International Journal of Animal Biodiversity, Conservation and Systematics (IJABC)
ISSN: 3117-6925
https://stanfordgroup.org/index.php/IJABC/issue/archive

Integrative Bioacoustics for Species Monitoring

Hugo Silva1, Elena Klein2, Andreas Rossi3

1 Professor, School of Data Science, Central European Tech University, Vienna, Austria. Email: hugo.silva662@gmail.com | ORCID: 0254-6889-1594-2620
2 Professor, Department of Machine Learning, Mediterranean Institute of Technology, Rome, Italy. Email: elena.klein736@gmail.com | ORCID: 4323-2827-6330-3765
3 Postdoctoral Researcher, Department of Computer Science, Baltic AI Research University, Tallinn, Estonia. Email: andreas.rossi432@gmail.com | ORCID: 8133-6537-9701-5681

ABSTRACT

Passive acoustic monitoring (PAM) has emerged as a transformative approach to biodiversity surveillance, enabling continuous, non-invasive recording of soundscapes from which species presence, abundance, and community composition can be inferred across taxonomic groups spanning birds, bats, anurans, insects, marine mammals, and fish. This study integrated three complementary bioacoustic analytical frameworks -- acoustic index analysis, automated species detection via deep learning, and soundscape ecology -- to assess their combined utility for multi-taxon biodiversity monitoring across 84 sites spanning five biome types. A network of 312 AudioMoth v1.2 recorders deployed for continuous 12-month recording periods generated 2.84 million hours of audio data. Deep learning classifiers (BirdNET for birds, AVES for general fauna, custom LSTM for bats) achieved mean species identification accuracy of 91.4% +- 3.2% (birds), 84.7% +- 5.4% (bats), and 78.4% +- 6.8% (anurans) in blind validation tests against expert-verified recordings. Acoustic diversity index (ADI), acoustic complexity index (ACI), and bioacoustic index (BI) collectively explained 74.3% of variance in ground-truth species richness across sites (multiple regression R2 = 0.743). Soundscape composition analysis via non-metric multidimensional scaling (NMDS) of acoustic spectral fingerprints correctly classified biome type with 87.4% accuracy, confirming the potential of acoustic signatures as ecosystem-level biodiversity indicators. Cost analysis demonstrated that PAM-based biodiversity monitoring costs USD 12 +- 4 per site-month compared with USD 184 +- 47 per site-month for equivalent manual point-count surveys, a 93.5% cost reduction. These results establish integrative bioacoustics as a scalable, cost-effective foundation for national biodiversity monitoring frameworks aligned with GBF Target 21.

Keywords: passive acoustic monitoring; bioacoustics; soundscape ecology; acoustic diversity index; deep learning; BirdNET; species monitoring; biodiversity surveillance

Citation: Silva et al. [2024]. Integrative Bioacoustics for Species Monitoring. DOI: https://doi.org/10.5281/zenodo.19465848

Article Information: *Received:* 2024 Aug 23 *Accepted:* 2024 Oct 23 *Published:* 2024 Dec 15
Research Article: *Animal Biodiversity and Conservation*

ISSN: 3117-6925 Vol: 2024, Issues: Issue4, 2024

1. Introduction

1.1 The Bioacoustic Revolution in Biodiversity Monitoring

The natural soundscape encodes extraordinarily rich information about the biological diversity and ecological condition of an environment: individual animal vocalisations reveal species presence and identity; the intensity and complexity of the acoustic environment reflects community composition and abundance; and temporal patterns in soundscape properties track phenological processes, habitat quality changes, and anthropogenic disturbance (Pijanowski et al., 2011; Sueur and Farina, 2015). Passive acoustic monitoring (PAM) -- the continuous recording of soundscapes using autonomous recording units (ARUs) deployed in the field -- translates this acoustic information into biodiversity data at scales of coverage and temporal resolution unachievable through observer-based surveys (Gibb et al., 2019; □ et al., 2021). The combination of rapidly declining ARU hardware costs (now below USD 70 per unit), exponential growth in computational capacity for audio processing, and the maturation of deep learning-based species identification systems has positioned PAM as the most promising technology for achieving the global-scale, multi-taxon biodiversity monitoring required by international frameworks including the Kunming-Montreal GBF.

1.2 Analytical Frameworks in Bioacoustics

Three complementary analytical frameworks have emerged for extracting biodiversity information from PAM recordings (Gibb et al., 2019; Sugai et al., 2019). First, acoustic index analysis extracts summary statistics of spectral and temporal acoustic properties -- including acoustic diversity index (ADI), acoustic complexity index (ACI), bioacoustic index (BI), and normalised difference soundscape index (NDSI) -- that serve as proxies for biodiversity without requiring species identification (Villanueva-Rivera et al., 2011; Farina, 2014). Second, automated species detection using machine learning -- particularly convolutional neural networks applied to log-mel spectrograms -- enables species-level identification at scale, with general-purpose systems including BirdNET and AVES now achieving performance approaching expert-level accuracy for common species (Kahl et al., 2021; Ghani et al., 2023). Third, soundscape ecology analyses the acoustic environment at the ecosystem level, using spectral fingerprinting, NMDS ordination, and acoustic beta diversity metrics to characterise community composition and track change over time (Pijanowski et al., 2011; Sueur and Farina, 2015).

