EUROPEAN PARTNERSHIP

The Automated Biodiversity Monitoring Stations (ABMS) Pilot draft report

A Transnational Sensor Network for Birds, Bats and Insects

Co-funded by
the European Union

Document Information

Grant Agreement number	101052342
Project acronym	Biodiversa+
Project full name	The European Biodiversity Partnership
Biodiversa+ duration	7 years
Biodiversa+ start date:	1 October 2021
For more information about Biodiversa+	Website: https://www.biodiversa.eu/ Email: contact@biodiversa.eu LinkedIn: Biodiversa+

Deliverable title	Automated Biodiversity Monitoring Stations (ABMS) Pilot draft report
Dissemination level	Public
Authors	Jamie Alison, Mark Gillespie, Jonáš Gaigr, Gerard Bota, Lluís Brotons, Luc De Bruyn, Stinna Danger, Domhnall Finch, Klara Grethen, Daniela Hamidovic, Guillaume Mougeot, Jarek Scanferla, Vladimír Nemček, Lars B. Pettersson, Jozef Šibík, Mária Šibíková, Mladen Zadravec, Rotem Zilber, Toke T. Høye
Work package title	WP2 Promote and support transnational biodiversity monitoring
Task or sub-task title	ST 2.6.5 Pilot Automated Biodiversity Monitoring Stations
Lead partner	SGAV
Date of publication	December 2025
Disclaimer	Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

What is Biodiversa+

The European Biodiversity Partnership, Biodiversa+, supports excellent research on biodiversity with an impact for policy and society. Connecting science, policy and practice for transformative change, Biodiversa+ is part of the European Biodiversity Strategy for 2030 that aims to put Europe's biodiversity on a path to recovery by 2030. Co-funded by the European Commission, Biodiversa+ gathers partners from research funding, programming and environmental policy actors in European and associated countries to work on 5 main objectives:

1. Plan and support research and innovation on biodiversity through a shared strategy, annual joint calls for research projects and capacity building activities
2. Set up a network of harmonised schemes to improve monitoring of biodiversity and ecosystem services across Europe
3. Contribute to high-end knowledge for deploying Nature-based Solutions and valuation of biodiversity in the private sector
4. Ensure efficient science-based support for policy-making and implementation in Europe
5. Strengthen the relevance and impact of pan-European research on biodiversity in a global context.

More information at: https://www.biodiversa.eu/

Table of contents

List of acronyms

ABMS	Automated Biodiversity Monitoring Stations
AI	Artificial Intelligence
AMI	Automated Monitoring of Insects (refers to insect cameras)
BON	Biodiversity Observation Network
BTO	British Trust for Ornithology
EBV	Essential Biodiversity Variable
ENV	Genus aggregate for bats (*Eptesicus, Nyctalus, Vespertilio*)
ERDA	Electronic Research Data Archive (at Aarhus University)
EUNIS	European Nature Information System
FAIR	Findable, Accessible, Interoperable, and Reusable (data principles)
GAM	Generalized Additive Model
GBIF	Global Biodiversity Information Facility
GEO BON	Group on Earth Observations Biodiversity Observation Network
ISO	International Organization for Standardization
LUCAS	Land Use/Cover Area frame Survey
SD	Secure Digital (referring to memory cards)
SFTP	Secure File Transfer Protocol
SSD	Solid State Drive
UKCEH	UK Centre for Ecology & Hydrology
YOLO	You Only Look Once (referring to the YOLOv8 insect segmentation model)

Executive Summary

Addressing biodiversity change means filling data gaps in space and time, building understanding of patterns, trends and drivers of species' spread and decline. Automated methods, including sensors and artificial intelligence (AI), promise to generate affordable and standardized data with extensive spatial, temporal and taxonomic coverage. However, their deployment has rarely been coordinated at transnational scales. The Automated Biodiversity Monitoring Stations (ABMS) pilot implements a transnational sensor network for bats, birds and night-flying insects, evaluating the potential of automation for scalable biodiversity monitoring across Europe and elsewhere.

Running throughout 2024 and 2025, the pilot deployed ~200 sensors at 70 sampling locations across 12 EU member states. Sensors comprised acoustic sensors for birds and bats, camera systems for night-flying insects, and loggers for temperature and moisture. Huge volumes of data were generated; 3 million audible acoustic files, adding up to 5.43 recording-years; 5 million ultrasonic acoustic files, adding up to 1.58 recording-years; over 200,000 time-lapse images, with millions more motion-triggered images, covering more than 8,000 recording-nights. Raw data were transferred to Aarhus University's Electronic Research Data Archive for centralized processing.

To enable centralized processing of images and sounds, we developed three novel open-source AI processing pipelines. We deployed BirdNet for bird detection and classification, BatDetect2 for bat detection and classification, flatbug for insect detection and several custom models for insect classification. These AI models generated over 1 million preliminary species records. However, recognizing the need to understand AI uncertainty, we also undertook an extensive verification programme. Experts from five countries checked 8,690 AI bird detections, while experts from seven countries checked 3,048 AI bat detections. Verifications authenticated records of 127 bird and 24 bat species, and helped to interpret AI outputs. For bats, they allowed us to generate bat activity metrics at appropriate taxonomic resolutions. For birds, they enabled confidence calibration curves, allowing summaries of AI-reliability for 86 European species.

Opportunities for analysis are vast, so we showcase the utility of the data to generate transnational Essential Biodiversity Variable (EBV) indicators. Ecosystem phenology indicators depict the annual rise and fall of bird, bat and insect activity across countries and habitats. Community abundance and taxonomic diversity indicators highlight dominant taxa, and estimate the species richness of each sampling location. Finally, we summarize the lessons learned during the pilot, aided by a partner questionnaire. We capture limitations of sensors, especially when technological readiness level is low, and the need for expert verification and local data processing options. Crucially, 100% of our partners said they could integrate ABMS approaches into their national monitoring; we conclude that sensor networks are a viable tool for transnational monitoring of birds, bats and insects, but that coordination and central data processing are fundamental to success.

1. Introduction

There is widespread concern over the state of biodiversity globally, supported by accumulating evidence of declines in the abundance and diversity of many taxonomic groups. However, patterns vary across time, space, and taxa, while there are significant gaps in the available data. Automated methods for biodiversity monitoring, such as insect camera trapping and passive acoustic monitoring of birds and bats, could help to fill data gaps and better understand trends in biodiversity. Specifically, automated methods may offer more extensive spatial, temporal and taxonomic coverage, with repeatable and standardised data to fill coverage gaps left by intensive traditional monitoring methods.

However, automated methods have rarely been rolled out on international scales - especially for cryptic and difficult to observe taxa such as insects, bats or nocturnal birds. For example, a variety of bat monitoring schemes deploying acoustics devices, but these usually operate at regional or national scales, and efforts are still under way to align protocols (López-Baucells *et al.*, 2025). In short, the feasibility and scalability of novel methods needs further evaluation. This Biodiversa+ pilot aimed to perform such an evaluation, implementing coordinated sampling across biogeographical regions to assess the potential of sensors and AI for European biodiversity monitoring. The taxonomic scope was nocturnal insects, birds and bats, including cryptic and nocturnal taxa, for which automated monitoring methods have been developed but need to be tested at scale. Specific objectives of the ABMS pilot were to:

1. Enable each active contributing partner to operate three multi-taxa sensor stations for at least three months during the 2024 and 2025 growing seasons
2. Develop and make publicly available user-friendly, online databases hosting audio and image data and associated metadata
3. Establish an operational pipeline for recognition of nocturnal insects from images and birds and bats from sound
4. Prepare tools to visualise the locations, identity and time of observation of insect, bird and bat species recognised from image and sound

In building centralized and scalable data infrastructure and analysis pipelines, the pilot attempts to measure the transnational potential of sensor systems. Similarly, coordinated testing and protocol refinement is a proof of principle for more joined-up transnational monitoring, including sharing of digital resources and expertise. The pilot produces recommendations for realistic yet ambitious sensor-based biodiversity monitoring across Europe.

2. Implementation

2.1. Equipment

Acoustic recorders were selected based on their battery life, sound quality, and previous user experiences. We specifically examined products from Wildlife Acoustics, including the Song Meter Mini 2 AA and Song Meter Mini Bat 2 AA, and compared them with the AudioMoth recorder and other devices (Mennill, 2024). Key points are captured below, while broad recommendations related to the equipment used in the ABMS pilot are captured in Table 1.

Song Meter devices:

- Quality: The frequency response is superior to cheaper alternatives.
- Battery life: Powered by 8 AA batteries, offering reasonable longevity for field use.
- Storage capacity: Support for up to a 2 TB SD card, so storage is not a limiting factor.
- Durability: Designed to be robust and well-suited for long-term installations.
- User friendly interface and outputs.

AudioMoth devices:

- Affordability: Priced lower than many competitors, but requires purchase of additional weather-resistant casing.
- Fragility: Based on previous experiences, it is more prone to damage or failure.
- Sensitivity: The AudioMoth is less able to capture faint sounds than Song Meter devices.

We also explored alternatives from Titley Scientific; however, their products were not accessible to all participating partners. In contrast, Wildlife Acoustics products are widely available and commonly used in research projects worldwide.

For automated monitoring of night flying insects, the primary option was Automated Monitoring of Insects (AMI) devices, which had begun to be batch produced by the UK-based companies Rototherm and UKCEH. These had been successfully purchased for the Invasive Alien Species pilot, though there had been delays and challenges - particularly related to a global shortage of Raspberry Pi devices. The only alternative produced and supplied internationally was the DIOPSIS device, which is not specialized for night-flying insects and did not have open-source data processing solutions.

Left: A Song Meter Mini 2 deployed in Denmark. Right: An AMI trap deployed in Belgium. Photo credit Jamie Alison (left) and Lucia Manzanares (right).

***Table 1.** Recommendations related to equipment used in the ABMS pilot.*

Device	**Topic**	**Recommendations**
	General	Song Meter devices were found to be accessible and reliable.
Song Meter Mini 2 & Song Meter Mini Bat 2	Placement	● Height: Post or tree (~1m), or tree (4-5m) if there is flood/wildlife/livestock disturbance risk ● Securing: Duct tape or self-locking plastic nylon ties, with a lock where additional security is needed
	Mic Direction	● With prevailing wind: For recording quality and repeatability ● Open spaces, pond, or streams: Maximize bat activity ● Into target habitat: Maximize occurrence of birds or bats
	For bats	● Sample frequency > 250 kHz ● Microphone capable of recording high-frequency sounds
	Batteries	● Disposable batteries are recommended for remote sites, for example ENERGIZER Ultimate Lithium AA FR6/4 1,5 V, 2950 mAh. These often last 2 months for audible and even ultrasound, unless bat activity is extremely high ● Rechargeable batteries can be used for very accessible sites, for example VARTA Recharge Accu Power AA 2100 mAh. These last 2-4 weeks for audible and ultrasound, depending on temperature and bat activity
	Storage	● 1TB means almost an entire season fits on one card, but 512GB is sufficient if emptied during the season. For example, SanDisk Micro SDXC Extreme Pro, read 200 MB/s, write max. 140 MB/s, U3, V30, A2 + SD adapter
AMI devices	General	AMI devices were difficult to obtain and sometimes faulty. If they have no internet connectivity upgrade, visit them once per month
	Placement	● A pedestal may be required in wet areas with flood risk (e.g. steel frame platform, wooden pallet)
	Batteries	● At least one rechargeable deep cycle battery, for example AGM or GEL Sealed Lead Acid batteries ● One 175-watt, or 2 100-watt, solar panel(s) with mounts, facing south, elevated to 20-40 degrees depending on latitude
	Storage	● A 500GB Samsung Portable SSD is large enough to hold an entire year of images, if the jpegs are compressed

2.2. Site selection

To showcase and evaluate novel technologies in diverse ecological contexts, each partner selected three sites representing each of three broad classes of habitats: forest, grassland and wetland. These were defined based on European Nature Information System habitat types (EUNIS; https://eunis.eea.europa.eu/habitats.jsp; Table 2). Forest sites were deciduous broadleaved forest (T1) unless this wasn't possible, in which case partners were advised to select a coniferous forest site (T3) instead. Partners were also instructed to prioritize sites inside Natura 2000 areas. Sites needed to contain large areas of the relevant habitat to establish up to two sampling locations separated by >200m (Fig. 1).

Accessibility and long-term viability was also an important consideration during site selection. The map of sites sampled at the end of 2025 is shown in Figure 2, while the distribution across biogeographical regions is shown in Table 3.

Table 2. *EUNIS habitat definitions used to define woodland, grassland and wetland habitats in the ABMS pilot. Names and codes are based on the EUNIS 2021 habitat types.*

Level	**Code**	**EUNIS name**	**ABMS habitat**
2	T1	Deciduous broadleaved forest	Forest
2	T3	Coniferous forest	Forest
1	R	Grasslands and lands dominated by forbs, mosses or lichens	Grassland
1	Q	Wetlands	Wetland

Top-left: Grassland site in Czechia. Bottom-left: Forest site in the Netherlands. Right: Forest site in the Province of Bolzano. Photo credit Jonas Gaigr (top-left), Rotem Zilber (bottom-left) and Jarek Scanferla (right).

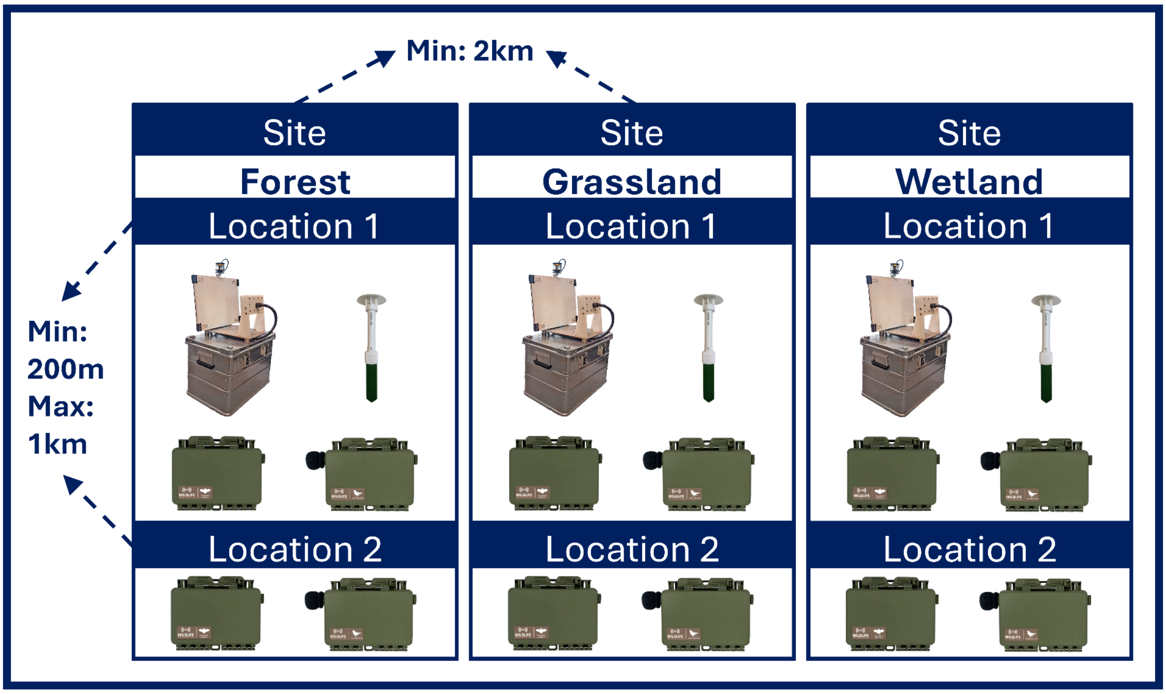


Figure 1. *Overall design of ABMS sampling with AMI devices, Song Meter devices and TMS-4 climate loggers. Up to two locations were sampled at sites associated with three broad habitat types: forest, grassland and wetland.*

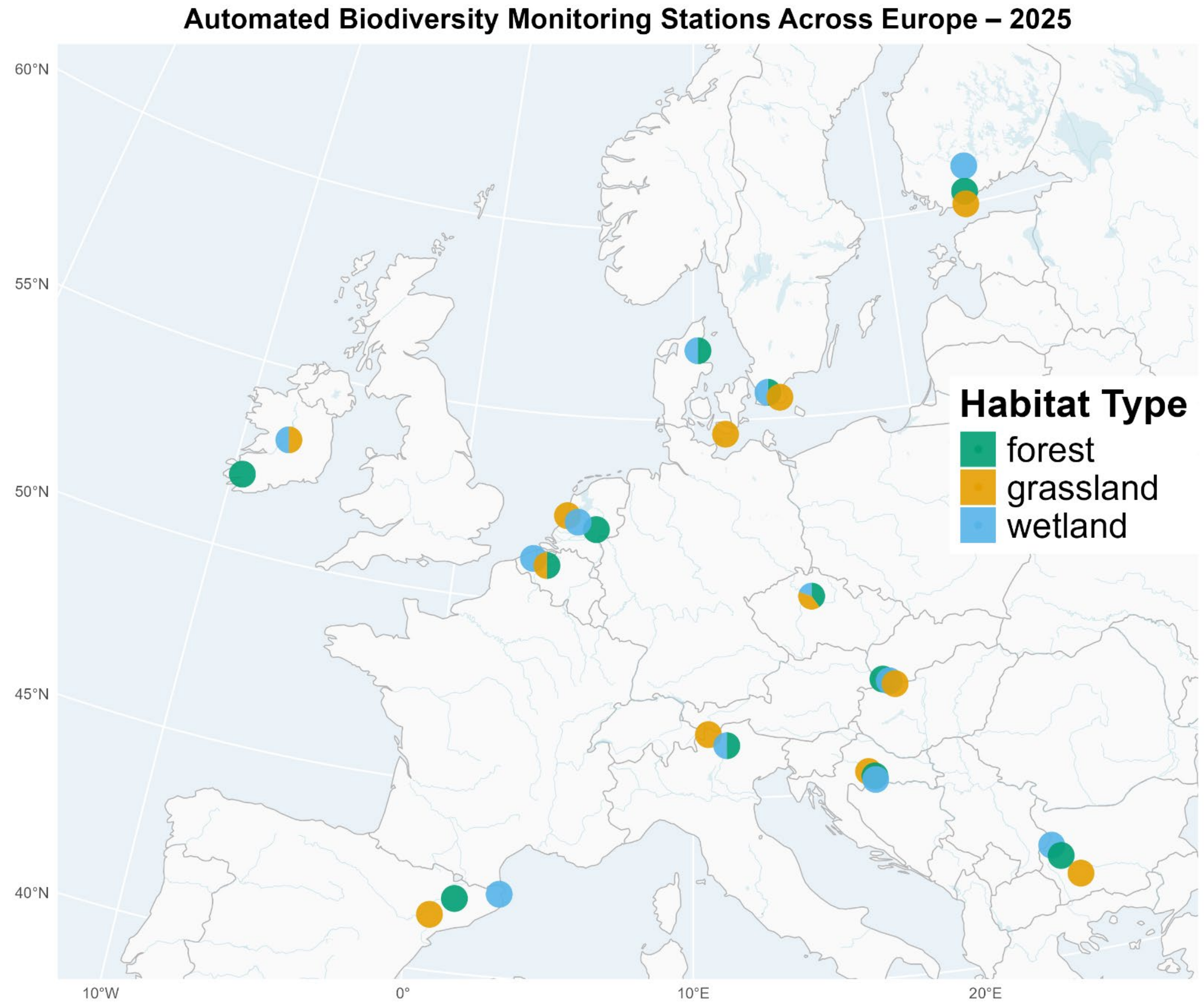


Figure 2. *Map of ABMS sampling sites distributed across Europe, selected to represent forest (green), grassland (amber) and wetland (blue) habitats. Up to two locations were monitored at each site. Where two or three sites are highly clustered, points are segmented based on the habitats of the underlying sampling locations - for example five locations in Czechia were very close to one another - two forest, two grassland and one wetland.*

Table 3. *Distribution of the ABMS sampling locations across biogeographical regions of Europe.*

Biogeographical Region	N locations	N grassland	N forest	N wetland
Continental	26	8	10	8
Atlantic	18	6	6	6
Alpine	8	4	2	2
Boreal	6	2	2	2
Mediterranean	6	2	2	2
Pannonian	6	2	2	2

2.3. Device deployment and maintenance

At each site, up to two sampling locations with similar characteristics, representing "replicates" for the site, were selected. One location would be sampled with an AMI trap, Song Meter Mini 2, Song Meter Mini Bat 2, and a TMS-4 logger, while the other would be sampled with only Song Meter devices. Locations were separated by at least 200m, following good practice guidelines for ecoacoustic monitoring (Metcalf *et al.*, 2022), ensuring the relevant habitat type was dominant within 100m of each location. To minimize placement bias, partners were asked to place locations at the closest intersection of the 100×100m EuroStat grid (EPSG:3035, ETRS89-extended / LAEA Europe). Furthermore, to enable a fair comparison of bat activity between locations with/without AMI traps, partners were asked to decide randomly which of the two locations would include the AMI trap and TMS-4 logger.

