



Comparaison de méthodes d'analyses multivariées pour la description de données de germination de semences

Ophélie Thierry, Rachid Boumaza, Julia Buitink, Claudine Landès, Olivier Leprince, Pierre Santagostini, Julie Bourbeillon

► To cite this version:

Ophélie Thierry, Rachid Boumaza, Julia Buitink, Claudine Landès, Olivier Leprince, et al.. Comparaison de méthodes d'analyses multivariées pour la description de données de germination de semences. Septième Rencontre R, Jul 2018, Rennes, France. . hal-01858677

HAL Id: hal-01858677

<https://institut-agro-rennes-angers.hal.science/hal-01858677>

Submitted on 10 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

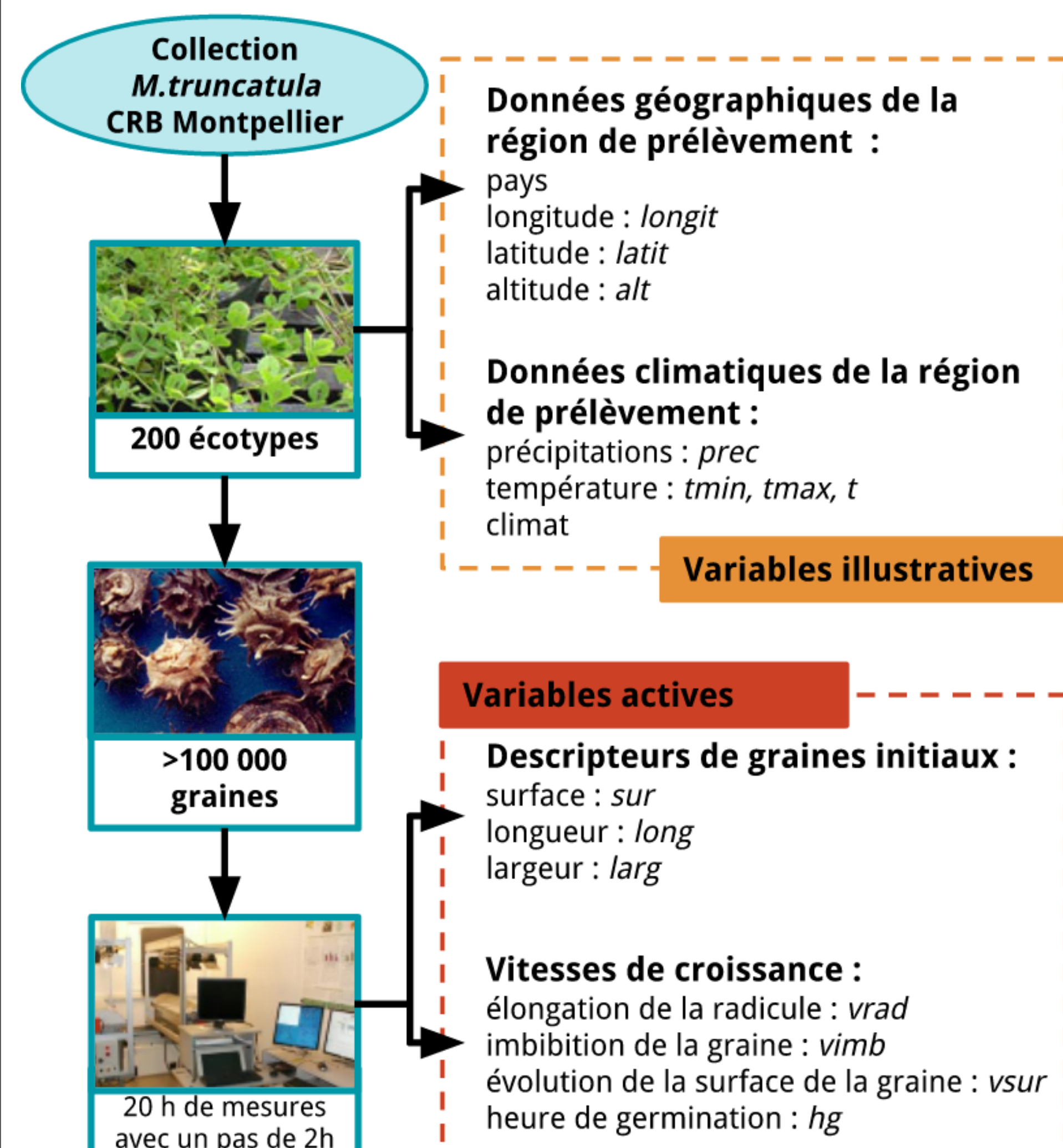
Introduction

- Les technologies d'imagerie haut-débit dans les études de germination permettent l'analyse d'un grand nombre de graines pour chaque plante, produisant de nombreuses données structurées en lots.
