

1 **Appendix S1: Supporting information for**

2

3 **Title:** A workflow to optimize spatial sampling in ecoacoustic studies

4

5 **Journal:** Landscape Ecology

6

7 **Authors:** Martínez-Arias, V. M.^{1,3}; Paniagua-Villada^{1,3}, C.; Guerrero, M.J.; Daza, J. M.¹

8

9 **Author affiliations at the time the work was conducted:**

10 1 Grupo Herpetológico de Antioquia GHA, Instituto de Biología, Facultad de Ciencias Exactas y

11 Naturales, Universidad de Antioquia UdeA, Calle 70 No. 52 - 21, Medellín, Colombia.

12 2 SISTEMIC, Facultad de Ingeniería, Universidad de Antioquia UdeA, Medellín,

13 Colombia

14 3 Corporación Merceditas, Calle 3 29A-11, casa 110, Medellín, Colombia

15

16

17 ***Corresponding author:** Victor M. Martínez Arias

18 Email: vmanuel.martinez@udea.edu.co

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

SUPPLEMENTARY INFORMATION S1: Detailed methods for landscape proxies construction

We gathered Sentinel-2A harmonized images for each locality's bounding box using the Google Earth Engine platform (Gorelick et al. 2017). Given the availability of Sentinel-2A images, we defined a specific temporal window for each bounding box and applied a 10% cloud cover filter. Next, we generated a best-pixel mosaic for each Sentinel-2A image set (Gorelick et al. 2017; Google Earth Engine 2023a), using an NDVI-based quality mosaic algorithm. Using the generated mosaic, we constructed two types of landscape proxies: continuous and discrete.

The construction of continuous landscape proxies aimed to reflect three key landscape attributes: composition, structure, and waterbody presence (Lymburner et al. 2000; Kontsiotis et al. 2023; Patle and Ghuge 2024). To do so, we selected three spectral indices specifically related to these attributes (Table 2) that are commonly used in scientific literature and then calculated them for each bounding box.

The next step involved performing an image fusion procedure to construct a single landscape proxy. To achieve this, we scaled the selected indices based on their significance in representing landscape complexity (Table 2), using a 1-100 scale to prevent any single variable with larger variance from dominating the resulting axes (Abdi and Williams 2010). Using the scaled indices, we then performed image fusion through Principal Component Analysis (PCA) (Mishra et al. 2014). We selected PCA because it provides a concise and optimal dataset representation (Sahu and Parsai 2012), making it particularly useful for satellite imagery analysis (Mishra et al. 2014). We obtained two Principal Components (PCs), which explained a percentage of image variability. As PC1 consistently explained the highest variability (>90%) in each bounding box, we selected it as the continuous landscape proxy for the corresponding locality.

Table S1-1. Spectral indices derived from satellite data used to construct the landscape proxies on each location.

Index	Formula	Purpose	Scaling reference value	
			1	100
Normalized Difference Vegetation Index (NDVI)	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$	Measures vegetation health and biomass by comparing near-infrared (NIR) and red light reflectance (Kriegler et al. 1969).	Bare soil	Forest units
Specific Leaf Area	NIR / SWIR	Estimates an approximate of the leaf	Less leaf	More leaf

Vegetation Index (SLAVI)		area (Lymburner et al. 2000).	area	area
Modified Normalized Difference Water Index (MNDWI)	$(\text{Green} - \text{SWIR}) / (\text{Green} + \text{SWIR})$	Improves water body extraction by reducing built-up land noise and enhancing water features, often used in flood monitoring and surface water mapping (Xu 2006).	Less humidity	Higher humidity, body water presence

64

65 Discrete landscape proxies were obtained by implementing a supervised classification using
66 Sentinel-2A bands. We performed the classification using the random forest algorithm in Google Earth
67 Engine (Gorelick et al. 2017; Google Earth Engine 2023a), defining four land cover units: Forest,
68 Built-up, Grass, and Tall Grass. The Forest land cover unit consists of dense vegetation with tall trees
69 forming a closed canopy. The Built-up land cover unit includes areas dominated by human-made
70 structures. The Grass land cover unit is characterized by a mix of herbaceous plants with occasionally
71 scattered trees. The Tall Grass land cover unit is dominated by extensive areas of tall herbaceous plants,
72 often reaching several meters in height.

73

74 Then, we involved generating sample points, starting with 10 per land cover class. We adjusted the final
75 number until classification accuracy approached 80%, measured as the proportion of correctly classified
76 pixels within land cover groups (Cansler and McKenzie 2012; Hejmanowska et al. 2021). We assessed
77 accuracy using a confusion matrix, implemented through the errorMatrix command in Google Earth
78 Engine (Gorelick et al. 2017; Google Earth Engine 2023b), for each classification generated within its
79 respective bounding box.

80

81

SUPPLEMENTARY INFORMATION S2: Detailed methods for soundscape estimation

Following the landscape proxies construction, we generated the soundscape proxies. After retrieving the recorders, we processed the recordings from each bounding box using an algorithm based on Power Spectrum Density (PSD) between 600 and 1200 Hz (Bedoya et al. 2017). PSD was calculated using the Welch technique, which segments the signal, applies a window to reduce discontinuities, performs a Fourier transform, and averages the periodograms (Bedoya et al. 2017). This process is used to identify and exclude corrupted and rainfall files, producing a filtered set of recordings that varied across sampling sites.

Soundscape properties were subsequently quantified using a set of acoustic indices implemented primarily through the Scikit-Maad Python package (Ulloa et al. 2021), complemented by custom routines for specific metrics. All indices were computed either directly from the PSD or from spectrogram-derived representations, ensuring consistency in time-frequency resolution across recordings. Consequently, the indices describe patterns of acoustic energy distribution and variability in the time-frequency domain, rather than explicitly identifying sound sources or taxa.

Specifically, we selected five acoustic indices, Acoustic Complexity Index in its temporal (ACItf) and spectral (ACIf) formulations, Bioacoustic Index (BI), Normalized Difference Soundscape Index (NDSI), and Number of peaks (NP) (see S2-1), to capture complementary aspects of soundscape structure, including temporal modulation, spectral variability, energy concentration, and spectral fragmentation (Sánchez-Giraldo et al. 2021; Luna-Naranjo et al. 2024). Although some of these indices are commonly interpreted in relation to biophonic or anthropogenic activity, such interpretations are conceptual generalizations rather than explicit source classifications. Accordingly, all indices were treated here as operational proxies of soundscape structure. In particular, indices based on predefined frequency bands (BI and NDSI), were interpreted as relative indicators of spectral dominance, acknowledging that frequency overlap among biological, anthropogenic and abiotic sounds is common and that source separation is not explicit.

Except for BI and NDSI, which rely on standard frequency ranges to preserve comparability with previous studies, we calculated the indices for the 0-24, 0-12, and 12-24 kHz bands to consider scale variation at the frequency level. We considered the temporal scale using the hour ranges defined by Rendon et al. (2025): 05:00-08:00 (dawn), 17:00-20:00 (evening), and the complete day. For each recorder, multiple recordings within each period were summarized using the median value of each index, providing a robust representation of site-level soundscape properties for subsequent spatial interpolation and analysis.

For the ultrasonic band (24-96 kHz) we calculated Number of peaks (NP) Bioacoustic Index (BI) and the Acoustic Complexity Index (ACItf, ACIf). We remark that the Acoustic Complexity Indices (ACItf,

ACIf_t) were computed using different spectral representations depending on the frequency domain, reflecting a precautionary and domain-specific analytical strategy. For the audible range, ACI's were calculated from power spectrograms (PSD). This choice was made to emphasize energetic contrasts in the audible spectrum and to evaluate the sensitivity of spatial sampling designs under the most conservative and contrasting conditions, thereby maximizing the detection of differences among sampling strategies. Power-based representations also ensured numerical stability and methodological consistency with other energy-based indices used in this study (e.g., BI, NP, and NDSI).

