7   **Authors:** Martínez-Arias, V. M.[1]; Paniagua-Villada[1], C.; Guerrero, M.J.; Daza, J. M.[1]

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

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

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

19   Tel: 573116797846

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	Scalling reference value	
			1	**100**
Normalized Difference Vegetation Index (NDVI)	(NIR - Red) / (NIR + 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 approxiate of the leaf	Less leaf	More leaf

Vegetation Index (SLAVI)		area (Lymburner et al. 2000).	area	area
Modified Normalized Difference Water Index (MNDWI)	(Green - SWIR) / (Green + 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.

We calculated soundscape properties using acoustic indices, implemented with the Scikit-Maad Python package (Ulloa et al. 2021). Specifically, we selected four acoustic indices (ACIt, ACIf, BI, NDSI; see Table 3), aiming to assess soundscape complexity in terms of spectrogram, envelope, spectrum, and signal level characteristics (Sánchez-Giraldo et al. 2021; Luna-Naranjo et al. 2024). Except for BI and NDSI, which have specific frequency ranges, 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.

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

Index	Description	Applications
Acoustic Complexity Index (ACItf)	Measures the variability in sound intensity across all spectral lines over time, distinguishing biotic sounds from human noise (Pieretti et al. 2011; Farina and Li 2022).	Biodiversity assessment, environmental monitoring (Alcocer et al. 2022).
Acoustic Complexity Index (ACIft)	Measures the variability in sound intensity for each spectral line at every temporal step (Farina and Li 2022).	Useful for detecting sound patterns in soundscapes and ecoacoustic research (Alcocer et al. 2022).
Bioacoustic Index (BI)	Evaluates the area under the frequency spectrum curve (typically 2-8 kHz), originally for bird abundance but adaptable for other taxa (Boelman et al. 2007).	Effective for analyzing biophonic sounds, robust to microphone gain (Alcocer et al. 2022).
Normalized Difference Sound Index (NDSI)	Compares biophony (2-12 kHz) to anthrophony (1-2 kHz) to assess anthropogenic disturbance (Kasten et al.	Effective for assessing human impact on soundscapes across

		2012).	ecosystems (Alcocer et al. 2022).

103

104

SUPPLEMENTARY INFORMATION S3: Detailed information on SLIC and Watershed tesellation 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 thelandcover reclassification is a segmentation itlself, 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 ($r = 200m$), then we calculate the total area ($A_{i} = 125644m^2$). We used this value for the other heterogeneity-based segmentation methods we applied: SLIC and watershed-derived tesellation.

The SLIC tesellation 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 tesellation 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).

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

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

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

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

Normality assessment for pearson correlation representing the samplie size effect, between the performanceand 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 S5-1. Shapiro-Wilks test for sample size effect for evaluate normality in the size dependence of the original sample.

Calc_stat	Subsample	variable	statistic	p
NORMALITY TEST FOR SAMPLE SIZE EFFECT
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

This table presents the results of Shapiro-Wilk tests applied to evaluate whether the distributions of the sample size effect (i.e., correlation between sample proportion and subsample performance) follow a normal distribution, across different sub sampling strategies and random sample sizes. Values are based on median statistics. Most tessellation-based strategies (Landcover, SLIC,

Basins, Halfbasins) show significant deviations from normality (p < 0.01), justifying the use of non-parametric tests in subsequent comparisons. In contrast, random sampling levels at 20%, 40%, and 60% did not significantly deviate from normality.

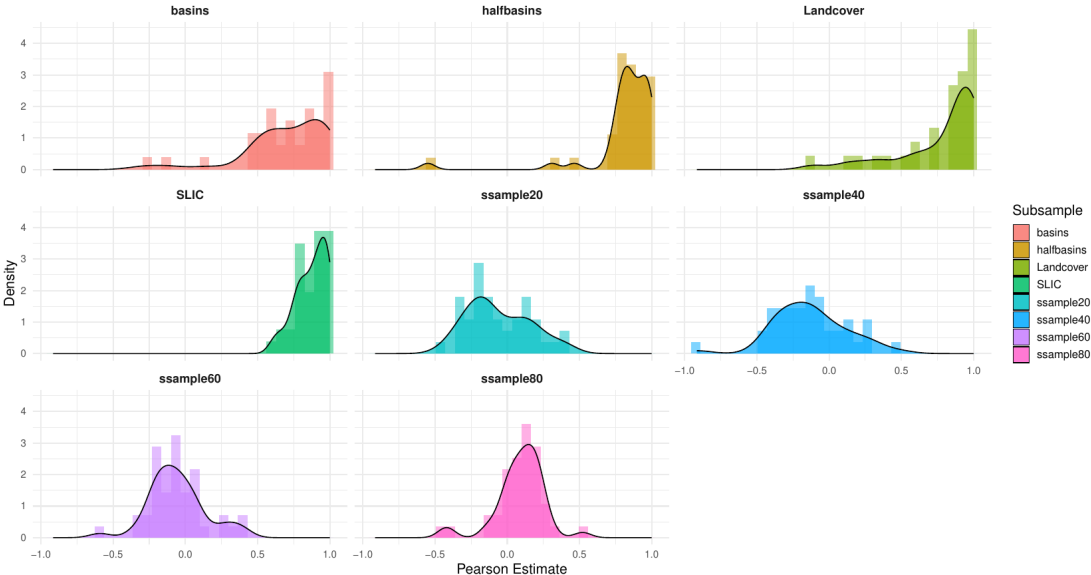



Figure S5-1 Pearson coefficient (Pcoeff2) value distributions representing sampling size effect for each subsample type, obtained with the median values of the original subsample performance.

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

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).

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

Evaluated metric	Raster's comparison Statistic	Subsample type	statistic	p
Subsample performance	MEDIAN	Landcover	0.9837	1.6579E-01
	MEDIAN	SLIC	0.8942	7.4565E-08
	MEDIAN	basins	0.9097	4.8741E-07
	MEDIAN	halfbasins	0.3123	1.1501E-21

240 Normality was assessed for each tessellation strategy using Pearson correlation values between interpolated surfaces and full

241 sampling grids. Results are based on median summary statistics of subsample performance. All strategies except Landcover

242 exhibited significant deviations from normality ($p < 0.05$), justifying the use of non-parametric analyses in subsequent tests.

243

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

245 performance)

Evaluated metric	Raster's comparison Statistic	n	statistic	df	p
Subsample performance	MEDIAN	489	194.444517	3	6.69E-42

246 A Kruskal-Wallis rank-sum test was performed to assess whether subsample performance differed significantly across

247 tessellation-based strategies (basins, halfbasins, SLIC, landcover). Median performance values were used. The results revealed

248 strong overall differences among groups ($\chi^2 = 194.44$, $p < 0.001$).

249
250
251
252 **Table S6-3** Results of Kruskal-Wallis Test on the efficiency of sampling strategy (Subsample

253 performance /Sample%)

Evaluated metric	Raster's comparison Statistic	n	statistic	df	p
Efficiency	MEDIAN	489	187.7701344	3	1.85E-40

254
255 This test evaluated differences in sampling efficiency—defined as the ratio of subsample performance to sampling proportion—

256 across the four tessellation strategies. Results based on median values indicated significant variation among strategies ($\chi^2 =$

257 187.77, $p < 0.001$), supporting subsequent pairwise post hoc analyses.

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

259

Evaluated metric	Raster's comparison Statistic	group1	group2	n1	n2	statistic	p	p.adj	p.adj. signif.
Subsample performance	MEDIAN	basins	halfbasins	123	125	3.1677	1.5363E-03	1.8435E-03	**
		basins	Landcover	123	118	-10.0837	6.5194E-24	1.9558E-23	****
		basins	SLIC	123	123	-0.5753	5.6510E-01	5.6510E-01	ns
		halfbasins	Landcover	125	118	-13.2579	4.0600E-40	2.4360E-39	****
		halfbasins	SLIC	125	123	-3.7453	1.8015E-04	2.7023E-04	***
		Landcover	SLIC	118	123	9.5145	1.8266E-21	3.6532E-21	****

260 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

261 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.

262

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

Evaluated metric	Raster's comparison Statistic	group1	group2	n1	n2	statistic	p	p.adj	p.adj. signif.
Efficiency (S. performance/Sample%)	MEDIAN	basins	halfbasins	123	125	-4.0125	6.0068E-05	7.2081E-05	****
		basins	Landcover	123	118	7.7825	7.1133E-15	1.4227E-14	****
		basins	SLIC	123	123	-4.2995	1.7118E-05	2.5678E-05	****
		halfbasins	Landcover	125	118	11.7835	4.7488E-32	1.4246E-31	****
		halfbasins	SLIC	125	123	-0.3043	7.6093E-01	7.6093E-01	ns
		Landcover	SLIC	118	123	-12.0371	2.2674E-33	1.3605E-32	****

264 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

265 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

266 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 pergformance), 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.

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

Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Variable
0.851861375	0.145319259	-0.149755129	-0.041427741	-0.015176126	Sample_percent
0.756625756	0.027853896	0.309265451	0.076822303	0.036161759	Sample performance

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

297

298 **Table S7-2** Coordinates of categorical variable categories in FAMD space.

Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Categoria
-0.406935276	-0.020188946	-0.279573264	0.789325407	0.2791168	AndeanRural
1.207530075	0.248893375	-0.396175078	-0.24790374	-0.17667598	AndeanUrban
-0.843706785	-0.240412178	0.708193997	-0.563839692	-0.105861135	Plains
-0.279759693	1.505889124	0.003419648	0.008974548	0.005827382	R0-2
0.235334756	-0.797792058	0.093675238	-0.487617078	1.089335401	R0-24
0.043448051	-0.807204002	-0.213912179	-0.119079686	-0.189419871	R12-24
0.170093991	-0.802450062	0.116554794	0.583323799	-0.888812578	R2-12
0.124887751	0.006917641	0.193329576	0.321322925	1.01940836	Dawn
-0.085304918	0.009631239	-0.377862385	-0.963833567	-0.452462316	Evening
-0.040598084	-0.016693736	0.184478507	0.644493282	-0.576738355	Full
0.276358092	-0.369306234	0.837968804	0.263920195	-0.215130152	ACIft
0.122899431	-0.387697235	0.420751064	-0.426283214	0.093408064	ACItf
-0.480916804	2.612844944	-0.021339035	-0.095240223	-0.038027496	BI
-0.43240671	2.615175135	0.101744418	0.076084087	0.075043022	NDSI
-0.175660923	-0.605101019	-1.349215271	0.179172311	0.11743916	NP

