



HAL
open science

Étude de la liaison entre banque de diaspores et végétation dans les milieux aquatiques

Thierry Wilhelm

► **To cite this version:**

Thierry Wilhelm. Étude de la liaison entre banque de diaspores et végétation dans les milieux aquatiques. Méthodologie [stat.ME]. 2010. dumas-00520271

HAL Id: dumas-00520271

<https://dumas.ccsd.cnrs.fr/dumas-00520271>

Submitted on 22 Sep 2010

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.

Rapport de stage

2010



Etude de la liaison entre banque de diaspores et végétation dans les milieux aquatiques

par Thierry WILHELM

Maître de stage : Dr Isabelle Combroux-Lazăr
Laboratoire d'Hydrologie et de Géochimie de Strasbourg

Encadrant UdS : Jean Luc Dortet

Remerciements

Je tiens tout d'abord à remercier le docteur Isabelle Combroux-Lazar de m'avoir accueilli au sein de son équipe et m'avoir fait confiance pour ce projet. Je tiens tout particulièrement à la remercier pour son aide, son écoute et sa disponibilité, et pour tous les conseils et explications qu'elle a pu me fournir durant ce stage.

Je tiens ensuite à remercier Mr Nobelis Photis et ses précieux conseils, sans qui la réalisation de ce projet n'aurait pu avoir lieu.

Je désire également remercier les différents Professeurs et Maîtres de conférence du laboratoire, Mme Trémolières Michèle, Mme Haan-Archipoff Gisèle et Mr Laurent Pascal, pour leur accueil chaleureux au sein du laboratoire.

Je tiens encore à remercier Mme Bertin Claudine de l'herbier et Mr De Ruffray Patrice de l'IBMP (Institut de biologie moléculaire des plantes) pour conseil dans le domaine de la botanique.

Enfin, un énorme remerciement à tous les stagiaires, entre autres Mr Meyer Albin (doctorant), Mlle Caridi Claudia et Mlle Koenig Sarah, pour leur gentillesse et leur bonne humeur quotidienne.

Table des matières

Remerciements	1
1. Introduction	5
1.1. L'organisme d'accueil : le LHyGeS	5
1.2. Objectifs du stage.....	6
1.3. Préambule.....	6
2. Revue mathématique	7
2.1. Les indices de similarité	7
2.1.1. Indices de similarité binaires asymétriques.....	7
2.1.2. Indices de similarité quantitatifs	9
2.1.3. Les coefficients de corrélation des rangs.....	11
2.2. L'analyse factorielle des correspondances (AFC)	12
2.2.1. Généralités	12
2.2.2. Analyse des profils lignes et colonnes	13
2.2.3. Distance du χ^2	14
2.2.4. Valeurs propres et inertie	14
2.2.5. Contributions et distances	15
2.2.6. AFC inter et intra classes	16
2.3. Classification ascendante hiérarchique (CAH).....	18
2.3.1. Indice de distance	18
2.3.2. Indice d'agrégation	19
2.3.3. Dendrogrammes	20
2.3.4. Qualité de partition	20
2.4. Le bootstrap (non paramétrique).....	21
2.4.1. Généralités	21
2.4.2. Principe théorique du bootstrap	21
2.4.3. Intervalles de confiance	22
2.4.4. Optimisation du bootstrap	23
3. Contextes d'application	24
3.1. Banque de diaspores et végétation	24
3.1.1. Banque de diaspores	24
3.1.2. La végétation.....	25
3.2. Anciens bras morts reconnectés au Rhin	26
3.2.1. Contexte historique :.....	26
3.2.2. Les sites	28
3.3. Bassin d'orage du Waldweg à Rouffach.....	31
3.3.1. Schéma et photographie du site du bassin d'orage	31
3.3.2. Généralités et description du bassin d'orage.....	32
3.4. Quelques exemples d'espèces.....	33
4. La réponse par la réalisation du programme	34

