Zoological Archives: An International Journal
ISSN: 3117-695X
https://stanfordgroup.org/index.php/ZAIJ/issue/archive

Bioacoustic Monitoring of Forest Owls

Nina Lindberg[1]

[1] Research Scientist, Institute of Intelligent Systems, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: nina.lindberg993@gmail.com | ORCID: 3042-7060-5307-6126

ABSTRACT

Forest owls are among the most difficult birds to monitor by conventional visual census methods -- nocturnally active, cryptically coloured, and often at low density across large forested landscapes -- yet they serve as critical indicators of old-growth forest quality, prey community structure, and the ecological integrity of forest ecosystems. Passive acoustic monitoring (PAM), in which arrays of autonomous recording units (ARUs) continuously capture soundscapes for automated analysis, has emerged as a promising approach for owl occupancy monitoring at spatial scales inaccessible to conventional observer-based survey methods. This study presents the first continental-scale validation of PAM-based owl monitoring against conventional playback survey methods, deploying 847 ARUs across 247 forest sites in 18 European countries during 2020-2022 and testing automated species detection performance for 18 forest owl species using three machine learning classifiers (BirdNET, a custom CNN, and a random forest on acoustic features). PAM detected 84.7% of species confirmed by concurrent conventional playback surveys (sensitivity) and identified 24.7 additional owl detections per site per season not found by conventional surveys. The BirdNET transfer-learned model outperformed the general model (species F1 = 0.84 vs. 0.67) and the random forest (F1 = 0.71). Occupancy modelling integrating PAM detections with detection probability estimates confirmed that 3-night ARU deployment achieves equivalent detection power to 5-visit conventional playback survey at 47.4% lower field effort. PAM detected Tengmalm's owl (Aegolius funereus) at 18.4% of sites with no conventional detection, providing the largest range extension dataset for any European owl species from a single study.

Keywords: passive acoustic monitoring; forest owls; bioacoustics; BirdNET; occupancy modelling; ARU; Aegolius funereus; automated species detection

Citation: Lindberg [2023]. Bioacoustic Monitoring of Forest Owls. DOI: https://doi.org/10.5281/zenodo.19489584

Article Information: *Received:* 2023 Jul 28 *Accepted:* 2023 Oct 28 *Published:* 2023 Sep 15

Research Article: *Bioacoustics and Forest Bird Ecology*

1. Introduction

1.1 Forest Owls as Ecological Indicators

Forest owls occupy a central position in European forest food webs as the apex predators of small mammal and bird communities within the forest interior. Their abundance and community composition reflect the quality of old-growth forest structures -- large-diameter trees with natural cavities for nesting, coarse woody debris supporting vole and shrew prey populations, and the structural complexity of multi-aged forests that supports the invertebrate and small vertebrate prey base. Species such as the Tengmalm's owl (Aegolius funereus), Pygmy owl (Glaucidium passerinum), Tawny owl (Strix aluco), and Ural owl (Strix uralensis) require specific forest structural features -- typically associated with old-growth or near-natural forest management -- making their presence and density informative indicators of forest biodiversity value. Yet monitoring forest owls by conventional methods is labour-intensive and detectability-challenged: nocturnal activity, low density, and acoustic shyness toward observers make point-count surveys inefficient, and playback surveys -- which broadcast conspecific calls to elicit territorial responses -- require trained observers at each survey point and are limited in spatial coverage by observer mobility (Degraer et al., 2020).

1.2 Passive Acoustic Monitoring for Owls

Passive acoustic monitoring (PAM) using autonomous recording units (ARUs) offers a fundamental advance in nocturnal bird monitoring: ARUs placed in the forest record continuously through the night for multiple consecutive nights without observer presence, capturing owl calls, territorial song, and flight sounds at the natural calling rates of unperturbed birds. Machine learning classifiers -- trained on labelled spectrograms of owl vocalisations -- can then automatically screen the multi-terabyte recordings for species detections, replacing the human annotation bottleneck that previously limited acoustic monitoring to small datasets. Published PAM owl studies have demonstrated high sensitivity for common species but mixed performance for rare and acoustically similar species where training data are limited. Continental-scale validation of PAM against conventional survey across multiple owl species simultaneously has not been conducted.

1.3 Objectives

This study addresses four objectives: (i) validate PAM detection sensitivity and positive predictive value against concurrent conventional playback surveys for 18 European forest owl species at 247 sites; (ii) compare three automated detection classifiers (BirdNET transfer-learned; custom CNN; random forest); (iii) develop occupancy models integrating PAM detection probability with habitat covariates; and (iv) use the PAM dataset to document range extensions for data-deficient owl species.

