

Supplementary Figures and Tables

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

12 juillet, 2022

Contents

1	Dataset	1
2	Biodiversity summary metrics	1
2.1	Community metrics	1
2.2	Exotic species	1
3	Environmental variables	1
4	Statistical analysis	2
4.1	Variable	2
4.2	Collinearity	2
4.3	Time modelling	5
4.4	Model validity	5
4.5	Clustering of temporal trends	5
5	Supplementary results	5
5.1	Figure with all the biodiversity facets	5
5.2	Predictions of the model	5
5.3	Random effect	5
•	Add plot species richness / chao richness	

1 Dataset

```
#> `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.
```

2 Biodiversity summary metrics

2.1 Community metrics

2.2 Exotic species

3 Environmental variables

Table S1: Databases included in the current study that were not included in RivFishTIME publication (Comte et al., 2021a). Apart from MARIS, these databases have already been used in a previous study (Comte et al., 2021b). N: number of sites.

Source id	Country	N	Reference
MARIS	USA	227	U.S. Geological Survey, Core Science Analytics and Synthesis Program, 20131201, Multistate Aquatic Resources Information System (MARIS): United States Geological Survey, http://dx.doi.org/10.5066/F7BZ641R .
Maryland	USA	73	Montgomery county monitoring program (2018). Available at https://www.montgomerycountymd.gov/water/streams/data.html
Ohio	USA	40	Ohio statewide monitoring program (2018). Available at https://www.orsanco.org/programs/fish-population/
RAMP	Canada	13	Regional Aquatics Monitoring Program (2018). Available at http://www.ramp-alberta.org/RAMP.aspx

Table S2: 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%

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

Table S3: Abundance unit across sites

Abundance unit	N	Percent
Individual number per 100m ²	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 S4: 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%

Table S5: 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)

multicollinearity.

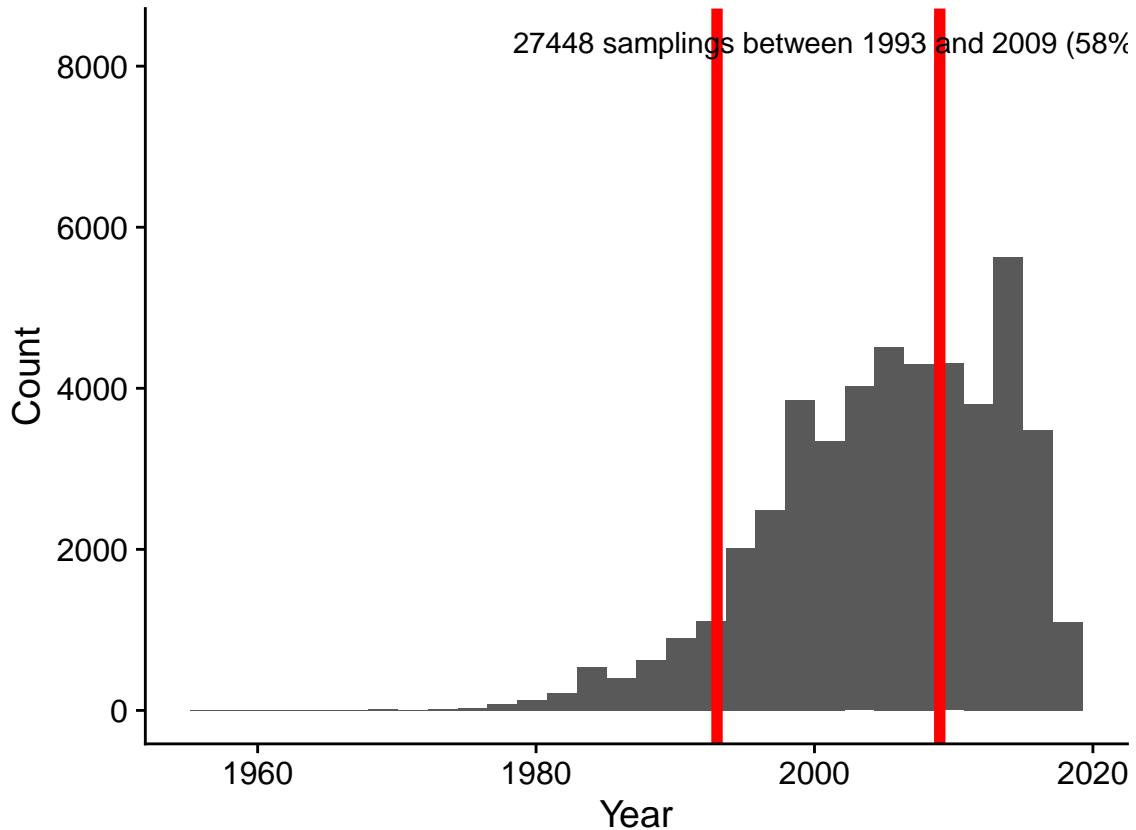


Figure S1: Distribution of year of samplings. Years 1993 and 2009, corresponding to the two human footprint measurements, are highlighted.

