

# Contamination par le mercure des communautés de poissons du Haut-Maroni (Guyane Française) \*

## 1 Les données <sup>1</sup>

Des études réalisées sur l'ensemble du territoire de la Guyane Française ont mis en évidence des imprégnations par le mercure supérieures à la norme OMS dans les cheveux des populations amérindiennes vivant dans la zone du Haut - Maroni. Cette contamination a été attribuée à la forte consommation par ces populations de poissons des rivières, ces derniers étant considérés comme les vecteurs privilégiés du métal. En effet, la contamination mercurielle peut être attribuée pour une part importante aux rejets métalliques liés aux activités d'orpaillage car le mercure est largement utilisé pour ses propriétés d'amalgame avec les micro particules d'or. Les données sont constituées de 80 poissons appartenant à 12 espèces et 4 régimes alimentaires:

- Carnivores :
  - *Ageneiosus brevifilis* (7)
  - *Cynodon gibbus* (7)
  - *Hoplias aimara* (10)
  - *Potamotrygon hystrix* (4)
- Omnivores :
  - *Leporinus fasciatus* (3)
  - *Leporinus frederici* (3)
- Détritivores :
  - *Doras micropoeus* (8)
  - *Platydoras costatus* (10)
  - *Pseudoancistrus barbatus* (7)
  - *Semaprochilodus varii* (8)
- Herbivores :
  - *Acnodon oliganthus* (8)
  - *Myleus rubripinnis* (5)

	ESPECE	REGIME	POID	MUSC	Foie	...	Estomac
1	<i>Ageneiosus brevifili</i>	Carnivore	525	2195	697	...	NA
	<i>Cynodon gibbus</i>	Carnivore	359	8737	333	...	439
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
	<i>Leporinus frederici</i>	Omnivore	350	255	1761,5	...	10880
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
	<i>Doras micropoeus</i>	Détritivores	351	965	1430	...	887
	<i>Doras micropoeus</i>	Détritivores	505	1947	1075	...	230
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
	<i>Acnodon oligacanthus</i>	Herbivores	100	44	190	...	NA
80	<i>Myleus rubripinis</i>	Herbivores	275	8	249	...	3256

Figure 1: Extrait du tableau de données

\*Article de la revue Modulad <http://www.modulad.fr/archives/numero-26/Chavent/contaminationpar.pdf>

<sup>1</sup>Les données ont été fournies par le laboratoire d'Ecophysiologie et d'Ecotoxicologie des Systèmes Aquatiques (UMR EPOC)

## 2 Etude avec les variables d'origine

Charger les données dans R :

```
load("../data/poissons.rda")
```

1. Faire des statistiques descriptives et regarder en particulier la distribution des variables (utiliser par exemple `ggpairs` du package `GGally`).
2. Réaliser une ACP avec la fonction `PCA` du package `FactoMineR` en mettant :
  - les 6 variables de concentration (MUSC, INTE, ESTO, BRAN, FOIR, REIN) en actives,
  - les variables longueur (LONG) et poids (POID) en illustratives,
  - l'espèce (ESPECE) et le régime alimentaire (REGIME) en illustratif.
3. Interpréter les résultats et plus particulièrement les contributions des individus.
4. Quel est l'effet de l'**asymétrie** des distributions ?

Exemples de graphiques que vous pouvez obtenir :