- Les analyses multivariées utilisées couramment par les biologistes pour analyser ces données représentent un lot de graines par une valeur unique (moyenne...) pour chaque variable mesurée. Cela masque l'hétérogénéité des lots, qui pourrait être caractéristique du comportement des systèmes biologiques étudiés. Il est donc nécessaire d'évaluer la pertinence de la prise en compte de la variabilité des lots.
- Pour ce faire, nous avons testé différentes méthodes de type ACP sur des données de germination de semences d'écotypes de *Medicago truncatula*. Ces données, issues d'analyses d'images, ont été acquises dans le cadre du projet ANR REGULEG.

Méthodes comparées

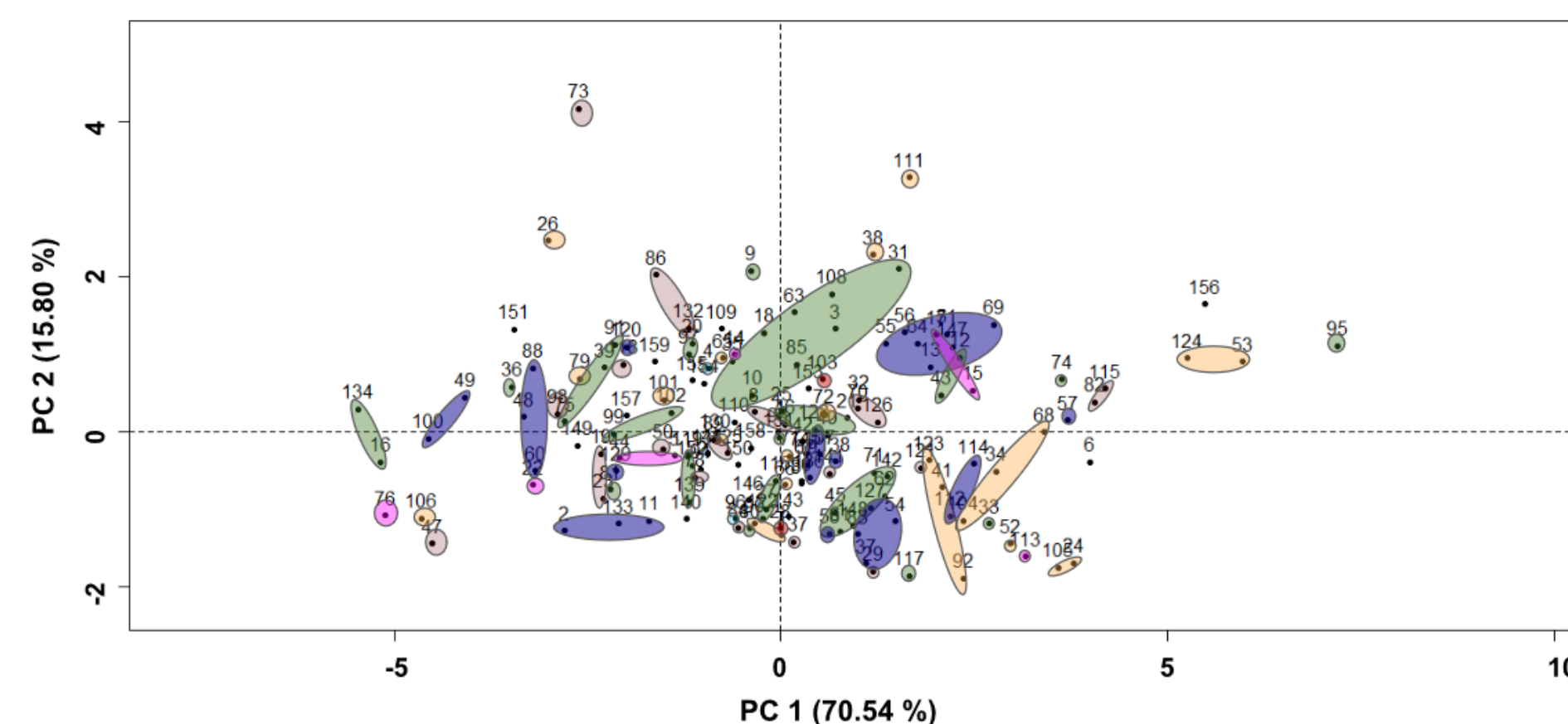
Objets mathématiques associés aux lots	Méthode (Package R)
Vecteur des Moyennes	ACP classique (FactoMineR)
Matrice de Covariance	première étape du STATIS-dual
Matrice de Corrélation	première étape du STATIS-dual
Densité de probabilité (fonction)	ACP de densités (dad)