In contrast, for the ultrasonic range, ACI's were computed from amplitude spectrograms rather than power. This decision was motivated by the properties of the ACI formulation and by the need to avoid spurious inflation of variability and false positives in high-frequency, low-energy signals. In this case, calculations followed the amplitude-based formulation with edge-effect correction as proposed by Farina and Li (2022), including temporal clumping and boundary correction, which are better suited for ultrasonic signals characterized by sparse, transient acoustic events.

Table S2-1. Summary of selected acoustic indices used in soundscape proxy interpolations, their definitions, applications, and methodological considerations

Index	Description	Applications	Frecuencies
Acoustic Complexity Index (ACIf_t)	Quantifies temporal variability in acoustic energy by measuring frame-to-frame changes in amplitude over time within a defined frequency band. Although originally proposed to capture biological activity, ACIf _t reflects general temporal modulation and may incorporate biotic, anthropogenic, or abiotic sounds depending on their temporal dynamics. Computed from the power spectral density, with no border correction for audible spectrum. Calculated from amplitude and border correction for ultrasound spectrum (Pieretti et al. 2011; Farina and Li 2022).	Indicator of temporal soundscape activity and variability in ecoacoustic and environmental monitoring studies (Alcocer et al. 2022).	A, U
Acoustic Complexity	Quantifies spectral variability by measuring changes in	Characterization and spatial and spectral	A, U

Index (ACIf)	acoustic energy across frequency bins at each temporal step. This index captures patterns of spectral modulation rather than sound source identity. Computed from the power spectral density, with no border correction for audible spectrum. Calculated from amplitude and border correction for ultrasound spectrum (Farina and Li 2022).	soundscape structure across landscapes (Alcocer et al. 2022).	
Bioacoustic Index (BI)	Measures the area under the mean frequency spectrum within a predefined band commonly 2–8 kHz or 2–12kHz, reflecting concentration of acoustic energy in mid-frequencies. Although often associated with biophonic activity, BI integrates all sound sources present in the selected band (Boelman et al. 2007). We also calculated it for ultrasonic range (24-96 kHz)	Effective for analyzing biophonic sounds, robust to microphone gain (Alcocer et al. 2022).	A, U
Normalized Difference Sound Index (NDSI)	Compares acoustic energy between two predefined frequency bands, traditionally interpreted as biophony and anthropophony. This separation is theoretical, as both biological and anthropogenic sounds can occur across overlapping frequencies. NDSI therefore represents a relative measure of spectral dominance rather than a strict source classification (Kasten et al. 2012).	Assessment of relative anthropogenic influence and large-scale soundscape gradients, with context-dependent interpretation (Alcocer et al. 2022).	A
Number of	Counts the number of prominent peaks in the mean	Description of spectral complexity	A, U

Peaks (NP)	frequency spectrum, within a defined frequency range, reflecting spectral fragmentation or tonal richness. NP i sensitive to the presence of multiple frequency components but does not discriminate among sound sources or taxa	and heterogeneity in comparative and spatial soundscape analyses.	
-------------------	--	---	--

138 Selected acoustic indices. A=Audible bands; U= Ultrasound bands.

139

SUPPLEMENTARY INFORMATION S3: Detailed information on SLIC and Watershed tessellation construction

We performed sub sampling from the original grid arrangement by removing points using two different criteria: random and stratified. To obtain random scenarios, we selected 20%, 40%, 60%, and 80% of the original grid and iterated this procedure 50 times for each scheme. For the stratified scenarios, we employed four different landscape segmentation approaches: landcover, SLIC (Simple Linear Iterative Clustering), and watershed-derived Basins and Halfbasins. We used the landcover segmentation obtained from the previously performed landcover reclassification, which served as the discrete landscape proxy.

Although the landcover reclassification is a segmentation itself, it does not define a minimum area size and is used as-is. Then, our next step was to define this area value by taking our distance of Audiomoth recorders influence as base parameter ($r = 200\text{m}$), then we calculate the total area ($A_i = 125644\text{m}^2$). We used this value for the other heterogeneity-based segmentation methods we applied: SLIC and watershed-derived tessellation.

The SLIC tessellation procedure is widely used in remote sensing (Achanta et al. 2012; Nowosad and Stepinski 2021). Is a spatially constrained variant of k-means, designed to segment an image into visually homogeneous regions (Achanta et al. 2012; Nowosad and Stepinski 2021). This method assigns search windows to regularly spaced centers, combining spatial and value distances for each cell (Nowosad and Stepinski 2021). We took the area (A_i) and divided it by the total bounding box area to calculate the K parameter, and then, construct the tessellation using the supercells R package (Nowosad & Stepinski, 2021).

To complement SLIC segmentation, we applied watershed tessellation, a method widely used for ecosystem boundary definition, flood risk assessment, and image analysis (Bormann and Likens 1979; Palomino and Hilares 2011; Gamarra et al. 2019; Moreira and Dos Santos 2020). This method is based on mathematical morphology and segments images by grouping pixels according to proximity, gradient, and texture homogeneity (Palomino and Hilares 2011). Here, we used the area (A_i) to define the basin threshold in pixels and divided this value by the continuous landscape proxy pixel size (100m^2). We then applied the `r.watershed` algorithm in GrassGIS to compute watersheds (GRASS Development Team 2022). We selected this algorithm because it generates not only Basins, but also generates Halfbasins, providing an additional tessellations for our analysis (Moreira and Dos Santos 2020).

176 **SUPPLEMENTARY INFORMATION S4: detailed information**
177 **regarding spatial interpolation of acoustic indices**

178 With the full grid and the selected subsamples, we generated soundscape surfaces using spatial kriging
179 interpolation of the acoustic indices median. Kriging is a variogram-based interpolation method that
180 accounts for spatial autocorrelation and data trends, allowing us to assess the effectiveness of different
181 sampling designs in capturing spatial variations (Zarco-Perello and Simões 2017; Amelia et al. 2023).
182 The first step of kriging interpolation consists in compute an empirical variogram to describe how spatial
183 dependence varies with distance, based on variance between paired observations. We did this using the
184 Hawkins-Cressie estimator with a Matérn model, applying a fixed range of 200m based on Audiomoth
185 influence. We selected the Hawkins-Cressie estimator for its robustness to outliers in noisy soundscape
186 data and the Matérn model for its flexibility in modeling spatial correlations, which is essential for
187 capturing soundscape variability (Cressie and Hawkins 1980; Fernández Casal and Cotos Yáñez 2019).
188 As result, we obtain the nugget (microscale variation) and the sill (total variance) parameters.

189
190 The second step involved using the generated nugget and sill parameters to calculate the theoretical
191 variogram with the Matérn model. To achieve this, we constructed a 10m resolution grid and performed
192 ordinary kriging interpolation, applying cross-validation to determine the Root Mean Square Error
193 (RMSE) (Fernández Casal and Cotos Yáñez 2019). We used the Scikit-GStat Python package to calculate
194 the empirical variogram and the GStat R package for the theoretical variogram and spatial interpolation
195 (Gräler et al. 2016; Mälicke 2022).

196
197

SUPPLEMENTARY INFORMATION S5: Associated statistics with objective 1: Evaluation of the sample size effect over sampling strategies

Sampling percent for each tessellation method

Table S5-1 reports the mean and standard deviation of the percentage of recorder locations retained by each subsampling strategy relative to the original sampling grid, summarized by locality. Values integrate variation across acoustic indices, frequency bands, and time periods. Minor differences among strategies and localities reflect quality-control filtering of recordings prior to temporal aggregation.