4.1. Les données	34
4.2. L'AFC et la CAH.....	34
4.3. Les indices	34
4.4. Le bootstrap	34
5. Discussions	35
5.1. L'analyse factorielle des correspondances	35
5.1.1. Les valeurs propres	35
5.1.2. Distances aux carrés des individus aux axes.....	37
5.1.3. Contribution des axes aux individus	39
5.1.4. Premier plan factoriel.....	41
5.1.5. AFC inter-stations.....	42
5.1.6. AFC intra-stations et intra-dates.....	46
5.2. La classification ascendante hiérarchique.....	49
5.2.1. Classification ascendante hiérarchique de la banque des stations	49
5.2.2. Classification ascendante hiérarchique de la banque des sites	50
5.2.3. Classification ascendante hiérarchique de la végétation des stations	51
5.3. Les indices et intervalles de confiance bootstrap associés	52
6. Adaptation du programme.....	55
7. Conclusions	56
7.1. Conclusions.....	56
7.2. Conclusions personnelles	56
ANNEXES.....	57
I. Les données des bras reconnectés au Rhin	57
I.1. La banque de graines	57
I.2. La végétation	62
II. Les analyses factorielles des correspondances	75
II.1. Premier plan factoriel de l'AFC de la banque de graines par stations	75
II.2. Deuxième plan factoriel de l'AFC de la banque de graines par stations	76
III. Indices de Ruzicka et intervalles de confiance bootstrap	77
IV. Bassin d'orage.....	78
IV.1. Banque de graines.....	78
IV.2. Végétation.....	80
IV.3. AFC de la banque de graines du bassin d'orage	83
IV.4. Classification ascendante hiérarchique	85
IV.4. Indices et intervalles de confiance bootstrap	86
V. Le script du programme	87
Mode d'emploi du programme	108
Liste des figures et tableaux	119
Bibliographie	120

1. Introduction

1.1. L'organisme d'accueil : le LHyGeS

Le *Laboratoire d'Hydrologie et de Géochimie de Strasbourg* (LHyGeS), créé le 1er janvier 2009, est une *Unité Mixte de Recherche* (UMR 7517) rattachée au *Centre National de la Recherche Scientifique* (CNRS) et à l'*Université de Strasbourg* (UdS). Une partie de ses membres sont hébergés par l'*institut de botanique*, situé 28 rue Goethe à Strasbourg, qui surplombe le jardin botanique et l'Observatoire de Strasbourg.

Le laboratoire résulte du regroupement de trois entités : le *Centre d'Ecologie Végétale et d'Hydrologie* (CEVH, créé en 2002 après la dissolution du CEREG en décembre 2001), les équipes "Géoparticules" - "Géochimie isotopique" et "Interactions Fluides-Roches" du *Centre de Géochimie de la Surface* (CGS) - et l'équipe "Hydrologie et Transferts en Milieux Poreux" de l'*Institut de Mécanique des Fluides et des Solides* (IMFS).

Ces trois groupes ont une démarche scientifique identique fondée sur une interprétation mécaniste des processus qui s'appuie sur des observations de laboratoire ou de terrain, afin d'élaborer des modèles de fonctionnement et d'évolution des hydro-systèmes à différentes échelles de temps et d'espace.

L'objectif général de l'UMR est de produire des connaissances, des théories et des méthodes permettant de caractériser, de comprendre et de quantifier le fonctionnement des hydro-systèmes continentaux et leurs évolutions, perturbées ou non par les actions anthropiques.

Le potentiel scientifique de cette unité dans les domaines des Sciences de la Terre et de l'Environnement porte sur trois orientations de recherche : bio-réactivité de la phase solide, géochimie isotopique et chimie de l'environnement, et processus élémentaires et modélisation.

Le LHyGeS s'occupe également de la gestion de l'*Observatoire Hydrogéochimique de l'Environnement* (OHGE) et participe activement au *Réseau Alsace de Laboratoires en Ingénierie et Sciences pour l'Environnement* (REALISE). (Gayoso, 2009)

1.2. Objectifs du stage

L'objectif principal de ce stage est l'écriture d'un programme qui permettrait d'établir la liaison entre deux compartiments biologiques sur différents sites. Cette liaison est basée sur le calcul d'indices de similarité et d'intervalles de confiance bootstrap autour de ces indices, après que les différents sites et stations aient été comparés entre eux à l'aide d'analyses factorielles de correspondances et de classifications ascendantes hiérarchiques. Les compartiments comparés durant de ce stage sont la banque de diaspores, c'est-à-dire une collection de tout élément permettant d'accomplir la dissémination d'espèces végétales (par reproduction ou par multiplication asexuée), et la végétation observée, c'est-à-dire l'ensemble des végétaux présents sur les sites (dans l'eau ou sur les berges), mais le programme a été écrit pour pouvoir être adapté à la comparaison d'autres compartiment. Le logiciel R ("R Development Core Team (2009) - R: A language and environment for statistical computing") a été utilisé lors de ce stage pour toute la partie statistique.

1.3. Préambule

L'année 2010 a été déclarée *année internationale de la biodiversité* par les Nations-Unies. Source d'innombrables richesses pour l'homme telles que l'agriculture, la recherche pharmaceutique ou encore l'énergie, la biodiversité est une des composantes essentielles de notre culture qui garantit la diversité des écosystèmes et des espèces végétales et animales. Mais aujourd'hui, cette biodiversité va mal : extinction de certaines espèces, diminution des ressources naturelles, ..., la prospérité de la nature s'affaiblit et laisse place à de nombreuses inquiétudes.

