Supplementary Figures and Tables

Danet, A, Giam, X, Olden, J, Comte, L

06 July, 2022

Contents

- Add plot species richness / chao richness

1 Data selection

```
#> function (string, locale = "en")
#> {
#>     stri_trans_totitle(string, opts_brkiter = stri_opts_brkiter(locale = locale))
#> }
#> <bytecode: 0x0000000024814d50>
#> <environment: namespace:stringr>
```

```
#> `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Table 1: Protocol occurrence across sites

Protocol	N	Percent
Electrofishing	1896	42.4%
Electrofishing backpack	1359	30.4%
Electrofishing partialsection	608	13.6%
Electrofishing wholesection	331	7.4%
Electrofishing boat	86	1.9%
Electrofishing and netting	67	1.5%
Seining	56	1.3%
Electrobackpack and netting	28	0.6%
Rotenone Lockchamber	20	0.4%
Trapnetting	9	0.2%
Gillnetting	6	0.1%
Seining and gillnetting	5	0.1%
Trapping	2	0.0%
Trawling	2	0.0%
Electrofishing shorebased	1	0.0%
Total	4476	100.0%

Table 2: Abundance unit across sites

Abundance unit	N	Percent
Individual number per 100m2	2349	52.5%
Count	2083	46.5%
Catch Per Unit Effort (CPUE)	40	0.9%
Leslie index	4	0.1%
Total	4476	100.0%

Table 3: Realm and country occurrence across sites

Realm	Country	N	Percent
Palearctic	GBR	1282	28.64%
Palearctic	FRA	935	20.89%
Palearctic	SWE	819	18.30%
Palearctic	FIN	126	2.82%
Palearctic	ESP	111	2.48%
Palearctic	JPN	33	0.74%
Palearctic	HUN	32	0.71%
Palearctic	BEL	18	0.40%
Palearctic	NOR	4	0.09%
Neotropics	BRA	3	0.07%
Nearctic	USA	784	17.52%
Nearctic	CAN	113	2.52%
Australasia	AUS	213	4.76%
Afrotropics	CIV	2	0.04%
Afrotropics	BWA	1	0.02%

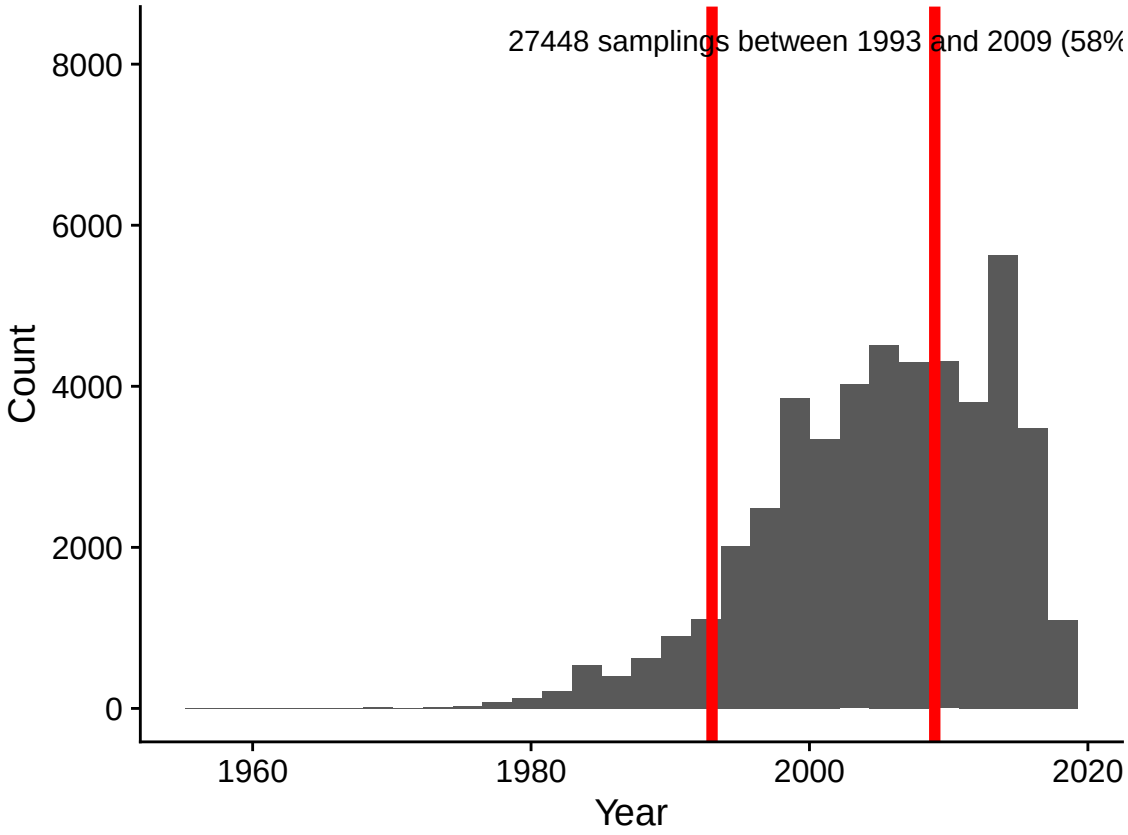

Figure 1: Distribution of year of samplings. Years 1993 and 2009 are highlighted because they correspond to the two human footprint measurements.