Table S6: 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 S7: 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 S8: Descriptive statistics for the site environmental descriptors.

Environmental variables	N	Median (Q1, Q3)	(Min, Max)
Annual average of discharge (m ³ /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 (km)	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)

4.3 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

```
#> [1] "log1_year_nb"           "hft_ix_c93"          "riv_str_rc1"        "hft_ix_c9309_log2
```

5.3 Random effect

Table S9: Description of the variables included in the model.

Category	Short name	Type	Unit	Signification
Response variable	Total abundance	Numeric	cf Abundance unit	Total number of individuals
Response variable	Species richness	Continuous		Coverage corrected richness
Response variable	Jaccard dissimilarity	Proportion		Temporal dissimilarity (presence/absence)
Response variable	Dissimilarity	Proportion		Temporal dissimilarity (relative abundance, Simpson index)
Response variable	Appearance	Proportion		Temporal species colonisation
Response variable	Disappearance	Proportion		Temporal species extirpation
Response variable	Turnover	Proportion		Temporal species replacement
Response variable	Nestedness	Proportion		Temporal nestedness
Response variable	Non-native richness	Proportion		Proportion of non-native species richness
Response variable	Non-native abundance	Proportion		Proportion of non-native species abundance
Predictor variable	Time	Continuous	Number of year	Time
Predictor variable	Stream gradient	Continuous		PCA axis related to average distance from source, Strahler order, stream size
Predictor variable	Past pressures	Continuous		Human footprint (1993)
Predictor variable	Recent pressures	Proportion		Human footprint ratio between 2009 and 1993
Predictor variable	Abundance unit	Factor	Factor with 4 levels: Count (reference), Number of individuals by 100 sq meters, Catch per Unit Effort (CPUE), Leslie index	Unit of abundance
Predictor variable	Basin	Character		Hydrographic basin
Predictor variable	Site	Character		Sampled site

Table S10: 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

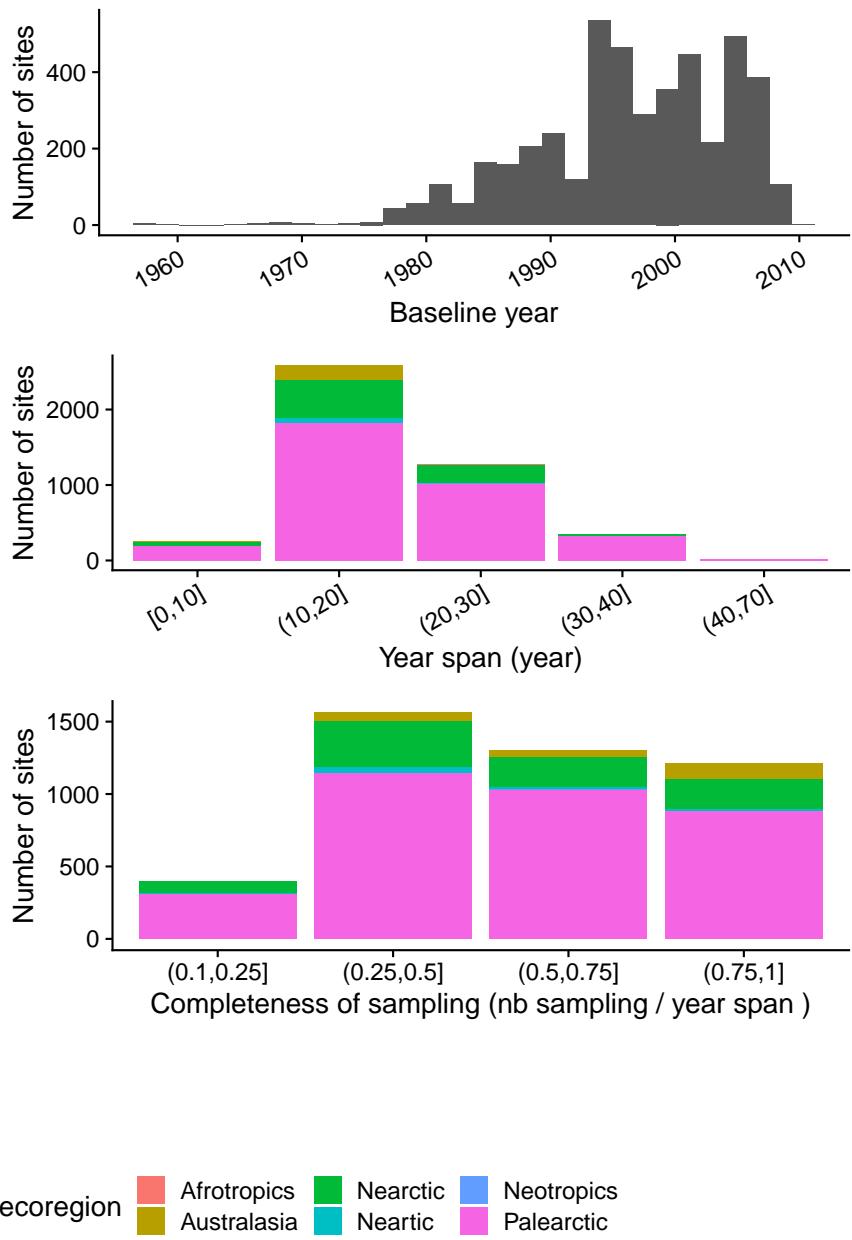


Figure S2: Distribution of (A) Baseline year, (B) year span and (3) completeness of sampling by site.