Table S5-1. Percent of the original sample, for each tessellation method as subsampling strategy

Locality	Strategy	Mean Percent of original sample	SD
AndeanRural	Landcover	55.92	0.49
AndeanUrban	Landcover	42.59	0.00
Plains	Landcover	38.87	0.98
AndeanRural	SLIC	72.05	1.66
AndeanUrban	SLIC	92.59	0.00
Plains	SLIC	73.78	1.36
AndeanRural	basins	67.86	2.31
AndeanUrban	basins	92.59	0.00
Plains	basins	59.39	1.50
AndeanRural	halfbasins	72.07	1.24
AndeanUrban	halfbasins	98.15	0.00
Plains	halfbasins	78.09	1.47
AndeanRural	ssample20	18.63	0.66
AndeanUrban	ssample20	18.52	0.00
Plains	ssample20	19.78	0.01
AndeanRural	ssample40	38.99	0.42
AndeanUrban	ssample40	38.89	0.00
Plains	ssample40	39.57	0.01

AndeanRural	ssample60	58.47	0.45
AndeanUrban	ssample60	59.26	0.00
Plains	ssample60	59.35	0.02
AndeanRural	ssample80	78.83	0.60
AndeanUrban	ssample80	79.63	0.00
Plains	ssample80	79.14	0.02

Normality assessment for pearson correlation representing the sample size effect, between the performance and sample percent

To quantify the effect of sample size on sampling design, we computed another Pearson coefficient, (Pcoeff2) between the percentage of sampled data for each subsample, and the previously calculated subsample performance. To assess effects different from those of the subsample design, we calculated this new Pearson coefficient, across combinations of locality, strategy, frequency range, time period, acoustic index, and summary statistic using the `ggstatsplot` package (Patil 2021). We did this, since we aimed to take into account with varying ecological and methodological conditions. To assess the statistical properties of the Pearson coefficient distributions representing the effect of sample size, we applied the Shapiro-Wilk test for normality within each sub sampling strategy using the `rstatix` package (Kassambara 2019). Since most distributions deviated from normality, we proceeded with non-parametric statistical tests (Supplementary information S1). sub sampling strategies were grouped into two major categories: "random" and "tessellation". We compared the sample size effect Pearson's values between these groups using the Wilcoxon rank-sum test and quantified differences in dispersion using the Fligner-Killeen test, which is robust to non-normal distributions (Sarkar et al. 1999).

Table S5-2. Shapiro-Wilks test for sample size effect for evaluate normality in the size dependence of the original sample.

NORMALITY TEST FOR SAMPLE SIZE EFFECT				
Calc_stat	Subsample	variable	statistic	p
median	Landcover	estimate	0.757726226	4.30567E-06
median	SLIC	estimate	0.918701454	0.007932343
median	basins	estimate	0.840735609	6.53985E-05
median	halfbasins	estimate	0.553122704	5.57313E-10
median	ssample20	estimate	0.95751788	0.120301316
median	ssample40	estimate	0.965232823	0.226478957

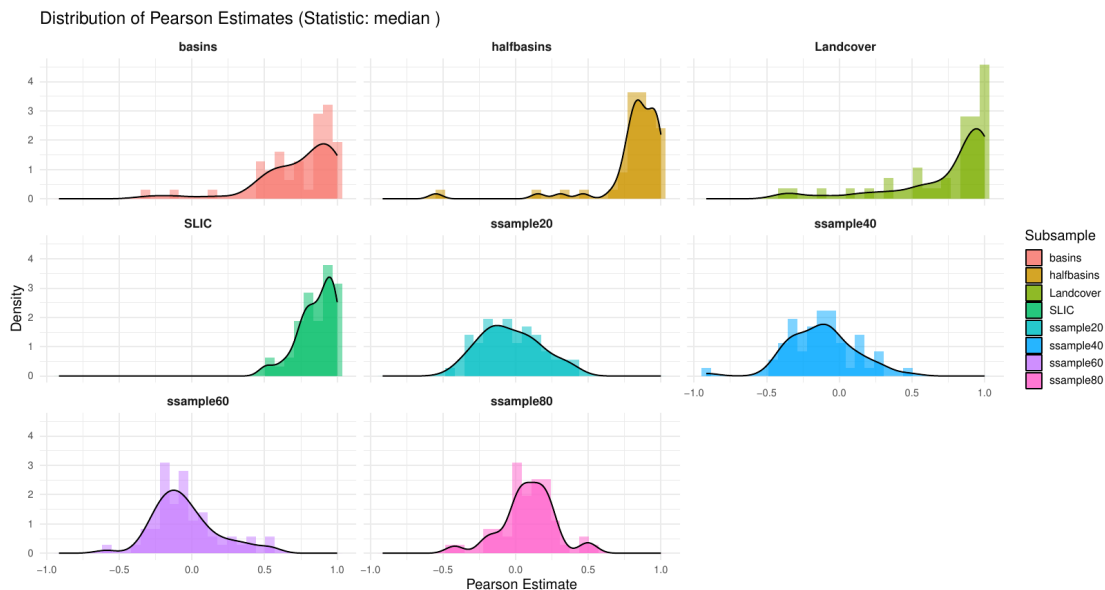
median	ssample60	estimate	0.959536953	0.142150633
median	ssample80	estimate	0.900798438	0.001521112

232

233 This table presents the results of Shapiro-Wilk tests applied to evaluate whether the distributions of the sample size effect (i.e.,
234 correlation between sample proportion and subsample performance) follow a normal distribution, across different sub sampling
235 strategies and random sample sizes. Values are based on median statistics. Most tessellation-based strategies (Landcover, SLIC,
236 Basins, Halfbasins) show significant deviations from normality ($p < 0.01$), justifying the use of non-parametric tests in subsequent
237 comparisons. In contrast, random sampling levels at 20%, 40%, and 60% did not significantly deviate from normality.

238

239



240

241 **Figure S5-1** Pearson coefficient (Pcoeff2) value distributions representing sampling size effect for each
242 subsample type, obtained with the median values of the original subsample performance.

243 Density plots show the distributions of Pearson correlation values calculated between sampling proportion and subsample
244 performance, stratified by subsample type. Tessellation-based strategies (basins, halfbasins, landcover, SLIC) and random sampling
245 levels (20%, 40%, 60%, 80%) are shown using kernel density estimation. Results are based on median performance values. Non-
246 normality is visually evident for most tessellation strategies, with skewed and mult imodal distributions, while random sub sampling
247 levels appear more symmetrically distributed. These patterns support the application of non-parametric tests for group comparisons
248 (see Table S5-1).

249

250

251

SUPPLEMENTARY INFORMATION S6: Detailed methods for the developemnt of the specific objective 2: Effectiveness of stratified sampling designs evaluation

We evaluated the effectiveness of stratified sampling strategies using two metrics: subsample performance and sampling efficiency. Subsample performance was previously defined as the Pearson correlation (ρ) between the interpolated surface from each subsample and the corresponding full grid. To quantify how well each strategy balances performance and effort, we calculated sampling efficiency as the ratio of performance to sampling proportion (Subsample performance / %sampled). This approach allowed us to identify designs that not only reproduce spatial patterns effectively, but also do so with fewer sampling units. For both metrics, we computed summary statistics (mean, median, standard deviation, and coefficient of variation). We did this to capture not only the central tendency of each strategy's performance, but also its variability and reliability across iterations and conditions. This allowed us to assess whether high performance was consistent or driven by outliers, and to identify strategies that offered both effectiveness and robustness.

