A Data-Driven Approach for Discovering Pure Clusters in Complex Multi-Source Terrestrial Soundscapes

Thomas Napier[1[0000−0003−4218−6731]], Euijoon Ahn[2], Ickjai Lee[2]

[1] James Cook University, Townsville, QLD 4811, Australia
[2] James Cook University, Cairns, QLD 4870, Australia
{thomas.napier,euijoon.ahn,ickjai.lee}@jcu.edu.au

Abstract. Unsupervised clustering is increasingly used in ecoacoustics to uncover patterns in complex natural soundscapes and support large-scale biodiversity monitoring. These approaches typically evaluate clustering quality using internal metrics such as compactness and separation, including the Silhouette Coefficient and Davies-Bouldin Index. However, these metrics do not account for ecological interpretability, meaning clusters can appear mathematically well-formed while mixing unrelated sound sources such as bird vocalisations, insect activity, rainfall, or human noise. This study presents a data-driven clustering framework for ecoacoustic recordings that combines configurable temporal segmentation, dimensionality reduction, and unsupervised learning. We apply the ecological clustering metrics Dominant Class Proportion and Mean Active Classes Per Cluster to select optimal clustering results from a grid search of 198 parameter combinations per segmentation strategy, comparing fixed-size and step-based windowing across four ecologically distinct sites from the Australian Acoustic Observatory. Results show that step-based windowing, when paired with ecological clustering metrics, consistently yields purer clusters than configurations selected using internal metrics alone. These findings highlight the limitations of conventional clustering metrics and demonstrate how the choice of segmentation and metrics grounded in ecological coherence can yield clusters that better reflect real sound sources in complex, multi-source soundscapes without requiring dense manual annotation.

Keywords: Biodiversity monitoring · Clustering validity · Ecoacoustics· Unsupervised learning

1 Introduction

Ecoacoustics is a field of study dedicated to comprehending the ecological patterns and processes in natural soundscapes [4]. With the rise of Autonomous Recording Units (ARUs), the volume of audio data collected in natural habitats has scaled exponentially, enabling unprecedented long-term, multi-site monitoring initiatives such as the Australian Acoustic Observatory (A2O) [13]. However, the exponential volume of data collected by ARUs has rendered manual annotation unsustainable [8]. To address this, unsupervised learning techniques are increasingly used to discover emergent patterns in unlabelled ecoacoustic datasets [9, 11, 12, 15, 7]. The goal is not only to reduce data complexity but to produce clusters that are meaningful from an ecological perspective. In this context, interpretable clusters are groupings that correspond to real biological or environmental phenomena, such as bird choruses, insect activity, rainfall, or anthropogenic noise.

However, standard evaluation practices often rely on Clustering Internal Validity Indices (CIVIs) such as the Silhouette Coefficient (SC) [14] and Davies–Bouldin (DB) indices [2], which assess clustering quality based on abstract geometric properties like compactness and separation. While CIVIs provide quantitative measures of clustering quality, a significant shortcoming is their inherent blindness to the ecological interpretability of the resulting clusters [12]. Consequently, while useful for internal comparisons, these application-agnostic metrics may favour clusters that are numerically coherent but biologically incoherent and therefore may fail to group clusters that correspond to expert knowledge [3, 11].

Another key factor influencing clustering outcomes is the segmentation of continuous recordings into smaller analysis units, a process known as temporal windowing. The choice of window size and spacing must align with distinct signal events; otherwise overly long windows risk combining unrelated sounds, while short or poorly spaced windows may fragment coherent acoustic events or introduce redundancy [17]. Despite its importance, windowing is rarely treated as a tunable component of ecoacoustic analysis pipelines. The lack of ecological congruence in clustering outputs and window size selections can therefore undermine the utility of unsupervised methods for drawing sound ecological conclusions. To address this gap, we propose a novel data-driven framework. Rather than relying on manual heuristics or expert-defined rules, our approach leverages patterns inherent in the data to guide segmentation choices, feature extraction, and clustering evaluation. The contributions of this work can be summarised as follows:

- We introduce a scalable, data-driven clustering framework for multi-source soundscapes, using a domain-aligned selection strategy to choose configurations that are quantitatively pure and ecologically coherent;
- We perform a large-scale grid search over 198 configurations per windowing strategy, demonstrating that purity and interpretability are maximised under sparse step-based sampling and ecological metric selection;
- We provide evidence-based guidance on segmentation and clustering configurations that maximise biological relevance, offering a practical path for unsupervised analysis of large-scale, multi-source soundscapes without dense manual labels;
- We systematically compare fixed-size and step-based segmentation strategies across four ecologically distinct sites, showing how temporal resolution affects the ecological coherence of clusters.

