Supplementary Figures and Tables

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

11 mai, 2022

Contents

1 Data selection

protocol	n	percent
Electrofishing_backpack	2242	41.8%
Electrofishing	1896	35.4%
Electrofishing_partialsection	608	11.3%
Electrofishing_wholesection	331	6.2%
Electrofishing_boat	86	1.6%
Electrofishing_and_netting	67	1.3%
Seining	56	1.0%
Electrobackpack_and_netting	28	0.5%
RotenoneLockchamber	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	5359	100.0%

unitabundance	n	percent
Ind.100m2	3232	60.3%
Count	2083	38.9%
CPUE	40	0.7%
Leslie_index	4	0.1%
Total	5359	100.0%

Characteristic	N = 5,359
country	
AUS	213 (4.0%)
BEL	18 (0.3%)
BRA	3 (<0.1%)
BWA	1 (<0.1%)
CAN	113 (2.1%)
CIV	2 (<0.1%)
ESP	111 (2.1%)
FIN	126 (2.4%)
FRA	935 (17%)
GBR	1,282 (24%)
HUN	32 (0.6%)
JPN	33 (0.6%)
NOR	4 (<0.1%)
SWE	1,702 (32%)
USA	784 (15%)
ecoregion	
Afrotropics	3 (<0.1%)
Australasia	213 (4.0%)
Nearctic	818 (15%)
Neartic	79 (1.5%)
Neotropics	3 (<0.1%)
Palearctic	4,243 (79%)

[1] n (%)