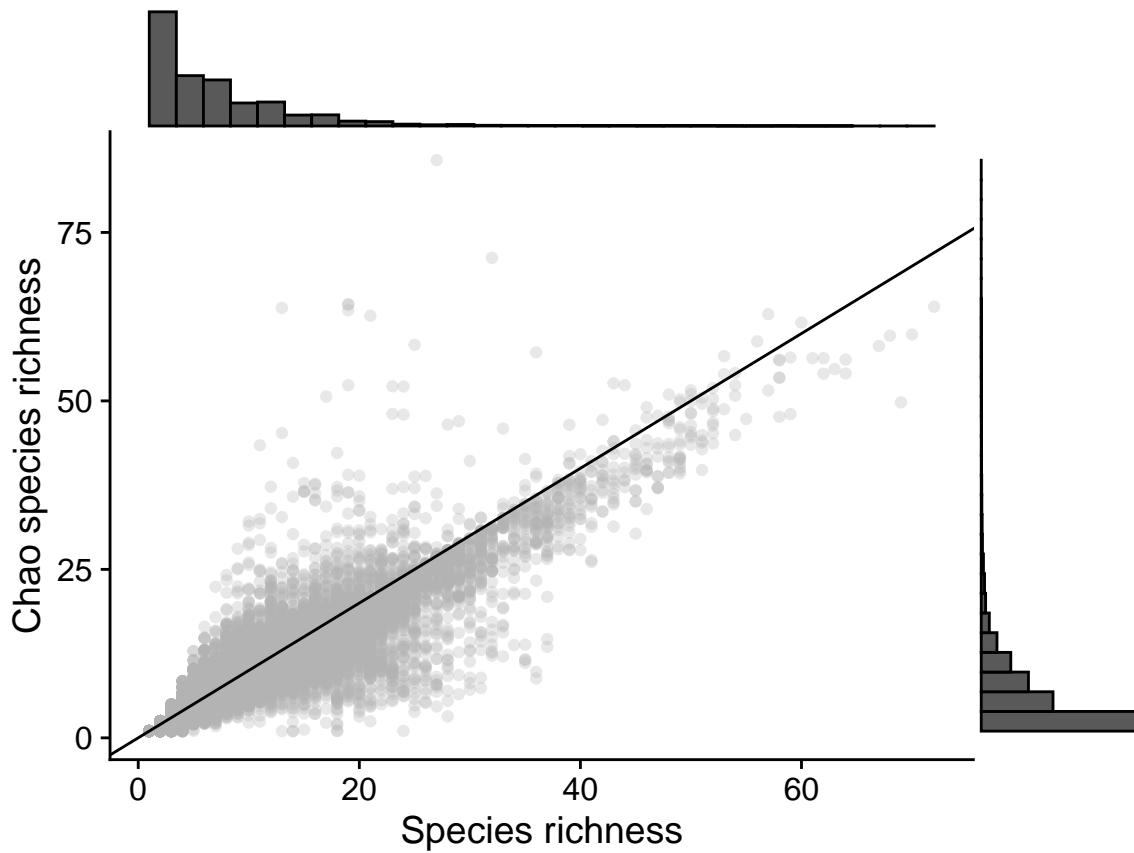


Figure S3: Relationship between Chao-corrected species richness and species richness, added to their respective distribution. The black line displays the bissection, i.e. the intercept and slope are respectively 0 and 1. Spearman's correlation is equal to 0.97.

Table S11: WAIC of the statistical models fitting temporal trends with either the number of year since the start of sampling or the log of the former variable added to one.

Response	WAIC			
	Log (Year nb + 1)	Year nb	Min WAIC	WAIC difference (to Year nb)
Jaccard (binary, dissimilarity)	4445	5316	Log (Year nb + 1)	-16.4%
Turnover (jaccard)	4042	4772	Log (Year nb + 1)	-15.3%
Nestedness (jaccard)	3967	4762	Log (Year nb + 1)	-16.7%
Dissimilarity (Simpson index)	4556	5608	Log (Year nb + 1)	-18.8%
Appearance	4589	5471	Log (Year nb + 1)	-16.1%
Disappearance	4417	5765	Log (Year nb + 1)	-23.4%
Log Chao species richness	5445	5386	Year nb	1.1%
Log total abundance	5416	5540	Log (Year nb + 1)	-2.2%
Proportion non-native species	5433	5388	Year nb	0.8%
Proportion non-native abundance	5469	5517	Log (Year nb + 1)	-0.9%

