

Analyse Factorielle des Correspondances : application à des données phytoécologiques

TD

Utilisation du logiciel libre R sur des données de présence/absence issues de relevés floristiques forestiers

Le but de cette session de Travaux Dirigés est de réaliser l'analyse factorielle des correspondances (AFC) d'un tableau floristique codé en présence/absence afin d'avoir une idée de la distribution des espèces et de la structuration des communautés de végétation spontanée dans le Massif du Lomont (Jura). On essaiera d'expliquer la variabilité des coordonnées factorielles des relevés sur les principaux axes factoriels avec des données écologiques issues du terrain pour avoir une idée des principaux gradients écologiques responsable de cette variabilité.

TABLE DES MATIERES

1	INTRODUCTION A L' AFC D' UN TABLEAU FLORISTIQUE.....	3
1.1	POURQUOI UTILISE-T- ON L' AFC EN PHYTO ECOLOGIE ?.....	3
1.2	DEFINITION ET PRINCIPES : APPROCHE PHYTO ECOLOGIQUE	3
1.3	DISTINCTION ENTRE ANALYSE EN COMPOSANTE PRINCIPALE (ACP) ET AFC EN PHYTO ECOLOGIE.....	5
2	JEU DE DONNEES	6
2.1	DONNEES FLORISTIQUES	6
2.1.1	<i>Présentation du tableau floristique brut.....</i>	6
2.1.2	<i>Codage en présence/absence.....</i>	7
2.2	DONNEES ECOLOGIQUES	7
2.2.1	<i>Localisation.....</i>	7
2.2.2	<i>Végétation</i>	8
2.2.3	<i>Conditions stationnelles</i>	8
2.2.4	<i>Conditions climatiques.....</i>	9
2.2.5	<i>Conditions édaphiques</i>	9
3	IMPORT DES DONNÉES DANS LE LOGICIEL LIBRE R	11
3.1	PREPARATION DE VOTRE REPERTOIRE DE TRAVAIL DANS R	11
3.2	IMPORTATION DU TABLEAU FLORISTIQUE SUR LEQUEL SERA REALISE L' AFC.....	11
3.3	IMPORTATION DU TABLEAU ECOLOGIQUE POUR L' INTERPRETATION DE L' AFC	12
4	AFC DU TABLEAU FLORISTIQUE ET INTERPRETATIONS ECOLOGIQUES	13
4.1	REALISATION DE L' AFC	13
4.1.1	<i>Quelques commandes utiles à connaître avant de lancer une AFC</i>	13
4.1.2	<i>Chargement du package ade4 de l' Université Lyon 1.....</i>	13
4.1.3	<i>Lancement de l' AFC.....</i>	14
4.1.4	<i>Accéder aux valeurs propres de l' AFC.....</i>	14
4.1.5	<i>Edition des cartes factoriels.....</i>	15
4.2	INTERPRETATION DE L' AFC.....	15
4.2.1	<i>Dépouillement de l' axe 1.....</i>	15
4.2.2	<i>Dépouillement de l' axe 2.....</i>	16

1 INTRODUCTION A L' AFC D' UN TABLEAU FLORISTIQUE

1.1 Pourquoi utilise-t-on l' AFC en phytoécologie ?

En phytoécologie, on cherche généralement à mettre en évidence des associations entre espèces (notion de communautés végétales) ou bien à faire des groupes écologiques de relevés (notion de typologies de station). De manière intuitive, cela revient à regrouper les espèces et les relevés qui se ressemblent, soit en associant les espèces présentes dans les mêmes relevés, ou bien en regroupant les relevés ayant des profils floristiques similaires. Effectuer ce travail à la main, dès l'instant où l'on dispose d'un nombre important de données, représente un travail long et fastidieux.

Le recours à l'outil statistique permet d'automatiser une procédure de tri par agrégations des espèces qui vont ensemble et des relevés qui se ressemblent. En ce sens, l'Analyse Factorielle des Correspondances (AFC) d'un tableau floristique est une méthode semi-automatique permettant de représenter géométriquement, dans un espace à plusieurs dimensions, les « distances écologiques » qui séparent les espèces suivant leurs dispositions dans les relevés ainsi que pour les relevés, suivant la composition de leurs profils floristiques.