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

#> `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.
```
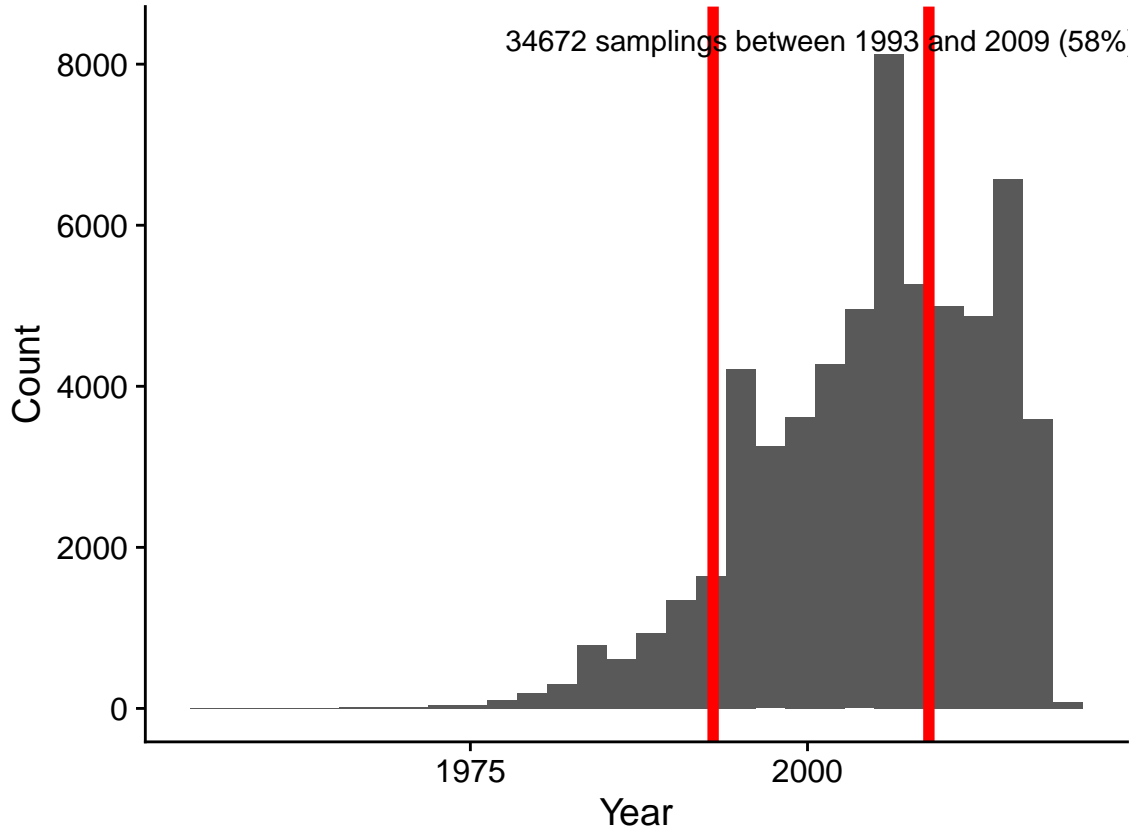
34672 samplings between 1993 and 2009 (58%

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

2 Biodiversity summary metrics

2.1 Biodiversity facets

Characteristic	N (no missing)	Median (1st Q, 3rd Q)	(Min, Max)
Chao species richness	59,904	3.6 (2.0, 7.2)	(1.0, 85.7)
Species richness	59,904	4 (2, 8)	(1, 72)
Log total abundance	59,904	4.19 (3.04, 5.40)	(-4.61, 11.60)
Log Chao species richness	59,904	1.28 (0.67, 1.97)	(0.00, 4.45)
Jaccard (binary, dissimilarity)	59,904	0.36 (0.00, 0.50)	(0.00, 1.00)
Dissimilarity (Simpson index)	59,904	0.23 (0.01, 0.59)	(0.00, 1.00)
Appearance	59,904	0.11 (0.00, 0.23)	(0.00, 0.92)
Disappearance	59,904	0.05 (0.00, 0.18)	(0.00, 0.91)
Turnover (jaccard)	59,904	0.00 (0.00, 0.33)	(0.00, 1.00)
Nestedness (jaccard)	59,904	0.10 (0.00, 0.33)	(0.00, 0.96)
Year	59,904	2,006 (1,999, 2,012)	(1,955, 2,019)

2.2 Exotic species

Characteristic	N (no missing)	Median (1st Q, 3rd Q)	(Min, Max)
species_nb_nat	59,904	4 (2, 7)	(0, 70)
species_nb_exo	59,904	0 (0, 1)	(0, 17)
perc_exo_sp	59,904	0.00 (0.00, 0.05)	(0.00, 1.00)
perc_nat_sp	59,904	1.00 (0.95, 1.00)	(0.00, 1.00)
perc_exo_abun	59,904	0.00 (0.00, 0.00)	(0.00, 1.00)
perc_nat_abun	59,904	1.00 (1.00, 1.00)	(0.00, 1.00)

Characteristic	N = 380,035
native_exotic_origin	
autofishbase	19,152 (5.0%)
handmade	298 (<0.1%)
tedesco	360,428 (95%)
usgs	157 (<0.1%)

[1] n (%)

Characteristic	N (no missing)	Median (1st Q, 3rd Q)	(Min, Max)
species_nb_nat	59,904	4 (2, 7)	(0, 70)
species_nb_exo	59,904	0 (0, 1)	(0, 17)
perc_exo_sp	59,904	0.00 (0.00, 0.05)	(0.00, 1.00)
perc_nat_sp	59,904	1.00 (0.95, 1.00)	(0.00, 1.00)
perc_exo_abun	59,904	0.00 (0.00, 0.00)	(0.00, 1.00)
perc_nat_abun	59,904	1.00 (1.00, 1.00)	(0.00, 1.00)

3 Environmental variables

Characteristic	N (no missing)	Median (1st Q, 3rd Q)	(Min, Max)
Distance from source	5,359	25 (13, 53)	(3, 2,798)
Strahler order			
1	1,649 (31%)	1,649 (31%)	1,649 (31%)
2	1,589 (30%)	1,589 (30%)	1,589 (30%)
3	1,123 (21%)	1,123 (21%)	1,123 (21%)
4	520 (9.7%)	520 (9.7%)	520 (9.7%)
5	287 (5.4%)	287 (5.4%)	287 (5.4%)
6	95 (1.8%)	95 (1.8%)	95 (1.8%)
7	82 (1.5%)	82 (1.5%)	82 (1.5%)
8	9 (0.2%)	9 (0.2%)	9 (0.2%)
9	5 (<0.1%)	5 (<0.1%)	5 (<0.1%)
Annual average of discharge (m3/s)	5,359	1 (1, 6)	(0, 7,831)
Average elevation (m)	5,359	170 (88, 287)	(-1, 2,531)
Average slope (degree)	5,359	31 (17, 49)	(0, 362)
Human footprint (1993)	5,359	15 (8, 26)	(0, 46)
Log2 Human footprint ratio (2009/1993)	5,359	-0.04 (-0.26, 0.02)	(-2.12, 4.18)

4 Statistical analysis

4.1 Collinearity

Term	VIF	SE_factor
Log (Year nb + 1)	1.004869	1.002432
PCA1 stream gradient	1.043300	1.021420
Log2 Human footprint ratio (2009/1993)	1.008678	1.004330
Human footprint (1993)	1.040998	1.020293

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. It is because main variables (X_1, X_2) and the interactions are colinear by construction, interactions being the products of main variables (i.e. $X_1 X_2$). The VIF values were all close to 1 (Table **??**), indicating absence of multicollinearity.

4.2 Selection of time modelling

