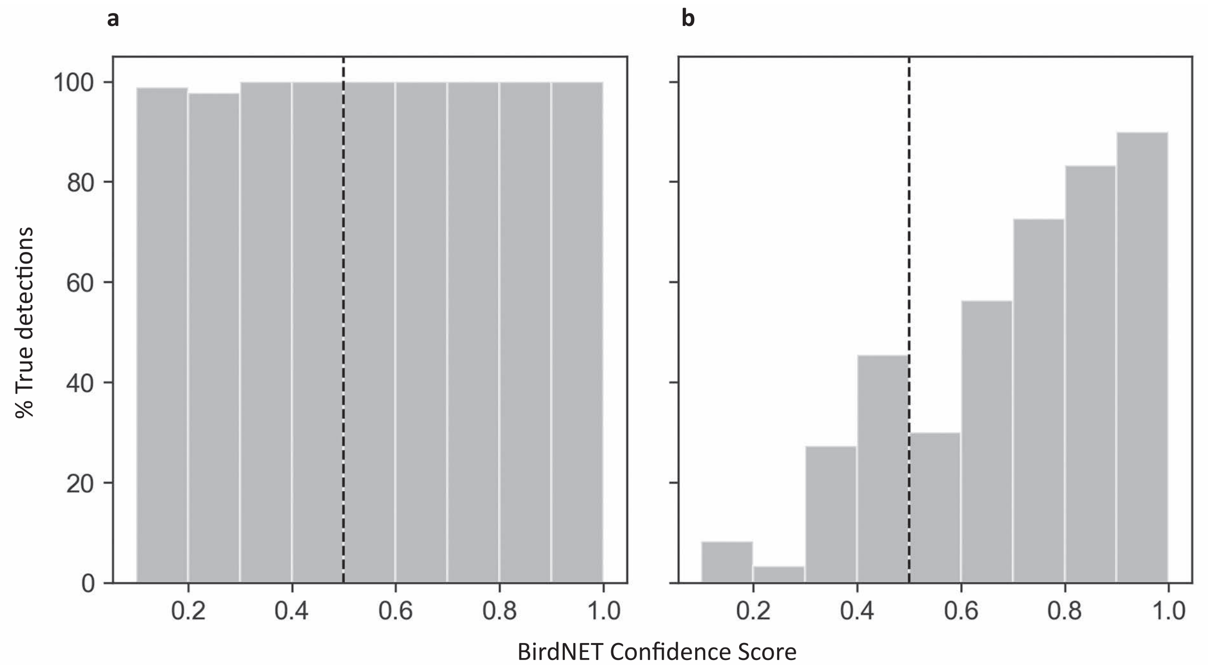


Figure 2. BirdNET-Analyzer precision for Chaffinch *Fringilla coelebs* (a) and Tree Pipit *Anthus trivialis* (b) in relation to true detection, based on manual verification. An arbitrary Confidence Score threshold of 50% (dotted line) would only retain correct Chaffinch detections but would result in 82% of correct Chaffinch detections being unnecessarily rejected. However, the same threshold would likely be too low for robust analysis of Tree Pipit identifications, as it would result in 36% of retained Tree Pipit identifications being incorrect.

man being confirmed as correct. Precision was generally greater for detections with high confidence scores, but this relationship varied substantially between species (Fig. 2). Only 27% of the calls detected and identified by a human expert were detected by the classifier. Among 34 species for which 50 or more calls were manually identified, the proportion of these calls that were also detected by BirdNET varied from 0% to 73%.

Comparison with BBS

During the period 2019 to 2023, 134 species were detected by BBS surveyors on the study squares, with a 60% of the 195 BirdNET species being recorded by the volunteers. When BirdNET data were refined — removing obvious errors and using Bird Atlas 2007–2013 data (Balmer et al. 2013) to exclude species that would not be expected to breed in these locations — 116 BirdNET species remained, of which 85% were recorded by BBS volunteers.

The degree to which BirdNET-derived call rates correlated with BBS counts varied between species and also depended on the number of years and spatial extent that BBS counts were drawn from. For Skylark *Alauda arvensis*, BirdNET call rate was positively correlated with the five-year mean BBS count, regardless of whether or not BirdNET detections were filtered according to confidence score (Fig. 3). However, for other species the correlations were improved by filtering. For common species with relatively small territory sizes that can be detected up to 100m or so away, like Willow Warbler *Phylloscopus trochilus*, BirdNET call rate was more strongly correlated with BBS counts from 200m sections than with counts from the whole BBS square (Fig. 4). For scarcer birds that range over larger distances and can be heard more than 1km away, such as Cuckoo *Cuculus canorus*, the reverse is true.

Discussion

Acoustics and monitoring abundance

We are just starting to learn how acoustic recording, coupled with automated classification technology, could contribute to the information we use to monitor bird abundance. Our 2023 data demonstrate that, at least for some species, the rate of calls classified from recordings reflects the variation in BBS counts across different sites. However, for acoustic monitoring to tell us about long-term changes in population, call rates captured by a recorder at a given location should (in the absence of sudden changes in population) be

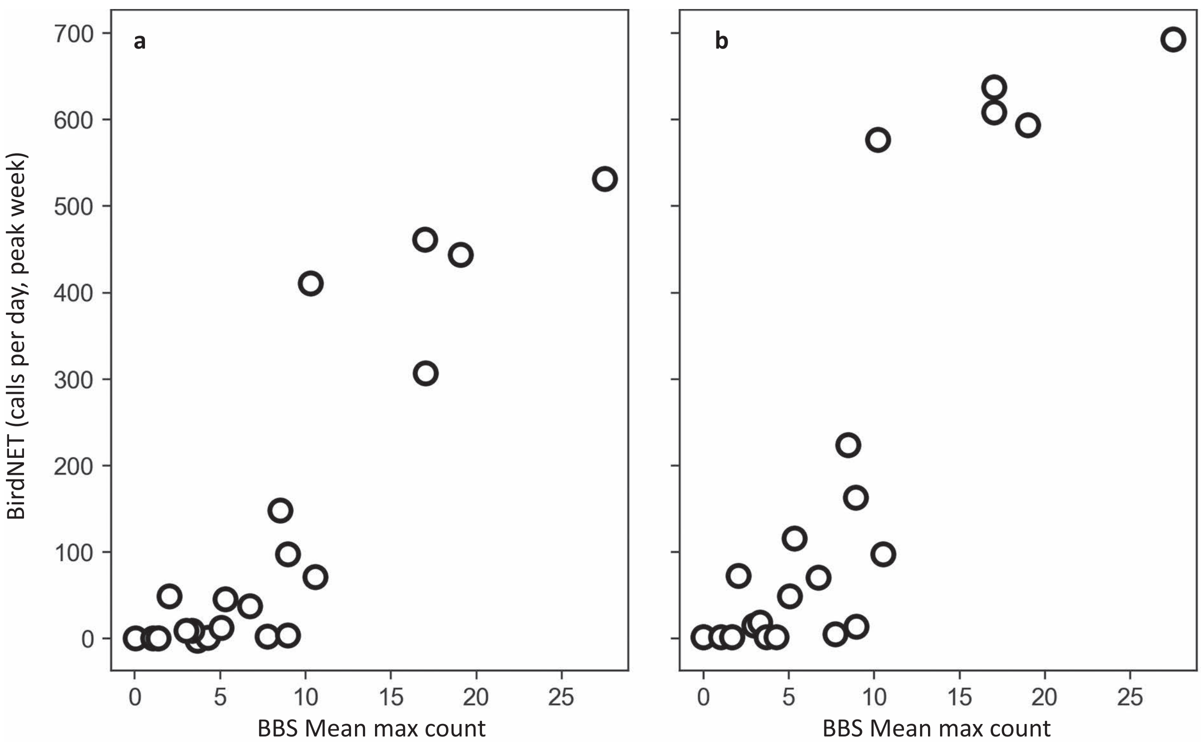


Figure 3. The mean BBS max count of Skylark *Alauda arvensis* over five years (a) correlates relatively well with an index of call output and continues to do so even after filtering by BirdNET confidence score (≥ 0.9) (b). In this case, BirdNET performs well and so relatively few recordings are filtered.

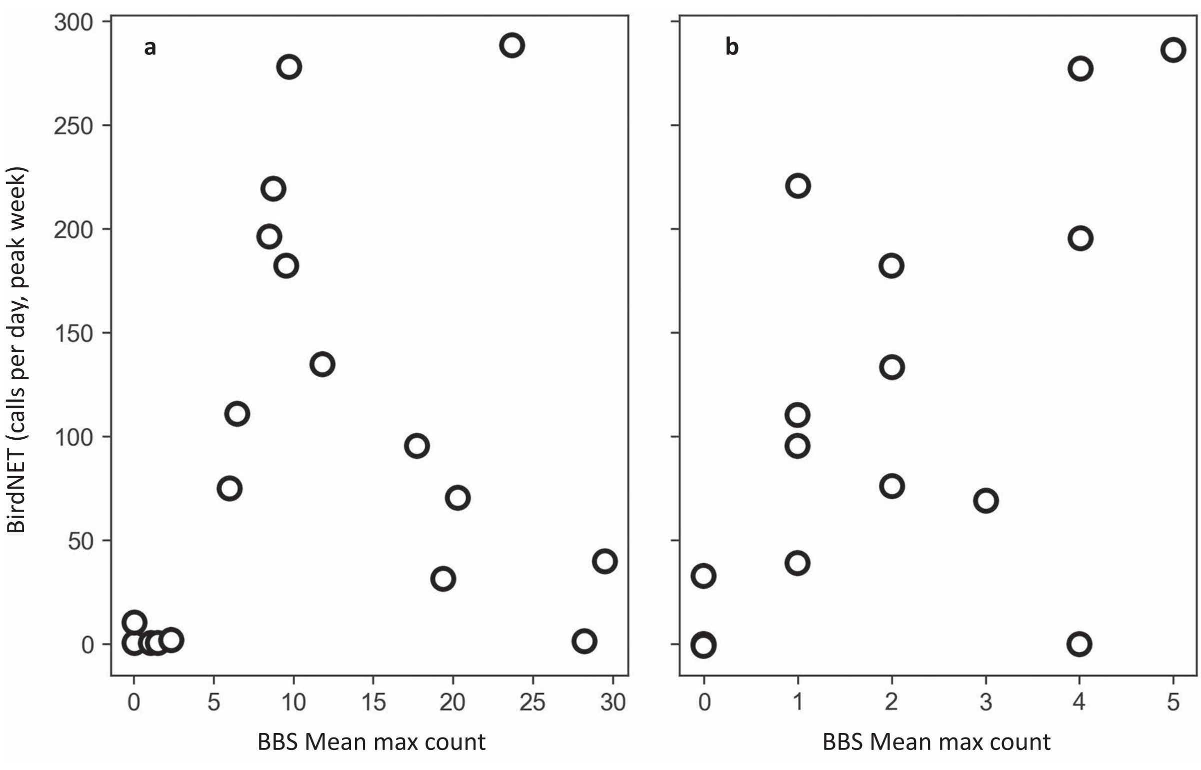


Figure 4. The mean BBS max count of Willow Warbler *Phylloscopus trochilus* over five years (a) does not correlate well with acoustic indices. However, when BBS data is filtered to only include counts from the same year (2023) and 200 m transect section (b), the correlation improves.

relatively consistent from year to year. This year (2025), 20 of the volunteers who took part in 2023 are deploying recorders on their BBS squares again. At 24 of the 28 sites that contributed to this study, this will give us a second year of acoustic data that can be compared directly with BBS counts from the same locations and time period. We can use these data to check how variation in call rates between years compares with inter-annual variability in BBS counts.

Another area of ongoing work aims to improve the ability of acoustic classifiers to correctly identify calls from the recordings. Much of this work focusses on developing single- and multi-species classifiers in the BTO's Acoustic Pipeline (BTO 2024), and we are also collaborating with researchers outside of BTO on various aspects of analysis and interpretation of acoustic data. Among the topics we want to get a better understanding of are:

- Whether classification of calls can be made more robust by drawing on information about other classified calls in the same recording.
- Whether call amplitude can be used to infer the distance of calling birds from a recorder.
- How recorder location and placement, and in particular the background noise this results in, affects the ability of automated classifiers to detect and correctly identify different species in acoustic recordings.
- How variation in recorder type/model influences the data collected. If acoustic recordings are going to contribute to long-term monitoring, we need to consider how improvements in hardware and software will affect the species that can be detected, and how this can be accounted for when working out how populations are changing over time.

Other applications

Acoustic methods could greatly enhance and improve on existing monitoring approaches, particularly in remote areas and/or for nocturnal or cryptic species. For example, acoustic monitoring could contribute information on breeding evidence, along with presence and abundance. As well as identifying calls made by different species, automated classifiers can also be trained to distinguish different sounds made by the same species. If song, alarm calls and juvenile begging calls can all be reliably identified, it may be possible to determine not only that a species was breeding in an area but to get some idea of variation in breeding success — something that is often very difficult and time-consuming to do through more traditional surveys and fieldwork (Jarrett et al. 2024). However, in order to realise acoustic monitoring's potential, further work is needed to improve our guidance for acoustic surveys, the ability of classifiers to identify calls from acoustic recordings and our interpretation of classifier outputs.

References

Balmer, D. E., Gillings, S., Caffrey, B. J., Swann, R. L., Downie, I. S. & R. J. Fuller. 2013. *Bird Atlas 2007–11: The Breeding and Wintering Birds of Britain and Ireland*. BTO Books, Thetford.

BTO 2024. BTO Acoustic Pipeline: Accurate specie identification and data management for acoustic monitoring in conservation, management and site assessment, Available at: www.bto.org/acoustic-pipeline

Defra 2024. Wild bird populations in the UK and England, 1970 to 2023. Available at: www.gov.uk/government/statistics/wild-bird-populations-in-the-uk

Heywood, J. J. N., Massimino, D., Baker, L., Balmer, D. E., Brighton, C. H., Gillings, S., Kelly, L., Noble, D. G., Pearce-Higgins, J. W., White, D. M., Woodcock, P., Workman, E. & S. Wotton. 2025. *The Breeding Bird Survey 2024*. BTO Research Report 787. British Trust for Ornithology, Thetford.

Jarrett, D., Lehikoinen, A. & S. G. Willis. 2024. Monitoring wader breeding productivity. *Ibis,* 166 (3): 780–800. doi: 10.1111/ibi.13298

Kahl, S., Wood, C. M., Eibl, M. & H. Klinck. 2021. BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61: 101236. doi.org/10.1016/j.ecoinf.2021.101236

Massimino, D., Baillie, S. R., Balmer, D. E., Bashford, R. I., Gregory, R. D., Harris, S. J., Heywood, J. J. N., Kelly, L. A., Noble, D. G., Pearce-Higgins, J. W., Raven, M. J., Risely, K., Woodcock, P., Wotton, S. R. & S. Gillings. 2024. The Breeding Bird Survey of the United Kingdom. *Global Ecology and Biogeography*, 34: e13943. doi.org/10.1111/geb.13943

Received: 6th June 2025
Accepted: 5th August 2025

Bird Census News 2025, 38/1–2: 31–36

Detection probability and abundance of a Long-eared Owl population in Switzerland

Tatiana Kornienko[1], Marc Kéry[1], Mathias Ritschard[2] and Thomas Sattler[1]

[1] Swiss Ornithological Institute, Seerose 1, CH-6204, Sempach, Switzerland
[2] Orniplan AG, Wiedingstrasse 78, CH-8045, Zurich, Switzerland
tetiana.korniienko@vogelwarte.ch

Abstract. Traditional Long-eared Owl survey methods like auditory detection by fieldworkers are time-intensive and often yield uncertain results due to low detectability and observer bias. We analysed a large playback-assisted survey in the Canton of Zurich, Switzerland, using occupancy models. Surveys using playback were more effective than passive listening, as detectability was increased, while Tawny Owl presence reduced it. The best way to detect Long-eared Owls was to use playback during late March and early April. The combination of playback and occupancy modelling has led to a substantial reassessment of population size in our study area. Our findings emphasize the value of playback for optimizing survey methods for some species and provide essential baseline information for future studies and potential conservation actions.