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

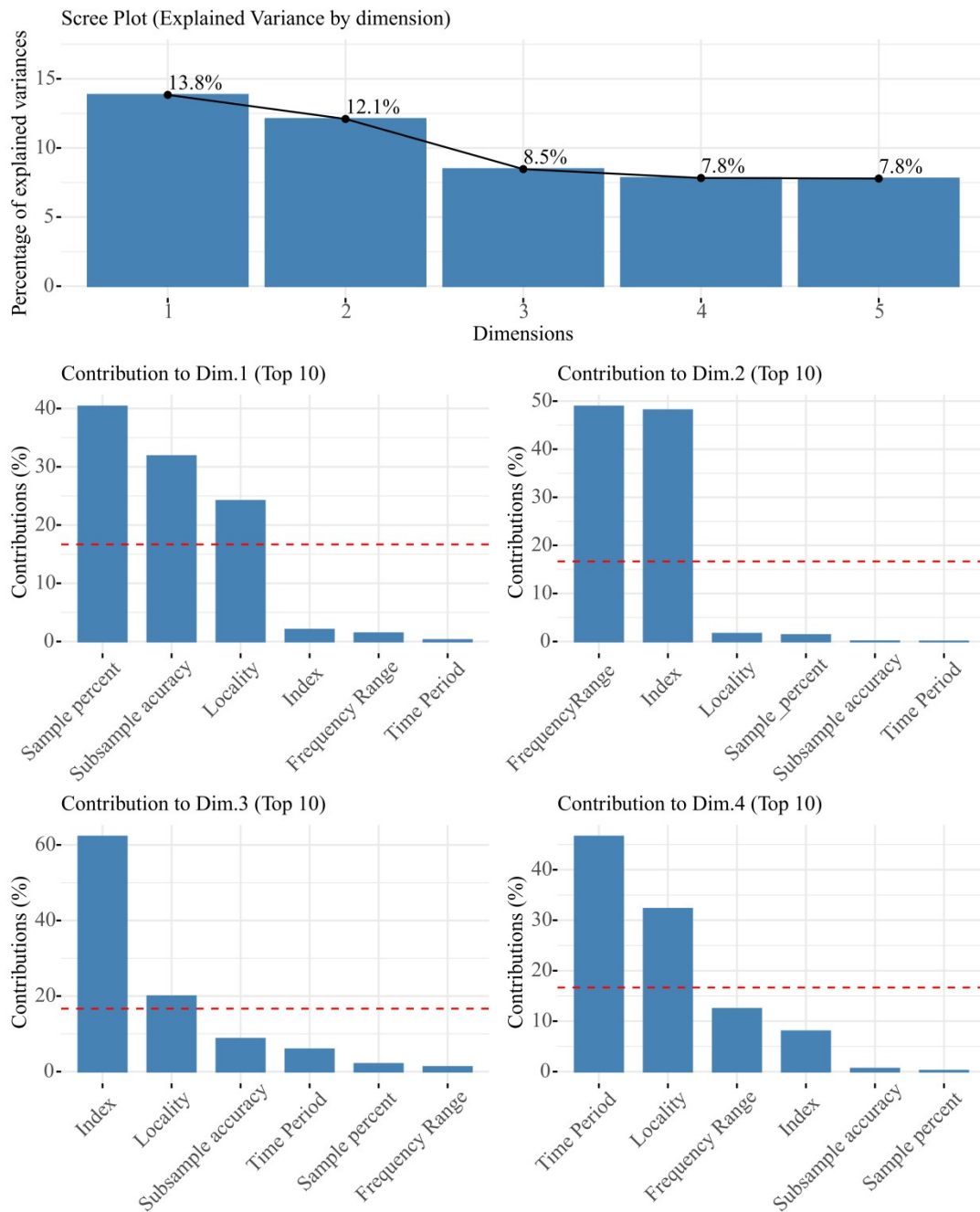
300

301 **Table S7-3** Explained variance per FAMD dimension

Dimension	Dimension	Variance (%)	Cumm. variance (%)
Dim.1	1.798800207	13.83692467	13.83692467

Dim.2	1.572358413	12.09506472	25.93198939
Dim.3	1.100659781	8.466613701	34.39860309
Dim.4	1.017269671	7.825151318	42.22375441
Dim.5	1.012287109	7.786823916	50.01057832

302   Proportion of variance (%) captured by each FAMD dimension and the cumulative variance explained across dimensions 1 to 5.
303
304

305 **Figure S7-1** FAMD SUMMARY PLOT

306
307 Upper panel: Scree plot of variance explained by each FAMD dimension (Dim.1-Dim.5).Middle & inferior pannels: Variable

308 contributions to dimensions 1 to 4, showing the top contributing continuous and categorical variables. Dimension 1 is

309 predominantly associated with performance-related variables (e.g., Sample percent, Subsample performance), while Dimensions

310 2-4 are shaped by landscape, acoustic, and temporal factors. The dashed red line indicates the expected average contribution

311 (threshold for relevance).

312
313
314
315

316 **Table S7-4** Statistical comparison of stratified sub sampling strategies: Dunn's Post-Hoc Test Across FAMD Dimensions

Comparissons								
.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

317   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

318   test statistics, adjusted p-values, and significance levels. Significant differences (p.adj < 0.05) indicate structural dissimilarities between strategies in the multivariate space.

319

320

321 **Table S7-6** Significance Groupings of sub sampling Strategies Across FAMD Dimensions

DIMENSION	Subsample	Significance_Letter
Dim.1	basins	a
Dim.1	halfbasins	b
Dim.1	Landcover	c
Dim.1	SLIC	d
Dim.2	basins	a
Dim.2	halfbasins	b
Dim.2	Landcover	c
Dim.2	SLIC	ab
Dim.3	basins	a
Dim.3	halfbasins	b
Dim.3	Landcover	b
Dim.3	SLIC	b
Dim.4	basins	a
Dim.4	halfbasins	b
Dim.4	Landcover	c
Dim.4	SLIC	c

322 Letter-based groupings resulting from post-hoc Dunn tests (adjusted with Benjamini-Hochberg correction) for each FAMD

323 dimension. sub sampling strategies sharing the same letter are not significantly different at $p < 0.05$. Distinct letters denote

324 statistically significant structural differences in multivariate space.

325

326 **Table S7-7** Kruskal-Wallis Test Results Across FAMD Dimensions

.y.	n	statistic	df	p	method
Dim.1	489	292.5488352	3	4.08E-63	Kruskal-Wallis
Dim.2	489	31.47879747	3	0.000000674	Kruskal-Wallis
Dim.3	489	58.06207889	3	1.52E-12	Kruskal-Wallis
Dim.4	489	127.2502701	3	2.12E-27	Kruskal-Wallis

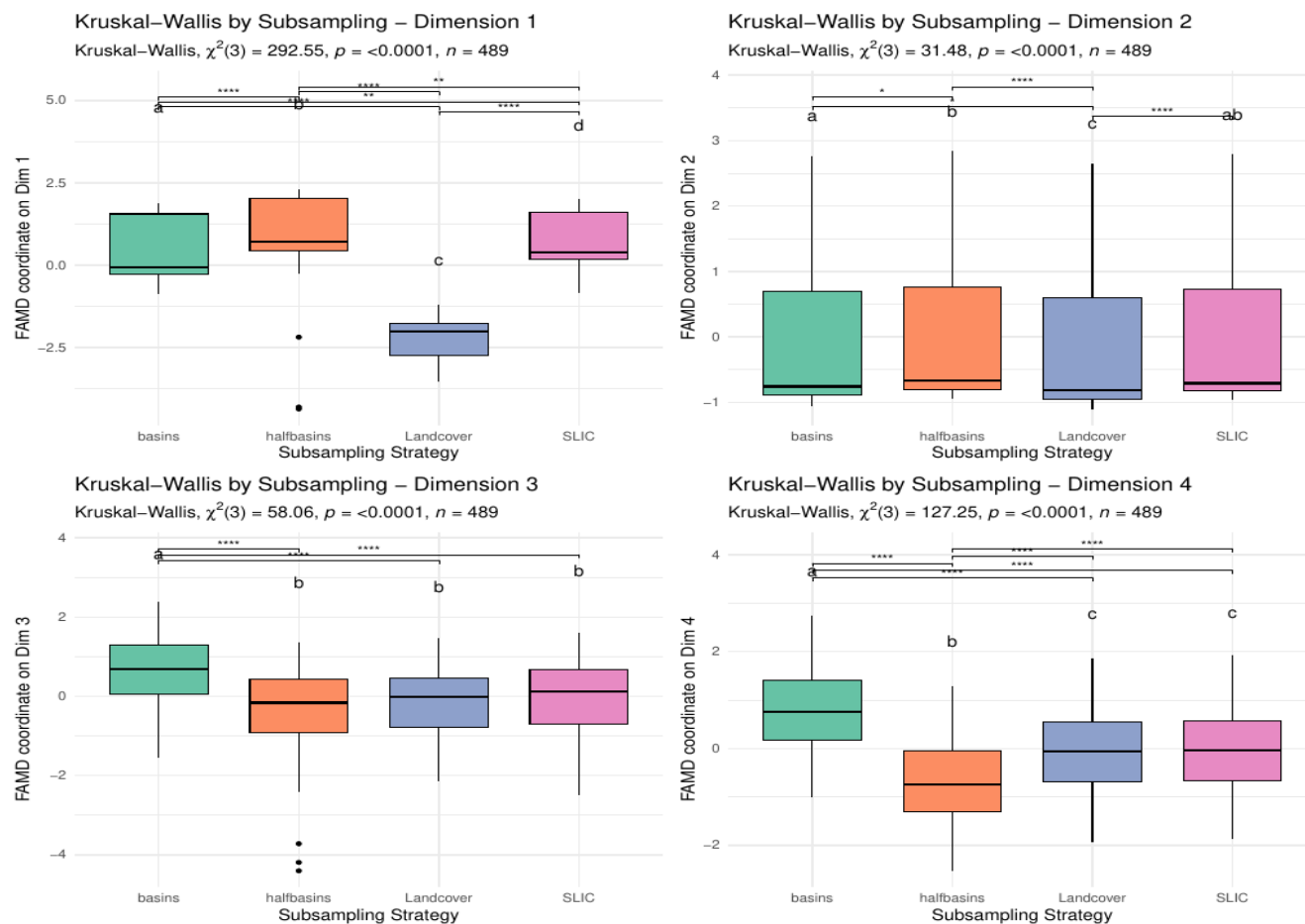
327 Summary statistics for Kruskal-Wallis tests evaluating differences among sub sampling strategies across the first four FAMD

328 dimensions. All tests were significant ($p < 0.001$), indicating structural differences in multivariate space depending on the

329 sampling strategy used.

330 **Figure S7-8** Kruskal-Wallis + Dunn Tests of each FAMD dimension graphic showing differences in FAMD dimensions across stratified sub sampling strategies.

331

332

333 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.

334 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

335   Benjamini-Hochberg correction. Dimension 1 primarily reflects performance variables, while Dimensions 2-4 capture structural variation linked to acoustic index, locality,

336   and temporal stratification.

SUPPLEMENTARY INFORMATION S8: Distributional comparison of landscape heterogeneity across stratified sub sampling strategies.

Kernel density estimates (Density plot B) and quantile-quantile plots (Q-Q plot A) illustrate how the mean landscape