2 Related Works

Unsupervised clustering has proven valuable in ecoacoustic analysis, allowing researchers to uncover latent ecological patterns in large unlabelled datasets. However, many studies continue to evaluate clustering quality using internal validation metrics such as the SC and DB indices. While assessing compactness and separation, these metrics do not guarantee ecological relevance. As a result, high internal scores may be misleading. An example of this can be found in [10], where coral reef sound events were clustered, yielding a moderate average SC of approximately 0.31, yet expert review revealed that several clusters contained multiple biologically distinct call types.

Another key methodological factor influencing clustering outcomes is how audio data is segmented before feature extraction. Fixed-length windowing is commonly used, where recordings are split into uniform chunks under the assumption that each segment reflects the local soundscape composition. While fixed-length windowing is common [11], the choice of window size can significantly affect cluster consistency. For example, long segments may merge distinct acoustic events, while short windows can fragment them or introduce noise [16]. This is further exemplified in [17], where the authors directly compared sliding-window segmentation with syllable-level segmentation in continuous frog recordings. They showed that sliding windows often break frog calls mid-syllable, whereas their domain-aligned approach achieved significantly higher F1-scores across species.

While unsupervised clustering represents a valuable tool for ecoacoustic analysis, current approaches tend to emphasise mathematical separation over ecological insight. Internal metrics alone are insufficient proxies for ecological coherence. There remains a need to incorporate ecological heuristics and improve segmentation using data-driven approaches. Only by addressing these gaps can clustering approaches achieve practical utility for biodiversity monitoring and ecological interpretation.

3 Proposed Methodology

In this study, we propose a data-driven clustering framework for use with large-scale, multi-source ecoacoustic soundscape recordings collected from the A2O, as visualised in Fig. 1. Raw audio files were paired with a curated set of one-hot encoded annotations indicating the sound class presence (e.g. birds, insects, wind, human speech), which were used solely for validation and not as input during clustering. To prepare the data for analysis, two distinct segmentation strategies were applied: fixed-size windowing and step-based windowing. Both strategies aimed to assess the impact of temporal coverage and resolution on cluster interpretability.

For each extracted segment, we computed a 13-dimensional Mel Frequency Cepstral Coefficient (MFCC) feature vector, chosen for its compact spectral representation, strong alignment with human auditory perception, and successful

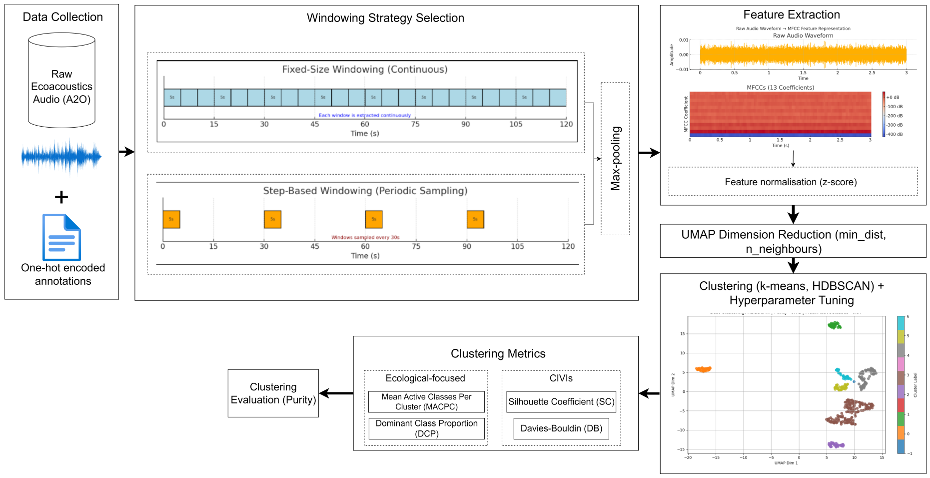


Fig. 1: A conceptual diagram for a data-driven clustering framework for ecoacoustic analysis.