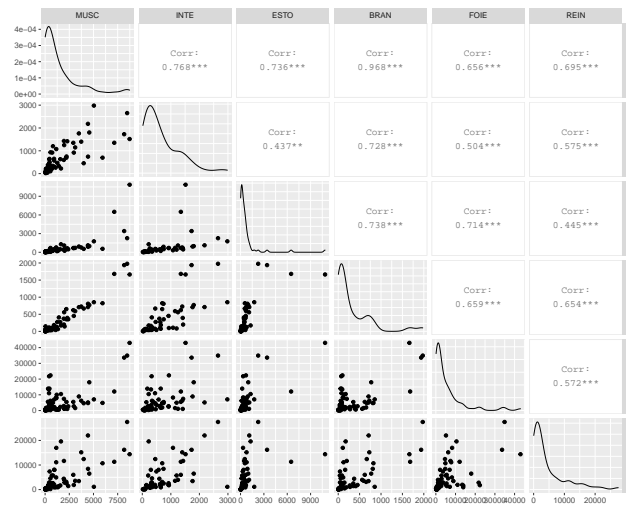


Figure 2: Statistiques descriptives des variables de concentration

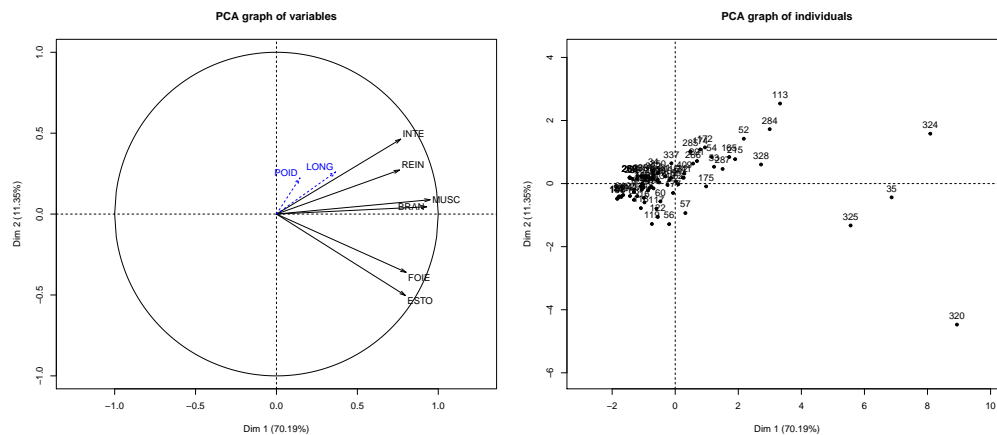


Figure 3: ACP normée des données d'origine

### 3 Etude avec les log des variables de concentration

L'idée est de transformer les variables de concentration par passage au log afin d'obtenir des distributions moins asymétriques.

- Créer un nouveau dataframe avec
  - les 4 variables ESPECE, REGIME, LONG, POID d'origine
  - le log des 6 variables MUSC, INTE, ESTO, BRAN, FOIE, REIN.
- Faire des statistiques descriptives et vérifier que les distributions sont moins asymétriques.
- Faire une ACP avec les log des 6 variables de concentration en actives et les autres en illustratives.
- On voit apparaître un **effet taille** sur le cercle des corrélations (variables toutes corrélées). Quelles sont les conséquences de cet effet taille sur l'interprétation du premier plan factoriel des individus.
- Ajouter une couleur aux poissons en fonction de leur régime alimentaire sur le premier plan factoriel des individus (en utilisant l'argument `habillage` de la méthode `plot.PCA`). Interpréter ce graphique. Pourquoi faudrait-il supprimer cet effet taille ?

Exemples de graphiques que vous pouvez obtenir :

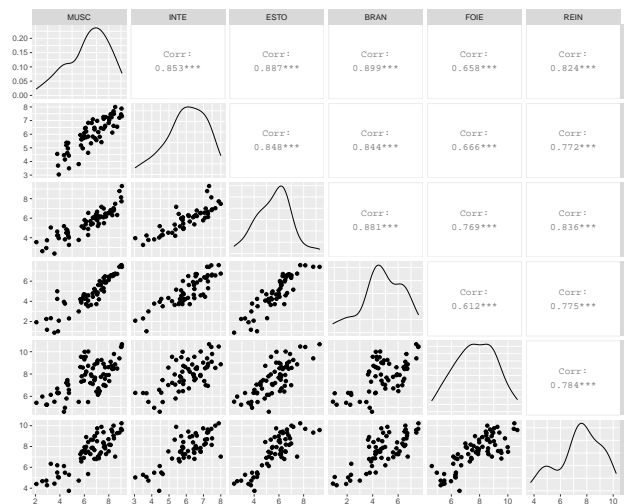


Figure 4: Statistiques descriptives des log des variables de concentration

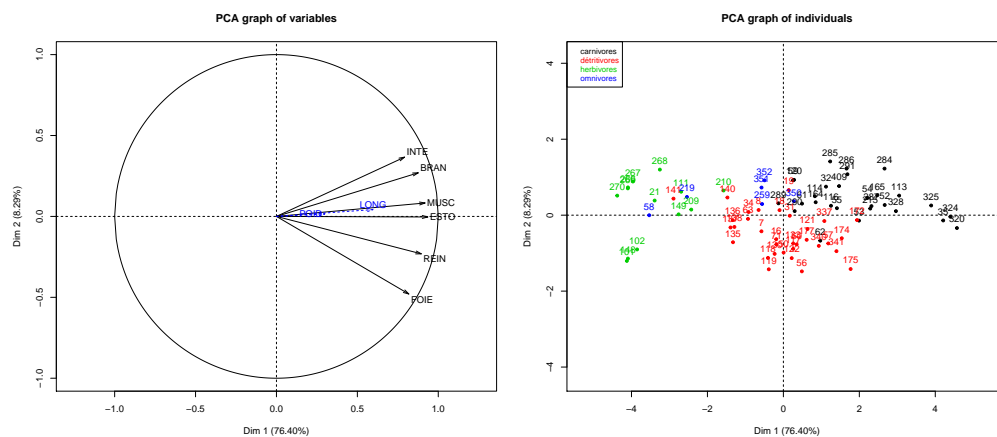


Figure 5: ACP normée des données après transformation log

## 4 Etude avec les variables “log(organe/muscle)”

Il faut identifier le facteur caché pour supprimer cet effet taille. Après avis d’expert, la variable concentration dans le muscle (MUSC) a été identifiée comme “facteur caché”. Afin de réduire l’effet taille, les concentrations dans les 5 organes (INTE, ESTO, BRAN, FOIE, REIN) sont divisées par la concentration dans le muscle (MUSC).