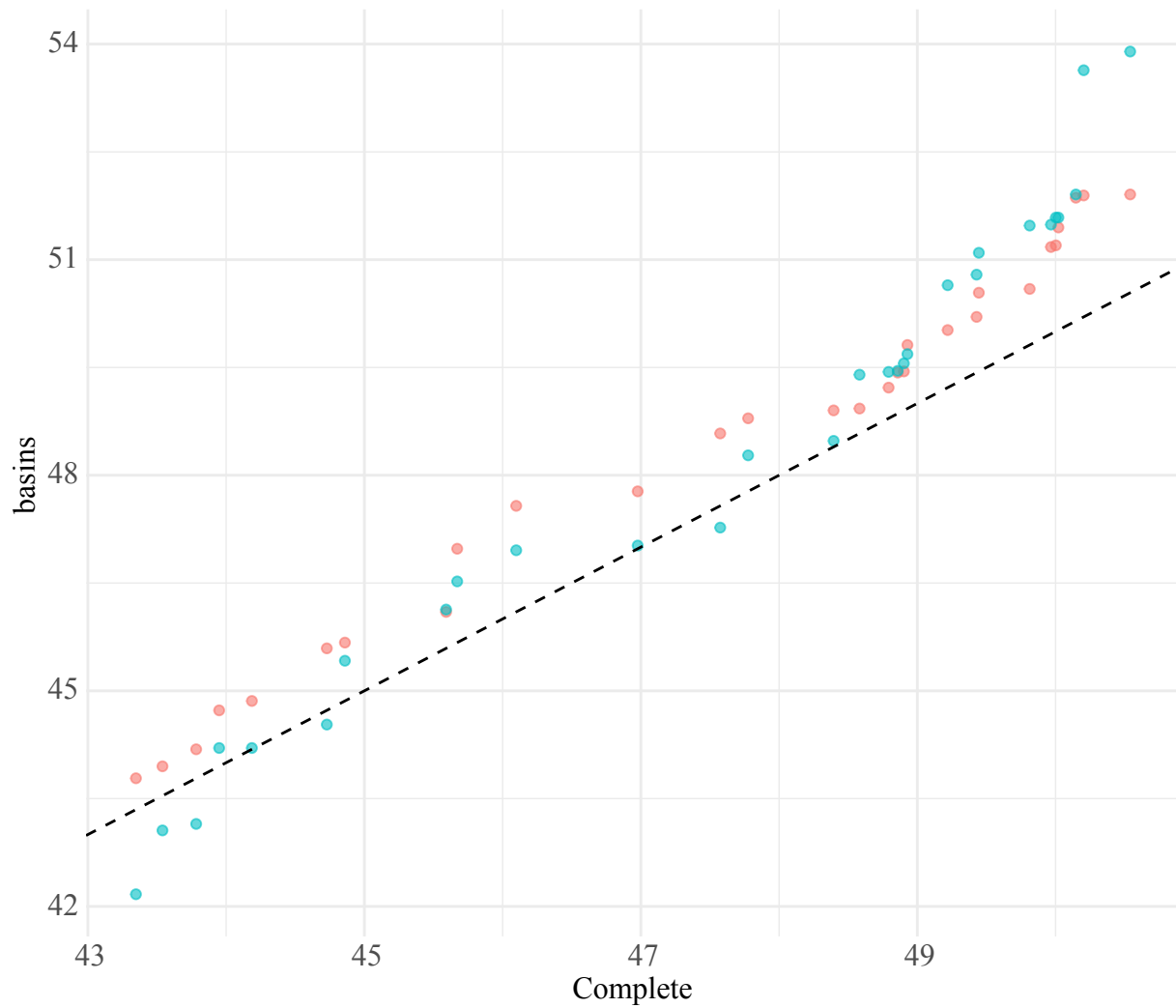
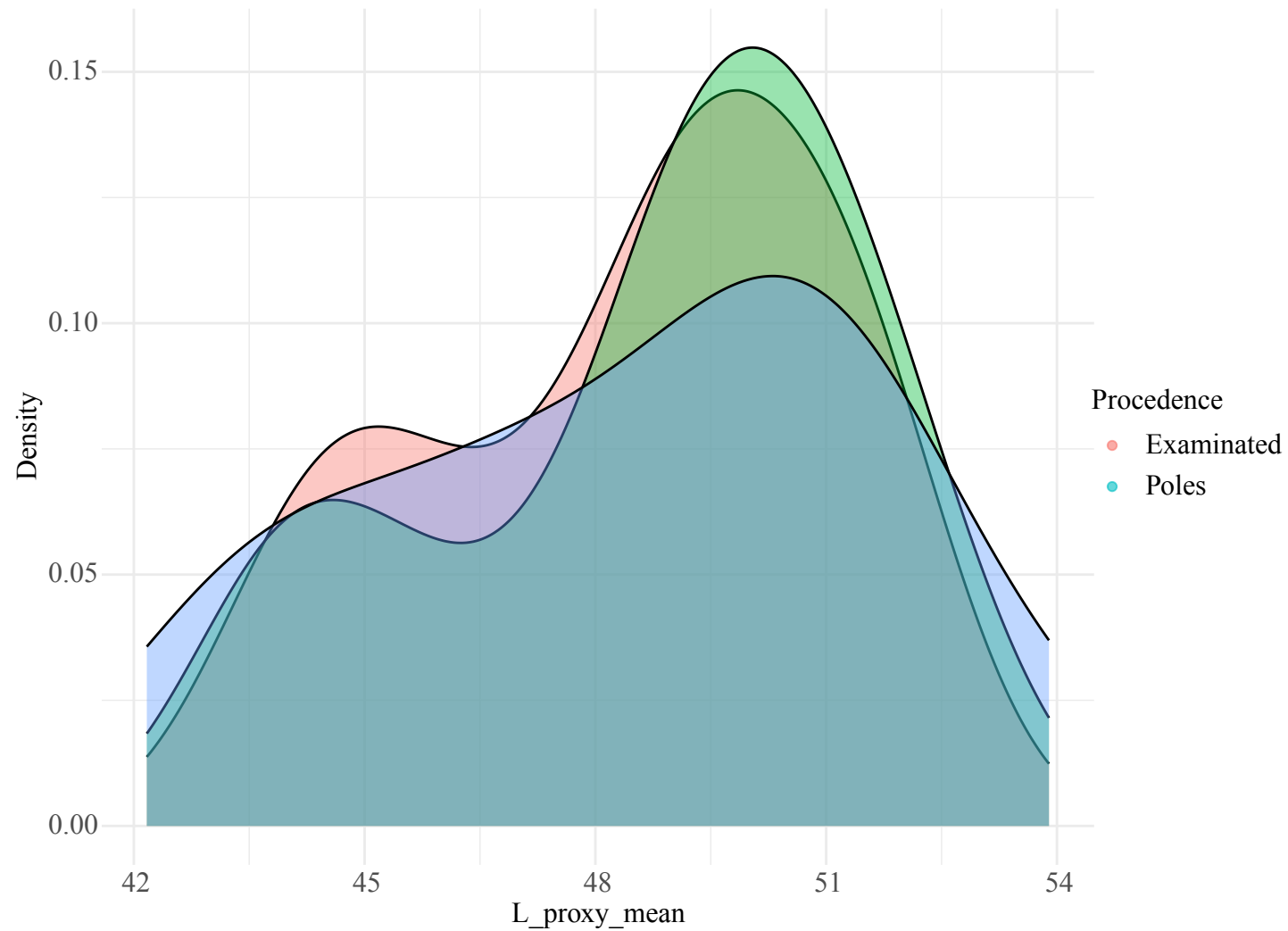
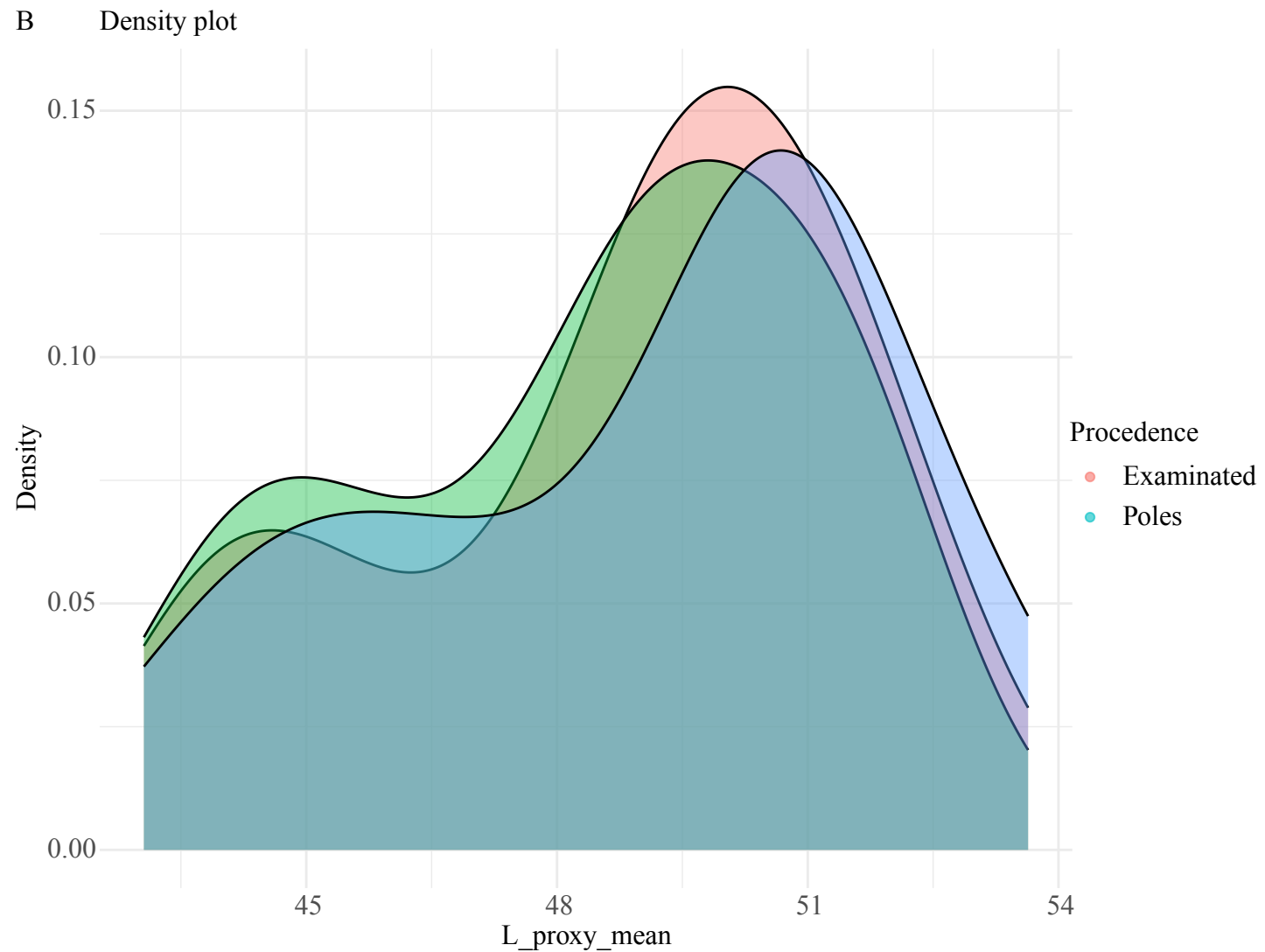
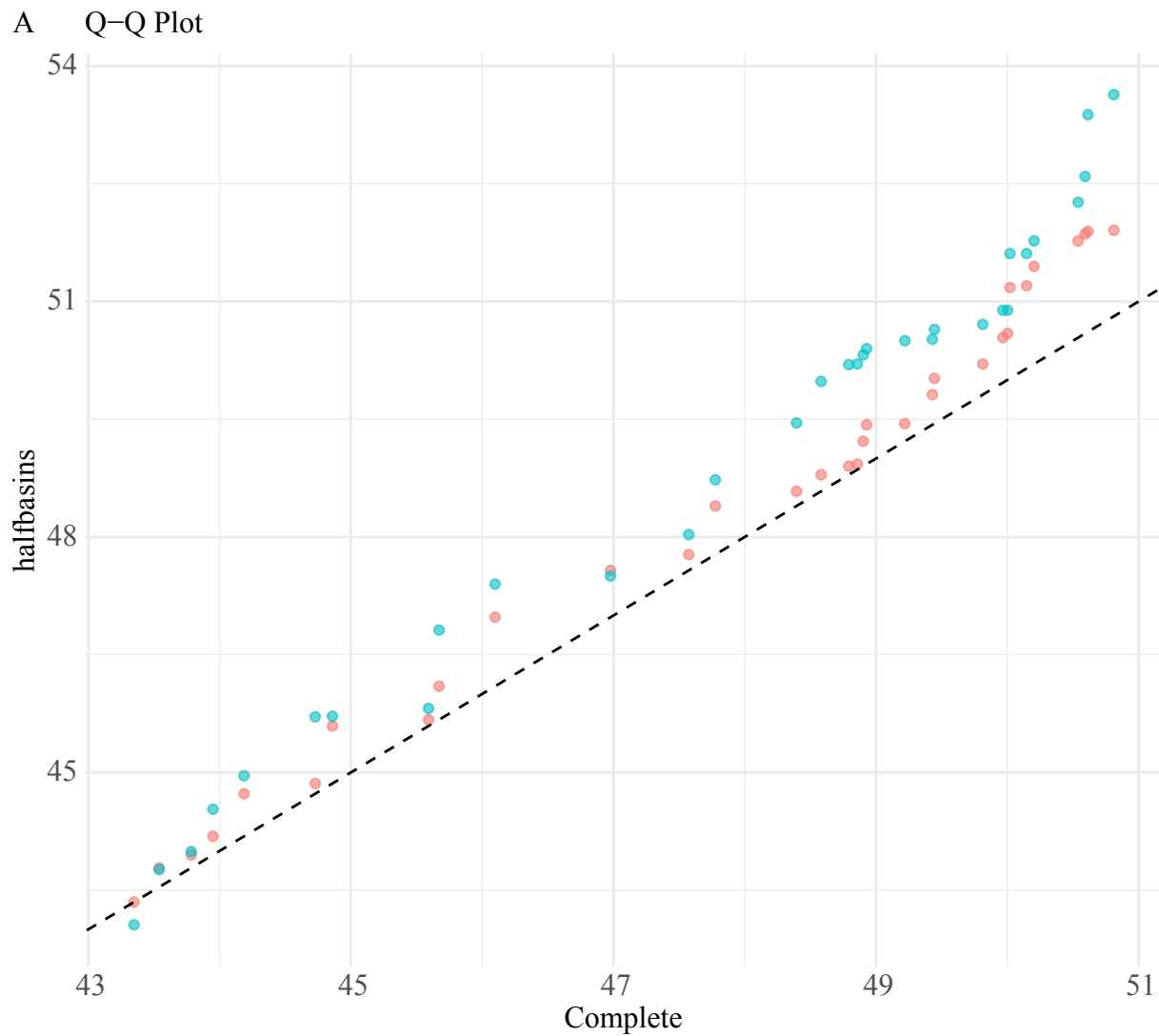