To assess the need for further subgroup analyses by factors such as time period, locality, or frequency range, we evaluated summary statistics for each tessellation strategy. Given the consistency in performance metrics across strategies (median Pearson > 0.89 , low to moderate coefficient of variation), we did not perform additional subgroup analyses by time period, frequency range, or locality (See supplementary information). The next step consisted in a normality assessment. To do so, we took separately the distribution of performance and the efficiency and tested individually for normality within each subsample group using the Shapiro-Wilk test (`shapiro_test()` from the `rstatix` package). Then, we generated violin plots with overlaid boxplots to compare distributions across strategies. Due to non-normality of our data, the differences between sub sampling strategies were evaluated using Kruskal-Wallis rank-sum tests, followed by Dunn's post hoc tests with Benjamini-Hochberg adjustment for multiple comparisons. All analyses were performed for both median and mean using R 4.4.0 "Puppy Cup", with the following packages: `dplyr` for data wrangling, `ggplot2` and `ggpubr` for visualization, `rstatix` for statistical tests, and `FSA` for post hoc comparisons (Wickham 2011; Wickham et al. 2014; Kassambara 2023; R Core Team 2024; Ogle et al. 2025).

Table S6-1 Normality assessment for tile's performance (raster's Pearson correlation)

Evaluated	Raster's Comparison Statistic	Subsample Type	Statistic	p
Subsample performance	MEDIAN	Landcover	0.715326335	8.28E-08
	MEDIAN	SLIC	0.904062478	8.53E-04
	MEDIAN	basins	0.805484608	2.11E-06
	MEDIAN	halfbasins	0.587809278	1.29E-10

Normality was assessed for each tessellation strategy using Pearson correlation values between interpolated surfaces and full sampling grids. Results are based on median summary statistics of subsample performance. All strategies except Landcover exhibited significant deviations from normality ($p < 0.05$), justifying the use of non-parametric analyses in subsequent tests.

Table S6-2 Results of Kruskal-Wallis Test on the raster precision of sampling strategy (Sub sample performance)

Evaluated	Evaluated Metric	Raster's Comparison Statistic	n	statistic	df	p
Audible	Subsample performance	MEDIAN	489	194.444517	3	6.69E-42
Ultrasound	Subsample performance	MEDIAN	107	53.9936393	3	1.13E-11

A Kruskal-Wallis rank-sum test was performed to assess whether subsample performance differed significantly across tessellation-based strategies (basins, halfbasins, SLIC, landcover). Median performance values were used. The results revealed strong overall differences among groups ($\chi^2 = 194.44$, $p < 0.001$).

Table S6-3 Results of Kruskal-Wallis Test on the efficiency of sampling strategy (Subsample performance /Sample%)

Evaluated	Evaluated Metric	Raster's comparison Statistic	n	statistic	df	p
Audible	Efficiency	MEDIAN	489	187.7701344	3	1.85E-40
Ultrasound	Efficiency	MEDIAN	107	42.1535577	3	3.72E-09

d						
---	--	--	--	--	--	--

301
302 This test evaluated differences in sampling efficiency—defined as the ratio of subsample performance to sampling proportion—
303 across the four tessellation strategies. Results based on median values indicated significant variation among strategies ($\chi^2 =$
304 187.77, $p < 0.001$), supporting subsequent pairwise post hoc analyses.

305 **Table S6-4** Post hoc Dunn test results for subsample performance across tessellation strategies.

Frequency	Evaluated Metric	Raster's Comparison Statistic	group1	group2	n1	n2	statistic	p	p.adj	p.adj. Signif.
Audible	Subsample performance	MEDIAN	basins	halfbasins	123	125	3.1677	1.54E-03	1.84E-03	**
			basins	Landcover	123	118	-10.0837	6.52E-24	1.96E-23	****
			basins	SLIC	123	123	-0.5753	5.65E-01	5.65E-01	ns
			halfbasins	Landcover	125	118	-13.2579	4.06E-40	2.44E-39	****
			halfbasins	SLIC	125	123	-3.7453	1.80E-04	2.70E-04	***
			Landcover	SLIC	118	123	9.5145	1.83E-21	3.65E-21	****
Ultrasound	Subsample performance	MEDIAN	basins	halfbasins	26	27	1.5283	1.26E-01	1.90E-01	ns
			basins	Landcover	26	27	-5.0740	3.89E-07	7.79E-07	****
			basins	SLIC	26	27	0.6856	4.93E-01	4.93E-01	ns
			halfbasins	Landcover	27	27	-6.6655	2.64E-11	1.58E-10	****
			halfbasins	SLIC	27	27	-0.8507	3.95E-01	4.74E-01	ns
			Landcover	SLIC	27	27	5.8148	6.07E-09	1.82E-08	****

306
307 Pairwise comparisons were conducted using Dunn's test with Benjamini-Hochberg correction following a significant Kruskal-Wallis test on subsample performance (median Pearson correlation). The results indicate
308 that Landcover significantly underperformed all other strategies, while Basins and SLIC showed no significant difference ($p_{adj} > 0.05$). Asterisks indicate the level of significance after adjustment.

309
310 **Table S6-5** Post hoc Dunn test results for sampling efficiency across tessellation strategies (Subsample performance/Sample%)

Frequency	Evaluated Metric	Raster's Comparison Statistic	group1	group2	n1	n2	statistic	p	p.adj	p.adj. Signif.
Audible	Efficiency (S. performance/Sample)	MEDIAN	basins	halfbasins	123	125	-4.0125	6.01E-05	7.21E-05	****
			basins	Landcover	123	118	7.7825	7.11E-15	1.42E-14	****

	%)		basins	SLIC	123	123	-4.2995	1.71E-05	2.57E-05	****
			halfbasins	Landcover	125	118	11.7835	4.75E-32	1.42E-31	****
			halfbasins	SLIC	125	123	-0.3043	7.61E-01	7.61E-01	ns
			Landcover	SLIC	118	123	-12.0371	2.27E-33	1.36E-32	****
Ultrasound	Efficiency (S. performance/Sample %)	MEDIAN	basins	halfbasins	26	27	-2.0266	4.27E-02	6.40E-02	ns
			basins	Landcover	26	27	3.7678	1.65E-04	3.29E-04	***
			basins	SLIC	26	27	-1.5445	1.22E-01	1.47E-01	ns
			halfbasins	Landcover	27	27	5.8499	4.92E-09	2.95E-08	****
			halfbasins	SLIC	27	27	0.4868	6.26E-01	6.26E-01	ns
			Landcover	SLIC	27	27	-5.3631	8.18E-08	2.45E-07	****

311

312 This table presents pairwise comparisons of sampling efficiency (subsample performance divided by sample proportion) among tessellation strategies. Dunn’s test with Benjamini-Hochberg adjustment was used

313 following a significant Kruskal-Wallis result. Basins consistently outperformed SLIC and Halfbasins in efficiency, while Landcover showed the highest values but also the greatest variability. Only the comparison

314 between Halfbasins and SLIC was not statistically significant.

SUPPLEMENTARY INFORMATION S7: Detailed methods for the development of the specific objective 3: Multivariate structure of stratified sub sampling strategies evaluation

To evaluate the multivariate structure and the influence of multiple variables in shaping the effectiveness of different stratified sampling designs, we applied a Factorial Analysis of Mixed Data (FAMD) using the FactoMineR package in R (Lê et al. 2008). FAMD was selected because our dataset includes both categorical variables (Locality, Index, Subsample, Frequency Range, with Subsample treated as a supplementary variable) and continuous variables (e.g., Sample percent, Subsample performance), which it integrates simultaneously within a unified ordination framework. FAMD combines the principles of Principal Component Analysis (PCA) for continuous data and Multiple Correspondence Analysis (MCA) for categorical data, allowing for the identification of underlying dimensions that best explain the global variance structure. Each resulting dimension captures a proportion of the total variance and is associated with the contribution of specific variables (Kassambara 2017).