```
#> `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
#> Warning in as_grob.default(plot): Cannot convert object of class character into
#> a grob.
```

Table 4: Summary descriptors of response variables distribution

Community metric	n	Median (Q1, Q3)	(Min, Max)
Proportion non-native abundance	46932	0 (0,0.01)	(0,1)
Proportion non-native species	46932	0 (0,0.11)	(0,1)
Turnover (jaccard)	46932	0 (0,0.33)	(0,1)
Disappearance	46932	0.07 (0,0.17)	(0,0.91)
Nestedness (jaccard)	46932	0.1 (0,0.29)	(0,0.96)
Appearance	46932	0.11 (0,0.22)	(0,0.92)
Dissimilarity (Simpson index)	46932	0.26 (0.03,0.6)	(0,1)
Jaccard (binary, dissimilarity)	46932	0.35 (0.09,0.5)	(0,1)
Log Chao species richness	46932	1.49 (0.69,2.14)	(0,4.45)
Log total abundance	46932	4.4 (3.12,5.68)	(-4.61,11.6)
Chao species richness	46932	4.43 (2,8.5)	(1,85.71)
Species richness	46932	5 (2,9)	(1,72)

Table 5: Data source for native species status in term of species occurrence

Source for native species status	N	Percent
Tedesco database (basin scale)	326032	94.35%
Fishbase (country scale)	19057	5.52%
Atlas, fishbase (country scale)	298	0.09%
NAS database (state scale, USA)	157	0.05%

Table 6: Summary distribution of native and non-native species metrics

Community metric	N	Median (Q1, Q3)	(Min, Max)
Proportion native abundance	46932	1 (0.99,1)	(0,1)
Proportion native species	46932	1 (0.89,1)	(0,1)
Proportion non-native abundance	46932	0 (0,0.01)	(0,1)
Proportion non-native species	46932	0 (0,0.11)	(0,1)
Species richness (native)	46932	4 (2,8)	(0,70)
Species richness (non-native)	46932	0 (0,1)	(0,17)

Table 7: Descriptive statistics for the site environmental descriptors.

Environmental variables	N	Median (Q1, Q3)	(Min, Max)
Annual average of discharge (m3/s)	4476	1.72 (0.61,6.79)	(0.01,7830.69)
Average elevation (m)	4476	174 (82,307)	(-1,2531)
Average slope (degree)	4476	30 (16,51)	(0,362)
Distance from source	4476	26.6 (13.8,57.5)	(2.7,2798.3)
Human footprint 1993 (index)	4476	18.1 (9.57,28.4)	(0.2,45.6)
Human footprint 2009 (index)	4476	15.7 (8.7,26.8)	(0.2,45.6)
Human footprint ratio 2009/1993	4476	0.98 (0.84,1.02)	(0.23,18.09)
Log2 Human footprint ratio (2009/1993)	4476	-0.04 (-0.26,0.02)	(-2.12,4.18)
Strahler order	4476	2 (1,3)	(1,9)

Table 8: Description of the variables included in the model.

Category	Short name	Type	Unit
Response variable	Total abundance	Factor	Count (reference), Number of individuals by 100 sq meters, Ca
	Species richness	Continuous	
	Jaccard dissimilarity	Proportion	
	Dissimilarity	Proportion	
	Appearance	Proportion	
	Disappearance	Proportion	
	Turnover	Proportion	
	Nestedness	Proportion	
	Non-native richness	Proportion	
	Non-native abundance	Proportion	
Predictor variable	Time	Continuous	Number of year
	Stream gradient	Continuous	
	Past pressures	Continuous	
	Recent pressures	Proportion	Factor with 4 levels
	Abundance unit	Factor	
	Basin	Character	
	Site	Character	

Table 9: Variance inflation factors on the variables included in the model. SE factor: inflation of the standard error of slope coefficients predicted by the multicollinearity of the variables.

Predictive variables	VIF	SE factor
Log (Year nb + 1)	1.003852	1.001924
PCA1 stream gradient	1.023532	1.011698
Log2 Human footprint ratio (2009/1993)	1.014258	1.007104
Human footprint (1993)	1.015634	1.007786

Table 10: AIC of the models fitting either temporal trends with the number of year since the start of sampling or the log of the former variable added to one.

Response variables	Year nb	Log (Year nb + 1)	Min AIC	Proportion difference AIC
Jaccard (binary, dissimilarity)	-19607	-27634	Log (Year nb + 1)	0.4093946
Turnover (jaccard)	-20626	-23518	Log (Year nb + 1)	0.1402114
Nestedness (jaccard)	-29776	-32139	Log (Year nb + 1)	0.0793592
Dissimilarity (Simpson index)	-4503	-10399	Log (Year nb + 1)	1.3093493
Appearance	-64809	NA	Year nb	NA
Disappearance	-80995	-85994	Log (Year nb + 1)	0.0617199
Log total abundance	126374	126530	Year nb	0.0012344

2 Biodiversity summary metrics

2.1 Community metrics

2.2 Exotic species

3 Environmental variables

4 Statistical analysis

4.1 Variable

4.2 Collinearity

The presence of multicollinearity among explicative variables was checked by measuring the Variance Inflation Factor (VIF) on a model formulation containing only the main effect, i.e. not the interactions. We did so because main variables (X_1, X_2) and the interactions are collinear by construction, interactions being the products of main variables (i.e. X_1X_2). The VIF values were all close to 1 (Table **??**), indicating absence of multicollinearity.