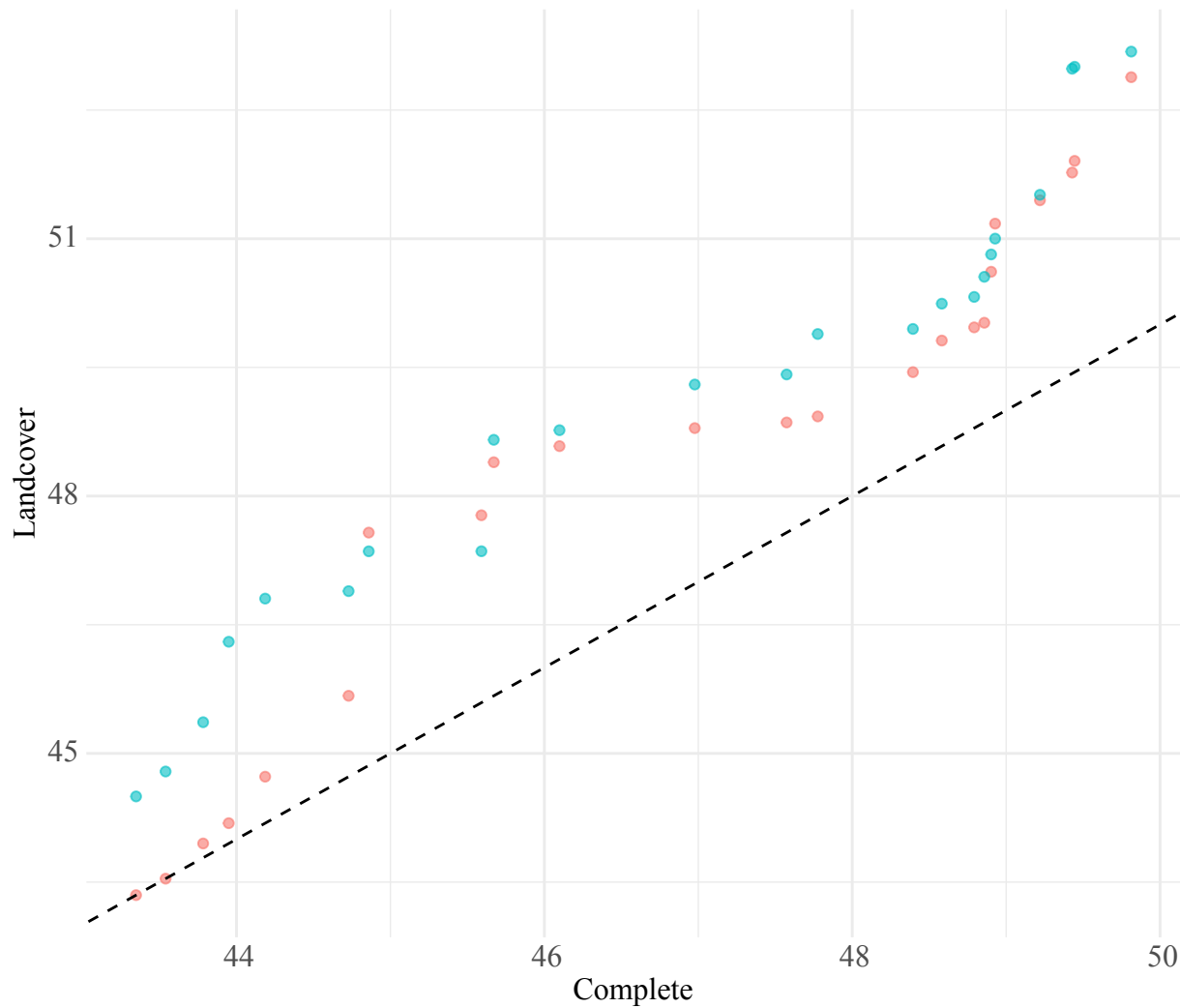
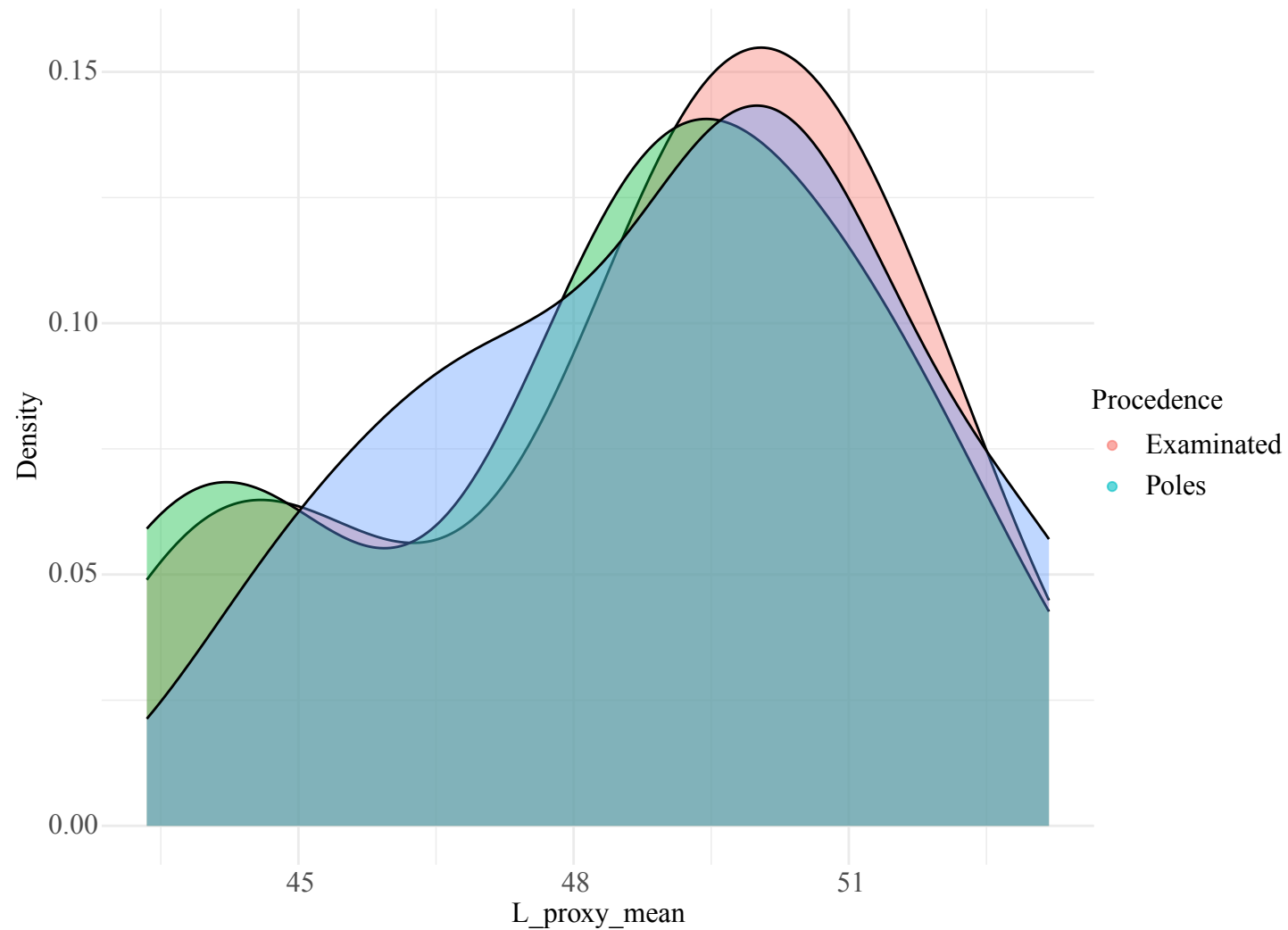
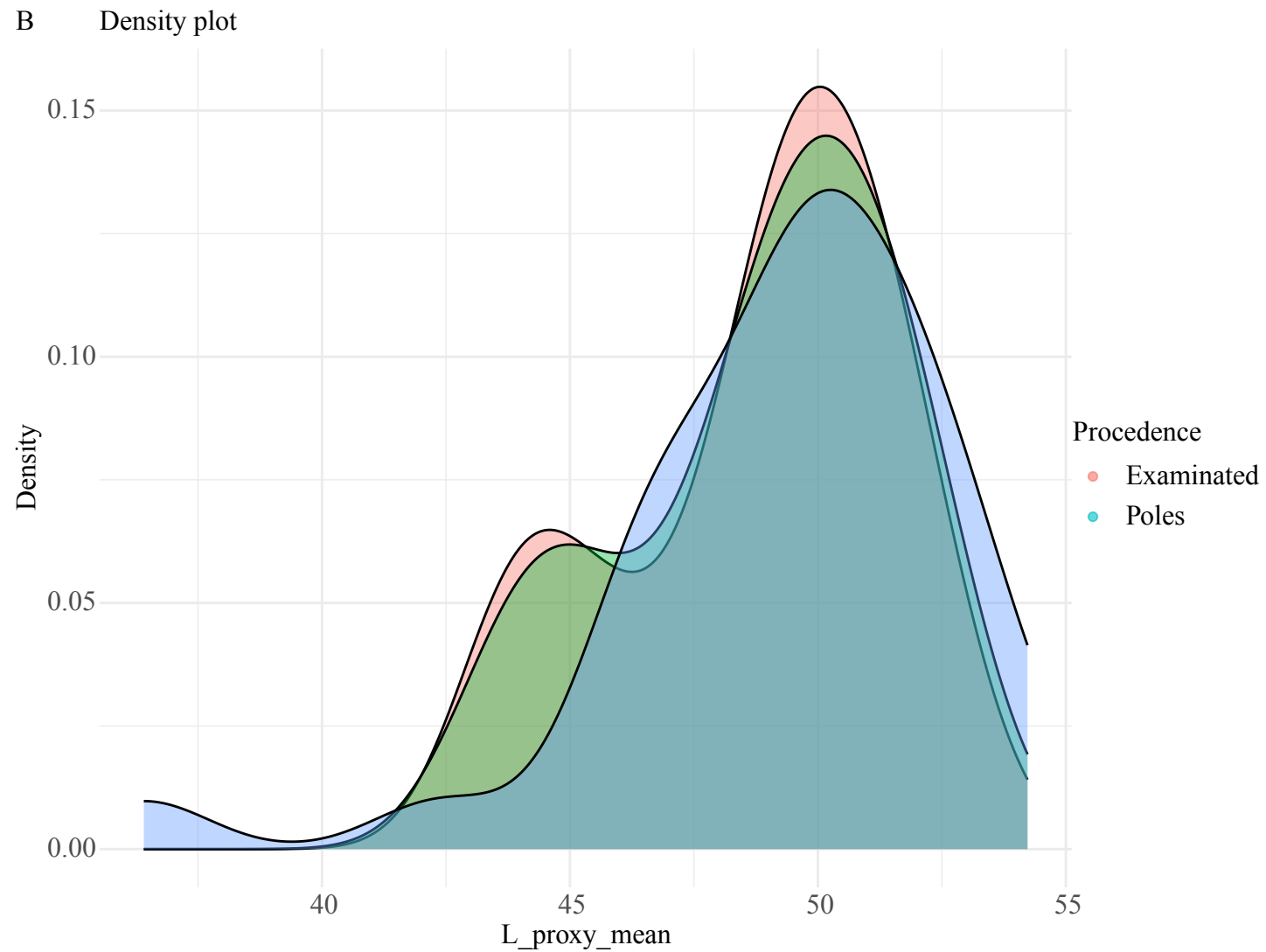
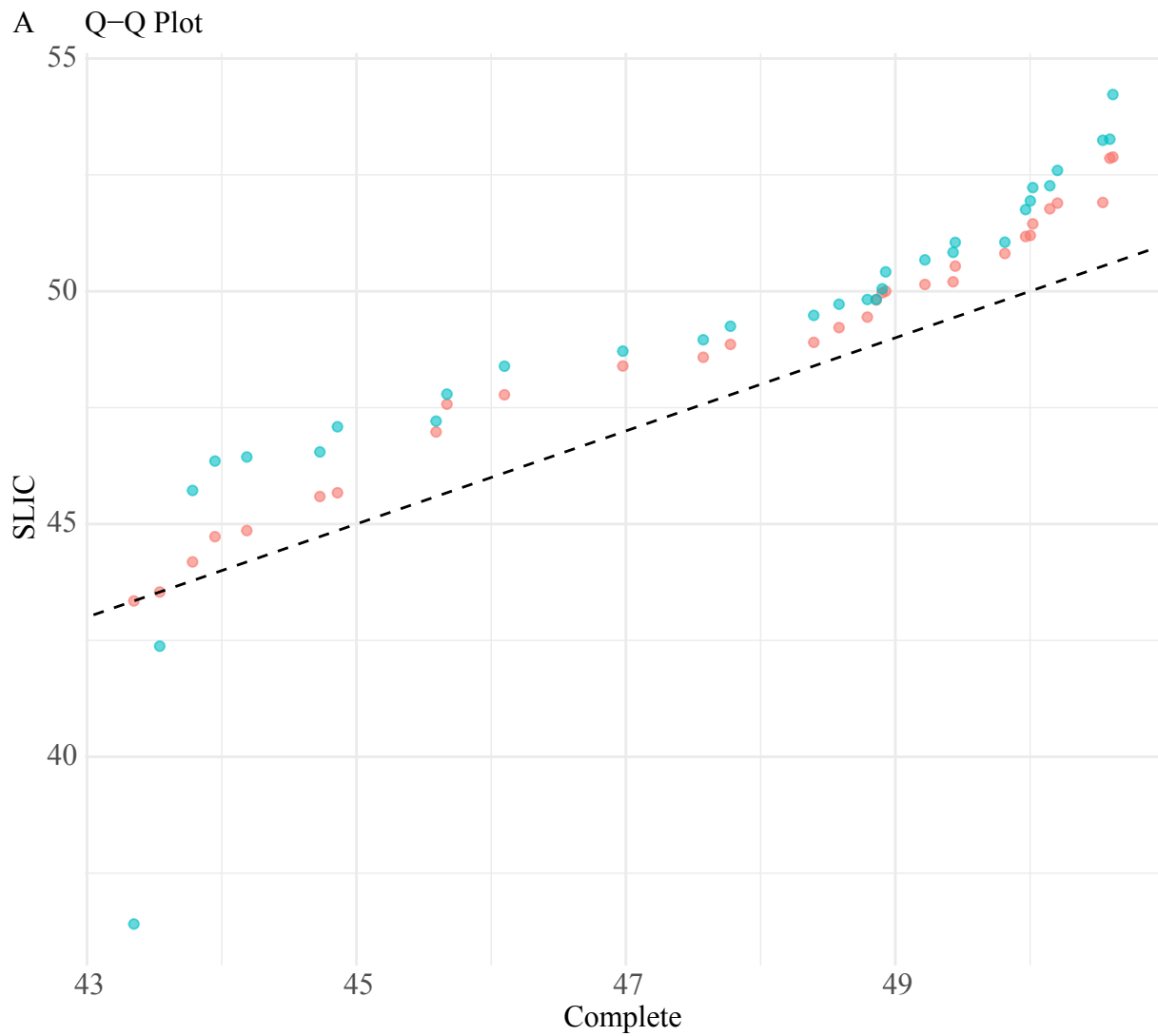
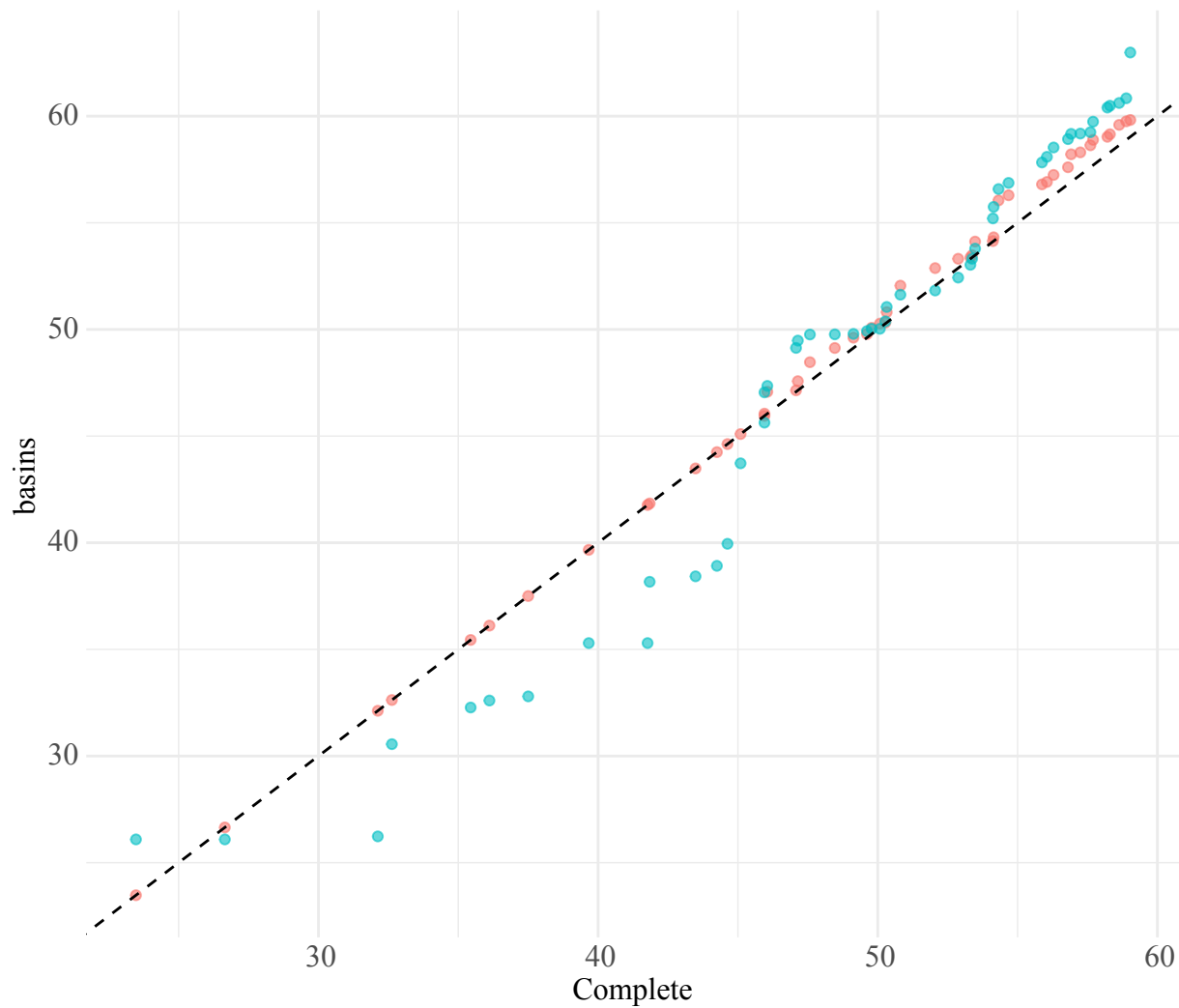
