

Problème pratique de statistique n° pps003  
Macrofaune et annexes d'un fleuve

D. Chessel

En souvenir des travaux de Monique Coulet, un difficile problème de liaison entre cortèges faunistiques et variables explicatives, soit expérimentales soit environnementales.

## 1 Source

```
pps003 <- read.table("https://pbil.univ-lyon1.fr/R/donnees/pps003.txt",
  header = TRUE)
dim(pps003)
[1] 163 97
names(pps003)
 [1] "lone"   "site"   "alim"   "link"   "season" "sampler" "expo"   "current"
 [9] "depth" "veg"    "Hy"     "Ne"     "Pl"     "STl"    "O1"     "Ac"
[17] "MBit"  "MVap"  "MVac"  "MGac"   "MRaa"   "MRap"   "MLit"   "MPha"
[25] "MPhf"  "MPlc"  "MBac"  "MAns"   "MAnl"   "MAn"    "MGya"   "Mhic"
[33] "MArc"  "MAcl"  "MFew"  "MSpc"   "MPio"   "MPic"   "Mpip"   "Mpin"
[41] "MPis"  "Mpim"  "MPpe"  "MPi"    "Mcam"   "Mcat"   "MEuf"   "Mvev"
[49] "Hyd"   "Gam"   "As"    "PIe"    "ECar"   "ECah"   "ECam"   "EClD"
[57] "Elep"  "Od"    "Sial"  "TPh"    "TLi"   "TLe"   "TPo"   "Tec"
[65] "THy"   "TPs"   "TSe"   "CHa"    "CDy"   "CGy"   "CHy"   "CHo"
[73] "CHe"   "CHp"   "CDr"   "CEL"    "CLi"   "C"     "PLa"   "Het"
[81] "DCh"   "DCe"   "DDi"   "DTa"    "DTi"   "DCa"   "DCu"   "DEm"
[89] "DSt"   "DPs"   "DSy"   "Dpt"    "DLi"   "DDo"   "DLm"   "DSc"
[97] "DRa"
```

## 2 Introduction

Les données numériques à étudier ont été proposées, il y a quelques années, par M<sup>me</sup> MONIQUE COULET (alors chercheur en hydrobiologie) à un groupe de travail intitulé « Analyse des données planifiées » dans le cadre d'un Groupement de Recherches Coordonnées « Analyse des données et Informatique ». Elles sont représentatives « des plans d'observations » largement en oeuvre en écologie factorielle et plus particulièrement en hydrobiologie où les contraintes de travail de terrain sont importantes.

On ignorait largement avant cette expérience la valeur typologique de la macrofaune benthique (ensemble des organismes non planctoniques vivant dans l'eau) relativement aux différents bras morts d'un grand fleuve, c'est-à-dire l'importance des variations de la liste faunistique entre les annexes du chenal principal. On ignorait en particulier à quelle échelle spatiale s'exprime cette variabilité, à quels moments du cycle saisonnier cette variabilité est-elle la plus sensible, dans quelle mesure cette variabilité dépend des caractéristiques morphologiques

des bras morts (appelés lônes dans le bassin rhodanien). La capture et la quantification des peuplements étudiés pose en outre des problèmes méthodologiques qui ont été abordés en même temps que le problème précédent. La méthode d'échantillonnage utilisée est celle des substrats artificiels : un objet présentant de nombreuses cavités est immergé pendant une durée déterminée dans l'eau puis récupéré avec la plus grande partie des animaux y « ayant élu domicile », On obtient ainsi l'image d'une partie des populations présentes dans la station étudiée.

Une dernière difficulté propre aux organismes utilisés ici comme indicateurs biologiques tient au problème de leur détermination. Un premier tri des animaux capturés dans un piège les répartit par grand groupe (niveau de l'ordre dans la classification linnéenne) : la détermination plus précise (aux niveaux famille, genre et espèce) des animaux dans chacun des groupes est l'affaire d'un spécialiste disponible au non suivant les laboratoires. L'auteur des données peut identifier les mollusques à l'espèce. Ses collègues l'ont aidée à déterminer les trichoptères, coléoptères et diptères au niveau de la famille, ce pour des raisons pratiques. Plusieurs groupes sont restés à l'état de généralités. Le tableau faunistique est donc structuré sur les deux marges de manière très dissemblables. Il compte 163 lignes correspondant à 163 prélèvements (substrats artificiels) et 87 colonnes correspondant à 87 unités taxonomiques.