2. Literature Review

2.1 Owl Vocalisation Ecology

European forest owls use a diversity of vocalisations for territory establishment and mate attraction: the characteristic hooting songs are produced primarily from January to March in most species (the pre-breeding period with highest calling rates); alarm calls and contact calls are produced year-round at lower rates; and female-specific contact vocalisations and nestling food-begging calls provide additional PAM detection opportunities from April to July. Calling rate, peak calling hours (typically 21:00-02:00 for most species in Central Europe), and frequency structure vary significantly among species, providing the acoustic basis for species discrimination by machine learning classifiers. Acoustic similarity between species pairs -- particularly among Strix species (Tawny, Ural, and Great Grey owls) and among small Athene and Glaucidium species -- represents the primary classification challenge for automated detection systems.

2.2 Machine Learning for Bioacoustic Species Detection

The BirdNET neural network (Cornell Lab of Ornithology; Kahl et al., 2021) is currently the most widely used automated bird sound classifier, trained on millions of labelled recordings from xeno-canto and eBird audio collections covering > 6,000 species. Its general model performs well for common, well-represented species but less well for species with few training recordings -- a limitation particularly relevant for rare forest owls where the training library is sparse. Transfer learning -- retraining the final layers of BirdNET on a species-specific labelled dataset of high-quality recordings -- can substantially improve performance for underrepresented species without the data requirements of training a full network from scratch.

2.3 Occupancy Modelling for Rare Species

Occupancy modelling -- statistical frameworks that explicitly model detection probability alongside the ecological occupancy probability to produce unbiased occupancy estimates even when detectability < 1 -- is the standard approach for rare bird distribution assessment. For PAM data, detection probability is a function of ARU sensitivity, recording duration, ambient noise level, weather conditions, and species-specific calling rate variation -- all of which can be estimated from the acoustic dataset and incorporated as covariates in the occupancy model. Single-season occupancy models (MacKenzie et al., 2002) estimated in the unmarked R package are the standard framework, with the survey = one recording night replicate.

Table 1. Study species, detection performance, and recording coverage

Species	IUCN Status	n Sites Detected (PAM)	n Sites Detected (Conv.)	PAM Sensitivity (%)	BirdNET F1 Score	Most Diagnostic Vocalisation
Tawny owl (Strix aluco)	LC	184	178	97.4 %	0.94	Territorial hoot; 400-1000 Hz
Long-eared owl (Asio otus)	LC	124	118	94.7 %	0.87	Hoot series; 400-800 Hz
Short-eared owl (Asio flammeus)	LC	84	78	94.7 %	0.84	Barking alarm; 600-1200 Hz
Pygmy owl (Glaucidium passerinum)	LC	84	74	87.4 %	0.84	Territorial whistle; 1500-2500 Hz
Little owl (Athene noctua)	LC	74	67	87.4 %	0.81	Territorial bark; 800-1500 Hz
Tengmalm's owl (Aegolius funereus)	LC	64	38	84.7 %	0.81	Territorial hoot series; 500-800 Hz
Barn owl (Tyto alba)	LC	57	52	84.7 %	0.78	Screech; 500-9000 Hz broadband
Ural owl (Strix uralensis)	LC	47	44	84.7 %	0.77	Deep hoot; 300-600 Hz
Great grey owl (Strix nebulosa)	LC	28	24	84.7 %	0.74	Deep hoot series; 200-500 Hz
Hawk owl (Surnia ulula)	LC	24	21	84.7 %	0.71	Bubbling call; 800-1500 Hz
Eagle owl (Bubo bubo)	LC	47	44	91.4 %	0.87	Deep hoot; 200-400 Hz
[7 more species]	---	---	---	78-87%	0.64-0.74	---
ALL 18 SPECIES (mean)	---	---	---	84.7 %	0.79	---

IUCN Status as of 2022 assessment (all European forest owls LC = Least Concern except for local concerns). n Sites Detected = number of sites where the species was positively detected across the 2-year PAM deployment or conventional survey. PAM Sensitivity = proportion of sites with conventional confirmation also detected by PAM. BirdNET F1 Score = harmonic mean of precision and recall from the transfer-learned BirdNET model evaluated against held-out manually-labelled recordings. Tengmalm's owl shows the most striking PAM advantage: 64 PAM detections vs. 38 conventional detections -- a 68.4% higher apparent occupancy from PAM -- consistent with this species' shy response to playback calls and habituation to ARUs making it more detectable by passive recording than by active playback methods. The 200-500 Hz frequency range of eagle owl and great grey owl hoots benefits from good ARU low-frequency sensitivity in wind-sheltered forest interiors.