```
#> [1] 0.2867481
```

response	year_nb	log1_year_nb	which_min	diff_ratio_log_raw
Jaccard (binary, dissimilarity)	-21492	-30846	log1_year_nb	0.4352317
Turnover (jaccard)	-23296	-26522	log1_year_nb	0.1384787
Nestedness (jaccard)	-31933	-34971	log1_year_nb	0.0951367
Dissimilarity (Simpson index)	-5720	-12805	log1_year_nb	1.2386364
Appearance	-78449	-81625	log1_year_nb	0.0404849
Disappearance	-100594	-106498	log1_year_nb	0.0586914
Log total abundance	159541	159633	year_nb	0.0005767

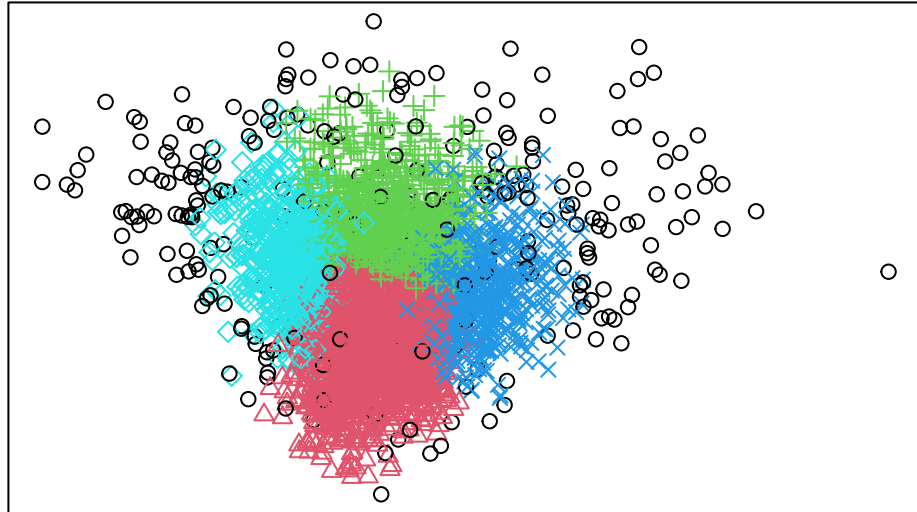## 4.3 Model validity

4.4 Clustering of temporal trends

```
#> Warning in as_grob.default(plot): Cannot convert object of class ctlcurves into a grob.
```