### 3 Code des taxa

On peut distinguer les groupes :

- Mollusques (détermination à l'espèce, 7-38)
- Epheméroptères ( 4 espèces et une famille, 43-47)
- Trichoptères (détermination à la famille, 50-57)
- Coléoptères ( - id- , 58-68)
- Diptères (-id-, 71 à 87)
- Autres groupes (généralités, 1-6, 39-42, 48-49, 69-70)

num	nom	label	type	num	nom	label	type
1	Hydres	Hy	gene	48	Odonates	Od	gene
2	Nématodes	Ne	gene	49	Megaloptères	Sial	gene
3	Planaires	Pl	gene	50	Phryganeidae	TPh	tricho
4	<i>Stylaria lacustris</i>	STl	gene	51	Limnophilidae	TLi	tricho
5	Oligochètes autres	Ol	gene	52	Leptoceridae	TLe	tricho
6	Achètes	Ac	gene	53	Polycentropodidae	TPo	tricho
7	<i>Bithynia tentaculata</i>	MBit	mollu	54	Ecnomidae	Tec	tricho
8	<i>Valvata pulchella</i>	MVap	mollu	55	Hydroptilidae	THy	tricho
9	<i>Valvata cristata</i>	MVac	mollu	56	Psychomyidae	TPs	tricho
10	<i>Galba corvus</i>	MGac	mollu	57	Sericostomatidae	TSe	tricho
11	<i>Radix auricularia</i>	MRaa	mollu	58	Haliplidae	CHA	coleo
12	<i>Radix peregra</i>	MRap	mollu	59	Dytiscidae	CDy	coleo
13	<i>Lymnaea truncatula</i>	MLit	mollu	60	Gyrinidae	CGy	coleo
14	<i>Physa acuta</i>	MPha	mollu	61	Hydraeinae	CHy	coleo
15	<i>Physa fontinalis</i>	MPhf	mollu	62	Hydrophilidae	CHO	coleo
16	<i>Planorbis carinatus</i>	MPlc	mollu	63	Helodidae	CHe	coleo
17	<i>Bathymophalus cantortus</i>	MBac	mollu	64	Helophoridae	CHp	coleo
18	<i>Anisus spirorbis</i>	MAns	mollu	65	Dryopidae	CDr	coleo
19	<i>Anisus leucostomus</i>	MAnl	mollu	66	Elmidae	CEl	coleo
20	<i>Anisus</i> sp.	MAn	mollu	67	Limnebiidae	CLi	coleo
21	<i>Gyraulus albus</i>	MGya	mollu	68	Chrysomelidae	C	coleo
22	<i>Hippeutis complanatus</i>	Mhic	mollu	69	Plannipennes	PLa	gene
23	<i>Armiger crista</i>	MArc	mollu	70	Hétéroptères	Het	gene
24	<i>Acroloxus lacustris</i>	MAcl	mollu	71	Chironomidae	DCh	dipt
25	<i>Ferrissia wautieri</i>	MFew	mollu	72	Ceratopogonidae	DCE	dipt
26	<i>Sphaerium corneum</i>	MSpc	mollu	73	Dixidae	DDi	dipt
27	<i>Pisidium obtusale</i>	MPio	mollu	74	Tabanidae	DTa	dipt
28	<i>Pisidium casertanum</i>	MPic	mollu	75	Tipulidae	DTi	dipt
29	<i>Pisidium pseudosphaerium</i>	Mpip	mollu	76	Chaoboridae	DCA	dipt
30	<i>Pisidium nitidum</i>	Mpin	mollu	77	Culicidae	DCu	dipt
31	<i>Pisidium subtruncatum</i>	MPis	mollu	78	Empididae	DEm	dipt
32	<i>Pisidium milium</i>	Mpim	mollu	79	Stratiomyidae	DSt	dipt
33	<i>Pisidium personatum</i>	MPpe	mollu	80	Psychodidae	DPs	dipt
34	<i>Pisidium</i> sp.	MPi	mollu	81	Syrphidae	DSy	dipt
35	<i>Carychium minimum</i>	MCam	mollu	82	Ptychopteridae	DPt	dipt
36	<i>Carychium tridentatum</i>	Mcat	mollu	83	Limnobiidae	DLi	dipt
37	<i>Euconulus fulvus</i>	MEuf	mollu	84	Dolichopodidae	DDo	dipt
38	<i>Vertigo antivertigo</i>	MVev	mollu	85	Limoniidae	DLM	dipt
39	Hydracariens	Hyd	gene	86	Sciomyzidae	DSc	dipt
40	Gammaridae	Gam	gene	87	Ragionidae	DRa	dipt
41	Asellus	As	gene				
42	Plecoptères	Plc	gene				
43	<i>Caenis rabusta</i>	ECar	ephe				
44	<i>Caenis horaria</i>	ECah	ephe				
45	<i>Caenis moesta</i>	ECam	ephe				
46	<i>Cloeon dipterum</i>	ECld	ephe				
47	Leptophlebiidae	Elep	ephe				

