



TD – PHYTOSOCIOLOGIE ET STATISTIQUES

AIDE À LA DIAGONALISATION DE GRANDS TABLEAUX DE RELEVÉS

Aurélie DARDILLAC

Conservatoire botanique national de Bailleul

Antenne Normandie

Script associé "Script_TD.R", peut s'ouvrir avec le bloc note.

Obtenir un tableau de donnée exploitable

Supprimer les espèces qui ont une présence inférieure ou égale à 5% dans les tableaux à tester (à adapter en fonction des tableaux). De la même manière supprimer les espèces présentes à plus de 75%, comme précédemment cela dépend du nombre d'échantillons = relevés phytosociologiques, à traiter. Cela permet de diminuer le bruit de fond et de rendre les graphiques plus lisibles. Les espèces supprimées ne doivent en aucun être négligée dans l'analyse phytosociologique ultérieure!

Mise en forme du tableau

- ✓ À partir du tableau Excel constitué entrer la formule suivante :

=remplacer (3 ! C3 ;2 ;2 ;) se placer dans un autre feuillet Excel (3 ! C3 = feuillet et ligne concerné). Cela permet de supprimer le coefficient de sociabilité ;

- ✓ Ensuite Copier et Coller les valeurs pour figer les formules, dans un autre feuillet ;
- ✓ Ensuite remplacer les coefficients numériques par une lettre de A à E pour éviter les sources d'erreur puis remplacer par l'échelle souhaitée. Dans le cas d'un traitement en présence/absence remplacez par 1 les différents coefficients puis remplacer les vides par 0. Il n'est pas utile dans ce cas de passer par l'étape des lettres.

➡ Le choix de l'échelle se fait selon la précision des résultats que l'on souhaite obtenir mais aussi suivant la conception de phytosociologie que l'on a (sigmatiste vs synusiale ...). Pour ma part je réalise les tests en présence/absence mais également avec l'échelle de Gillet qui représente bien les différents coefficients. Lire la synthèse de Meddour 2011 pour une meilleure compréhension (le pdf est dans le dossier ci-joint).

Échelle de Gillet (2000)

Abondance/Dominance	Gillet
5	90
4	57
3	32
2	14
1	3
+	0,3
r	0,03

- ✓ Simplifier les noms d'espèce en prenant les 3 premières lettres du nom de genre et les 3 premières du nom d'espèce. **Attention aux doublons.** Scinder les noms d'espèces en utilisant la fonction convertir puis appliquer la formule suivante : =STXT(B3;1;3) puis CONCATENER.
- ✓ Simplifier les noms de relevés en R1..... ;
- ✓ Après avoir réalisé toutes les étapes précédentes, copier et coller en transposant la matrice.
- ✓ Pour finir enregistrer au format Texte (DOS) = .txt pour que ce soit lisible pour le logiciel, R dans notre cas.

Il faut compter entre 30min et 45min pour obtenir un tableau de donnée exploitable s'il y a un nombre de relevés important. Toujours enregistrer chaque étape pour éviter de devoir tout reprendre au début.

Exploiter les données sur R

Identifier les différents syntaxons

Différentes méthodes existent mais il semble que la plus utilisée à ce jour soit la suivante :

Coupler un AFC avec une CAH

Il s'agit de réaliser une AFC puis d'effectuer une Classification sur les résultats obtenus. Les distances seront calculées à partir de variables euclidiennes, suivant la méthode de Ward.

Le rendu pourra être sous forme de dendrogramme grâce à la fonction `plot()` du package `ade4` soit un graphique plaçant les classes obtenues sur les axes de l'AFC avec la fonction `s.class()`, soit sous la forme d'un autre graphique avec le package `cluster`.

➡ À mon avis les trois représentations sont complémentaires, la troisième est plus parlante et permet de voir si les "classes" définies se recoupent. Le premier permet de comprendre la hiérarchisation et d'identifier les relevés concernés de manière plus claire.

Aide à l'interprétation

AFC, pour choisir le nombre d'axe il faut stopper au plus grand saut (Fig1, ci-contre).

Normalement R sélectionne automatiquement le bon nombre d'axe. Ne pas oublier de préciser le pourcentage d'inertie que l'AFC nous confère. Il est égal à l'addition des contributions à la variance de chaque axe retenu.

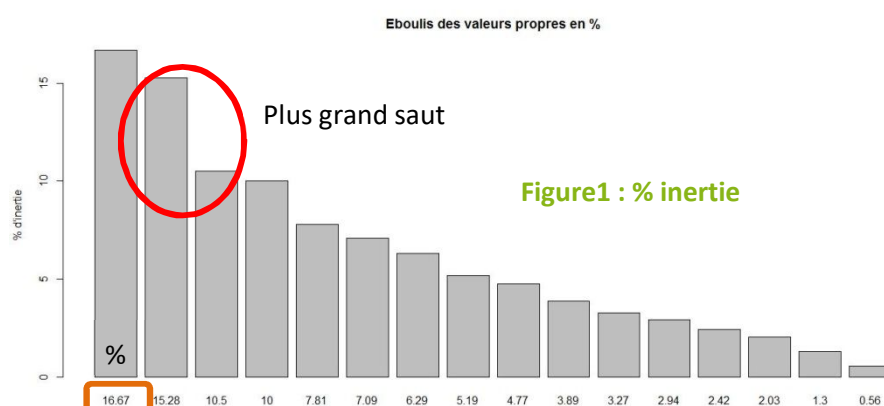
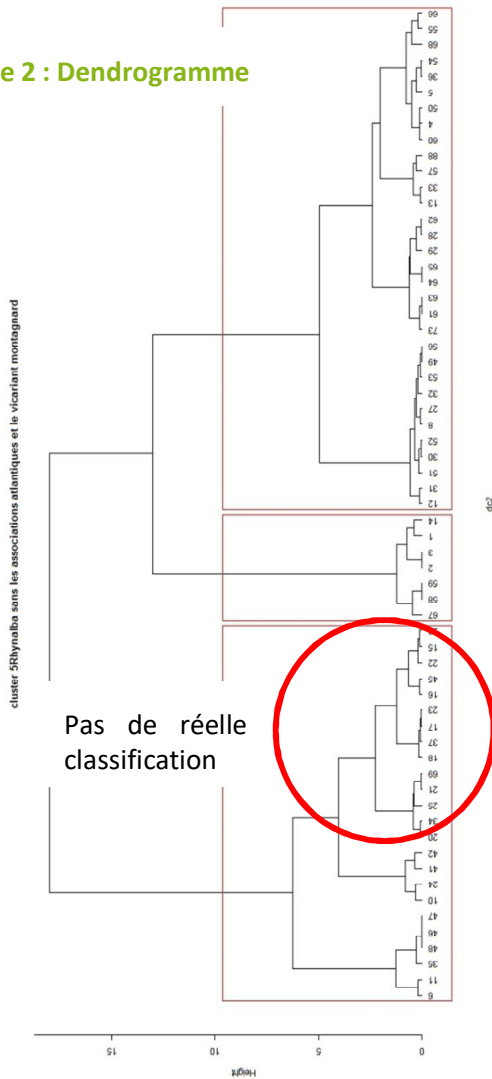