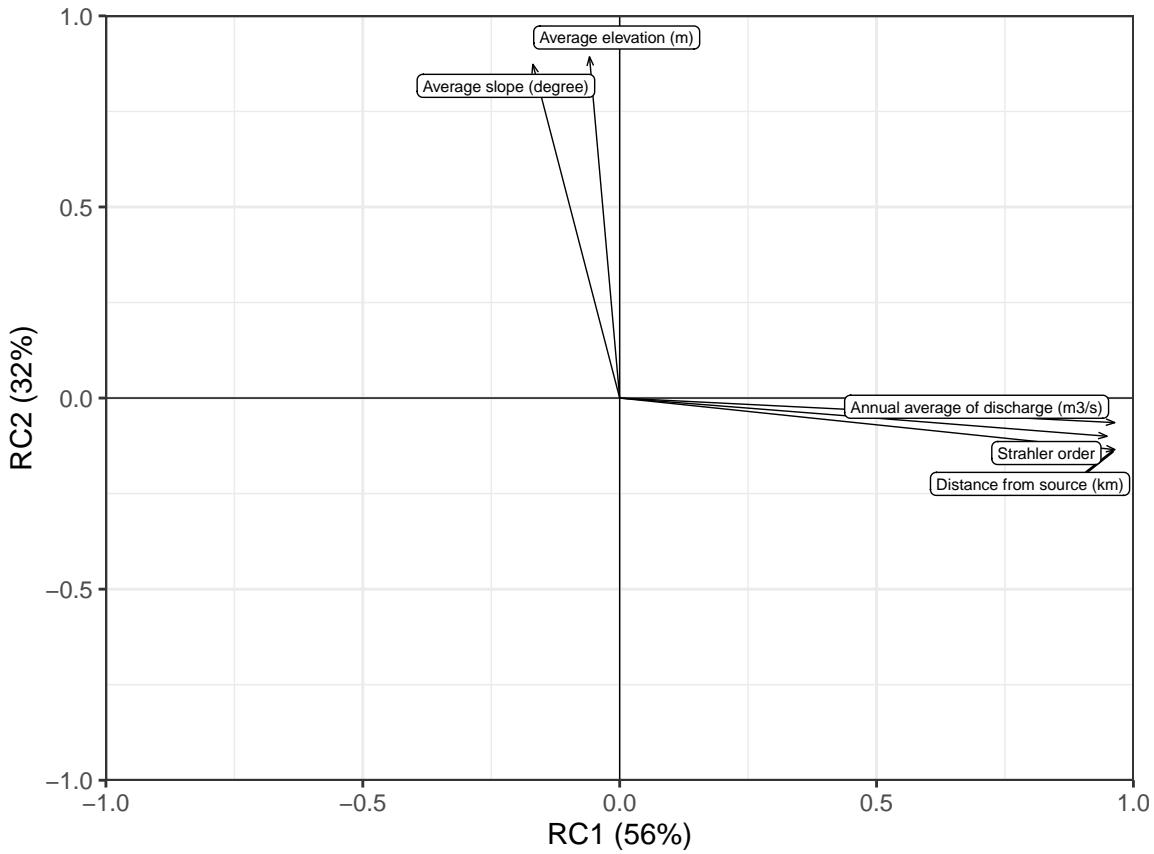


Figure S4: Rotated PCA over the physical and hydrological structure of streams. The first axis is related the average discharge, Strahler order and the distance from source. This first axis was used in the statistical modelling as a composite variable to summarise stream gradient from upstream to downstream.

Table S12: Bayesian marginal and conditional R-squared from the models with ecological drivers. Details of computations are presented in methods.

Response variable	Marginal (s.d.)	Conditional (s.d.)
Non-native richness	0.02 (0.02)	0.76 (0)
Non-native abundance	0.03 (0.02)	0.75 (0)
Jaccard dissimilarity	0.05 (0.03)	0.41 (0.02)
Turnover	0.06 (0.05)	0.14 (0.04)
Nestedness	0.05 (0.04)	0.21 (0.03)
Dissimilarity	0.06 (0.04)	0.29 (0.03)
Appearance	0.06 (0.05)	0.2 (0.04)
Disappearance	0.06 (0.05)	0.21 (0.04)
Species richness	0.02 (0.01)	0.8 (0)
Total abundance	0.03 (0.02)	0.7 (0.01)