usage across varied approaches in ecoacoustics [5, 6, 15]. These features were normalised using z-score transformation to ensure consistency across recordings of varying amplitude scales. We then applied Uniform Manifold Approximation and Projection (UMAP) to reduce the feature space while preserving local and global structure [1]. UMAP parameters, including the number of neighbours and minimum distance, were varied as part of the experimental grid search to evaluate their impact on cluster separability.

Following dimensionality reduction, clustering was performed using both k-means and HDBSCAN algorithms. HDBSCAN, in particular, was selected for its ability to identify noise and irregularly shaped clusters without requiring a predefined number of groups. To assess the quality of the resulting clusters, we computed both internal and ecological-focused evaluation metrics. This framework enables a systematic comparison of how temporal segmentation and feature transformations influence the formation of semantically meaningful acoustic clusters in natural soundscape recordings.

3.1 Feature Representation, Dimensionality Reduction, and Clustering

Each audio segment was represented using 13 MFCCs, standardised using z-score normalisation. Dimensionality reduction was applied using UMAP, with a grid search conducted over `n_neighbors` $\in \{5, 10, 15\}$ and `min_dist` $\in \{0.0, 0.1, 0.5\}$, resulting in 9 UMAP configurations. For each UMAP embedding, clustering was performed using both k-means and HDBSCAN. The k-means approach varied the number of clusters between 3 and 15, yielding 13 configurations per UMAP projection. For HDBSCAN, the grid search explored `min_cluster_size` $\in \{5, 10, 20\}$

and `min_samples` $\in \{1, 5, 10\}$, yielding an additional 9 configurations per UMAP projection. Each resulting clustering configuration was evaluated using internal metrics and ecological heuristics. A total of 198 unique configurations per windowing strategy were considered, and the top-performing runs were selected based on either internal or ecological selection criteria per site.

3.2 Clustering Validation Metrics

Clustering Metrics To assess clustering performance, we use CIVIs as well as the newly proposed ecological metrics tailored to ecoacoustics. We compute the SC and DB indices, which are widely used to assess the compactness and separation of clusters in feature space. However, these metrics do not capture ecological interpretability or annotation utility, particularly in multi-label, polyphonic soundscapes.

To address this gap, we propose using two biologically informed metrics based on multi-class annotation distributions within each cluster: *Mean Active Classes Per Cluster* and *Dominant Class Proportion*.

Mean Active Classes Per Cluster (MACPC). MACPC quantifies within-cluster biological diversity by computing the average number of active (i.e., present) classes per sample in each cluster and then averaging across all clusters:

$$MACPC = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i \in C_k} \sum_{j=1}^{L} 1[y_{ij} > \tau]. \tag{1}$$

Here, K denotes the number of clusters, C_k the set of indices in cluster k, L the number of possible classes, y_{ij} the binary indicator of presence for class j in sample i, and τ a threshold (typically 0.5). A MACPC value closer to 1 may indicate greater acoustic homogeneity, which is generally a desirable trait for downstream labelling and ecological interpretation.

Dominant Class Proportion (DCP). DCP measures the extent to which a single class dominates each cluster, defined as the proportion of samples assigned to the most frequent class within that cluster:

$$DCP = \frac{1}{K} \sum_{k=1}^{K} \frac{\max_j \sum_{i \in C_k} y_{ij}}{|C_k|}. \tag{2}$$

Higher DCP values suggest that clusters may correspond more strongly to interpretable units, such as consistent sound types or species vocalisations.

Evaluation Metric: Purity Purity is used as an external evaluation metric to assess how well each cluster aligns with ground-truth labels. For each cluster, it is defined as the proportion of samples belonging to the most frequent class:

$$Purity = \frac{1}{|C|} \sum_{c \in C} \frac{\max_k |\{x_i \in c : y_i = k\}|}{|c|}. \quad (3)$$

Here, C is the set of clusters and y_i is the label of sample x_i. In our implementation, labels are stored as a binary multilabel matrix. For each cluster, we sum class occurrences, divide the maximum count by the total, and average across clusters. Purity is computed only post-hoc and is not used for selecting configurations.

