

Analyse de d'inventaire: Approches basées sur les graphes bipartites

11 juin 2018

1

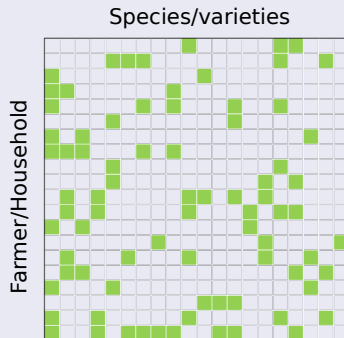
- Tableaux d'inventaire
- Représentation des données
- Questions/hypothèses

- Prémices
- Analyse des marginales

- Interpretation des communautés

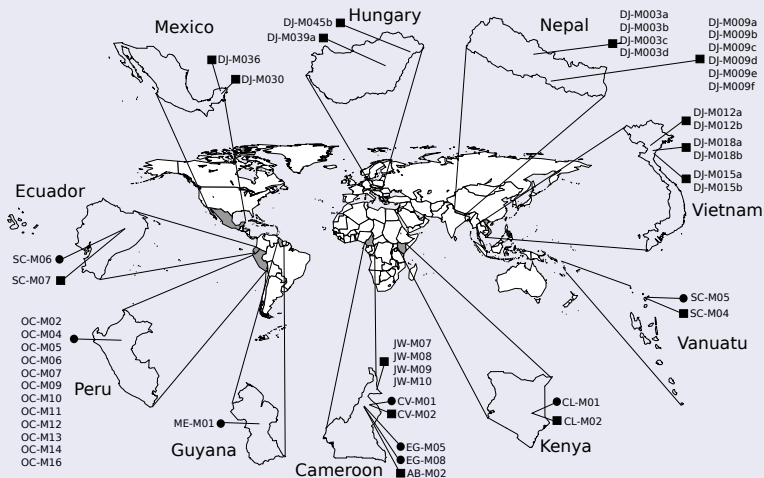
Présence-Absence

Données de Présence/absence (inventaire)



Méta-analyse

Données de Présence/absence (inventaire d'espèces et de variétés)



cercle = espèce, carré = variété

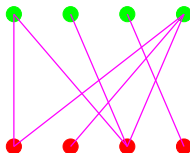
Focus : données de diversité variétales de riz au Vietnam



Protocole de collecte (Jarvis et al., 2008)

- Communauté : Dabac ; Province : Nam Dinh ; district : Nghia Hung
- Enquête préliminaire pour identifier :
 - 2 communes représentatives de l'agrosystème ciblé : Nghia Lac et Nghia Loi
 - liste des 86 variétés de riz cultivées dans ces 2 villages
- Enquête auprès de 117 chefs d'exploitation

Représentations sous la forme Graphe Bipartite



Noeud rouge = Variété; Noeud vert = Ferme;

Intérêt= Transposer les méthodes d'analyse de réseaux aux tableaux d'incidence

Données de comptage

Exemples :

- Données d'inventaire d'espèce : nombre de variétés cultivées
- Données d'inventaire d'espèce ou de variété : Surface cultivée...

Questions/hypothèses ?

Exemples de Questions :

- Les fermes exhibent-elles différents profil de cultures ?
- Certaines variétés sont-elles préférentiellement cultivées par certains agriculteurs ?

Triple objectif de l'analyse statistique :

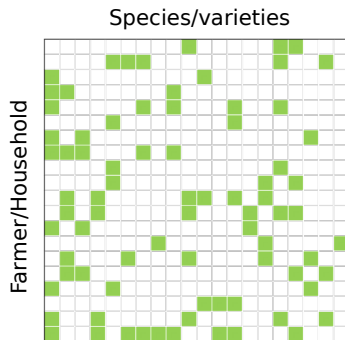
- Visualisation des données
- Répondre à des questions/hypothèses
- Analyse plus exploratoire (recherche de structure, formulation d'hypothese)

Allers-Retours nécessaire entre ces 2 approches.