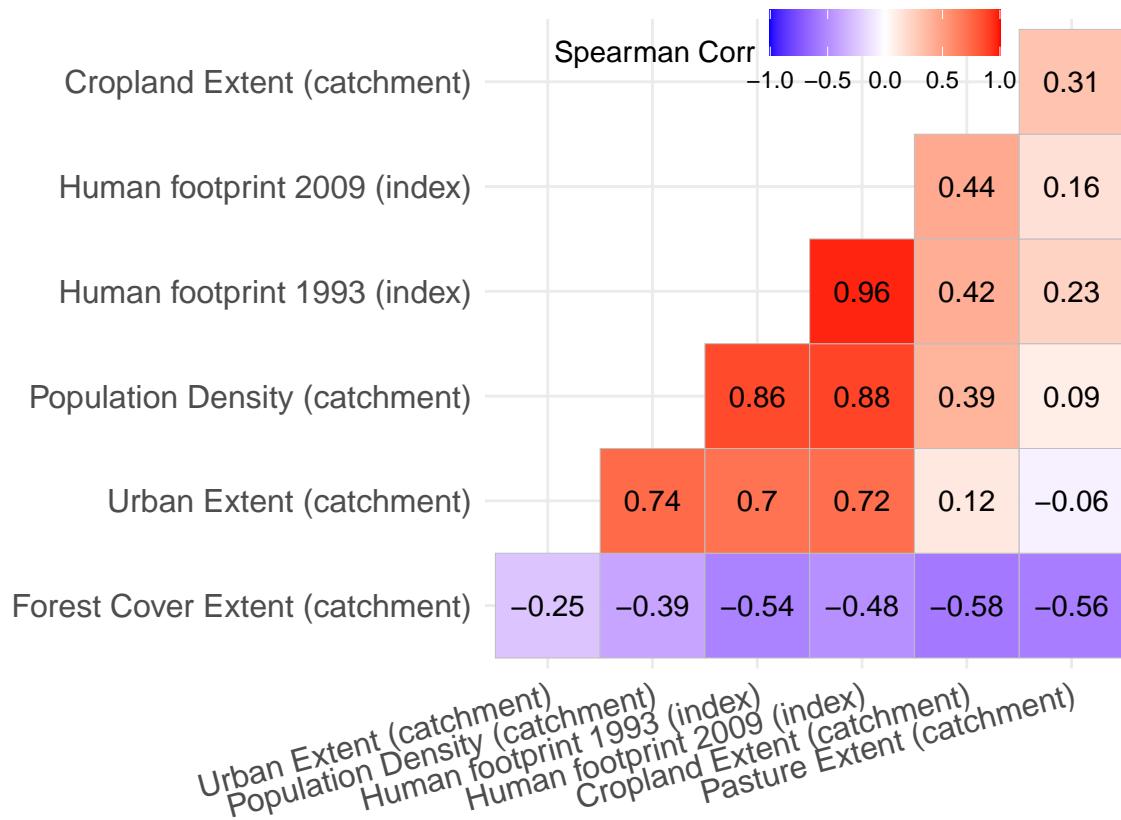


Figure S5: Spearman correlation between land uses and human footprint indexes

Table S13: Values and signification of predictor variables used in the main text.

Driver	Predictor variable	Value	Signification
Time	Log (Year nb + 1)	0.0	Year nb of sample for a site = 0, i.e. baseline year
		2.4	Year nb of sample for a site = 10, i.e. a decade
Past pressures	Human footprint (1993)	2.5	Median intact site (human footprint index between 1 and 4, Williams et al. 2020)
		45.6	Most degraded site in the dataset, i.e. highest human footprint index
Stream gradient	PCA1 stream gradient	-0.2	Median stream gradient
Recent pressures	Log2 Human footprint ratio (2009/1993)	-1.0	Anthropogenic pressures divided by two between 1993 and 2009
		1.0	Anthropogenic pressures multiplied by two between 1993 and 2009