```
#> Warning in as_grob.default(plot): Cannot convert object of class ctlcurves into a grob.
```
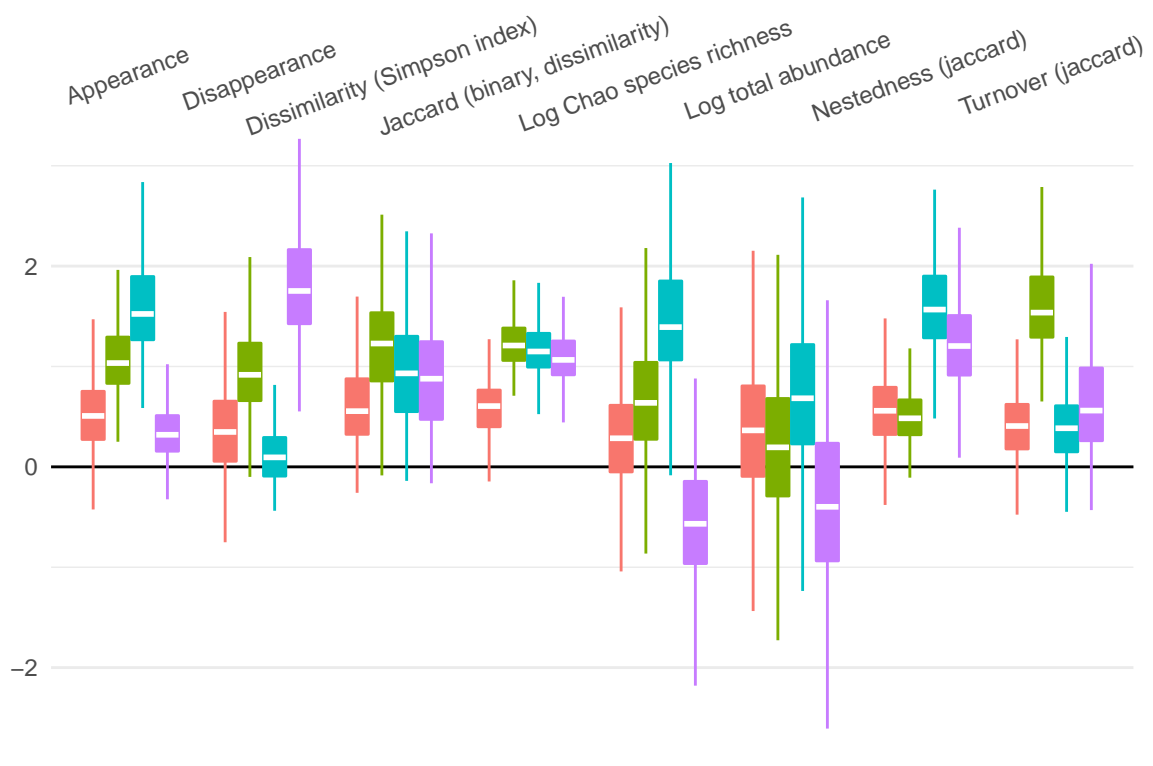
Classification

k = 4, α = 0.05

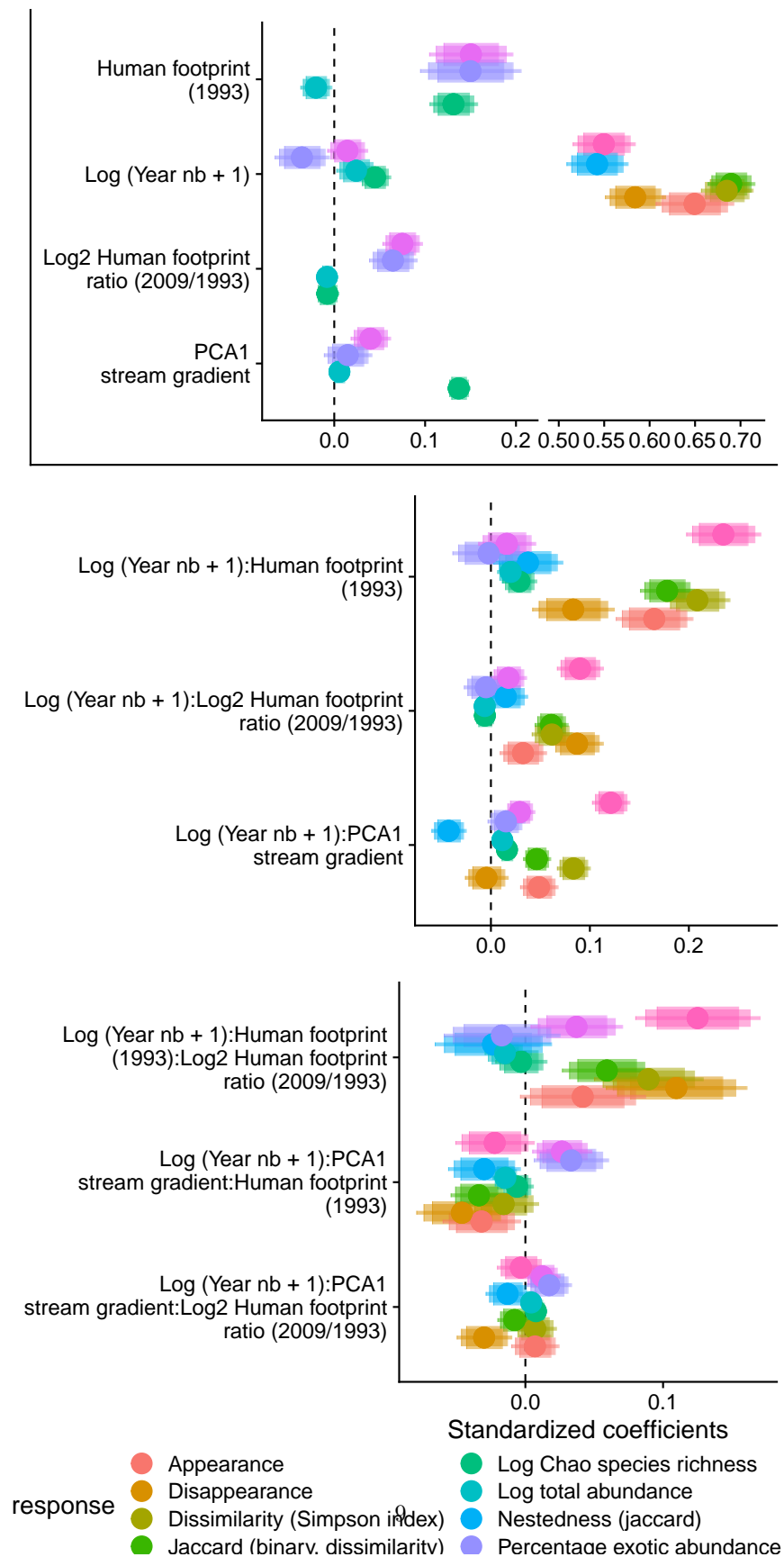

Second discriminant coord.

First discriminant coord.