Pictured: Deploying an AMI trap in the field in Bulgaria. Photo credit Radoslav Stanchev.

Devices were placed in clusters at sampling locations (Fig. 3). The Song Meter Mini 2 and Song Meter Mini Bat 2 were placed on a post or tree (sometimes the same post or tree), 1m from the ground, facing the same direction, ideally with the prevailing wind. The exception was in Croatia, where Song Meter devices were placed higher in the canopy to prevent damage by animals.

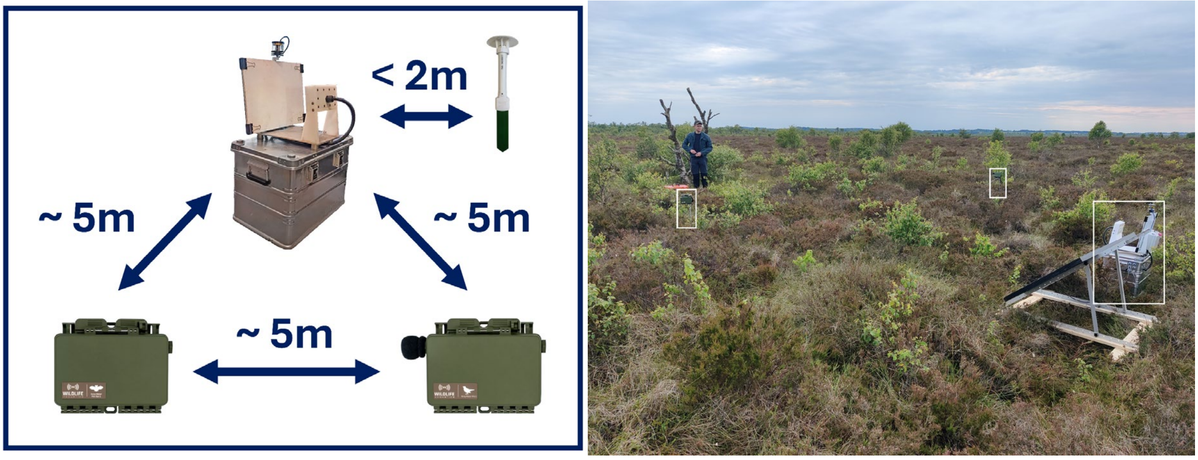


Figure 3. Device deployment in the ABMS. At sites with AMI devices, Song Meter devices were placed around 5m away, while TMS-4 loggers were placed <2m away. Right: Layout of Song Meter devices and an AMI trap (white boxes) at the wetland site in Denmark (photo credit Jamie Alison).

The pilot aimed to keep devices active every day from the beginning of May to the end of October in both 2024 and 2025. As such, we chose recording schedules and duty cycles that enabled acoustic devices to record for weeks, and AMI devices for months, without servicing battery or storage. Recording parameters for acoustic devices are presented in Table 4. AMI devices operated from 11pm to 3am local time, taking a “snapshot” image every 10 minutes (2 minutes in Sweden during 2025). Furthermore, an image was recorded every 2 seconds while motion was detected in the frame.

Table 4. *Key parameters used for acoustic monitoring in the ABMS. Minimum frequency to trigger ultrasound recording was lowered from 16kHz in 2024 to 10kHz in 2025 (9 kHz in Croatia), to increase potential to detect Tadarida spp. To further align with parameters used in other bat monitoring schemes (López-Baucells et al., 2025), Croatia adopted a max recording length of 5s. Audible monitoring parameters draw on advice from Metcalf et al. (2022).*

Parameter	**Song Meter Mini 2**	**Song Meter Mini Bat 2**
Schedule	24 hour	Dusk -1h to dawn +1h
Trigger	None	Above 10kHz (3 second trigger window)
Duty cycle	Recording 1 minute in 5	Always (triggered)
Max recording length	1 minute	15 seconds (min 3 seconds)
Sample rate	48kHz	256kHz (Full-spectrum)

2.4. Processing pipelines

The pilot aimed to design and deploy low-cost, replicable, open-source pipelines to extract ecological data from images and sounds. Importantly, these pipelines should be deployable without any further training. Silva del Pozo *et al.* (2023) highlighted several approaches to harmonized biodiversity monitoring, differing in whether raw data are processed at national or transnational scales. Our pipeline achieves centralized, transnational processing of raw data (see section 3. Data Management), providing the easiest comparison of results. However, considering the need to boost local knowledge, motivation and ownership, we only considered pipelines that partners could feasibly run free-of-charge with local computing resources. While paid or subscription-based processing services were not used for central processing in the pilot, they could minimize technical barriers to local data processing in future. We created a pipeline for acoustics data, with BirdNet (Kahl *et al.*, 2021) and Batdetect2 (Aodha *et al.*, 2022) models for audible and ultrasound respectively, and a pipeline for AMI image data, with custom insect detection models (Svenning *et al.*, 2025) and classifiers (Bjerge, Karstoft & Høye, 2024; Jain *et al.*, 2025).

2.4.1. Acoustics pipeline

Automated processing is crucial to make effective use of acoustic monitoring data. Many software tools are available to efficiently sort and label sounds, facilitating biodiversity analyses. While several indices exist to quantify biodiversity and ecosystem health - for example the Acoustic Complexity Index (ACI), the Acoustic Diversity Index (ADI), and the Bioacoustic Index (BI) - we focus on tools that produce species-level information. Similarly, while commercial software, such as Kaleidoscope Pro from Wildlife Acoustics, offers advanced features for professional labelling of bird songs, frog calls, and bat vocalizations, we focus on tools that are freely available and open-source, at least in terms of model weights.

Birds

Code related to the AI pipeline for birds can be accessed here.

Data management for acoustics was the same for audible and ultrasound data, but AI models and downstream processing differed (Figure 4). For audible acoustic data, BirdNet was the main available open-source solution to detect and classify bird species across Europe (Kahl *et al.*, 2021). Alternatives discussed during the pilot included the BTO Acoustic Pipeline, which requires payment for large scale processing, and AvesEcho, an extension of BirdNet utilizing a wider range of training data (Ghani *et al.*, 2025). We adopted BirdNet, as a published, tried and tested model that was easy to set up. Developed by the Cornell Lab of Ornithology and Chemnitz University of Technology, BirdNet uses machine learning to locate bird sounds in recordings and identify the species. Before analysis, you can adjust various settings, such as the minimum confidence level (affecting the sensitivity of the algorithm), or the candidate species list. The latter is particularly important to prevent the algorithm from searching for all bird species in training data at any given location (over 6000). BirdNET gives the option to generate a reasonable species list, based on eBird, from coordinates provided by the user.

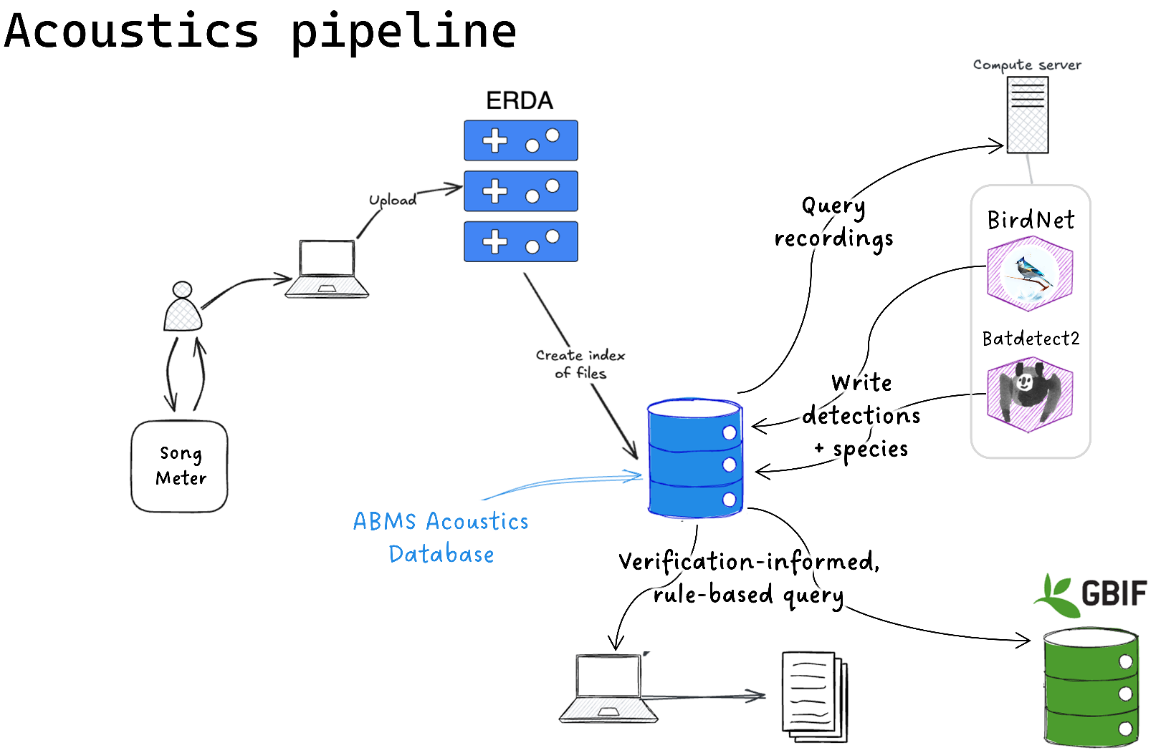


Figure 4. The acoustics data processing pipeline in the ABMS.

We ran BirdNet v2.2 using cloud computing resources at Aarhus University. Raw .wav files were passed to the BirdNet model, along with the location of the recording device, to classify based on a bespoke list of possible species for each site. We set a minimum confidence threshold of 0.1 for BirdNet, opting to produce a full list of predictions for later confidence-based subsetting. We used the default parameters for *Overlap* = 0 and *Sensitivity* = 1, although these could be varied in future iterations to improve performance (Pérez-Granados *et al.*, 2025). BirdNet provides text-based table outputs, where each row is a predicted bird detection within a 3-second snippet of the input .wav file. Each prediction has an associated confidence, which was used to subset data used for downstream analysis. Specifically, for 86 species receiving expert verification from two partner regions during our evaluation of BirdNet (see 2.5. Model evaluation), we used species-specific confidence thresholds to achieve a precision of 0.90, but enforced minimum and maximum thresholds of 0.5 and 0.9. For the remaining species we used a confidence threshold of 0.7, based on previous verifications of the BirdNet model (Wood & Kahl, 2024; Scanferla *et al.*, 2025; Pérez-Granados *et al.*, 2025).

Bats

The current AI pipeline for bats can be accessed here.

For ultrasound data, processing options without payment or further training were even more limited than for birds. BatDetect2 was the main option (Aodha *et al.*, 2022), as we excluded paid services such as Kaleidoscope or the BTO Acoustic Pipeline which are widely deployed in other bat monitoring schemes (López-Baucells *et al.*, 2025). We also excluded alternatives such as ANIMAL-SPOT (Bergler *et al.*, 2022), which does not come pretrained. We opted for BatDetect2, although the most relevant BatDetect2 model is trained to detect and classify just 17 species of bats occurring in the UK, while upwards of 30 bat species can be found in ABMS partner regions (Battersby, 2010). It uses a machine learning algorithm for joint detection and classification of bats, and provides confidence scores for detection (corresponding to the presence or absence of a bat call) and classification (corresponding to the species for that bat call).

We ran BatDetect2 with the python package, run from the command line (with bash) using cloud computing resources at Aarhus University. Raw .wav files were passed to BatDetect2, which output text based tables corresponding to detections of bats within each file. As with BirdNet, we set a minimum confidence threshold of 0.1, opting to produce a full list of bat call predictions for later confidence-based subsetting. BatDetect2 output tables comprise one row per detected call, of which there were often hundreds for each file, each associated with a bounding box (lower and upper frequency limits, as well as start and end times), a detection confidence and a classification confidence score. These outputs were not immediately biologically meaningful, with many calls from many species in each file.

A rule-base was used to generate file-level detections and classifications from hundreds of calls, lasting for fractions of a second. This rule-base was informed by manual verifications across 7 countries (see 2.5. Model evaluation). This involved iterating over 3,048 verified files, varying the threshold detection and classification confidence, as well as the minimum number of calls above that confidence. Detection was optimized by assuming bats were present where there were 2 or more detected calls above a threshold detection confidence of 0.495. For classification, we accepted the species identity with the highest summed class confidence in calls above the detection threshold, and considered the file to contain multiple species if there were at least 6 calls from each of two species above a threshold classification confidence score of 0.65.

2.4.2. AMI pipeline

Code related to the AI pipeline for insects can be found here, here and here.

Processing of AMI data built on the pipeline developed for AMI devices in the Invasive Alien Species pilot (Høye *et al.*, 2024). First, images were passed through a detection algorithm known as flatbug (Svenning *et al.*, 2025). Flatbug is a YOLOv8 instance segmentation model, pre-trained on images from several other automatic insect localization projects, including >150 AMI trap images. Instance segmentation presents a major advantage compared to bounding boxes, in that when two or more insects are close to each other, bounding boxes can contain several insects, while polygons extract individual insects more precisely. This facilitates the classification process, which is carried out in two stages: Order-level classification of insects and spiders, and species-level classification of Lepidoptera (moths). The order-level classifier is trained on annotations of a wide variety of arthropods seen in AMI trap images from Denmark, and classifies Araneae, Coleoptera, Diptera, Ephemeroptera, Hemiptera, Hymenoptera, Lepidoptera, Neuroptera, Opiliones and Trichoptera (Bjerge *et al.*, 2024). Lepidoptera are further classified as moths that belong to Macroheterocera (i.e. macro-moths) and those that do not (micro-moths). The species-level classifier was trained on a dataset derived from GBIF, covering 2,530 species distributed across the UK and Denmark (Jain *et al.*, 2025). Classifiers trained on larger subsets of European Lepidoptera were trained as a part of the ABMS, but these models were not sufficiently refined to use for processing for this report (see 2.5. Insect classifiers).

2.5. Insect classifiers).

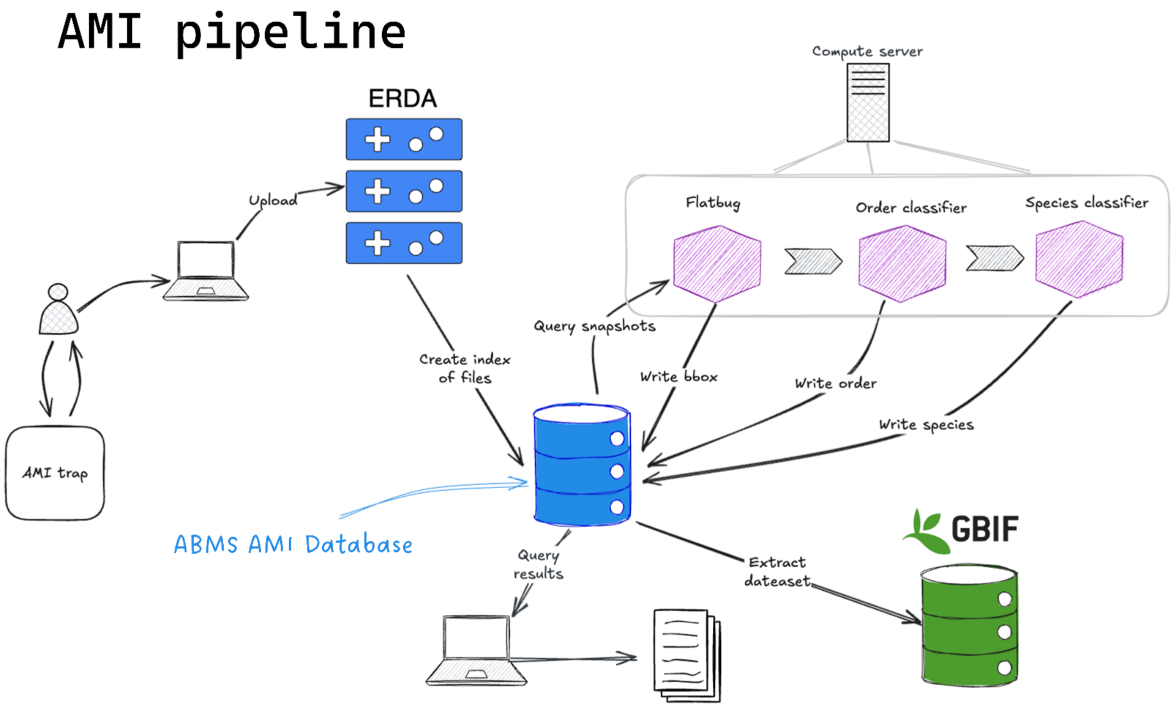


__Figure 4.__ The AMI data processing pipeline in the ABMS pilot.

2.4.3. Model evaluation

When deploying AI models in the pilot, it was considered vital to verify a subset of AI outputs, and quantify model performance in the context of the pilot. For birds and bats, we sent a strategic subset of data from 2024 for expert verification. For moths, we used a dataset from AMI traps in Denmark to evaluate the species classifier. We also developed an interactive tool to visualize and label insects in images (see 5.2. Data visualization and smart annotation with AnnFlux).

BirdNet

When verifying outputs of BirdNet, we aimed to quantify uncertainty, but also build confidence calibration curves (Wood & Kahl, 2024) to improve interpretation of BirdNet outputs for Europe-wide monitoring. Six partners were able to make experts available to verify AI predictions from 2024, of which five were available for this report. We selected species for verification based on the presence of at least three detections in each of nine confidence bands, ranging from 0.1-0.2, to 0.9-1.0. We then took a random sample of 27 3-second recordings of those species: three recordings from each of the nine confidence bands (note, for many species, verifiers expressed that they would have preferred significantly longer recordings to be sure about their IDs; see 5.2. Working with sensor data). The number of species sent for verification ranged from 29 for Ireland, to 133 for Denmark, and depended on the amount of recordings taken in 2024 as well as local diversity.

The verification process was made as efficient as possible, with a focus on identification of false positive detections. Given a single page of guidance on bird verification, the verifier simply answered the question: *Can the AI-suggested species be identified in the recording?* A yes or no answer was provided by entering a “1” or “0” in a provided table, as well as the initials of the verifier and some notes where applicable. Not all species sent to each partner were able to be verified in the time available, and partners prioritized species that were most common or most interesting for local monitoring needs.

Experts from Belgium, the Province of Bolzano, Denmark, Ireland and Slovakia verified 8,690 files with bird species predictions, of which 4,909 were deemed correct. Of 142 species proposed by BirdNet and verified by at least one partner, 127 were considered correct in at least one recording. Only six species were fully verified by all five partners (of which only four were confirmed in all regions): *Ardea cinerea, Athene noctua, Erithacus rubecula, Strix aluco, Troglodytes troglodytes* and *Turdus iliacus*. Given 27 verifications per species per country, it was possible to produce provisional confidence calibration curves for each species in each country (e.g. Fig. 5). This process hinted at distinct curves between countries, which could relate to variation in calls and call types, background noise, confusion species, and observer effects.