La valeur descriptive devrait normalement croître avec la précision de la détermination et on peut se proposer globalement de mesurer la valeur typologique de chaque groupe (éventuellement la variation de cette valeur au cours de l'année).

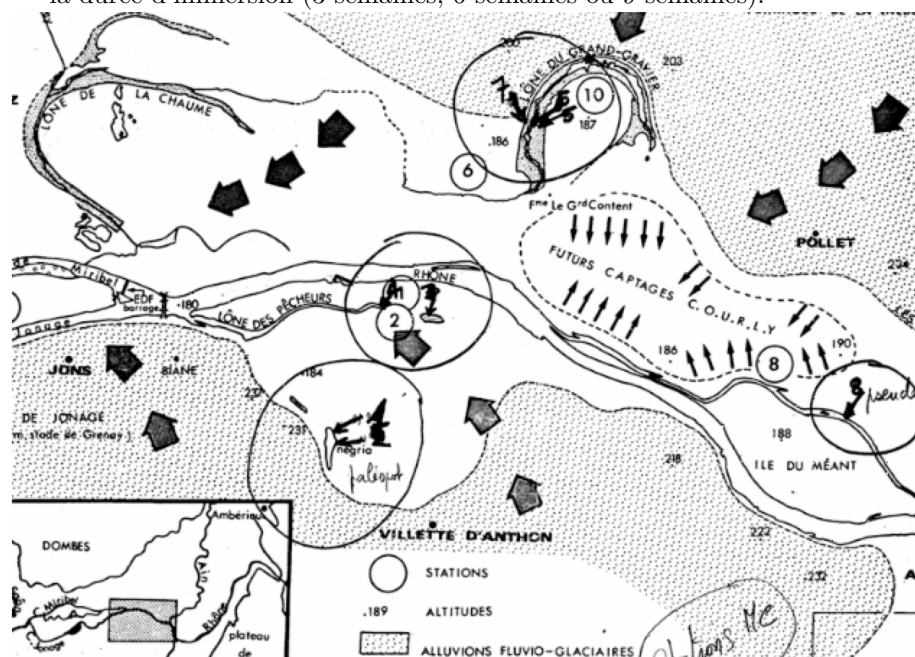
## 4 Code des relevés

Chacun des 163 relevés présentent des caractéristiques de deux ordres. Ils sont d'une part situés dans une station donnée à une saison donnée. Il y a 8 stations, numérotées de 1 à 8, réparties dans 4 îlots. Leurs caractéristiques essentielles sont :

n° station	Lône	Alimentation Liaison avec			
		en eau	le fleuve	Courant	Végétation
1	Négria	sans	sans	sans	immergée
2	Négria	sans	sans	sans	sans
3	Pêcheurs	avec	épisodique	sans	émergée
4	Pêcheurs	avec	par l'aval	sans	émergée
5	Grand Gravier	avec	sans	sans	sans
6	Grand Gravier	avec	sans	sans	émergée
7	Grand Gravier	avec	sans	faible	émergée
8	Méant	avec	amont/aval	moyen	immergée