4.3 Selection of time modelling

4.4 Model validity

4.5 Clustering of temporal trends

5 Supplementary results

5.1 Figure with all the biodiversity facets

```
#> Warning: Removed 2 rows containing non-finite values (stat_ellipse).
#> Warning: Removed 2 rows containing missing values (geom_point).
```

```
#> Warning: Removed 2 rows containing non-finite values (stat_ellipse).
```

```
#> Warning: Removed 2 rows containing missing values (geom_point).
#> Warning: Removed 2 rows containing non-finite values (stat_ellipse).
#> Warning: Removed 2 rows containing missing values (geom_point).
```

5.2 Predictions of the model

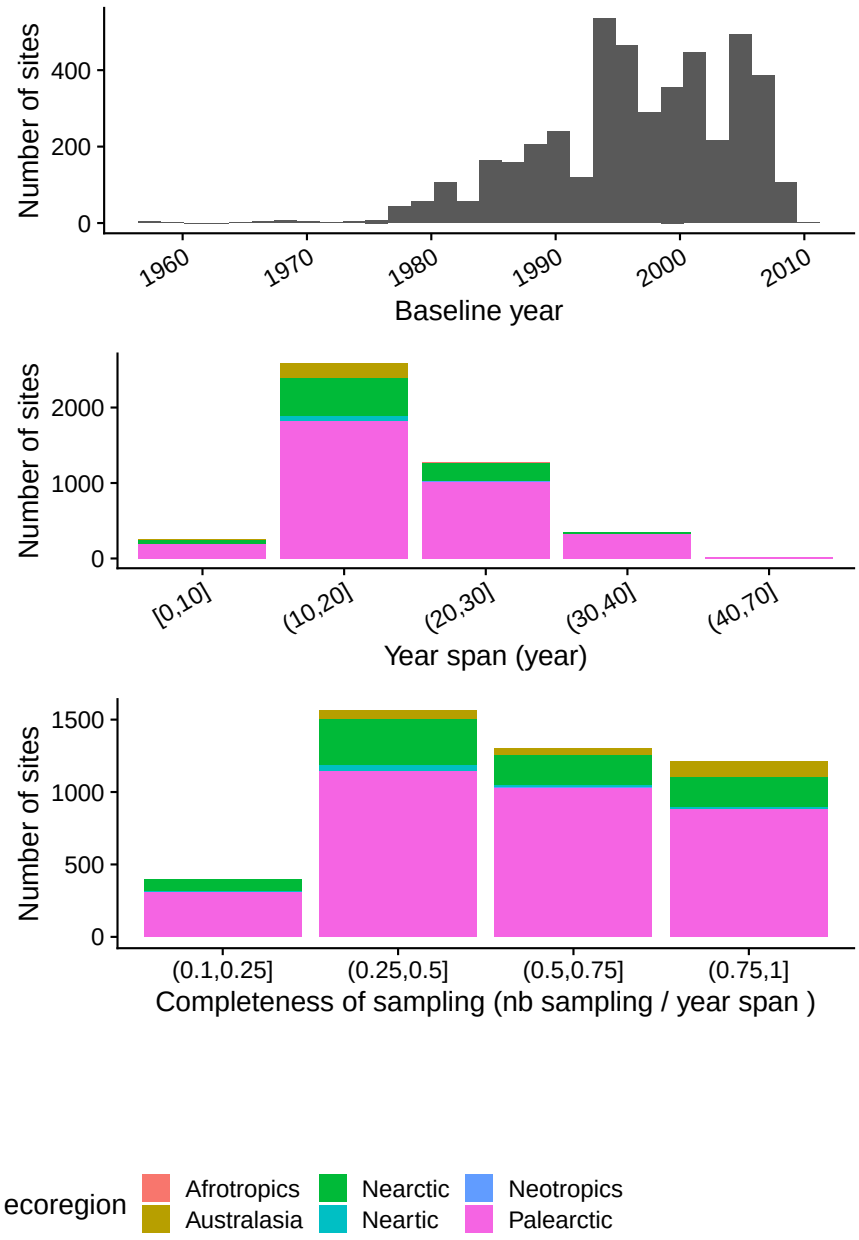

Figure 2: Distribution of (A) Baseline year, (B) year span and (3) completeness of sampling by site.