Pairwise external validity measures such as Adjusted Mutual Information are less straightforward to interpret in our setting because annotations are incomplete, strongly imbalanced, and often multi-species within a window. We therefore focus on Purity as a simple external summary, complemented by DCP and MACPC as task-specific ecological heuristics, while making the selection–evaluation separation explicit to minimise target leakage.

4 Experiments and Results

4.1 Datasets

Table 1: Structured breakdown of class labels by site with groupings, subcategories, and totals. Zero values shown as '-'.

Category / Label	Duval	Mourachan	Rinyirru	Undara	Total
Biophony	139 (34.0%)	394 (81.1%)	236 (95.9%)	275 (58.3%)	1044
- *birds*	103 (25.2%)	150 (30.9%)	93 (37.8%)	79 (16.7%)	425
- *frogs*	-	1 (0.2%)	-	26 (5.5%)	27
- *insects*	29 (7.1%)	241 (49.6%)	142 (57.7%)	170 (36.0%)	582
- *mammals*	7 (1.7%)	2 (0.4%)	1 (0.4%)	-	10
Geophony	206 (50.4%)	81 (16.7%)	4 (1.6%)	183 (38.8%)	474
- *wind (strong)*	-	30 (6.2%)	-	49 (10.4%)	79
- *wind (light)*	14 (3.4%)	41 (8.4%)	4 (1.6%)	134 (28.4%)	193
- *rain (heavy)*	2 (0.5%)	9 (1.9%)	-	-	11
- *rain (light)*	190 (46.5%)	1 (0.2%)	-	-	191
Anthrophony	52 (12.7%)	5 (1.0%)	2 (0.8%)	13 (2.8%)	72
- *human speech*	20 (4.9%)	-	-	11 (2.3%)	31
- *vehicles (aircraft/cars)*	32 (7.8%)	5 (1.0%)	2 (0.8%)	2 (0.4%)	41
Other/Silence	12 (2.9%)	6 (1.2%)	4 (1.6%)	1 (0.2%)	23
- *background silence*	-	-	-	-	0
- *misc/uncertain*	12 (2.9%)	6 (1.2%)	4 (1.6%)	1 (0.2%)	23

This study uses 24-hour continuous recordings from four ecologically diverse sites within the A2O: Duval, Mourachan, Rinyirru, and Undara. Recordings were collected using ARUs and sampled to span a full diurnal cycle, capturing both daytime and nocturnal acoustic activity. Annotation was performed using the LEAVES software tool, where approximately 10% of samples within each cluster were manually annotated, and the dominant label was propagated to the remainder [8]. Labels were encoded in one-hot format across a set of common classes.

Table 1 provides a breakdown of the annotated classes per site. Mourachan and Rinyirru were dominated by biophony (especially insects and birds), while Duval exhibited high levels of geophony, including light and heavy rain. Anthrophonic sounds (human origin) were less common but appeared more frequently at Duval and Undara. This variation in soundscape composition provides a biodiverse and realistic testbed for assessing ecological interpretability across clustering configurations.