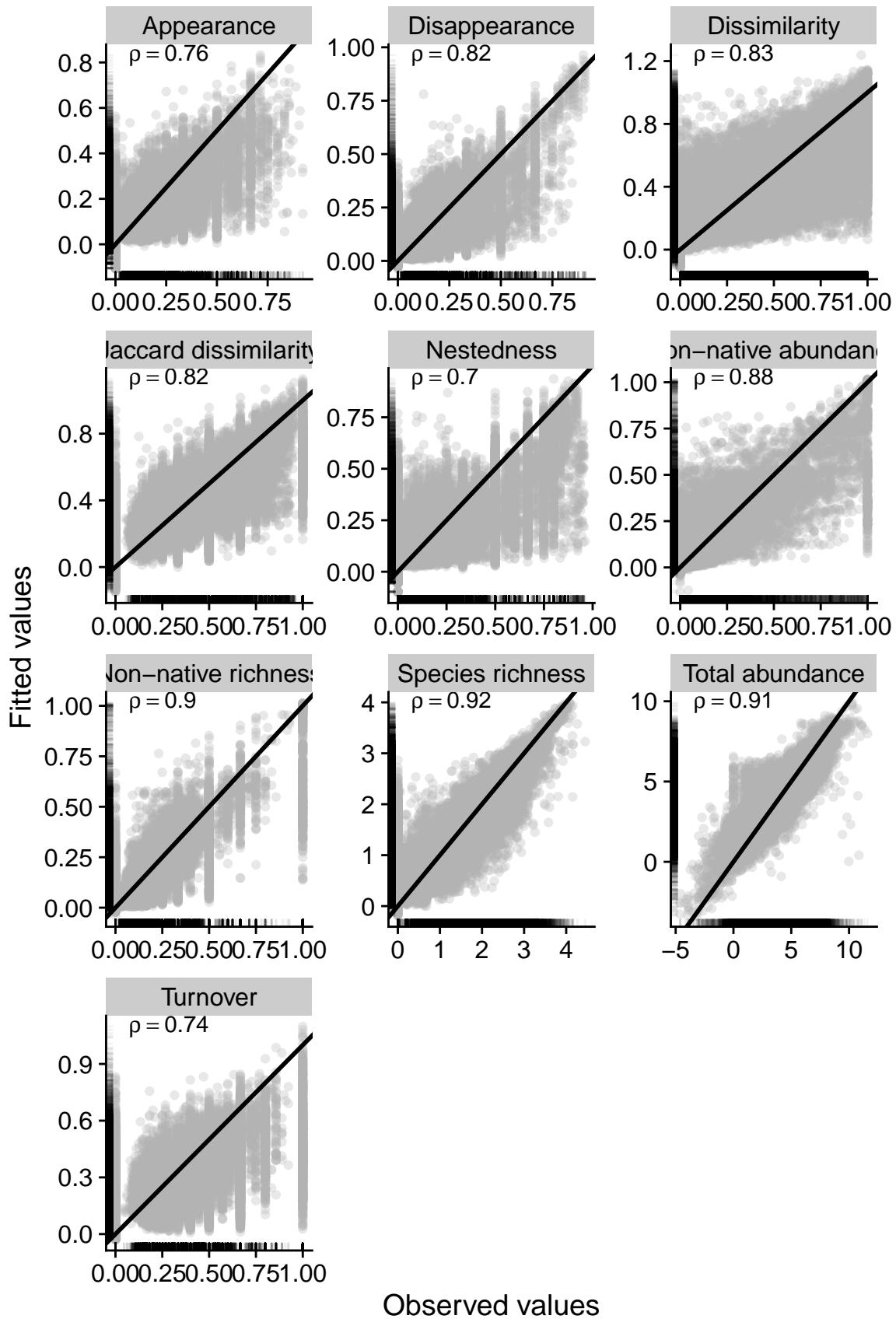


Figure S6: Fitted values by the models vs observed values. The black line displays the bisection, i.e. the intercept and slope are respectively 0 and 1. Sticks located at the abscissa and ordinate axis display value concentration. The Pearson's correlation coefficient is displayed in each panel.

Figure S7: Objective function quantifying the goodness of k-means clustering according the percentage of removed data (most outliers data in the multidimensional 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

Second discriminant coord.

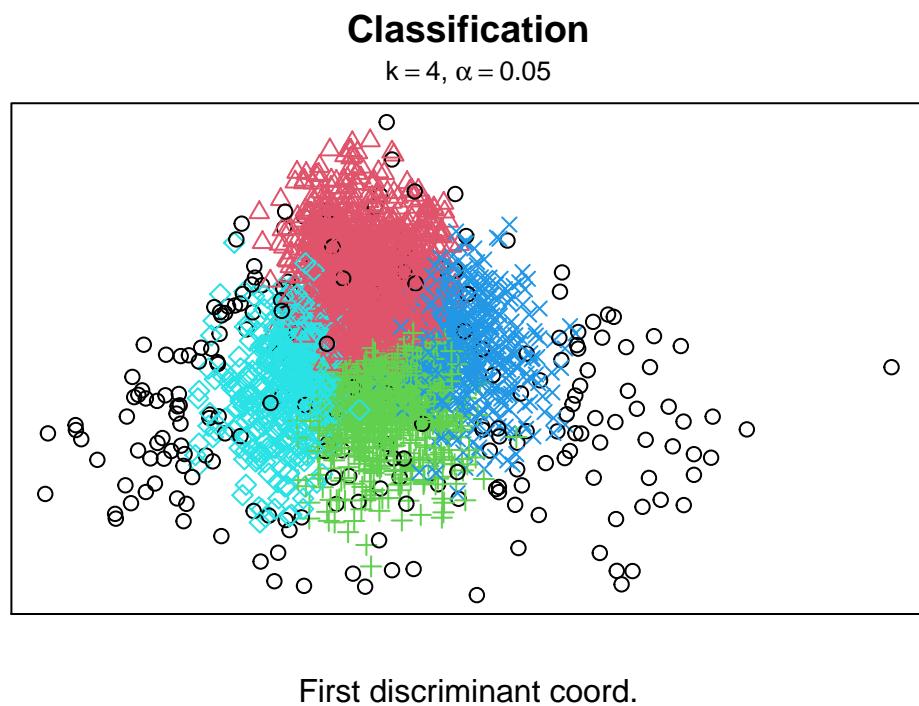


Figure S8: Display of the sites in the multidimensional space, colored by their cluster belonging. Number of clusters = 4