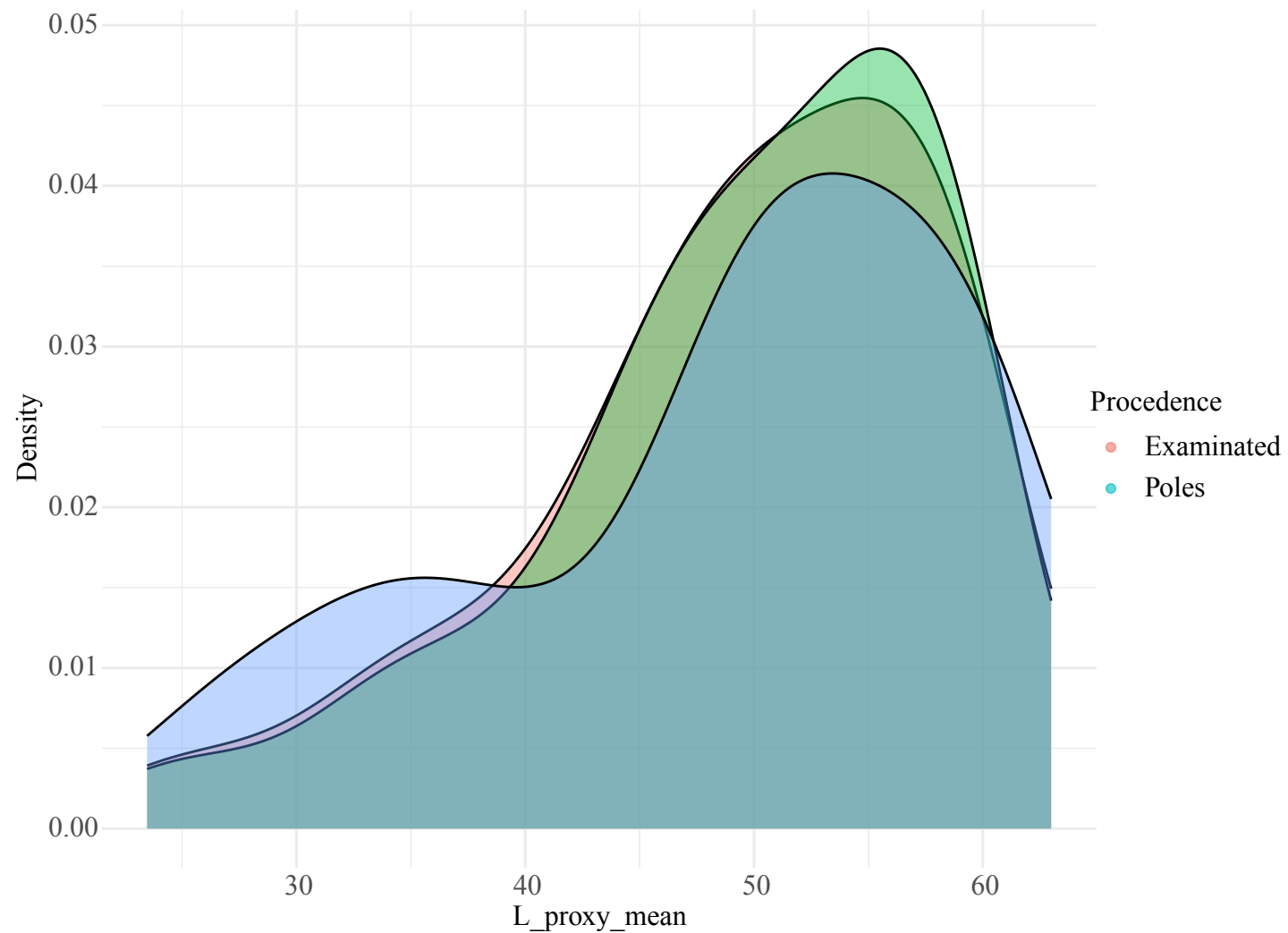
proxy values (L_proxy_mean) from each subsample compare with the full sampling grid across different strategies