1. Créer un nouveau dataframe avec :

- les 4 variables ESPECE, REGIME, LONG, POID d’origine,
- les 5 variables  $\log(\text{INTE}/\text{MUSC})$ ,  $\log(\text{ESTO}/\text{MUSC})$ ,  $\log(\text{BRAN}/\text{MUSC})$ ,  $\log(\text{FOIE}/\text{MUSC})$ ,  $\log(\text{REIN}/\text{MUSC})$ .

2. Faire une ACP avec les 5 variables “log(organe/muscle)” en actives et les autres en illustratives. Regarder en particulier les résultats pour les herbivores.

3. Refaire une ACP sans les poissons herbivores et interpréter les résultats.

Exemples de graphiques que vous pouvez obtenir :

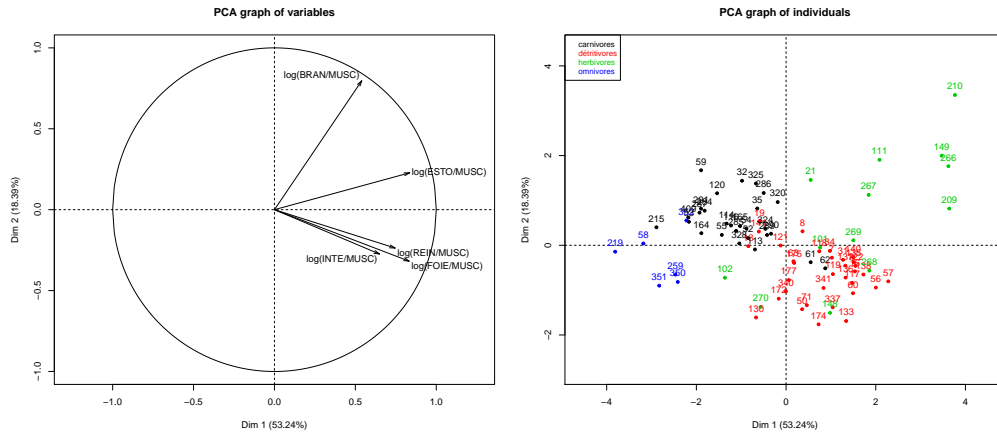


Figure 6: ACP normée des 80 poissons après transformation  $\log(\text{organe}/\text{muscle})$

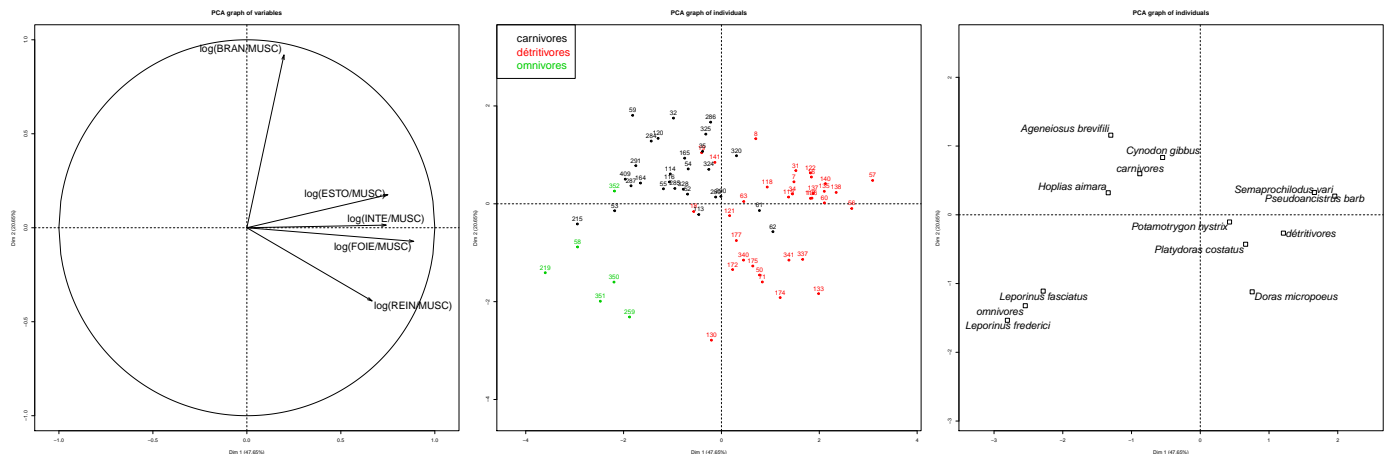


Figure 7: ACP normée des 67 poissons non herbivores après transformation  $\log(\text{organe}/\text{muscle})$

## 5 Définir une typologie des poissons et des espèces

On veut maintenant construire une partition du tableau des 67 poissons (non herbivores) décrits par les 5 variables continues (logarithme népérien du rapport entre la concentration dans l'organe et la concentration dans le muscle).