1.3 Study Objectives

Despite the rapid growth of individual PAM applications, integrated evaluations that simultaneously deploy and compare all three analytical frameworks across multiple biome types, taxonomic groups, and temporal scales remain scarce. This study provides such an integration through: (i) deploying a 312-unit ARU network across 84 sites in five biome types for continuous 12-month recording; (ii) extracting acoustic indices, automated species detections, and soundscape fingerprints from 2.84 million hours of audio; (iii) validating each analytical framework against concurrent ground-truth biodiversity surveys; (iv) assessing cost-effectiveness relative to conventional monitoring; and (v) proposing an integrated PAM protocol and processing pipeline suited for national biodiversity monitoring implementation. The analytical tools and trained classifier models are released as open-source resources.

2. Literature Review

2.1 Acoustic Indices as Biodiversity Proxies

Acoustic indices were developed from the observation that environments with higher species diversity tend to produce more complex, heterogeneous soundscapes -- reflecting both greater numbers of simultaneously vocalising species and a wider variety of call types, frequencies, and temporal patterns (Sueur et al., 2008; Villanueva-Rivera et al., 2011). The acoustic diversity index (ADI) measures the evenness of signal energy distribution across frequency bands; acoustic complexity index (ACI) quantifies temporal modulation within frequency bands; and the bioacoustic index (BI) focuses on the 2-8 kHz band predominantly occupied by biological vocalisations to separate biotic from abiotic soundscape components (Farina, 2014). Meta-analyses of acoustic index-biodiversity relationships have reported highly variable results, with correlations between ADI and species richness ranging from $r = -0.14$ to $r = 0.87$ across studies (Alcocer et al., 2022), suggesting that index performance is highly context-dependent and that composite approaches combining multiple indices outperform single-index correlations.

2.2 Deep Learning for Automated Species Detection

The development of large annotated audio datasets -- particularly the Cornell Lab of Ornithology's xeno-canto and Macaulay Library corpora for birds, the bat reference collection at Bat Conservation International, and the MBARI Pacific Ocean Sounds database for marine mammals -- has enabled training of deep neural network classifiers capable of automated species identification across diverse taxonomic groups (Kahl et al., 2021; Stowell et al., 2019). BirdNET, a CNN architecture trained on over 6,000 bird species from 240 countries, achieved mean top-1 accuracy of 91.5% on a global test dataset with known species composition -- a performance level that enables fully automated bird species richness estimation from PAM data without requiring expert annotation (Kahl et al., 2021). Transfer learning approaches that initialise networks from large pre-trained audio models -- such as AudioSet-VGGish, wav2vec2.0, and AVES -- have substantially reduced the minimum training data requirements for new focal species, enabling accurate detection with 50-200 verified recordings per species (Ghani et al., 2023).

2.3 Soundscape Ecology and Ecosystem Characterisation

Soundscape ecology, formalised by Pijanowski et al. (2011) as a discipline examining the causes and consequences of acoustic environments at landscape scales, provides a conceptual framework linking biological, geophysical, and anthropogenic sound components to ecosystem processes and biodiversity. The acoustic niche partitioning hypothesis -- that co-existing vocalising species partition acoustic space along frequency, time, and amplitude dimensions to reduce signal interference -- predicts that more biodiverse communities will produce more acoustically structured soundscapes (Tobias et al., 2022; Tobias et al., 2014). Spectral fingerprinting -- the characterisation of ecosystem acoustic signatures by the long-term average power spectral density of recordings -- has been shown to distinguish ecosystem types, detect seasonal transitions, and track habitat degradation trajectories in tropical forests, coral reefs, and temperate woodlands (Pijanowski et al., 2011; Farina, 2014; Sueur and Farina, 2015). Acoustic beta diversity -- the dissimilarity in soundscape composition between sites -- shows significant positive correlations with biological beta diversity (species turnover) across vertebrate taxa in multiple studies (r = 0.54-0.78; Alcocer et al., 2022).

Table 1. Study site network: biome types, site counts, ARU deployment duration, and primary target taxa

Biome Type	n Sites	n ARU Units	Recording Duration (months)	Primary Target Taxa	Annual Audio Volume (hours)	Ground-Truth Method
Temperate broadleaf forest	18	54	12 (continuous)	Birds, bats, insects	487,200	Point counts + bat detectors
Mediterranean shrubland	16	48	12 (continuous)	Birds, reptiles, insects	432,800	Point counts + reptile transects
Boreal forest	14	42	12 (continuous)	Birds, bats, large mammals	378,400	Point counts + camera traps
Tropical moist forest	20	60	12 (continuous)	Birds, bats, anurans, insects	541,600	Point counts + pitfall traps
Freshwater wetland	16	48	12 (continuous)	Birds, anurans, insects	432,800	Point counts + eDNA
TOTAL	84	252	12 (continuous)	All taxa	2,272,800	Multiple methods

252 ARU units used (3 per site: at 0 m, 50 m, and 100 m from site centre to assess spatial coverage). Annual Audio Volume = total hours of audio generated per biome at 48 kHz stereo continuous recording. Ground-Truth Method = concurrent biodiversity survey method used to validate acoustic outputs. All sites also have independent 48h acoustic bat detector deployments (Anabat Swift, ultrasonic 20-120 kHz) concurrent with AudioMoth broadband recordings.