L'étude de la biodiversité nécessitant l'analyse, souvent complexe, de nombreux facteurs, c'est tout naturellement que les mathématiques en général, les statistiques en particulier, sont devenues incontournables pour toute étude ou recherche dans le domaine de l'écologie, et ce, grâce à l'essor de la bioinformatique et de la bio-statistique. L'objectif à terme étant d'élaborer de nouveaux modèles sur l'évolution probable des écosystèmes, les statistiques fournissent une aide concrète aux chercheurs, aménageurs et autres décideurs impliqués dans la gestion des espaces naturels.

2. Revue mathématique

2.1. Les indices de similarité

2.1.1. Indices de similarité binaires asymétriques

Le calcul d'un indice de similitude binaire permet de quantifier le degré d'association de deux espèces, ou encore le niveau de similitude entre deux sites, pour des données binaires (de type présence / absence). Il est calculé en comptant les occurrences du couple d'espèces ou de sites considérés. (Pielou, 1984 ; Borcard, 2009)

Nombre d'espèces		Compartiment 2		
		Présence	Absence	Total
Compartiment 1	Présence	a	b	a+b
	Absence	c	d	c+d
	Total	a+c	b+d	a+b+c+d

Tableau 1 : occurrences pour données binaires

- a représente le nombre d'espèces présentes sur les deux sites comparés (co-occurrence)
- b représente le nombre d'espèces présentes sur le premier site uniquement
- c représente le nombre d'espèces présentes sur le deuxième site uniquement
- d représente le nombre d'espèces absentes des deux sites (co-absence, qui est écologiquement non significative)

Le facteur d (co-absence) introduit une dépendance à la taille du corpus, et il est quasiment impossible d'éviter que soient privilégiés soit les espèces très rares, soit les espèces les plus fréquentes. Benzécri conclut à la meilleure représentation par des indices qui ne font pas intervenir d (Benzécri & al. 1973a, §C.2.3) : les indices de similarité asymétriques.

On élimine donc d'emblée les indices de similarité symétriques (qui prennent d en compte), comme les indices de Sokal et Michener (1958) et de Russel et Rao (1940), définis par :

$$SM = \frac{a+d}{a+b+c+d} \qquad RR = \frac{a}{a+b+c+d}$$

Une famille d'indices de similarité binaires classiques asymétrique (qui ne font pas intervenir d) est constituée d'indices de la forme suivante (pour α , β et γ rationnels positifs) :

$$\frac{\alpha a}{\beta a + \gamma(b+c)}$$

Prendre β non nul permet d'avoir un indice borné, ce qui nous intéresse ; on écarte donc par exemple l'indice binaire de Kulczynski (1927), qui prend $\alpha = \gamma = 1$ et $\beta = 0$. Il reste alors diverses propositions qui prennent $\alpha = \beta (= 1$ par exemple) pour avoir un indice borné à valeur entre 0 (les sites sont complètement différents) et 1 (les sites sont strictement identiques), ce qui est le cas pour tous les indices suivants. Reste γ .

2.1.1.a. L'indice de Jaccard

γ est pris égal à 1 pour l'indice de Jaccard (1908) ou indice de communauté, qui exprime une proportion globale d'accords :

$$J = \frac{a}{a + b + c}$$

La formule de Jaccard est la même que celle de Tanimoto, elle peut donc être désignée sous cet autre nom.

2.1.1.b. L'indice de Dice

γ est pris égal à $\frac{1}{2}$ pour l'indice de Dice (1945), qui exprime une densité moyenne d'accords :

$$S = \frac{2a}{2a + b + c}$$

L'indice de Dice est encore appelé indice de Czekanowski (1913), ou indice de Sørensen (1948).

Il existe de nombreuses variantes de l'indice de Dice, par exemple en prenant $\gamma = \frac{1}{3}$ ou $\gamma = \frac{1}{4}$.

2.1.1.c. Autres indices binaires

D'autres familles d'indices de similarité binaires asymétriques existent, mais ces indices sont très peu utilisés en écologie. On peut citer, par exemple :

l'indice de Ochiai (1957):

$$O = \frac{a}{\sqrt{(a+b)(a+c)}}$$

l'indice de distance :

$$D = \frac{b+c}{a+b+c}$$

2.1.2. Indices de similarité quantitatifs

Ces indices, eux aussi, varient entre 0 (les compartiments sont totalement différents) et 1 (les compartiments sont identiques).

2.1.2.a. L'indice de Ruzicka

$$R = \frac{\sum_{i=1}^s \min(x_{ij}, x_{ik})}{\sum_{i=1}^s \max(x_{ij}, x_{ik})}$$

avec s le nombre d'espèces, j le premier site comparé et k le deuxième.