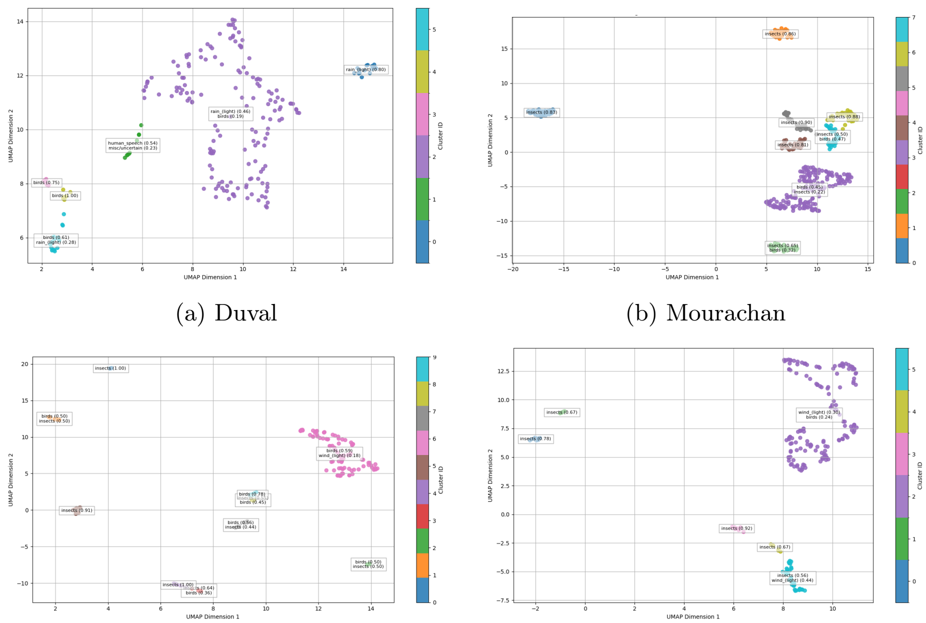
4.2 Windowing Strategies

Two temporal segmentation strategies were employed to investigate how windowing impacts cluster formation and ecological coherence. In the fixed-size strategy, recordings were divided into consecutive, non-overlapping segments of 1, 3, 5, or 10 seconds. In the step-based strategy, 5- and 10-second segments were sampled periodically every 15, 30 or 60 seconds (e.g., 10s every 60s), mimicking duty-cycled sampling. These configurations were selected to explore the trade-off between temporal coverage and temporal resolution, particularly with respect to polyphonic sound scenes. This enabled the assessment of whether sparser but broader coverage (step-based) improves cluster separability and ecological interpretability compared to dense, but potentially redundant, fixed-step windowing.

4.3 Clustering Configuration Outcomes Across Sites

To evaluate the quality of unsupervised clustering outcomes, we visualised the optimal configurations per site using UMAP embeddings coloured by HDBSCAN cluster assignment and annotated with the dominant ecological class in Fig. 2. Across all four A2O sites, the selected configurations using HDBSCAN with step-based windowing (10s every 60s) produced well-separated clusters, many of which aligned with single dominant class labels. Notably, the Duval site exhibited strong separation between geophonic and biophonic-dominant clusters, while Mourachan and Rinyirru displayed tighter groupings corresponding to insect and bird activity. Undara showed more diffuse clusters, potentially reflecting higher soundscape variability or overlap between wind and biotic signals. Overall, these results support the ability of the chosen clustering pipeline to isolate ecologically distinct sound types in diverse environments.

To assess whether clustering outcomes aligned with meaningful ecological patterns, we selected four representative clusters from the Duval site for detailed analysis as illustrated in Fig. 3. Each cluster's internal consistency was examined using mel spectrograms of exemplar samples. As shown, Cluster 0 was dominated by low-frequency, low-intensity patterns associated with wind and light rain, which are common geophonic features. Cluster 1 included broadband vertical striations characteristic of human speech, particularly in the mid- to high-frequency range. Cluster 2 displayed continuous broadband energy with minimal fluctuation, consistent with heavy rainfall. Cluster 3 exhibited pulsed



(a) Duval (b) Mourachan

(c) Rinyirru (d) Undara

Fig. 2: Best clustering configurations for four A2O monitoring sites using HDBSCAN and UMAP. Each subplot visualises the 2D embedding of step-wise windowed acoustic features, coloured by cluster and annotated with dominant class labels. Higher dominance proportions suggest more ecologically coherent clusters.

bird vocalisations partially overlapping with low-frequency geophony, highlighting the proposed framework's ability to capture polyphonic mixtures. These observations reinforce the idea that the resulting clusters are not just statistically separable, but also acoustically and semantically coherent.