Introduction

Due to their cryptic behaviour, the distribution and abundance of many owl species is often incompletely known (Aebischer 2008, Robb 2009). For example, in Switzerland, despite a good knowledge of many other breeding birds (Strebel et al. 2024), assessing population status and trends remains a challenge for owls. Auditory detection by fieldworkers is time-intensive and often suffers from a high incidence of false-negative errors (Robb 2009). This means, that a species is overlooked in the field, despite being present, i.e. imperfect detection (Kissling et al. 2010, Zuberogoitia et al. 2019, Kéry et al. 2024). For species like the Long-eared Owl *Asio otus* these limitations hamper our ability to collect reliable data on distribution, population size and long-term trends, all crucial for population studies and conservation planning.

Previous research has shown that playback increases the detection probability of many bird species (McGregor 2000, Grinde et al. 2018), especially for owl species (McGarigal & Fraser 1985, Martinez & Zuberogoitia 2003, Hannah 2009, Mori et al. 2014, Vrezec & Bertoncelj 2018, Zuberogoitia et al. 2019, Kéry 2023). Birrer (2014) has previously demonstrated the effectiveness of playback for Long-eared Owl surveys.

The Long-eared Owl is a characteristic species of forest-open areas ecotones and of various types of mosaic landscapes. Main threats include intensive agriculture and the associated decline in prey abundance and availability (Birrer 2009, Keller et al. 2020). In Switzerland, the Long-eared Owl population size is estimated to be between 2,000–3,000 territories (Knaus et al. 2018). This assessment has a large uncertainty for Swiss standards. In the Canton of Zurich, 31 territories were reported in 1988 and 59 in 2008 (Weggler et al. 2009). However, standardized, comprehensive surveys have been missing so far.

To bridge this gap, in 2020 we started a long-term project that aims at updating distribution and abundance information on Long-eared Owls in the Canton of Zurich with a systematic and standardized method. We used playback to increase the detection probability and compare it to auditory-only detection. Specifically, occupancy modelling was used to analyse the influence of playback, to find out whether there are interspecific interactions (especially with respect to the Tawny Owl *Strix aluco*), to assess temporal effects in detection probability, and to derive a rigorous population size estimate that accounts for imperfect detection. At the time of writing, half of the Canton has been covered by our surveys. Here we report our first findings.

Material and methods

The study was conducted in the Swiss Canton of Zurich. This area is situated in the eastern part of the Swiss plateau (1729 km²; altitude 330–1291 m a.s.l.) and for the most part composed of low-elevation river valleys, with forests covering 43% of the area, agriculture 30%, and urban areas 20% (Swiss Land Use Statistics, Bundesamt für Statistik, Bodennutzung, 2024).

Our study has been conducted in five years (2020–2024). Fieldwork was done by 90 volunteers from the Avimonitoring program Zurich (Ritschard 2020, Ritschard 2022, Ritschard 2023a, Ritschard 2023b, Ritschard 2024). They visited 2,218 point-count surveys (2.3 surveys/km²), covering 54% of the Canton of Zurich. Each year, a different set of communities were selected and all points within their corresponding administrative boundaries were surveyed.

Survey points had been a-priori selected by the coordination office in such a way that no forest edge or grove in the study area is further than 300 metres from any survey point. By doing so, we aim at a complete population survey of the Canton of Zurich. On average, 440 points were surveyed per year (ranging from 368 in 2023 to 539 in 2021). Volunteers were allowed to move the listening points by a maximum of a few dozen meters if this was suggested by local conditions. In hilly terrain, it was important to ensure that no obstacle (topography, building or similar) hindered the acoustic survey of the nearby forest edges. Therefore, the survey points were often placed at exposed locations.

Volunteers surveyed each point twice, once in March and again in April. The first visit ideally took place before March 20th, and the second before April 20th, with at least 10 days between the two visits. Surveys started approximately 30 minutes after sunset, i.e., at 7:00 p.m. in mid-March and 8:45 p.m. in mid-April. Up to 16 points were surveyed in one suitable evening.

In one survey, observers were asked to follow the following steps:

1. 120 sec. of passive listening (i.e. traditional auditory detection): if a Long-eared Owl was heard calling spontaneously, the survey was stopped, and the fieldworker moved to the next survey point. The main aim of the study was to detect territories with as little disturbance as possible and not to run the complete protocol at each survey point.
2. 15 sec. of playback (male hooting, 5 calls): once the owl answered, the playback was stopped, no additional survey on the point was done on that day, and the fieldworker moved to the next survey point.
3. 30 sec. listening in case there was no answer in step 2.
4. 15 sec. of playback: stopped if answer received, as in step 2.
5. 300 sec. of listening in case there was no answer in step 4.

In addition to male and female calls of Long-eared Owls, volunteers also registered whether they heard other owl species, especially Tawny Owls as potential predators or competitors of the smaller Long-eared Owls.

Although the initial passive listening period was stipulated to last 120 seconds, we realized that some people also recorded owls detected during the final approach to a prescribed listening point and during the initial preparation phase for playback. Thus, the actual duration of step 1 was typically between 180 and 240 seconds. We explicitly considered this imprecision in the modelling part (see below).

We analysed survey data with an occupancy model for a removal design (MacKenzie et al. 2017). We treated step 1 as the first occasion, steps 2 and 3 combined as the second, and steps 4 and 5 as the third one. Occasion 1 was listening-only, while occasions 2 and 3 were playback. Since occasion length was different, we expressed detection probability per 1-minute unit as our 'common currency' of comparison. Thus, detection probability is the probability to detect an owl during a 1-minute survey at a site that is in fact occupied by at least one owl.

Every survey point was only surveyed in a single year out of the five and we did not have any specific hypotheses or information about drivers of occupancy probability. Hence, we treated this parameter as a constant. Our main interest lay in the elucidation of spatial and temporal patterns in detection probability. Our detection model had fixed effects of the year treated as a factor, linear, quadratic and cubic effects of survey date as a continuous explanatory variable, and site-level detection of Tawny Owl and playback versus passive listening as two further factors. We dealt with the

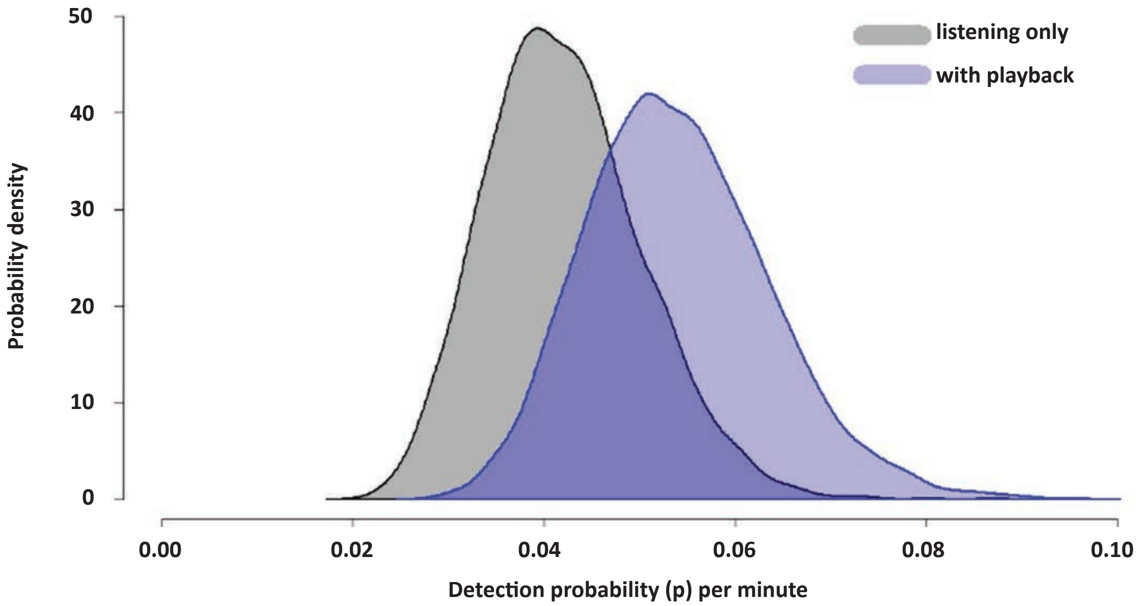


Figure 1. Effect of playback on the detection probability of the Long-eared Owl *Asio otus* (posterior distributions shown of per-minute detection probability for passive listening vs. playback).

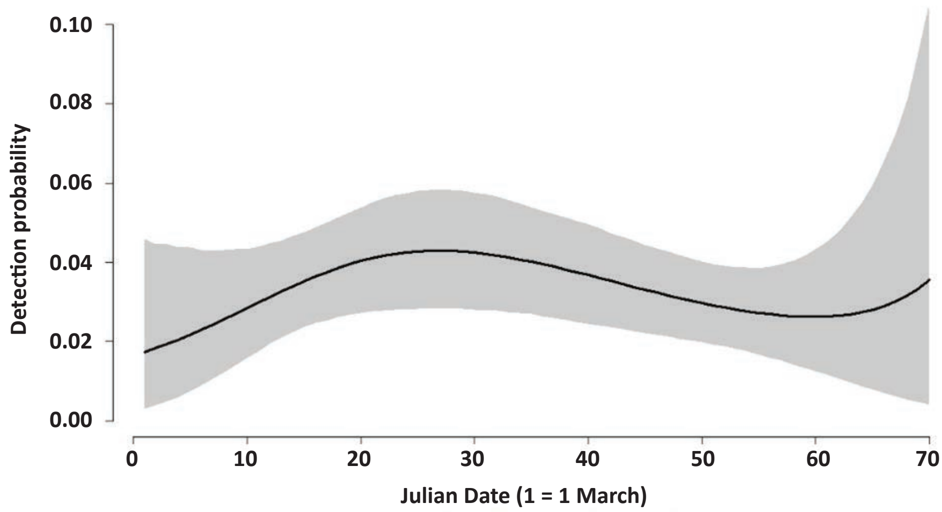


Figure 2. Effect of survey date on the detection probability of the Long-eared Owl *Asio otus* (posterior mean and 95% CRI shown).

imprecision in the duration of occasion 1 by placing a uniform prior on it in our Bayesian analysis (see below), with bounds 180 and 240 seconds, which properly accommodated the additional uncertainty stemming from this lack of knowledge.

We used Bayesian inference with MCMC methods implemented with program JAGS (Plummer 2003) and run from R via the package 'jagsUI' (Kellner 2024). We ran 4 chains for as long as needed for convergence to be reached for all parameters. This was assessed by visual inspection of traceplots and by values of Rhat that were at most 1.1 (Kéry & Kellner 2024). We present posterior means as point estimates of parameters and other estimates, and 95% quantiles of posterior distributions as 95% Bayesian credible intervals (CRIs) for an uncertainty assessment.

Results and discussion

In total, Long-eared Owl territories were found at 207 (9%) survey points. This corresponds to an observed value of the density of one territory per 3.9 km².

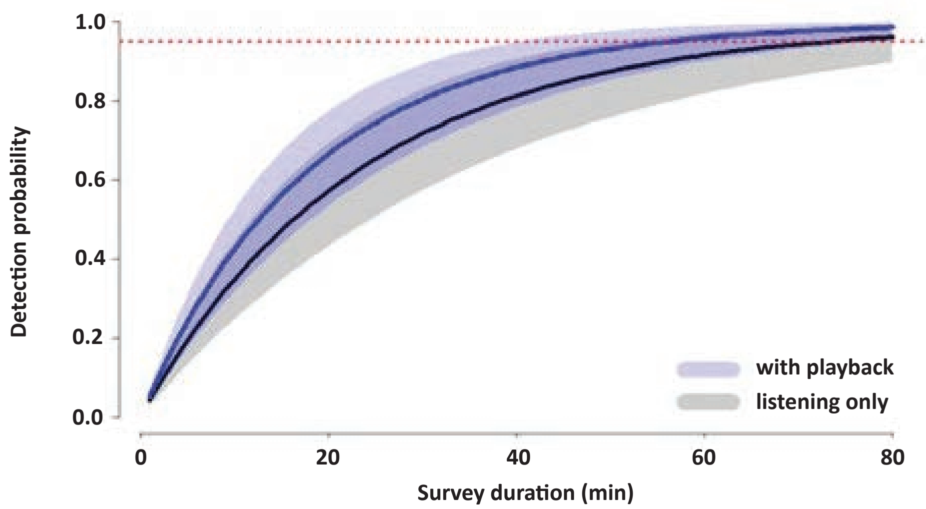


Figure 3. Effect of survey duration on detection probability of the Long-eared Owl *Asio otus*. Posterior means and 95% CRIs are shown for an occupied site with and without playback, without Tawny Owl *Strix aluco*, and averaged over all five years and on 1 April. The dotted red line indicates a detection probability of 0.95.

Occupancy modelling showed that the use of playback increased the average detection probability to 0.056 per minute (CRI 0.030–0.102) versus 0.044 (CRI 0.023–0.081) of the passive listening (Fig. 1).

Detection probability peaks from the last week of March to the first week of April (Fig. 2). The increase of detection probability at the end of the survey period around days 65–70 (early May) is an artefact caused by a quasi-lack of surveys then and the behaviour of polynomial terms in a regression model; therefore, it should be ignored.

For the design of field surveys, it is important to know how long the observer should wait at a listening point without playback to detect the species at e.g. 50% of all occasions. Figure 3 shows the model results for an average survey on 1st of April (i.e. at the peak of the vocalisation activity, Fig. 2): based on our study conditions, a survey without playback lasting 16 minutes detects approximately 50% of present territorial individuals and to detect 95% of the individuals the surveyor must spend around 74 min. With hypothetical playback of approximately 13 minutes, we would be expected to detect 50% and around 57 minutes until 95% of resident owls are detected. These numbers highlight the efficiency gains from this method; however we highly discourage the use of playback for more than a couple of minutes, otherwise it is to be feared that the territorial Long-eared Owl is stressed in an irresponsible manner or even scared off completely. With two minutes of playback, the model predicts a detection probability of app. 10%.

Long-eared Owl detection probability was reduced when Tawny Owls were present at a site (Fig. 4), likely due to avoidance behaviour by Long-eared Owls (Scott 1997, Keller et al. 2020). This is also known for other owl species (Zuberogoitia et al. 2008, Sergio et al. 2009, Lourenco et al. 2013, Theux et al. 2022). When a Tawny Owl was present at a site, detection probability was estimated to peak at 0.044 (see Fig. 1 and text above), while with Tawny Owl it was 0.035 (CRI 0.016–0.070).