We observed that Dimension 1 was largely driven by variables related to sampling performance (Sample percent, Subsample performance), and thus reflects outcome effectiveness rather than structural differences between sampling strategies (Supplementary information S3). Consequently, although we analyzed Dim 1 too, our comparative analysis of sampling strategies focused on subsequent dimensions (e.g., Dim 2 to Dim 4), which were more strongly influenced by design-related and landscape-related variables (e.g., Index, Locality, Time_Period). To statistically assess differences in the latent FAMD dimensions across sub sampling strategies (Basins, Halfbasins, SLIC, Landcover), we performed Kruskal-Wallis tests independently on each relevant dimension. When a significant effect was detected ($p < 0.05$), we followed up with post-hoc pairwise comparisons using Dunn's test with Benjamini-Hochberg correction to adjust for multiple testing, as implemented in the rstatix package in R (Chen and Sarkar 2020; Ogle et al. 2025). Significance groupings were also extracted and visualized to facilitate interpretation of strategy-specific contrasts, to provide a clear overview of which sub sampling strategies differed structurally across FAMD dimensions for both mean and median statistics.

343 **Table S7-1** Coordinates of continuous variables in FAMD space.

Frequency range	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Variable
Audible	0.85186137 5	0.145319259	-0.14975512 9	-0.04142774 1	-0.01517612 6	Sample_percent
Audible	0.75662575 6	0.027853896	0.309265451	0.076822303	0.036161759	Subsample_performance
Ultrasound	0.95174460 9	-0.00607177 1	0.025013897	-0.01499731 7	-1.55534E-2 3	Sample_percent
Ultrasound	0.92327655 7	0.007009913	-0.00378703 8	0.020140896	3.3756E-16	Subsample_performance

344 Numeric variables (Sample percent and Subsample performance) and their loading values across the first five FAMD dimensions.

345
346
347 **Table S7-2** Coordinates of categorical variable categories in FAMD space.

Categoria	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Categoria
Audible	-0.4069352 76	-0.0201889 46	-0.2795732 64	0.7893254 07	0.2791168	AndeanRural
Audible	1.2075300 75	0.2488933 75	-0.3961750 78	-0.2479037 4	-0.1766759 8	AndeanUrban
Audible	-0.8437067 85	-0.2404121 78	0.7081939 97	-0.5638396 92	-0.1058611 35	Plains
Audible	-0.2797596 93	1.5058891 24	0.0034196 48	0.0089745 48	0.0058273 82	R0-2
Audible	0.2353347 56	-0.7977920 58	0.0936752 38	-0.4876170 78	1.0893354 01	R0-24
Audible	0.0434480 51	-0.8072040 02	-0.2139121 79	-0.1190796 86	-0.1894198 71	R12-24
Audible	0.1700939 91	-0.8024500 62	0.1165547 94	0.5833237 99	-0.8888125 78	R2-12
Audible	0.1248877 51	0.0069176 41	0.1933295 76	0.3213229 25	1.0194083 6	Dawn
Audible	-0.0853049	0.0096312	-0.3778623	-0.9638335	-0.4524623	Evening

Categori a	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Categoria
	18	39	85	67	16	
Audible	-0.0405980 84	-0.0166937 36	0.1844785 07	0.6444932 82	-0.5767383 55	Full
Audible	0.2763580 92	-0.3693062 34	0.8379688 04	0.2639201 95	-0.2151301 52	ACIf
Audible	0.1228994 31	-0.3876972 35	0.4207510 64	-0.4262832 14	0.0934080 64	ACIf
Audible	-0.4809168 04	2.6128449 44	-0.0213390 35	-0.0952402 23	-0.0380274 96	BI
Audible	-0.4324067 1	2.6151751 35	0.1017444 18	0.0760840 87	0.0750430 22	NDSI
Audible	-0.1756609 23	-0.6051010 19	-1.3492152 71	0.1791723 11	0.1174391 6	NP
Ultrasou nd	-0.3604525 84	-0.0100902 39	-1.1701623 97	0.1489310 84	-0.0099818 31	AndeanRur al
Ultrasou nd	1.2726760 49	-0.0494045 42	0.3102220 81	-0.1842240 72	0.0099818 31	AndeanUr ban
Ultrasou nd	-0.9382869 93	0.0611946 32	0.8845100 39	0.0363013 6	1.48512E- 14	Plains
Ultrasou nd	0.0730104 78	-0.4771186 92	0.3406504 88	1.2207590 83	-0.0925464 13	Dawn
Ultrasou nd	0.1130746 64	1.0032857 31	-0.5588831 95	-0.0354030 43	1.08235E- 15	Evening
Ultrasou nd	-0.1829441 79	-0.4982979 91	0.2027081 74	-1.1863394 58	0.0925464 13	Full
Ultrasou nd	-0.0203644 03	0.5147117 47	0.2419044 47	-0.0881855 05	-1.2155026 27	ACIf
Ultrasou nd	-0.0142878 46	0.5166471 57	0.2402500 85	0.0978230 06	1.2155026 27	ACIf
Ultrasou nd	0.0356423 14	-1.0608263 01	-0.4959303 76	-0.0099128 58	-1.59346E- 14	NP

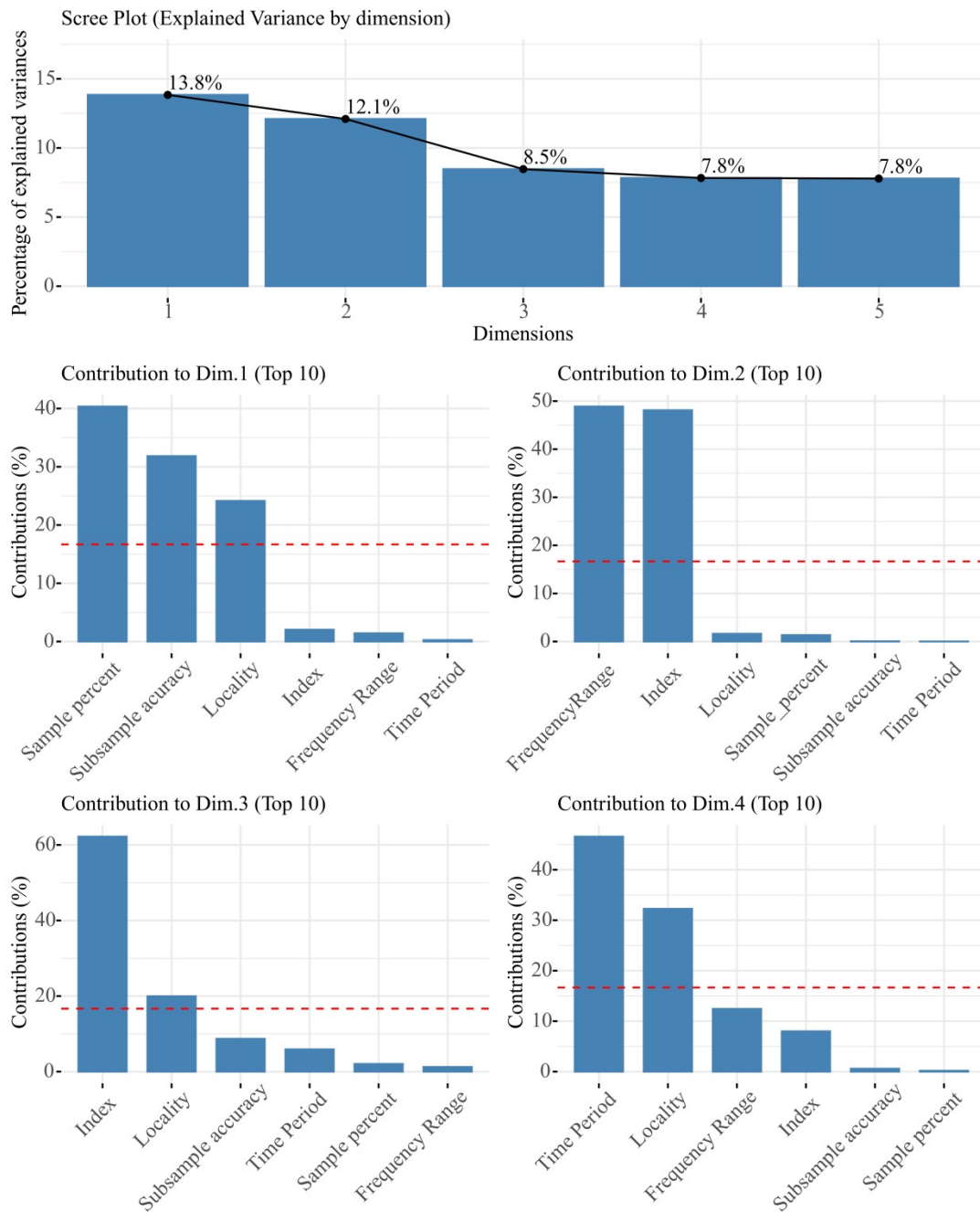
Individual category coordinates (e.g., Locality, Index, Frequency Range) showing their location along FAMD dimensions 1 to 5.