4.4 Metric-Based Comparison of Selection Strategies

Table 2 compares clustering performance across all sites under different configuration-selection strategies, contrasting internal metrics (SC, DBI) with ecological metrics (MACPC, DCP) using Purity as the evaluation criterion. While internal metrics favoured configurations with compact and well-separated clusters (e.g., an SC of 0.829 at Mourachan), these did not consistently correspond to high purity. In contrast, configurations selected using ecological metrics produced stronger purity scores across most sites. For example, at Mourachan, the configuration selected by DCP achieved a MACPC of 1.296 and a purity of 0.724, compared to 0.644 when selected by SC which is a 12.4% improvement. Across all sites, purity improvements from SC- to DCP-selected configurations ranged

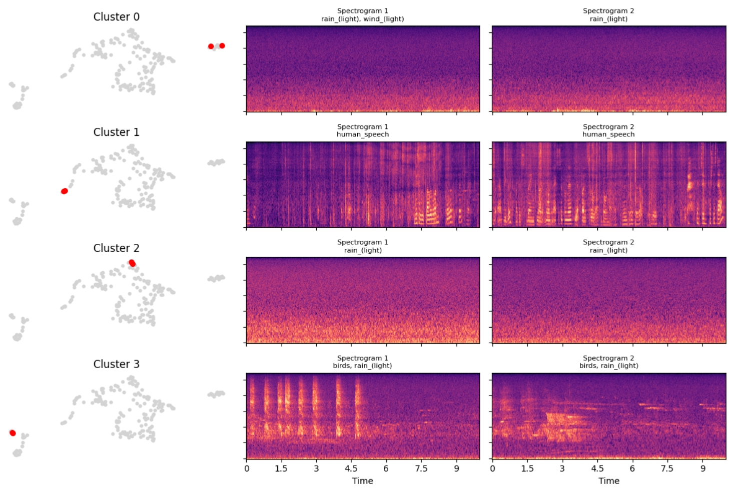


Fig. 3: UMAP projections and spectrograms of four acoustic clusters from the Duval site. Each row displays a distinct cluster, with two sample points highlighted in red on the UMAP embedding (left) and their corresponding mel spectrograms (right).

from 3.8% to 12.4% (mean 7.13%). This pattern suggests that the proposed ecological metrics and data-driven windowing strategies are better suited for identifying clusters that preserve dominant class structure and semantic clarity within complex soundscapes.

There is a single exception at Duval, where DBI outperformed DCP. This possibly reflects metric non-monotonicity: internal and ecological metrics optimise different properties, but occasional cross-overs may occur when their objectives coincide. This does not undermine the broader pattern that ecological metrics provide a more stable alignment with semantically meaningful cluster structure.

4.5 Temporal Performance of Step-Based Windowing

Each timeline shown in Fig. 4 represents the sequence of dominant class labels per time window, with colour denoting its prevailing sound categories. Results indicate that step-based sampling provides a coarse but effective summary of the soundscape. While it does sacrifice some temporal granularity, it captures the major acoustic events with less interference from short-duration noise. In Duval, for example, long sequences of wind and rain were preserved, while Mourachan and Rinyirru highlighted insect-dominant periods clearly. In this light, step 10/60 windowing serves as a scalable and effective approximation for large-scale ecoacoustic tasks, especially where the goal is to extract broad patterns efficiently rather than capture every transient detail.

Table 2: Summary of cross-metric performance across sites. Each row shows the configuration selected by a given metric and bold values indicate the highest subsequent score per column.

Site	Configuration			Clustering Metrics				Evaluation
	SelectedBy	Method	Windowing	SC	DBI	DCP	MACPC	Purity
Duval	SC	hdbscan	Step 10/30	–	**0.316**	0.528	**0.899**	0.633
	DBI	hdbscan	Fixed 3s	**0.555**	–	0.113	0.147	**0.733**
	DCP	hdbscan	Step 10/60	0.341	0.476	–	1.280	0.667
	MACPC	hdbscan	Step 10/60	**0.555**	0.346	**0.793**	–	0.634
Mourachan	SC	kmeans	Step 10/60	–	**0.223**	0.812	1.356	0.644
	DBI	kmeans	Step 10/15	**0.829**	–	0.349	0.463	0.669
	DCP	hdbscan	Step 10/60	0.498	0.659	–	**1.296**	**0.724**
	MACPC	hdbscan	Step 10/30	0.541	0.714	**0.856**	–	0.645
Rinyirru	SC	kmeans	Step 10/60	–	**0.352**	0.204	0.295	0.702
	DBI	kmeans	Step 10/15	**0.735**	–	0.174	0.285	0.692
	DCP	hdbscan	Step 10/30	0.704	1.329	–	**0.838**	**0.729**
	MACPC	kmeans	Step 10/30	0.697	0.441	**0.545**	–	0.642
Undara	SC	kmeans	Step 10/60	–	**0.275**	0.641	**1.216**	0.560
	DBI	kmeans	Step 10/60	**0.753**	–	0.689	1.265	0.579
	DCP	hdbscan	Step 10/60	0.404	0.483	–	1.373	**0.614**
	MACPC	hdbscan	Step 10/60	0.660	0.436	**0.725**	–	0.543