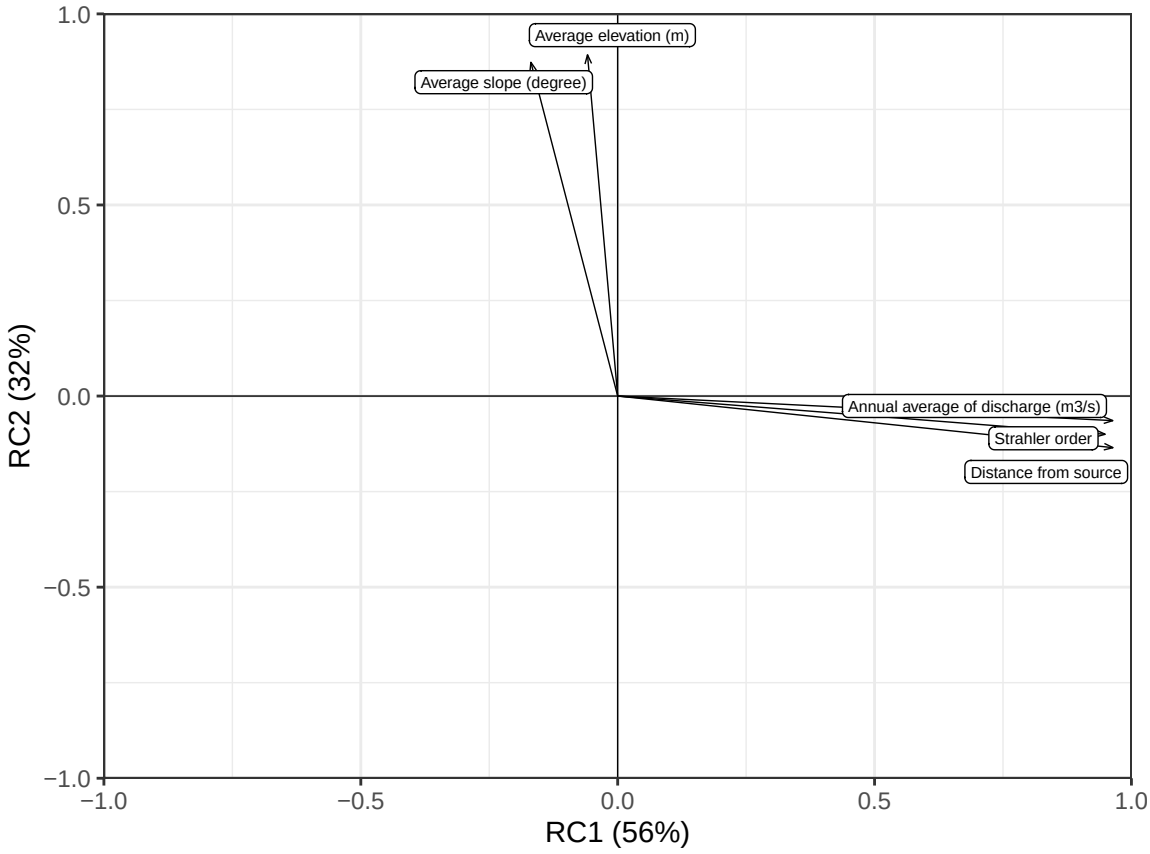

Figure 3: Rotated PCA over the physical and hydrological structure of streams. The first axis is related the average discharge, Strahler order and the distance from source. This first axis was used in the statistical modelling as a composite variable to summarise stream gradient from upstream to upstream.

