

Épreuve MADG

J.R. Lobry

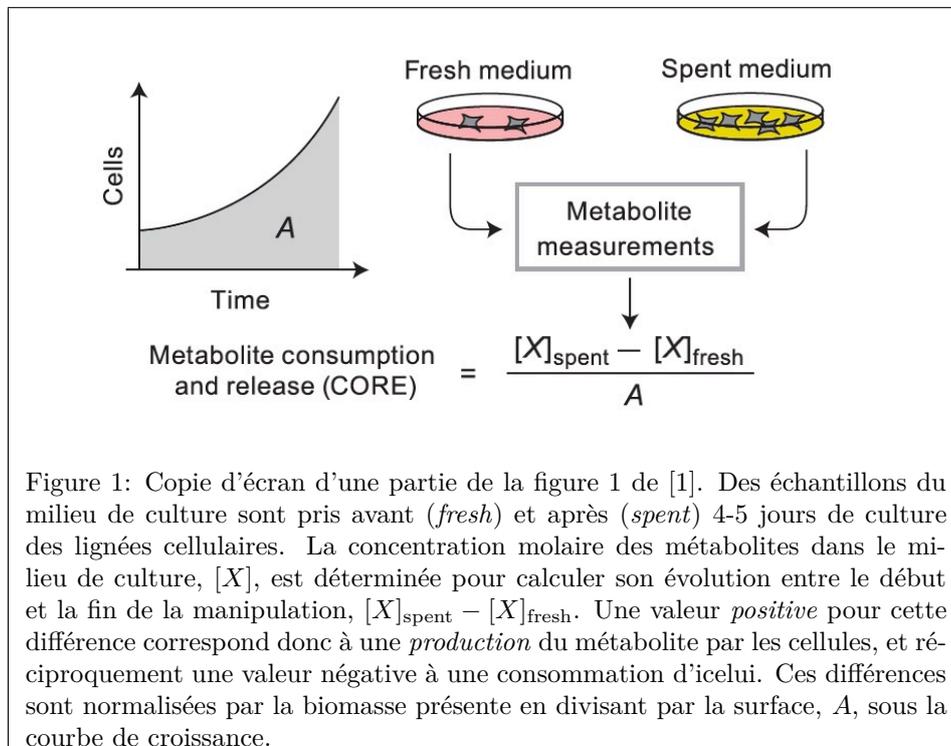
Première session - automne 2022

Tous documents autorisés - échanges strictement interdits

Numéro d'intercalaire :

1 Introduction

LES données [1] sont des flux pour 140 métabolites chez 60 lignées cellulaires cancéreuses. Le principe de la manipulation est donné dans la figure 1. Dans le fichier utilisé ici (Jean-Pierre MAZAT, communication personnelle) les expériences répliquées ont été fusionnées et le temps de doublement en phase de croissance exponentielle a été ajouté en première colonne.



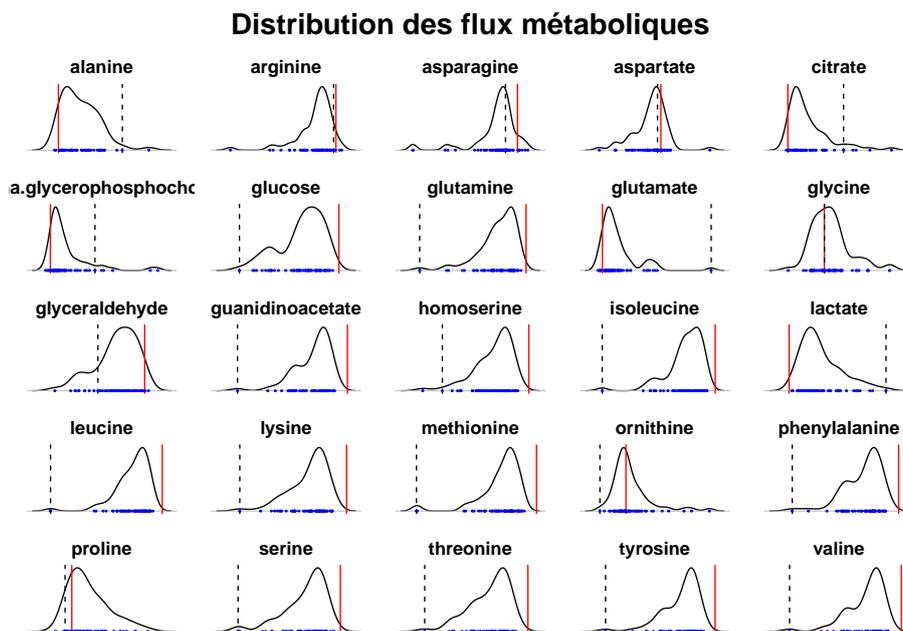
On retiendra ici les 25 métabolites présentant les flux les plus importants.

```
# Importation des données
load(url("http://pbil.univ-lyon1.fr/R/donnees/cni.Rda"))
dim(cni)
[1] 60 141

# Sélection des flux les plus importants
cni <- cni[ , which(log10(colMeans(abs(cni))) > 0)]
dim(cni)
[1] 60 26
```