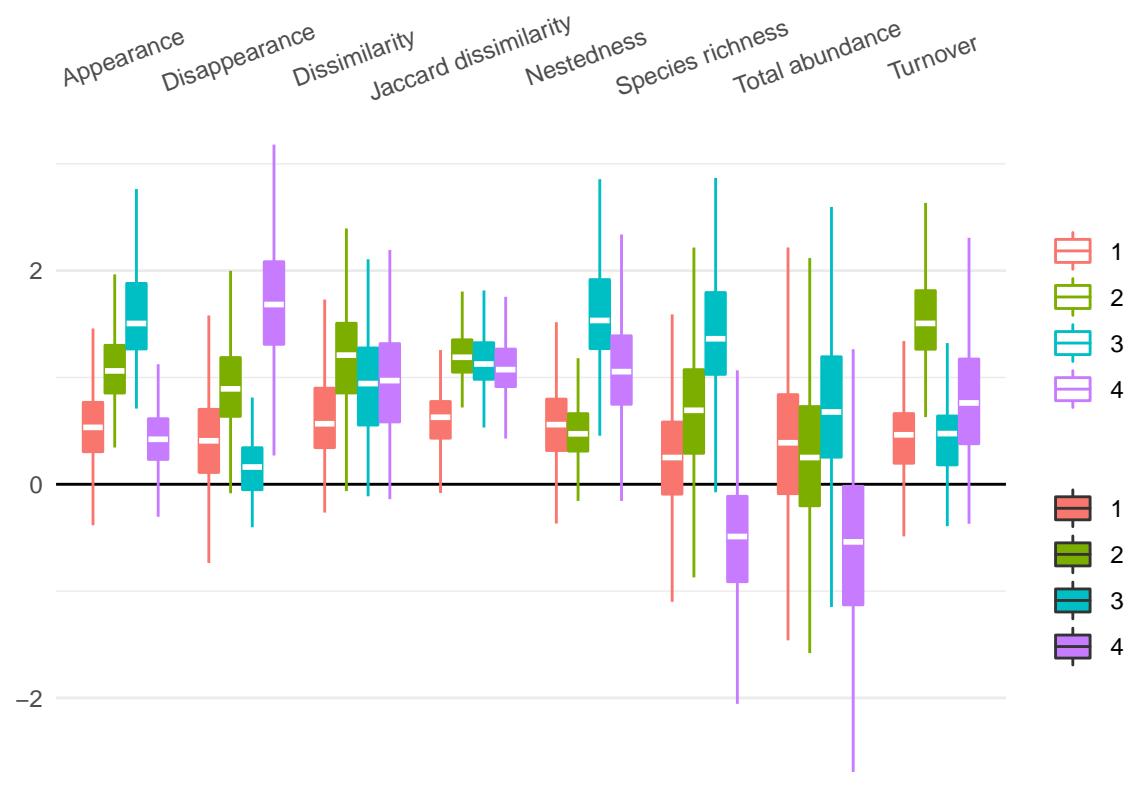


Figure S9: Distribution of scaled temporal trends by variable and cluster, for four clusters. With four cluster instead of six in the main text, the clustering groups temporal trends of species richness and of total abundance and the cluster characterising low change (Fig. 2, main text) is not present.