La difficulté associée à cette méthode d'ordination réside dans la représentation que l'on se fait d'un espace à plusieurs dimensions (autant que d'espèces ou de relevés). Néanmoins, il est possible de simplifier cet espace par projection sur une, deux ou trois dimensions afin de mieux « voir » les distances qui séparent chaque espèce ou chaque relevés. Le jeu consiste à bien choisir la projection la plus pertinente de cet espace à plusieurs dimensions dans un espace réduit à un plan, ou un axe, de façon à résumer au mieux la variabilité, ou inertie du nuage des espèces (resp. des relevés).

Par analogie, si l'on veut décrire un chameau à quelqu'un qui n'a jamais vu l'animal, le message le plus explicite sera de dessiner sur le papier une représentation de profil, c'est le plan factoriel qui décrit le mieux l'animal en question. L'objectif de l'AFC appliquée au tableau floristique est identique à celui de vouloir expliquer par le dessin ce que l'on a retenu de l'allure générale du chameau (simplification de notre réalité tridimensionnelle sur une feuille de papier plane).

En définitive, l'AFC permet d'**analyser** un espace multidimensionnel, afin d'y trouver un axe ou bien un plan dit « **factoriel** » qui maximise simultanément la **correspondance** entre les espèces et les relevés.

1.2 Définition et principes : approche phytoécologique

L'AFC a été conçu au départ comme une méthode de statistique descriptive destinée à étudier les relations entre deux **variables qualitatives** à partir d'un « **tableau de contingence** ». Cette table de contingence correspond au croisement des différentes modalités qui composent chacune des deux variables qualitatives : à chaque case du tableau est affecté l'effectif des individus présentant les modalités considérées. Par exemple, soit une variable peuplement à trois modalités, et une variable humus à trois modalités. On effectue l'AFC sur la table de contingence suivante :

	Moder	Hemimoder	Eumoder	Total
Peuplement feuillus	11	9	10	30
Peuplement résineux	13	10	7	30
Mélange feuillus/résineux	10	10	10	30
Total	34	29	27	90

Tableau 1 : Table de contingence Humus/Peuplement

Or un tableau floristique n'est pas un tableau de contingence : on parle plutôt de « **tableau disjonctif semi-complet** » : chaque case du tableau contient une information codée en présence/absence pour l'espèce j dans le relevé i :

	Espèce1	Espèce2	Espèce3	Espèce j	Espèce P
Relevé n°1	0	1	1	...	1
Relevé n°2	1	0	0	...	0
Relevé n°3	1	0	1	...	1
Relevé n° i
Relevé n° N	0	0	1	...	1

Tableau 2 : Tableau floristique codé en présence/absence

Si en phytoécologie, l'objectif de l'AFC est l'étude de la correspondance entre les espèces, alors il faut attribuer à chaque espèce du tableau floristique un rôle de variable qualitative à deux modalités (présence ou absence) et à chaque relevé le rôle d'individu. Par analogie avec l'AFC classique, cela reviendrait à construire une table de contingence très complexe sur un nombre important de variables bimodales. Pour retomber dans le cas de l'AFC classique, il faudrait que sur l'ensemble des relevés étudiés, le nombre total d'espèces atteigne deux (cas très simple, non rencontré dans la réalité). Dans ce cas très simplifié, on pourrait transformer le tableau floristique en une table de contingence du type :

	Espèce1 (présence)	Espèce1 (absence)
Espèce2 (présence)	24	12
Espèce2 (absence)	10	14

Tableau 3 : Table de contingence Espèce1/Espèce2

Pour trois espèces, on peut aussi représenter la table de contingence dans un cube à deux modalités par côté. Au delà de trois, la représentation de la table de contingence est impossible. Néanmoins, lorsque le nombre de variables qualitatives est supérieur à deux, il est possible d'effectuer une AFC non plus sur la table de contingence, mais sur un « **tableau disjonctif complet** ». Il s'agit d'effectuer ce que l'on appelle un codage disjonctif des variables. Reprenons l'exemple de l'AFC classique sur les variables Humus/Peuplement : dans ce cas, chaque variable dispose de trois modalités. Parmi les 90 relevés issus de l'effectif total, considérons 5 relevés à titre d'exemples et contenant les modalités Humus/Peuplement suivantes :