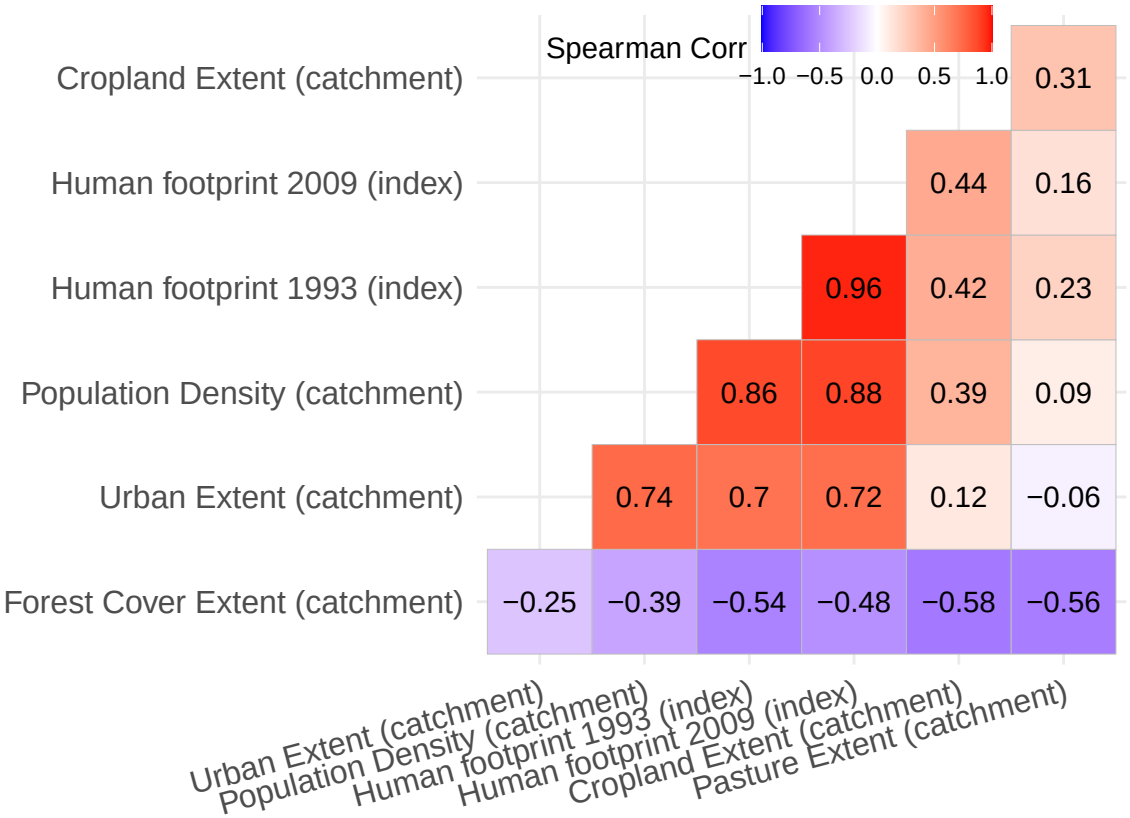

Figure 4: Spearman correlation between land uses and human footprint indexes

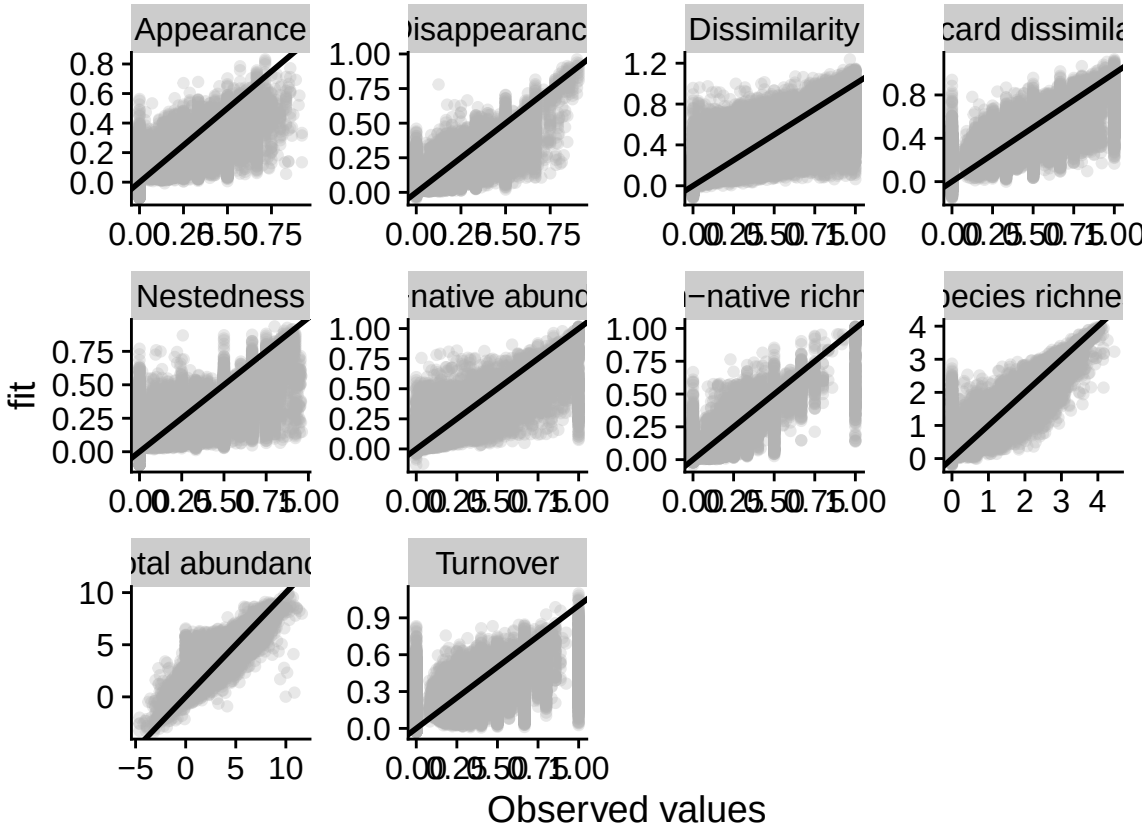

Figure 5: Fitted values by the models vs observed values. The black line displays the bissection, i.e. the intercept and slope are respectively 0 and 1.