Données

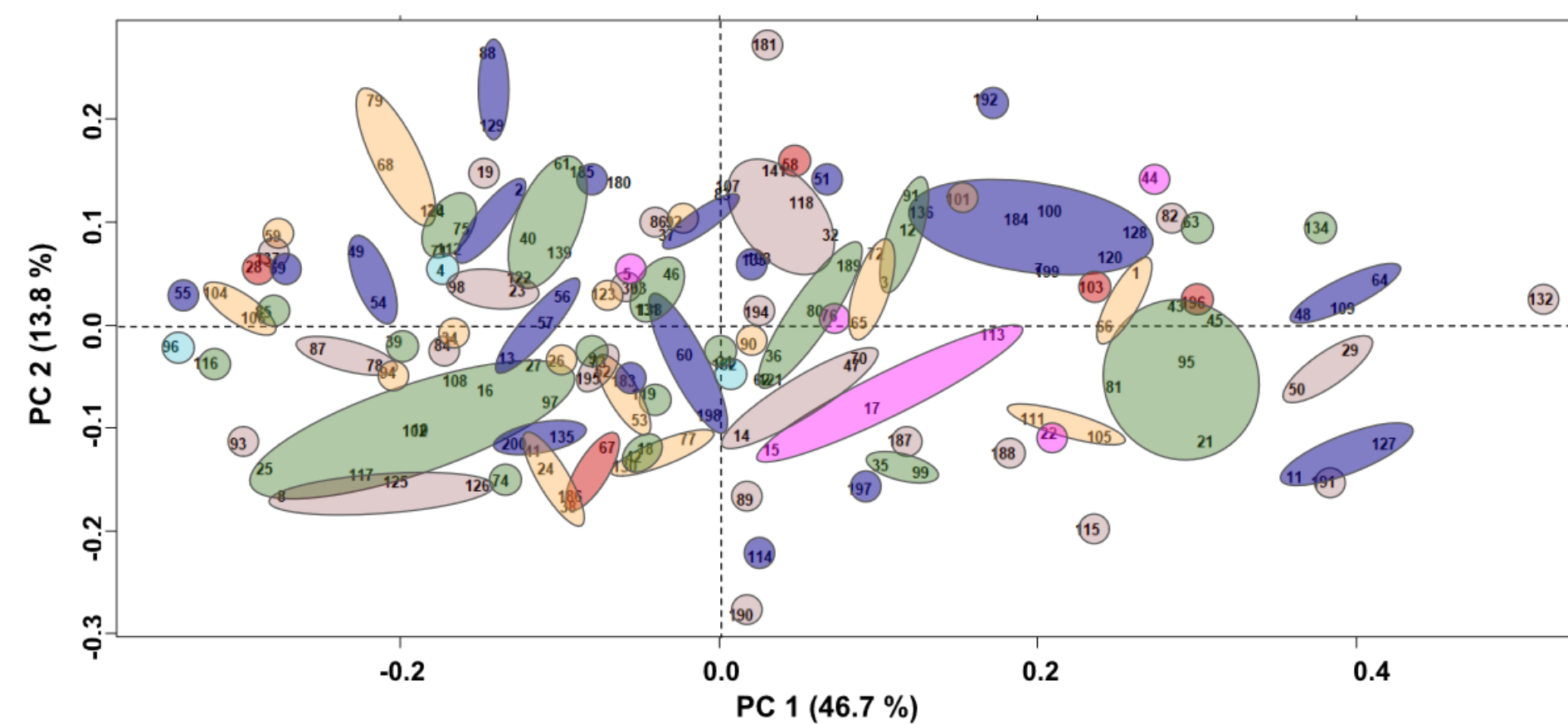


Résultats

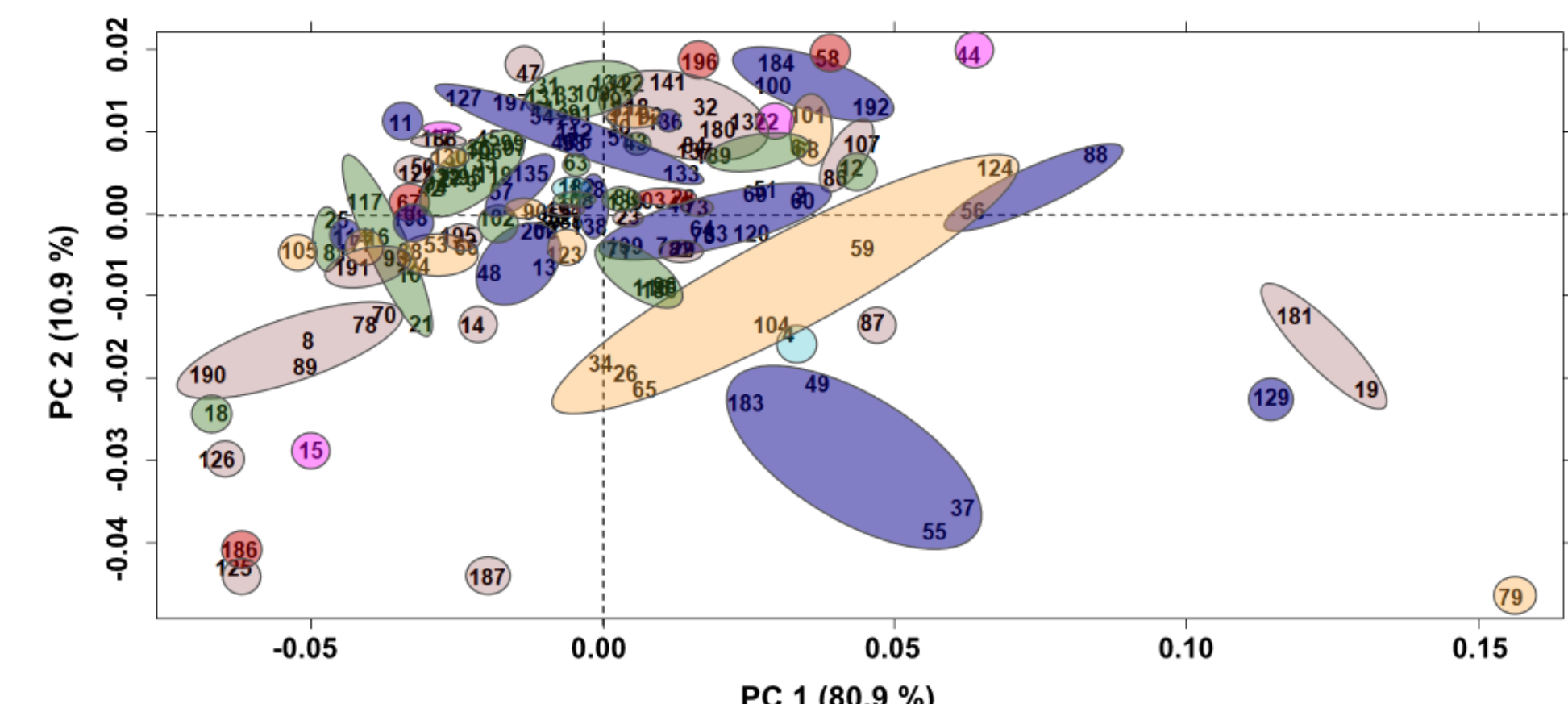
Vecteur des Moyennes