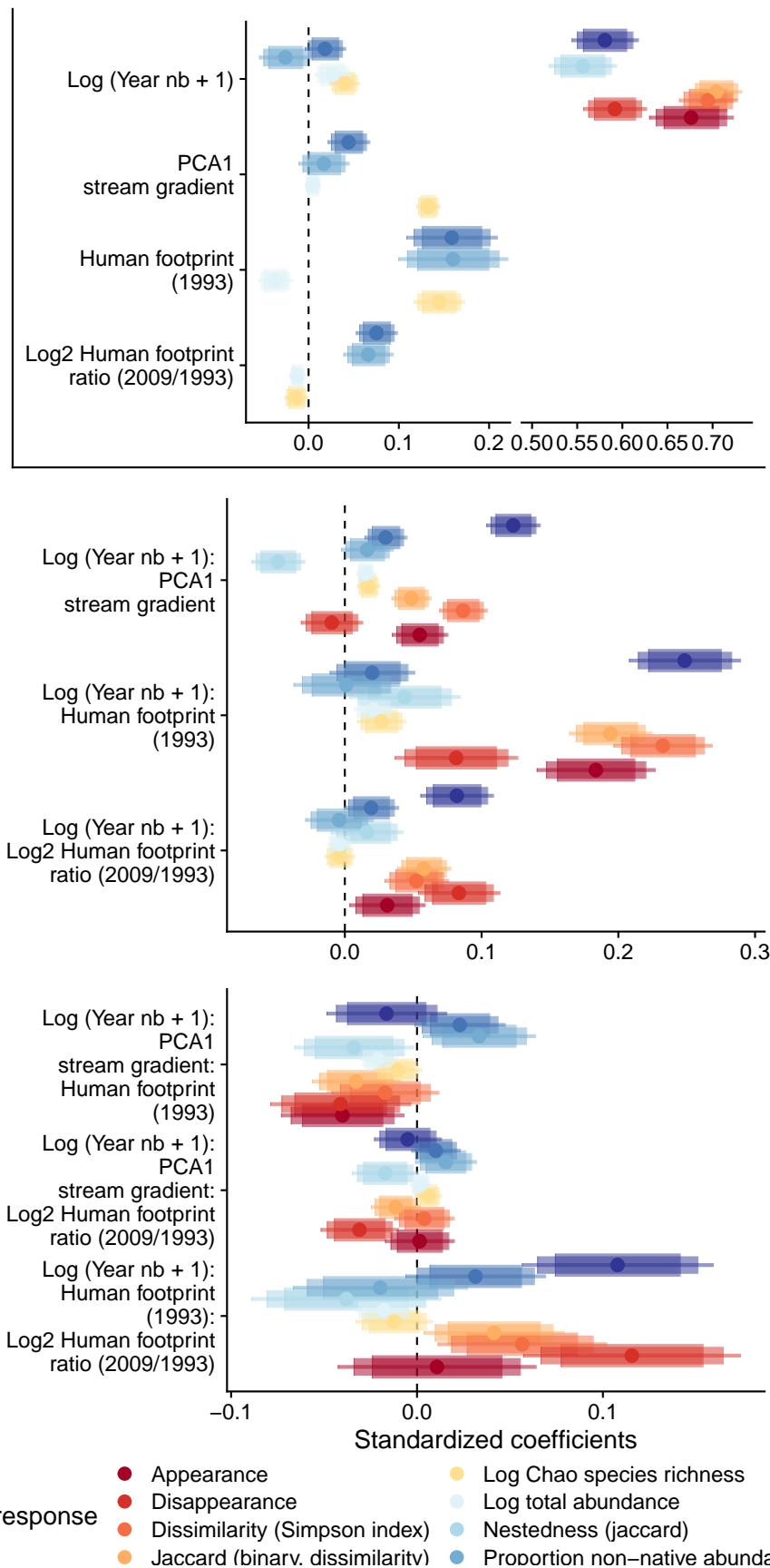


Figure S10: **Responses of community metrics to time, local anthropogenic pressures, stream gradient and their interaction**. See Fig. 1 in the main text for details. To the difference with the Fig. 2, we added Jaccard dissimilarity index and its partitions in two sets of complementary indices, respectively nestedness/turnover, and appearance/disappearance.

Table S14: Standard deviation estimated by the models for the random effects and the error term.

Response variable	Term	Mean [95% CI]
Appearance	Time (site nested in basin)	0.045 [0.044,0.046]
	Time (basin)	0.026 [0.023,0.03]
	Intercept (site nested in basin)	0.015 [0.011,0.021]
	Intercept (basin)	0.042 [0.036,0.055]
	Error	0.106 [0.105,0.107]
Disappearance	Time (site nested in basin)	0.047 [0.046,0.048]
	Time (basin)	0.019 [0.017,0.022]
	Intercept (site nested in basin)	0.022 [0.02,0.027]
	Intercept (basin)	0.029 [0.028,0.029]
	Error	0.084 [0.083,0.084]
Dissimilarity (Simpson index)	Time (site nested in basin)	0.088 [0.085,0.091]
	Time (basin)	0.043 [0.04,0.046]
	Intercept (site nested in basin)	0.067 [0.063,0.073]
	Intercept (basin)	0.083 [0.075,0.095]
	Error	0.191 [0.189,0.192]
Jaccard (binary, dissimilarity)	Time (site nested in basin)	0.066 [0.064,0.068]
	Time (basin)	0.028 [0.024,0.034]
	Intercept (site nested in basin)	0.059 [0.054,0.065]
	Intercept (basin)	0.111 [0.101,0.124]
	Error	0.16 [0.159,0.161]
Log Chao species richness	Time (site nested in basin)	0.097 [0.092,0.103]
	Time (basin)	0.072 [0.063,0.084]
	Intercept (site nested in basin)	0.509 [0.497,0.523]
	Intercept (basin)	0.525 [0.478,0.586]
	Error	0.366 [0.363,0.368]
Log total abundance	Time (site nested in basin)	0.223 [0.212,0.239]
	Time (basin)	0.169 [0.15,0.196]
	Intercept (site nested in basin)	0.892 [0.869,0.917]
	Intercept (basin)	0.783 [0.702,0.913]
	Error	0.8 [0.794,0.806]
Nestedness (jaccard)	Time (site nested in basin)	0.056 [0.054,0.057]
	Time (basin)	0.024 [0.021,0.029]
	Intercept (site nested in basin)	0.022 [0.018,0.029]
	Intercept (basin)	0.066 [0.058,0.077]
	Error	0.156 [0.155,0.157]
Proportion non-native abundance	Time (site nested in basin)	0.027 [0.025,0.028]
	Time (basin)	0.009 [0.008,0.012]
	Intercept (site nested in basin)	0.112 [0.109,0.117]
	Intercept (basin)	0.092 [0.085,0.102]
	Error	0.084 [0.084,0.085]
Proportion non-native species	Time (site nested in basin)	0.02 [0.019,0.022]
	Time (basin)	0.007 [0.005,0.009]
	Intercept (site nested in basin)	0.093 [0.091,0.096]
	Intercept (basin)	0.095 [0.088,0.104]
	Error	0.075 [0.075,0.076]
Turnover (jaccard)	Time (site nested in basin)	0.063 [0.061,0.064]
	Time (basin)	0.026 [0.022,0.033]
	Intercept (site nested in basin)	0.029 [0.024,0.037]
	Intercept (basin)	0.043 [0.037,0.053]
	Error	0.171 [0.17,0.172]