La figure ci-après est une analyse univariée de la table d'intérêt. On représente pour chaque métabolite la distribution des flux observés chez les 60 lignées cellulaires. La ligne rouge verticale représente la valeur zéro, donc l'absence de flux. La ligne verticale en pointillé donne la position de la lignée SF-295.

```
par(mfrow = c(5, 5), mar = c(1, 1, 2, 1), xaxt = "n", yaxt = "n",
    bty = "n", oma = c(0, 0, 3, 0))
qui <- which(rownames(cni) == "SF-295")
for(i in 2:26){
  x <- cni[ , i]
  plot(density(x), main = colnames(cni)[i])
  abline(v = 0, col = "red")
  abline(v = cni[qui, i], lty = 2)
  points(x, rep(0, length(x)), pch = 19, col = "blue", cex = 0.25)
}
title(main = "Distribution des flux métaboliques", outer = TRUE, cex.main = 2)
```



Au vu de cette représentation, que pouvez-vous dire de la distribution des flux pour ces 25 métabolites ? Autrement dit, si vous deviez classer ces métabolites, combien de classes définiriez-vous et à quoi correspondraient-elles ?

Réponse:

2 Les flux dans l'absolu

DANS un premier temps, on ne s'intéresse qu'à l'intensité des flux, indépendamment du fait qu'ils soient entrants ou sortants. On utilise une ACP centrée-réduite pour résumer l'information.

```
cnia <- abs(cni)
library(ade4)
cnia.pca <- dudi.pca(cnia, scannf = FALSE)
summary(cnia.pca)
Class: pca dudi
Call: dudi.pca(df = cnia, scannf = FALSE)
Total inertia: 26

Eigenvalues:
  Ax1   Ax2   Ax3   Ax4   Ax5
15.385  2.597  1.970  1.505  1.043

Projected inertia (%):
  Ax1   Ax2   Ax3   Ax4   Ax5
59.175  9.988  7.576  5.788  4.012

Cumulative projected inertia (%):
  Ax1  Ax1:2  Ax1:3  Ax1:4  Ax1:5
 59.17  69.16  76.74  82.53  86.54

(Only 5 dimensions (out of 26) are shown)
```

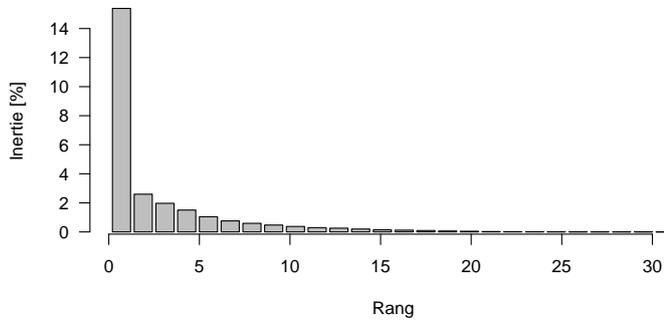
QUEL est le pourcentage de la variabilité initiale (inertie) qui est conservé si on ne conserve que deux facteurs ?