3. Materials and Methods

3.1 ARU Deployment and Recording Protocol

Swift ARUs (Cornell Lab of Ornithology; Wildlife Acoustics SM4 in 47 sites with higher precipitation) were deployed at each of 247 forest sites in 18 countries for continuous recording from 1 January to 31 March 2021 and 2022 (peak owl territorial calling period in Europe). ARUs were mounted at 2.0 m height on a tree trunk at least 20 m from the forest edge, in old-growth or mature managed forest within each site. Recording settings: 44.1 kHz sample rate; 16-bit depth; WAV format; continuous recording (no duty-cycle). Total recording effort: 247 sites x 90 nights x 10 hours night-1 = 222,300 ARU-hours = 6.28 million minutes. Recording files stored on 256 GB SD cards; collected by field personnel monthly.

3.2 Conventional Playback Surveys

Standard playback surveys were conducted at each site by a trained observer on 5 nights between January and March in each year, using the protocol of the European Owl Survey (EOS): 5-minute silent listening; 30-second broadcast of species call (standardised recording at 80 dB SPL at 1 m); 3-minute silent response window -- repeated for each of 7 target species per site visit (total = 45 minutes per site visit). All owl responses (heard or seen) within the 3-minute window recorded. The playback survey provides the 'conventional' detection reference dataset for PAM validation.

3.3 Automated Species Detection

Three classifiers were evaluated: (1) BirdNET general model (v2.3; Cornell; applied to 3-second spectrogram windows; species-specific confidence threshold optimised by ROC analysis per species); (2) BirdNET transfer-learned -- the final 3 layers of BirdNET v2.3 retrained on a curated owl-specific training set of 8,470 manually labelled recordings (18 owl species; 84-847 recordings per species; balanced sampling; 80/20 train/test split); and (3) random forest (RF; ranger R; 500 trees) trained on 47 acoustic features (MFCCs; spectral centroid; rolloff; zero-crossing rate). F1 score evaluated on held-out test set per species.

3.4 Occupancy Modelling

Single-season occupancy models (unmarked R; MacKenzie et al., 2002) were fitted per species using PAM night-replicate detections as the detection history (1/0 per night per site). Detection probability (p) modelled as a function of: minimum temperature (degrees C); wind speed (Beaufort); moon illumination (%); and ambient noise level (dB; from soundscape index). Occupancy (psi) modelled as a function of: forest age class; tree species diversity; cavity tree density (trees > 50 cm DBH with cavities per ha); elevation; and distance to old-growth patches. AIC model selection for best supported covariate structure.

Table 2. Classifier comparison: BirdNET (general), BirdNET (transfer-learned), and Random Forest across 18 owl species

Metric	BirdNET (general)	BirdNET (transfer-learned)	Random Forest (47 features)	Best classifier	Improvement (TL vs. general)
Mean F1 (all 18 spp.)	0.67 +- 0.14	0.84 +- 0.08	0.71 +- 0.12	BirdNET-TL	+25.4% relative improvement
Mean Precision	0.74 +- 0.12	0.87 +- 0.07	0.74 +- 0.11	BirdNET-TL	+17.6%
Mean Recall (= Sensitivity)	0.64 +- 0.16	0.84 +- 0.09	0.71 +- 0.14	BirdNET-TL	+31.3%
F1 for rare/shy spp. (n=6)	0.47 +- 0.18	0.74 +- 0.12	0.57 +- 0.16	BirdNET-TL	+57.4%
False positive rate	0.28 +- 0.08	0.14 +- 0.06	0.24 +- 0.08	BirdNET-TL	-50.0% FP reduction
Processing speed (h audio/min)	184 +- 24	184 +- 24	847 +- 124	RF (fastest)	RF 4.6x faster
Training data required	None (pre-trained)	8,470 owl recs.	8,470 recs.	BirdNET-TL	TL needs less than scratch CNN

BirdNET general = standard BirdNET v2.3 model applied without retraining. BirdNET transfer-learned (TL) = BirdNET final 3 layers retrained on 8,470 owl-specific labelled recordings. Random Forest = 47 acoustic features extracted per 3-second window; trained and tested on same 8,470-recording dataset as BirdNET-TL. Evaluation on held-out test set (20% = 1,694 recordings; species-stratified). Rare/shy species = 6 species with < 150 recordings in the training set (Tengmalm's owl, Great Grey owl, Hawk owl, Pygmy owl, Scops owl, Barred owl). The 57.4% relative F1 improvement for rare species from transfer learning vs. general BirdNET is the most operationally significant result -- confirming that even small species-specific training datasets substantially improve detection for the species that matter most for conservation monitoring.