- Add box plot of cluster values distribution by variables

5 Supplementary results

5.1 Figure with all the biodiversity facets

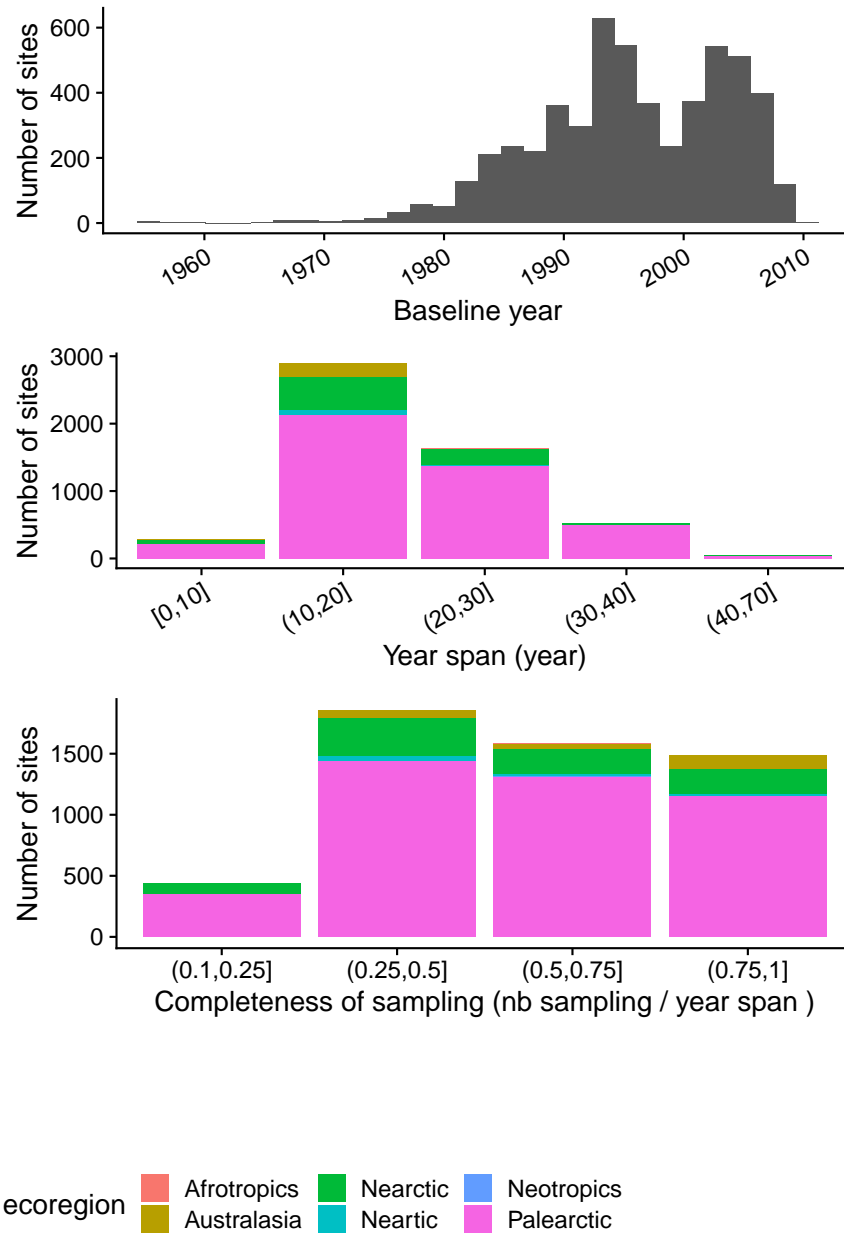**5.2   Unstandardized coefficients**

5.3   Predictions of the model

Figure 2: Distribution of (A) Baseline year, (B) year span and (3) completeness of sampling by site.