- R1 : Moder/**Résineux**
- R2 : Eumoder/**Feuillus**
- R3 : Moder/**Résineux**
- R4 : Hemimoder/**Résineux**
- R5 : Eumoder/**Mélange**

Pour ces 5 relevés, le tableau disjonctif complet correspond à un éclatement des modalités de chacune des deux variables Humus et Peuplement en un ensemble de 6 variables binaire (0/1) :

	Moder	Hemimoder	Eumoder	Résineux	Feuillus	Mélange
R1	1	0	0	1	0	0
R2	0	0	1	0	1	0
R3	1	0	0	1	0	0
R4	0	1	0	1	0	0
R5	0	0	1	0	0	1

Tableau 4 : Tableau disjonctif complet sur 5 relevés

Le résultat de l'AFC sur le tableau disjonctif complet contenant les 90 relevés de l'effectif total donne les mêmes résultats que l'AFC effectué sur la table de contingence présentée plus haut. Cette représentation permet cependant de travailler sur un nombre plus grand de variables qualitatives, autant que d'espèces contenues dans l'ensemble des relevés. C'est sur la base de ce constat que l'on s'appuie pour dire que l'AFC d'un tableau floristique correspond à réaliser une AFC sur une table de contingence à n variables (n étant le nombre d'espèces). Cependant dans le cas d'un tableau floristique, le tableau disjonctif n'est pas complet, étant donné le rôle symétrique de la variable espèce (présence/absence). En toute logique, il faudrait compléter le tableau floristique des présences par celui des absences pour obtenir un tableau disjonctif complet (information redondante).

Par conséquent le lien qui existe entre AFC classique et AFC sur tableau floristique, c'est le passage par un **codage disjonctif** des variables.

1.3 Distinction entre Analyse en Composante Principale (ACP) et AFC en phytoécologie

L'AFC est une méthode proche de l'ACP, que l'on emploie lorsque l'on dispose de données qualitatives. Comme l'AFC, l'ACP est utilisée à des fins descriptives pour étudier les ressemblances entre variables et les proximités entre individus. Cependant, l'ACP étudie les ressemblances entre les variables de nature quantitative par l'intermédiaire des **distances euclidiennes**, tandis que l'AFC s'intéresse à la relation entre deux variables (ou plus) de nature qualitative par comparaison des **distances du Chi2**.

2 JEU DE DONNEES

Le jeu de données est extrait de l'exercice phytoécologique 2006 réalisé par la 16^{ème} promotion de la Formation des Ingénieurs Forestiers (FIF) dans le massif du Lomont (Jura). Le but de cet exercice était la réalisation d'une typologie de stations forestières dont l'objectif est d'aider les gestionnaires forestiers à établir un diagnostic rapide des conditions écologiques existant sur une parcelle forestière. Un tel diagnostic permet d'adapter la gestion aux conditions écologiques en conciliant production de bois de qualité et respect de l'environnement dans le contexte actuel de production durable. La **méthode d'échantillonnage** employée par les élèves au cours de l'exercice est celle des **transects**.

On a cherché ensuite, à équilibrer *a posteriori* notre échantillon de relevés en utilisant un plan stratifié suivant l'altitude et l'exposition étant donné l'importance de ces deux facteurs dans un massif forestier comme celui du Lomont :

- L'altitude est un facteur très structurant de la végétation : on peut distinguer un étage collinéen aux altitudes inférieures à 500m et un étage montagnard aux altitudes supérieures à 500m ;
- L'influence microclimatique est également très importante dans les massifs du Jura, avec de fortes oppositions de versants (Nord-Sud et Ouest-Est) : on distingue 5 modalités d'exposition différentes ;

Pour cette séance de travaux dirigés, nous avons sélectionné 125 relevés (afin de limiter la taille du jeu à manipuler et d'équilibrer le plan d'échantillonnage) parmi l'ensemble des relevés effectués par les élèves de la 16^{ème} promotion de la FIF.

On obtient ainsi un plan comprenant 10 strates :

	Exposition N	Exposition E	Exposition S	Exposition O	Plateau	Total
Etage collinéen	8	9	12	12	11	52
Etage montagnard	25	7	23	10	8	73
Total	33	16	35	22	19	125

Tableau 5 : Plan d'échantillonnage stratifié suivant l'altitude et l'exposition

2.1 Données floristiques

- Le tableau floristique se trouve dans le fichier **flo.txt**.