Table S7-3 Explained variance per FAMD dimension

Dimension	Dimension	Dimension	Variance (%)	Cumm. variance (%)
Audible	Dim.1	1.798800207	13.83692467	13.83692467
Audible	Dim.2	1.572358413	12.09506472	25.93198939
Audible	Dim.3	1.100659781	8.466613701	34.39860309
Audible	Dim.4	1.017269671	7.825151318	42.22375441
Audible	Dim.5	1.012287109	7.786823916	50.01057832
Ultrasound	Dim.1	2.170393004	27.12991255	27.12991255
Ultrasound	Dim.2	1.019117891	12.73897363	39.86888618
Ultrasound	Dim.3	1.01204083	12.65051037	52.51939655
Ultrasound	Dim.4	1.000566314	12.50707892	65.02647548
Ultrasound	Dim.5	1	12.5	77.52647548

Proportion of variance (%) captured by each FAMD dimension and the cumulative variance explained across dimensions 1 to 5.

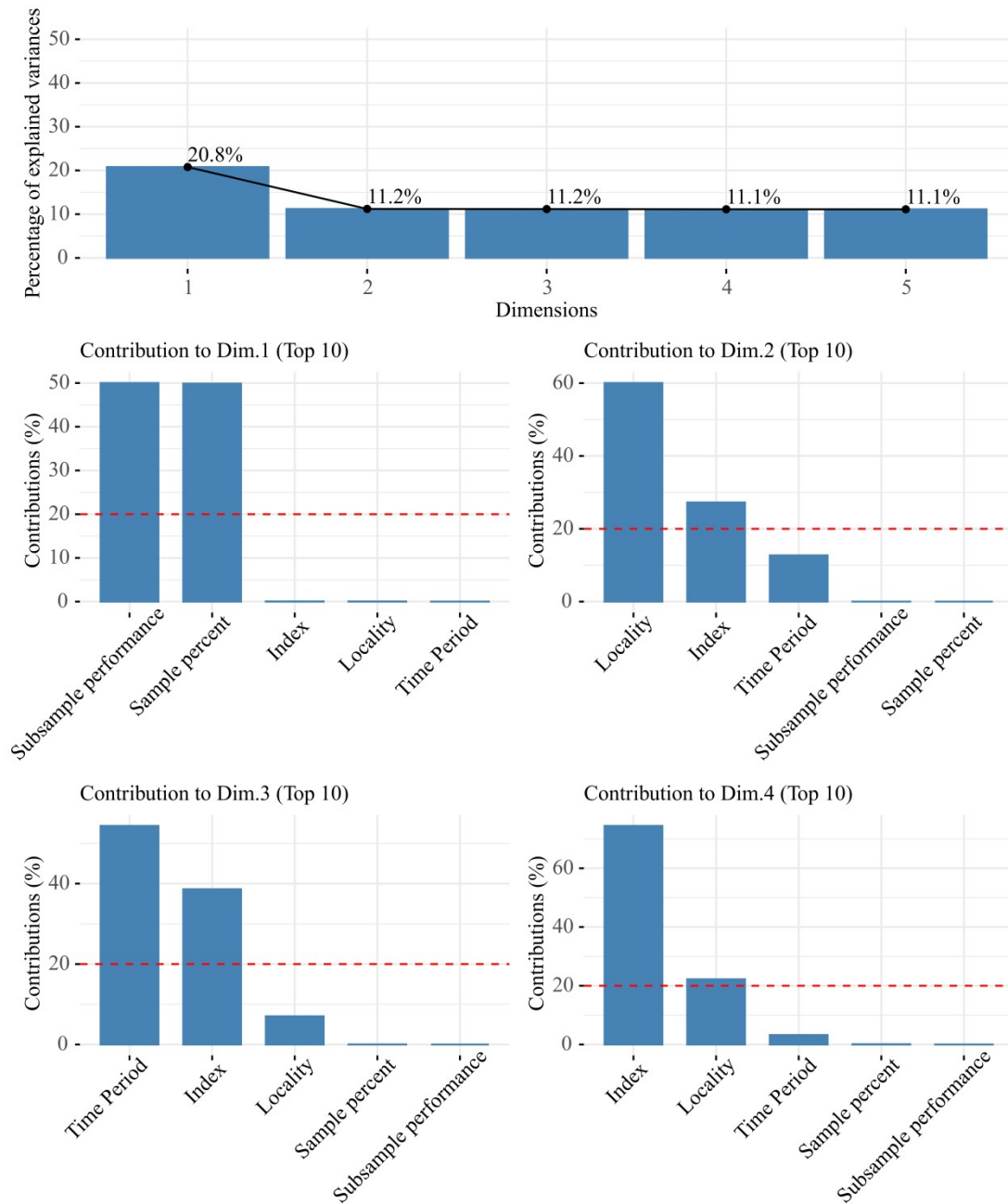
356 **Figure S7-1 FAMD SUMMARY PLOT FOR AUDIBLE FREQUENCYBANDS**
 FAMD – MEDIAN



357 Upper panel: Scree plot of variance explained by each FAMD dimension (Dim.1-Dim.5) for audible bands. Middle & inferior
 358 panels: Variable contributions to dimensions 1 to 4, showing the top contributing continuous and categorical variables.
 359 Dimension 1 is predominantly associated with performance-related variables (e.g., Sample percent, Subsample performance),
 360 while Dimensions 2-4 are shaped by landscape, acoustic, and temporal factors. The dashed red line indicates the expected
 361 average contribution (threshold for relevance).
 362

363
 364
 365
 366
 367
 368

369 **Figure S7-2** FAMD SUMMARY PLOT FOR ULTRASOUND FREQUENCY BANDS
FAMD – ULTRASOUND - MEDIAN



370 Upper panel: Scree plot of variance explained by each FAMD dimension (Dim.1-Dim.5) for ultrasound bands (24-96
371 kHz) .Middle & inferior panels: Variable contributions to dimensions 1 to 4, showing the top contributing continuous and
372 categorical variables. Dimension 1 is predominantly associated with performance-related variables (e.g., Sample percent,
373 Subsample performance), while Dimensions 2-4 are shaped by landscape, acoustic, and temporal factors. The dashed red line
374 indicates the expected average contribution (threshold for relevance).
375

376
377

378 **Table S7-4** Statistical comparison of stratified sub sampling strategies: Dunn's Post-Hoc Test Across FAMD Dimensions for audible frequencies.