5 Discussion

Our findings reveal a consistent disconnect between conventional internal clustering metrics and ecologically meaningful outcomes in the majority of cases. This was evident at several sites, where the configurations selected based on SC yielded clusters that were internally coherent but acoustically heterogeneous, and thus had reduced purity. These results align with those found in other studies [10] and reinforce the limitations of relying solely on internal metrics for ecoacoustic clustering. In contrast, the ecological heuristics used in this study demonstrated superior alignment with semantically meaningful groupings. Across multiple sites, configurations selected using the domain-aligned metrics produced clusters with higher purity.

However, MACPC did not always yield strong purity performance, particularly at sites with strong multi-species overlap such as Rinyirru. This behaviour is somewhat to be expected because the two metrics optimise different properties: Purity rewards single-label dominance, whereas MACPC emphasises multi-species co-occurrence patterns that are common in polyphonic soundscapes. In assemblages with strong overlap, clusters that better represent ecological structure may appear less pure under a dominance-based external metric. At more homogeneous sites such as Mourachan, this trade-off is minimal. These results suggest that ecological and internal metrics capture complementary dimensions of clustering quality, and that divergences may reflect ecological realism rather than reduced clustering performance.

Step-based windowing (10s every 60s) was found to be effective in balancing the trade-off between resolution and interpretability. Compared to fixed-length windowing, this approach captured dominant acoustic events while filtering out low-activity or ambiguous periods. When paired with HDBSCAN and appropriate UMAP settings, this sampling strategy consistently yielded pure clusters across diverse soundscapes. The findings suggest that temporal segmentation

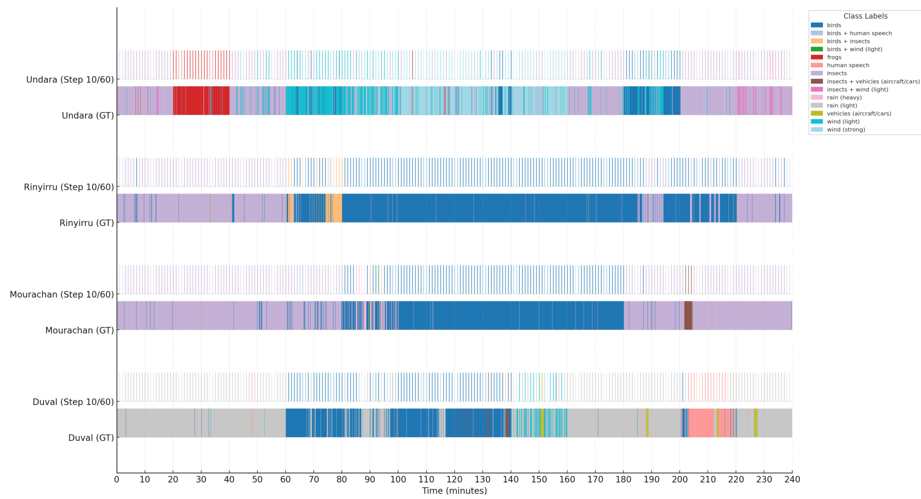


Fig. 4: Comparison of the best temporal annotation strategy across four ecoacoustic monitoring sites: Duval, Mourachan, Rinyirru, and Undara. For each site, two rows are shown: (i) **Step 10/60** represents 10-second windows sampled every 60 seconds and; (ii) **Ground Truth (GT)** represents ground truth labelled events over time. Each segment is colour-coded according to the dominant acoustic class present. Step-based sampling provides coarser temporal resolution but often captures dominant events effectively while reducing noise, balancing annotation resolution and label reliability.