2.1.1 Présentation du tableau floristique brut

Il est le fruit d'une numérisation des relevés de terrain. Les données saisies à l'issue de la phase de terrain sont très rarement utilisables en l'état et il est souvent nécessaire de les retravailler pour pouvoir les analyser. La meilleure façon de présenter les données pour l'analyse est de disposer les relevés en lignes et les espèces en colonnes. Le tableau « flo » comporte 125 lignes (une ligne = un relevé) et 143 colonnes (une colonne = une espèce végétale spontanée). Seules la flore spontanée, issue de la strate herbacée ou de la strate arbustive basse, est traitée ici car peu ou moins influencée par la gestion forestière que les espèces issues de la strate arboré. Font partie de la strate herbacée ou de la strate arbustive basse, les bryophytes, les ptéridophytes et les plantes vasculaires. Le cas des plantes grimpantes comme le lierre terrestre est considéré comme appartenant à la flore spontanée. Le nom des espèces est indiqué par un radical à 4 lettres. Pour

obtenir le nom scientifique complet en latin, voir le fichier **espece.txt** qui donne les correspondances entre code radical à 4 lettres et nom complet.

2.1.2 Codage en présence/absence

Les données sont codées de manière binaire (0/1) :

0. absence

1. présence

2.2 Données écologiques

• Le tableau écologique se trouve dans le fichier **eco.txt**.

Il représente la numérisation des données écologiques disponibles sur les 125 placettes (données relevés sur le terrain, ou issues des valeurs indicatrices des espèces, ou encore issues de croisements sous système d'information géographique).

2.2.1 Localisation

- **ID** : Identifiant unique de la placette. Correspond à la concaténation du numéro du groupe d'élèves, du numéro du transect effectué, ainsi que du numéro du relevé effectué sur le transect.

- **St1** : Strate altitudinal ou étage d'appartenance du relevé, première strate d'échantillonnage à deux modalités :

1. étage collinéen (alt<500m)

2. étage montagnard (alt>500m)

- **St2** : Exposition du site, deuxième strate d'échantillonnage à cinq modalités :

1. exposition nord N (entre 350 et 50 grad)

2. exposition est E (entre 50 et 150 grad)

3. exposition sud S (entre 150 et 250 grad)

4. exposition ouest O (entre 250 et 350 grad)

5. situation de plateau P (aucune exposition)

- **Stnum** : Résultat de la concaténation des strates 1 et 2, 10 modalités croisées.

- **XLamb2** : Coordonnées longitudinales (m) en X Lambert II (étendu) pointées sur carte.

- **YLamb2** : Coordonnées latitudinales (m) en Y Lambert II (étendu) pointées sur carte.

- **Alt** : Altitude du site par rapport au niveau moyen de la mer, mesuré à l'aide d'un altimètre étalonné (m).

- **Expo** : Exposition du site (grad).

2.2.2 Végétation

- **Habitat** : Code Corinne (CB) des habitats forestiers rencontrés. Habitats identifiés à l'aide de la clé de détermination des habitats forestiers du Nord-Est de la France.
- **RecArbre** : Recouvrement de la strate arborée (%).
- **RecArbu** : Recouvrement de la strate Arbustive (%).
- **RecHerb** : Recouvrement de la strate Herbacée (%).
- **RecMous** : Recouvrement de la strate Muscinale (%).
- **Traitement** : Traitement sylvicole de la placette en six modalités :
 1. **FI** : peuplement traité en futaie irrégulière
 2. **FR** : peuplement traité en futaie régulière
 3. **T** : taillis
 4. **TC** : taillis à courte rotation
 5. **TSF** : taillis sous futaie
 6. **TSFV** : taillis sous futaie vieilli