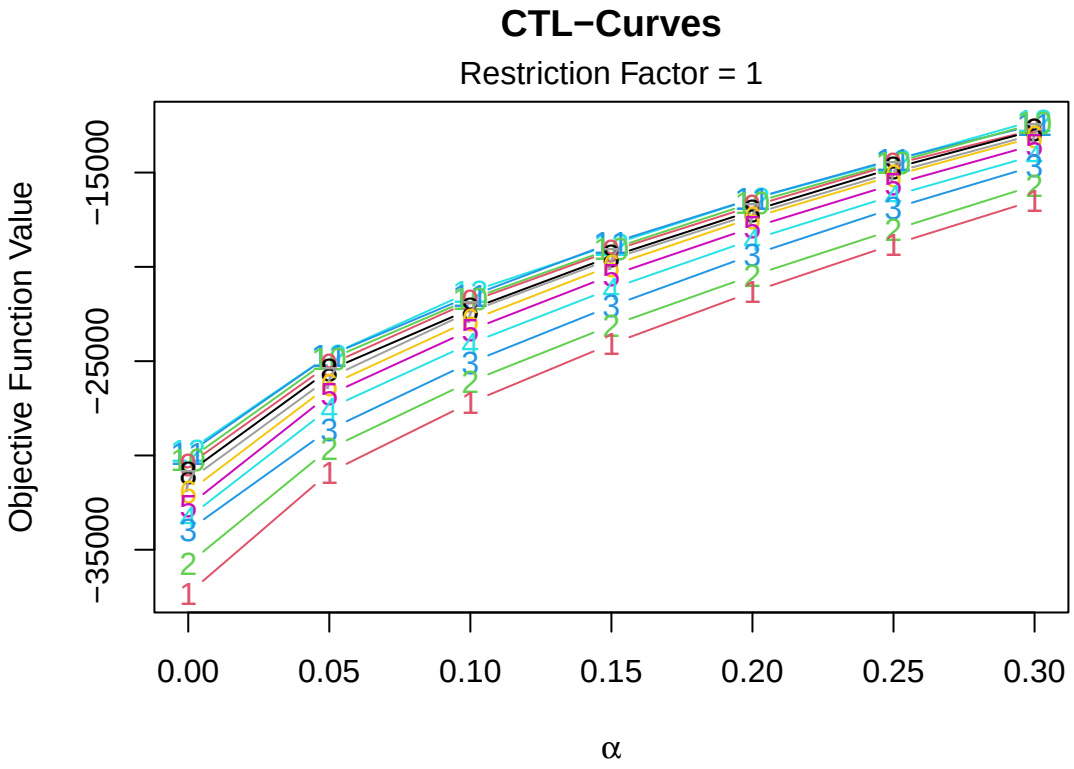

Figure 6: Objective function quantifying the goodness of k-means clustering according the percentage of removed data (most outliers data in the multidimentional space) removed. In the results presented in the main text, we removed 5factors of 1 and 50 respectively. Restricted factor of 1 was used for the main text

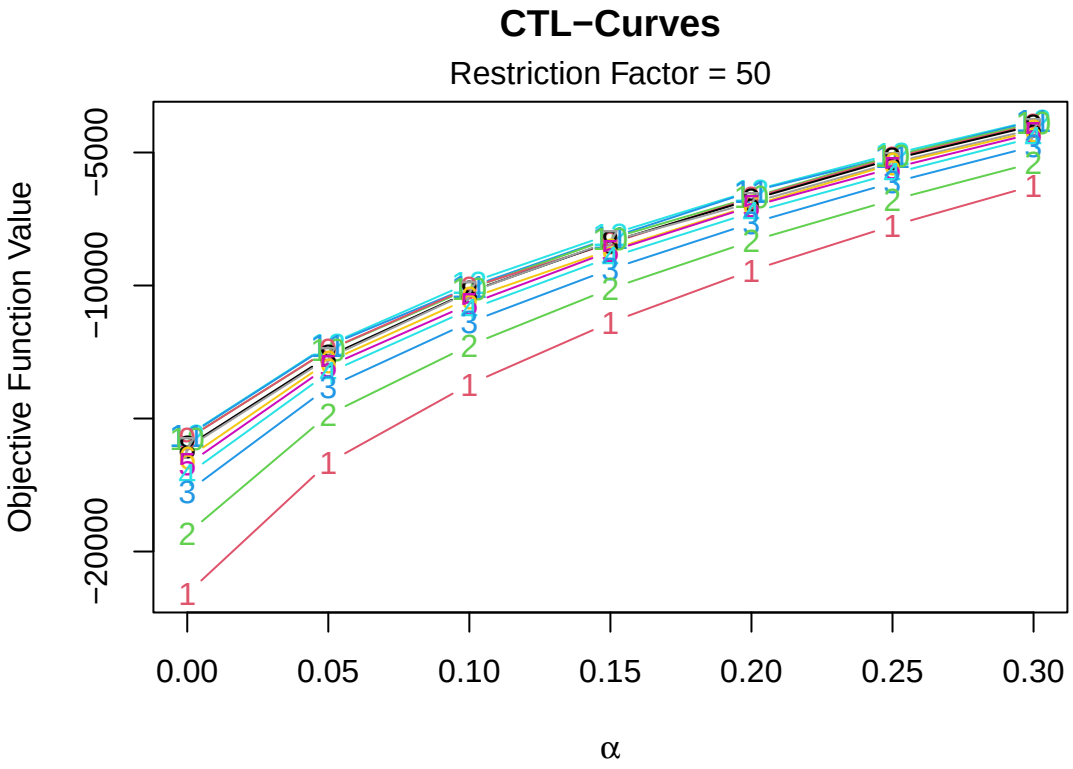

Figure 7: Objective function quantifying the goodness of k-means clustering according the percentage of removed data (most outliers data in the multidimentional space) removed. In the results presented in the main text, we removed 5factors of 1 and 50 respectively. Restricted factor of 1 was used for the main text