For the whole studied area (covering 54% of the canton, excluding lake surface) our occupancy model estimated 415 territories (95% CRI 345–504). This number exceeds earlier estimates by Weggler et al. (2009) — 59 territories for the entire canton by far. While an increase in populations size of the Long-eared Owls may explain the increased abundance, it is highly unlikely that the species has increased thirteen times in 12–15 years. We attribute the higher abundance mainly to the systematic and species-specific surveys in combination with occupancy analysis that takes imperfect detection into account.

Results of our study have confirmed the above-mentioned earlier studies that playback-based surveys can substantially improve our capacity to monitor some cryptic owl species. In combination with systematic and broad-scale surveys and analysed with appropriate statistical methods such as occupancy models, they maximise their potential to improve fundamental

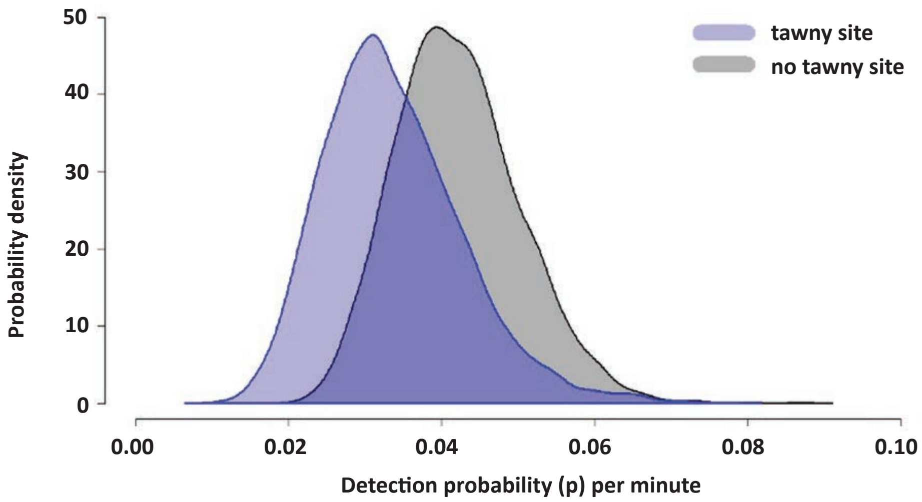


Figure 4. Posterior distributions of the effect of Tawny Owl *Strix aluco* on the detection probability of the Long-eared Owl *Asio otus* during an average year, without playback and on 1 April.

baseline information of cryptic species. Reliable estimates of distribution and abundance form the cornerstone of understanding species status, trends and habitat use, which are important in the context of population biology, as well as conservation. On the other hand, it is important to emphasize that playback can cause bias in population estimates due to changing bird behaviour (Anich et al. 2009), so it is important to follow the methodology, as well as ethical aspects (Sibley 2011, Davis et al. 2024) strictly.

These methods provide a repeatable framework for long-term monitoring and can be adapted for other cryptic species with low natural detectability. The findings also underscore the need for carefully planned survey timing and for the awareness of interspecies interactions, which can likewise bias survey results if not accounted for.

Acknowledgments

The authors thank the numerous dedicated volunteers — only their efforts made this study possible. We also thank the Cantonal Nature Conservation Office (Fachstelle Naturschutz) of the canton of Zurich for financial support.

References

Aebischer, A. 2008. Eulen und Käuze. Auf den Spuren der nächtlichen Jäger. Bern, Stuttgart, Wien.

Anich, N. M., Benson, T. J. & J. C. Bednarz. 2009 Estimating territory and home-range sizes: do singing locations alone provide an accurate estimate of space use? *Auk*, 126: 626–634.

Birrer, S. 2009. Synthesis of 312 studies on the diet of the Long-eared Owl *Asio otus*. *Ardea*, 97 (4): 615–624.

Birrer, S. 2014. Reaction of the Long-eared Owl *Asio otus* on imitations of wing-clapping and playback of mating calls — consequences for the population censuses. *Vogelwarte*, 52: 111–117.

Davis, R. A., Greenwell, C., Davis, B. J. & P. W. Bateman. 2024. Liked to death: the impacts of social media and photography on biodiversity. *Science of The Total Environment,* 949: 1–7.

Grinde, A. R., Niemi, G. J. & M. A. Etterson. 2018. Effects of experimental playbacks on availability for detection during point counts. *Journal of Field Ornithology*, 89 (1): 93–104.

Hannah, K. 2009. Call Playbacks increase detection rates of Northern Hawk Owl in recent burns. *Journal of Raptor Research*, 43 (3): 241–244.

Keller, V., Herrando, S., Voříšek, P., Franch, M., Kipson, M., Milanesi, P., Martí, D., Anton, M., Klvaňová, A., Kalyakin, M. V., Bauer, H.-G. & R. P. B. Foppen. 2020. European Breeding Bird Atlas 2: Distribution, Abundance and Change. European Bird Census Council & Lynx Editions, Barcelona.

Kellner, K. 2024. jagsUI: A Wrapper Around 'rjags' to Streamline 'JAGS' Analyses. doi: 10.32614/CRAN.package.jagsUI, R package version 1.6.2

Kéry, M., Royle, J. A., Hallman, T., Robinson, W. D., Strebel, N. & K. F. Kellner. 2024. Integrated distance sampling models for simple point counts. *Ecology*, 105 (5): e4292.

Kéry, M. & K. F. Kellner. 2024. Applied Statistical Modeling for Ecologists — A practical guide to Bayesian and likelihood inference using R, JAGS, NIMBLE, Stan, and TMB. Elsevier, London.

Kéry, M. 2023. Occupancy and Nmix models for the French national PCM survey (Pygmy Owl, Boreal Owl, Tawny Owl). Unpublished report. Sempach, Vogelwarte. 143 p.

Kissling, M. L., Lewis, S. B. & G. Pendleton. 2010. Factors influencing the detectability of forest owls in south-eastern Alaska. *Condor*, 112: 539–548.

Knaus, P., Antoniazza, S., Wechsler, S., Guélat, J., Kéry, M., Strebel, N. & T. Sattler. 2018. Schweizer Brutvogelatlas 2013–2016. Sempach.

Lourenco, R., Goytre, F., del Mar Delgado, M., Thornton, Michael; Rabaça, J. E., V. Penteriani. 2013. Tawny owl vocal activity is constrained by predation risk. *Journal of Avian Biology*, 141: 461–468.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Hines, J. E. & L. L. Bailey. 2017. Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence (Second Edition). San Diego: Elsevier.

Martinez, J. A. & I. Zuberogoitia. 2003. Factors affecting the vocal behaviour of Eagle Owl *Bubo bubo*: effects of season, density and territory quality. *Ardeola*, 50: 255–258.

McGarigal, K. & J. Fraser. 1985. Barred Owl responses to recorded vocalizations. *Condor*, 87 (4): 552–553.

McGregor, P. K. 2000. Playback experiments: design and analysis. *Acta ethologica*, 3: 3–8.

Mori, E., Menchetti, M. & F. Ferretti. 2014. Seasonal and environmental influences on the calling behaviour of Eurasian Scops Owls. *Bird Study*, 61: 277–281.

Plummer, M. 2003. JAGS: A program for analysis of bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), Proceedings of the 3rd International Workshop in Distributed Statistical Computing (DSC 2003) (pp. 1–10). Vienna, Austria: Technische Universität, March 20–22.

Ritschard, M. 2020. Waldohreulen-Monitoring Kanton Zürich. Ergebnisse 2020. Bericht z. Hd. der Fachstelle Naturschutz Kanton Zürich.

Ritschard, M. 2022. Waldohreulen monitoring Kanton Zürich. Ergebnisse 2021. Bericht z. Hd. der Fachstelle Naturschutz Kanton Zürich.

Ritschard, M. 2023a. Bestandserhebung Waldohreule Kanton Zürich. Ergebnisse 2022. Bericht z. Hd. der Fachstelle Naturschutz Kanton Zürich.

Ritschard, M. 2023b. Bestandserhebung Waldohreule Kanton Zürich. Ergebnisse 2023. Bericht z. Hd. der Fachstelle Naturschutz Kanton Zürich.

Ritschard, M. 2024. Bestandserhebung Waldohreule Kanton Zürich. Ergebnisse 2024. Bericht z. Hd. der Fachstelle Naturschutz Kanton Zürich.

Robb, M. 2015. Undiscovered owls. A sound approach guide. UK.

Scott, D. 1997. The Long-eared Owl. London.

Sergio, F., Marchesi, L. & P. Pedrini. 2009. Conservation of Scops Owl *Otus scops* in the Alps: relationships with grassland management, predation risk and wider biodiversity. *Ibis*, 151: 40–50.

Sibley, D. 2011. The Proper Use of Playback in Birding. https://www.sibleyguides.com/2011/04/the-proper-use-of-playback-in-birding.

Strebel, N., Antoniazza, S., Auchli, N., Birrer, S., Bühler, R., Sattler, T., Volet, B., Wechsler, S. & M. Moosmann. 2024. *Zustand der Vogelwelt in der Schweiz: Bericht 2024*. Schweizerische Vogelwarte, Sempach.

Swiss Land Use Statistics 2024: land cover (NOLC04) by district and commune, in hectares. Bundesamt für Statistik, Bodennutzung https://www.bfs.admin.ch/bfs/de/home/statistiken/raum-umwelt/bodennutzung-bedeckung.html (v. 30.08.2024).

Theux, C., Klein, N., Garibaldi, E., Jacot, A., Eichhorn, S., Guisan, A. & J.-N. Pradervand. 2022. Food and habitats requirements of the Eurasian Scops Owl (*Otus scops*) in Switzerland revealed by very high-resolution multi-scale models. *Ibis*, 164: 240–254.

Vrezec, A. & I. Bertoncelj. 2018. Territory monitoring of Tawny Owls *Strix aluco* using playback calls is a reliable population monitoring method. *Bird Study*, 65: sup1: 52–62.

Weggler, M., Baumberger C., Widmer M., Schwarzenbach Y. & R. Bänziger. 2009. Zürcher Brutvogelatlas 2008 — Aktuelle Brutvogelbestände im Kanton Zürich 2008 und Veränderungen seit 1988. ZVS/BirdLife Zürich.

Zuberogoitia, I., Martinez, J.E., Zabala, J., Martinez, J.A., Azkona, A., Castillo, I. & S. Hidalgo. 2008. Social interactions between two owl species sometimes associated with intraguild predation. *Ardea*, 96: 109–113.

Zuberogoitia, I., Burgos, G., Gonzales-Oreja, J. A., Morant, J., Martinez, J. E. & J.Z. Albizua. 2019. Factors affecting spontaneous vocal activity of Tawny Owl *Strix aluco* and implications for surveying large areas. *Ibis*, 161: 495–503.

Received 18rd June 2025
Accepted 13th August 2025

Bird Census News 2025, 38/1–2: 37–42

Maximizing call detections of Boreal Owl by combining Kaleidoscope PRO and BirdNET

Marina Klaas-Fábregas[a*], Alba Estrada[a], Iván Afonso[b], Irene Figueroa[c], Martí Franch[d], Rafael López del Río[e], José María Martínez[f], Blas Molina[g], Juan Carlos Del Moral[g], Diego Villanúa[h] and Jesús Martínez-Padilla[a]

[a] Instituto Pirenaico de Ecología (CSIC). Avenida Nuestra Señora de la Victoria, S/N. 22700, Jaca, Spain
[b] Conselh Generau d'Aran. Plaça d'Aran, 1-2. 25530, Vielha, Spain
[c] Associació Paisatges Vius. Carrer Padró, 10. 08511 L'Esquirol, Spain
[d] Servei de Fauna i Flora, Generalitat de Catalunya. Carrer del Foc, 57. 08038, Barcelona, Spain
[e] Servicio de Biodiversidad, Dirección General de Medio Natural, Caza y Pesca. Departamento de Medio Ambiente y Turismo. Gobierno de Aragón. Plaza de San Pedro Nolasco, 7. 50071, Zaragoza, Spain
[f] Sección de Biodiversidad. Servicio Provincial de Medio Ambiente y Turismo. Avenida de la Paz, 5. 22004, Huesca, Spain
[g] SEO/BirdLife. Calle de Melquíades Biencinto, 34. 28053, Madrid, Spain
[h] Gestión Ambiental de Navarra (Orekan). Calle del Padre Adoáin, 219 bajo. 31015, Pamplona, Spain

*Corresponding author: m.klaas@ipe.csic.es; mklaasfabregas@gmail.com

Abstract. Ecoacoustics is an effective technique for monitoring elusive species, but generates large datasets that require processing and validation. In 2024, we deployed passive audio recordes to study the distribution of Boreal Owl (*Aegolius funereus*) in the Spanish Pyrenees. We assessed the performance of Kaleidoscope PRO and BirdNET against 25 manually-reviewed audio recorders. By combining both software, we detected 95% of positive audio recorders identified by manual revision. To maximize the detections of Boreal Owl calls, we recommend a two-step protocol based first on Kaleidoscope PRO with an advanced classifier, followed by BirdNET. We recommend developing a species-specific approach when aiming to study distribution or behavioural traits of different taxa.

Introduction

Ecoacoustics is a relatively new field of research that is constantly evolving (Stowell & Sueur 2020). In this discipline, Passive Acoustic Monitoring (PAM) has proven to be a useful tool for wildlife monitoring, and can help answer questions within the fields of biogeography, population dynamics, or animal behaviour (Ross et al. 2023; Stowell & Sueur 2020). More precisely, PAM has become a widely used technology to monitor uncommon or cryptic species (Campos-Cerqueira & Aide 2016; Freitas et al. 2023; Gibb et al. 2019).

The Boreal Owl (*Aegolius funereus*) is an elusive nocturnal raptor, and the Spanish Pyrenees is the meridional limit of its distribution (Keller et al. 2020). The low detectability and scattered distribution constrain its monitoring at population and individual level. As a consequence, knowledge about its distribution, abundance and ecology in Spain is very limited, affecting its conservation strategies.

During 2024, we designed and carried out the first national survey of Boreal Owl in Spain. Passive audio recorders were set along the Spanish Pyrenees, in order to update its distribution and abundance (Martínez-Padilla et al. 2024). PAM has proven to be a very useful technique to monitor owl species and to increase their detection rates in other regions (Freitas et al. 2023, Shonfield et al. 2018). Nevertheless, PAM has some constraints concerning the management of large amounts of data (Gibb et al. 2019). Audio recorders can record for several months with high frequency intervals accumulating large quantity of audio data, which requires a considerable amount of time and resources for data processing (Nieto-Mora et al. 2023). The use of artificial intelligence (AI) has considerably increased in the last few years in order to maximize the detection of targeted species, optimizing the time spent reviewing and improving data processing (Nieto-Mora et al. 2023). However, the reliability of AI differs between species (Ruff et al. 2021, Nieto-Mora et al. 2023, Manzano-Rubio et al. 2022, Kahl et al. 2021), and enhancing the specificity and sensibility of species detection is essential to improve the robustness of the analysis outcomes. Kaleidoscope PRO is a commercial software designed for bioacoustics analysis. One of its functionalities, cluster analysis, can be used for targeted searches of acoustic patterns using Hidden Markov Models, detecting and sorting signals of given characteristics within the audio files. Cluster analysis can also be combined with the use of "bait files", improving the detection rate by augmenting the statistical density of the target sounds. In addition, cluster analysis allows advanced classifiers to be built from training data, enabling a better isolation of the target species and labelling the detections with custom tags (Wildlife Acoustics, Inc. 2023). Alternatively, BirdNET is a deep neural network trained for the sound detection of more than 900 bird species (Kahl et al. 2021). One of its multiple features include the detection of target species by providing a specific species list, with a differing performance depending on the species to detect, its distance to the call source and the recorder itself (Manzano-Rubio et al. 2022, Pérez-Granados 2025, Wood & Kahl 2024). Once analysed, BirdNET provides detections in three seconds splits, with a confidence score for each one (Kahl et al. 2021, Wood & Kahl 2024). Despite its usefulness, both Kaleidoscope PRO and BirdNET require human validation of the results, since it is usual to find false-positive detections (Pérez-Granados et al. 2024, Tseng et al. 2025, Wood & Kahl 2024). Therefore, the reliability of results of both methods demands human validation.