2.2.3 Conditions stationnelles

- **Cosexpo** : Cosinus de l'exposition (converti en radians), mesure les oppositions de versants Sud-Nord (vaut -1 pour l'exposition Sud et 1 pour l'exposition Nord, annulation pour les expositions Est et Ouest).
- **Sinexpo** : Sinus de l'exposition (converti en radians), mesure les oppositions de versants Ouest-Est (vaut -1 pour l'exposition Ouest et 1 pour l'exposition Est, annulation pour les expositions Nord et Sud).
- **Pente** : Pente du site à l'endroit du relevé (%).
- **Confin** : Confinement du site à l'endroit du relevé, calculé à l'aide du masque (%).
- **Drainage** : Bilan des circulations latérales d'eau du sol liées à la topographie du site. Quatre modalités de drainage latéral :
 0. départs d'eau
 1. apports et départs équilibrés

2. apports faibles à moyens

3. apports importants

2.2.4 Conditions climatiques

- **PA** : Somme des précipitations annuelles (mm). Données issues du modèle AURELHY au pas de 1km qui couvre l'ensemble du territoire français. Récupération de l'information par croisement SIG des relevés géoréférencés avec la couche AURELHY des précipitations.
- **TmoyA** : Température moyenne annuelle (°C). Données issues du modèle AURELHY au pas de 1km qui couvre l'ensemble du territoire français. Récupération de l'information par croisement SIG des relevés géoréférencés avec la couche AURELHY des températures.
- **Tmin01** : Minimum des températures du mois de janvier (°C). Données issues du modèle AURELHY au pas de 1km qui couvre l'ensemble du territoire français. Récupération de l'information par croisement SIG des relevés géoréférencés avec la couche AURELHY des températures.
- **DM** : Indice de De Martonne : $P/(T+10)$, mesure l'aridité du site. Données issues d'un croisement SIG au pas de 1km sur la France.
- **BH06** : Bilan Hydrique du mois de juin, correspond au bilan entre apports d'eau (précipitations) et départs d'eau (évapotranspirations). Données issues d'un croisement SIG au pas de 1km sur la France.

2.2.5 Conditions édaphiques

- **ProfArret** : Profondeur d'arrêt (cm) de la fosse pédologique.
- **CauseArret** : Cause d'arrêt de la prospection. Quatre modalités :

1. horizon compact

2. éléments grossiers

3. dalle

4. volontaire

- **Profondeur disparition des racines fines** : Profondeur (cm) à partir de laquelle les racines fines ne sont plus visibles sur le profil pédologique. Correspond à la profondeur prospectable par les racines (interprétation du profil).
- **ProfHCL** : Profondeur (cm) d'apparition de l'effervescence à l'acide chlorhydrique (dans la terre fine).
- **Oln** : Présence d'une couche Oln dans l'Humus. Trois modalités :

1. Oln absent

2. Oln discontinu

3. Oln continu

- **Olv** : Présence d'une couche Olv dans l'Humus. Trois modalités :

1. **Olv absent**
2. **Olv discontinu**
3. **Olv continu**

- **OF** : Présence d'une couche OF dans l'Humus. Trois modalités :

1. **OF absent**
2. **OF discontinu**
3. **OF continu**

- **OH** : Présence d'une couche OH dans l'Humus. Trois modalités :

1. **OH absent**
2. **OH discontinu**
3. **OH continu**

- **OHmm** : Epaisseur de l'horizon OH (cm).

- **Agrum** : Présence d'un horizon A grumeleux. Deux modalités :

0. **absence**
1. **présence**

- **Ahum** : Présence d'un horizon A humifère. Deux modalités :

0. **absence**
1. **présence**

- **Humus** : Type d'humus. Sept modalités :

1. **amphimull**
2. **mor**
6. **dysmull**
7. **oligomull**
8. **mesomull**
9. **eumull**
10. **hydromull**

- **RU** : Réserve utile du sol (mm). Calculée à partir de la profondeur prospectable par les racines, de la texture par horizon, ainsi que de la charge en éléments grossiers par horizons.

3 IMPORT DES DONNÉES DANS LE LOGICIEL LIBRE R

3.1 Préparation de votre répertoire de travail dans R

- Lancez le logiciel libre R.

- Demandez à R le répertoire de travail actuel.

- Pour cela, utilisez la fonction « **getwd()** » :

```
> getwd() # Indique le chemin d'accès au répertoire de travail actuel
```

```
[1] "C:/Users/admin2/Documents"
```

- Indiquez à R le chemin d'accès à votre nouveau répertoire de travail contenant l'ensemble des jeux de données à analyser (exemple dossier intitulé « **Données** »).