Figure 8: Objective function quantifying the goodness of k-means clustering according the percentage of removed data (most outliers data in the multidimentional space) removed. In the results presented in the main text, we removed 5factors of 1 and 50 respectively. Restricted factor of 1 was used for the main text

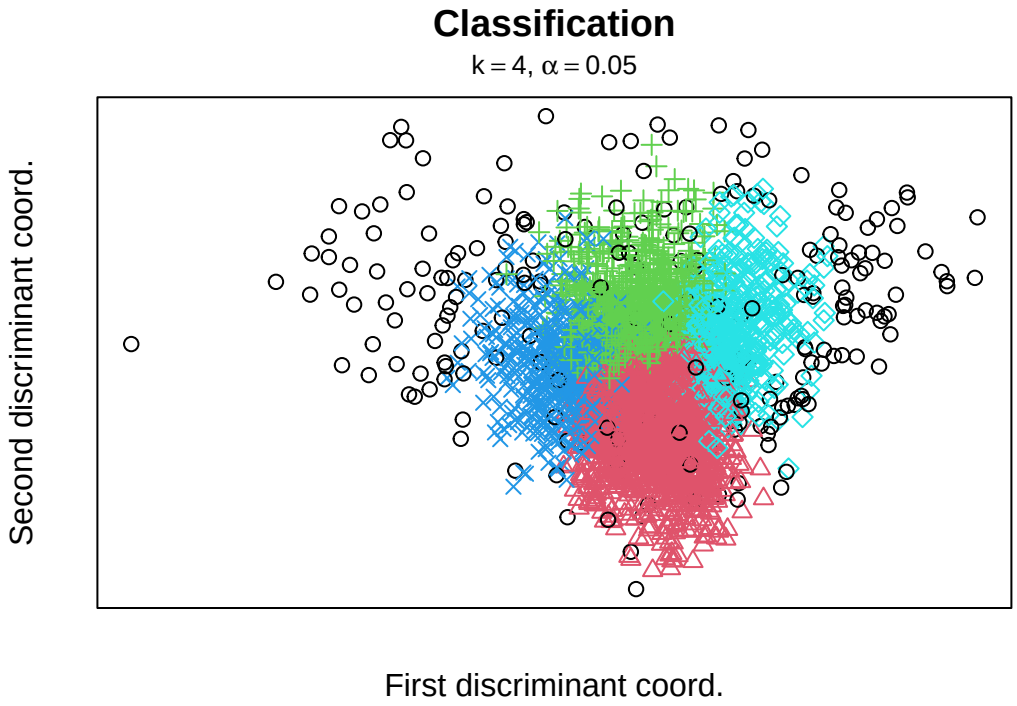

Figure 9: Display of the sites in the multidimentional space, colored by their cluster belonging. Number of clusters = 4

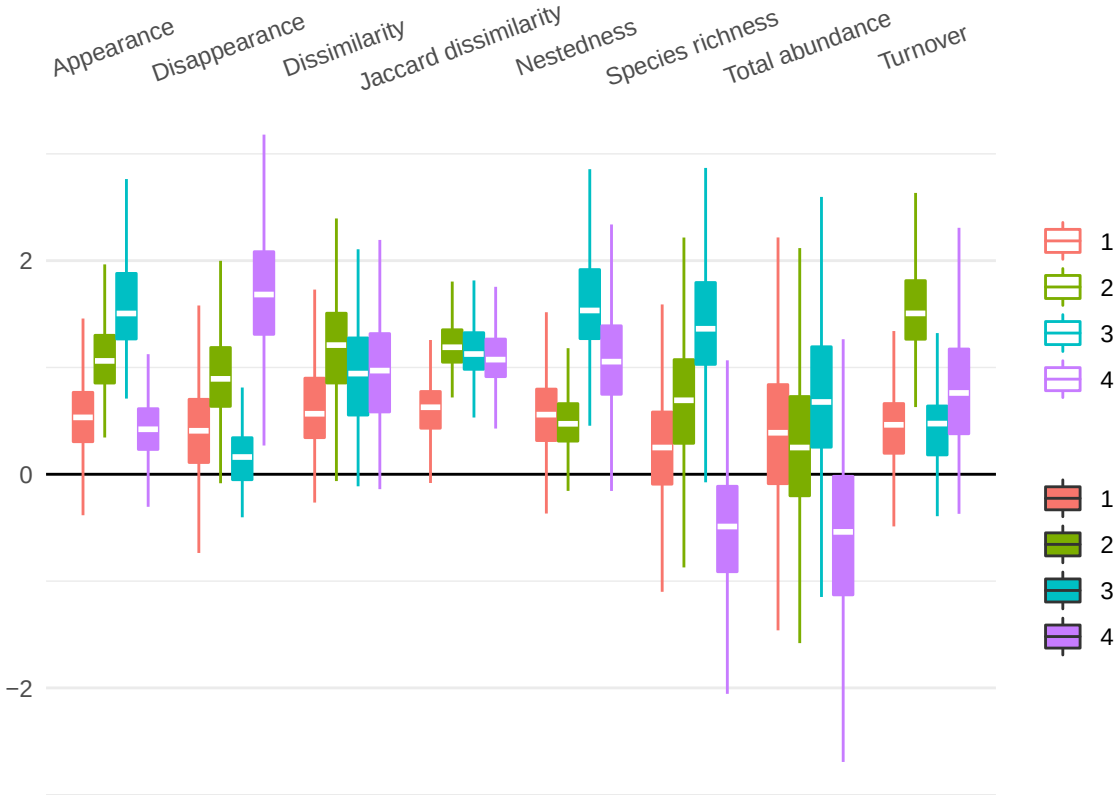

Figure 10: Distribution of scaled temporal trends by variable and cluster, for four clusters. With four cluster instead of six in the main text, the clustering groups temporal trends of species richness and of total abundance and the cluster characterising low change (Fig. 2, main text) is not present.