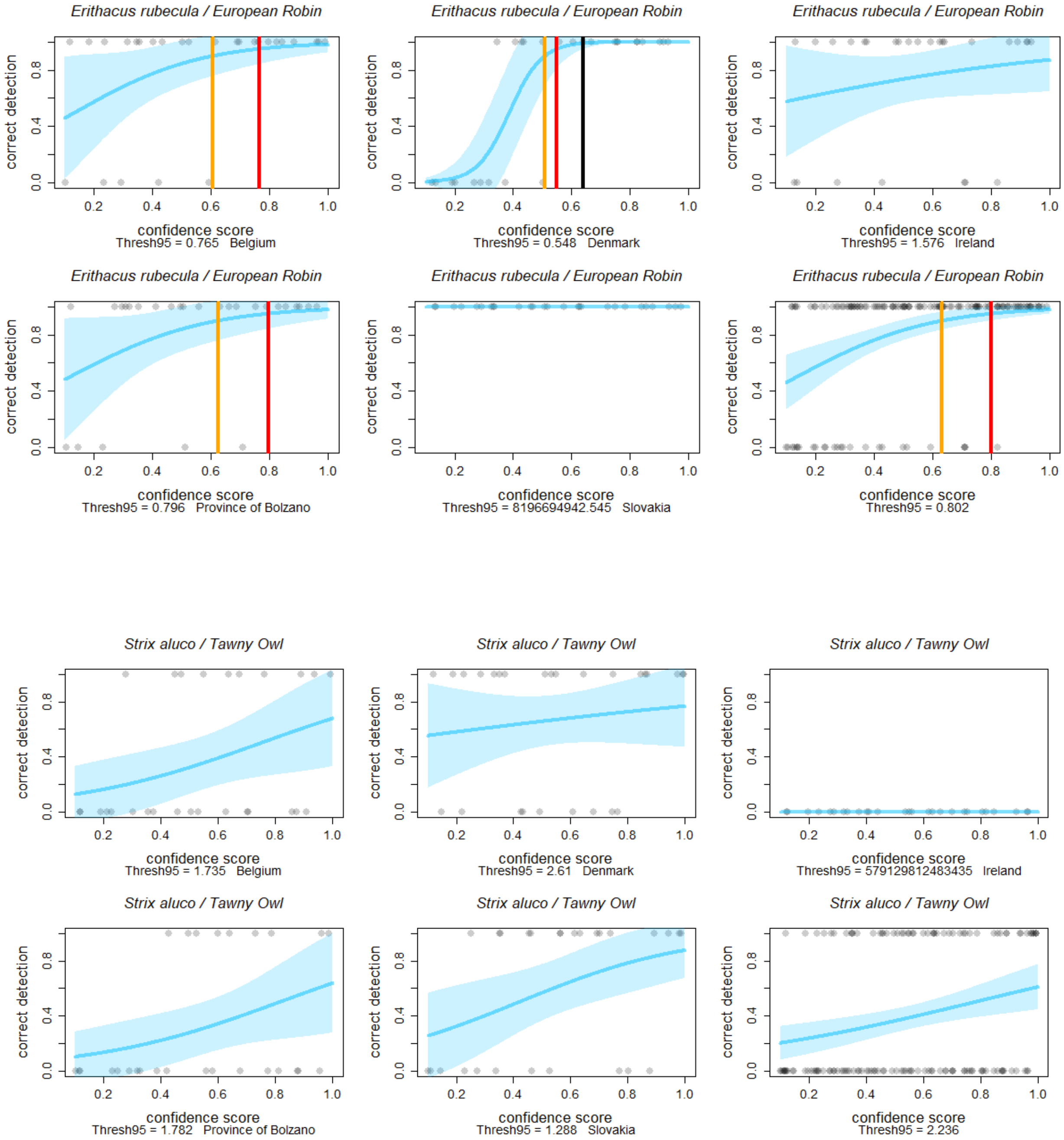


Figure 5. *Confidence calibration curves for Erithacus rubecula (top six) and Strix aluco (bottom six) across Belgium, Denmark, Ireland, the Province of Bolzano, Slovakia, and all five countries combined. The graphs show precision, or the proportion of BirdNet detections that are correct, across confidence scores from 0.1-1. For Erithacus rubecula in Slovakia, all calls were confirmed, regardless of confidence, while Strix aluco was never confirmed in Ireland. For Erithacus rubecula, it is possible to set confidence thresholds to achieve precision of 0.90 (yellow line) and 0.95 (red line) across the five countries, while in Denmark it was even possible to set a threshold for 0.99 precision (black line).*

A total of 86 species had at least 54 verifications across two or more partner regions for a more reliable confidence calibration. This allowed the assignment of species to traffic light groupings (Table 5) based on precision, which represents the proportion of classifications of a species that are correct. For example, precision didn't reach 0.9 at any confidence in any region for *Strix aluco* (Fig. 5, bottom). However, we can achieve 50% precision with a threshold of 0.778, thus *Strix aluco* is considered an amber species by the ABMS. Clearly detections of this species should be interpreted with caution. On the other hand, *Erithacus rubecula* is a green species, as 90% precision can be ensured by setting a confidence threshold of 0.633 (Fig. 5, top). Finally, *Athene noctua* is an example of a red species, as even 50% precision cannot be achieved; this species was commonly mistaken for dogs, cats and foxes according to the verifier from Slovakia.

Transnational confidence curves were further explored and used to improve data processing. Specifically, for 86 species with verification data across two or more regions (Table 5) we applied a custom confidence threshold to achieve a precision of 90%, but enforced minimum and maximum thresholds of 0.5 and 0.9. Recent work suggests that confidence thresholds above 0.5 can remove false-positives, but increase false-negatives (Pérez-Granados *et al.*, 2025). While we favoured precision, we used the 90% over the 95% precision threshold, because the latter would lead to a disproportionate increase in false-negative bird detections.

Table 5. *Results of confidence calibration for species identification with BirdNet. Confidence calibration used 8,690 expert verifications from five regions: Belgium, Province of Bolzano, Denmark, Ireland and Slovakia. Species shown here had at least 54 3-second snippets verified across a minimum of two partner regions. Threshold confidence scores to achieve 50%, 90% and 95% precision are shown, and species are assigned to traffic light (green, amber and red) groupings based on this. Coloeus monedula is referred to as Corvus monedula, in keeping with BirdNet naming.*

Species	**N snippets**	**N regions**	**0.50 prec. threshold**	**0.90 prec. threshold**	**0.95 prec. threshold**
Merops apiaster	54	2	0.000	0.000	0.000
Certhia familiaris	108	4	0.000	0.208	0.000
Sitta europaea	108	4	0.097	0.126	0.135
Phylloscopus collybita	108	4	0.057	0.214	0.267
Alauda arvensis	54	2	0.107	0.292	0.355
Cyanistes caeruleus	117	4	0.013	0.363	0.482
Cuculus canorus	81	3	0.029	0.412	0.542
Corvus monedula	81	3	0.189	0.464	0.558
Dendrocopos major	108	4	0.159	0.467	0.572
Corvus corax	108	4	0.135	0.477	0.593
Prunella modularis	54	2	0.023	0.449	0.593
Linaria cannabina	54	2	0.000	0.420	0.623
Turdus merula	108	4	0.000	0.451	0.633
Fringilla coelebs	108	4	0.000	0.437	0.674
Troglodytes troglodytes	135	5	0.000	0.432	0.702
Phasianus colchicus	81	3	0.000	0.518	0.718
Garrulus glandarius	108	4	0.000	0.530	0.725
Cettia cetti	54	2	0.313	0.636	0.746
Gallinago gallinago	81	3	0.115	0.589	0.750

Parus major	96	4	0.000	0.498	0.785
Erithacus rubecula	135	5	0.135	0.633	0.802
Motacilla alba	81	3	0.352	0.688	0.803
Pica pica	81	3	0.123	0.640	0.815
Certhia brachydactyla	108	4	0.288	0.686	0.822
Regulus regulus	108	4	0.142	0.660	0.837
Dryocopus martius	81	3	0.234	0.703	0.862
Phylloscopus trochilus	54	2	0.295	0.719	0.863
Curruca communis	54	2	0.202	0.727	0.905
Columba palumbus	81	3	0.071	0.701	0.915
Sylvia atricapilla	108	4	0.000	0.608	0.916
Apus apus	54	2	0.603	0.855	0.941
Emberiza schoeniclus	81	3	0.165	0.771	0.977
Carduelis carduelis	108	4	0.000	0.709	0.987
Anthus pratensis	108	4	0.000	0.543	NA
Delichon urbicum	81	3	0.000	0.639	NA
Hirundo rustica	108	4	0.000	0.723	NA
Turdus viscivorus	54	2	0.000	0.747	NA
Aegithalos caudatus	108	4	0.055	0.789	NA
Anser anser	81	3	0.330	0.835	NA
Poecile palustris	81	3	0.000	0.846	NA
Spinus spinus	54	2	0.448	0.869	NA
Turdus philomelos	81	3	0.271	0.965	NA
Pernis apivorus	54	2	0.293	0.980	NA
Tringa nebularia	54	2	0.497	0.987	NA
Picus viridis	108	4	0.155	NA	NA
Buteo buteo	108	4	0.242	NA	NA
Muscicapa striata	108	4	0.257	NA	NA
Alcedo atthis	54	2	0.318	NA	NA
Anas platyrhynchos	108	4	0.475	NA	NA
Corvus frugilegus	108	4	0.489	NA	NA
Ficedula hypoleuca	81	3	0.495	NA	NA
Nucifraga caryocatactes	54	2	0.548	NA	NA
Rallus aquaticus	108	4	0.607	NA	NA
Dendrocoptes medius	54	2	0.620	NA	NA
Fringilla montifringilla	54	2	0.648	NA	NA
Anas crecca	54	2	0.678	NA	NA
Ardea cinerea	135	5	0.705	NA	NA
Tringa ochropus	81	3	0.707	NA	NA
Anthus trivialis	108	4	0.769	NA	NA
Strix aluco	135	5	0.778	NA	NA
Dryobates minor	81	3	0.783	NA	NA
Fulica atra	54	2	0.846	NA	NA
Motacilla cinerea	81	3	0.847	NA	NA
Columba oenas	54	2	0.872	NA	NA
Numenius arquata	54	2	0.879	NA	NA

Mareca strepera	54	2	0.922	NA	NA
Phoenicurus phoenicurus	81	3	NA	NA	NA
Botaurus stellaris	81	3	NA	NA	NA
Caprimulgus europaeus	54	2	NA	NA	NA
Asio otus	54	2	NA	NA	NA
Bubo bubo	54	2	NA	NA	NA
Tyto alba	54	2	NA	NA	NA
Athene noctua	135	5	NA	NA	NA
Coturnix coturnix	54	2	NA	NA	NA
Scolopax rusticola	108	4	NA	NA	NA
Ixobrychus minutus	108	4	NA	NA	NA
Coccothraustes coccothraustes	108	4	NA	NA	NA
Calidris alpina	54	2	NA	NA	NA
Pyrrhula pyrrhula	54	2	NA	NA	NA
Turdus iliacus	135	5	NA	NA	NA
Phoenicurus ochruros	54	2	NA	NA	NA
Nycticorax nycticorax	108	4	NA	NA	NA
Cygnus cygnus	54	2	NA	NA	NA
Grus grus	81	3	NA	NA	NA
Locustella naevia	54	2	NA	NA	NA
Ficedula parva	54	2	NA	NA	NA

BatDetect2

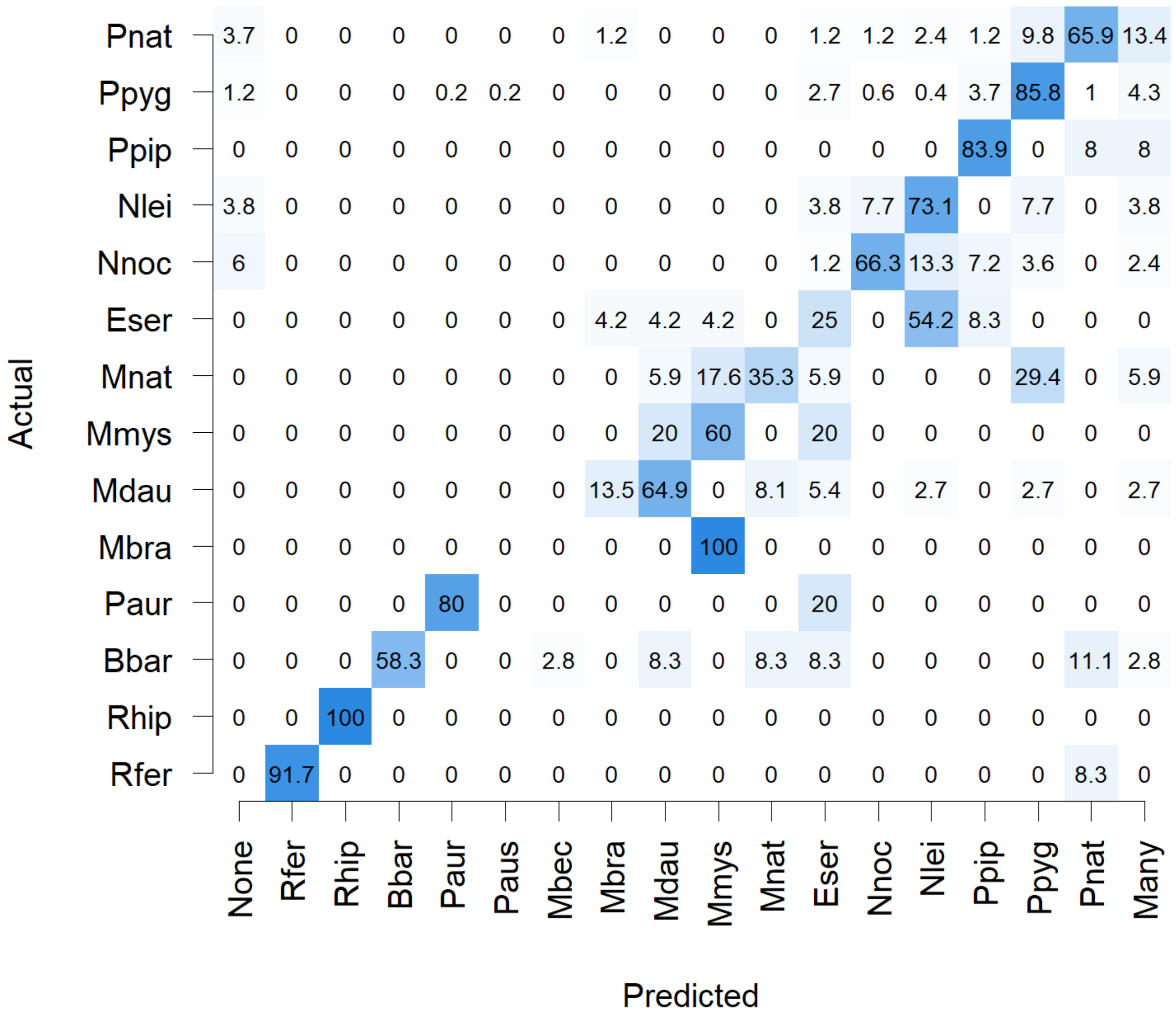
When verifying outputs of BatDetect2, we aimed to quantify uncertainty, but also to refine a rule-base to translate BatDetect2 outputs, with hundreds of detected calls, to species- or genus-level occurrences at the file-level. We also aimed to determine the taxonomic level at which it is sensible to interpret AI outputs. Experts from Belgium, Bulgaria, Denmark, Finland, Ireland, Slovakia and Sweden were able to make experts available to verify AI predictions from 2024. As only 17 species can be predicted by BatDetect2, we sent examples of all species to each expert to be verified. For each species, files were selected by randomly selecting up to four predicted calls in each of nine confidence bands, ranging from 0.1-0.2, to 0.9-1.0. The file corresponding to each call was sent for verification. Given there were very few calls in the highest confidence band of 0.9-1.0, this generally resulted in a maximum of 32 files, of a maximum of 15 seconds, per species.

For bat verification, we focused heavily on detection as well as classification, aiming to capture not only false-positive detections, but also quantify misclassification. Given a single page of guidance on bat verification, the verifier answered three questions: *Does the window contain a call from the suggested species? If not, what would you identify the call in the window as? Does the file contain multiple bats (of any species)?* The answers to the first two questions were used to evaluate BatDetect2. Verifications ultimately corresponded to the entire file - not just the small time window in which the call was detected. Verifiers provided bat identifications (or absences) for all 3,048 files according to their own taxonomies, which were aligned and unified in order to ensure interoperability of the seven datasets (Tables 6 & 7).

Table 6. *Bat verifications across 3,048 files, focussing on species-level IDs. While 1762 files had no species ID, most of these contained bats that were not identified to species resolution (see Table 7). Verifiers confirmed 24 species in total, including one individual of Vespertilio murinus in Sweden, included here under "Many bat species" (it occurred in the same file as another Pipistrellus pygmaeus). Species codes, corresponding grouping under a Genus/ENV aggregate, and number of verified files are also shown. Seven species not featured in the BatDetect2 training set are in bold.*

Code	Species name	Genus/ENV aggregate	Verified
Many	Many bat species	-	208
Pkuh	***Pipistrellus kuhlii***	***Pipistrellus***	**1**
Pnat	*Pipistrellus nathusii*	*Pipistrellus*	82
Ppip	*Pipistrellus pipistrellus*	*Pipistrellus*	174
Ppyg	*Pipistrellus pygmaeus*	*Pipistrellus*	513
Nlas	***Nyctalus lasiopterus***	**ENV**	**1**
Nlei	*Nyctalus leisleri*	ENV	26
Nnoc	*Nyctalus noctula*	ENV	83
Enil	***Eptesicus nilssonii***	**ENV**	**34**
Eser	*Eptesicus serotinus*	ENV	24
Hsav	***Hypsugo savii***	***Hypsugo***	**15**
Malc	*Myotis alcathoe*	*Myotis*	0
Mbec	*Myotis bechsteinii*	*Myotis*	0
Mbra	*Myotis brandtii*	*Myotis*	1
Mdas	***Myotis dasycneme***	***Myotis***	**2**
Mdau	*Myotis daubentonii*	*Myotis*	37
Mema	***Myotis emarginatus***	***Myotis***	**7**
Mmys	*Myotis mystacinus*	*Myotis*	5
Mnat	*Myotis nattereri*	*Myotis*	17
Paur	*Plecotus auritus*	*Plecotus*	5
Paus	*Plecotus austriacus*	*Plecotus*	1
Bbar	*Barbastella barbastellus*	*Barbastella*	36
Tten	***Tadarida teniotis***	***Tadarida***	**1**
Rfer	*Rhinolophus ferrumequinum*	*Rhinolophus*	12
Rhip	*Rhinolophus hipposideros*	*Rhinolophus*	2
None	No species ID	-	1762

The rule-base to convert call detections into file-level verifications was determined based on the verification data (See 2.4.1. Acoustics pipeline). With a detection rule of two or more calls above 0.495 detection confidence, we achieved a detection macro-F1 score of 0.924. Of files confirmed to contain bats, 97.4% were also predicted to contain bats. However, there were still a number of false positives, as 13.9% of files confirmed to contain no bats were still predicted to contain bats. For files where the verifier confirmed one of the 17 BatDetect2 species, classification performance was reasonable for e.g. *Pipistrellus* but poor for e.g. *Myotis* (Fig. 6). Poor performance for *Myotis* was expected; experts often cannot separate species in this genus, for example *M. brandtii* and *M. mystacinus*. Of 61 files with bats identified as species not in the BatDetect2 training data, 59 were successfully detected, but classified as other bat species (Fig A1).