The aim of this study is assessing the performance of Kaleidoscope PRO and BirdNET in detecting Boreal Owl calls compared to human-expert detection, in order to outline a species-specific protocol for its monitorization using passive audio recorders (PAM). Specifically, we aimed to i) evaluate the performance of several cluster analyses in Kaleidoscope PRO; ii) develop an advanced classifier with Kaleidoscope PRO; iii) assess the performance of BirdNET at different confidence thresholds; iv) evaluate the performance of BirdNET in recorders not detected by Kaleidoscope PRO; and v) quantify the performance of both software together.

Material and methods

Collection of acoustic data

We used two types of passive audio recorders, Audiomoth and Song Meter Micros, in order to record the Boreal Owl male calls in the Spanish Pyrenees. The devices were deployed in subalpine forests during the breeding season (March – April) for a mean of 19 days (range: 6 to 42 days). They were programmed to record 90 and 270 minutes before and after sunset, respectively, and 150 and 45 minutes before and after sunrise, respectively. From all the recorders installed, we selected 25 and reviewed them manually (visualization and earing) in order to look for Boreal Owl calls. After human validation, 20 of them were positive and five of them negative to the presence of the species. The amount of manual validated files was equivalent to 1 TB.

Kaleidoscope PRO

We assessed the performance of three cluster analysis methods of Kaleidoscope PRO 5.6.6 (Wildlife Acoustics, 2024) in the 25 manually reviewed recorders. The methods evaluated were: i) signal extraction and clustering; ii) signal extraction and clustering with baits; and iii) advanced classifier. Each analysis took a rough of 12–15 minutes (not including the validation process). The signal parameters settled were: 400 and 900 Hz (minimum and maximum frequency range, respectively); 0.1 and 7.5 seconds (minimum and maximum length of detection, respectively); a maximum inter-syllable gap of 0.35 seconds; a maximum distance from cluster centre of 1.0; a Fast Fourier Transform Window of 21.33 ms; and default settlings for: maximum states, maximum distance to cluster centre for building clusters, and maximum clusters. Likewise, the advanced classifier was built with the same parameters and 553 audio files with calls from the same study area. The bait files were selected from a dataset of recordings collected in 2023 from the same study area.

For each method and each recorder, we checked the detections of the first four clusters, and com-

Table 1. Comparison of performance of cluster analysis tools in Kaleidoscope PRO for positive audio recorders. Method 1: signal extraction and clustering; Method 2: signal extraction and clustering with baits; Method 3: advanced classifier.

Kaleidoscope PRO methods	Method 1	Method 2	Method 3
Positive audio recorders detected	6	4	12
Positive audio recorders not detected	14	16	8
Success rate	30	20	60

pared the results with the results of the manual review. If a Boreal Owl call was found in at least one detection of the first four clusters, the recorder was considered positive to the presence of the species. If not, the recorder was considered negative to the presence of the species.

BirdNET

Using the same 25 audio recorders, we analysed the accuracy of the detections of Boreal Owl in BirdNET in a range from 0 to 0.99 confidence. We manually reviewed the three-second detections delivered by the programme, and noted: 1) whether the detection was a *true positive* (a Boreal Owl present in the audio fragment and correctly identified by BirdNET as a Boreal Owl) or a *false positive* (any sound different to a Boreal Owl but incorrectly identified by BirdNET as a Boreal Owl), and 2) the confidence score given by BirdNET for both true and false positive detections. We then evaluated the relationship between the confidence score of the detections and their classification as true positive or false positive, by applying a logistic regression model using a generalized linear model (GLM) with a binomial distribution in R 4.2.2 (R Core Team 2024). In such models, “detection” (1: true positive; 0: false positive) was the dependent variable and confidence the explanatory term.
Secondly, we selected the *false negative* audio recorders (Boreal Owl presence confirmed manually, but considered negative by Kaleidoscope PRO) from the 25 sample. We analysed them with BirdNET and a custom species list to compare its performance in relation to Kaleidoscope PRO and manual reviewing. Each analysis took about three hours, and the inference settings were: 0.5 minimum confidence threshold, 1 sensitivity, 0 overlap, 0 Hz minimum bandpass frequency and 1500 Hz maximum bandpass frequency. If a true positive was found in at least one detection of BirdNET, the recorder was considered to be correctly identified as positive to the presence. To the contrary, a recorder was considered negative to the presence if there were no detections of Boreal Owl, or no true positive detections.

Results

Kaleidoscope PRO

The advanced classifier built with Kaleidoscope PRO was the cluster analysis method with the best performance, by detecting the 60% (n=12 out of 20 positive audio recorders; Table 1) of the positive audio recorders of the sample. The signal extraction and clustering tool was not efficient, as it only detected the 30% (n=6) of the positive audio recorders. The use of baits did not improve the detection rates, as it detected the 20% (n=4) of the positive audio recorders. As a consequence, eight out of 20 positive audio recorders were incorrectly considered by Kaleidoscope PRO as negatives to the presence of Boreal Owl. Furthermore, none of the three methods found Boreal Owl calls in the five audio recorders that were negative to the presence of the species by manually reviewing, validating the negative results.

BirdNET

Concerning the detections of BirdNET and its confidence, we found that detections were associated with confidence interval (estimate 10.108 ± 0.264, $z=38.34$, $p<0.001$). Specifically, both true positives and false positives could be found at both very low and very high confidence scores. Noteworthy, in our analysis, the number of true positive detections reached a maximum specially from a 0.5 confidence (Fig. 1).
On the other hand, BirdNET was able to detect the 87.5% (n=7 out of 8) of the positive audio audio recorders that were not detected by Kaleidoscope PRO.
The combination of the best method of Kaleidoscope PRO (in this case, the advanced classifier)

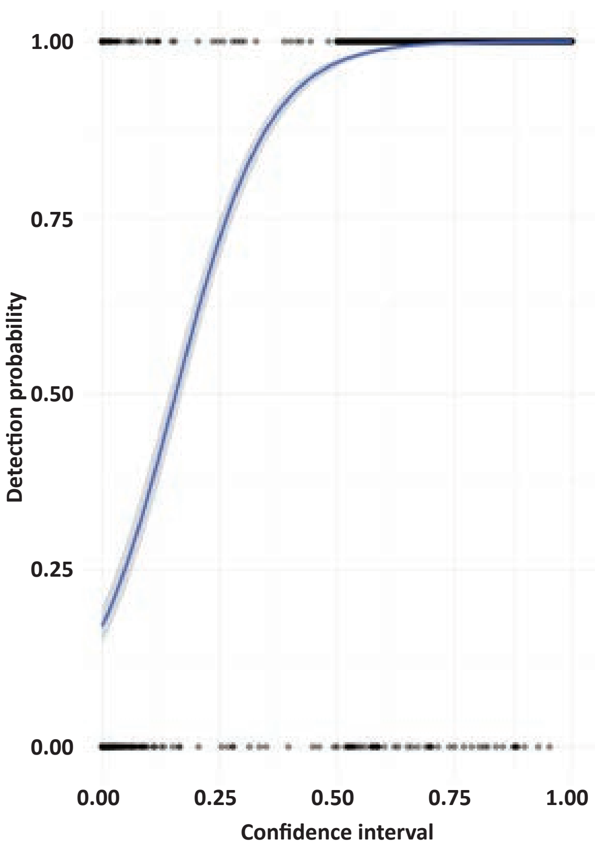


Figure 1. The positive relationship between BirdNET confidence scores and the probability of a true detection for Boreal Owl vocalisations. The blue line represents the predicted probability of detection as a function of confidence. The individual points represent observed detections classified as true positives (1) and false positives (0).

and BirdNET resulted in a correct classification of the positive audio recorders of 95% (n=19).

Discussion

The fast development of AI is providing researchers and managers an increasing number of efficient tools to work with. However, AI needs qualified personnel, big training datasets and a large amount of time to be trained (Manzano-Rubio et al. 2022; Pérez-Granados et al. 2024, Wood et al. 2023). Moreover, the large amount of data generated by PAM leads to a trade-off between precision and time investment. In this context, resources like Kaleidoscope PRO and BirdNET emerge as useful tools when facing constraints in time and/or specialised staff. Nevertheless, it is important to emphasize that the performance of each software varies depending on the species (Ruff et al. 2021, Nieto-Mora et al. 2023, Manzano-Rubio et al. 2022, Kahl et al. 2021), for example, between an owl that hoots or trills and a songbird that warbles. Therefore, our study stresses the need to properly characterise the calls of the target species to avoid false positive detections of the target species. Particularly, the length of the call, the pauses or gaps and the range of frequencies are crucial traits that need to be carefully determined (Wildlife Acoustics, Inc. 2023). In our study, the most frequent false positives were: 1) Tawny Owl detections (*Strix aluco*; similar range of frequencies), 2) environmental noise (wind, storms, screeching trees), 3) drumming of several species of woodpeckers (similar range of frequencies), and 4) anthropic sounds (church bells, machines, cars or humans). Regarding methodological aspects, the quality of an advanced classifier in Kaleidoscope PRO relies on the good characterisation of the sound, the performance of the signal extraction and the quality and source of origin of the audio files used for building it. Building an advanced classifier with sound files from the same study area can partially tackle these issues (Wildlife Acoustics, Inc. 2023). In the case of BirdNET, the training dataset is not publicly available, so the characteristics and quality of vocalisations used to train the algorithm remain unknown (Pérez-Granados et al. 2024). This is especially relevant when aiming to study ecological patterns, as it is necessary to determine whether the software detects the full range of vocal variation in the call of the species. In our study, BirdNET only detected the so-called ‘staccato’ song and the prolonged staccato song of male Boreal Owls. It remains unclear whether this is because the BirdNET algorithm has only been trained with these types of calls, or because other vocalisations were absent from our recordings. We did not find other types of vocalisations described in the literature, such as the delivery call, screech, peeping call, weak call, ‘chuuk’ call or hiss (Korpimäki & Hakkarainen 2012) by manual reviewing. Besides BirdNET not detecting all the calls of each positive audio recorders manually reviewed, it allowed us to correctly classify as positives the 87.5% of the positive audio recorders that had not been detected by Kaleidoscope PRO, resulting in a 95% of correct classification when using both software. The remaining underestimation of the detections of positive audio recorders by the two software may be due to the quality of the recordings; the distance between the sound source and the re-

corder; and the differences between the Boreal Owl calls of the study area and the calls used to train the advanced classifier in Kaleidoscope PRO, and the algorithm in BirdNET. However, we expect that the underestimation rate of Boreal Owl calls will decrease in the near future, with the improvement of AI and new technologies.
Given the differences in computing time between Kaleidoscope PRO and BirdNET, we recommend to design a workflow in which the audio recorders are first analysed with the advanced classifier in Kaleidoscope PRO, and secondly analysed with BirdNET if they are negative in the first stage. This combination of both tools was also recommended by Pérez-Granados et al. (2024) to detect calls of an amphibian species. Above all, when combining the use of PAMs and AIs, human validation is crucial to quantify the reliability of the detection outcome for each species, as these tools may produce a biased understanding of species distribution or individual behaviour (Pérez-Granados 2025). In conclusion, we support the use of advanced classifiers with Kaleidoscope PRO followed by BirdNET as an effective way of maximizing the detection of Boreal Owl calls in passive acoustic recorders.

Acknowledgements

We acknowledge the help of the forestry rangers of Aragón, Aran, Catalonia and Navarra, the technicians of protected areas in Catalonia, and those of Science for Forest Management, Biodiversity and Bioeconomy (CTFC) for the deployment of the passive audio recorders; Ramón Jato and Juan Carlos Albero for the bait files; Alberto Pastoriza for the configuration of the devices; and José Lahoz Monfort for help with BirdNET.
This study was funded by the European Union through an agreement between MITECO and TRAGSATEC. Job3088153_Ref. P2.C4I1.P1.S000.A2.E1, “Fauna Terrestre y Aves marinas, Especies autóctonas y Exóticas Invasoras. Mejora de Conocimiento del Estado de Conservación. TRAGSATEC agreed to SEO/BirdLife to carry out this study”. Such European funds are associated within the Recovery Plan, Transformation and Resilience (NEXTGENERATIONEU) under the following project: Conservation state of Spanish avian species included in Bird Directive, invasive species (article 24), and Spanish breeding bird atlas.

References

Campos-Cerqueira, M. & T. M. Aide. 2016. Improving distribution data of threatened species by combining acoustic monitoring and occupancy modelling. *Methods in Ecology and Evolution*, 7: 1340–1348.

Freitas, B., Bas, Y., Robert, A., Doutrelant, C. & M. Melo. 2023. Passive acoustic monitoring in difficult terrains: The case of the Principe Scops-Owl. *Biodiversity and Conservation*, 32: 3109–3132.

Gibb, R., Browning, E., Glover-Kapfer, P. & K. E. Jones. 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10: 169–185.

Kahl, S., Wood, C. M., Eibl, M. & H. Klinck. 2021. BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61: 101236.

Keller, V., Herrando, S., Voríšek, P., Franch, M., Kipson, M., Milanesi, P., Martí, D., Anton, M., Klvanová, A., Kalyakin, M. V., Bauer, H.-G. & R. P. B. Foppen. 2020. *European Breeding Bird Atlas 2: Distribution, Abundance and Change*. European Bird Census Council & Lynx Editions.

Korpimäki, E. & H. Hakkarainen. 2012. *The Boreal Owl. Ecology, Behaviour and Conservation of a Forest-Dwelling Predator*. Cambridge University Press.

Manzano-Rubio, R., Bota, G., Brotons, L., Soto-Largo, E. & C. Pérez-Granados. 2022. Low-cost open-source recorders and ready-to-use machine learning approaches provide effective monitoring of threatened species. *Ecological Informatics*, 72: 101910.

Martínez-Padilla, J., Estrada, A. & M. Klaas-Fábregas. 2024. Distribución y estima de abundancia poblacional de Mochuelo boreal (*Aegolius funereus*) en el Pirineo. *Report on the agreement between CSIC and SEO-BirdLife.*

Nieto-Mora, D. A., Rodríguez-Buritica, S., Rodríguez-Marín, P., Martínez-Vargaz, J. D. & C. Isaza-Narváez. 2023. Systematic review of machine learning methods applied to ecoacoustics and soundscape monitoring. *Heliyon*, 9: e20275.

Pérez-Granados, C. 2025. BirdNET confidence scores decrease with bird distance from the recorder: Revisiting Pérez-Granados 2023. *Ardeola*, 72: 149–159.

Pérez-Granados, C., Feldman, M. J. & M. J. Mazerolle. 2024. Combining two user-friendly machine learning tools increases species detection from acoustic recordings. *Canadian Journal of Zoology*, 102: 403–409.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing* [Software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ross, S. R. P.-J., O'Connell, D. P., Deichmann, J. L., Desjonquères, C., Gasc, A., Phillips, J. N., Sethi, S. S., Wood, C. M. & Z. Burivalova. 2023. Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. *Functional Ecology*, 37: 959–975.