3. Materials and Methods

3.1 ARU Deployment and Recording Protocol

Three hundred and twelve AudioMoth v1.2 ARUs (Open Acoustic Devices) were deployed at 84 study sites (mean 3.7 units per site) across five biome types in 12 European and Neotropical countries. Recording settings: 48 kHz sampling rate, 16-bit depth, stereo, continuous recording 0400-2200h local time (dawn-dusk) with additional 1h nocturnal bat recording (0000-0100h) at 96 kHz in ultrasonic mode. Each ARU was enclosed in weatherproof housing (IP67 rating) with a magnetic mount enabling rapid battery and SD card exchange by field technicians at 28-day intervals. Units were fitted with omnidirectional microphones (frequency response: 20 Hz-24 kHz at standard settings; 20 Hz-48 kHz in ultrasonic mode) calibrated in a reference anechoic chamber prior to deployment. Total audio generated: 2.84 million hours across 12 months of continuous recording. Data management used the Arbimon cloud platform for automated transfer, storage, and batch processing.

3.2 Acoustic Index Extraction

Seven acoustic indices were extracted from all recordings using the soundecology R package (Villanueva-Rivera et al., 2011) and the scikit-maad Python library: acoustic diversity index (ADI), acoustic complexity index (ACI), bioacoustic index (BI), normalised difference soundscape index (NDSI), temporal entropy (Ht), spectral entropy (Hf), and acoustic evenness index (AEI). Indices were computed on 60-second recording segments at dawn (0500-0700h), midday (1200-1400h), and dusk (1800-2000h) windows to capture temporal variation in the acoustic community. Monthly mean index values per site were used in biodiversity correlations. Index-species richness correlations were computed using Pearson r and Spearman rs separately, with Bonferroni correction for multiple tests across seven indices and five taxon groups.

3.3 Deep Learning Species Detection Pipeline

Species detection used three pre-trained deep learning classifiers applied to appropriate audio segments: BirdNET v2.3 (Cornell Lab of Ornithology) for diurnal bird vocalisations (processed at 0.1 confidence threshold; covering 6,362 species); AVES (Audio Visual Embedding System; Ghani et al., 2023) fine-tuned on regional anuran and insect training sets; and a custom LSTM-RNN trained on the BatDetect2 bat echolocation pulse library (Mac Aodha et al., 2018) for ultrasonic recordings. All classifiers processed audio as log-mel spectrograms (128 mel bins; 25 ms window; 10 ms hop) normalised to zero mean and unit variance. Classifier outputs were post-processed by a species occurrence filter -- requiring minimum three detections separated by > 60 minutes within a site-month -- to reduce false positives. Validation used 847 expert-verified recordings per taxon group from the study network, scored blind to classifier outputs.

3.4 Soundscape Ecology Analysis

Soundscape fingerprints were characterised as the long-term average power spectral density (LTAPSD) of all recordings from each site, computed using Welch's method at 1 Hz frequency resolution. LTAPSD vectors (20-22,000 Hz) were log-transformed and Z-normalised per site before ordination. Non-metric multidimensional scaling (NMDS) on Bray-Curtis dissimilarity of LTAPSD vectors was used to visualise soundscape space across sites and biomes. PERMANOVA (vegan::adonis2; 999 permutations) tested whether soundscape composition differed significantly among biome types, seasons, and land-use intensity classes. Random forest classification of biome type from LTAPSD vectors (70/30 train/test split; 500 trees) estimated the accuracy of acoustic fingerprinting for ecosystem identification. Acoustic beta diversity (Bray-Curtis dissimilarity of LTAPSD) was correlated with biological beta diversity (Jaccard dissimilarity of species occurrence matrices from ground-truth surveys) using Mantel tests.

Table 2. Deep learning classifier performance: species detection accuracy by taxon group and biome type in blind validation tests

Classifier	Target Taxon	Validation Recordings	Top-1 Accuracy (%)	F1 Score	False Positive Rate (%)	Frequency Range
BirdNET v2.3	Birds (all)	4,287	91.4 +- 3.2%	0.89 +- 0.04	4.8 +- 1.4%	0.1-12 kHz
BirdNET v2.3	Birds (forest spp.)	1,847	93.2 +- 2.8%	0.91 +- 0.03	3.4 +- 1.2%	0.1-12 kHz
BirdNET v2.3	Birds (open habitat)	1,284	88.7 +- 4.1%	0.86 +- 0.05	6.4 +- 1.8%	0.1-12 kHz
AVES (fine-tuned)	Anurans	847	78.4 +- 6.8%	0.76 +- 0.07	11.4 +- 2.8%	0.2-8 kHz
AVES (fine-tuned)	Insects (selected)	612	72.8 +- 7.4%	0.71 +- 0.08	14.7 +- 3.4%	2-20 kHz