	None	Rfer	Rhip	Bbar	Paur	Paus	Mbec	Mbra	Mdau	Mmys	Mnat	Eser	Nnoc	Nlei	Ppip	Ppyg	Pnat	Many
Pnat	3.7	0	0	0	0	0	0	1.2	0	0	0	1.2	1.2	2.4	1.2	9.8	65.9	13.4
Ppyg	1.2	0	0	0	0.2	0.2	0	0	0	0	0	2.7	0.6	0.4	3.7	85.8	1	4.3
Ppip	0	0	0	0	0	0	0	0	0	0	0	0	0	0	83.9	0	8	8
Nlei	3.8	0	0	0	0	0	0	0	0	0	0	3.8	7.7	73.1	0	7.7	0	3.8
Nnoc	6	0	0	0	0	0	0	0	0	0	0	1.2	66.3	13.3	7.2	3.6	0	2.4
Eser	0	0	0	0	0	0	0	4.2	4.2	4.2	0	25	0	54.2	8.3	0	0	0
Mnat	0	0	0	0	0	0	0	0	5.9	17.6	35.3	5.9	0	0	0	29.4	0	5.9
Mmys	0	0	0	0	0	0	0	0	20	60	0	20	0	0	0	0	0	0
Mdau	0	0	0	0	0	0	0	13.5	64.9	0	8.1	5.4	0	2.7	0	2.7	0	2.7
Mbra	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0
Paur	0	0	0	0	80	0	0	0	0	0	0	20	0	0	0	0	0	0
Bbar	0	0	0	58.3	0	0	2.8	0	8.3	0	8.3	8.3	0	0	0	0	11.1	2.8
Rhip	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rfer	0	91.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8.3	0

Figure 6. *Species-level confusion matrix for BatDetect2, for files where verifiers provided a single species-level ID (see Table 6 for numbers of files per species). Paus and Mbec were not recorded in any files, so the diagónal of the confusion matrix is shifted. A large proportion of errors occur within Myotis, and within the ENV Genus aggregate. “Many” indicates where several bat species were detected in a given file.*

Given the mixed performance of BatDetect2 at species-level, as well as the fact that some species cannot be reliably distinguished by experts, we opted to interpret BatDetect2 outputs at Genus level. However, we further lumped easily-mistaken genera *Eptesicus, Nyctalus* and *Vespertilio* into the ENV aggregate, giving a class that could be classified with an F1 of 81.3% (Fig. 7). The overall performance of our BatDetect2 classification pipeline at the level of genus/ENV aggregate was 82% (Macro-F1; Fig. 7). While 100% of *Rhinolophus* and 96.8% of *Pipistrellus* were identified correctly by the pipeline, this was only true for 58.3% of *Barbastella* and 65.2% of *Plecotus*. During further analysis, to retain detail within *Pipistrellus*, which was both abundant and relatively well classified by BatDetect2 (Fig. 6), we treated *Pipistrellus pipistrellus/pygmaeus* and *P. nathusii* separately.

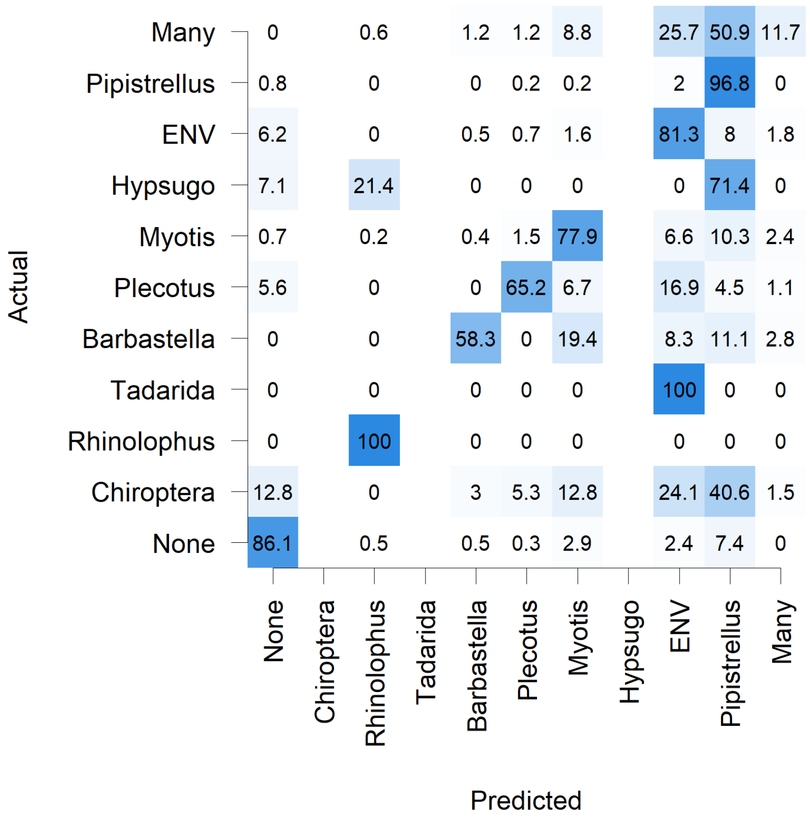


Figure 7. Genus/ENV aggregate-level confusion matrix for BatDetect2 across all verified files. While Hypsugo, Tadarida and Chiroptera cannot be identified correctly by BatDetect2, they were at least detected by the model.

Table 7. *Number of Genus/ENV level bat verifications across 3048 files*

Genus/ENV aggregate	**Verified records**
Many genera	171
Pipistrellus	1063
ENV	439
Hypsugo	14
Myotis	457
Plecotus	89
Barbastella	36
Tadarida	1
Rhinolophus	14
Chiroptera (bat)	133
None (no bat)	631

Insect classifiers

We used a published classifier trained to identify 2,530 species occurring in the UK and Denmark for this first iteration of ABMS processing (UK Denmark Classifier; Jain *et al.*, 2025). This model achieved an accuracy of 86.45% on a test set of GBIF data. However, when tested on a dataset that more fairly represented the ABMS use case, with images of 244 species at AMI traps in the UK and Denmark, accuracy fell to 79.39% (with a per-class accuracy of 72.84%). This performance is probably still inflated, as the evaluation only included examples where a verifier could confidently produce a species ID; in reality, for most moths, experts could not provide a species ID (Jain *et al.*, 2025). While the scope of the UK Denmark classifier is limited, we are not aware of any existing open-source moth species classifiers for the whole of Europe. The species coverage of the UK Denmark classifier, as well as our “in progress” European classifier, is shown in Table 8 for 13 ABMS partner countries.

Table 8. *Moth species coverage per ABMS country for each model. Species coverage is calculated as the proportion of a country's listed moth species (including micromoths) that are included in the model training data. The European classifier has improved species coverage across all countries, but did not yet achieve adequate performance for deployment in the ABMS. The moth species list per country is maintained by Jurrien van Deijk at the De Vlinderstichting ("The Butterfly Foundation") in the Netherlands.*

Country	**UK Denmark classifier (deployed in ABMS)**	**European classifier (in progress)**
Belgium	85.81%	**95.73%**
Bulgaria	53.07%	**83.95%**
Czech Republic	72.37%	**93.25%**
Germany	69.87%	**91.96%**
Denmark	90.04%	**94.70%**
Spain	46.13%	**87.51%**
Finland	80.70%	**91.03%**
Croatia	60.02%	**90.27%**
Ireland	92.75%	**96.37%**
Italy	48.71%	**86.74%**
Netherlands	88.66%	**95.88%**
Sweden	83.25%	**91.88%**
Slovakia	67.02%	**91.49%**

During this project, we have trained European classifiers to better reflect the species pool of countries such as Croatia, which have a richer moth fauna. To do so, we downloaded all Lepidoptera images from GBIF flagged with European country codes. After some quality filtering, removing non-adults images, background images and blurry images, we trained models on this dataset. The resulting data had a high number of species and an extremely long-tailed distribution: Of 7121 species in total, 351 had 500 or more images and 2030 had 5 or less images. As such, the European classifier has not yet achieved adequate performance for deployment in the ABMS.

Models perform well on GBIF data, but we have found results that can differ in the context of AMI traps. As such, we carried out further evaluation of UK Denmark and European classifiers using perhaps the largest expert-annotated AMI dataset in Europe. It includes 58,640 images with 522 moth species detected across a network of AMI traps in Denmark. Performances of both the UK Denmark model and the European and Global classifiers are represented in Table 9. Among the two currently integrated models, the UK Denmark model was the most performant one on AMI data. Furthermore, the model performs particularly well for the most common species in the Danish AMI dataset (Table 10). Furthermore, some of the species in the Danish AMI dataset were also highly abundant in forest, grassland or wetland habitats in the ABMS pilot. Four of these had reasonable model performance, but predictions of *Eilema complana* across Europe should be interpreted with caution (F1=0.406, Table 8).

Table 9. *Performance of the UK Denmark classifier deployed in the ABMS pilot, and the in progress European classifier, for an expert annotated AMI dataset in Denmark.*

Classifier	Macro F1	Macro Precision	Macro Recall	Micro F1	Micro Precision	Micro Recall
UK Denmark	**61.14%**	**64.24%**	**67.10%**	**74.29%**	72.08%	**72.08%**
European	46.43%	53.22%	52.06%	60.67%	**76.89%**	53.08%

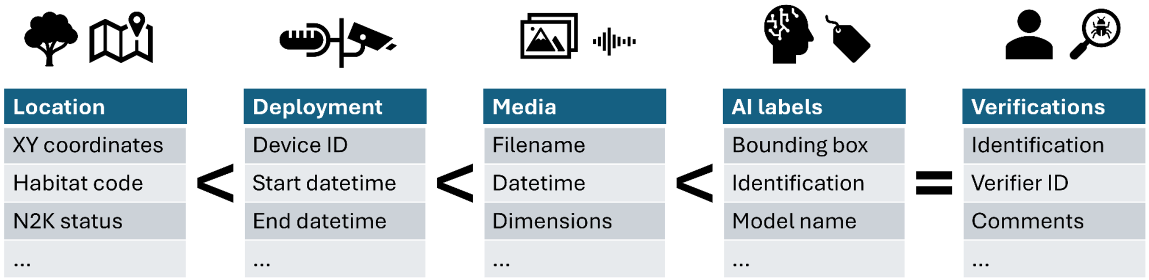
Table 10. *Performance of the UK Denmark classifier for the 30 most abundant species in the expert-annotated AMI dataset, making up just over half of the verification data. While per-species F1 is 61.14%, performance is high for 26 of these 30 common species. Three of the species for which the model performed poorly were not present in the training data for the UK Denmark classifier (underlined). Species in bold are among the top 10 species in at least one habitat type across the ABMS dataset in 2025 (Fig. 18). The specific habitats are in parentheses (F = Forest, G = Grassland, W = Wetland).*

Species	**Count**	**F1 score**
Deltote pygarga **(F, W)**	**3441**	**0.970**
Pelosia muscerda **(F, G, W)**	**2843**	**0.962**
Spilosoma lubricipeda	2074	0.982
Chrysoteuchia culmella	2067	0.861
Eilema complana **(F, G, W)**	**1901**	**0.406**
Calliteara pudibunda	1335	0.959
Eilema lutarella **(G)**	**1232**	**0.876**
Hypena proboscidalis	1146	0.937
Pheosia tremula	1068	0.952
Eupithecia nanata	953	0.976
Ectropis crepuscularia	905	0.000
Collita griseola	878	0.000
Colocasia coryli	837	0.917
Parapoynx stratiotata	728	0.925
Stauropus fagi	712	0.954
Ecliptopera silaceata	698	0.964
Nyea lurideola	694	0.000
Biston betularia	688	0.945
Sphinx pinastri	682	0.925
Epirrhoe alternata	678	0.987
Acronicta megacephala	639	0.955
Pleuroptya ruralis	617	0.936

Rivula sericealis	607	0.958
Thera obeliscata	570	0.824
Pheosia gnoma	567	0.930
Arctia caja	518	0.964
Apoda limacodes	486	0.889
Ochropacha duplaris	483	0.871
***Euproctis similis* (F)**	**476**	**0.843**
Cybosia mesomella	460	0.926

3. Data management

The project generated many types of data, primarily images and recordings collected by insect camera systems and acoustics systems for birds and bats. Beyond images and audio files, we collected data about the locations and deployments during implementation of the pilot. The structure of the metadata tables created during the project are highlighted in Fig. 8. Centralized data management was coordinated by Jamie Alison and Lars Dalby at Aarhus University, with local data management by national or regional data controllers named in the data management plan. Relevant information for data generation is contained in the site selection protocol, the device deployment protocol and the data transfer protocol. In particular, the data transfer protocol captures how partners remotely transfer media files to ERDA (electronic research data archive) at Aarhus University, using a Secure File Transfer Protocol (SFTP), according to standard naming conventions. Other metadata for locations and deployments were collated using online forms, and standardized using quality control scripts written in R.



Location		Deployment		Media		AI labels		Verifications
XY coordinates		Device ID		Filename		Bounding box		Identification
Habitat code	<	Start datetime	<	Datetime	<	Identification	=	Verifier ID
N2K status		End datetime		Dimensions		Model name		Comments
...		...		...		...		...

Figure 8. *Linkages between data and metadata tables within the ABMS. Tables are generated for locations, deployments, and media, with a variety of fields. Following data processing, tables of AI labels and human-generated verifications are produced. The "<" symbol indicates a one-to-many relationship between tables, while "=" indicates a one-to-one relationship.*

3.1. Description of data & metadata

Data summaries represent only data available in the ABMS databases as of October 2025; upload of some data from Ireland was delayed.

Image data: AMI traps collected 45,476 snapshot images (4096 x 2160 pixels in jpg format) in 2024 and 169,523 in 2025, representing more than 8,000 trapping nights in total. The date and time of recording is stored as part of the filename of each image. Significantly more images were captured due to the motion trigger, but these were not processed in order to make efficient use of computing resources, and because motion triggering failed for several devices.

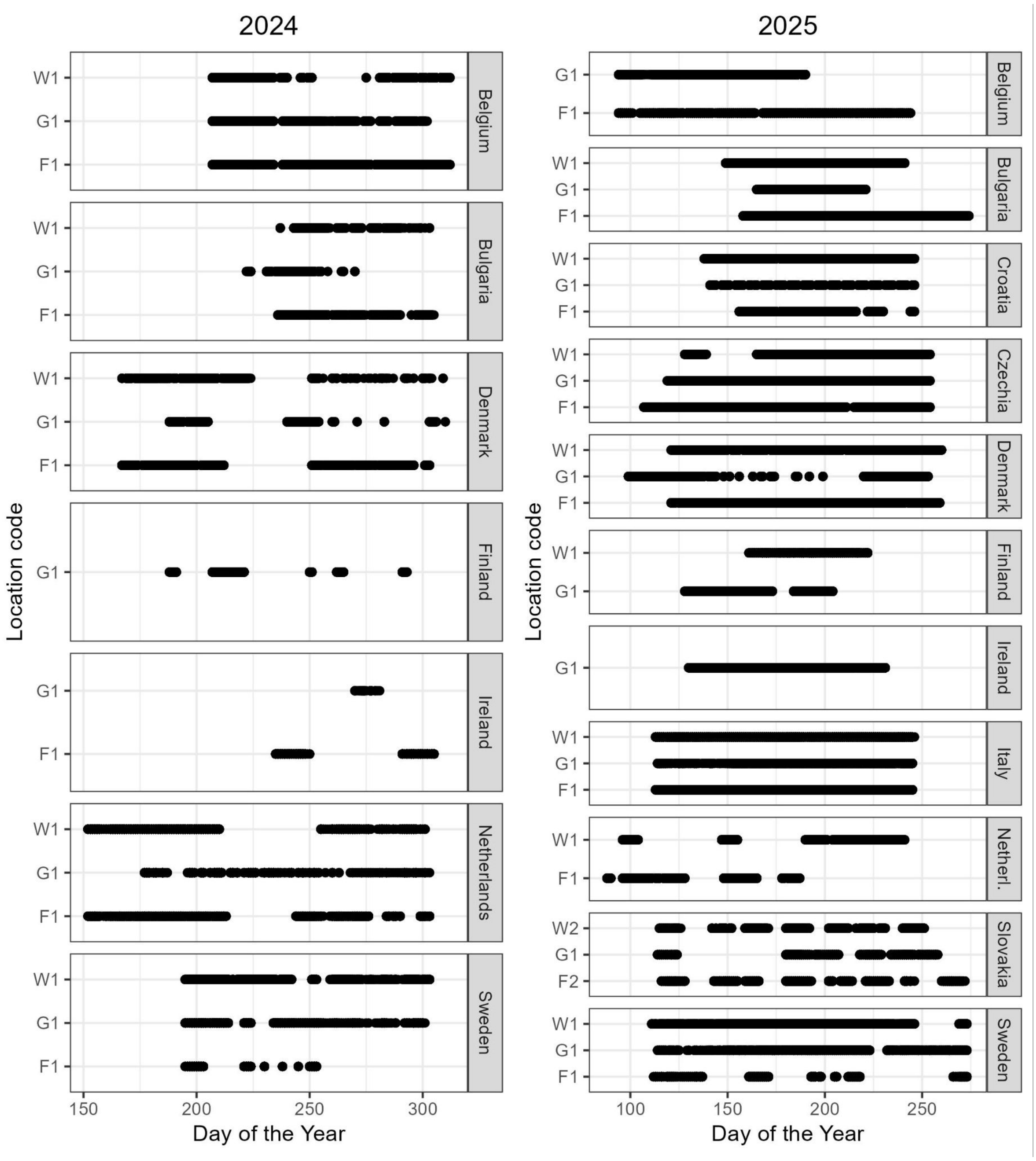


Figure 9. Timeline of image data across all partner countries with active AMI devices in 2024 (left) and 2025 (right). Represents the data submitted to the ABMS AMI database as of October 2025; upload of some data from Ireland was delayed. Gaps correspond to device maintenance constraints.

Sound data: Song Meter Mini 2 systems recorded 2,818,453 audible acoustic files (1 minute long, 48kHz sample rate, in wav format) adding up to 5.43 recording-years, and equating to 26.81 deployment years. Each file is also referenced in a summary text file containing details such as location and temperature.

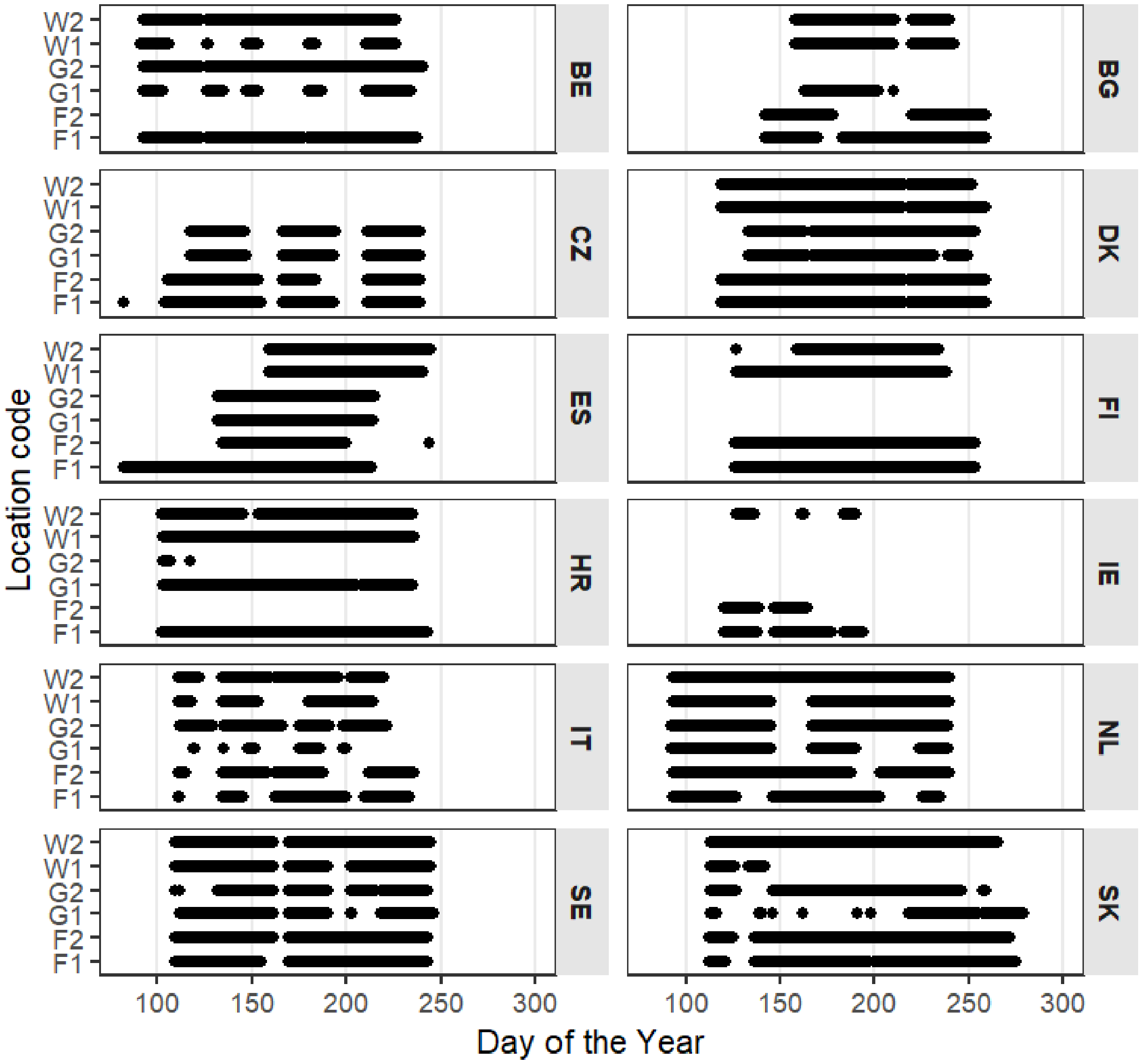


***Figure 10.** Timeline of audible acoustic data collection for 2025. Represents the data available in the ABMS acoustics database as of October 2025; upload of some data from Ireland was delayed. Gaps correspond to device maintenance constraints. Countries denoted using ISO 3166-1 alpha-2 abbreviations.*

Song Meter Mini Bat 2 systems recorded 4,924,603 ultrasonic acoustic files (up to 15 seconds long, 256kHz sample rate, in wav format) adding up to 1.58 recording-years. Each file is also referenced in a summary text file containing details such as location and temperature.