Ruff, Z. J., Lesmeister, D. B., Appel, C. L. & C. M. Sullivan. 2021. Workflow and convolutional neural network for automated identification of animal sounds. *Ecological Indicators*, 124: 107419.

Shonfield, J., Heemskerk, S. & E. M. Bayne. 2018. Utility of automated species recognition for acoustic monitoring of owls. *Journal of Raptor Research*, *52*: 42–55.

Stowell, D. & J. Sueur. 2020. Ecoacoustics: Acoustic sensing for biodiversity monitoring at scale. *Remote Sensing in Ecology and Conservation*, 6: 217–219.

Tseng, S., Hodder, D. P. & K. A. Otter. 2025. Setting BirdNET confidence thresholds: Species-specific vs. universal approaches. *Journal of Ornithology,* 166: 1123–1135.

Wildlife Acoustics. 2024. *Kaleidoscope PRO* (Version 5.6.6) [Software]. Inc.: Maynard, MA, USA.

Wildlife Acoustics, Inc. 2023. *Kaleidoscope User Guide*.

Wood, C. M. & S. Kahl. 2024. Guidelines for appropriate use of BirdNET scores and other detector outputs. *Journal of Ornithology*, 165: 777–782.

Wood, C. M., Kahl, S., Barnes, S., Van Horne, R. & C. Brown. 2023. Passive acoustic surveys and the BirdNET algorithm reveal detailed spatiotemporal variation in the vocal activity of two anurans. *Bioacoustics*, 32: 532–543.

Received: 17th June 2025
Accepted: 28th October 2025

Bird Census News 2025, 38/1–2: 43–47

A Novel Approach to Estimate the Number of Territorial Birds in a Network of Spatially Distributed Acoustic Recording Units

Leonhard Brüggemann[1], Daniel Otten[1], Frederik Sachser[2] and Nils Aschenbruck[1]

[1] Osnabrück University – Institute of Computer Science, 49076 Osnabrück, Germany
[2] Austrian Research Centre for Forests, 1130 Vienna, Austria
Corresponding author: brueggemann@uos.de

I. Introduction

Modern technologies, such as the development of low-cost passive acoustic monitoring devices (PAM), e.g., AudioMoth (Hill et al., 2019), placed in natural habitats of various species, are increasingly becoming a valuable complement to traditional field studies (Pérez-Granados, 2023). At the same time, significant progress is being made in computer-based, acoustic species recognition of birds. Various innovative approaches and models, e.g., BirdNET (Kahl et al., 2021), contribute to this and are now able to identify bird species based on their call or song (relatively) reliably. Currently, the focus is mainly on identifying whether a species is present at the location where the recording was taken.

We lack methods to acquire the number of individuals of a species in recordings, as automated identification of individuals of the same species is still difficult (Knight et al., 2024). In short, we are only at the beginning of a more sophisticated monitoring approach, the potential of which can already be assessed as immense. In our research, we are taking steps towards a more complex method of bird monitoring that does not require individual identification. We fused biology and computer science knowledge to develop the algorithm TASE—Territorial Acoustic Species Estimation — published in Ecological Informatics (Brüggemann et al., 2025). Using multiple acoustic recording devices and AI-based species identification that form an acoustic sensor network, theoretical computer scientists thought (in the past) about a way to count bird individuals (Stattner et al., 2011). We put their approach into practice and encountered various difficulties or found that assumptions made, are unrealistic. However, their approach inspired us to develop a generalized, easy-to-use method — even with limited computer science knowledge. By applying that algorithm to bird acoustics, we aim to determine the number of breeding birds in spring in a specific area. Our approach was tested in practice on an area of approximately 12 hectares. Initial promising results have been shared in our presentation.

II. Deployment and problem definition

We deployed 29 AudioMoth PAMs in a nature reserve in North Rhine-Westphalia, Germany on 3 June 2023, recording soundscapes from 4:00 to 10:00 (Fig. 1). Due to their dense outdoor placement, the devices formed an Acoustic Sensor Network (ASN). Their close proximity created overlapping recording ranges, allowing us to capture vocalizations across a large, continuous area. Applying a species classifier to these recordings yields unitless classification scores (ranging from 0 to 1) for each identifiable species, e.g., Wood &

Figure 1a. Deployment area in a nature reserve in Germany, the icons refer.

Figure 1 b, c. Deployment area in a nature reserve in Germany. (b) Exemplary Scene within the forest. (c) PAM in its waterproof to the ARU AudioMoth case.

Kahl (2024), in the following referred to as *confidence scores*. A score of 0 indicates the species is undetected, while a score of 1 indicates a strong species detection. Across distributed recorders in an ASN, this generates spatially distributed confidence scores for each species.

Figure 2 shows a glimpse of the challenge we are facing: The figure illustrates the confidence score sets over time using bird data from an actual deployment. Each circle represents a recorder that is coloured by the classifier's confidence score. One bird (here Common Redstart *Phoenicurus phoenicurus*) is active in Figs. 2a and 2b in the top right corner, resulting in high confidence scores.

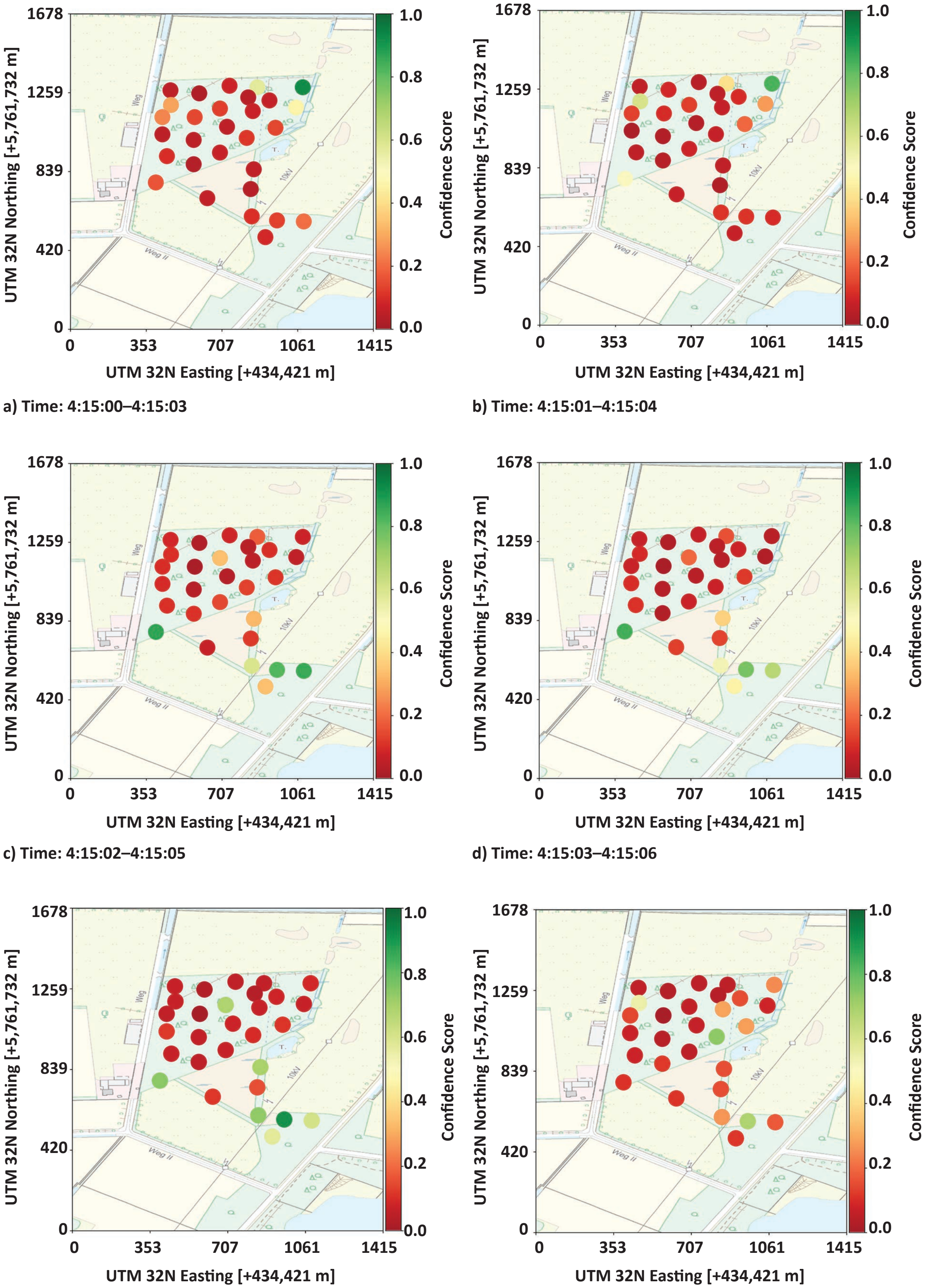
Then, it is silent in Figs. 2c to 2f, and active again in Figs. 2g and 2h. Similarly, another bird in the bottom right is initially inactive, vocalizes, and then becomes silent. The third Common Redstart in the southeast causes high confidence scores at neighboring recorders. Some recorders, such as the green recorder in the west in Fig. 2b, show high confidence scores while their neighbours do not, potentially indicating errors in the classification. The challenge is how to estimate the number of birds from such temporal confidence score sets while coping with potential misclassifications.

III. Methods

Our algorithm, called TASE, is formalized in Brüggemann et al. (2025), including requirements that must be met. Its concept is based on the observation that the classifier's confidence scores decrease with distance. In other words, PAMs closest to a vocalizing source have higher confidence scores than those farther away. If two birds vocalise simultaneously, the intermediate PAM between the two birds will have a lower confidence score than those closest to the birds. That forms a pattern that we can exploit. We can separate our recorders into distinct groups corresponding to individual vocal sources. Repeating that over long periods, we identify occupied areas, the territories. Eventually, TASE acquires a spatio-temporal point cloud and, by calculating a kernel density estimate, captures territorial spatial patterns.

IV. Exemplary results

Figure 3 shows an exemplary result, comparing the expert's field monitoring (in red and gray coloured circles) and the density map of our approach (green colors). Notably, we see that the expert assessment and the density map overlap to a large extent with the high-density areas, showing our method's efficiency. However, some green spots lie outside the circles for three reasons: (1) The circles are only a crude estimate of the individual territory, whose actual borders are unknown. Territorial behavior is quite complex, and the species might move more dynamically in the area. (2) Non-territorial birds being present



a) Time: 4:15:00–4:15:03

b) Time: 4:15:01–4:15:04

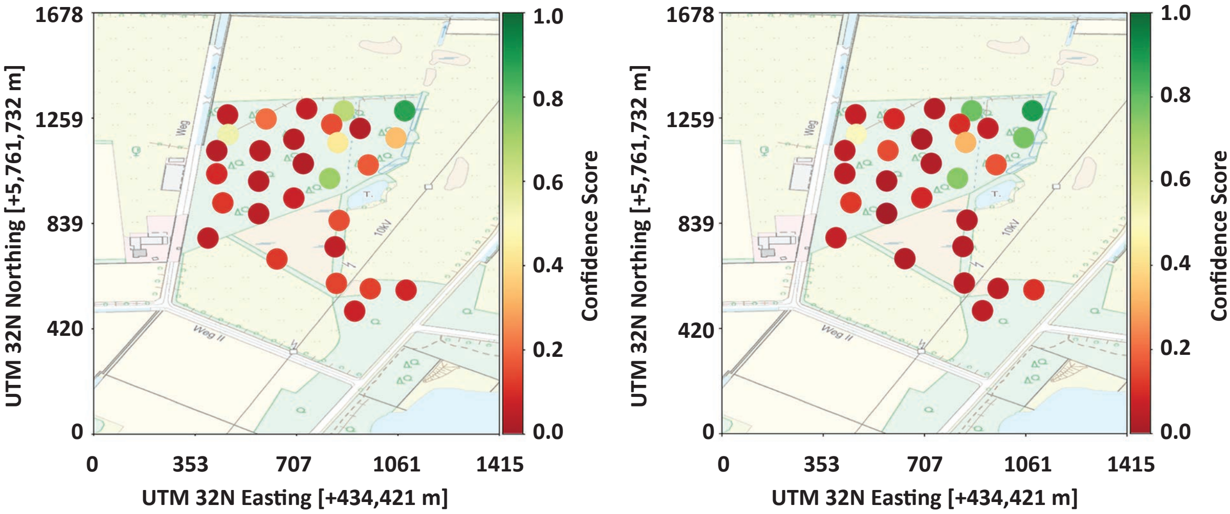
c) Time: 4:15:02–4:15:05

d) Time: 4:15:03–4:15:06

e) Time: 4:15:04–4:15:07

f) Time: 4:15:05–4:15:08

Figure 2. Illustration of the Challenge: A real-world ASN of Common Redstarts *Phoenicurus phoenicurus* observations and their identification confidence scores in North Rhine-Westphalia, Germany on 3 June 2023 in three second intervals.



g) Time: 4:15:06–4:15:09

h) Time: 4:15:07–4:15:10

Figure 2 continued. Illustration of the Challenge: A real-world ASN of Common Redstarts *Phoenicurus phoenicurus* observations and their identification confidence scores in North Rhine-Westphalia, Germany on 3 June 2023 in three second intervals.

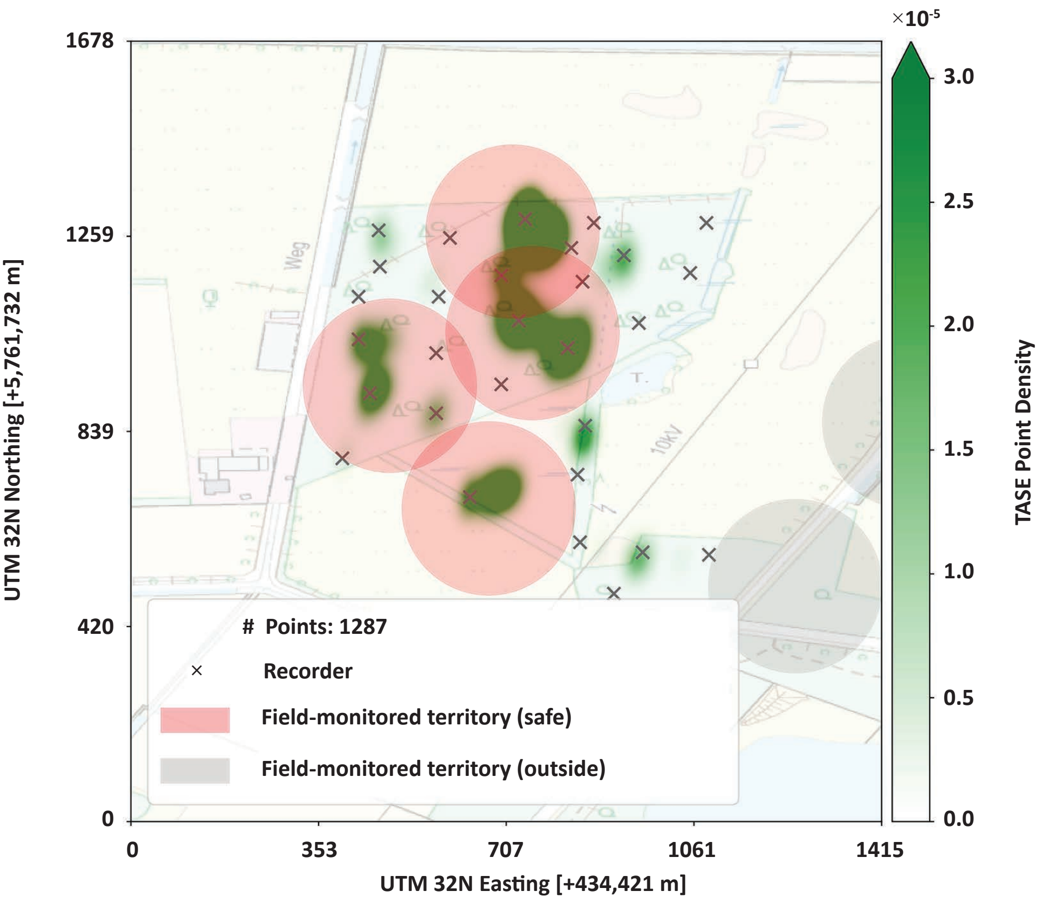


Figure 3. Kernel density estimate (green) of spatio-temporal detections for the Blackcap *Sylvia atricapilla* generated by TASE, with the expert's territorial assessment shown by red and gray circles.

in the area vocalize. (3) Our approach requires setting some parameters that might cause methodological errors.