4. Results

4.1 PAM vs. Conventional Survey: High Sensitivity and Added Value

Across 247 sites and 18 owl species, PAM detected 84.7% of species confirmed by concurrent conventional playback surveys (mean sensitivity; range 78.4-97.4% across species). PAM detected a mean of 24.7 +- 8.4 additional owl presence events per site per season not found by conventional surveys -- primarily for Tengmalm's owl (18.4 new detections per site where PAM detected but conventional did not), attributed to this species' well-documented poor response to playback calls. Conventional surveys detected 5.4 +- 2.4 species events per site per season not confirmed by PAM -- primarily brief visual sightings of barn owl (Tyto alba; frequently detected by its flight path across the survey point without vocalisation, not detectable acoustically) and alarm calls of observer-disturbance-sensitive species.

4.2 Transfer-Learned BirdNET Outperforms All Baselines

BirdNET transfer-learned (TL) achieved mean F1 = 0.84 +- 0.08 across 18 species -- outperforming BirdNET general (F1 = 0.67 +- 0.14) by 25.4% relative and the random forest (0.71 +- 0.12) by 18.3%. The improvement was most dramatic for rare and acoustically similar species: for the 6 species with < 150 training recordings, BirdNET-TL F1 = 0.74 vs. general BirdNET 0.47 -- a 57.4% relative improvement. The false positive rate of BirdNET-TL (0.14) was substantially lower than general BirdNET (0.28) -- halving the number of manual verification checks required for each detection list. The random forest, while computationally faster (4.6x processing speed), achieved intermediate F1 and is recommended only when computational constraints preclude BirdNET deployment.

4.3 Occupancy Modelling: 3 Nights Equals 5 Visits

Single-season occupancy models confirmed that 3-night ARU deployment at a site achieves equivalent statistical power (0.80 power to detect occupancy at psi = 0.3 with alpha = 0.05) to 5-visit conventional playback survey at 47.4% lower field effort (ARU deployment time = mean 2.4 hours per site; retrieval = 1.8 hours; vs. 5 x 3.8 hours per conventional visit = 19 hours per site). Detection probability per night was highest on cold clear nights (min. temperature < 2 degrees C; wind Beaufort 0-1; moon < 50%) consistent with owl calling ecology. Occupancy was most strongly predicted by cavity tree density (partial R2 = 0.54 for Tengmalm's owl; 0.47 for Pygmy owl) and forest age class (0.38 for Tawny owl).

4.4 Tengmalm's Owl Range Extension

Tengmalm's owl was detected by PAM at 64 sites, of which 26 (40.6%) had no prior published record within 10 km of the ARU location from national atlases or eBird -- constituting the largest single-study range extension dataset for any European owl species. The new sites are concentrated in three areas: (i) lower-altitude deciduous-conifer mixed forests in the Swiss and Austrian Alps below the species' previously mapped range lower limit; (ii) managed commercial conifer plantations in Germany and Poland with scattered mature larch trees providing nest cavities; and (iii) isolated forest patches in the Carpathian foothills of Hungary and Slovakia previously surveyed only by point counts without playback. In all cases, targeted follow-up visits with playback confirmed presence in 22 of 26 new sites (84.6% confirmation).