Comparisons (Audible frequencies)								
.y.	group1	group2	n1	n2	statistic	p	p.adj	p.adj.signif
Dim.1	basins	halfbasins	123	125	5.655714108	1.55E-08	2.33E-08	****
Dim.1	basins	Landcover	123	118	-10.38615614	2.87E-25	5.73E-25	****
Dim.1	basins	SLIC	123	123	2.97025992	0.002975479	0.003570575	**
Dim.1	halfbasins	Landcover	125	118	-16.02334867	8.78E-58	5.27E-57	****
Dim.1	halfbasins	SLIC	125	123	-2.673501384	0.007506397	0.007506397	**
Dim.1	Landcover	SLIC	118	123	13.32544274	1.65E-40	4.94E-40	****
Dim.2	basins	halfbasins	123	125	2.694278568	0.007054118	0.013372203	*
Dim.2	basins	Landcover	123	118	-2.615304186	0.008914802	0.013372203	*
Dim.2	basins	SLIC	123	123	1.58597351	0.112745319	0.135294382	ns
Dim.2	halfbasins	Landcover	125	118	-5.291569068	1.21E-07	7.28E-07	****
Dim.2	halfbasins	SLIC	125	123	-1.101922845	0.270495216	0.270495216	ns
Dim.2	Landcover	SLIC	118	123	4.18473946	2.85E-05	8.56E-05	****
Dim.3	basins	halfbasins	123	125	-7.061336111	1.65E-12	9.89E-12	****
Dim.3	basins	Landcover	123	118	-5.806997543	6.36E-09	1.91E-08	****
Dim.3	basins	SLIC	123	123	-5.182501206	2.19E-07	4.38E-07	****
Dim.3	halfbasins	Landcover	125	118	1.15723917	0.247174659	0.29660959	ns
Dim.3	halfbasins	SLIC	125	123	1.857979685	0.063171899	0.094757849	ns
Dim.3	Landcover	SLIC	118	123	0.678538494	0.497430326	0.497430326	ns
Dim.4	basins	halfbasins	123	125	-11.2371515	2.68E-29	1.61E-28	****
Dim.4	basins	Landcover	123	118	-6.367047969	1.93E-10	5.78E-10	****
Dim.4	basins	SLIC	123	123	-6.222972021	4.88E-10	9.76E-10	****
Dim.4	halfbasins	Landcover	125	118	4.726907166	2.28E-06	2.74E-06	****
Dim.4	halfbasins	SLIC	125	123	4.989137241	6.06E-07	9.10E-07	****

Dim.4	Landcover	SLIC	118	123	0.20896794	0.834473266	0.834473266	ns
-------	-----------	------	-----	-----	------------	-------------	-------------	----

Pairwise comparisons of sub sampling strategies (basins, halfbasins, Landcover, and SLIC) were performed for each FAMD dimension using Dunn's test with Benjamini-Hochberg correction. Reported values include test statistics, adjusted p-values, and significance levels. Significant differences ($p_{adj} < 0.05$) indicate structural dissimilarities between strategies in the multivariate space.

Table S7-5 Statistical comparison of stratified sub sampling strategies: Dunn's Post-Hoc Test Across FAMD Dimensions for ultrasonic frequencies.

Comparisons (Ultrasonic frequencies)								
Dimension	group1	group2	n1	n2	statistic	p	p.adj	p.adj.signif
Dim.1	basins	halfbasins	26	27	2.756707132	0.005838662	0.008757993	**
Dim.1	basins	Landcover	26	27	-4.849004576	1.24083E-06	2.48165E-06	****
Dim.1	basins	SLIC	26	27	1.46664638	0.142472276	0.170966732	ns
Dim.1	halfbasins	Landcover	27	27	-7.678495297	1.60968E-14	9.65808E-14	****
Dim.1	halfbasins	SLIC	27	27	-1.302406113	0.192777594	0.192777594	ns
Dim.1	Landcover	SLIC	27	27	6.376089184	1.81667E-10	5.45001E-10	****
Dim.2	basins	halfbasins	26	27	-6.781923885	1.18586E-11	7.11516E-11	****
Dim.2	basins	Landcover	26	27	-5.357210327	8.45167E-08	1.69033E-07	****
Dim.2	basins	SLIC	26	27	-6.612521968	3.77827E-11	1.13348E-10	****
Dim.2	halfbasins	Landcover	27	27	1.438347491	0.150335483	0.225503225	ns
Dim.2	halfbasins	SLIC	27	27	0.171023025	0.864205661	0.864205661	ns
Dim.2	Landcover	SLIC	27	27	-1.267324467	0.205039296	0.246047155	ns
Dim.4	basins	halfbasins	26	27	2.212248703	0.026949485	0.053898971	ns
Dim.4	basins	Landcover	26	27	-0.558992913	0.57616655	0.63578359	ns

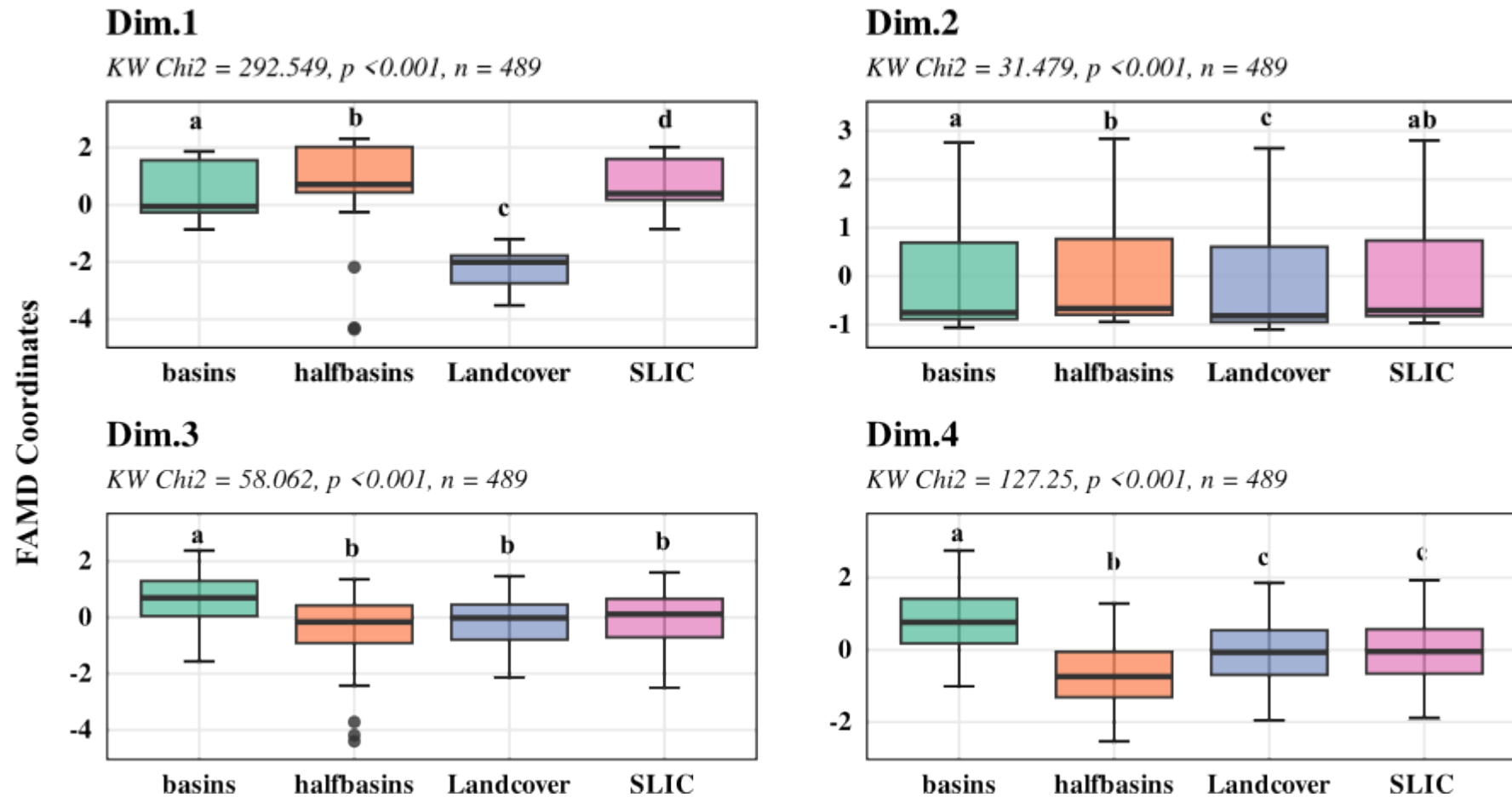
						7		
Dim.4	basins	SLIC	26	27	1.743135702	0.081309914	0.12196487	ns
Dim.4	halfbasins	Landcover	27	27	-2.797761279	0.005145813	0.030874878	*
Dim.4	halfbasins	SLIC	27	27	-0.473602223	0.63578359	0.63578359	ns
Dim.4	Landcover	SLIC	27	27	2.324159056	0.020116971	0.053898971	ns