In Brüggemann et al. (2025), we provide a detailed proof-of-concept evaluation of our approach, focusing on eleven species and discussing the opportunities and limitations. The next steps include scaling up to larger, longer deployments and introducing "soft" territorial boundaries with automated spatio-temporal clustering. With these refinements, TASE could deliver reliable data in abundance estimates, whereas conventional monitoring is hard or impossible.

Map data sources and licensing

Base map data in this paper were provided by Geobasis NRW under the Data License Germany—Zero—Version 2.0 (DL-DE-Zero-2.0; https://www.govdata.de/dl-de/zero-2-0).

References

Brüggemann, L., Otten, D., Sachser, F. & N. Aschenbruck. 2025. Territorial Acoustic Species Estimation using Acoustic Sensor Networks. *Ecological Informatics*, vol. 91, doi: 10.1016/j.ecoinf.2025.103281

Hill, A. P., Prince, P., Snaddon, J. L., Doncaster, C. P. & A. Rogers. 2019. AudioMoth: a low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX*, 6: 1–19.

Kahl, S., Wood, C. M., Eibl, M. & H. Klinck. 2021. BirdNET: a deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61: 101236.

Knight, E., Rhinehart, T., de Zwaan, D. R., Weldy, M. J., Cartwright, M., Hawley, S. H., Larkin, J. L., Lesmeister, D., Bayne, E. & J. Kitzes. 2024. Individual identification in acoustic recordings. *Trends in Ecology & Evolution*, 39: 947–960.

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

Stattner, E., Hunel, P., Vidot, N. & M. Collard. 2011. Acoustic scheme to count bird songs with wireless sensor networks. In: Proceedings of the 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, IEEE, pp. 1–3.

Wood, C. M. & S. Kahl. 2024. Guidelines for appropriate use of BirdNET scores and other detector outputs. *Journal of Ornithology*, 165: 777–7.

Received: 27th May 2025
Accepted: 26th November 2025

Bird Census News 2025, 38/1–2: 48–55

Building future monitoring together — towards improved acoustic monitoring of European birds and bats

Otso Ovaskainen[1], Ari Lehtiö[2], Patrik Lauha[3], Panu Somervuo[3], Ossi Nokelainen[1,4], Tomas Roslin[5] and Aleksi Lehikoinen[6]

[1]Department of Biological and Environmental Science, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland
[2]Digital Services, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland
[3]Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, P.O. Box 65, 00014 Helsinki, Finland
[4]Open Science Centre, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland
[5]Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland
[6]Finnish Museum of Natural History, University of Helsinki, P.O Box 17 (Pohjoinen Rautatiekatu 13), 00014 University of Helsinki

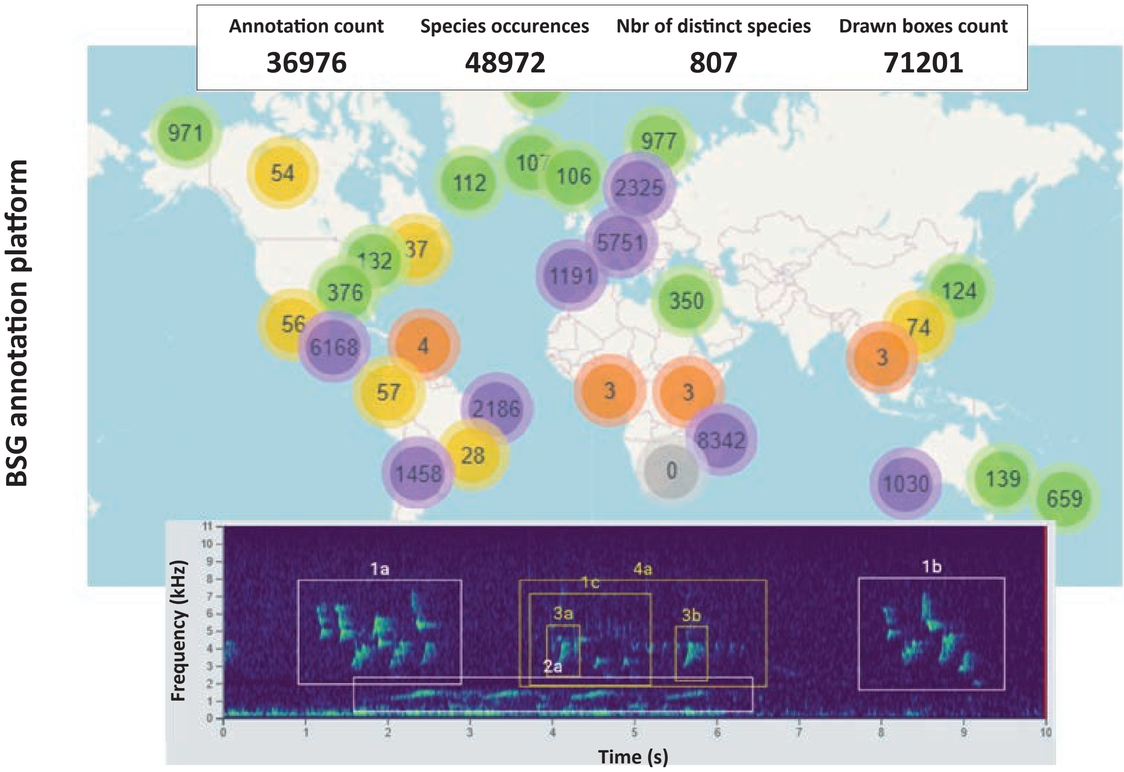
Corresponding author Otso Ovaskainen: otso.t.ovaskainen@jyu.fi

Abstract. Acoustic biodiversity monitoring is increasingly popular, but its wide and effective implementation across Europe remains challenging. To clear current obstacles, we propose new forms of collaboration among European bird and bat researchers. First, we suggest building improved bird and bat classifiers through a portal that facilitates joint annotation efforts and end-to-end workflows for model building. Second, we suggest using new technology for streamlining the prevailing "collect first and analyze later" -paradigm to achieve truly real-time audio monitoring. Third, we suggest using new smartphone apps and digital twinning to enable also non-expert citizens to contribute to real-time monitoring.

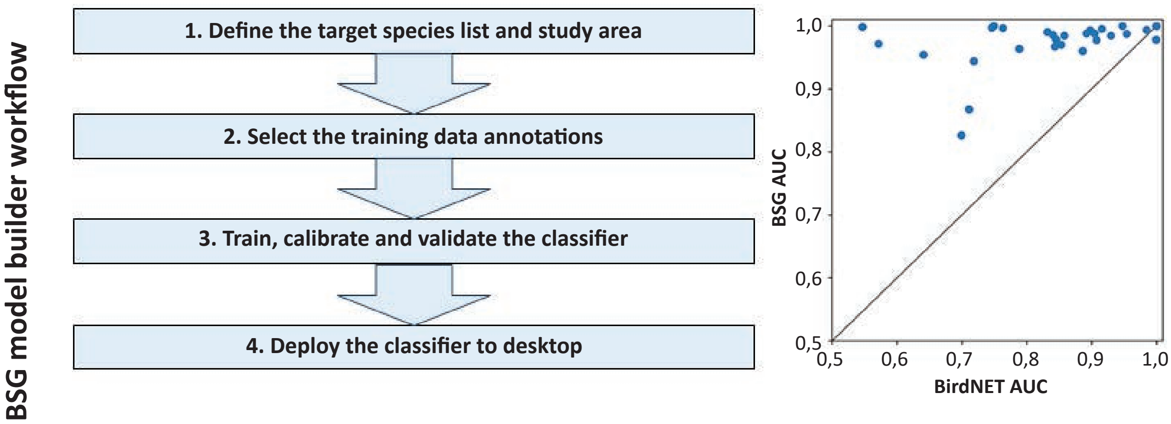
We present our suggestion for increased collaboration within the European (and worldwide) acoustic community in terms of three challenges and three potential solutions. These address (1) the need for more reliable classifiers for acoustic monitoring; (2) the need for more streamlined workflows for rapidly converting acoustic data to research and monitoring outputs; and (3) untapping the full potential of citizen science in (real-time) biodiversity monitoring. We believe that all these challenges are solvable but reaching them requires coordinated joint efforts. We hope you are interested in joining these efforts in one way or another — if so, we warmly invite you to contact any of the authors (e.g., by emailing to the corresponding author) to express your interest.

Challenge 1: We need more reliable classifiers for acoustic monitoring

Passive acoustic monitoring (PAM) provides unprecedented potential for research and monitoring of birds, bats and other vocal animals. PAM can provide massive data, such as >100 years of audio generated by the global biodiversity sampling campaign LIFEPLAN (Hardwick et al. 2024; Somervuo et al. 2025). The massive size of the data is a blessing and a curse. It is a blessing, because the information content of the data increases with its size. It is a curse, because the logistic challenges related to storing, managing, and analyzing audio also increase with the size of the data. The primary challenge in processing PAM audio is how to reliably classify the species that vocalize in the audio. Manual analysis requires expensive expert time and results in loss of information if (and when) one needs to resort to subsampling. For large-scale studies, the use of automated classifiers is in practice the only feasible option for processing the audio. Generic and openly accessible classifiers such as BirdNet (Kahl et al. 2021) are widely used, but their performance varies greatly from case to case. Specific performance depends on how well the data used to train the classifier corresponds to the audio to be analyzed in terms of the species to be classified, the type of background noise, as well as the recording devices used. To resolve this, finetuned



Example soundscape annotation made by a citizen scientist at Bird Sound Global (BSG): Pied Flycatcher (1), Arctic Loon (2), Chaffinch (3) and Willow Warbler (4).



The performance of BSG model builder approach exemplified by the baseline Bird-NET model finetuned for Central Europe birds.

Figure 1. The BSG annotation platform and BSG model builder. The BSG annotation platform enables expert users to annotate 10-second soundscapes by listing the species that vocalize in the soundscape and drawing bounding boxes around the vocalizations. These annotations are fed into the BSG model builder, which finetunes the baseline BirdNET model to the desired species community. As exemplified in the figure with a model finetuned for Central European birds, the approach leads to improved classification performance.

models can be constructed to optimize the performance for each specific case study (Lauha et al. 2022). However, developing finetuned models has remained tedious, as it requires substantial annotation effort to generate the training data, as well as substantial expertise in machine learning to successfully train an optimally performing classifier.

Solution 1: Bird and Bat Sound Global — model builder

We suggest developing improved acoustic classifiers through community efforts. To facilitate such a process, we have developed the Bird and Bat Sound Global (BSG; https://bsg.laji.fi/) annotation platform, where anyone who has the expertise to identify bird or bat species from their vocalizations can annotate sound clips originating from various audio sources, such as those resulting from the LIFEPLAN sampling campaign. The annotators can generate so-called strong labels by not only specifying the species vocalizing in the audio clips but also drawing bounding boxes around the vocalizations in time-frequency space (Fig. 1). Such strong labels are essentially useful for finetuning classifiers (Hershey et al. 2021). Currently, the bird section of BSG has resulted in annotations of 37,000 ten-second soundscape clips, involving 71,000 bounding boxes that represent 800 bird species. The bat section of BSG has resulted in annotations of 4,000 ultrasound clips, involving 5,600 bounding boxes that represent 30 bat species. The bird annotations have global coverage, representing especially Europe and Madagascar, whereas the bat annotations are currently restricted to Europe only. The BSG platform follows an open science model, with all annotations being published through the Xeno-canto database (Xeno-canto Foundation, 2005). Most importantly, BSG is not only an annotation platform, but contains a systematized model builder workflow for finetuning classifiers (Lauha et al. 2025). We have developed and tested the BSG model builder workflow through six bird classifiers (targeted to Finland, Central Europe, the Iberian Peninsula, Madagascar, Mexico and the Dry Chaco of Argentina; Lauha et al. 2025) and a European bat classifier (Meramo et al. 2025), in each case demonstrating improved performance compared to alternative classifiers. By enabling each user to annotate soundscapes for their own research needs, while simultaneously contributing to a shared global database of reference annotations, BSG fosters a global community effort to advance automated bird sound classification. ***We invite researchers to take full advantage of these new resources, to thereby jointly develop more reliable classifiers for birds and bats within and outside Europe.***

Challenge 2: The workflow from collecting acoustic data to research and monitoring outputs needs streamlining

The lack of sufficiently reliable classifiers is not the only bottleneck with audio monitoring. Another, yet related, bottleneck is the resource intensity of the entire process from field work to the delivery of the actual monitoring outputs. For example, collecting the >100 years of global audio data in the LIFEPLAN project required (1) hundreds of researchers to visit the sampling sites on a weekly basis to change the memory cards and batteries of the AudioMoth devices (Hill et al. 2018); (2) the upload of the data to a common server through internet connection or shipping hard drives; (3) excessive amount of storage capacity to host ca. 1000 GB of audio; (4) several months of computational time in a high-performance environment to classify the audio in terms of soundscape indices and species contents; (5) manual validation of subsets of classifications to assess their reliability. Taking all these steps took several years before we could eventually statistically analyze the data to address the research questions that motivated collecting the data (e.g., Somervuo et al. 2025). Clearly, the long delay from data collection to research outputs is not rewarding, and the resource intensity of the process, in terms of researcher time and access to storage and computational facilities, limits the broad applicability of large-scale passive audio monitoring.