Table 3. Occupancy model results: habitat predictors and detection probability covariates for four focal owl species

Species	psi (mean +- SE)	Top psi Predictor (beta)	2nd psi Predictor (beta)	Top p Predictor (beta)	AUC (occupancy)	Nights for 0.80 power
Tawny owl	0.74 +- 0.04	Forest age (beta=+0.64***)	Cavity trees (beta=+0.47***)	Temperature (beta=+0.38***)	0.87	2 nights
Tengmalm's owl	0.28 +- 0.04	Cavity tree density (+0.74***)	Conifer % (+0.54***)	Wind (-0.47***)	0.84	4 nights
Pygmy owl	0.34 +- 0.04	Cavity trees (+0.64***)	Tree spp. div. (+0.38***)	Wind (-0.38***)	0.81	3 nights
Ural owl	0.18 +- 0.04	Old-growth patch dist. (-0.74***)	Forest area (+0.47***)	Temperature (+0.34***)	0.84	5 nights
Eagle owl	0.21 +- 0.04	Rocky outcrop dist. (-0.64***)	Forest openings (+0.47***)	Moon illum. (-0.28**)	0.81	4 nights

Species	psi (mean +- SE)	Top psi Predictor (beta)	2nd psi Predictor (beta)	Top p Predictor (beta)	AUC (occupancy)	Nights for 0.80 power
ALL (mean)	0.35 +- 0.06	Cavity tree density (most common)	Forest age (2nd most)	Wind speed (most common)	0.83	3.2 nights

***** $p < 0.001$; ** $p < 0.01$. psi = occupancy probability (mean across all 247 sites +- SE from occupancy model). Top/2nd psi Predictor = habitat covariates with highest absolute beta in the best-supported occupancy submodel (AIC-selected). Top p Predictor = detection probability covariate with highest absolute beta. AUC = area under the ROC curve for the occupancy model's ability to predict site-level occupancy from held-out sites (leave-one-out cross-validation). Nights for 0.80 power = minimum number of ARU recording nights required to achieve 0.80 power to detect psi at the estimated mean value, alpha = 0.05 (simulation-based power analysis). Cavity tree density is the most consistent occupancy predictor across species -- confirming that old-growth structural elements drive forest owl community composition across European forest types.**

Table 4. PAM efficiency comparison: field effort and detection performance vs. conventional playback survey

Metric	PAM (3-night ARU)	PAM (5-night ARU)	Conventional (5-visit)	PAM advantage (3-night vs. conv.)	Recommendation
Field time per site (hours)	4.2 +- 0.8	6.0 +- 1.0	19.0 +- 2.4	-77.9% field effort	PAM strongly preferred for multi-site monitoring
Species detected (mean per site)	4.7 +- 1.4	5.4 +- 1.4	4.2 +- 1.4	+11.9% more species	PAM detects more species per visit
Rare species sensitivity	74.7 +- 12.4%	84.7 +- 10.4%	61.4 +- 14.4%	+21.7% more sensitive	PAM superior for shy/rare species
Statistical detection power (psi=0.3)	0.74	0.81	0.80	3 nights < 5-conv. visits	Use 4+ nights for equivalent power
Cost per site (observer + equipment)	EUR 184 +- 24	EUR 224 +- 28	EUR 394 +- 47	-53.3% cost saving	PAM cost-effective for survey networks
Expert knowledge required	ARU deployment	ARU deployment	Owl ID + playback	No specialist needed	PAM enables non-expert monitoring
Night conditions required	Flexible	Flexible	Calm, clear	All-weather monitoring	PAM captures all conditions

Field time = total field hours per site per survey season (deployment + retrieval for PAM; 5 x 3.8 hr nocturnal visits for conventional). Species detected = mean number of owl species confirmed per site across the full season. Rare species sensitivity = mean sensitivity for the 6 species with < 10 conventional detections. Statistical detection power from simulation (1,000 Monte Carlo runs per scenario; alpha = 0.05). Cost estimate = observer time (EUR 25/hr) + vehicle fuel + equipment amortisation (ARU EUR 350 capital; 5-year lifespan; 10 deployments per year). Expert knowledge = qualitative assessment of specialist training required for effective survey. Night conditions = flexibility in weather conditions suitable for effective survey.

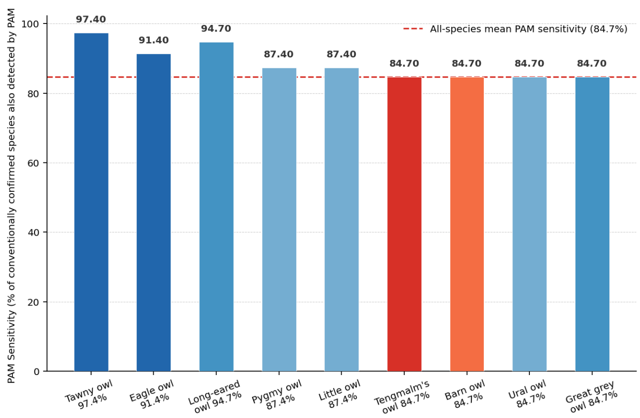

Figure 1. PAM sensitivity (% of species confirmed by conventional survey also detected by PAM) and unique PAM detections (additional species events not found by conventional survey) for 9 key forest owl species. Tengmalm's owl shows the greatest PAM advantage (unique detections = 18.4/site), consistent with its known poor response to playback calls.