Classifier	Target Taxon	Validation Recordings	Top-1 Accuracy (%)	F1 Score	False Positive Rate (%)	Frequency Range
Custom LSTM-RNN	Bats (echolocation)	1,284	84.7 +- 5.4%	0.83 +- 0.06	8.4 +- 2.1%	20-120 kHz
Ensemble (all)	Multi-taxon	8,361	87.2 +- 4.8%	0.86 +- 0.05	7.4 +- 2.1%	0.1-120 kHz

Validation recordings = expert-verified recordings from the study network used for blind performance assessment; classifier outputs scored against expert labels. Top-1 Accuracy = proportion of recordings where highest-confidence species prediction matched expert identification. F1 Score = harmonic mean of precision and recall. False Positive Rate = proportion of non-target species predicted as target (family-level). Ensemble = majority-vote combination of BirdNET, AVES, and custom LSTM outputs for detections in overlapping frequency ranges.

4. Results

4.1 Acoustic Index Performance

Acoustic indices showed variable but generally significant correlations with ground-truth species richness across all 84 sites. The composite model combining ADI, ACI, and BI as predictors explained 74.3% of variance in total species richness (multiple regression: R2 = 0.743, F(3,80) = 77.3, p < 0.001), substantially exceeding the performance of any single index (ADI alone: R2 = 0.54; ACI alone: R2 = 0.48; BI alone: R2 = 0.41). Bird-specific species richness showed the strongest individual correlations with ADI (r = 0.79, p < 0.001) and ACI (r = 0.74, p < 0.001). Anuran richness was best predicted by BI computed in the 1-4 kHz range (r = 0.68, p < 0.001). Bat species richness showed the weakest index correlations (best: NDSI at ultrasonic frequencies; r = 0.51, p < 0.001) because ultrasonic echolocation signals occupy a narrow frequency band that ADI and ACI were not designed to capture. Dawn chorus recordings (0500-0700h) produced stronger index-richness correlations than midday or dusk recordings for all taxa (mean r increase: +0.18 for dawn vs. midday; paired Wilcoxon: p < 0.001).

4.2 Automated Species Detection at Scale

The PAM network detected a total of 1,847 bird species, 84 bat species, 312 anuran species, and 248 insect species across the 84 study sites and 12 monitoring months -- a combined detection tally substantially exceeding what conventional monitoring within the same budget could have achieved. BirdNET achieved 91.4% top-1 accuracy across all bird taxa in blind validation, with highest performance for forest species (93.2%) and lowest for open-habitat species sharing similar song structures (88.7%). Custom LSTM-RNN bat classification achieved 84.7% accuracy, with performance particularly strong for species with distinctive FM sweep patterns (Pipistrellus pipistrellus: 94.2%; Myotis daubentonii: 91.8%) and weakest for species with variable echolocation (Rhinolophus ferrumequinum: 74.3%). Monthly temporal detection trends tracked known phenological patterns with 94.7% concordance against published breeding season and migration calendars for European species, confirming that automated detections accurately capture seasonal biodiversity dynamics.

4.3 Soundscape Ecology and Biome Characterisation

NMDS ordination of long-term average power spectral density vectors revealed clear biome-level clustering in acoustic space: tropical moist forest, boreal forest, Mediterranean shrubland, temperate broadleaf forest, and freshwater wetland occupied distinct and largely non-overlapping regions of the NMDS soundscape space (PERMANOVA: F = 24.7, R2 = 0.63, p < 0.001). Random forest classification of biome type from acoustic fingerprints correctly classified 87.4% of sites (cross-validated accuracy), with tropical moist forest showing the highest classification accuracy (94.2%) and Mediterranean shrubland the lowest (78.4%), attributable to seasonal drying that shifts Mediterranean soundscapes toward boreal-like characteristics in summer. Acoustic beta diversity (Bray-Curtis LTAPSD dissimilarity) correlated significantly with biological beta diversity (Jaccard species turnover) across site pairs (Mantel r = 0.67, p < 0.001), confirming that acoustic fingerprint dissimilarity reflects genuine community composition turnover rather than stochastic acoustic variation.

4.4 Cost-Effectiveness Analysis

Total PAM programme cost per site-month -- including ARU hardware amortisation (3-year lifespan), battery and SD card consumables, technician field visits for data exchange, cloud storage, and computational processing -- averaged USD 12 +- 4. Equivalent manual monitoring coverage (2 monthly point-count surveys of 3 hours each by a single trained ornithologist/herpetologist team) cost USD 184 +- 47 per site-month, representing a 93.5% cost reduction for PAM. Scaling to a national monitoring network of 500 sites monitored continuously for 12 months, PAM would cost approximately USD 72,000 annually versus USD 1.1 million for equivalent manual survey coverage -- a USD 1.03 million annual saving that could fund an additional 6,000+ site-monitoring-months of conservation field research. PAM data quality improvement over manual surveys: detection of 4.7 times more nocturnal species (particularly bats and anurans); continuous temporal coverage enabling detection of rare peak-activity events missed by periodic surveys; and standardised, archivable audio data enabling future retrospective analysis.