1. Le tableau de données possède des données manquantes. Appliquer d'abord la méthode basique (et généralement pas la meilleure) qui consiste à remplacer les données manquantes par la moyenne de la colonne.

##	log(INTE/MUSC)	log(ESTO/MUSC)	log(BRAN/MUSC)	log(FOIE/MUSC)	log(REIN/MUSC)
## 7	-0.1517265	-0.6309049	-1.560252	2.0044510	1.70253826
## 8	-0.1129125	-0.9799727	-1.266596	1.4388722	0.94550505
## 16	0.2564257	NA	-1.467893	2.9464436	1.57750467
## 18	-0.8638853	-1.2471816	-1.891242	0.4137228	1.18466065
## 19	-0.3511585	NA	-1.600711	0.6119301	-0.01637384
## 31	0.3747047	NA	-1.413361	1.8865751	1.69041832

2. Appliquer une méthode de clustering à ces données.

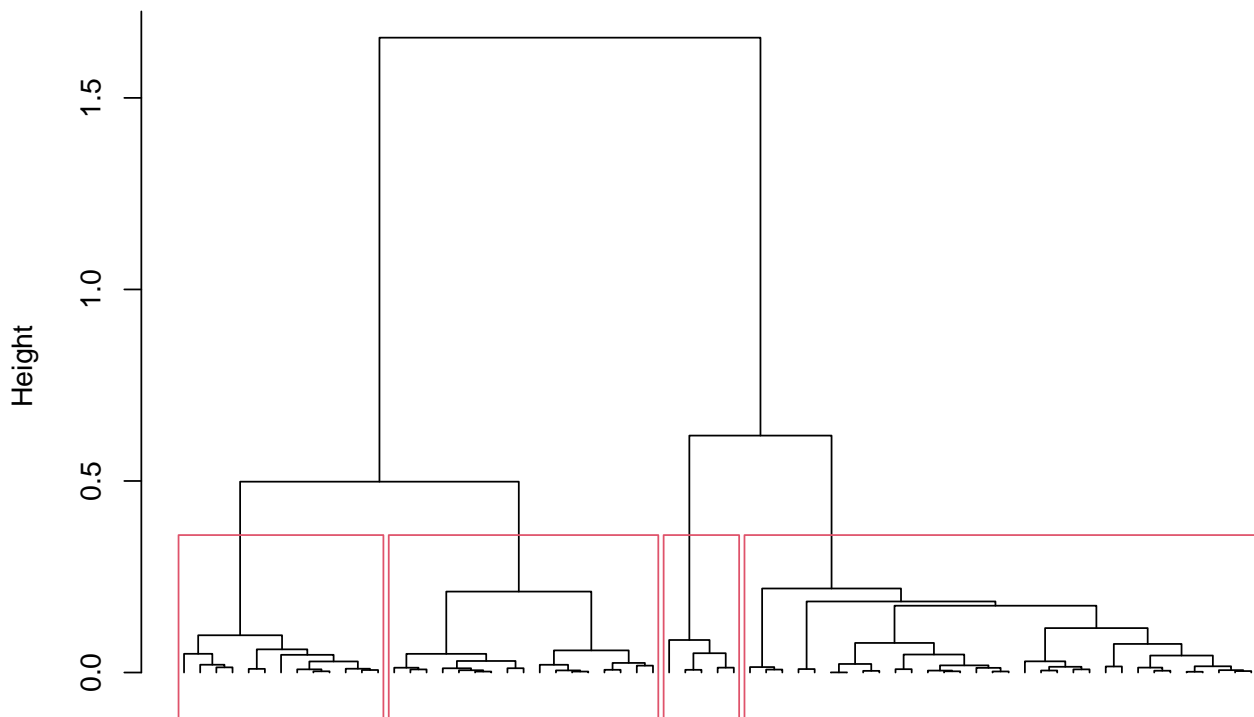


Figure 8: CAH de Ward sur données centrées-réduites

3. Interpréter les classes obtenue en fonction des 5 variables quantitatives utilisées et des variables qualitatives espèces et régime alimentaire.
5. Imputez cette fois les données manquantes avec le package missMDA. Les résultats du clustering sont-ils très différents ?
6. Comment pourriez-vous déduire de ces résultats une partition des espèces ? Une partition "floue" ?