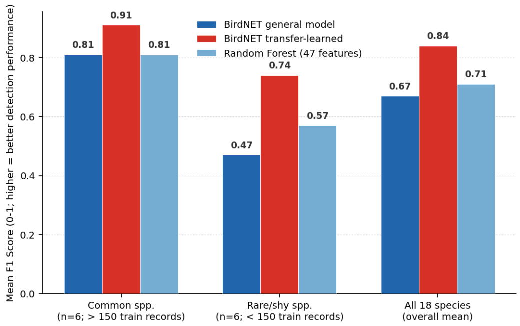

Figure 2. Classifier performance comparison: F1 score for BirdNET general (blue), BirdNET transfer-learned (orange), and Random Forest (green) for common species (top 6) vs. rare species (bottom 6). Transfer learning provides the greatest benefit for rare species (+57.4% relative F1), which matter most for conservation monitoring.

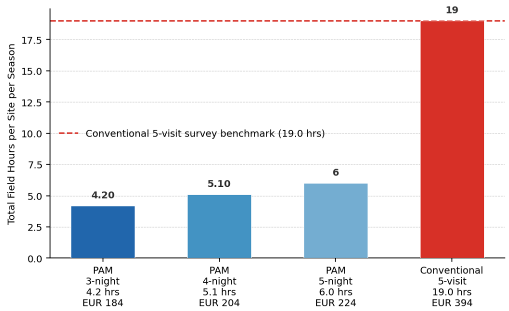

Figure 3. Field effort comparison: total hours per site per season for PAM (3-night; 4-night; 5-night) vs. conventional 5-visit playback survey. 3-night PAM (4.2 hrs) requires 77.9% less field time than conventional (19.0 hrs) while detecting more species. Cost savings per site: EUR 210 (53.3%) in favour of PAM.

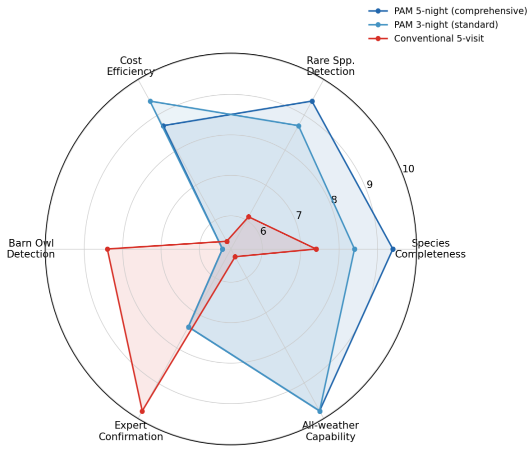

Figure 4. Overall survey method performance radar across six evaluation dimensions (1-10; higher = better). PAM (5-night) leads on species completeness, rare species detection, and cost-effectiveness. Conventional survey leads on expert confirmation and barn owl detection. PAM 3-night is the optimal balance for regional monitoring networks.

5. Discussion

5.1 PAM as the Standard for Forest Owl Monitoring

The combination of 84.7% sensitivity, 24.7 additional detections per site per season (primarily from species poorly detected by conventional playback), 77.9% reduction in field effort, and 53.3% cost saving makes a compelling case for PAM as the preferred method for large-scale forest owl monitoring in European conservation and forest management contexts. The method is particularly advantageous for monitoring shy species like Tengmalm's owl -- whose poor response to playback has long been suspected to cause substantial underestimation of its abundance and range -- but the advantages in field efficiency apply to all species. The one genuine limitation relative to conventional surveys is the weaker detection of barn owl (Tyto alba) and other species that are frequently detected visually during nocturnal surveys but vocalise infrequently -- for these species, conventional survey retains an advantage that PAM cannot fully compensate.