- Pour cela, utilisez la fonction « **setwd()** » en indiquant le chemin d'accès à vos données :

```
> setwd("C:/.../TD AFC-CAH/Données") # Paramètre le nouveau répertoire de travail
```

- Vérifiez le contenu de votre nouveau répertoire de travail.

- Pour cela, utilisez la fonction « **dir()** » :

```
> dir("C:/.../TD AFC-CAH/Données") # Affiche le contenu du répertoire de travail
```

```
[1] "eco.txt"          "flo.txt"          "espece.txt"
```

3.2 Importation du tableau floristique sur lequel sera réalisé l'AFC

- Importez le fichier **flo.txt** dans R.

- Pour cela, utilisez la fonction « **read.table()** » :

```
> flo <- read.table("flo.txt", header=TRUE, sep="\t") # Créer un objet « flo » contenant les données floristiques
```

- Attention lors de l'importation des données, il est toujours conseillé de vérifier la nature des variables (integer, double, factor, character, boolean, etc.) et la structure de l'objet ainsi créé par le logiciel R.

- Pour cela, utilisez la fonction « **str()** » :

```
> str(flo) # Donne des informations précieuses sur l'objet « flo » et la nature des variables qu'il contient
```

```
'data.frame': 125 obs. of 144 variables:
```

```
$ Releve: int 10101 10102 10103 10104 10105 10106 10107 10601...  
$ acsp: int 0 0 0 0 0 0 0 0 0 0...  
$ adal: int 0 0 0 0 0 0 0 0 0 0...  
[list output truncated]
```

3.3 Importation du tableau écologique pour l'interprétation de l'AFC

- Importez le fichier **eco.txt** dans R et créez un objet nommé « eco » contenant les données écologiques à utiliser pour l'interprétation de l'AFC.

- Pour cela, utilisez les fonctions que nous venons de voir ensemble.

♦ *Quelle est la nature de chaque variable ?*

4 AFC DU TABLEAU FLORISTIQUE ET INTERPRETATIONS ECOLOGIQUES

- Vérifier que le tableau floristique (objet « **flo** ») et le tableau écologique (objet « **eco** ») ont bien été importés dans R.

- Pour cela, utilisez la fonction « **ls()** » :

```
> ls() # Liste l'ensemble des objets importés et créés au cours de cette session R
```

4.1 Réalisation de l'AFC

4.1.1 Quelques commandes utiles à connaître avant de lancer une AFC

```
> flo[, 1] # Appel la première colonne du tableau floristique
```

```
> flo[, "aepo"] # Appel la colonne nommé « aepo » contenant le vecteur d'information présence/absence par relevés de l'espèce Aegopodium podagraria
```

```
> flo[1, ] # Appel la première ligne
```

```
> flo[flo$Releve=="10101", ] # Appel la ligne nommé « 10101 » contenant le premier relevé du groupe 1 sur le transect numéro 1
```

```
> flo[, 1:2] # Appel les 2 premières colonnes du tableau
```

```
> flo[, c("acsp", "adal")] # Appel les 2 premières colonnes du tableau
```

```
> flo$aepo # Appel la colonne nommé « aepo »
```

4.1.2 Chargement du package ade4 de l'Université Lyon 1

Pour réaliser une AFC avec le logiciel libre R, il est possible d'utiliser plusieurs fonctions disponible dans différents package :

1. dans le package **MASS**, fonction « **corresp()** »
2. dans le package **ade4**, fonction « **dudi.coa()** »
3. dans le package **FactoMineR**, fonction « **CA()** »

Dans le cadre de ce TD, nous allons utiliser le package **ade4** qui dispose de nombreuses fonctions graphiques pour visualiser et interpréter les résultats de l'AFC.

- Vérifiez que vous disposez du package **ade4** dans votre librairie de modules et sinon importer le package **ade4**.