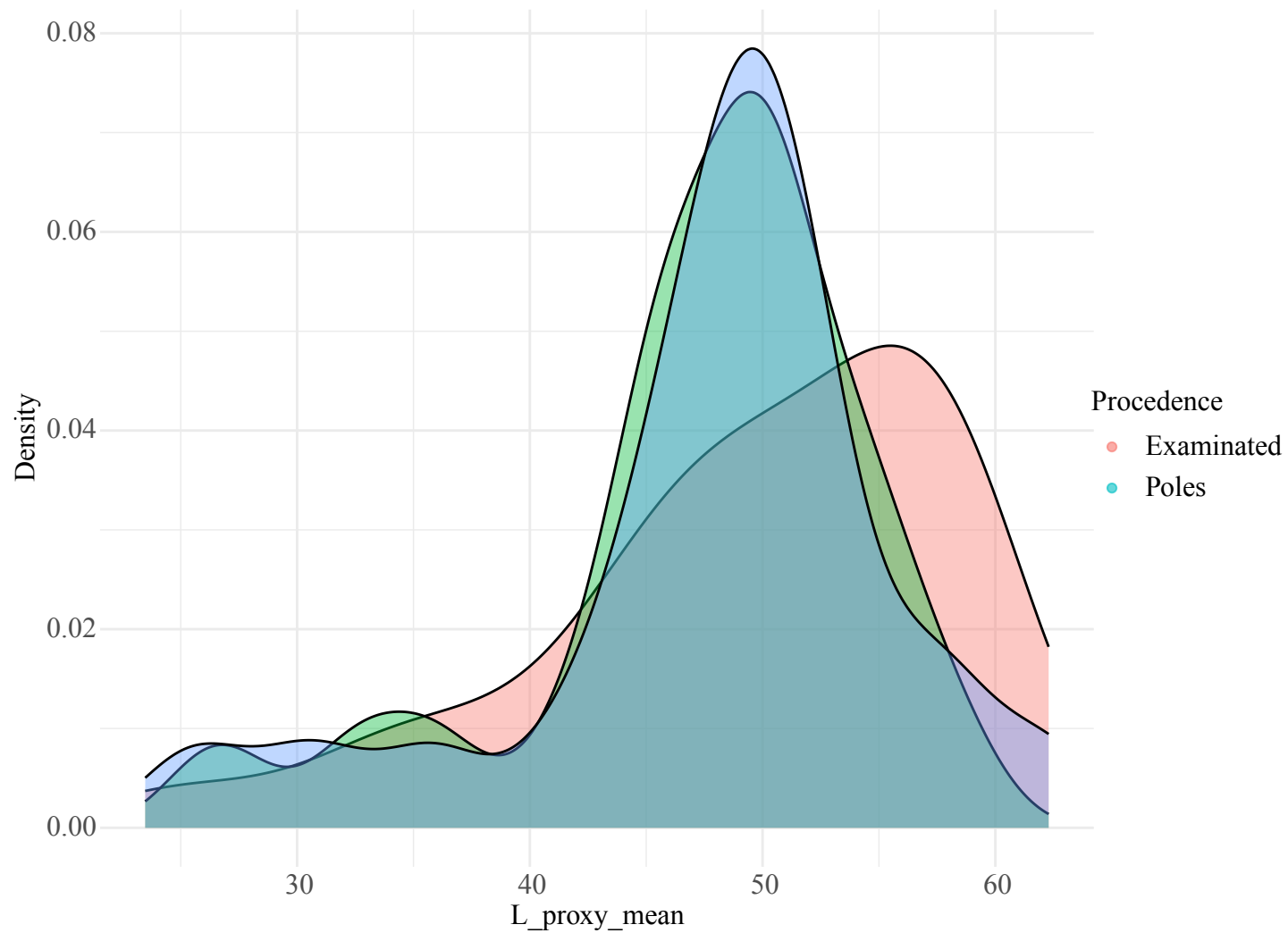
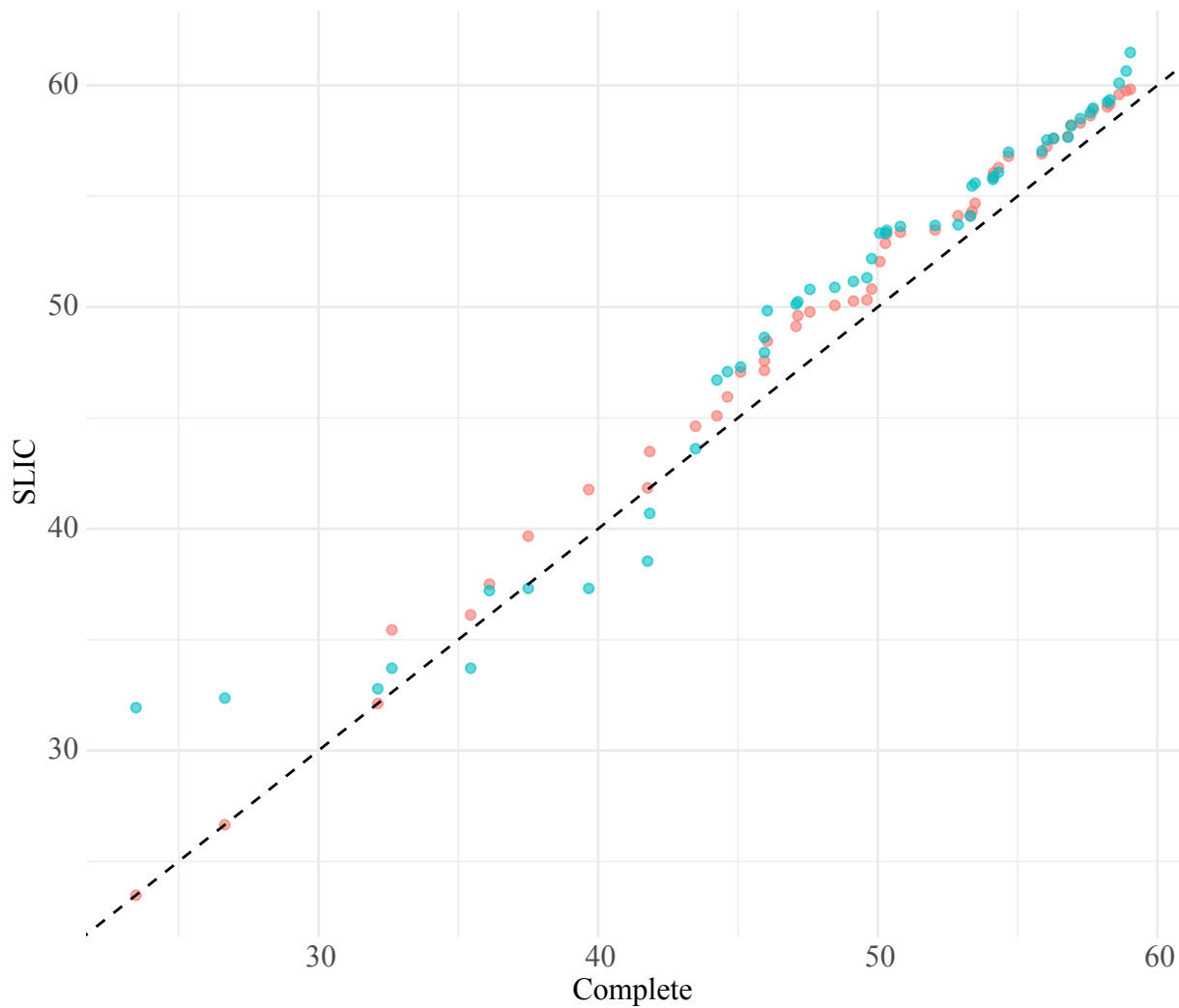
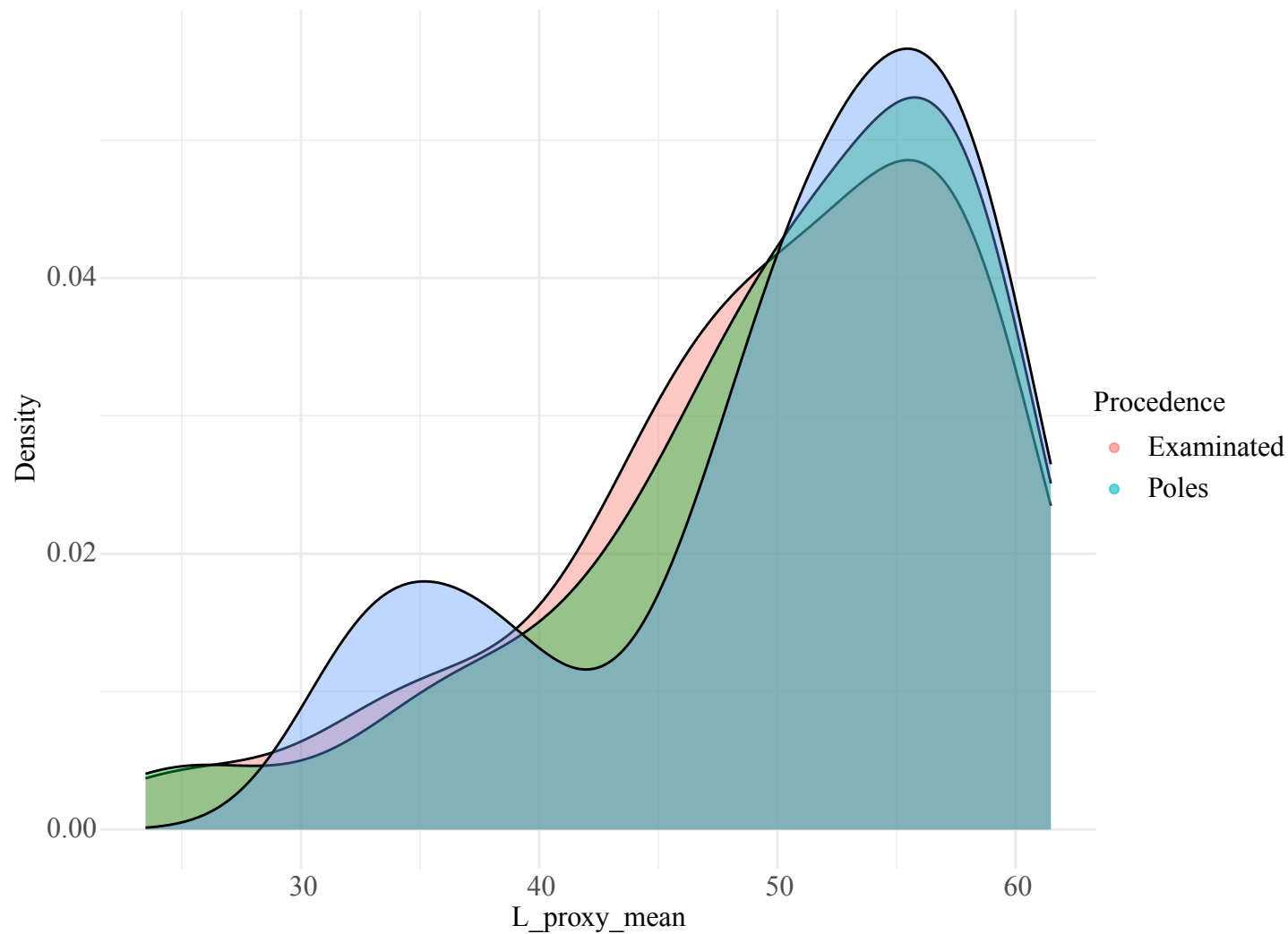
(Basins, Halfbasins, Landcover, SLIC) and localities (Andean Rural, Andean Urban, Plains). Sub samples were