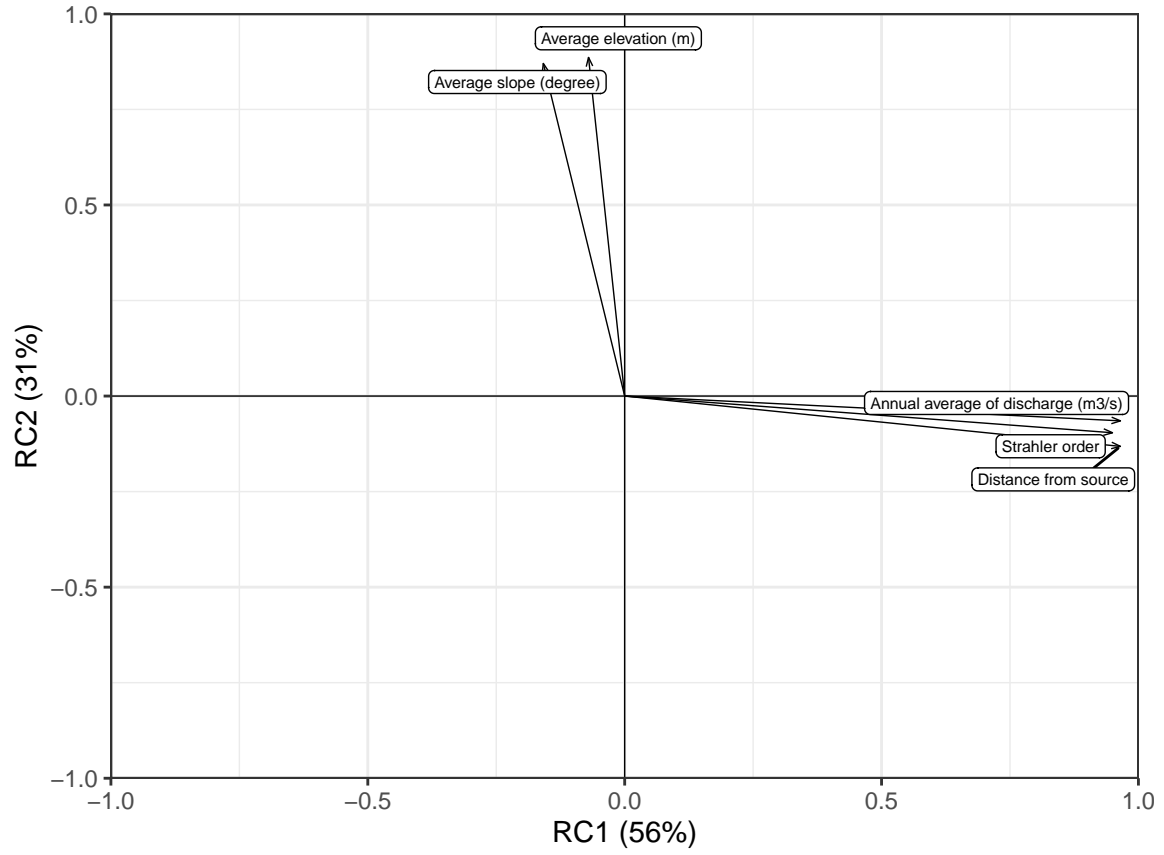
Figure 3: Rotated PCA over the physical and hydrological structure of streams. The first axis is related the average discharge, Strahler order and the distance from source. This first axis was used in the