Réponse:

AU vu du graphe des valeurs propres, combien de facteurs seriez-vous tenté de chercher à interpréter ?

```
barplot(cnia.pca$eig, las = 1, ylab = "Inertie [%]",
        main = "Graphe des valeurs propres",
        xlab = "Rang")
axis(1)
```

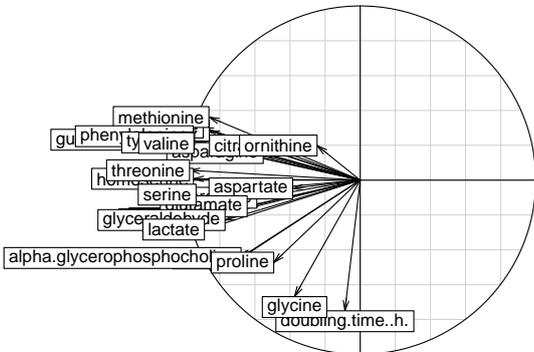
Graphe des valeurs propres



Réponse:

DANS le cercle des corrélations ci-dessous, quel phénomène est-il parfaitement illustré pour les variables métaboliques de ce jeu de données ? Donnez son nom technique et expliquez succinctement de quoi il s'agit.

```
s.corcircle(cnia.pca$co)
```



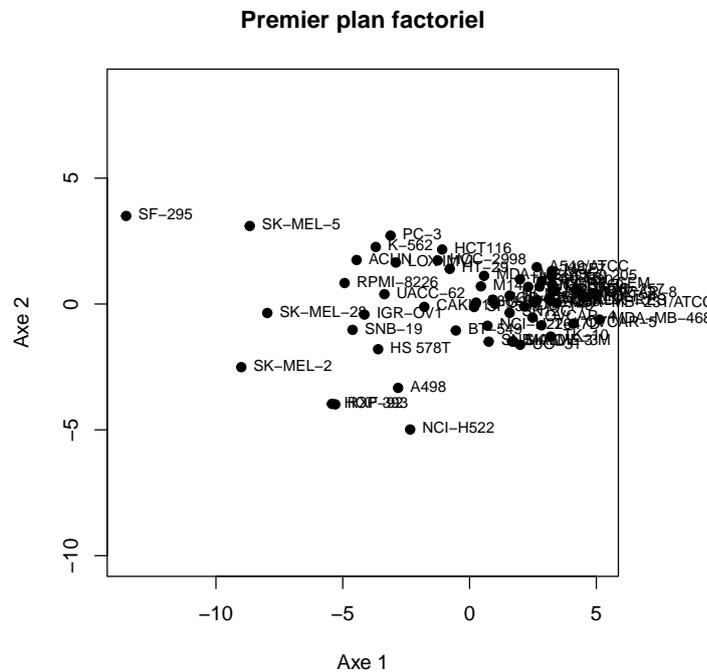
Réponse:

QU'EST-CE que cela signifie concrètement d'un point de vue biologique pour les flux de métabolites de ces lignées cellulaires ?

Réponse:

À la vue du premier plan factoriel pour les individus, que pouvez-vous dire de la lignée cellulaire SF-295 ?

```
par(pty = "s")
x <- cnia.pca$li[, 1] ; y <- cnia.pca$li[, 2]
plot(x, y, asp = 1, pch = 19,
      xlab = "Axe 1", ylab = "Axe 2", main = "Premier plan factoriel")
text(x, y, rownames(cnia), pos = 4, cex = 0.75, xpd = NA)
```



Réponse:

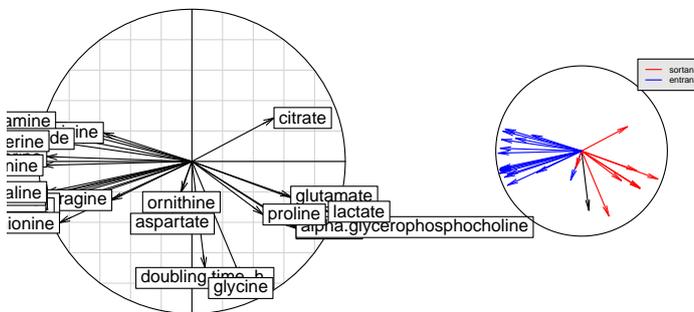
3 Les flux entrants et sortants

On cherche maintenant à ne pas perdre l'information sur le sens des flux (entrant ou sortant). Pour faciliter la lecture on coloriera en rouge les flux qui sont en moyenne positifs (sortants) et en bleu ceux qui sont en moyenne négatifs