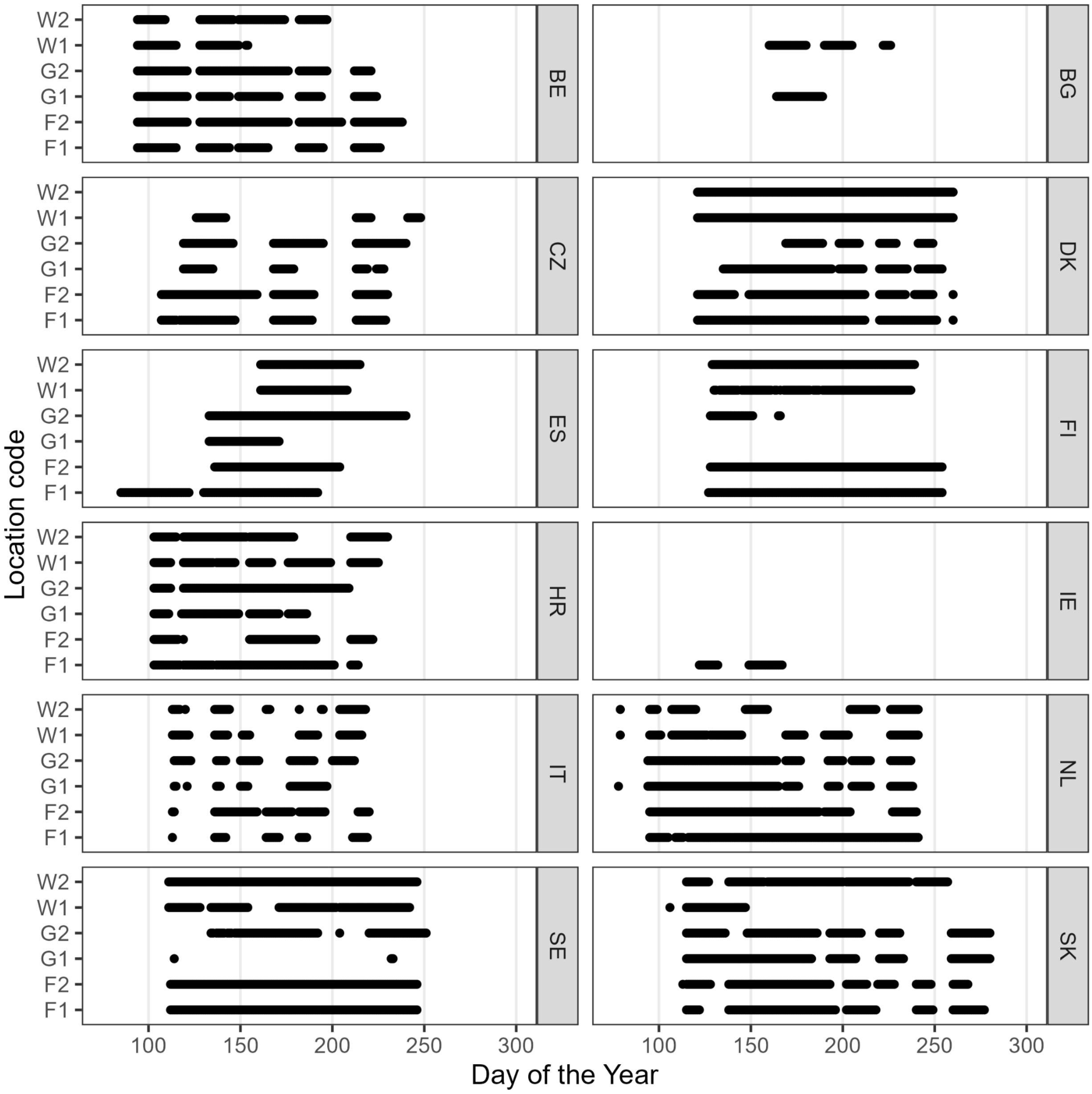


Figure 11. Timeline of ultrasound acoustic data collection for 2025. Represents the data available in the ABMS acoustics database as of October 2025; upload of some data from Ireland was delayed. Gaps correspond to device maintenance constraints. Countries denoted using ISO 3166-1 alpha-2 abbreviations.

Location data: Coordinates, associated descriptors and notes about 51 locations in 2024 and 70 locations in 2025 are tracked in an online spreadsheet, quality checked and written to a central database during indexing of image and sound data.

Deployment data: Partners provide details of devices, locations and time periods that comprise 326 individual ABMS deployments in an online spreadsheet. These data are useful to capture sampling effort and contextual information.

AI model predictions: Images and sounds in the central AMI and acoustics databases are processed using AI models (see 2.4. Processing pipelines). AI predictions are stored in a bespoke database, and output as csv files for analysis and EBV generation. These files refer to source images or sounds, and provide identifications related to regions of the media where species occurred.

Verification data from experts: A subset of 8690 audible and 3048 ultrasound acoustics files were presented to taxonomic experts for review. Outputs were formatted according to the verification protocols, and unified into a single csv file of image names and bat or bird species verifications.

Climate logger data: TMS-4 climate loggers provide data on temperature, moisture and other variables in a single table for each deployment. Data are downloaded from the device as a csv with predefined columns described in the device manual. These are uploaded to the project’s cloud storage, to be unified and analysed at a later stage.

3.2. Ensuring the FAIR data principle

All image and sound data collected as part of the ABMS pilot is stored in the Electronic Research Data Archive (ERDA) at Aarhus University, Denmark. All generated data will be made publicly available at the end of the pilot through a Danish implementation of the dataverse. AI labels will be made available with clear acknowledgement of model uncertainty, as quantified in 2.5. Model evaluation.

Expert verifications for acoustic data will be given special attention, with the creation of an annotated dataset of bird and bat taxa on xeno-canto and an associated publication. Following this, confirmed occurrences will be submitted to the Global Biodiversity Information Facility (GBIF).

4. Deriving EBV Indicators from Sensor Data

The analytical possibilities of ABMS data are vast. To showcase the potential of sensor-derived data to produce high-level indicators across Europe, we centred our approach on Essential Biodiversity Variables (EBVs). Specifically, we identified and produced indicators relevant to several EBV categories defined by the Group on Earth Observations Biodiversity Observation Network (GEO BON; https://geobon.org/ebvs/what-are-ebvs/). Across insects, birds, and bats, we derived three indicators to efficiently capture ecosystem function and community composition (Table 11). These indicators also relate to a wide range of indicators described by EuropaBON for Europe (Quoss, Junker & Wendt, 2024).

Table 11. *Summary of three EBVs defined by GEO BON and their implementation in the ABMS. GEO BON descriptions retrieved from https://geobon.org/ebvs/what-are-ebvs/ on 04/12/2025.*

GEO BON EBV	GEO BON Description	ABMS Description
Ecosystem Phenology	Duration and magnitude of cyclic processes observed at the ecosystem level, such as in vegetation activity, phytoplankton blooms, etc.	Duration, fluctuation and magnitude of bat, bird and insect activity throughout the year.
Community Abundance	The abundance of organisms in ecological assemblages.	The abundance of insect orders, bat genera, and moth and bird species, where activity is a proxy for abundance.
Taxonomic diversity	The diversity of species identities, and/or phylogenetic positions, of organisms in ecological assemblages.	The diversity of moth and bird species identities, as estimated with species accumulation curves.

Beyond EBVs, it is important to note the impressive depth and temporal resolution of the data collected in the ABMS. Exciting questions could be addressed about day-night cycles in activity of bats, birds and insects, impacts of temperature and moisture, and species interactions - especially between insectivorous birds, bats and nocturnal insects. These could include sophisticated methods of imputation for missing data, as well as estimating density from activity data. The analysis presented in this report only scratches the surface of the underlying data, and further work should delve deeper into such questions.

4.1. Ecosystem phenology indices

Regular observations collected throughout the season enable statistical modelling techniques designed to estimate complex non-linear patterns in time. Species activity periods typically exhibit non-linear, hump shaped curves, and Generalised Additive Models (GAM) are suitable to distinguish the seasonal trend from day-to-day “noise”. Furthermore, by modelling many species (or other taxonomic groupings), habitats and countries together, traps with sparse data due to technical difficulties can “borrow” statistical strength from more data-rich traps. Resulting phenology curves can be drawn for most country and habitat combinations, allowing precise predictions of metrics such as peak activity and magnitude, activity onset, activity end and seasonal duration, both at species and community levels. All of these metrics are thought to be sensitive to climate change and intensive management, and multi-year tracking will provide critical information about species and community responses to stressors.

While phenology curves from order-level to species level are also possible, we applied GAMs to all arthropod detections from AMI traps in 2025, modelling total arthropod activity. The resulting curves (Fig. 12) show how arthropod activity increases or decreases over time, after controlling for sampling effort and other confounding factors. These curves have the potential to provide important insights into the timing of activity across habitats within and between countries, even at the community level. For example, in Bulgaria, the timing of peak activity is relatively conserved across habitats but the magnitude of the peaks varies greatly. Conversely, in Denmark, the forest and grassland habitats have long activity periods with multiple peaks, whereas the wetland habitat appears to have a short activity season. Further research into the drivers of these and other patterns, such as vegetation patterns, weather conditions and background species pools could enhance ecological understanding (Gillespie *et al.*, 2025), particularly if patterns can be monitored over long time periods. If produced at species-level, phenology curves could also identify indicator species sensitive to climate changes at levels of precision that exceed traditional monitoring methods.

As for arthropod activity, we also produced indicators of bird and bat activity over time (Fig. 13 & 14).

Importantly, these phenology curves show not only the timing, but the magnitude of activity across countries and habitats (but not axis scales vary between countries). Interestingly, the curves for birds do not always show an expected spring and early summer distribution, highlighting that BirdNet is bird calls, and not just bird songs. For bats, activity is generally higher in forest than in grassland and wetland sites (Fig. 14). Especially in the case of the nocturnal taxa, comparisons between prey and predator species' activity patterns could reveal important intra-species interactions within habitats. Although there are gaps in the raw data (Fig. 10 & 11; large confidence intervals for bat activity in Fig.14), several overlaps in peak activity can be observed between bats and arthropods such as for the Netherlands or the Province of Bolzano. Such multi-taxon analysis of phenology could be used to generate cumulative impact assessments in future.

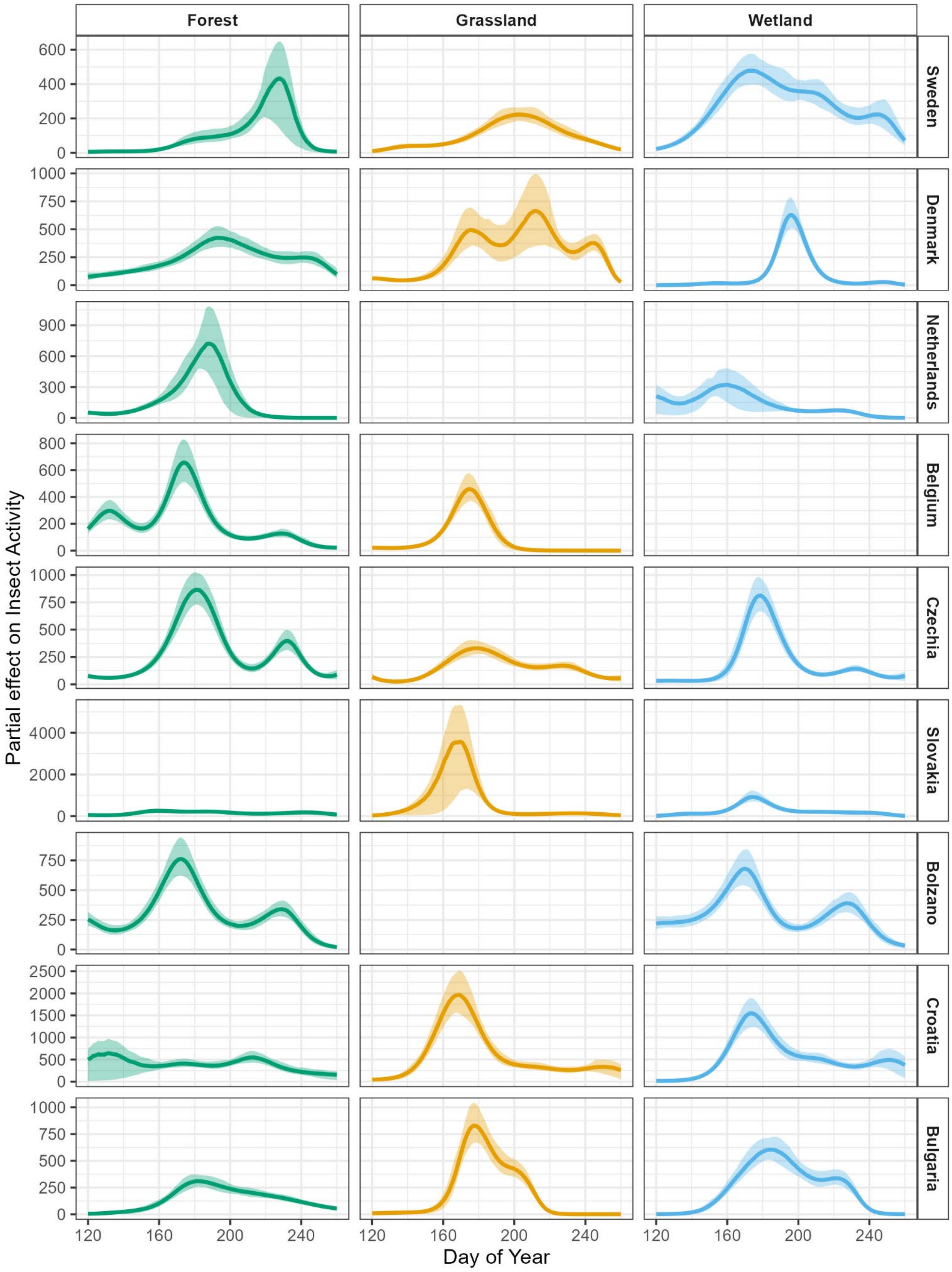


Figure 12. Ecosystem phenology index for activity of night-active arthropods at forest (green), grassland (amber) and wetland (blue) sites across seven ABMS partner regions in 2025. Solid lines represent average activity over time and shaded areas are 95% credible intervals. Sufficient data were not available to construct the indicator for all region and habitat combinations (blank panels) and in some cases, activity was relatively low resulting in "flat" curves that lack detail at this plotting scale.

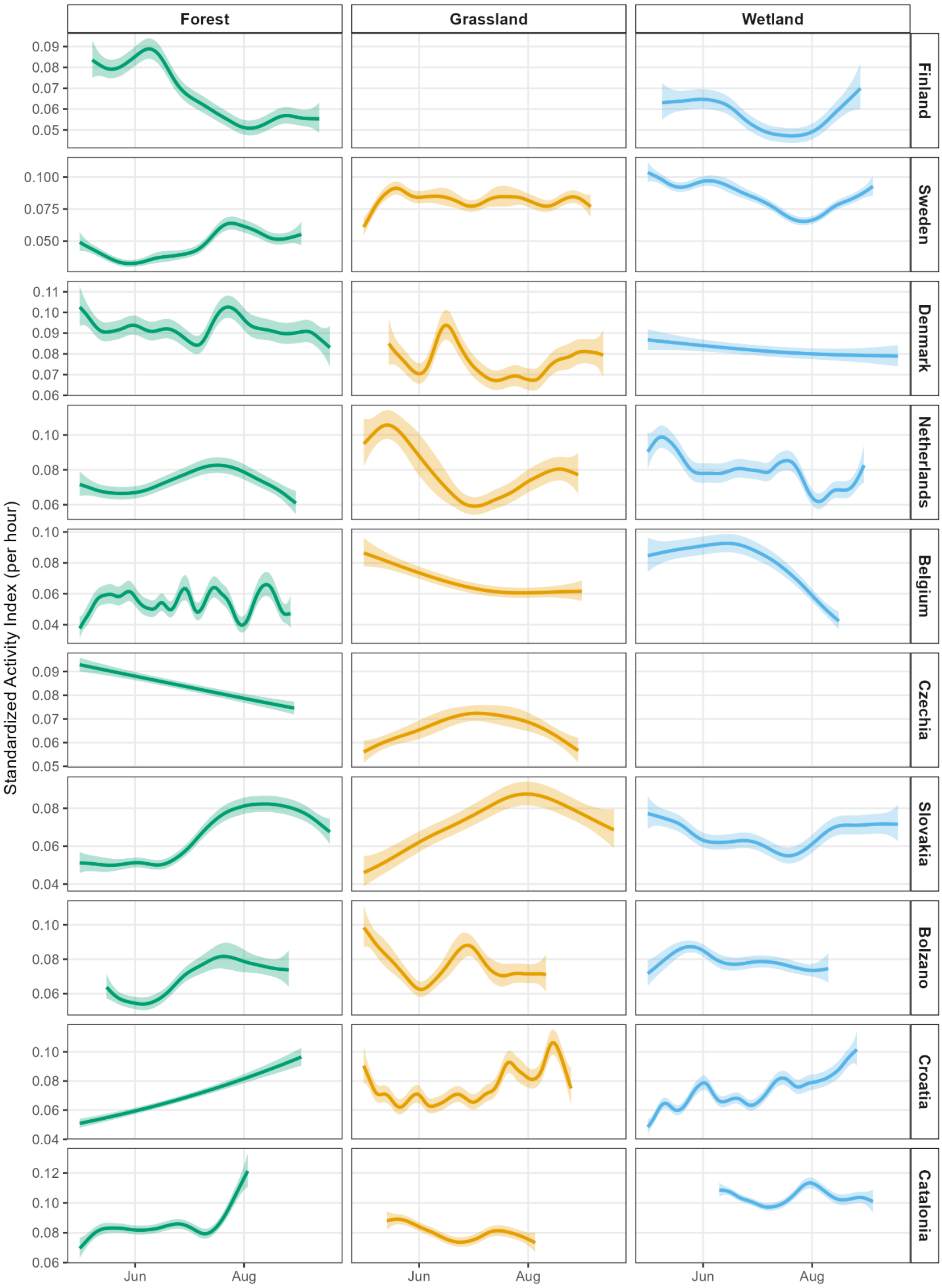


__Figure 13.__ Ecosystem phenology index for activity of birds at forest (green), grassland (amber) and wetland (blue) sites across seven ABMS partner regions in 2025. Solid lines represent average activity over time and shaded areas are 95% credible intervals. Sufficient data were not available to construct the indicator for all region and habitat combinations (blank panels).