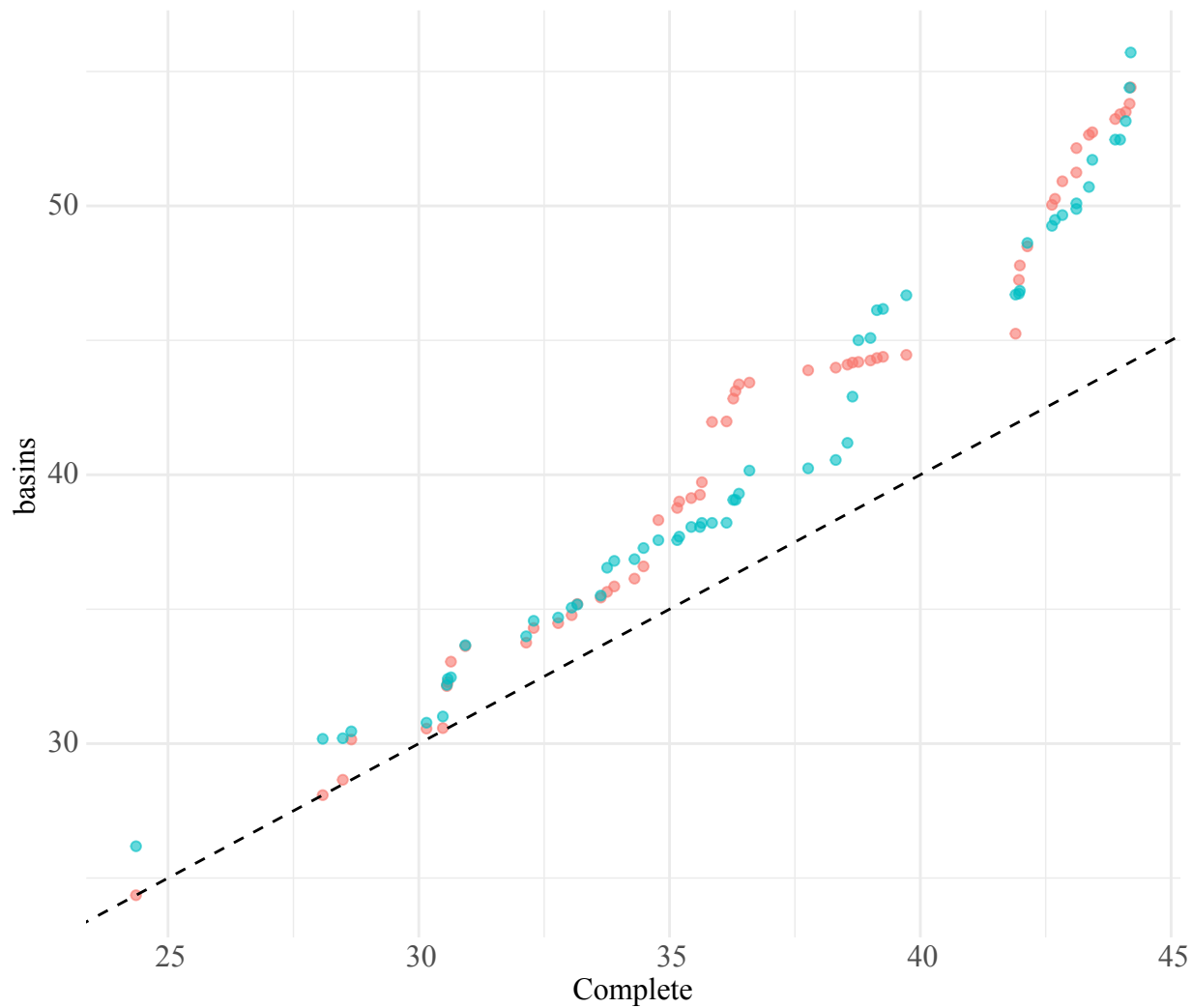
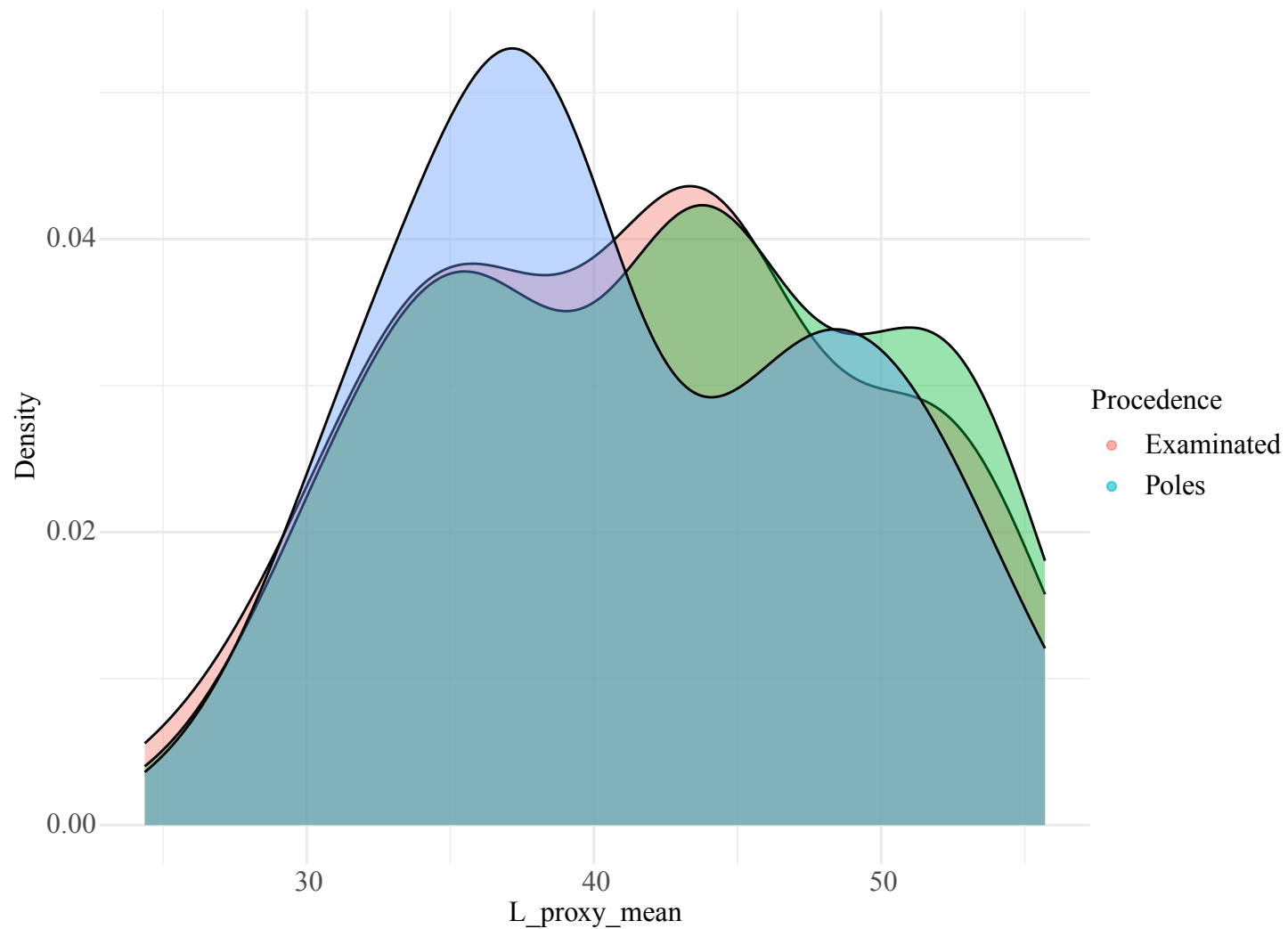
grouped by point selection origin: poles of inaccessibility or directly examined locations. Statistical differences in