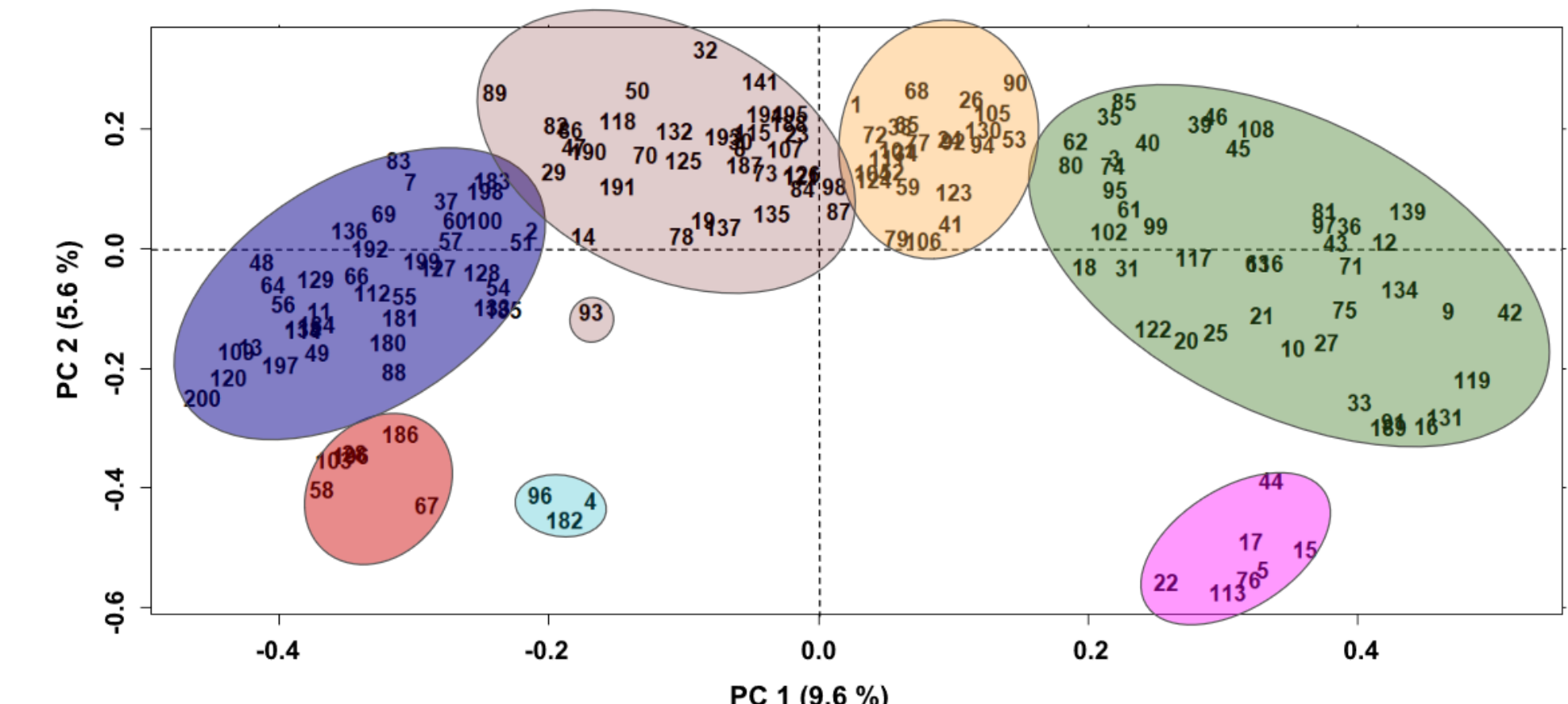
Matrice de Corrélation



Matrice de Covariance



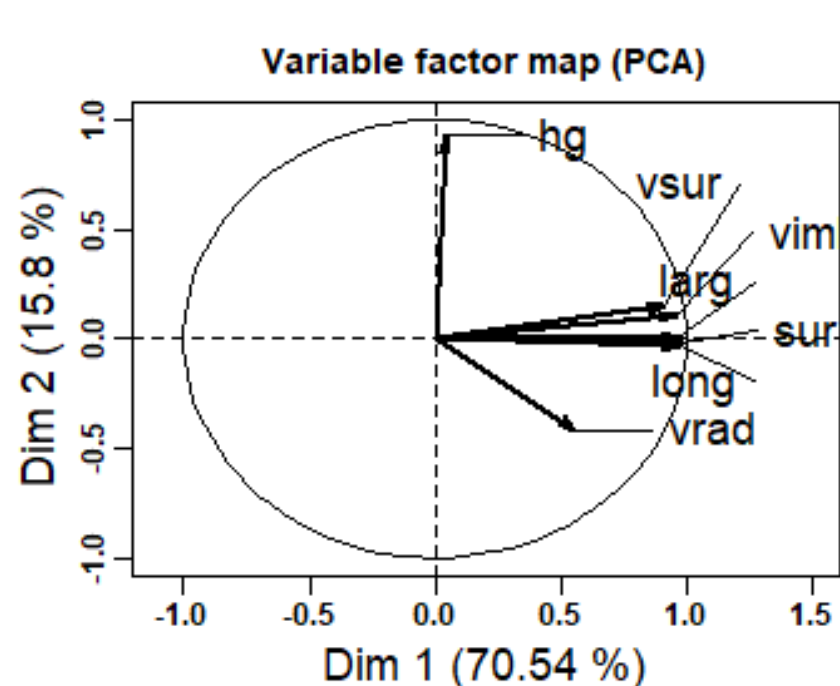
Densité de Probabilité



► Une dispersion des écotypes différente et des regroupements qui émergent selon la méthode.