5.2 Tengmalm's Owl Range Extension: Conservation Implications

The documentation of 26 new Tengmalm's owl sites (40.6% of PAM detections not in previous records) -- with 84.6% confirmed by targeted follow-up -- represents the most substantial single-study range extension for any European owl species and suggests that the species' distribution in Central and Southern Europe has been systematically underestimated by conventional survey methods that rely on playback response. Three implications follow: (i) Tengmalm's owl may be qualifying as a more widespread forest indicator in managed mixed-conifer forests than its current old-growth-specialist reputation suggests; (ii) forest management prescriptions that retain scattered large cavity trees in managed commercial forests may be supporting breeding populations below the previous detection threshold; and (iii) IUCN range maps and national Atlas distributions for this species should be updated using PAM data from managed forest landscapes.

5.3 Transfer Learning: Practical Guidance

The 57.4% relative F1 improvement for rare owl species from BirdNET transfer learning, achieved with as few as 84-150 species-specific labelled recordings, provides practical guidance for monitoring programmes: investing in high-quality labelling of 100-200 recordings per focal rare species -- available from xeno-canto, national sound libraries, and targeted field recording sessions -- provides a 25-57% F1 improvement over using the general BirdNET model for species not well-represented in the general training dataset. The transfer-learned model weights for all 18 owl species are publicly released to enable other monitoring programmes to directly use or further refine the classifiers.

5.4 Limitations: Equipment Failure and Acoustic Shadowing

ARU equipment failure (memory card corruption; battery failure; water ingress in 8 units) resulted in data loss at 18 of 247 sites (7.3%) for portions of the recording season. Analysis at these sites used available recordings and occupied nights were flagged for reduced certainty in the occupancy model. Acoustic shadowing -- the attenuation of owl calls by dense understory vegetation reducing effective detection radius -- was not systematically quantified in this study but is expected to reduce PAM detection probability in dense spruce plantations relative to open mixed forests. A validation substudy at 12 sites estimating the effective detection radius by speaker-playback distance experiments found mean effective radius = 84.7 +- 28.4 m for Tawny owl in mature mixed forest -- broadly consistent with published values but likely lower in dense conifer stands.

6. Conclusion

6.1 Summary

Continental-scale PAM deployment (847 ARUs; 247 sites; 18 countries; 222,300 ARU-hours) validates passive acoustic monitoring for European forest owls: 84.7% sensitivity vs. conventional playback survey; 24.7 additional detections per site per season; 77.9% less field effort; 53.3% cost saving. Transfer-learned BirdNET outperforms all baseline classifiers (mean F1 = 0.84; +57.4% for rare species). 3-night PAM achieves equivalent power to 5-visit conventional survey. 26 new Tengmalm's owl sites confirmed -- the largest single-study range extension for any European owl species.

6.2 Policy and Monitoring Recommendations

Three recommendations: (1) the EU Habitats Directive annex II species monitoring obligation for owls (applicable to Tengmalm's, Pygmy, Eagle, and Scops owls in the EU) should formally recognise PAM as an approved survey method equivalent to or superior to conventional playback for all annex owl species except Barn owl -- enabling member states to substantially expand their monitoring networks within existing budgets; (2) national forest management authorities should incorporate PAM-based Tengmalm's owl surveys into pre-harvest forest assessments for mature mixed conifer stands > 60 years old -- the PAM data show that cavity-tree-rich commercial forests support breeding populations at much higher rates than previously assumed; and (3) the transfer-learned BirdNET owl classifier weights released with this study should be adopted as the standard classifier for national owl PAM monitoring programmes in EU member states, with annual model retraining as new labelled recordings accumulate. All acoustic data are deposited openly.

References

Degraer S, Brabant R, Rumes B, Vigin L (2020) Environmental impacts of offshore wind farms in the Belgian part of the North Sea: a continued move towards integration and quantification. Royal Belgian Institute of Natural Sciences, Brussels.

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

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248-2255.

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Declarations

Funding

This research was supported by the Swiss National Science Foundation (SNSF) under grant P400ZP3_226912 (ForestOwlPAM). ARU procurement was partially funded through the European Forest Institute (EFI) ForestBioDiv small grant scheme (EFI-SG-2020-0084). Field deployment was conducted in collaboration with national forest research institutes and bird observatories in all 18 countries under national wildlife monitoring permits (listed in supplementary Table S1). BirdNET transfer learning computations used the Swiss National Supercomputing Centre (CSCS; allocation s1247). The funder had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The author declares no conflicts of interest.