Table S7-6 Significance Groupings of sub sampling Strategies and Kruskal-Wallis Test results Across FAMD Dimensions

Frequency	Dimension	KW_Statistic	KW_df	KW_n	KW_p_value	Basins	Halfbasins	Landcover	SLIC
Audible	Dim.1	292.549	3	489	<0.001	a	b	c	d
Audible	Dim.2	31.479	3	489	<0.001	a	b	c	ab
Audible	Dim.3	58.062	3	489	<0.001	a	b	b	b
Audible	Dim.4	127.25	3	489	<0.001	a	b	c	c
Ultrasound	Dim.1	67.605	3	107	<0.001	a	b	c	ab
Ultrasound	Dim.2	60.526	3	107	<0.001	a	b	b	b
Ultrasound	Dim.3	3.935	3	107	0.269	a	a	a	a
Ultrasound	Dim.4	10.874	3	107	0.0124	ab	a	b	ab

Letter-based groupings resulting from post-hoc Dunn tests (adjusted with Benjamini-Hochberg correction) for each FAMD dimension. sub sampling strategies sharing the same letter are not significantly different at $p < 0.05$. Distinct letters denote statistically significant structural differences in multivariate space. Summary statistics for Kruskal-Wallis tests evaluating differences among sub sampling strategies across the first four FAMD dimensions. Most of the tests were significant ($p < 0.001$), indicating structural differences in multivariate space depending on the sampling strategy used

390 **Figure S7-3** Kruskal-Wallis + Dunn Tests of each FAMD dimension graphic showing differences in FAMD dimensions across stratified sub sampling strategies for audible
 391 frequencies.
 392



393

394 Boxplots show the distribution of FAMD coordinates for each strategy (Basins, Halfbasins, Landcover, and SLIC) across dimensions 1 to 4, based on the median statistic.
395 Kruskal-Wallis tests indicate significant differences across strategies in all dimensions ($p < 0.0001$). Letters denote significance groupings from Dunn's post-hoc tests with
396 Benjamini-Hochberg correction. Dimension 1 primarily reflects performance variables, while Dimensions 2-4 capture structural variation linked to acoustic index, locality,
397 and temporal stratification.

398

399

400

401

402

403

404

405

406

407

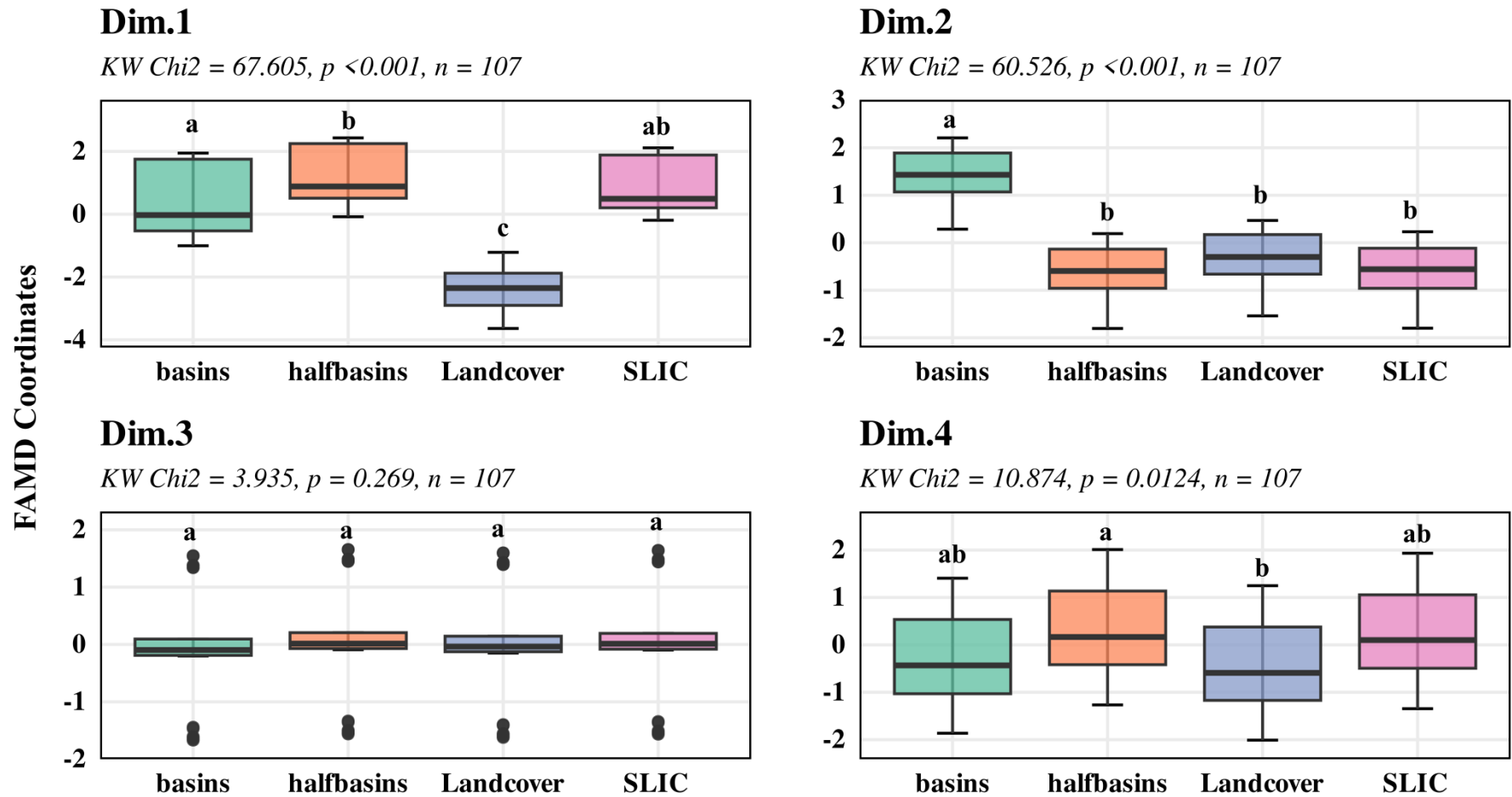
408

409

410

411

412 **Figure S7-3** Kruskal-Wallis + Dunn Tests of each FAMD dimension graphic showing differences in FAMD dimensions across stratified sub sampling strategies for ultrasonic
413 frequencies.



414

415 Boxplots show the distribution of FAMD coordinates for each strategy (Basins, Halfbasins, Landcover, and SLIC) across dimensions 1 to 4, based on the median statistic.
416 Kruskal-Wallis tests indicate significant differences across strategies in most of dimensions ($p < 0.0001$). Letters denote significance groupings from Dunn's post-hoc tests
417 with Benjamini-Hochberg correction. Dimension 1 primarily reflects performance variables, while Dimensions 2-4 capture structural variation linked to acoustic index,
418 locality, and temporal stratification.

419

420

421

422