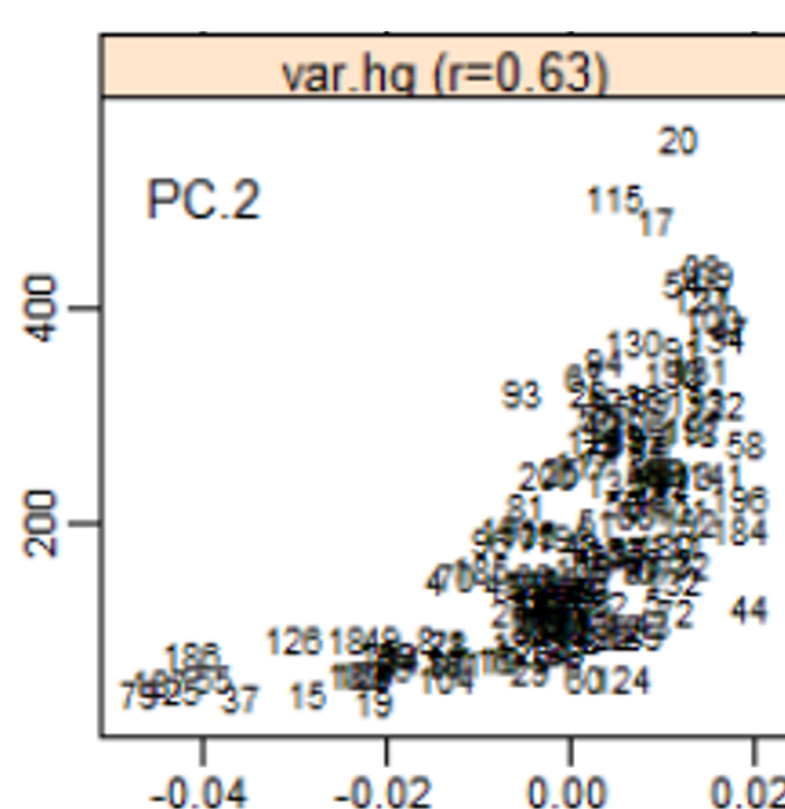
Interprétation

Vecteur des Moyennes



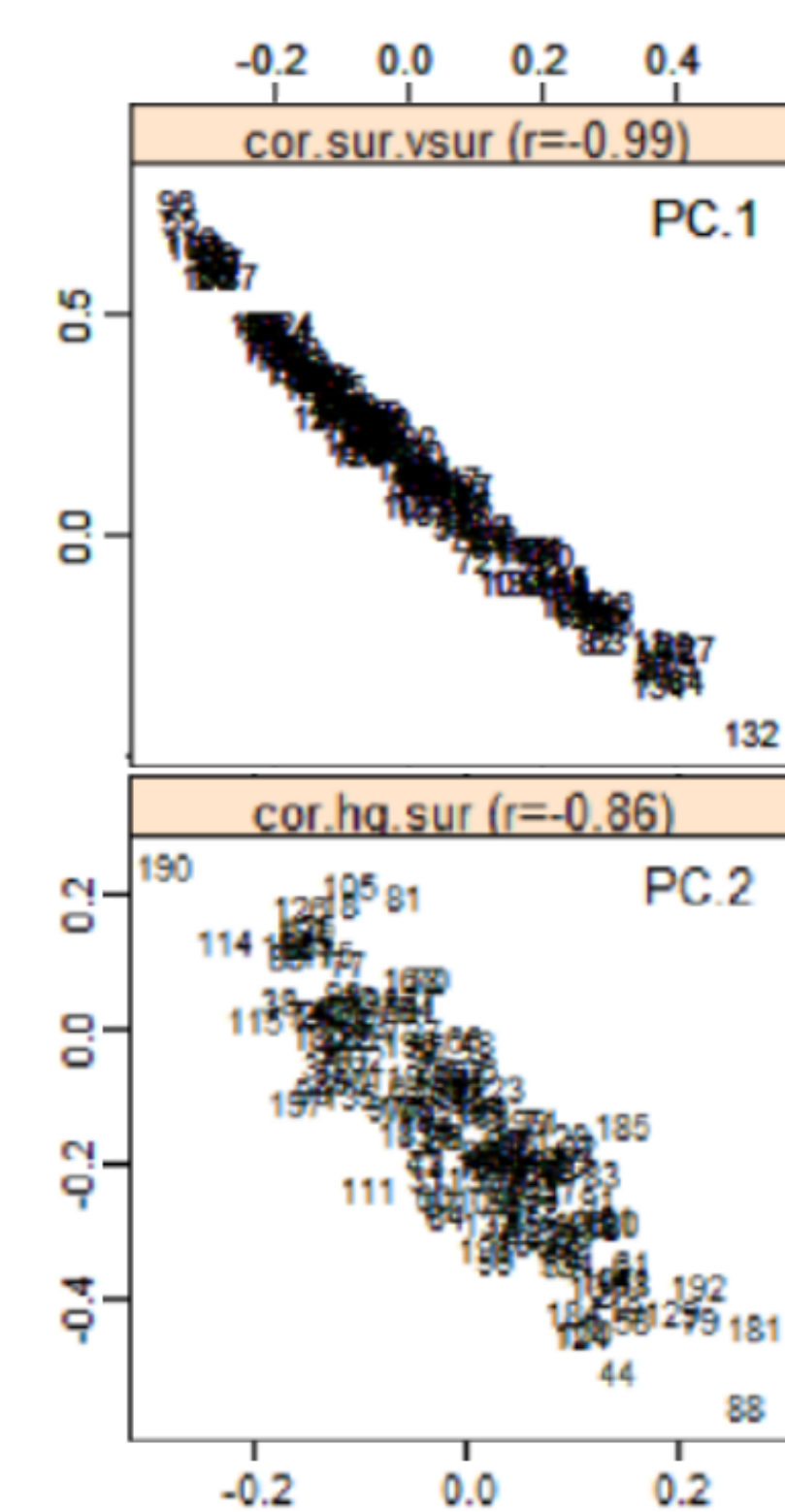
- Corrélation entre les descripteurs initiaux et PC1
- *vimb* = f(taille graine initiale)