statistical modelling as a composite variable to summarise stream gradient from upstream to upstream.
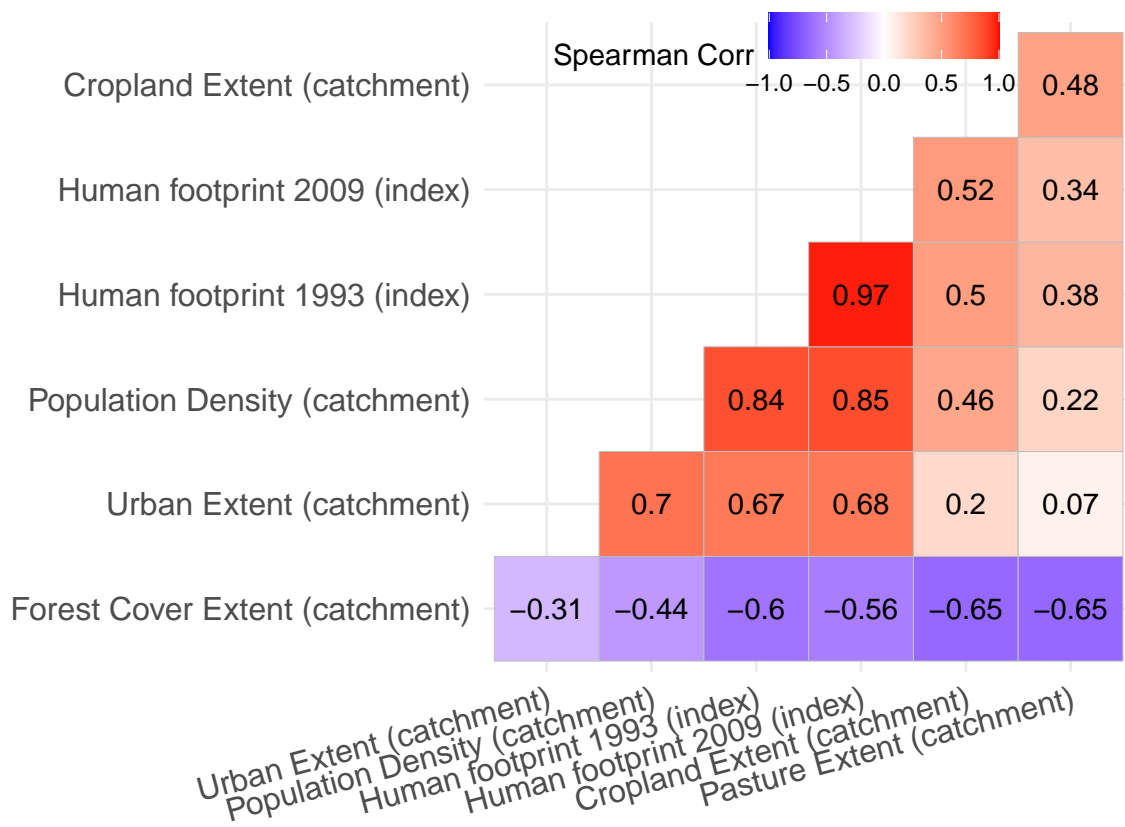
Figure 4: Spearman correlation between land uses and human footprint indexes

Figure 5: 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.