- 1 Quelles données ?
 - Tableaux d'inventaire
 - Représentation des données
 - Questions/hypothèses
- 2 Explorations
 - Prémices
 - Analyse des marginales
- 3 Analyses en composantes principales
- 4 Latent Block Models et interprétation
 - Interpretation des communautés
- 5 Intégration de covariables
- 6 Pour aller plus loin

En Guise de préliminaire

- Y a t-il des données manquantes ?
Les 0 représentent-ils une absence de l'espèce ou une non-observation ?
Questionnaire sur liste **libre** ou sur liste **fermée** ?
- Quelle est l'**unité d'observation** ? (i.e. à quoi correspond une ligne ?) *une ferme ? un village ? un foyer ?*



Analyse en ligne

```
library(blockmodels)
library(ade4)
library(vegan)
# Some annex functions
source("annex_functions.R")

# Importing the data
DJM12= read.table("data/DJ-M012.csv",sep="\t",header=T,row.names = 1)
```

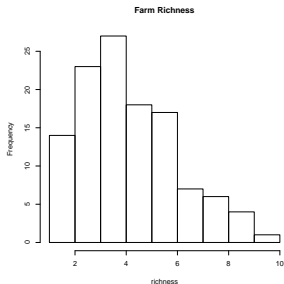
```
# Cleaning step
dim(DJM12)

## [1] 117 85

DJM12=DJM12[rowSums(DJM12)>0,colSums(DJM12)>0]
DJM12= DJM12[order(rowSums(DJM12)),order(colSums(DJM12),decreasing = T)]
```

Analyse en ligne

```
# Diversity in each Farm
richness= rowSums(DJM12)
hist(richness, main="Farm Richness")
```

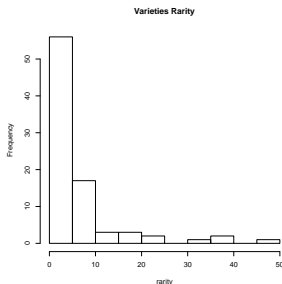


```
table(richness)
```

```
## richness
##  1  2  3  4  5  6  7  8  9 10
##  6  8 23 27 18 17  7  6  4  1
```

Analyse en colonne

```
# For each variety
rarity= colSums(DJM12)
hist(rarity, main="Varieties Rarity")
```

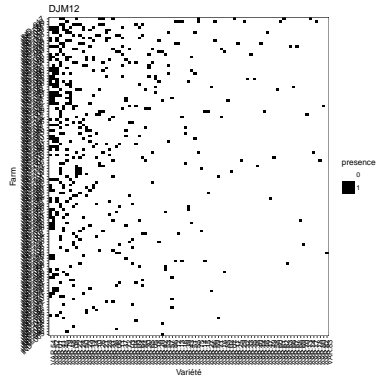


```
table(rarity)
```

```
## rarity
##  1  2  3  4  5  6  7  8  9 10 11 15 17 18 20 23 24 31 36 39 48
## 31  8  8  7  2  5  4  4  2  2  1  2  1  1  1  1  1  1  1  1  1
```

Représentation des données

```
plotpa(DJM12,mytitle="DJM12",var1="Farm",var2="Variété")
```



- 1 Quelles données ?
 - Tableaux d'inventaire
 - Représentation des données
 - Questions/hypothèses
- 2 Explorations
 - Prémices
 - Analyse des marginales
- 3 Analyses en composantes principales
- 4 Latent Block Models et interprétation
 - Interpretation des communautés
- 5 Intégration de covariables
- 6 Pour aller plus loin

ACP

Objectifs :

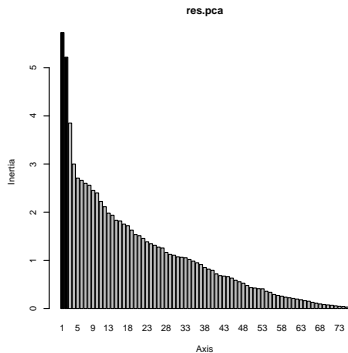
- visualiser des corrélations entre fermes, entre variétés
- Détection de données anormales

Risque : Le résultat de l'ACP sera fortement influencé par la présence de ces variétés très rares.

Remarque : Dans ce cas particulier, l'ACP est équivalent à un ACM.

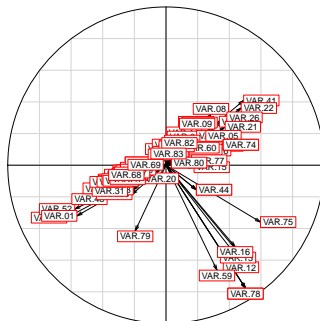
PCA

```
# PCA
res.pca = dudi.pca(DJM12, nf= 2,scannf = F)
screeplot(res.pca)
```