strategies can play an important role in shaping clustering outcomes and should be carefully considered as part of the optimisation pipeline. We show that configurations chosen using ecological heuristics not only produce purer clusters but also demonstrate generalisability across distinct ecosystems. Importantly, this was achieved without relying on dense manual annotations or species-specific classifiers. Instead, our framework offers a clear data-driven pathway for generating biologically relevant groupings in large, heterogeneous ecoacoustic datasets.

6 Conclusion

Natural soundscapes contain valuable ecological information, yet many existing clustering approaches struggle to produce biologically meaningful groupings based on existing CIVIs. In this study, we addressed this limitation by proposing an unsupervised clustering framework that integrates tunable segmentation, feature transformation, and biologically informed clustering. Using recordings from four ecologically distinct sites, we found that step-based, HDBSCAN configurations selected by ecological clustering metrics produced purer and more interpretable clusters than those selected using internal validity metrics. Across the majority of cases, DCP scores aligned well with dominant acoustic patterns and improved the clarity of resulting clusters. Given the biodiversity of the environments tested and the consistency of results across sites, we believe our framework is generalisable. It provides a practical foundation for large-scale ecoacoustic analysis and offers strong potential for supporting future applications in biodiversity monitoring and ecological discovery in large-scale datasets.

References

1. Allaoui, M., et al.: Considerably improving clustering algorithms using umap dimensionality reduction technique: a comparative study. In: Int. Conf. Image and Signal Process. pp. 317–325. Springer (2020)
2. Davies, D.L., et al.: A cluster separation measure. IEEE Trans. Pattern Analysis and Mach. Intell. (2), 224–227 (1979)
3. Gagolewski, M., et al.: Are cluster validity measures (in) valid? Inf. Sci. **581**, 620–636 (2021)
4. Gibb, R., et al.: Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. **10**(2), 169–185 (2019)
5. Mcloughlin, M.P., et al.: Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring. J. R. Soc. Interface **16**(155), 20190225 (2019)
6. Mesaros, A., et al.: Detection and classification of acoustic scenes and events: Outcome of the dcase 2016 challenge. IEEE/ACM Trans. Audio Speech Lang. Process. **26**(2), 379–393 (2018)
7. Napier, T., Ahn, E., Allen-Ankins, S., Lee, I.: An optimised grid search based framework for robust large-scale natural soundscape classification. In: AJCAI. pp. 468–479. Springer (2023)
8. Napier, T., et al.: Leaves: An open-source web-based tool for the scalable annotation and visualisation of large-scale ecoacoustic datasets using cluster analysis. Ecol. Inform. **87**, 103026 (2025)
9. Nieto-Mora, D.A., et al.: Soundscape characterization using autoencoders and unsupervised learning. Sensors **24**(8) (2024)
10. Noble, A.E., et al.: Unsupervised clustering reveals acoustic diversity and niche differentiation in pulsed calls from a coral reef ecosystem. Front. Remote Sens. **5**, 1429227 (2024)
11. Phillips, Y.F., et al.: Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualisation. PLOS ONE **13**(3), 1–27 (03 2018)
12. Rendon, N., et al.: Uncertainty clustering internal validity assessment using fréchet distance for unsupervised learning. Eng. Appl. Artif. Intell. **124**, 106635 (2023)
13. Roe, P., et al.: The australian acoustic observatory. Methods Ecol. Evol. **12**, 1802–1808 (07 2021)
14. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
15. Schneider, S., et al.: Introducing the software case by comparing different clustering methods and audio transformation techniques using animal vocalizations. Animals **12**(16) (2022)
16. Stowell, D., Clayton, D.: Acoustic event detection for multiple overlapping similar sources. In: 2015 IEEE WASPAA. pp. 1–5. IEEE (2015)
17. Xie, J., et al.: Bioacoustic signal classification in continuous recordings: Syllable-segmentation vs sliding-window. Expert Syst. Appl. **152**, 113390 (2020)