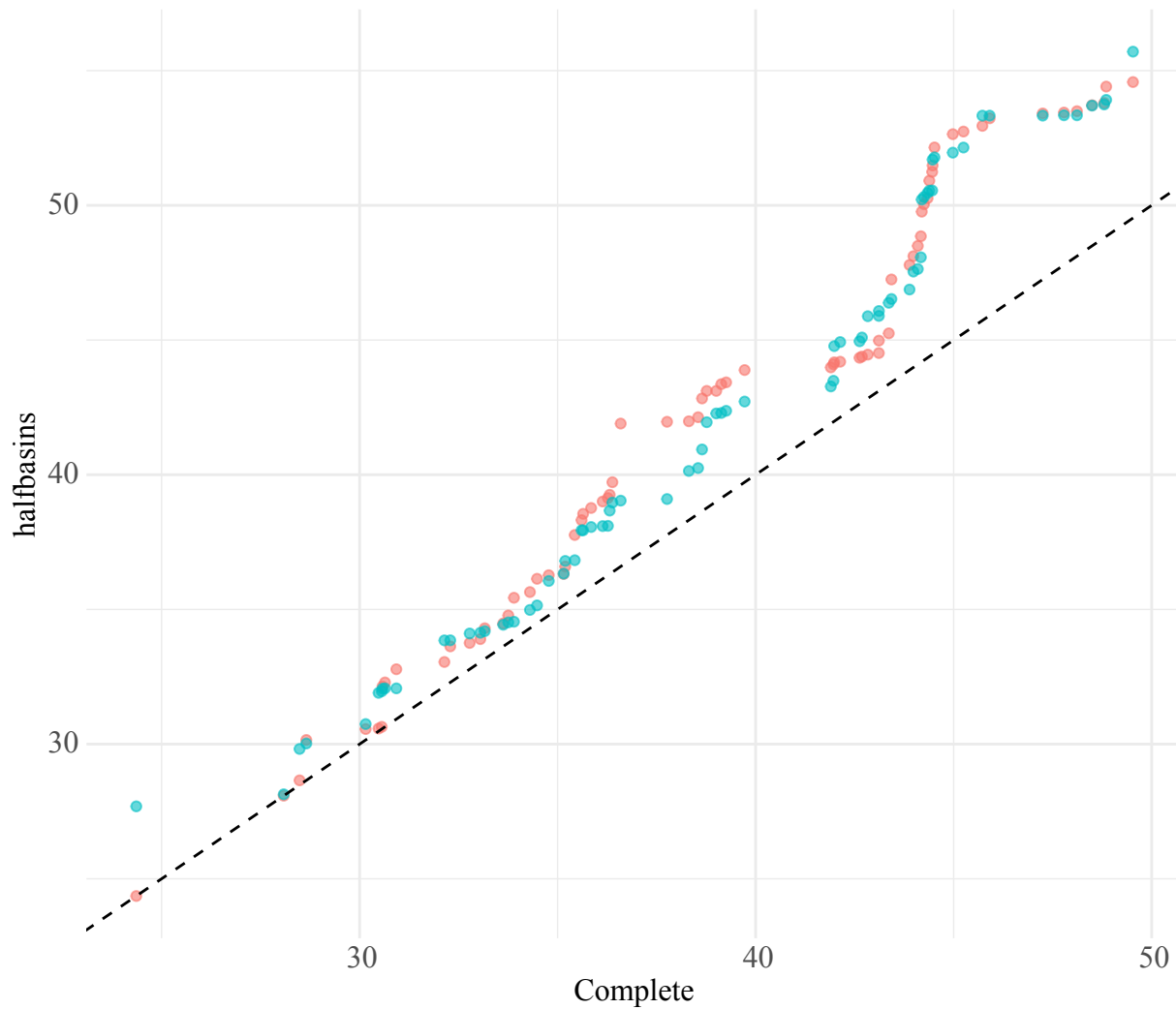
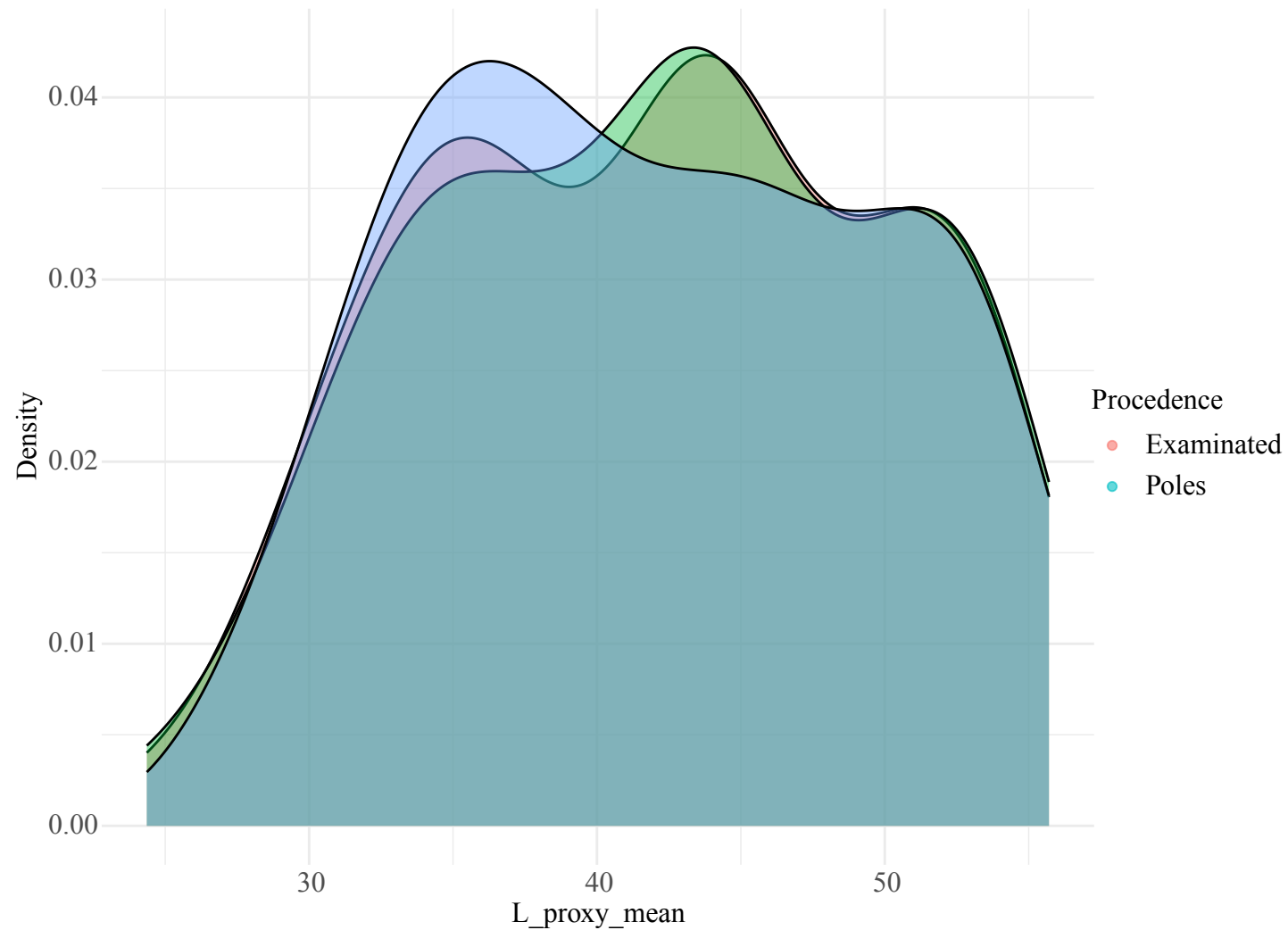
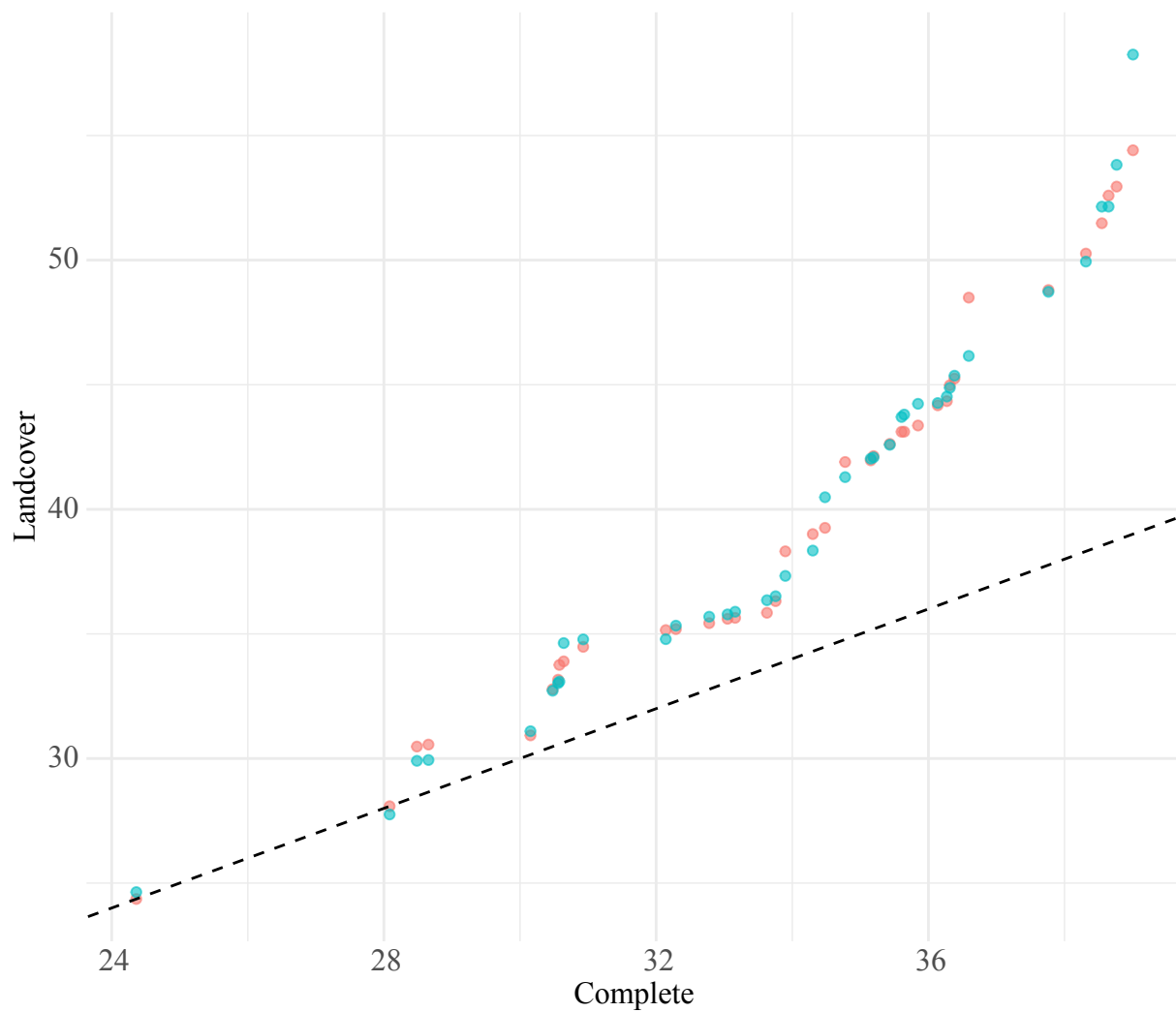
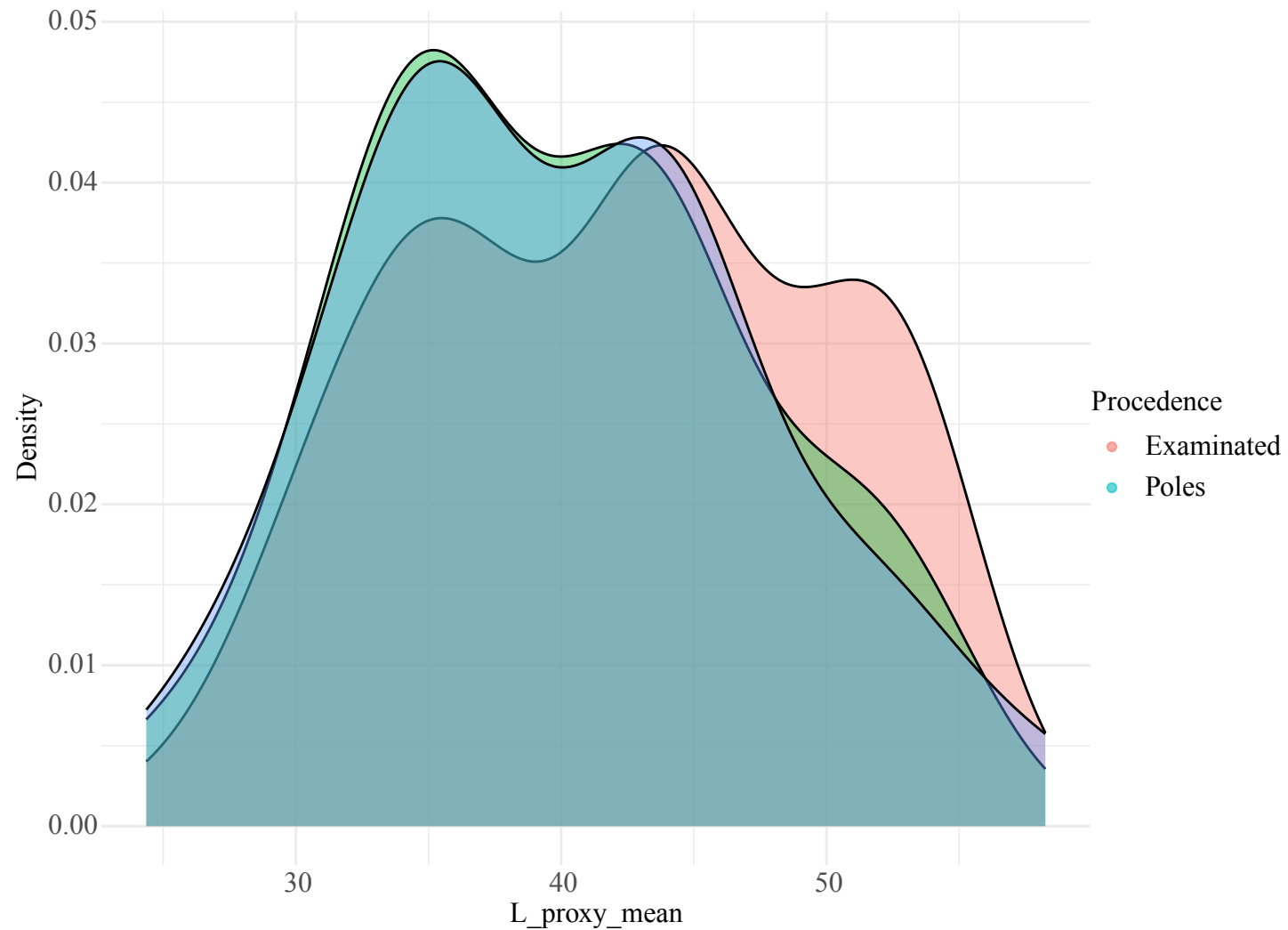
distributions were assessed using Kolmogorov-Smirnov tests and Wasserstein distances.

AndeanRural / halfbasins / Comparison with Examinated & Poles

AndeanRural / Landcover / Comparison with Examinated & Poles

A Q−Q Plot

B Density plot

AndeanUrban / halfbasins / Comparison with Examinated & Poles

AndeanUrban / SLIC / Comparison with Examinated & Poles

A Q−Q Plot

B Density plot

Plains / basins / Comparison with Examinated & Poles

A   Q−Q Plot

B   Density plot

Plains / halfbasins / Comparison with Examinated & Poles

Plains / Landcover / Comparison with Examinated & Poles

Plains / SLIC / Comparison with Examinated & Poles