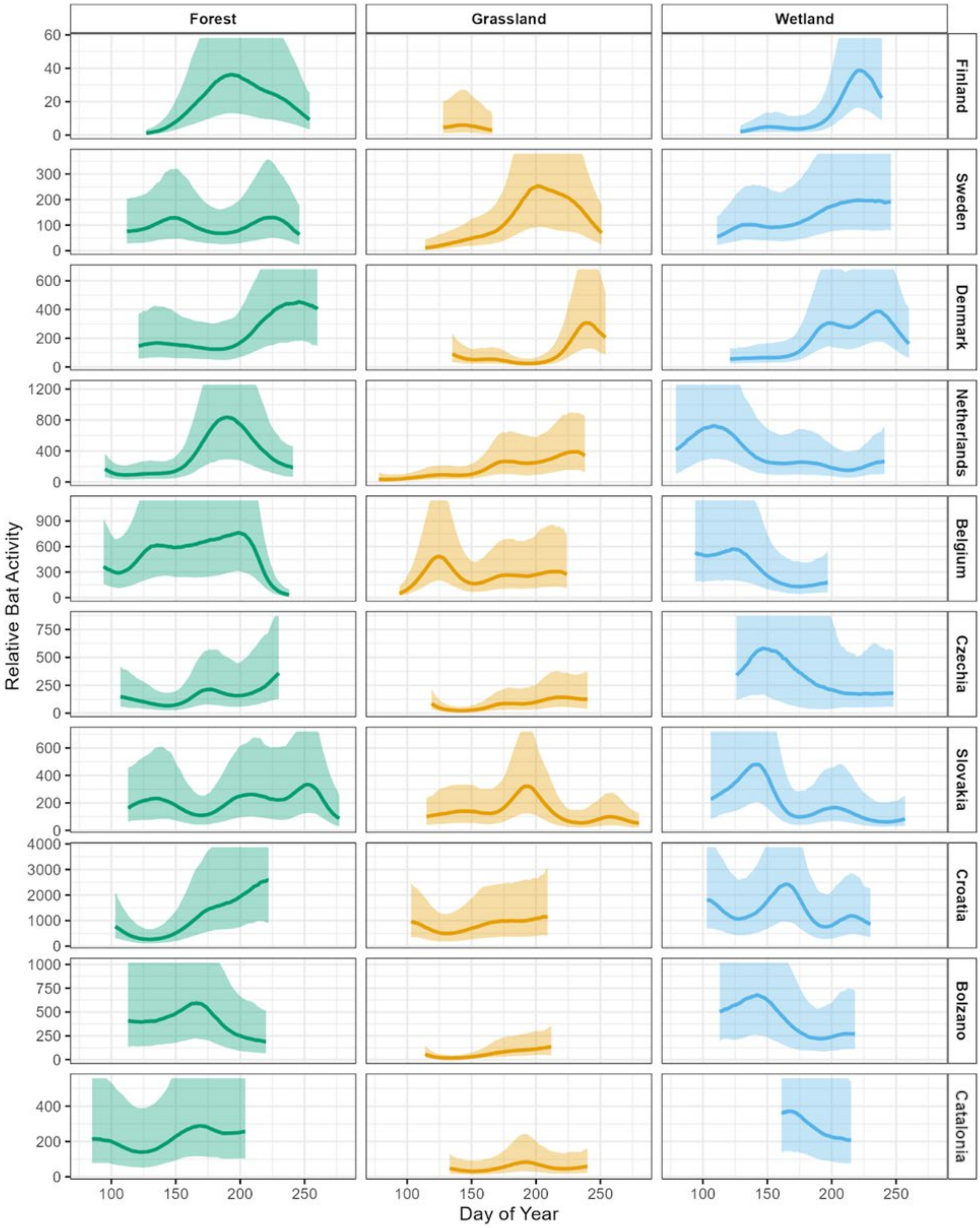


Figure 14. Ecosystem phenology index for activity of bats at forest (green), grassland (amber) and wetland (blue) sites across 10 ABMS partner regions in 2025. Solid lines represent average activity over time and shaded areas are 95% credible intervals (upper intervals are often cut off by the y-axis limits, to better visualize trends). Sufficient data were not available to construct the indicator for all region and habitat combinations.

4.2. Community abundance indices

The community composition class of EBVs includes “community abundance”, which can be interpreted in several ways. At its most basic, an index of the relative activity of different taxa detected by sensors, across habitat and country combinations, provides an informative overview of communities that can also be compared across time and space. This can sometimes be done at species-level, but the simplest presentation of this is shown in Fig. 15 for bats and Fig. 16 for arthropods. For bats, the dominance of *Pipistrellus* species is evident in many habitats and countries, with notable exceptions close to the northern range edge. Long-term monitoring of these relative activity patterns will provide important insights for managers aiming to improve diversity or track dispersal across spatial gradients. Further work might take steps to better convert relative activity metrics into relative abundance, by better accounting for the possibility of double-counting. For bats, this might include developing coefficients of detectability, that relate acoustic activity to the actual number of individuals (Roemer *et al.*, 2025).

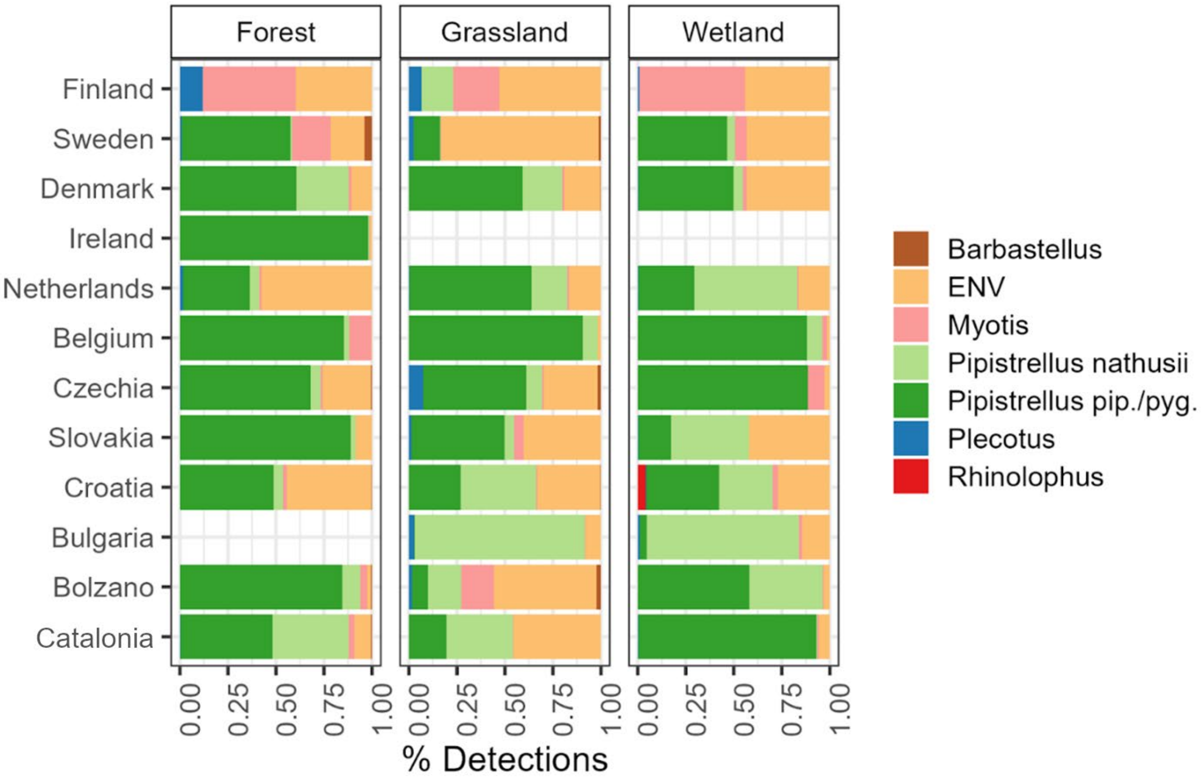


***Figure 15.** Relative abundance of seven bat taxa across three habitats and 11 ABMS partner regions in 2025. Fractions of bars represent the proportion of acoustic files identified to species in a given taxon by the BatDetect2 model. ENV represents a genus aggregate for three genera that are easily mistaken for one another: Eptesicus, Nyctalus and Vespertilio. The separation between Pipistrellus pipistrellus/pygmaeus and Pipistrellus nathusii is retained, to show variation within Pipistrellus which was relatively well classified (Fig. 6). Species outside the listed taxa cannot be classified by BatDetect2, and are often subsumed into ENV (Fig. A1). Data were not available to construct the indicator for all region and habitat combinations; upload of some data from Ireland was delayed, for example.*

For night active arthropods in 2025 (Fig. 16), relative activity levels of different orders vary greatly across habitats and countries, with some northern grasslands and wetlands dominated by Diptera, and diversity increasing further south. The interpretation of these patterns are subject to an understanding of the location of the traps, but with replication across multiple sites and environmental data about vegetation, management and weather, researchers and managers could explore the drivers of community abundance patterns across ecological gradients.

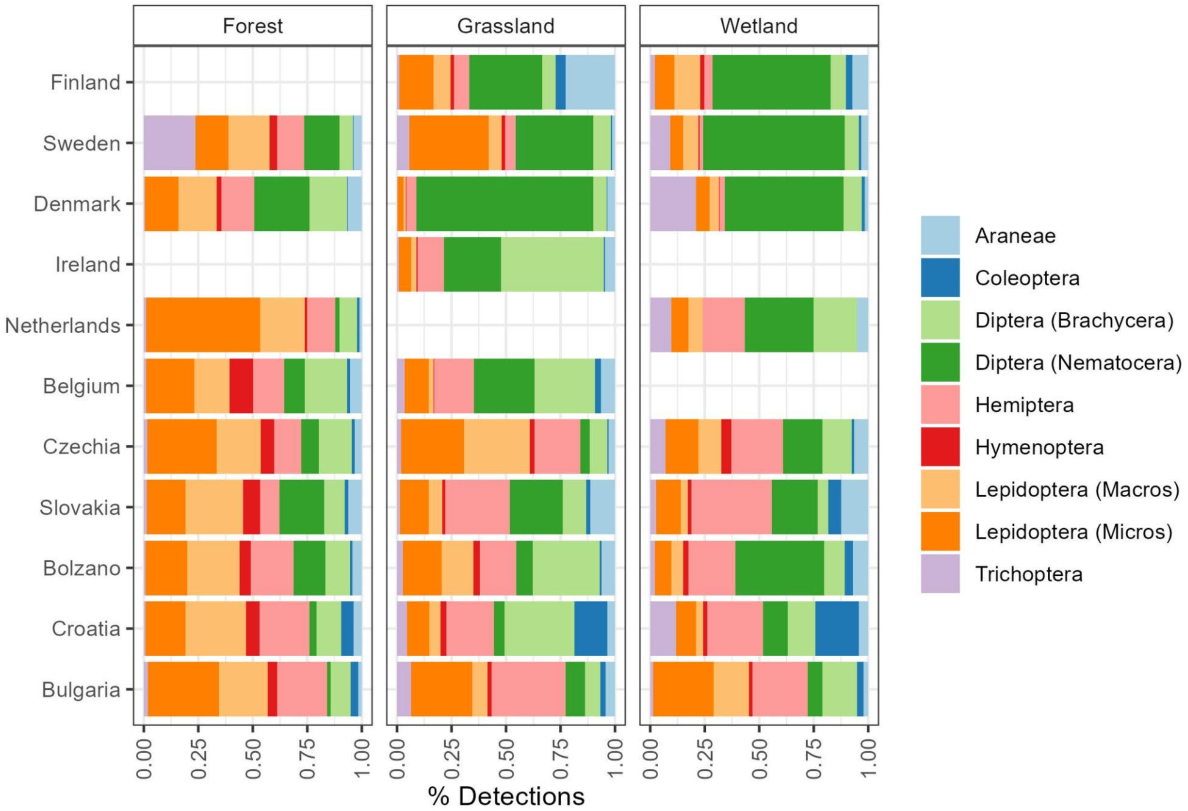


Figure 16. *Relative abundance of nine arthropod taxa across three habitats and 11 ABMS partner regions in 2025. Fractions of bars represent the proportion of recorded instances of arthropods belonging to each taxon. Images were recorded every ten minutes between 2300 and 0300 local time. Data were not available to construct the indicator for all region and habitat combinations; upload of some data from Ireland was delayed, for example.*

Community abundance can be further broken down to lower taxonomic levels, subject to classification confidence. For example, the top 10 bird species detected with high confidence across countries in the three habitats (Fig. 17) reveal insights into the species that dominate European bird communities, which could be tracked over time and further broken down by spatial units. *Phylloscopus collybita* emerges as the most abundant species in both forest and wetland locations. This pattern reflects how this species mainly occupies different types of forests during the breeding season, while it can be found in abundance in wetlands with reed beds during migration and wintering periods. On the other hand, *Alauda arvensis* is the most frequently detected species in grasslands. This finding is consistent with the ecology of the species, as it specializes in grassland habitats throughout the year. Still, some results would benefit from further investigation. For example, *Anthus trivialis* appears as the second most detected species in grasslands, and the third in wetlands. However, this is typically considered a forest species, and it was an “amber” species in our verification (Table 5; not possible to reach 90% precision). Thanks to our verification, we might suspect that other *Anthus* species are being mistaken for *Anthus trivialis* in this case.

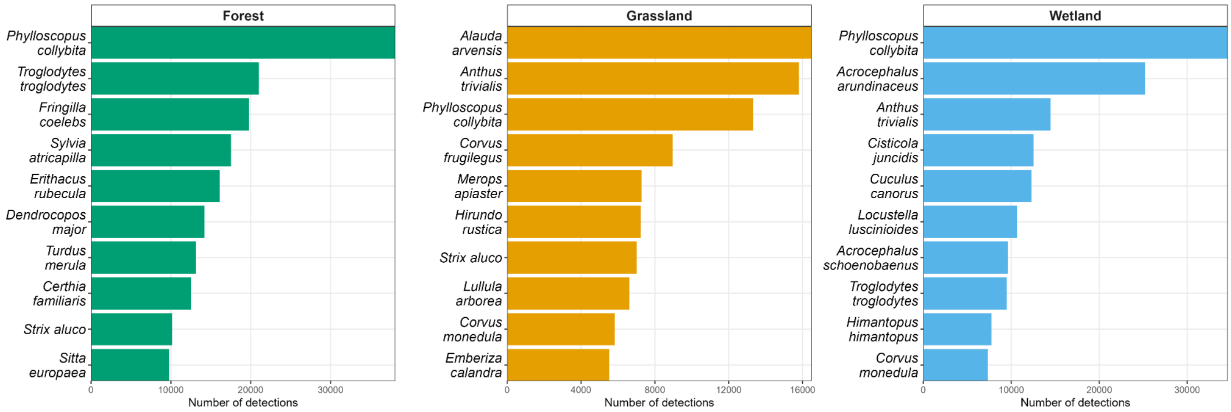


Figure 17. *The number of AI (BirdNet) detections of the top 10 bird species in each habitat type summed across all countries..*

A summary of the most abundant moth species across Europe (Fig. 18), shows some ever-present species such as *Pelosia muscerda* (the dotted footman), known for an association with mild climates, wet woodlands and wetlands. However, the habitat types also have distinct species, for example *Chiasmia clathrata* (the latticed heath) which is found in open habitats and was highly abundant at grassland locations. While monitoring the relative abundance of common species will reveal much about community abundance changes over time, biodiversity sensors will also provide vital information about rare and threatened species. Still, as for birds, species-level detections from AI should be interpreted with caution; *Eilema complana* (the scarce footman) is commonly detected in all three habitats, being widespread across Europe and found in a range of habitats. However, our verification in Denmark suggested the results for this species should be interpreted with caution, while those for *Deltote pygarga* and *Pelosia muscerda* are probably much more reliable (Table 10).

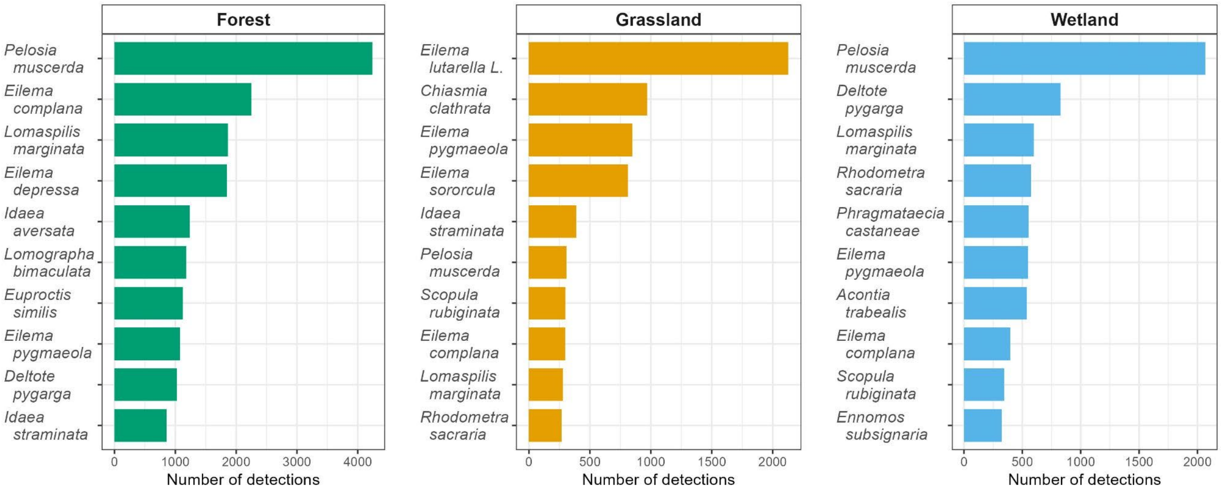


Figure 18. *The number of AI detections of the top 10 night-flying macro moth species in each habitat type summed across all countries.*

Exploring further than the focal EBV, relative abundance of species in a community can be viewed collectively in terms of their associations with each other and their habitats. A common way to do so is with ordination techniques, where community composition is plotted by measures of “dissimilarity”. Sampling locations that are close together on the ordination graph tend to exhibit bird or insect communities that are more similar, while those that are far apart are made up of very different species. This is important for monitoring, especially to determine whether communities are becoming more alike (i.e. “homogenizing”) over time - perhaps becoming increasingly dominated by a few common species.

Interestingly, for the 2025 data, species community ordinations for moths and birds (Fig. 19) reveal that across Europe, forest communities tend to be more similar from country to country, while grasslands and wetlands vary widely in their community makeup. Further study can help to identify why this might be the case, for example by comparing the context of each site, or by identifying ecological gradients associated with these patterns.

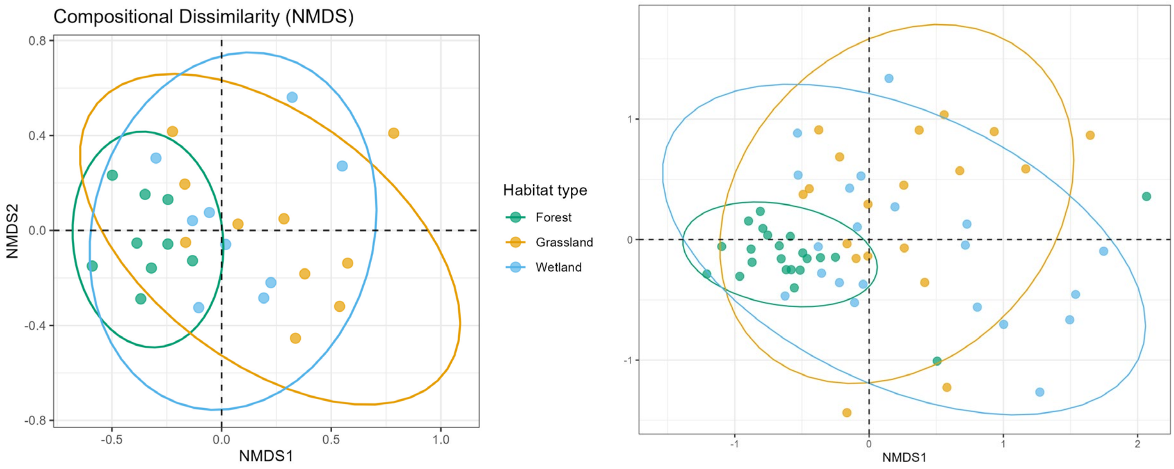


Figure 19. Compositional dissimilarity of the macro moth (left) and bird (right) species communities across Europe, separated by habitat. Each point represents a sampling location in the given habitat, while ellipses represent the 95% confidence interval around the group’s centroid. For both taxa, communities in forest habitats are much more homogenous than those in wetland or grassland habitats.