Ces caractéristiques ne sont pas strictement constantes à cause des modifications de la hauteur d'eau dépendant pour partie du débit du fleuve (assèchement partiel). C'est pourquoi elles ont été enregistrées directement au niveau du prélèvement sous forme de variables qualitatives. Elles correspondent d'autre part à une pratique expérimentale à savoir :

- le type de substrat (bois ou ficelle, cf. articles cités);
- la durée d'immersion (3 semaines, 6 semaines ou 9 semaines).



La perte de quelques substrats (cruie et vandalisme) rend le plan d'observation non orthogonal.

Le code relevés est donc enregistré par un ensemble de 10 variables qualitatives dont les unes (station ou lône, saison, méthode, durée) sont fondamentales, les autres décrivant les caractères plus ou moins permanent de la station d'appartenance du relevé. Le code de ces variables est dans le suivant :

N° d'ordre	Identificateur	Nombre de mod.	N°	Identificateur	N° d'ordre
1	Lône	4	1	Négria	1
			2	Pêcheurs	2
			3	Grand Gravier	3
			4	Méant	4
2	Station	8	1	Négria 1	5
			2	Négria 2	6
			3	Pêcheurs 1	7
			4	Pêcheurs 2	8
			5	Grand Gravier 1	9
			6	Grand Gravier 2	10
			7	Grand Gravier 3	11
			8	Méant	12
3	Alimentation en eau souterraine	2	1	oui	13
			2	non	14
4	Liaison avec le fleuve	4	1	Amont-aval	15
			2	Avall	16
			3	Épisodique	17
			4	Sans	18
5	Saison d'échantillonnage	4	1	Hiver	19
			2	Printemps	20
			3	Été	21
			4	Automne	22
6	Méthode d'échantillonnage	2	1	Bois	23
			2	Ficelle	24
7	Durée d'immersion	3	1	3 semaines	25
			2	6 semaines	26
			3	9 semaines	27
8	Courant	3	1	moyen	28
			2	faible	29
			3	nul	30
9	Profondeur	5	1	> 1 mètre	31
			2	50 cm - 1 m	32
			3	20 cm - 50 cm	33
			4	< 20 cm	34
			5	à sec	35
10	Végétation	3	1	Émergée	36
			2	Immergée	37
			3	Sans	38

Table 1. List and description of the 10 qualitative variables and their different modalities.

Variable	Modality	Number of samples	Frequency of the modality	
n° 1 Old beds	N Négria	48	0.294	
	P Pêcheurs	48	0.294	
	G Grand Gravier	59	0.362	
	M Méant	8	0.049	
n° 2 Sampling sites	N1 Négria	24	0.147	
	N2 Négria	24	0.147	
	P1 Pêcheurs	24	0.147	
	P2 Pêcheurs	24	0.147	
	G1 Grand Gravier	21	0.129	
	G2 Grand Gravier	16	0.098	
	G3 Grand Gravier	22	0.135	
n° 3 Water supplying type	1 groundwater	115	0.706	
	2 rain	48	0.294	
	n° 4 Linking type with the river	1 up and downstream	8	0.049
		2 downstream	24	0.147
3 episodic		24	0.147	
n° 5 Season	4 without	107	0.656	
	1 winter	34	0.209	
	2 spring	44	0.270	
	3 summer	44	0.270	
n° 6 Sampler type	4 autumn	41	0.252	
	1 wood sampler	81	0.497	
	2 rope sampler	82	0.503	
	n° 7 Exposure type	1 3 weeks	47	0.288
2 6 weeks		62	0.380	
3 9 weeks		54	0.331	
n° 8 Water current	1 moderate	8	0.049	
	2 weak	22	0.135	
	3 without	133	0.816	
n° 9 Depth	1 1 m	10	0.061	
	2 0.5 - 1 m	32	0.196	
	3 0.20 - 0.50	42	0.258	
	4 0.20	68	0.417	
	5 drought	11	0.067	
n°10 Vegetation	1 helophytes	86	0.528	
	2 hydrophytes	32	0.196	
	3 without	45	0.276	