Solution 2: Real-time audio monitoring with BirdPipe

The above-mentioned challenges can be solved with recent technological developments. For example, the open-source real-time audio monitoring device BirdPipe that we have developed classifies audio through edge computing and transmits both the classifications and the raw audio wirelessly to the researcher (Fig. 2). As a result, the researcher will receive a species detection essentially in real time, e.g. within a minute since the bird or bat vocalization actually took place in the field. In some environments, such as urban environments, recording sound may be problematic from the legislative point of view because of risk of eavesdropping. In such a case, BirdPipe can be used in pure edge computing mode, in which case the audio is recorded and classified directly in RAM-memory without storing audio information into any external memory such as disk or SD card from where it could be recovered. BirdPipe is very easy to use: all one has to do is to set up the device in the desired location and turn it on. BirdPipe will then figure out where and when it is and integrate these central metadata into the classifications. The R-package BirdPipe enables users to monitor the data flow in real time, thus facilitating quality control and

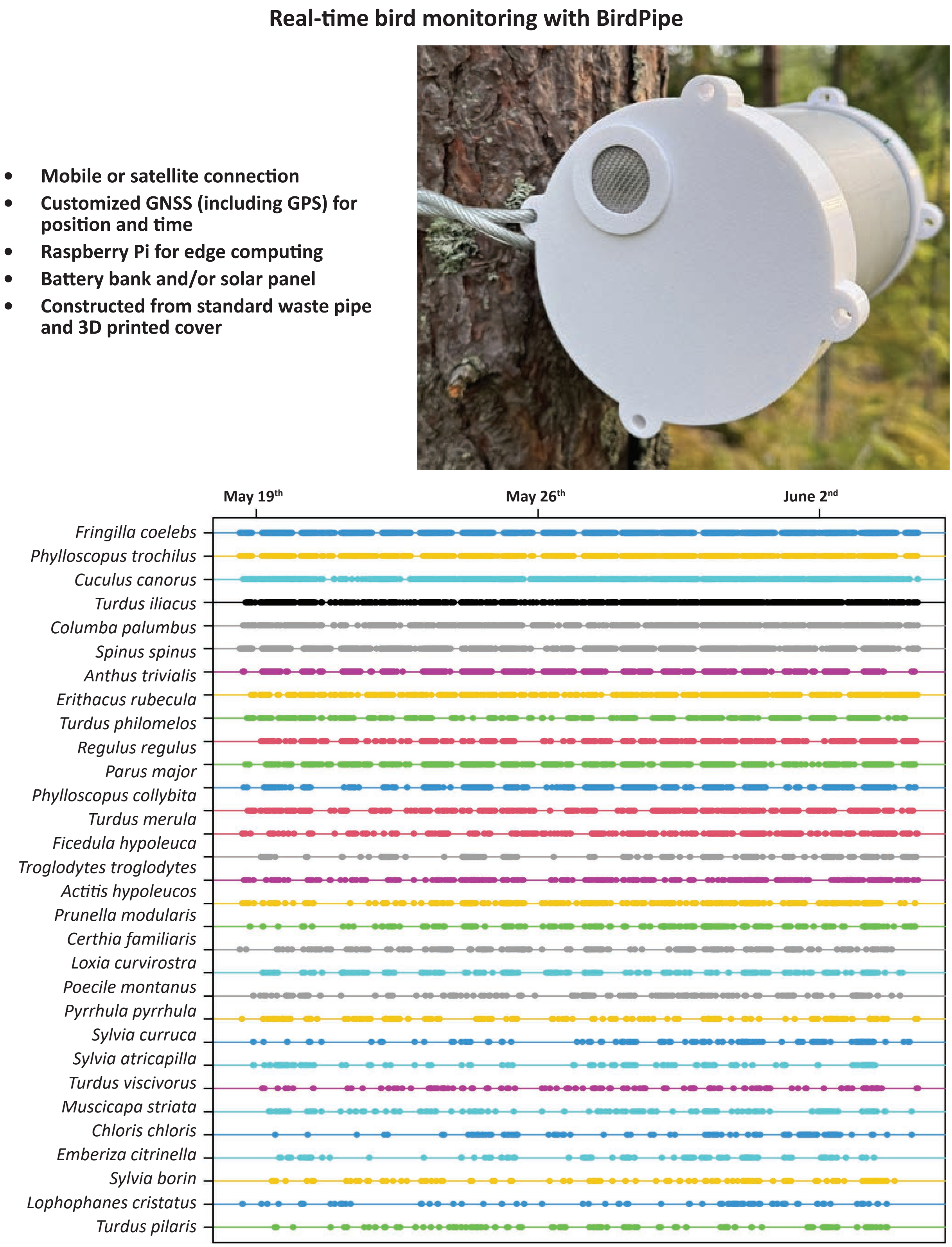


Figure 2. The open-source real-time monitoring device BirdPipe. BirdPipe is easy to use, as the user just needs to deploy it to the desired location and turn it on. BirdPipe will then find out where and when it is and start classifying birds from the soundscape using edge computing. The classifications are transmitted wirelessly, enabling the user to monitor the data flow in real time, as illustrated here for detections that accumulated over a 17-day pilot study that we conducted in 2025. The data shown are restricted to the most common 30 species.

fast response to potential problems (Fig. 2). As the BirdPipe device performs classification with machine learning, it surely makes mistakes — but less so in the future as the classifiers will improve (see Solution 1). Importantly, also the optional use of manual validation has been streamlined. As default option, the R-package BirdPipe picks only the vocalizations of highest confidence for manual validation, resulting in substantial saving in expert time. We tested BirdPipe by rotating 10 devices across 116 locations that varied in terms of forest management, using a 24-hour recording cycle in each location. Species classification by edge computing revealed a total of 23,478 detections with at least 0.3 classification probability. Utilizing BirdPipe, it took only one working day of expert time to manually validate the detection of each species at each location. Thus, the approach yielded unvalidated classifications in real-time, and a manually validated species times locations data matrix one day after the study was finished. ***We invite researchers to take full advantage of this newly emerging technology, moving from the still prevailing "collect first and analyze later" -paradigm towards audio monitoring in real time.***

Challenge 3: The full potential of citizen science in real-time biodiversity monitoring remains partially untapped

The use of citizen science in biodiversity research and especially in bird research has long traditions. This is because people are generally interested in birds, and many citizens are very skillful identifying birds in the field. A large number of initiatives have focused on converting these citizen science activities into data that are valuable for research and monitoring. Thanks to these efforts, bird data has become very widely available. For one example, the Global Biodiversity Information Facility (GBIF) database is actually dominated by bird observations (Hughes et al. 2021), and these data are widely used for research purposes (Heberling et al. 2021). For another example, the eBird platform has recruited some 1.1 million birdwatchers, who collect data using procedures that involve systematic quality control and quantification of user skills (Sullivan et al. 2009). The eBird data enables many kinds of research and monitoring outputs, such as status and trends -products provided by the Cornell Lab of Ornithology data scientists (Fink et al. 2025). In spite of all this progress, there are two perspectives from which the full potential of citizen science in bird research and monitoring still remains untapped. First, most data are currently contributed by seasoned birdwatchers, even if less skilled citizens may also spend much effort in observing birds. While many smartphone apps exist for automatically classifying birds from their audio, these apps have not yet been fully integrated into research and monitoring. For example, Merlin Bird ID leverages eBird data to develop the classification algorithms and create lists of likely birds for given locations, but the information only flows one way from monitoring outputs to citizens using the ID app. Thus, a key question concerns how citizens who themselves are unable to reliably identify birds may best partake in distributed efforts of audio monitoring. Second, there is typically a long delay from citizen science observations to research and monitoring outputs. For example, the status and trends -products based on eBird data are provided periodically by utilizing data that has accumulated over recent years (Fink et al. 2025). We argue that it is possible to streamline the full workflow from citizen science data collection to research and monitoring outputs into real-time, providing great added value not only from the point of view of research and monitoring, but also for the citizen who takes part in data collection.

Solution 3: MK smartphone app and digital twinning enables ordinary citizens to contribute to real-time monitoring

To enable ordinary citizens to take part in real-time bird monitoring, we developed a research-oriented smartphone app (Nokelainen et al. 2024) and integrated it into a digital twinning environment (Ovaskainen et al. 2026) (Fig. 3). We call the app here by the acronym MK, which refers to the app's Finnish name 'Muuttolintujen Kevät', meaning Spring of Migratory Birds. This name of the app was inherited from a nature themed program run by the Finnish national broadcasting company Yle, a key collaborator in this project. In spite of its name, the MK app is targeted not just at migratory birds but at all the 263 bird species regularly occurring in Finland, and not just for spring but for year-around use. The MK app was launched in 2023, and rapidly gained popularity among Finns, with 300,000 citizens (5% of the national population) having submitted >43 million bird detections. A key feature of the MK app is that all detections are based on a machine learning based classifier rather than users identifying the birds themselves. This en-

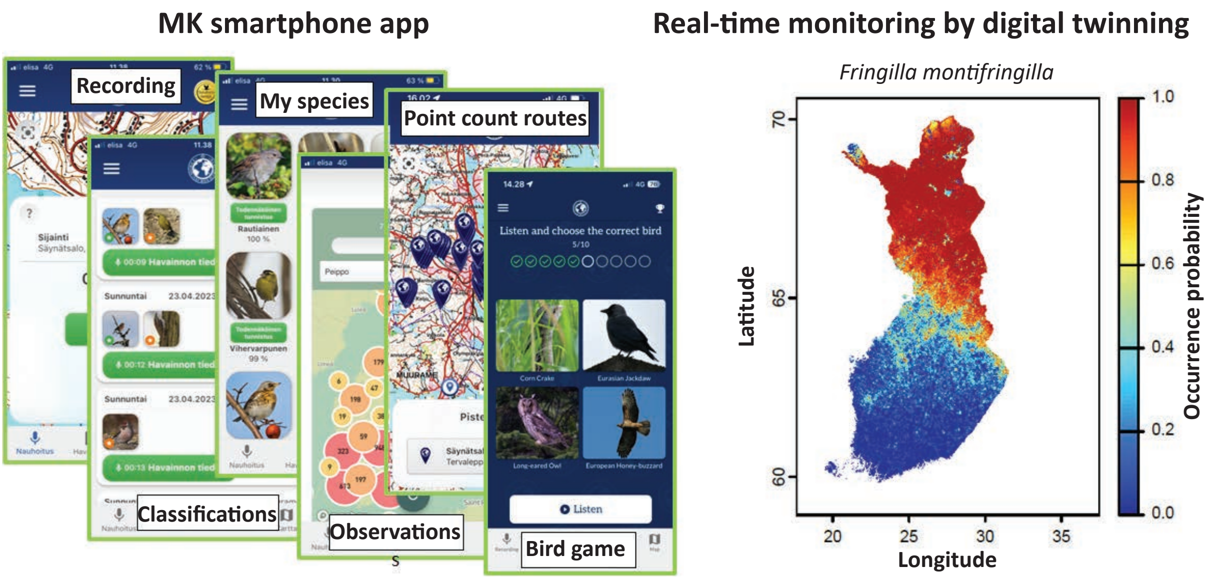


Figure 3. Real-time bird monitoring with smartphone-based citizen science and digital twinning. The MK smartphone app implements three recording modes: direct recording, interval recording, and systematic point count recording. In addition to providing engaging feedback to the user, the app submits the audio to a computational backend, which combines the data with long-term bird census data to provide daily updating species distribution maps at 1ha resolution, exemplified here by the Brambling (*Fringilla montifringilla*).

ables any citizen to provide equally valid data, whether they are able to classify the birds themselves or not. To counteract mistakes in machine learning based classifications, the backend of the database stores all the audio, enabling quality control by manual validation, and reclassifying the data with continuously improving classifiers (see Solution 1). While most detections are opportunistic (the citizens make direct recordings triggered by an interesting bird vocalization), the MK app also includes two other recording modes that are targeted to producing higher quality data that avoid some of the detection biases of opportunistic observations. First, the MK app facilitates interval recordings, enabling users to detect birds that, for example, vocalize overnight in their yard. Second, in collaboration with Finnish national parks and municipalities, we have implemented a permanent citizen science point counting network, which currently includes 580 pre-selected locations where citizens can conduct a systematic five-minute recording. To streamline the workflow from citizen scientists making observations to monitoring outputs, we have implemented a digital twinning approach, which integrates the newly accumulating MK app detections with long-term breeding bird survey data on birds (line transects), resulting in daily updating species distribution maps (Ovaskainen et al. 2026). We have successfully validated these predictions with independent data obtained by bird experts conducting manual point counts in the field (Ovaskainen et al. 2026).

If there is sufficient interest from European bird researchers to utilize this technology in citizen science projects, we would be happy to upscale the MK app and the digital twinning approach to the European scale, and to effectively integrate these data with other data sources such as EuroBirdPortal.

References

Fink, D., Auer, T., Johnston, A., Strimas-Mackey, M., Ligocki, S., Robinson, O., Hochachka, W., Jaromczyk, L., Crowley, C., Dunham, K., Stillman, A., Davis, C., Stokowski, M., Sharma, P., Pantoja, V., Burgin, D., Crowe, P., Bell, M., Ray, S., Davies, I., Ruiz-Gutierrez, V., Wood, C. & A. Rodewald, A. 2025. eBird Status and Trends, Data Version: 2023; Released: 2025. Cornell Lab of Ornithology, Ithaca, New York.

Hardwick, B., Kerdraon, D., Rogers, H. M. K., Raharinjanahary, D., Rajoelison, E. T., Mononen, T., Lehikoinen, P., Banelyte, G., Farrell, A., Fisher, B. L., Roslin, T. & O. Ovaskainen. 2024. LIFEPLAN: A worldwide biodiversity sampling design. *PLOS ONE,* 19: e0313353.

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B. & D. Schigel. 2021. Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences,* 118: e2018093118.

Hershey, S., Ellis, D. P., Fonseca, E., Jansen, A., Liu, C., Moore, R. C. & M. Plakal 2021. The benefit of temporally-strong labels in audio event classification. In ICASSP 2021–2021 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 366–370.

Hill, A. P., Prince, P., Covarrubias, E. P., Doncaster, C. P., Snaddon, J. L. & A. Rogers. 2018. AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolution,* 9: 1199–1211.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C. & H. Qiao. 2021. Sampling biases shape our view of the natural world. *Ecography*, 44: 1259–1269.

Kahl, S., Wood, C.M., Eibl, M. & H. Klinck 2021. BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics,* 61: 101236.

Lauha, P. M., Somervuo, P., Lehikoinen, P., Geres, L., Richter, T., Seibold, S. and O. Ovaskainen. 2022. Domain-specific neural networks improve automated bird sound recognition already with small amount of local data. *Methods in Ecology and Evolution,* 13: 2799–2810.