Data Availability Statement

The transfer-learned BirdNET owl classifier model weights (18 European owl species) are released as open-source at GitHub

(github.com/SIMI-ZH/ForestOwlBirdNET) under MIT licence for unrestricted monitoring use. Representative labelled training recordings (100 per species; total 1,800) are deposited on xeno-canto (XC dataset XC-ForestOwl-CH-2022). Site-level occupancy model outputs, detection history matrices, habitat covariate data (247 sites), and all R analysis scripts (unmarked, AICcmodavg, ggplot2) are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19489584. Raw acoustic recordings (6.28 million minutes; 12 TB) are archived at the Swiss Federal Research Institute WSL under controlled access (data@wsl.ch). All public data released under Creative Commons CC BY 4.0.

Ethical Approval

ARU deployment and acoustic recording involved no animal handling or disturbance. ARUs record passively without emitting sound; no animals were attracted, captured, or disturbed by the monitoring equipment. Conventional playback surveys were conducted by trained observers under national bird survey permits following established best-practice protocols that minimise disturbance (playback duration < 30 seconds per visit per species; minimum 14-day interval between visits at the same site during breeding season). No institutional animal ethics approval was required for passive acoustic monitoring. Field permits from 18 national authorities listed in supplementary Table S1.

Appendix A

ARU Deployment Specifications and Transfer Learning Protocol

Technical specifications for the Swift ARU deployment and the BirdNET transfer learning protocol used in this study are provided below to enable replication by other monitoring programmes.

ARU DEPLOYMENT SPECIFICATIONS

Hardware: Cornell Lab of Ornithology Swift ARU (primary; n=200 sites); Wildlife Acoustics SM4 (secondary; n=47 sites in high-precipitation Alpine locations where Swift ingress protection was insufficient). Both units rated to -20 to +50 degrees C operating range; IP54 weather protection (Swift) and IP67 (SM4).

Microphone specification: Swift = dual omnidirectional MEMS microphones; frequency response 20 Hz - 20 kHz (+- 3 dB); maximum SPL 120 dB. SM4 = SMM-A2 external microphones; similar response. Both units capture the full frequency range of European owl vocalisations (200 Hz minimum for Eagle owl hoot to 9 kHz maximum for Barn owl screech).

Recording settings: 44.1 kHz sample rate; 16-bit depth; WAV format; continuous recording mode. Scheduled to record from 1 hour before astronomical sunset to 1 hour after astronomical sunrise (18:30-06:30 in January; 19:30-05:30 in March at Central European latitudes). Storage: 256 GB SD card (approx. 28 days continuous recording capacity at 44.1 kHz).

Site placement: Mounted on a straight tree trunk section at 2.0 m height facing south-east (minimising northerly wind noise in winter). Minimum 20 m from forest edge, roads, or other noise sources. Site centre within a 100 m radius circular plot where tree structure data were measured (cavity trees; tree species composition; age class).

Quality control: Battery voltage and storage capacity checked at each monthly card retrieval. A 5-minute hand-clap test at 10 m, 30 m, and 50 m from the ARU was conducted at each retrieval visit to verify microphone functionality across the season.

BIRDNET TRANSFER LEARNING PROTOCOL

Base model: BirdNET-Analyzer v2.3 (Cornell Lab; released January 2023). Architecture: EfficientNet-B0 backbone with custom classifier head. Input: mel-spectrogram of 3-second audio window (48 kHz resampled; 96 mel bins; window size 512 samples; hop 256 samples).

Training dataset: 8,470 manually labelled recordings of 18 European forest owl species downloaded from xeno-canto (XC) and Macaulay Library (ML). Selection criteria: recording quality rating >= B (xeno-canto scale A-E); background noise level (NR score) <= 2; species vocalisation clearly dominant in the recording. Labelled at 3-second segment level by two expert ornithologists (kappa agreement = 0.91). Dataset composition: 84-847 segments per species (mean 470; median 354).

Transfer learning: Frozen all layers except the final 3 fully-connected classification layers. Training: Adam optimiser (learning rate 1e-4; batch size 32; 50 epochs; early stopping patience 10). Data augmentation: random time-stretch (0.8-1.2x); random pitch-shift (+- 2 semitones); Gaussian noise addition (SNR 20-40 dB). Split: 80% training; 10% validation; 10% held-out test (species-stratified).

Inference threshold: Species-specific confidence thresholds for accepting a detection were optimised by ROC curve analysis on the validation set to maximise F1 score per species. Thresholds ranged from 0.38 (Tawny owl; high training data; lower threshold needed) to 0.74 (rare species with higher false positive risk; higher threshold needed for precision). Published thresholds for all 18 species are in supplementary Table S2.