Matrice de Covariance



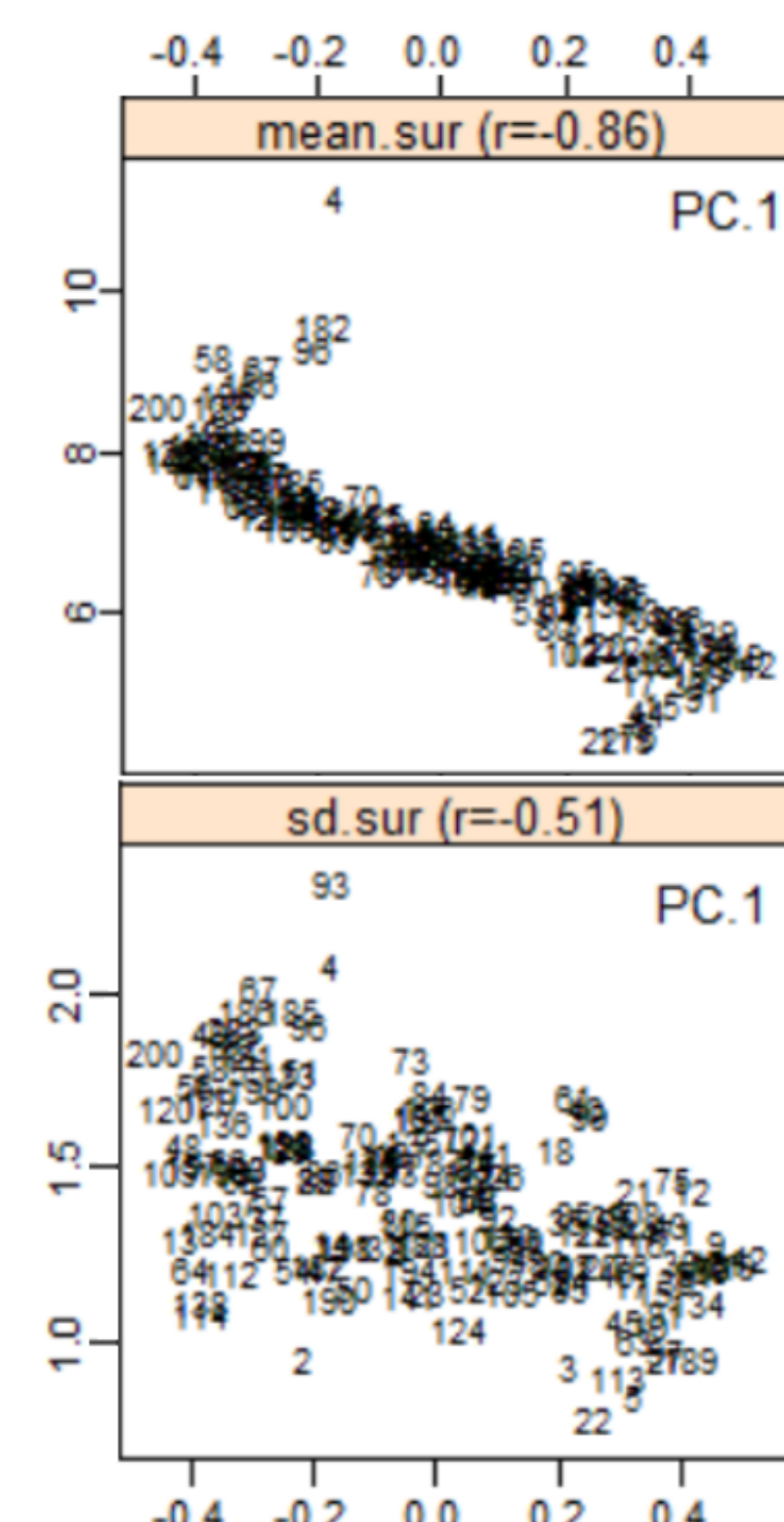
- écotype(faible PC2) = faible variance *hg*