- Pour vérifier la liste des packages installés dans votre librairie (cf. en dur sur votre disque), utilisez la fonction « **library()** » et si le package **ade4** n'y figure pas, utilisez les fonctions « **chooseCRANmirror()** » et « **install.packages()** » pour sélectionner un miroir à partir duquel il vous sera possible de télécharger le package **ade4** :

> `library()` # Une liste détaillant le contenu des packages installés sur votre ordinateur apparaît

> `chooseCRANmirror()` # Une liste de miroir apparaît, sélectionner le miroir « France (Lyon1) » par exemple (NB : une connexion Internet est nécessaire)

> `install.packages("ade4")` # Télécharge et installe le package `ade4` dans votre librairie (NB : une connexion Internet est nécessaire)

- Une fois installé dans votre librairie de packages, il faut charger le package **ade4** dans R.

- Pour cela, utilisez la fonction « `library()` » :

> `library("ade4")` # Charge le package `ade4` dans votre espace de travail actuel

> `?dudi.coa` # Affiche l'aide de la fonction « `dudi.coa()` »

4.1.3 Lancement de l'AFC

- Lancer l'AFC sur le jeu de données « **flo** » (**relevés en lignes et espèces en colonnes**) en faisant bien attention d'exclure la première colonne contenant les identifiants des relevés.

- Pour cela, utilisez la fonction « `dudi.coa()` » et créez un objet qui va recevoir le résultat de votre AFC :

> `monafc <- dudi.coa(flo[, 2:144], scannf=F, nf=3)` # Lance l'AFC sur le jeu de données « `flo` » et stock le résultat dans l'objet « `monafc` » (NB : on peut indiquer le nombre d'axes que l'on souhaite conserver à l'aide l'argument « `nf` »)

> `monafc` # Affiche un résumé contenant l'ensemble des résultats de votre AFC

> `summary(monafc)` # Edite le rapport de votre AFC contenu dans l'objet « `monafc` »

4.1.4 Accéder aux valeurs propres de l'AFC

- Récupérez les valeurs propres de chaque axe factoriel qui renseignent sur le pourcentage d'inertie (variance) du nuage de points (relevés) portée par chaque axe factoriel.

- Pour cela, utilisez et décortiquez l'objet « **monafc** » qui contient le vecteur de valeurs propres ou « **eigenvalues** » en anglais :

> `monafc$eig` # Renvoie le vecteur de valeurs propres issues de votre AFC

> `inertie <- (monafc$eig/sum(monafc$eig))*100` # Créer un objet « `inertie` » qui contient les valeurs en pourcentage d'inertie et dont la somme des éléments vaut 100%

> `barplot(inertie, ylab="% d'inertie", names.arg=round(inertie, 2))` # Affiche l'histogramme des valeurs propres classées de manière décroissante

> `title("Eboulis des valeurs propres en %")` # Ajoute un titre à votre histogramme

> `screepplot(monafc)` # Une façon plus rapide d'obtenir la même chose

♦ **Quel est le nombre d'axes à conserver ?**

4.1.5 Edition des cartes factoriels

- Editez la position des espèces et des relevés dans le plan factoriel F1/F2 de votre AFC.

- Pour cela, utilisez la fonction « `scatter.coa()` » :

```
> scatter.coa(monaafc, clab.row=0, clab.col=0.5, sub="Espèces
(radicaux) et relevés (points)", posieig="none")# Affiche le graphique du nuage
de points des espèces et des relevés dans le plan factoriel F1/F2 avec les espèces situées aux
barycentres des relevés dans lesquelles elles sont présentes et les relevés situés aux barycentres des
espèces qui les composent
```

```
> par(mfrow=c(1, 2))# Découpe la fenêtre graphique en 1 ligne et 2 colonnes
```

```
> s.label(monaafc$co, clab=0.7, label=colnames(flo[, 2:144]),
boxes=FALSE, sub="Espèces")# Affiche la position des espèces dans le plan factoriel
F1/F2
```

```
> s.label(monaafc$li, clab=0.7, label=flo$Releve, boxes=FALSE,
sub="Relevés")# Affiche la position des relevés dans le plan factoriel F1/F2
```

4.2 Interprétation de l'AFC

4.2.1 Dépouillement de l'axe 1

4.2.1.1 Analyse des espèces : interprétation des gradients écologiques sous-jacents

- Déterminez les contributions des espèces à la construction de l'axe F1.