4.3. Taxonomic diversity indices

A third EBV derived rapidly from sensor-based data represents taxonomic diversity, the simplest measure of which is species richness. This can be a simple count of the number of species detected, subject to classification confidence, but techniques are available to account for differences in sampling effort. For example, sampling effort can be measured as the number of nights that an AMI trap is active over 2025, which ranges from 55 days in Finland to 150 days in Sweden, such that total species richness is incomparable. However, using a method known as rarefaction, we can draw species accumulation curves for each country and habitat combination. These derive the rate of new species detection with increasing sampling effort, up to the observed level (the solid line in Fig 20). Beyond the observed level of species richness, the technique can extrapolate the rate of new species detection if the device continued to operate. We can then more easily compare the Finnish grasslands and wetlands to the Swedish sites, at different levels of sampling effort.

Within each country it is also informative to explore the relative species richness between habitat types. In most countries, particularly in northern Europe, forest sites are the most species rich, while in southern Europe the sites chosen for monitoring are more closely matched. Following these patterns over longer time scales will provide critical insights into the impact of factors such as land management on diversity. These metrics can also be adjusted to account for the relative abundance of each species (e.g. Shannon or Simpson diversity indices) to study diversity patterns in conjunction with the community composition methods above. While comparisons between habitats are robust, comparisons between countries should be made with caution until a European classifier is available; the best available AI moth classifier is currently trained with images from the UK and Denmark (See 2.5. Model evaluation), so that species numbers in southern European countries are probably underestimated.

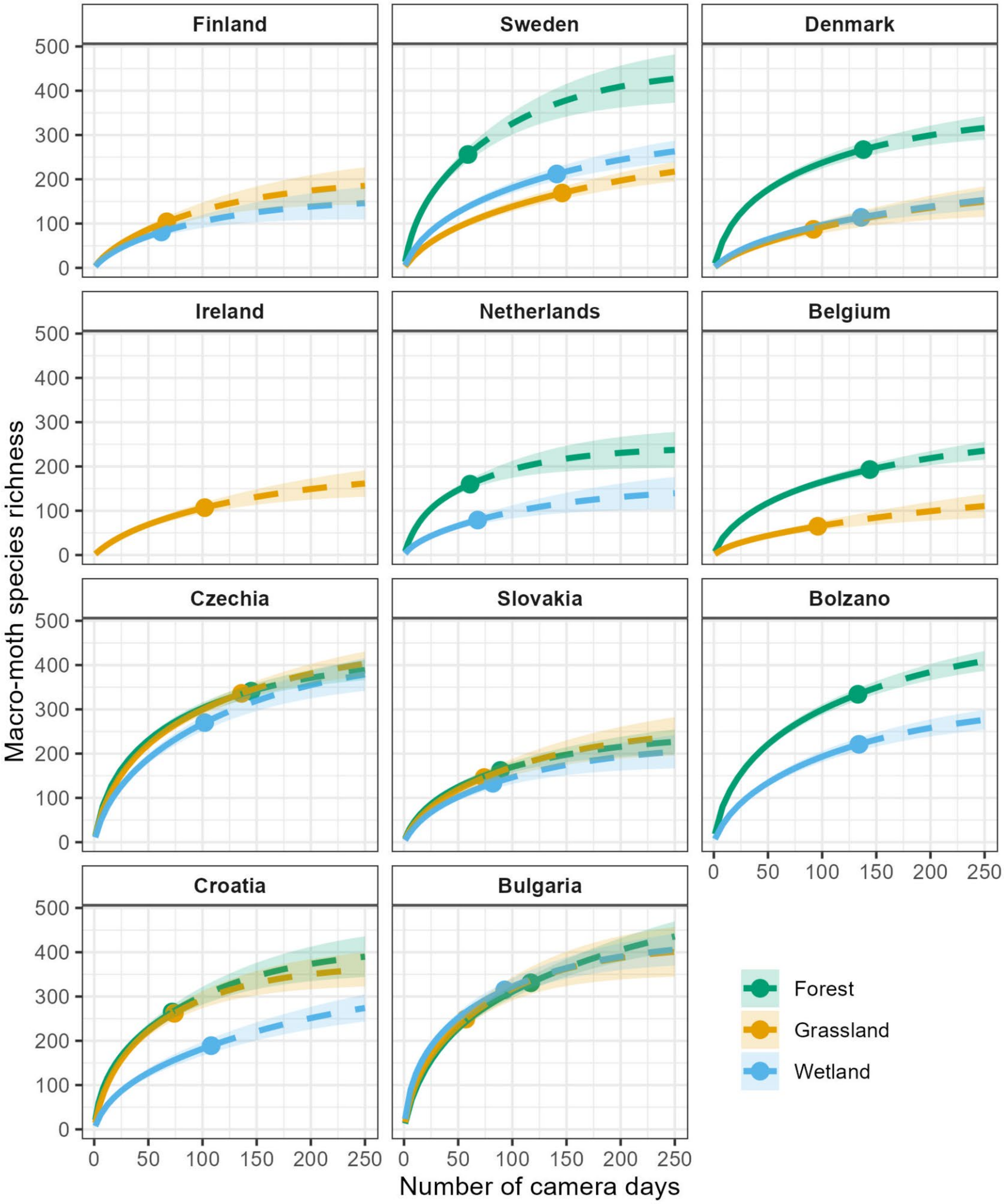


Figure 20. *Species accumulation curves for macro moths at forest (green), grassland (amber) and wetland (blue) sites across 11 partner countries. Species were classified using an AI model. The point symbol depicts the observed number of species detected, the solid line depicts the interpolated number of species likely to be detected with reduced sampling effort (active camera days), and the dashed line shows the extrapolated number of species likely under increased effort. The shaded error shows uncertainty in the extrapolation - uncertainty increases when fewer observations are available.*

5. Towards a European Biodiversity Sensor Network: Lessons learned from the ABMS pilot

Beyond a rich dataset and suite of transnational monitoring sites, the ABMS pilot has yielded an active community of practice, with strong ties to national biodiversity monitoring networks, and clear lessons learned. To capture these, a survey was completed by all 13 ABMS partners. Answers to the most pertinent questions are presented below, under headings related to six key lessons learned, with answers from 15 respondents representing 13 regions and partner organizations.

5.1. Sensors are viable for transnational biodiversity monitoring

Our dataset highlights the efficiency and scalability of sensors for biodiversity monitoring across borders, and this was also reflected in our survey. When asked “Could ABMS-like monitoring be integrated with existing monitoring schemes in your country?” every respondent replied “yes”. For acoustics sensors, 13 of 14 respondents agreed that “working with sensors and sensor-derived biodiversity data has been a positive experience” (10 of 14 for AMI data). Furthermore, 11, 10 and 8 respondents (69%) agreed that biodiversity data derived from bat recorders, bird recorders and AMI devices were more cost-effective than traditional alternatives.

5.2. Costs related to sensors can exceed expectations

While offering good return on investment, we experienced unpredictable costs when using sensors for biodiversity monitoring. Half of our respondents found that equipment for the pilot exceeded the budgeted amount, as the cost of devices differed between countries. Furthermore, we experienced major problems with the purchasing of AMI devices, a “bleeding edge” and thus risky technology which turned out to have a very poor chain of distribution; the devices are not yet marketed by retailers in the European Union. Almost half (7 of 15) of our respondents rated the ease of purchasing of AMI devices as 1 or 2 out of 5. Rototherm was the UK supplier, chosen on behalf of UKCEH, and many partners experienced issues related to significant delays in delivery, tax, customs clearance and condition of the devices. Many arrived with damage, needing to be repaired before deployment; respondents reported that 16 of 36 AMI devices had to be repaired over the course of the pilot, while 3 devices failed completely. Some other devices also failed: 2 bird recorders, 3 bat recorders and 4 climate loggers failed completely, while 1 bird recorder and 1 climate logger were stolen. Clearly, replacement costs must be budgeted throughout any sensor-based monitoring programme.

Investment in sensors goes much further than replacing and repairing them; we found that human resources to understand, deploy and maintain devices can also exceed expectations. Again, these costs were increased for AMI devices; 9 of 14 respondents (64%) rated the ease of deployment of AMI traps as 1 or 2 out of 5. Furthermore, half of our partners found that more than 35% of the time spent working with the AMI devices, including transport, was spent on troubleshooting the device (usually 10% or less for other sensors). This represents a considerable human resource investment in the use of these sensors. When using pioneering technologies, it is necessary to budget extra time to deploy them, but also to recruit or train technicians who specialize in their maintenance. In order to be safe and sure that custom sensors are functional, field technicians may need training in basic electronics, including use of a multimeter, soldering, and identifying short circuits.

5.3. Sensors constrain sampling and site selection

When selecting sites for the ABMS, we kept our criteria broad, but tried to introduce some rules to increase consistency across partners (see 2.2. Site selection). However, when selecting sites and sampling locations, ABMS partners faced a variety of constraints (Fig. 21). The most important related to permission, accessibility, and habitat quality, which are common constraints to ecological monitoring. However, in the context of sensors, constraints such as security from people, local perceptions and privacy were particularly important. Partners were generally concerned about the theft of their equipment, or that land users would not perceive the monitoring positively (Fig. 21). Some partners were also concerned about terrain, distance, and security from wildlife and livestock, but others less so. Thankfully, light pollution from equipment was not a prevalent concern, as sites were often remote, and power provision was not an issue when using battery powered devices.

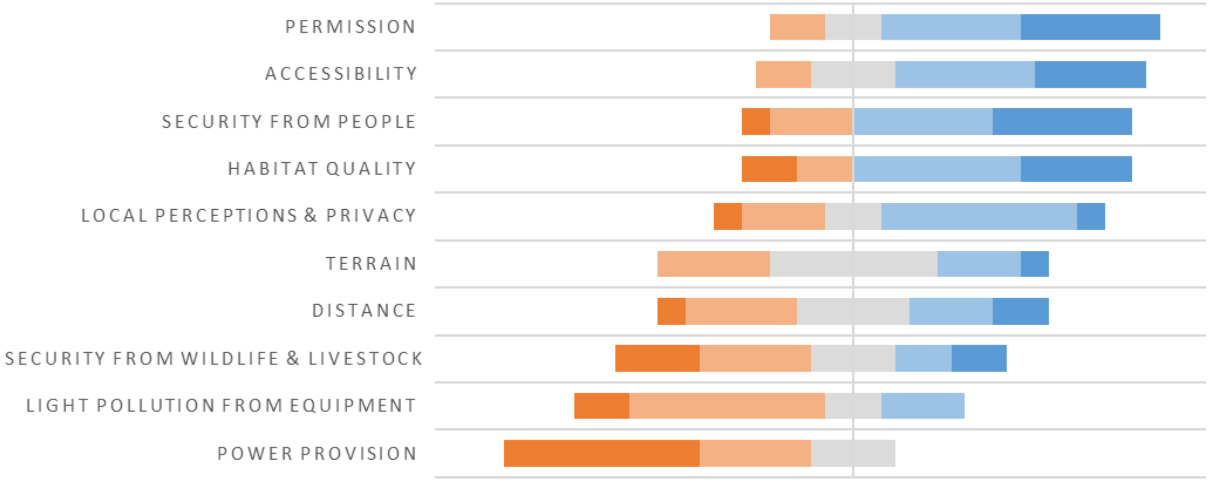


Figure 21. *Constraints on selection of sites and locations for sensor-based monitoring, ranked from most to least important. Ratings from 1-5 were provided for each constraint by 14 partners, with a rating of 1 (red) being least important and 5 (blue) being most important.*

These constraints made it difficult to meet strict sampling criteria over short timescales. In 2024, Finland could not access a wetland site, and sampled another available site that did not meet habitat criteria. Furthermore, for practical reasons, 10% of locations needed to be moved between 2024 and 2025, showing the importance of keeping location metadata for analysis and interpretation. Just over 75% of locations were able to be placed within Natura 2000 sites. Given site selection constraints, it makes sense that almost two thirds of locations were based on sites where monitoring or research partnerships already existed. This highlights synergies with existing research and monitoring across partner regions. However, it also highlights the importance of existing networks and partnerships for access and continuity of sensor data.

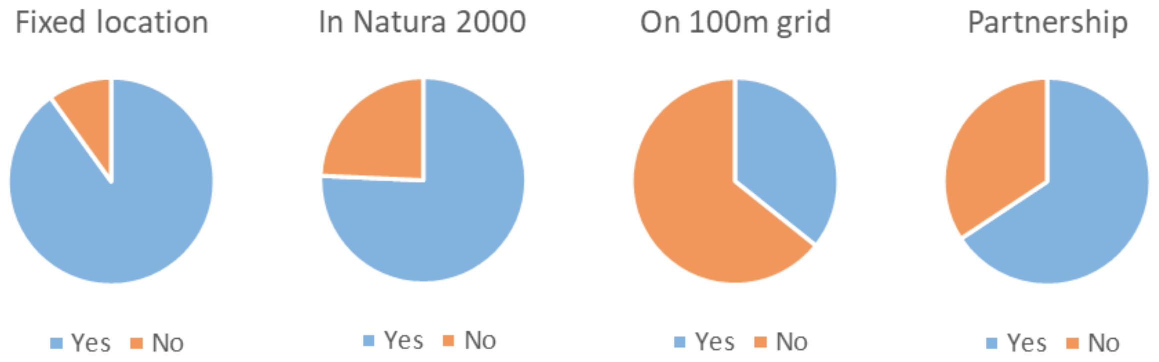


Figure 22. *Characteristics of 70 ABMS sampling locations for 2025. Each of 12 partners reported the number of locations that (1) did not need to be moved 20m or more between 2024 and 2025, (2) were inside Natura 2000 areas, (3) were successfully placed on the 100m EuroStat grid, and (4) were placed on sites where there was an existing partnership in place.*

Finally, initiative was taken in the pilot to test the possibility of sampling predetermined points using sensors, much like the LUCAS soil survey (Fernandez-Ugalde *et al.*, 2022). To that end, partners were asked to sample points at the closest intersection of the 100×100m EuroStat grid (EPSG:3035, ETRS89-extended / LAEA Europe). However, almost two thirds of locations failed to meet this criterion, highlighting the difficulty of probability sampling with sensors that are heavy, costly, or require attachment to stakes or trees. This is a shame, as probability samples are very useful for statistical inference. However, nonprobability samples are common in biodiversity monitoring, and methods exist to make best use of them (Boyd, Powney & Pescott, 2023).

5.4. Continuity of sensors is high - but expect gaps in the data

A major strength of sensor-based approaches, confirmed during this pilot, is the continuity of data collection. Collecting data every few minutes ensures that most passing organisms can be counted, and allows intricate study of trends over short or long timescales (see 4.1. Ecosystem phenology indices). For this reason, among others, on average 13 of 14 respondents agreed that data from AMI, bird, and bat sensors offer added value compared to traditional alternatives. Beyond this, sensors offer consistent sampling and data structures, enabling robust comparisons of sites and habitats (see 4.2. Community abundance indices).

However, this pilot highlights that while sensor data are continuous in nature, gaps are difficult to avoid (See Figures 9-11). For continued operation, 10 of 12 partners agreed that acoustic sensors needed at least monthly maintenance; in the Province of Bolzano in 2025, battery life even dropped below two weeks for acoustics devices. Biweekly visits cannot always be upheld in a busy field season. Even for the AMI traps, which usually had solar panels, most ABMS partners agreed that one visit every two months is prudent. However, even with frequent visits, equipment can fail unexpectedly. As well as mentioned AMI trap failures, respondents reported that 5 bird recorders, 5 bat recorders and 3 climate loggers failed temporarily during the pilot.

Of course, some data gaps are preventable. Acoustic sensors supported 1tb SD cards, so storage was not an issue, but gaps in acoustics data could be minimized by investing more in batteries. Most partners opted for rechargeable alkaline batteries, partly on the grounds of environmental impact. However, in reality, these batteries increased the frequency of site visits, and were probably a false economy - both in terms of finance and environmental footprint. We would recommend high capacity lithium batteries in future. For AMI devices, gaps could be minimized through the installation of internet-of-things infrastructures, allowing real-time browsing of data via an internet connection. These have been well-received in the Invasive Alien Species pilot, helping prioritize field visits to times when sensors really need maintenance. However, some data gaps are difficult to quantify, let alone prevent. For example, heavy wind, rain and other interference can severely impact the ability of recorders to detect birds and bats. Methods are needed to quantify and control for the effects of weather.

However, this pilot clearly highlights how processing and analysis pipelines must be able to deal with gaps in the data. An important element of this is retaining information on sensor downtime and failures, so that sampling effort can be properly controlled for. The presence of media files is an initial indicator of the status of sensors, but some failures result in invalid media - for example, a dysfunctional microphone or a spider web or a blurry camera. Once these phenomena are documented, statistical tools exist to recognise and cope with gaps in the data. For example, when constructing ecosystem phenology indicators, our models interpolate gaps by borrowing information from other sites in the same country or habitat. The result is continuous indicators with increased uncertainty during data gaps (Fig. 12).

5.5. Strategic expert verification helps deal with AI uncertainty

We selected AI models that we considered the most available, reputable and suitable for running on cloud computing resources. However, we anticipated from an early stage that models would make identification errors, especially in regions for which they had not been trained. The clearest example of this was for the BatDetect2 model, which was only trained to identify 17 species. Partners in Croatia provided a list of 18 species that they would expect to pick up in bat recordings, of which only 8 are captured in the BatDetect2 list; *Hypsugo savii, Miniopterus schreibersii, Myotis capaccinii, Myotis emarginatus, Nyctalus lasiopterus, Pipistrellus kuhlii, Rhinolophus blasii, Rhinolophus euryale, Tadarida teniotis* and *Vespertilio murinus* are possible species at the Croatian sites which cannot be detected by BatDetect2. Furthermore, even experts cannot always distinguish species that form phonic groups, such as *Eptesicus/Nyctalus/Vespertilio*, *Myotis* spp., or *Hypsugo savii/Pipistrellus kuhlii/nathusii*. Given similar issues for birds and insects, it is not surprising that all respondents agreed that human verification is necessary to make use of AI outputs.

But how can verification help us make use of AI outputs, without negating the benefits of rapid AI processing? As we do not expect to verify all the data, it is important to strategically subset data for verification, and the best strategy may depend on the overall goal (Fig 23). In the ABMS, our verification was focussed on error quantification and confidence calibration. For moth species, we used existing verifications from Denmark to understand the performance of AI models. However, partners expressed interest in generating many more verification data for insects across Europe. This is especially important for the southern parts of the continent, where there are even more similar looking moth species that can not be reliably identified using AI models for northern Europe and the UK. For bats, we carried out a more comprehensive verification in which seven partners delivered verifications of bat records across 17 species. This helped us capture the variable performance of BatDetect2 at species level (Fig. 6), allowing us to select a more appropriate taxonomic level at which to interpret the data (Fig. 15). The verifications also allowed us to set thresholds to move from call-level to file-level identifications. Similarly, for birds, we used verifications from five countries to build confidence calibration curves for 86 species, converting BirdNet confidence scores into ABMS probabilities (Wood & Kahl, 2024). We identified some species that were identified very reliably, such as the European Robin *Erithacus rubecula*, and others that were not, such as the Tawny Owl *Strix aluco* (Fig. 5, Table 3). Crucially, we were able to use adjusted confidence thresholding for those 86 species to increase the accuracy of our EBV indicators. When published, ABMS verifications will also contribute to improving the performance of AI models across Europe.