On a utilisé pour noms des variables le code :

```
names(pps003)[1:10]
[1] "lone" "site" "alim" "link" "season" "sampler" "expo" "current"
[9] "depth" "veg"
```

## 5 Questions ouvertes

Le tableau de données rassemble les 10 variables de contrôle et les 87 taxons. Il a 163 lignes et 97 colonnes. L'abondance de chaque taxon dans chaque piège est un entier (comptage à l'unité). On peut considérer que ce sont des mesures d'abondance (la transformation est en général utilisée réduire l'effet de l'agrégativité des insectes).

Elles concernent l'efficacité de chaque type de substrat (sélectivité relative, diversité ou richesse des captures, préférendum éventuel de certaines espèces), l'efficacité de chaque durée d'immersion (richesse et diversité maximale, éventuel effet d'accumulation ou « effet substrat », le piège devenant un système autonome), l'efficacité de l'échantillonnage à diverses saisons en terme de possibilités de séparer les stations sur leur contenu faunistique. Globalement on cherche à savoir à quelle période, avec quel piège et quelle durée d'immersion, pour quel groupe taxonomique on obtient une valeur typologique optimale.

Ce jeu de données a été en partie analysé dans Richardot-Coulet *et al.* (1986) [5]. Il est utilisé comme illustration méthodologique dans Lebreton *et*

*al.* (1988) [3], Sabatier (1987) [7] et Sabatier *et al.* (1989) [8]. Il fait partie des travaux de l'équipe (Richardot-Coulet *et al.* (1983) [6], Castella *et al.* (1984) [1], Richardot-Coulet *et al.* (1987) [4], Castella *et al.* (1991) [2]). Une version réduite est également disponible par `data(mollusc)` dans le paquet `ade4`.

## Références

- [1] E. Castella, M. Richardot-Coulet, C. Roux, and P. Richoux. Macroinvertebrates as "describers" of morphological and hydrological types of aquatic ecosystems abandoned by the rhône river. *Hydrobiologia*, 119 (3) :219–225, 1984.
- [2] E. Castella, M. Richardot-Coulet, C. Roux, and P. Richoux. Aquatic macroinvertebrate assemblages of two contrasting floodplain : the rhône and ain rivers, france. *Regulated rivers : Research and Management*, 6 :289–300, 1991.
- [3] J.D. Lebreton, D. Chessel, R. Prodon, and N. G. Yoccoz. L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. i. variables de milieu quantitatives. *Acta Oecologica, Oecologia Generalis*, 9 :53–67, 1988.
- [4] M. Richardot-Coulet, E. Castella, and C. Castella. Classification and succession of former channels of the french upper rhône alluvial plain using mollusca. *Regulated rivers : Research and Management*, 1 :111–127, 1987.
- [5] M. Richardot-Coulet, D. Chessel, and M. Bournaud. Typological value of the benthos of old beds of a large river. methodological approach. *Archiv für Hydrobiologie*, 107 :363–383, 1986.
- [6] M. Richardot-Coulet, P. Richoux, and C. Roux. Structure et fonctionnement des écosystèmes du haut-rhône français. 29 - structure des peuplements de macroinvertébrés benthiques d'un ancien méandre. *Archiv für Hydrobiologie*, 96 :363–383, 1983.
- [7] R. Sabatier. *Méthodes factorielles en analyse des données : approximations et prise en compte de variables concomitantes*. PhD thesis, 1987.
- [8] R. Sabatier, J.D. Lebreton, and D. Chessel. Principal component analysis with instrumental variables as a tool for modelling composition data. In R. Coppi and S. Bolasco, editors, *Multway data analysis*, pages 341–352. Elsevier Science Publishers B.V., North-Holland, 1989.