Lauha, P., Rannisto, M., Somervuo, P., Heikkinen, M., Hardwick, B., Kerdraon, D., Nokelainen, O., Andrejeff, S., Lehikoinen, P., Acevedo-Charry, O., Adalsteinsson, S. A., Agger, C., Alcobia, S., Aldará, J., Alvarez, M., Andersen, H. J. S., Andrew, N. R., Anslan, S., Antonelli, A., Aragón, R., Arena, J. S., Aster, I., Balestrieri, R., Banelyte, G. G., Barrett, A., Barrio, I. C., Beaudreault, S. R., Beja, P., Bergerot, B., Biro, E. G., Birot, H., Björk, R. G., Björkman, M. P., Blendinger, P.G., Bollache, L., Bou Dagher Kharrat, M., Boyer, S., Bridell, E., Brotherson, M., Brown, L. R., Buckley, H. L., Buscardo, E., Buthelezi, N., Cagnolo, L., Calvente, A., Capobianco, G., Carreón-Palau, L., Carrière, S., Casale, F., Case, B. S., Chang, J., Chaparro, J. M., Chen, C., Chicoine, C., Christensson, M., Chumo, S. K., Colom Montero, W., Colón-Piñeiro, Z., Costa, S. D. R., Cowan, L., Cross, J., Da Fonseca, R. D. S., Da Silva, W. C., De Francesco, M. C., Del Ángel-Rodríguez, J. A., Díaz-Delgado, R., Diaz-Pulido, A., Dirnböck, T., Djukic, I., Dladla, P., Domínguez Masciale, J., Dorigo, T. A., Douwes, E., Eckelt, A., Ekrem, T., El Chamy, L., Enderskog, H., Falahati Anbaran, M., Farrell, A., Fisher, B. L., Freiberga, I., Frosch-Radivo, A., Ganchozo Intriago, L. A., Gandoy, F., Garzoli, L., Gaviraghi Mussoi, J., Gerlich, S., Gianuca, A. T., Gilg, O., Gilg, V., Gili, F., Gísladóttir, F., Glowacki, R., Gordo, O., Gottsberger, B., Griem, L., Groner, E., Guedes, P., Haase, P., Haidegger, M., Hämäläinen, K., Hänninen, J., Hansen, L. H., Hansen, L. J., Harrop, R. A., Havnås, H., Ho, C. C., Höglund, L., Honnavar, P., Høye, T. T., Huszarik, M., í Haraldsstovu, K., Inkinen, J., Ishihara, M. I., Jackson, A. C., Jacot, A., Jansson, G., Jaroszewicz, B., Jeffery Brown, H., Jha, R., Kahale, R., Kalttopää, O., Kandlikar, G. S., Karai, E. W., Karlsson, D., Kaus-Thiel, A., Keurulainen, S. M., Kimoto, K., Klante, C., Klinth, M., König, S. T., Koskinen, J., Kotakorpi, M., Kreiling, A., Krisai-Greilhuber, I., Kuemmerle, T., Kupagme, J. Y., Lecomte, E., Lecomte, N., Leitman, P. M., Leppänen, E., Li, D., Lindgren, A., Lönnqvist, W., Lopez-Vaamonde, C., Mangini, G. G., Marchetto, A., Maroschek, M., Mata, V. A., Matsuoka, S., Mcclean, L. A., McDonald, P. G. M., Meléndez-González, M., Melville, H. I. A., Mendoza-Henao, A. M., Merkujeva, M., Molin, A. D., Moore, D., Moreau, S., Moyer, E., Mrazova, A., Mteshane, S., Müller, J., Mungai, N. W., Musazzi, S., Mutanen, M., Nagy, L., Ndlovu, M., Nicolai, A., Niessner, A., Nordén, J., Nurse, O., O'Connell, E., Ochoa, Y., Oja, J., Övergaard, R., Paavola, R. K., Pagnon, T., Pantoja, Y., Pastoriza, A., Paulaniemi, N., Pentinsaari, M., Phillips, A., Pinder, A., Pinheiro, T., Põlme, S., Powell, L. L., Průchová, A., Queralt, E., Quilantang, M., Quinlan, K., Raharinjanahary, D., Rajoelison, E. T., Randrianaina, L. A., Rankinen, J., Rasino, M. D. V., Rebelo, R., Reneerkens, J., Renne, V. M. A., Ridha, H., Rogers, H. M. K., Rosário, I. T., Ruohomäki, A., Rütting, T., Sabard, B., Saenz-Paez, V., Saliba, C., Sam, K., Santos, D., Santos-Reis, M., Schattanek-Wiesmair, B., Schiestl-Aalto, P., Schmidt, N. M., Schwandner, I. A., Seibold, S., Seidl, R., Seoane, J., da Silva, L. P., Singer, C., Sithole, M., Sivault, E., Smith, R. J., Soda, A., Sousa-Lima, R. S., Stanisci, A., Steenkamp, D., Stewart, J., Stewart, T., Stoll, S., Strauss, W. M., Sujala, M., Sundell, J., Svavarsdóttir, J., Tedersoo, L., Tuominen, E., Tweraser, S., Ulloa, C., Uusitalo, J., Valle Piñuela, C. A., Vallinmäki, M., Vannier, F., Varela, A., Vatka, E., Veikkolainen, S., Vernes, K., Vlasata, T., Watts, P. C., Weslien, P., Westin, J., Wilson, J., Wirth, C. F., Wisniewski, J. H., Young, A. B., Zwerts, J. A., Lehikoinen, A., Roslin, T. and Ovaskainen, O. 2025. Bird Sounds Global — model builder: An end-to-end workflow for building locally fine-tuned bird classifiers. *Authorea,* https://doi.org/10.22541/au.176599468.88578746/v1

Meramo, K., Somervuo, P., Rannisto, M., Lauha, P., López-Baucells, A., Gili, F., Roslin, T., Lilley, T. & O. Ovaskainen. 2025. BSG-BATS: an open-source classifier and data annotation portal for European bat vocalizations. *Methods in Ecology and Evolution, 00: 1–9,* https://doi.org/10.1111/2041-210x.70220

Ovaskainen, O., Winter, S., Tikhonov, G., Lauha, P., Lehtiö, A, Nokelainen, O, Abrego, N., Aroluoma, A., Harrison, J.P., Heikkinen, M., Kallio, A., Koliseva, A., Lehikoinen, A., Roslin, T., Somervuo, P., Souza, A. T., Tahir, J., Talaskivi, J., Turunen, A., Vancraeyenest, A., Zuquim, G., Autto, H., Hänninen, J., Inkinen, J., Kalttopää, O., Koskinen, J., Kotakorpi, M., Kuntze, K., Loehr, J., Mutanen, M., Oranen, M., Paavola, R., Renkonen, R., Schiestl-Aalto, P., Sipilä, M., Sujala, M. Sundell, J., Tepsa, S., Tuominen, E.-P., Uusitalo, J., Vallinmäki, M., Vatka, E., Veikkolainen, S., Watts, P.C. & D. A. Dunson. 2026. A digital twin for real-time biodiversity forecasting with citizen science. *Nature Ecology & Evolution,* in press. https://www.nature.com/articles/s41559-025-02966-3

Somervuo, P., Roslin, T., Fisher, B. L., Hardwick, B., Kerdraon, D., Raharinjanahary, D., Rajoelison, E. T., Lauha, P., Griem, L., Lehikoinen, P., Niittynen, P., Piirainen, E., Lumme, M., Riihikoski, V., Acevedo-Charry, O., Adalsteinsson, S. A., Ahmad, M., Alcobia, S., Aldará, J., Andrew, N. R., Anslan, S., Antonelli, A., Arena, J. S., Arroyo Almeida, S., Aster, I., Autto, H., Aviles Gamboa, A., Baixeras, J., Baldauf, M., Balestrieri, R., Banelyte, G. G., Barrett, A., Beja, P., Berg, T. O., Bergerot, B., Biro, E. G., Blendinger, P. G., Bollache, L., Bou Dagher Kharrat, M., Boyer, S., Bridell, E., Brotherson, M., Brown, L. R., Buckley, H. L., Buscardo, E., Buthelezi, N., Cagnolo, L., Calvente, A., Capobianco, G., Carreón-Palau, L., Carriere, S., Case, B. S., Chang, J., Chaparro, J. M., Chen, C., Chicoine, C., Christensson, M., Collado Rosique, F., Colom Montero, W., Da Fonseca, R. D., Da Silva, L. P., Dal Molin, A., Dallas, T., de Francesco, M., Del Ángel-Rodríguez, J. A., Díaz-Delgado, R., Dirnböck, T., Djukic, I., Dladla, P., Domínguez Masciale, J., Dorigo, T., Douwes, E., Ekrem, T., Enderskog, H., Erefur, C., Fahad, M., Falahati-Anbaran, M., Farrell, A., Ferland, G., Ferrari, E., Figueiredo, A., Forero, F., Freiberga, I., Frosch-Radivo, A., Ganchozo Intriago, L. A., Garzoli, L., Giacomotti, P., Gianuca, A. T., Gilg, O., Gilg, V., Gísladóttir, F., Glowacki, R., Gottsberger, B., Gregoire, J., Groner, E., Guedes, P., Guile, A. M., Haase, P., Hadi, F., Haidegger, M., Hansen, L., Hansen, L. H., Harrop, R., Havnås, H., Herrera Báez, D., Ho, C. C., Hohenbühel, D., Houska Tahadlova, M., Hänninen, J., Höglund, L., Í Haraldsstovu, K., Imbeau, E., Inkinen, J., Ishihara, M. I., Jackson, A. C., Jansson, G., Jha, R., Kager, G., Kahale, R., Kalttopää, O., Karai, E. W., Karlsson, D., Kaus-Thiel, A., Khan, A., Khan, Q., Kimoto, K., Kipngetich, S. C., Klante, C., Klemedtsson, L., Klinth, M., Koskinen, J., Kotakorpi, M., Kreiling, A., Krisai-Greilhuber, I., Kristensen, E., König, S., Langenheder, S., Laurila, K., Le Gouar, P., Lecomte, N., Lecomte, E., Leitman, P. M., León-Cortés, J. L., Li, D., Loehr, J., Lopez-Vaamonde, C., Makari, M., Mangini, G. G., Maroschek, M., Mata, V. A., Matsuoka, S., Mazzafera, T., McDonald, P. G., Meinert, L., Meléndez-González, M., Mendoza-Henao, A. M., Moreau, S., Moreau, J., Mosbacher, J., Moyer, E., Mrazova, A., Mteshane, S., Mungai, N. W., Muñoz Herraiz, G., Murillo-Vázquez, A., Musazzi, S., Mutanen, M., Müller, J., Navarro Canales, R., Ndlovu, M., Nicolai, A., Niessner, A., Nordén, J., Nowak, P., O'Connell, E., Orru, A., Pagnon, T., Pantoja-Diaz, Y. N., Pentinsaari, M., Pilloni, S., Pinder, A., Pinheiro, T. A., Põlme, S., Powell, L. L., Pröll, G., Pulido-Santacruz, P., Queralt, E., Quilantang, M. T., Quinlan, K., Ramirez, R., Rankinen, J., Rasino, M. D., Rebelo, R., Remmers, W., Retz, F., Reyes, E., Rivas Torres, G., Rogers, H. M., Rosário, I. T., Rosário Da Costa, S., Rütting, T., Sahlstén, J., Saliba, C., Salmirinne, T., Sam, K., Santos, D., Santos-Reis, M., Sawan, M., Schattanek-Wiesmair, B., Schiestl-Aalto, P., Schmidt, N. M., Seibold, S., Seidl, R., Seifert, L., Sithole, M., Sivault, E., Smart, J., Smerczyński, I., Soda, A., Sousa-Lima, R. S., Stanisci, A., Stanley, M. C., Steenkamp, D., Stengel, E., Stoll, S., Strauss, W. M., Stur, E., Sujala, M., Sundell, J., Svavarsdóttir, J., Tedersoo, L., Tepsa, S., Tikochinsky, M. T., Tuominen, E., Tweraser, S., Ulloa Espinosa, C., Uusitalo, J., Vallinmäki, M., Vannier, F., Varela, A., Vatka, E., Veikkolainen, S., Vernes, K., Watts, P. C., Weslien, P., Wirth, C., Wisniewski, J. H., Young, A. B., Övergaard, R. & O. Ovaskainen. 2025. Human contributions to global soundscapes are less predictable than the acoustic rhythms of wildlife. *Nature Ecology & Evolution,* 9: 1585–1598.

Sullivan, B. L. Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D. & S. Kelling. 2009. eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation,* 142: 2282–2292.

Xeno-canto Foundation. 2005. *Xeno-canto: Sharing wildlife sounds from around the world*. Available at https://xeno-canto.org/ [Last accessed 1 November 2023].

Received: 15th October 2025
Accepted: 11th December 2025

Bird Census News 2022, 38/1–2: European Monitoring News 56–57

EUROPEAN MONITORING NEWS

Introducing the EBCC board: Dawn Balmer

Danae Portolou

What is your title and the current working position?

Head of Surveys at the British Trust for Ornithology (BTO) in the United Kingdom

What is your current role in the EBCC?

I joined EBCC as an Observer on the Board in autumn 2018 and joined the Board formally at the AGM in 2019 and I have the role as Secretary.

What is your current role in the BTO and role in using monitoring data?

At BTO, I'm the Head of Surveys and oversee many of our long-term monitoring schemes and projects. I have worked for BTO for nearly 33 years and have organised several of our long-term projects in the past and was the Atlas Coordinator for Bird Atlas 2007–2011. I joined BTO on a short contract as an analyst on our territory mapping project called the Common Birds Census and have since been involved with many monitoring schemes!

Can you tell more about how you are using citizen science data in your work?

Volunteers and our monitoring work is at the very heart of BTO. We are fortunate to have long-term schemes such as the Common Birds Census, Breeding Bird Survey, Wetland Bird Survey, Heronries Census, Seabird Monitoring Programme which provide long-term data sets so we can present trends in abundance and produce indicators and use the data in research, policy and conservation action. We share our information with the public through BirdFacts (https://www.bto.org/learn/about-birds/birdfacts). We're also able to run periodic and short-term surveys related to specific species which are not well monitored through our long-term surveys or to answer questions about habitat use. BirdTrack is our popular web/phone app project for birdwatchers to record their sightings and receives around 10 million records annually.

Do you still manage to take part in fieldwork, and if so, in which monitoring programs do you personally participate?

I really enjoy volunteering and taking part in surveys. I have two Breeding Bird Survey squares in Thetford Forest and have been covering these since 1994 and 1997. I also monitor a small inland breeding colony of Black-headed Gulls *Chroicocephalus ridibundus*. This year I've been taking part in the Heathland Birds Survey and surveying for Woodlark *Lullula arborea*, Dartford Warbler *Sylvia undata* and Nightjar *Caprimulgus europaeus* and have just started fieldwork for the Winter Birds Survey — both periodic surveys that we organise. I'm a really keen birder and log my records into BirdTrack — so far I've logged 240 species, 535 submissions and 398 complete lists in the UK. I'm a fairly active bird ringer and particularly enjoy helping out with a local project on Nightjars in the forest. I'm very much looking forward to volunteering for the next bird atlas in 2027–2031!

Do you have a favourite bird or birding habitat/location?

My favourite bird is the Mediterranean Gull *Ichthyaetus melanocephalus* — I love all their plumages and enjoy trying to find them amongst large flocks of gulls. I have a local patch inland near home called East Wretham Heath where I try to visit a few times each week but I do love birding on the coast, and especially at migration times.

Your text in the next issue?

Bird Census News is meant as a forum for everybody involved in bird census, monitoring and atlas studies. Therefore we invite you to use it for publishing articles and short reviews on your own activities within this field such as (preliminary) results of a regional or national atlas or a monitoring scheme, species-specific inventories, reviews or activity news of your country (as a delegate: see also below).

Instructions to authors

– Text in MS-Word.
– Author name should be with full first name. Add address and email address.
– Add short abstract (max 100 words).
– Figures, pictures and tables should not be incorporated in the text but attached as separate files.
– Provide illustrations and figures both in colour.
– The length of the papers is not fixed but should preferably not exceed more than 15 pages A4 (including tables and figures), font size 12 pt, line spacing single (figures and tables included).
– Authors will receive proofs that must be corrected and returned as soon as possible.
– Authors will receive a pdf-file of their paper.
– References in the text: Aunins (2009), Barova (1990a, 2003), Gregory & Foppen (1999), Flade et al. (2006), (Chylarecki 2008), (Buckland, Anderson & Laake 2001).
– References in the list: Gregory, R.D. & Greenwood, J.J.D. 2008. Counting common birds. In: A Best Practice Guide for Wild Bird Monitoring Schemes (eds. P. Voříšek, A. Klvaňová, S. Wotton & R.D. Gregory), CSO/RSPB, Czech Republic; Herrando, S., Brotons, L., Estrada, J. & V, Pedrocchi, V. 2008. The Catalan Common bird survey (SOCC): a tool to estimate species population numbers. Revista Catalana d'Ornitología, 24: 138–146.

Send contributions in digital format by email to: dimitrije.radisic@dbe.uns.ac.rs