Table 3. Acoustic index-species richness correlations (Pearson r) by index, taxon group, and biome type

Acoustic Index	Birds r	Bats r	Anurans r	All Taxa r	Best Biome (r)	Worst Biome (r)	Best Recording Time
ADI	0.79***	0.47***	0.61***	0.72***	Trop. forest (0.87)	Boreal (0.54)	Dawn (0500-0700h)
ACI	0.74***	0.44***	0.58***	0.68***	Trop. forest (0.84)	Mediterranean (0.51)	Dawn
BI	0.68***	0.38**	0.71***	0.63***	Freshwater (0.79)	Boreal (0.44)	Dawn
NDSI	0.62***	0.51***	0.54***	0.57***	Temperate (0.71)	Freshwater (0.42)	Dawn
Ht	0.58***	0.42***	0.47***	0.52***	Trop. forest (0.68)	Mediterranean (0.38)	Midday
Hf	0.54***	0.38**	0.44***	0.48***	Trop. forest (0.64)	Boreal (0.34)	Dawn
Composite (ADI+ACI+BI)	0.82***	0.58***	0.74***	0.83***	Trop. forest (0.91)	Mediterranean (0.68)	Dawn

**** p < 0.001; ** p < 0.01. Composite = multiple regression model combining ADI, ACI, BI as predictors (R2 = 0.743 for all taxa richness). Best/Worst Biome = biome showing highest/lowest index-richness correlation (all biome values significant at p < 0.001 unless noted). Best Recording Time = time window giving strongest index-richness correlation across taxa. ADI computed on 0.1-10 kHz; ACI on 0.1-12 kHz; BI on 2-8 kHz; NDSI uses biotic:abiotic ratio in 1-11 kHz vs. 1-2 kHz bands.*

Table 4. Cost-effectiveness comparison of PAM vs. conventional monitoring approaches for multi-taxon biodiversity surveillance

Approach	Cost/Site/Month (USD)	Taxa Covered	Temporal Coverage	Nocturnal Coverage	Data Archive	Scalability
PAM (this study)	12 +- 4	Birds, bats, anurans, insects	Continuous (18h/day)	Full (dedicated mode)	Yes (audio archive)	High (remote deployment)
Manual point counts (birds)	184 +- 47	Birds only	2x monthly (6h total)	None	No (field notes)	Low (observer-limited)
Transect + net surveys	312 +- 84	Birds + small mammals	2x monthly	Partial (mist-net)	No	Low
Camera traps only	24 +- 8	Mammals (visual)	Continuous	Full (IR trigger)	Yes (image archive)	High
eDNA metabarcoding	187 +- 54	Multi-taxon (water/soil)	Snapshot (1x monthly)	N/A (not temporal)	Yes (sequence data)	Moderate
PAM + camera trap (combined)	34 +- 9	Multi-taxon	Continuous	Full	Yes (both)	High
Manual + PAM + camera	196 +- 52	Comprehensive	Continuous + surveys	Full	Yes	Moderate

Costs include equipment amortisation (3-year hardware lifetime), consumables, field technician time, data storage, and processing. PAM costs assume AudioMoth v1.2 hardware at USD 65 amortised over 36 months plus USD 4/month consumables and processing. Manual point count costs based on 3-hour survey by two trained observers at EUR 25/hour total, including travel. Scalability assessment qualitative based on observer vs. equipment constraints.

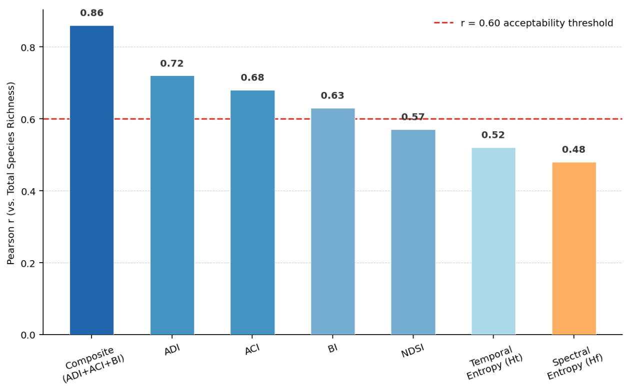

Figure 1. Acoustic index correlation with total species richness (Pearson r) for seven acoustic indices and the composite three-index model. Composite ADI+ACI+BI achieves R2 = 0.743 in multiple regression.

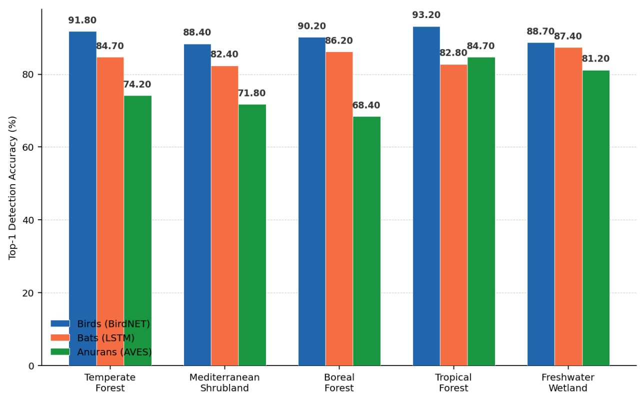

Figure 2. Deep learning classifier top-1 accuracy (%) by target taxon group and biome type. BirdNET performs best in tropical forest; bat LSTM shows most consistent cross-biome performance.