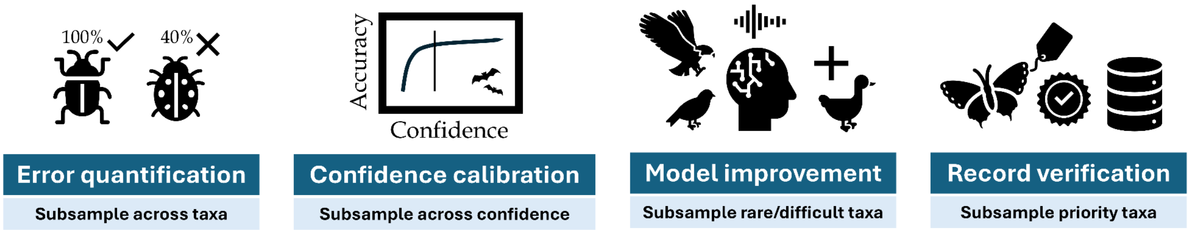


Figure 23. *Different verification goals require different strategies for subsampling data. Experts can help to determine how and where a model tends to fail across a range of taxa, aiding interpretation of uncertain AI outputs. Furthermore, if verifications are made across a range of AI confidence levels, they can be used to calibrate confidence to true probabilities for a given context, making more accurate predictions using the same AI model. Alternatively, AI models can be improved using additional verified data capturing difficult taxa or contexts. Finally, experts can offer final verification of records of species of interest, such as invasives, pests or priority species, focussing expert time and attention to maximize impact.*

Beyond interpreting and improving AI models, verification must be a requirement for a species-level biological record to be considered valid. A perspective from the Nature Conservation Agency of the Czech Republic emphasizes that occurrences of protected and red-listed species may have an impact on area development. When an observation is made by a human expert, the expert can be held accountable, and this is not yet the case for AI; a human expert still needs to validate red-listed and protected species and inform management. This is important when sensors generate records of rare or elusive species. For example, the black-crowned night heron (*Nycticorax nycticorax*) was detected at the grassland site in Ireland in 2024. This species is a rare visitor to Ireland with a few records every year, normally on coastal sites around the south or east of the country. Files with *N. nycticorax* detections were sent to two experts as a part of the verification process, and two detections, recorded between 1 and 3am on the same day, were deemed correct. Given there were only eight other records for *N. nycticorax* in the whole country in 2024, and none in the midlands of Ireland, verification was crucial to confirm that these records were not false-positive detections.

While verification was extremely valuable for the pilot, we also determined areas of improvement for the verification process. In particular, we found that verifiers benefit from having more information than the AI model in order to reach reliable identifications. That may include access to longer recordings, but also information on the location and ecological context. Previous studies with bird verification have used snippets of 5s (Scanferla *et al.*, 2025) or ≥3s (Wood & Kahl, 2024); we opted for 3s to avoid the complexity of pooling confidence scores across snippets. However, verifiers found that for a considerable number of species, particularly *Sylvia borin* and *Anthus spinoletta* in Belgium, 3s was simply too short to identify with real certainty. Future efforts might provide longer snippets, or even entire 1 minute files (Pérez-Granados *et al.*, 2025), giving verifiers more information to work with. The downside is that annotation of longer files is much more intensive, as exact start and end times must be provided for bird calls. For bat verification we asked for identifications at file-level, but also call-level to match the BatDetect2 predictions. The 15 second snippets were apparently sufficient for file-level verification, but call-level identifications were not feasible or useful for most of our verifiers, highlighting the importance of longer time windows for species identification. Finally, we found it was advantageous to have verifications across as many taxa and regions as possible, but that larger sample sizes for each species were also highly desirable. In general, development of consistent and robust verification or annotation protocols is a priority for comparable sensor-based monitoring.

5.6. Local processing fosters engagement and understanding

Centralized processing has been a huge draw for partners of the ABMS (see recommendation 7). At the same time, partners have shown a collective interest in AI processing solutions. Throughout the pilot, it became clear that there is significant interest in capacity building across partner regions in terms of data processing, analysis and verification. This could include training in how to install, use, filter, and analyze outputs from BirdNet, BatDetect2 or insect classifiers. It could also include workshops and provision of user-friendly code to generate EBV indicators from sensor data. Experience from the ABMS supports the principle laid out by Silva del Pozo *et al.* (2023); Specifically, there is a role for both centralized and localized data processing in transnational monitoring, and localized processing is crucial to build an equitable network in which all partners are empowered and motivated.

While the pilot will make resources available for AI data processing, a main objective was to make tools available to visualise the locations, identity, and time of observation of species recognized by sensors. We focussed on images with insect species, aiming to facilitate creation of annotated datasets of insects across all partner regions. We explored efficient ways of doing so, including software tools such as MothBot, Antenna, and AnnFlux. AnnFlux in particular is a research tool for exploring and annotating large datasets with Active Learning, developed by Laurens Hogeweg at Naturalis. It uses a few annotations from a human expert to infer labels across the rest of the dataset, aiming to target annotations to maximize model improvements.

For the ABMS pilot, an instance of the AnnFlux application was created for each partner country. Clustering of detected animals can be visualised in a 2D scatterplot based on features derived from an AI model (Fig. 24). The distance between samples in this so-called “embedding space” is used to label clusters of points that are visually similar, and likely to belong to the same taxon. Users are actively prompted for annotations where uncertainty is highest, and the internal model is periodically fine-tuned, updating the clustering. Our other ideas for AnnFlux or similar tools include streamlining verification and adjustment of AI outputs, loading them in as preliminary annotations for the user to explore and correct. Further development and evaluation is needed to integrate such tools, but early testing indicates that they are crucial to maximize engagement and access to AI-generated data, as well as to generate annotations in regions with poor coverage.

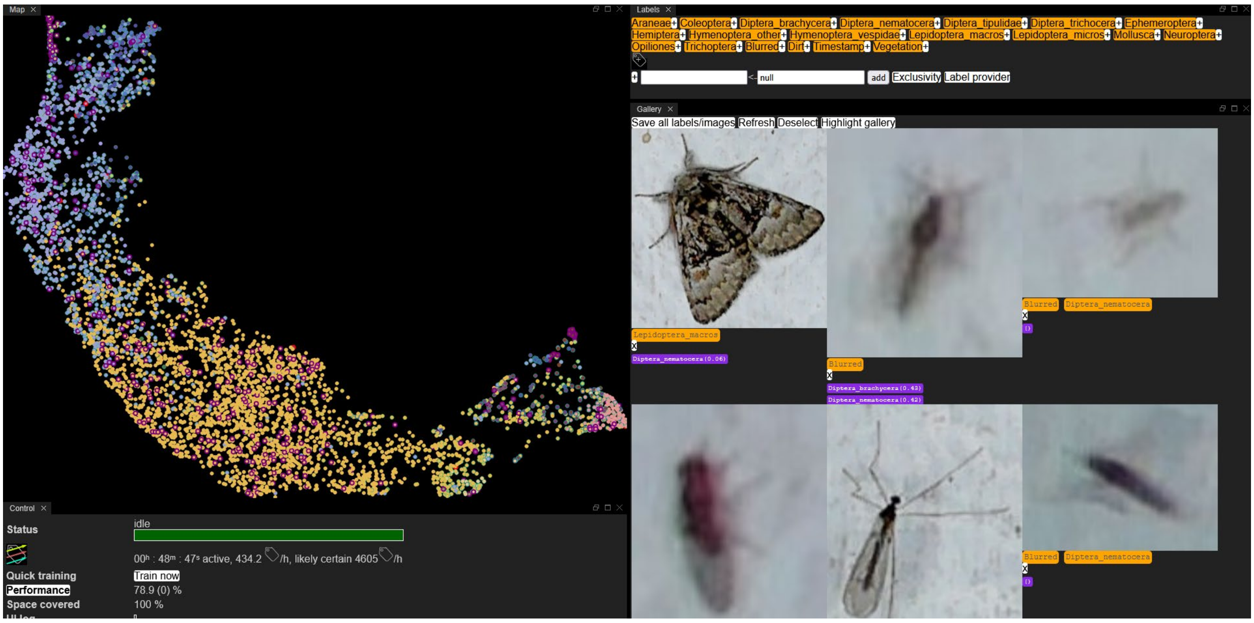


Figure 24. *Screen capture of the AnnFlux-application's user interface for a random selection of 5000 detections from data collected from AMI-traps in Denmark in 2025. The left panel shows clustering of detected insects, while the right panel allows a user to provide annotations. So far, the user has provided 432 annotations, allowing the model to infer provisional labels for all 5000 samples.*

5.7. Coordination and centralized processing enable sensor networks

In their guide on harmonizing biodiversity monitoring across scales, Silva del Pozo *et al.* (2023) highlighted potential roles for both centralized and localized data processing in transnational monitoring. Local data processing options are desirable for pilot partners (see recommendation 6). However, the ABMS has trialled centralized AI processing for sensor-based monitoring, successfully generating a huge number of provisional species records, as well as spatially and temporally replicable EBV indicators. Furthermore, when survey respondents were asked about minimum requirements to continue sensor-based monitoring, the most frequent response was the provision of AI processing and outputs (Fig. 25). Remarkably, this response was more popular than three which related directly to funding for monitoring, whether for existing locations, new locations or equipment. Clearly a centralized data management and processing solution is highly desirable and motivating for partners to continue sensor-based monitoring, and should receive investment in future. This means communication and collaboration with an ever-growing number of projects and initiatives in image and sound processing, ensuring that the community converges on solutions that are most accurate, but also most repeatable.

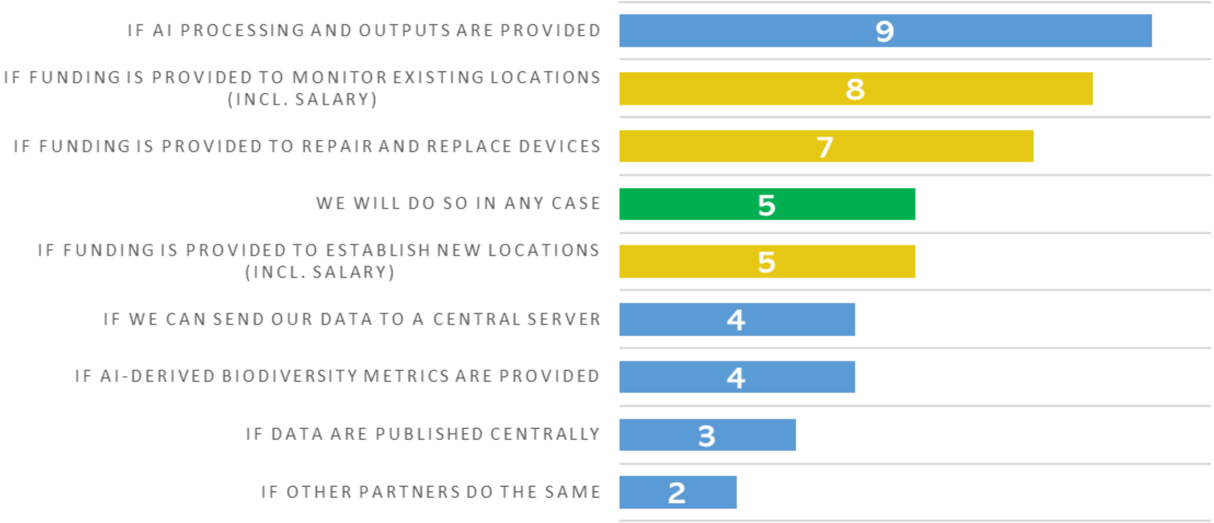


Figure 25. *Conditions that facilitate sensor-based monitoring at ABMS locations. Respondents were asked to select a set of minimum requirements to continue ABMS monitoring at their locations. Of the 14 respondents, none selected two other options: "We will not do so" or "If data rights are more clearly defined".*

However, the benefits of coordination go beyond centralized data management and processing. Silva del Pozo *et al.* (2023) highlight the relative roles of strict vs. general protocols, in that the former can lack flexibility, but generate data that are easy to process and compare. In the ABMS pilot we tried to develop relatively strict protocols - specifically protocols for site selection, device deployment and data transfer. Strict protocols for device deployment and data transfer were relatively easy to uphold; exceptions involved the recording frequency of AMI devices, for which Sweden preferred a 2 minute interval, and deployment of acoustic devices in Croatia, where recording parameters were adjusted and devices placed further from the ground. While relatively strict protocols for device deployment and data transfer provided consistent data flows from partners across Europe, a strict protocol for site selection was difficult to adhere to (see 3. Sensors constrain sampling and site selection). Collaboration and coordination to develop protocols was crucial in the pilot, and continue to be necessary going forwards.

From the perspective of the pilot, sensor-based monitoring should provide opportunities to coordinate efforts across Europe. Further work is needed to finalize protocols and produce automated monitoring handbooks, and funding is needed to maintain, repair, upgrade or replace devices being used to collect baseline data. Initiatives and tools are needed to strategically annotate sensor data, whether for model evaluation or to efficiently generate biological records (Fig. 23). Importantly, general AI and EBV pipelines need further development, including new classifiers for insects and amphibians in acoustic data. In particular, options should be made available for centralized data management and processing (Fig. 25). Furthermore, pipelines need to be replicable and well-disseminated, ensuring that localized processing is also an option (Silva del Pozo *et al.*, 2023). In general, there are opportunities to better exploit both new and existing streams of sensor data for ecological and policy insights. Meanwhile, continuity of biodiversity trends is best ensured by deploying novel and traditional monitoring methods in tandem.

Thanks to connections with existing and emerging automated monitoring initiatives, the ABMS pilot has built a strong foundation for a transnational biodiversity sensor network. While the pilot officially ends in

2025, five respondents stated that they will continue monitoring the same locations regardless of any further support (Fig. 25). While centralized funding has not been made available, further alignment of regional and national monitoring goals is the primary avenue to maintain joined-up transnational sensor networks. We hope that support for capacity building and networking around sensing of bats, birds and night-flying insects will provide both economic and collaborative incentives to continue monitoring at as many locations as possible, building on the strong foundation of the pilot.

References

AODHA, O.M., BALVANERA, S.M., DAMSTRA, E., COOKE, M., EICHINSKI, P., BROWNING, E., BARATAUD, M., BOUGHEY, K., COLES, R., GIACOMINI, G., G, M.C.M.S., OBRIST, M.K., PARSONS, S., SATTLER, T. & JONES, K.E. (2022) Towards a General Approach for Bat Echolocation Detection and Classification. bioRxiv. https://www.biorxiv.org/content/10.1101/2022.12.14.520490v1 [accessed 18 November 2025].

BATTERSBY, J. (2010) Guidelines for Surveillance and Monitoring of European Bats. EUROBATS Publication Series No. 5. UNEP / EUROBATS Secretariat, Bonn, Germany, 95 pp.

BERGLER, C., SMEELE, S.Q., TYNDEL, S.A., BARNHILL, A., ORTIZ, S.T., KALAN, A.K., CHENG, R.X., BRINKLØV, S., OSIECKA, A.N., TOUGAARD, J., JAKOBSEN, F., WAHLBERG, M., NÖTH, E., MAIER, A. & KLUMP, B.C. (2022) ANIMAL-SPOT enables animal-independent signal detection and classification using deep learning. Scientific Reports 12, 21966. Nature Publishing Group.

BJERGE, K., KARSTOFT, H. & HØYE, T.T. (2024) Towards edge processing of images from insect camera traps. bioRxiv. https://www.biorxiv.org/content/10.1101/2024.07.01.601488v1 [accessed 18 December 2024].

BOYD, R.J., POWNEY, G.D. & PESCOTT, O.L. (2023) We need to talk about nonprobability samples. Trends in Ecology & Evolution 38, 521–531.

FERNANDEZ-UGALDE, O., SCARPA, S., ORGIAZZI, A., PANAGOS, P., VAN LIEDEKERKE, M., MARECHAL, A. & JONES, A. (2022) LUCAS 2018 Soil Module. Presentation of dataset and results, EUR 31144 EN. Publications Office of the European Union, Luxembourg.

GHANI, B., KALKMAN, V.J., PLANQUÉ, B., VELLINGA, W.-P., GILL, L. & STOWELL, D. (2025) Impact of transfer learning methods and dataset characteristics on generalization in birdsong classification. Scientific Reports 15, 16273. Nature Publishing Group.

GILLESPIE, M.A.K., BJERGE, K., ALISON, J., GERLICH, H.S., HELSING-NIELSEN, F., MOUGEOT, G., SVENNING, A., WOGRAM, S.F.A. & HØYE, T.T. (2025) Automated insect monitoring with camera traps is transforming ecological understanding. EcoEvoRxiv. https://ecoevorxiv.org/repository/view/10577/ [accessed 10 December 2025].

HØYE, T.T., DALBY, L., MELLERUP, K., SVENNING, A. & PINOY, N. (2024) Monitoring of Invasive Alien Species: Main Findings of the First Year of the Biodiversa+ Pilot: ‘Monitoring of Invasive Alien Species with Image-Based Methods’. Biodiversa+.

JAIN, A., CUNHA, F., BUNSEN, M.J., CAÑAS, J.S., PASI, L., PINOY, N., HELSING, F., RUSSO, J., BOTHAM, M., SABOURIN, M., FRÉCHETTE, J., ANCTIL, A., LOPEZ, Y., NAVARRO, E., PIMENTEL, F.P., ET AL. (2025) Insect Identification in the Wild: The AMI Dataset. In Computer Vision – ECCV 2024 (eds A. LEONARDIS, E. RICCI, S. ROTH, O. RUSSAKOVSKY, T. SATTLER & G. VAROL), pp. 55–73. Springer Nature Switzerland, Cham.

KAHL, S., WOOD, C.M., EIBL, M. & KLINCK, H. (2021) BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics 61, 101236.

LÓPEZ-BAUCELLS, A., LÓPEZ-BOSCH, D., BLANCH, E., HAMIDOVIC, D., ORTEGA CASTAÑO, A., AIHARTZA, J., AULAGNIER, S., BARTONICKA, T., BAS, Y., BELLÈ, A., BRABANT, C., BRIGGS, P., BÜCS, S.L., CSORBA, G., DEBERNARDI, P., ET AL. (2025) Bat monitoring programmes and protocols for European bats. Zenodo. https://zenodo.org/records/15346727 [accessed 10 December 2025].

MENNILL, D.J. (2024) Field tests of small autonomous recording units: an evaluation of in-person versus automated point counts and a comparison of recording quality. Bioacoustics 33, 157–177. Taylor & Francis.

METCALF, O., ABRAHAMS, C., ASHINGTON, B., BRADFER-LAWRENCE, T., BROWNING, E., CARRUTHERS-JONES, J., DARBY, J., DICK, J., ELDRIDGE, A., ELLIOTT, D., HEATH, B., HOWDEN-LEACH, P., JOHNSTON, A., LEES, A., MEYER, C., ET AL. (2022) Good practice guidelines for long-term ecoacoustic monitoring in the UK. UK Acoustics Network.

PÉREZ-GRANADOS, C., FUNOSAS, D., MORANT, J., MARÍN GÓMEZ, O.H., MENDOZA, I., MOHEDANO-MUNOZ, M.A., SANTAMARÍA, E., BASTIANELLI, G., MÁRQUEZ-RODRÍGUEZ, A., BUDKA, M., BOTA, G., DE LA PEÑA-RUBIO, J.M., GARCÍA DE LA MORENA, E., SANTA-CRUZ, M., DE LA NAVA, P., ET AL. (2025) Optimization of passive acoustic bird surveys: a global assessment of BirdNET settings. Ibis Early access.

QUOSS, L., JUNKER, J. & WENDT, E. (2024) EuropaBON/EBV-Descriptions: EuropaBON EBVs List Descriptions. Zenodo.

ROEMER, C., HAQUART, A., LÓPEZ-BAUCELLS, A. & BESNARD, A. (2025) Current frontiers in the passive acoustic monitoring of bats. Methods in Ecology and Evolution 16, 2534–2544.

SCANFERLA, J., BRAMBILLA, M., BRAMBILLA, G., HILPOLD, A., MARCHETTI, A.E., PUFF, F., TAPPEINER, U. & ANDERLE, M. (2025) Determining species-specific thresholds to improve precision in passive acoustic monitoring. Ecological Informatics 91, 103423.

SILVA DEL POZO, M., BODY, G., RERIG, G. & BASILLE, M. (2023) Guide on harmonising biodiversity monitoring protocols across scales. Biodiversa+ report.

SVENNING, A., MOUGEOT, G., ALISON, J., CHEVALIER, D., MOLINA, N.L.C., ONG, S.-Q., BJERGE, K., CARRILLO, J., HOEYE, T.T. & GEISSMANN, Q. (2025) A General Method for Detection and Segmentation of Terrestrial Arthropods in Images. bioRxiv. https://www.biorxiv.org/content/10.1101/2025.04.08.647223v1 [accessed 15 April 2025].

WOOD, C.M. & KAHL, S. (2024) Guidelines for appropriate use of BirdNET scores and other detector outputs. Journal of Ornithology 165, 777–782.