Figure 2 : Dendrogramme



CAH, pour couper le dendrogramme et définir le nombre de classe, il faut là encore couper au milieu de la branche la plus longue (Fig 2). Mais il n'y a pas encore de méthode fixée, donc il faut tester le nombre de classe qui nous paraît le plus approprié. On peut s'aider du graphique généré par la package cluster (Fig 3).

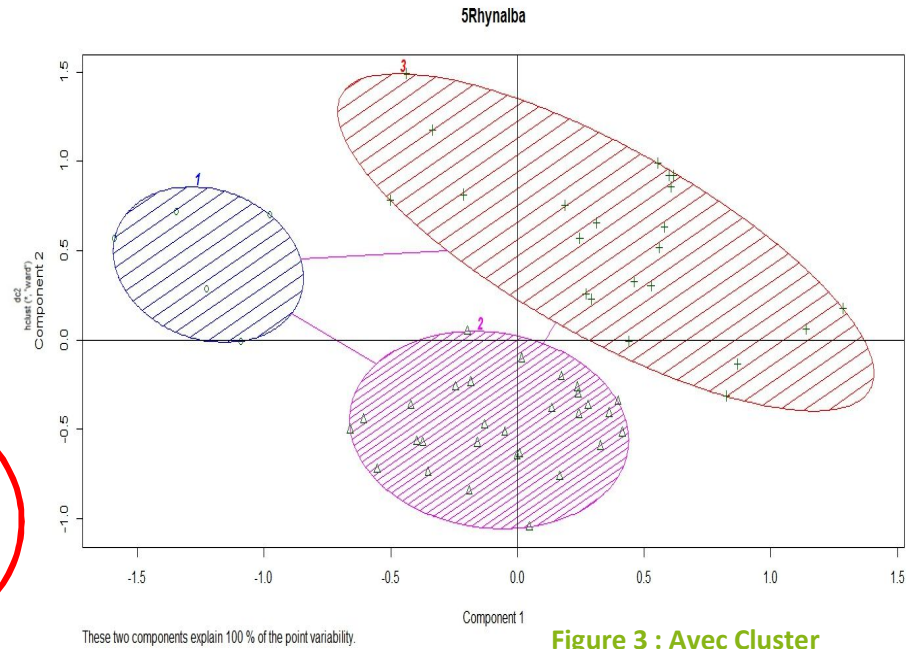


Figure 3 : Avec Cluster

Dans le cas de figure où les distances sont toutes de la même taille cela signifie qu'on ne pas réellement classer (Fig 2).

Une troisième représentation peut être utilisée, elle consiste à représenter les classes de la CAH en esquissant un ovale centré sur les relevés les plus proches et à rattacher les autres avec des traits proportionnels à l'éloignement relatif de "cette classe centrée" (Fig 4).

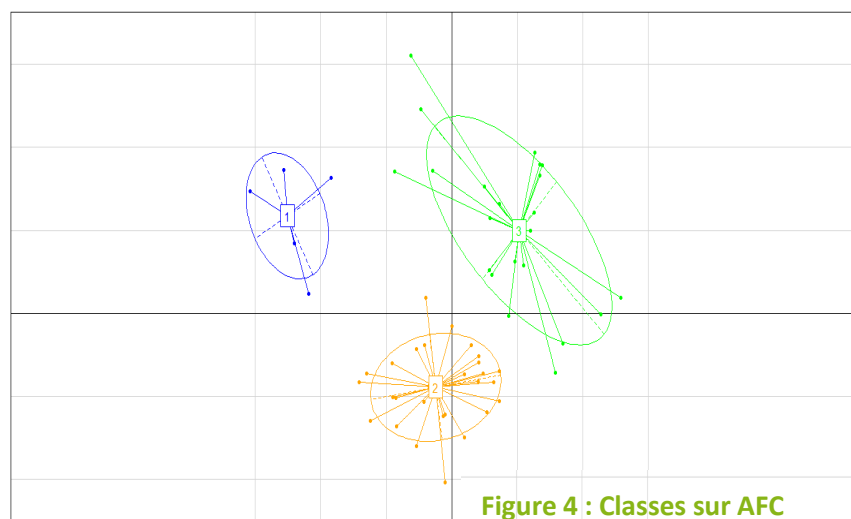


Figure 4 : Classes sur AFC