Représentation des individus

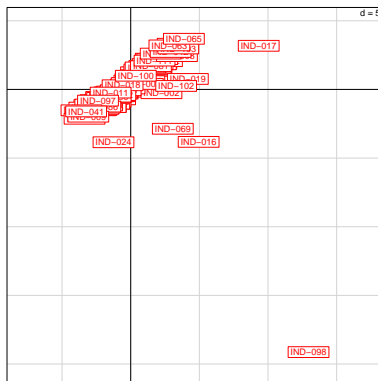
```
# Projection des variétés  
s.corcircle(res.pca$co)
```



Représentation des fermes

```
# Projection des Fermes
```

```
s.label(res.pca$li, xax = 1, yax = 2)
```



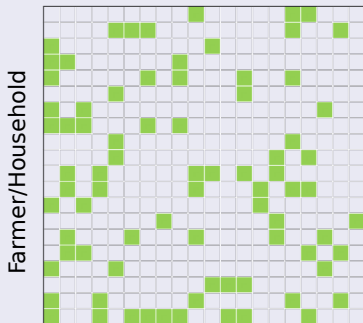
Attention à l'effet des plantes rares

- 1 Quelles données ?
 - Tableaux d'inventaire
 - Représentation des données
 - Questions/hypothèses
- 2 Explorations
 - Prémices
 - Analyse des marginales
- 3 Analyses en composantes principales
- 4 Latent Block Models et interprétation**
 - Interpretation des communautés
- 5 Intégration de covariables
- 6 Pour aller plus loin

Latent block models

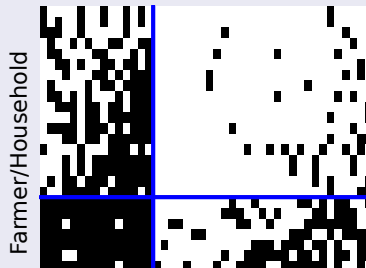
Presence/absence data (inventories)

Species/varieties

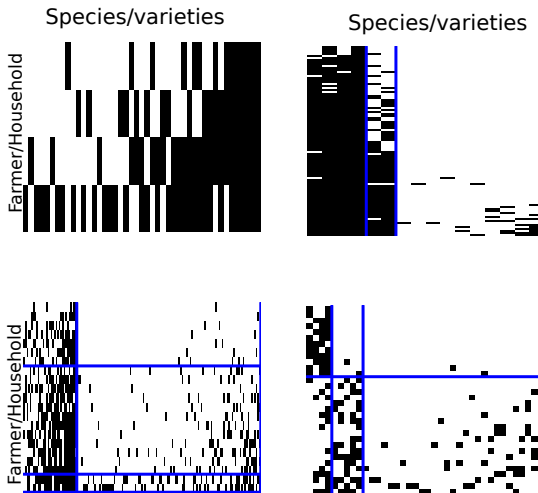


Latent Block Model

Species/varieties



Quelques résultats de la méta-analyse



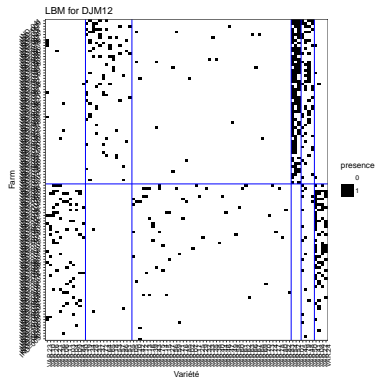
Analyse

```
# Création du LBM
bm <- BM_bernoulli("LBM",as.matrix(DJM12),verbosity=0,plotting="")
# Ajustement et choix du modèle
res <- bm$estimate()
```

```
num <- which.max(bm$ICL)
prob.row <- bm$memberships[[num]]$Z1
mbrshp.row <- apply(prob.row,1,which.max)
names(mbrshp.row)= rownames(DJM12)
prob.col <- bm$memberships[[num]]$Z2
mbrshp.col <- apply(prob.col,1,which.max)
names(mbrshp.col)= colnames(DJM12)
```

```
# Représentation des résultats
```

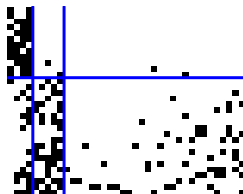
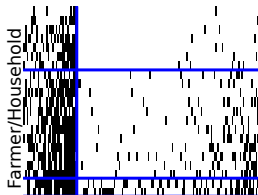
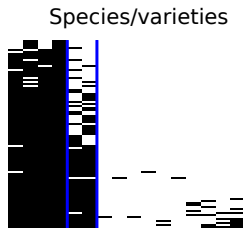
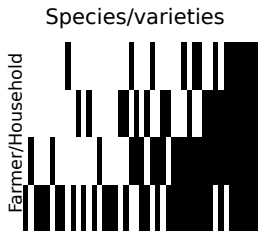
```
plot.classif(DJM12,var1="Farm",var2="Variété", group1=mbrshp.row,group2= mbrshp
```