Matrice de Corrélation



- écotype(faible PC1) = forte corrélation *sur.vsur*
- écotype(fort PC2) = forte anti-corrélation *hg.sur*

Densité de Probabilité



- écotype(fort PC1) = faible *sur* avec un faible écart-type

Corrélation de Spearman : données climatiques et valeurs des composantes pour l'analyse multivariée utilisant les densités de probabilités.

Mois	<i>tmin</i>	<i>tmax</i>	<i>prec</i>
1	-0.18	-0.19	-0.21
2	-0.21	-0.22	-0.21
3	-0.23	-0.18	-0.27
4	-0.19	-0.11	-0.2
5	-0.14	-0.07	0.01
6	-0.11	-0.08	0.12
7	-0.09	-0.11	0.18
8	-0.13	-0.12	0.2
9	-0.17	-0.13	0.11
10	-0.16	-0.1	-0.2
11	-0.15	-0.1	-0.34
12	-0.15	-0.15	-0.25

PC.1	
<i>alt</i>	0.17
<i>latit</i>	0.05
<i>longit</i>	0.21

- écotype(fort PC1) = tendance à être originaire de régions aux hivers froids
- écotype(fort PC1) = tendance à être originaire de régions aux étés pluvieux et hivers secs
- écotype(fort PC1) = tendance à être originaires de longitudes élevées au sein de la zone de prélèvement

Références

- Jolliffe, I.T., Principal Component Analysis, Springer Series in Statistics, Springer, NY, 2nd Edition, 2002.
- Husson, F., Josse, J., Le, S., Mazet, J., (2018). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining. R package, version 1.4.0.
- Lavit, C., Escoufier, Y., Sabatier, R., Traissac, P. (1994). The ACT (STATIS method). Computational Statistics and Data Analysis, (18), 97-119.
- Eslami A., Qannari E.M., Bougeard S., Sanchez G., (2015). multigroup: Multigroup Data Analysis. R package version 0.4.4.
- Boumaza, R., Youfi, S., Demotes-Mainard, S. (2015). Interpreting the principal component analysis of multivariate density functions. Communications in statistics. Theory and methods, 44(16): 3321-3339.
- Boumaza, R., Santagostini, P., Youfi, S., Hunault, G., Bourbeillon, J., Pumo, B., Demotes-Mainard, S. (2018). dad: Three-Way Data Analysis Through Densities. R package, version 3.1.0.

Conclusion

- Des méthodes complémentaires qui permettent d'apporter des éléments supplémentaires par rapport à une ACP "classique" : proximité entre écotypes, liens entre variables, etc.
- Pertinence de ne pas se contenter de l'ACP "classique" pour l'analyse préliminaire de données structurées en lots.
- Intérêt du package dad dont le paramétrage permet de reproduire les différentes méthodes comparées ici et qui propose des outils de visualisation pour évaluer la corrélation entre les variables et les scores calculés.

Remerciements

Ce travail a été partiellement financé par le projet REGULEG ANR-15-CE20-0001 et le projet RFI Objectif Végétal DIVIS. Nous remercions la plateforme PHENOTIC pour l'acquisition et l'analyse des images de germination de semences.