(entrants). On laissera en noir la variable donnant le temps de doublement. On commence par une ACP classique centrée-normée.

```
cni.pca <- dudi.pca(cni, scannf = FALSE)
par(mfrow = c(1, 2), oma = c(0, 0, 2, 0), xpd = NA)
s.corcircle(cni.pca$co)
col <- ifelse(colMeans(cni) > 0, "red", "blue")
col[1] <- "black"
plot.new() ; plot.window(xlim = c(-1, 1), ylim = c(-1, 1), asp = 1)
arrows(0, 0, cni.pca$co[, 1], cni.pca$co[, 2], angle = 10, le = 0.1, col = col)
library(seqinr) ; circle(0, 0, 1)
legend("topright", inset = 0, legend = c("sortant", "entrant"), lty = 1,
      col = c("red", "blue"), cex = 0.5, bg = grey(0.9))
title(main = "Le premier plan factoriel dans l'espace des variables",
      outer = TRUE)
```

Le premier plan factoriel dans l'espace des variables



Sur ce même plan factoriel mais au niveau des individus, où va se situer la lignée SF-295 ? À droite ou bien à gauche ? Justifiez votre réponse.

Réponse:

Sur le plan factoriel des variables les flux sortants sont orientés à droite et les flux entrants à gauche. Est-ce que cela ne devrait pas être le contraire ?

Réponse:

Si on cherche à conserver dans l'absolu l'information sur l'orientation (entrant ou sortant) des flux, utiliser une ACP centrée-réduite n'est pas forcément très satisfaisant. Expliquez pourquoi.

Réponse:

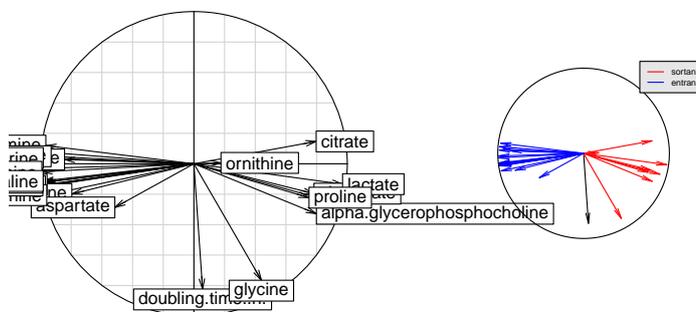
4 Une dernière analyse

ON effectue une ACP centrée sur la valeur 0 pour les flux métabolites et à la valeur moyenne pour le temps de doublement.

```
acp0 <- dudi.pca(cni, center = c(mean(cni[, 1]), rep(0, ncol(cni) - 1)),
                scannf = FALSE)
```

```
par(mfrow = c(1, 2), oma = c(0, 0, 2, 0), xpd = NA)
s.corcircle(acp0$co)
plot.new() ; plot.window(xlim = c(-1, 1), ylim = c(-1, 1), asp = 1)
arrows(0, 0, acp0$co[, 1], acp0$co[, 2], angle = 10, le = 0.1, col = col)
circle(0, 0, 1)
legend("topright", inset = 0, legend = c("sortant", "entrant"), lty = 1,
      col = c("red", "blue"), cex = 0.5, bg = grey(0.9))
title(main = "Le premier plan factoriel dans l'espace des variables",
      outer = TRUE)
```

Le premier plan factoriel dans l'espace des variables



QUELLE est votre interprétation du premier facteur de l'ACP ? Quel lien faites-vous avec la typologie que vous aviez proposée lors de l'analyse univariée ?

Réponse:

Le temps de doublement est un indicateur de l'agressivité des lignées cellulaires cancéreuses : plus il est faible plus elles sont invasives. Le titre de l'article dont sont issues les données était « [m]etabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. » Le second facteur de l'ACP est-il en accord avec cette assertion ?

Réponse:

References

- [1] M. Jain, R. Nilsson, S. Sharma, N. Madhusudhan, T. Kitami, A.L. Souza, R. Kafri, M.W. Kirschner, C.B. Clish, and V.K. Mootha. Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science*, 336(6084):1040–1044, 2012.