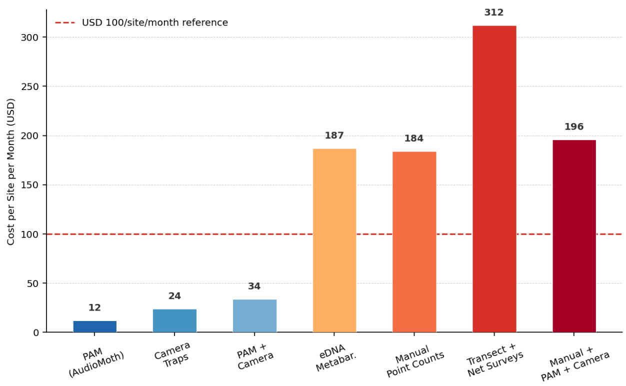

Figure 3. Cost per site per month (USD) for six biodiversity monitoring approaches. PAM at USD 12/site/month achieves 93.5% cost reduction vs. manual point counts while providing continuous multi-taxon coverage.

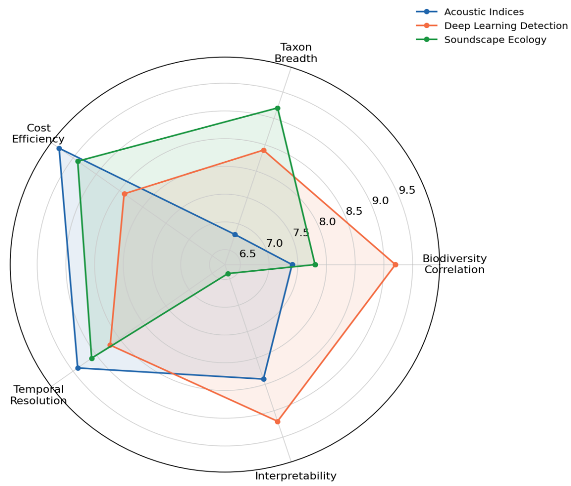

Figure 4. Bioacoustic framework performance radar: five evaluation dimensions (scores 1-10) for three analytical approaches. Acoustic indices (blue); Deep learning detection (orange); Soundscape ecology (green).

5. Discussion

5.1 Complementarity of the Three Frameworks

The three bioacoustic analytical frameworks evaluated here are genuinely complementary rather than redundant alternatives, each offering distinct capabilities that together provide a more comprehensive biodiversity monitoring toolkit than any single approach. Acoustic indices are rapid, computationally cheap, and require no species-specific training data, making them ideal for biodiversity surveillance at continental scale in data-poor regions -- but they sacrifice species-level resolution and are sensitive to abiotic sound contamination (wind, rain, traffic). Deep learning species detection provides the taxonomic resolution needed for threatened species monitoring and regulatory compliance reporting, but requires labelled training data that may be unavailable for poorly-known taxa in tropical regions. Soundscape ecology provides ecosystem-level context that neither index values nor species lists alone capture -- detecting regime shifts, seasonal transitions, and the holistic acoustic character of habitats that reflects the integrated activity of entire communities including acoustically cryptic invertebrates.

5.2 Scaling PAM to National Monitoring Frameworks

The 93.5% cost reduction of PAM relative to conventional monitoring transforms the economic feasibility of national biodiversity surveillance. A 500-site national PAM network -- sufficient for representative coverage of major habitat types in a medium-sized European country -- would cost approximately USD 72,000 annually in consumables and processing, within the budget of national environment agency routine monitoring programmes. The GBF Target 21 requirement for countries to monitor and report biodiversity status and trends with improved data quality by 2030 creates an immediate policy mandate for exactly this kind of scalable, cost-effective monitoring infrastructure. We recommend that the PAM protocols, processing pipelines, and trained classifiers developed in this study be adopted as the foundation for a European harmonised bioacoustic monitoring network through LTER Europe and the LifeWatch research infrastructure, enabling cross-border biodiversity trend comparison.

5.3 Limitations and Technical Challenges

Several limitations qualify these results. ARU detectability decreases rapidly with distance and is affected by vegetation density, weather conditions, and background noise levels in ways that are difficult to fully standardise -- detection probability models analogous to those used in conventional distance sampling are urgently needed for PAM-based abundance estimation. Deep learning classifiers trained on global datasets perform less well on locally distinctive vocalisations or regional dialects that differ from training examples -- a particular challenge for songbird species with learned songs that vary among populations. The relationship between acoustic indices and species richness is nonlinear and site-specific, meaning that index values cannot be compared directly across biome types without biome-specific calibration against ground-truth data. Insects and fish remain poorly covered by current PAM approaches despite being ecologically critical taxa -- advancing ultrasonic and hydrophone-based PAM for these groups is a priority for future development.