Variantes pour des données de comptage

```
# Représentation des résultats  
bm_poisson <- BM_poisson("LBM", as.matrix(DJM12), verbosity=0, plotting="")
```

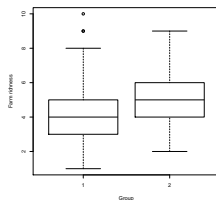
Interprétation des groupes



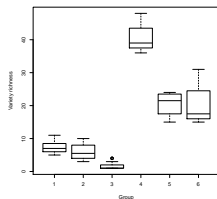
Interprétation des groupes

```
# Looking at the groups
```

```
boxplot(richness~mbrshp.row, xlab="Group", ylab="Farm richness")
```



```
boxplot(rarity~mbrshp.col, xlab="Group", ylab="Variety richness")
```



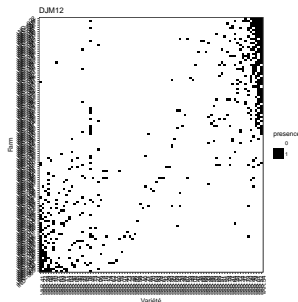
ACP sur les résidus

Question : les Groupes sont-ils seulement un artefact de l'hétérogénéité de richesse des fermes/variétés.

```
# Testing a configuration Model
test.c= func.conf.test(DJM12, nsim=5000)
test.c$test
```

```
##          sv mean.sv.sim      p.value
##          9.04         5.94         0.00
```

```
plotpa(DJM12[test.c$ordre.ligne,test.c$ordre.colonne],mytitle="DJM12",var1="Far"
```



- 1 Quelles données ?
 - Tableaux d'inventaire
 - Représentation des données
 - Questions/hypothèses
- 2 Explorations
 - Prémices
 - Analyse des marginales
- 3 Analyses en composantes principales
- 4 Latent Block Models et interprétation
 - Interpretation des communautés
- 5 Intégration de covariables
- 6 Pour aller plus loin

Utilisation d'information jointe

Intégration de covariables

- [Ligne] Pays (si l'unité statistique est un village)
- [Ligne] Indicateur de richesse du foyer (si l'unité statistique est un foyer)
- [Colonne] Type d'espèce cultivé (céréale, oléagineux)

Questions :

- Les variations de cultures entre agrosystèmes sont-elles corrélées au ?
- Les Différents types d'espèces ont-ils des profils de culture différentes ?
- Les foyers riches ont-il des profils de culture similaire ?

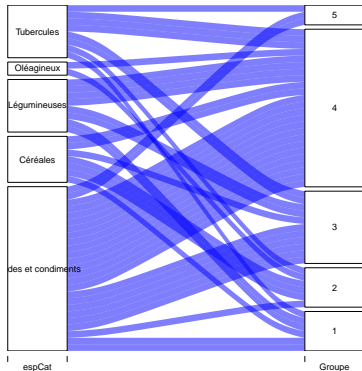
Exemple : Type d'espèce cultivé

```
library(alluvial)
cat= unique(plantadiv[,c("espCatListes", "espCat")])
cat[, "Groupe"] = mbrshp.col[cat$espCatListes]
table(cat[,2:3])
```

##		Groupe				
##	espCat	1	2	3	4	5
##	Brèdes et condiments	2	1	6	14	2
##	Céréales	1	3	1	2	0
##	Légumineuses	2	0	2	4	0
##	Oléagineux	0	1	0	1	0
##	Tubercules	1	1	2	3	1

Exemple : Type d'espèce cultivé

```
alluvial(cat[,2:3],freq=rep(1,dim(cat)[1]),col=c("blue"))
```



```
chisq.test(table(cat[,2:3]))$p.value
```

```
## [1] 0.4844796
```


- 1 Quelles données ?
 - Tableaux d'inventaire
 - Représentation des données
 - Questions/hypothèses
- 2 Explorations
 - Prémices
 - Analyse des marginales
- 3 Analyses en composantes principales
- 4 Latent Block Models et interprétation
 - Interpretation des communautés
- 5 Intégration de covariables
- 6 Pour aller plus loin

Extensions

- Analyse jointe d'inventaires et de réseaux d'échanges [[Barbillon et al.\('18\)](#)]
- Analyse conjointe de plusieurs (≥ 3) effets?