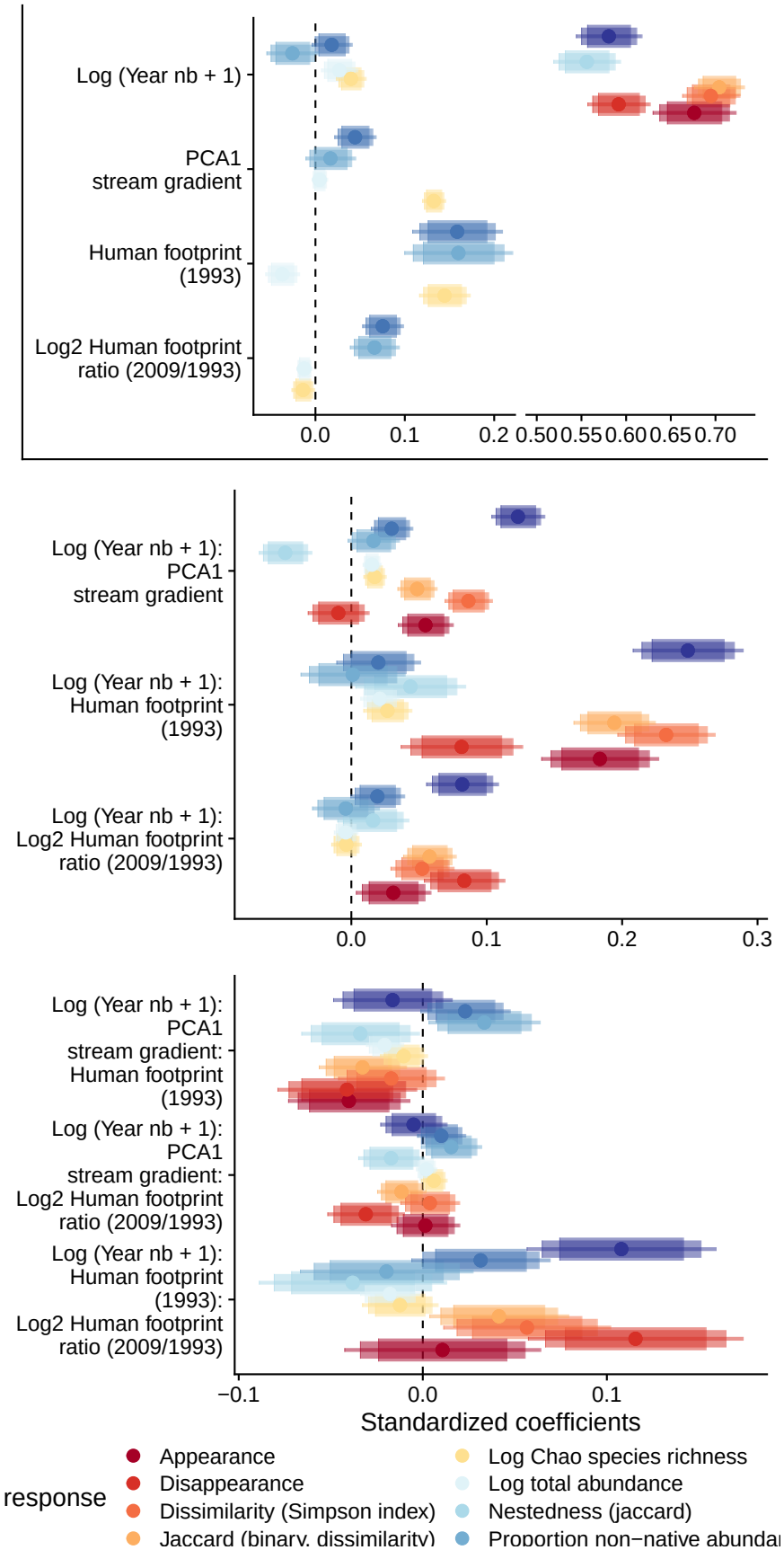

Figure 11: **Responses of community metrics to time, local anthropogenic pressures, stream gradient and their interaction**. See Fig. 1 in the main text for details. To the difference with the Fig. 2, we added Jaccard dissimilarity index and its partitions in two sets of complementary indices, respectively nestedness/turnover, and appearance/disappearance.