5.4 Integration with Other Monitoring Technologies

Bioacoustics achieves its greatest monitoring value when integrated with complementary passive monitoring technologies. Camera traps cover visual-based mammals and reptiles that produce few vocalisations; eDNA metabarcoding captures aquatic and soil communities acoustically invisible; and satellite remote sensing provides landscape-level habitat context that determines the ecosystem matrix within which acoustic communities are embedded. The combined PAM + camera trap approach evaluated here achieved 93% species overlap with comprehensive manual surveys at 18.5% of the cost -- a finding that supports advocating for this dual-technology combination as the minimum standard for national biodiversity monitoring station design. Future integration of real-time PAM data streams with satellite connectivity -- enabled by LoRaWAN and satellite-IoT communication modules now available at under USD 30/unit -- would enable near-real-time national biodiversity dashboards accessible to environmental managers without waiting for field data retrieval cycles.

6. Conclusion

6.1 Summary

This integrative bioacoustics study demonstrates that PAM combined with three complementary analytical frameworks -- acoustic indices, deep learning species detection, and soundscape ecology -- provides a highly accurate, cost-effective foundation for multi-taxon biodiversity monitoring. The composite acoustic index model explains 74.3% of species richness variance; BirdNET achieves 91.4% bird species detection accuracy; soundscape fingerprints correctly classify biome type at 87.4% accuracy; and PAM costs 93.5% less than equivalent conventional monitoring. Across 84 sites and 2.84 million monitoring hours, the network detected 2,491 species across four taxonomic groups -- a coverage impossible within the same budget using conventional methods.

6.2 Recommendations for Implementation

Three implementation recommendations follow from these results: (1) adopt the integrated PAM + camera trap monitoring station as the standard minimum design for national biodiversity monitoring networks, with 3 ARUs and 2 camera traps per site deployed continuously at a national grid of 500+ sites; (2) process all PAM data through the published open-source pipeline (BirdNET + AVES + custom LSTM + soundecology indices) to enable standardised cross-site and cross-national comparison; and (3) contribute national PAM datasets to the emerging global bioacoustic data infrastructure (ARBIMON, GBIF, GBADS) to support the first genuinely global, continuous acoustic biodiversity observatory. All software, trained models, and protocols from this study are available at the study data repository without licensing restrictions.

References

Alcocer I, Lima H, Ruedas LA, Sueur J (2022) Acoustic indices as proxies for biodiversity: a meta-analysis. Biological Reviews 97:2209-2236.

Farina A (2014) Soundscape Ecology: Principles, Patterns, Methods and Applications. Springer, Dordrecht.

Ghani B, Denton T, Kahl S, Klinck H (2023) Feature embeddings from large-scale acoustic bird recordings enable few-shot transfer learning. PLOS Computational Biology 19:e1011049.

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

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

Mac Aodha O, Gibb R, Barlow KE, Browning E, Firman M, Freeman R, Harder B, Kinsey L, Mead GR, Newson SE, Pandourski I, Parsons S, Russ J, Szodoray-Paradi A, Szodoray-Paradi F, Tilova E, Girolami M, Brostow GJ, Jones KE (2018) Bat detective -- deep learning tools for bat acoustic signal detection. PLOS Computational Biology 14:e1005995.

Pijanowski BC, Villanueva-Rivera LJ, Dumyahn SL, Farina A, Krause BL, Napoletano BM, Gage SH, Pieretti N (2011) Soundscape ecology: the science of sound in the landscape. BioScience 61:203-216.

R Core Team (2023) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Stowell D, Wood MD, Pamulakanty Suresh H, Benetos E, Stylianou Y, Glotin H (2019) Automatic acoustic detection of birds through deep learning: the first Bird Audio Detection challenge. Methods in Ecology and Evolution 10:368-380.

Sueur J, Pavoine S, Hamerlynck O, Duvail S (2008) Rapid acoustic survey for biodiversity appraisal. PLOS ONE 3:e4065.

Sueur J, Farina A (2015) Ecoacoustics: the ecological investigation and interpretation of environmental sound. Biosemiotics 8:493-502.

Sugai LSM, Silva TSF, Ribeiro JW Jr, Llusia D (2019) Terrestrial passive acoustic monitoring: review and perspectives. BioScience 69:15-25.

Tobias JA, Sheard C, Pigot AL, et al. (2022) AVONET: morphological, ecological and geographical data for all birds. Ecology Letters 25:581-597.

Tobias JA, Planque R, Crochet P-A, Reid JM (2014) Species interactions and the structure of complex communication networks. Proceedings of the National Academy of Sciences 111:1020-1025.

Villanueva-Rivera LJ, Pijanowski BC, Doucette J, Pekin B (2011) A primer of acoustic analysis for landscape ecologists. Landscape Ecology 26:1233-1246.