- Pour cela, utilisez la fonction « `inertia.dudi()` » :

```
> inertia.dudi(monaafc, col.inertia=T)$col.abs[, 1]# Affiche les valeurs
absolues des contributions des espèces le long de l'axe F1 de votre AFC
```

```
> plot(monaafc$co[, 1], monaafc$co[, 2], type="n", xlab="Axe
F1", ylab="Axe F2")# Affiche un graphique vide prêt à accueillir vos valeurs
```

```
> text(monaafc$co[, 1], monaafc$co[, 2], lab=colnames(flo[, 2:144]),
cex=(inertia.dudi(monaafc, col.inertia=T)$col.abs[,
1]/max(inertia.dudi(monaafc, col.inertia=T)$col.abs[, 1])+0.1))#
Affiche les espèces avec une taille de texte proportionnelle à leurs contributions sur l'axe F1
```

◆ *Quelles espèces contribuent le plus à l'axe F1 ?*

- Observez l'autécologie de ces espèces dans une Flore pour vous aider à interpréter la signification de l'axe F1.

◆ *Quelles sont les caractéristiques communes des espèces situées au pôle négatif, qui s'opposent aux caractéristiques communes des espèces situées au pôle positif de l'axe F1 ?*

- Quelques lignes de code très utiles pour visualiser le nom complet des espèces le long de l'axe F1 :

```

> esp <- read.table("espece.txt", header=TRUE, sep="\t") # Créer un objet
« esp » contenant le nom complet des espèces

> sco.distrib(score=monafc$li[, 1], df=flo[, 2:144],
labels=esp[match(colnames(flo[, 2:144]), esp$Radical), "Espece"],
clab=0.5) # Range les espèces le long de l'axe F1

> par(mfrow=c(1,2)) # Découpe la fenêtre graphique en 1 ligne et 2 colonnes

> sco.label(monafc$co[, 1], lab=esp[match(colnames(flo[, 2:144]),
esp$Radical), "Espece"], reverse=TRUE, boxes=FALSE,
horizontal=FALSE, pos.lab=0.3, clab=0.5, lim=c(-4, 4)) # Range les
espèces le long de l'axe F1

> sco.label(monafc$li[, 1], horizontal = FALSE, boxes=FALSE,
lab=flo$Releve, clab=0.5, lim = c(-4, 4)) # Range les relevés le long de l'axe F1

```

4.2.1.2 Analyse des relevés : confirmation de l'interprétation par des variables mesurées sur le terrain

- Utilisez les variables mesurées sur le terrain et disponible dans le jeu de données « eco » pour confirmer votre interprétation de la signification écologique de l'axe F1.

- Pour cela, utilisez la fonction « **s.class()** » pour les variables qualitatives (type factor) et la fonction « **s.value()** » pour les variables quantitatives (type double) :

```

> s.class(monafc$li, as.factor(eco$Drainage), label=c("xero",
"meso", "frais", "hydro")) # Affiche les relevés dans le plan factoriel F1/F2 et leurs
appartenances à une ellipse représentant chacune une modalité de la variable drainage

> s.value(monafc$li, eco$Alt, method="greylevel", csize=0.2) # Affiche
les relevés dans le plan factoriel F1/F2 avec des niveaux de gris fonction de l'altitude des relevés
(i.e., couleurs claires pour les basses altitudes et couleurs foncées pour les hautes altitudes)

> plot(eco$Alt, monafc$li[, 1], pch=20) # Affiche les coordonnées factorielles des
relevés le long de l'axe F1 (Y) en fonction de l'altitude (X)

> summary(lm(monafc$li[, 1]~eco$Alt)) # Edite les résultats d'un modèle linéaire
simple pour expliquer la position des relevés le long de l'axe F1 en fonction de l'altitude

> pred <- predict(lm(monafc$li[, 1]~eco$Alt)) # Stock les prédictions de votre
modèle linéaire dans un vecteur objet nommé « pred »

> lines(eco$Alt[order(eco$Alt)], pred[order(eco$Alt)], lty=2,
lwd=2) # Ajoute une droite de tendance au graphique précédent

```

♦ **Quelles sont les variables qui expliquent le mieux l'axe F1, pour lesquelles on maximise le R^2 ?**

4.2.2 Dépouillement de l'axe 2

♦ **Que dire de l'axe F2 ?**