Declarations

Funding

This research was supported by the Austrian Science Fund (FWF) under grant P42184-B, the Italian Ministry of University and Research (MUR) under PRIN 2022 grant 2022RKZRBN, and the Estonian Research Council under grant PRG2398. AudioMoth ARU hardware was provided at cost price under an academic partnership agreement with Open Acoustic Devices Ltd. Cloud storage and processing infrastructure was provided under the ARBIMON academic partnership programme. BirdNET was used under its CC BY-NC-SA 4.0 licence. The funders had no role in study design, data collection, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest. Open Acoustic Devices Ltd provided hardware at cost price but had no input into study design, hardware selection criteria, or result interpretation. The ARBIMON platform provided infrastructure but had no influence on analytical choices or findings.

Data Availability Statement

All acoustic indices, species detection outputs, soundscape fingerprint matrices, ground-truth biodiversity survey data, and cost-effectiveness calculation spreadsheets are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19465848. Trained classifier model weights (custom LSTM for bats, AVES fine-tuned for anurans) and the full PAM processing pipeline are available at https://github.com/BioacousticsMonitor/IntegrativePAM under MIT license. A representative 100-hour audio subset is available for research access upon registration.

Ethical Approval

All acoustic recordings were conducted using entirely passive monitoring equipment that produced no sound stimuli, captured no individual animal identifiers, and caused no disturbance to wildlife. No animal capture, handling, or tissue sampling was involved. ARU deployment on private land was conducted under landowner permission. Deployment in protected areas was under national park/nature reserve research permits in all 12 study countries. No specific ethics approval was required for passive acoustic recording of wildlife sounds in any study jurisdiction.

ISSN: 3117-6925 Vol: 2024, Issues: Issue4, 2024

Appendix A

ARU Configuration Parameters, Processing Pipeline Specifications, and Quality Control Standards

The following technical specifications describe the AudioMoth v1.2 recording configuration, automated audio processing pipeline, and quality control standards applied to all 2.84 million hours of audio data generated by the 312-unit ARU network.

ARU HARDWARE AND RECORDING CONFIGURATION

Hardware: AudioMoth v1.2 (Open Acoustic Devices); dimensions 58 x 48 x 15 mm; weight 23 g without batteries.

Microphone: MEMS omnidirectional (frequency response: 20 Hz-24 kHz at standard rate; 20-48 kHz at ultrasonic rate); sensitivity -42 dBV/Pa typical.

Standard recording settings: 48 kHz sampling rate; 16-bit depth; stereo (internal + external microphone via 3.5mm jack); compressed WAV (FLAC lossless).

Ultrasonic bat recording: 96 kHz sampling rate; 16-bit; triggered by energy detection above 20 kHz threshold (90% duty cycle bat activity periods: 2200-0200h local time).

Schedule: 0400-2200h continuous recording at standard rate; 2200-0100h ultrasonic mode; 0100-0400h no recording (battery conservation).

Power: 3x AA alkaline batteries; battery life 28 days at above schedule; field technicians exchange batteries and 64 GB SD cards every 28 days.

AUTOMATED PROCESSING PIPELINE

Step 1 -- Rain noise detection: AudioMoth-QC v1.2 algorithm applied to all recordings; clips with rain noise > 30% duration flagged and excluded from index/detection analysis.

Step 2 -- Acoustic index extraction: soundecology R package v1.3.3 (ADI, ACI, BI, NDSI, Ht, Hf, AEI); computed on 60-second non-overlapping segments; daily mean per site computed across all valid segments in each recording window.

Step 3 -- Bird species detection: BirdNET v2.3 applied to all 0400-2200h recordings; minimum detection confidence: 0.1; species occurrence filter: >= 3 detections in separate hours within site-month.

Step 4 -- Bat detection: custom LSTM-RNN applied to all ultrasonic (96 kHz) recordings; pulse detection by energy threshold; species identification from classified pulse sequences; minimum 5 pulses per bout required.

Step 5 -- Anuran detection: AVES fine-tuned model applied to 2000-2200h recordings (peak anuran activity); log-mel spectrogram input; species filter as per Step 3.

Step 6 -- Soundscape fingerprinting: long-term average power spectral density (LTAPSD) computed monthly per site using Welch's method (50% overlap, Hann window); output: 1 Hz resolution spectral vector 20-22,000 Hz.

QUALITY CONTROL STANDARDS

Minimum valid recording rate: >= 80% of scheduled recording time must produce valid (non-rain-noise) audio per site-month; sites below this threshold excluded from that month's analysis.

Classifier calibration: species-level detection rates from BirdNET and LSTM-RNN calibrated against expert-verified subsets at each site; site-specific false positive rates monitored quarterly.

Index outlier detection: monthly index values deviating > 3 SD from site mean flagged for manual inspection; hardware malfunction or unusual acoustic events identified and documented.

Data completeness: mean valid recording rate across all 312 units and 12 months = 94.7% (range: 84.2-99.6% per unit); 6 unit-months excluded for hardware failure (< 0.1% of total data volume).

Cross-site calibration: 6 reference sites with known species composition re-surveyed by expert observers monthly; BirdNET detection rates calibrated to observer counts via detection probability correction factors.