2.1.2.b. L'indice de Steinhaus

$$St = \frac{2W}{A + B}$$

avec A et B la somme des abondances pour les deux sites et W la somme des minimums d'abondance (c'est-à-dire $W = \sum_{i=1}^s \min(x_{ij}, x_{ik})$ avec s le nombre d'espèces, j le premier site et k le deuxième).

2.1.2.c. L'indice quantitatif de Kulczynski

$$K = \frac{W}{2A} + \frac{W}{2B}$$

avec A et B la somme des abondances pour les deux sites et W la somme des minimums d'abondance.

2.1.2.d. L'indice de Bray-Curtis (ou Percentage Similarity index)

$$BC = \frac{2 \sum_{i=1}^s \min(x_{ij}, x_{ik})}{\sum_{i=1}^s (x_{ij} + x_{ik})}$$

avec s le nombre d'espèces, j le premier site comparé et k le deuxième.

2.1.2.e. L'indice de Baroni-Urbani & Buser

$$BUB = \frac{\sum_{i=1}^s \min(x_{ij}, x_{ik}) + \sqrt{\left(\sum_{i=1}^s \min(x_{ij}, x_{ik})\right) \left(\sum_{i=1}^s \max_j(x_{ij}) - \max(x_{ij}, x_{ik})\right)}}{\sum_{i=1}^s \max(x_{ij}, x_{ik}) + \sqrt{\left(\sum_{i=1}^s \min(x_{ij}, x_{ik})\right) \left(\sum_{i=1}^s \max_j(x_{ij}) - \max(x_{ij}, x_{ik})\right)}}$$

avec s le nombre d'espèces, j le premier site comparé et k le deuxième.

2.1.2.f. L'indice de Jaccard modifié (ou similarity ratio)

$$SR = \frac{\sum_{i=1}^s x_{ij} x_{ik}}{\sum_{i=1}^s x_{ij}^2 + \sum_{i=1}^s x_{ik}^2 - \sum_{i=1}^s x_{ij} x_{ik}}$$

avec s le nombre d'espèces, j le premier site comparé et k le deuxième.

2.1.2.g. L'indice de Chord

$$C = 1 - \sqrt{1 - \frac{\sum_{i=1}^s x_{ij} x_{ik}}{\sqrt{\sum_{i=1}^s x_{ij}^2 \sum_{i=1}^s x_{ik}^2}}}}$$

avec s le nombre d'espèces, j le premier site comparé et k le deuxième.

2.1.2.h. Les autres indices quantitatifs

Il existe plusieurs dizaines d'indices de similarité quantitatifs, qui n'ont bien entendu pas tous été étudiés. On peut citer, par exemple : l'indice de Canberra, celui de Horn, celui de Moritsa, celui de Roberts, celui de Yule, la similarité du Chi-carré, le coefficient de corrélation, et ainsi de suite. (Boyce R.-L., 2001)

2.1.3. Les coefficients de corrélation des rangs

Les coefficients de corrélation des rangs sont des statistiques utilisées pour mesurer l'association entre deux quantités. Il s'agit d'une mesure de corrélation de rangs : on étudie la similitude des classements des données (et non les observations en elles-mêmes). Contrairement aux indices précédemment cités, qui varient entre 0 et 1, les coefficients de corrélation des rangs de Spearman (1904) et Kendall (1938), qui ne sont en fait qu'une version non paramétrique du coefficient de corrélation de Pearson, varient entre -1 (les deux classements sont inversés l'un par rapport à l'autre) et 1 (les classements sont identiques), en passant par 0 (les classements sont indépendants).

2.1.3.a. Le coefficient de corrélation ρ de Spearman

Le coefficient de corrélation des rangs de Spearman ou rho de Spearman, nommé d'après Charles Spearman (1863 – 1945), psychologue anglais connu pour son travail en statistique (analyse factorielle (1904) et corrélation), est une mesure non-paramétrique de la dépendance statistique entre deux variables. Il évalue dans quelle mesure la relation entre deux variables peut être décrite en utilisant une fonction monotone.

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

avec d_i la différence des rangs et n le nombre d'observations.

2.1.3.b. Le coefficient de corrélation τ de Kendall

Le coefficient de corrélation des rangs de Kendall ou tau de Kendall, tient son nom de Sir Maurice George Kendall (1907 – 1983), statisticien anglais.

$$\tau = \frac{2(n_c - n_d)}{n(n-1)}$$

avec n_c le nombre de paires concordantes (c'est-à-dire, pour $\{X_1, X_2\}$ et $\{Y_1, Y_2\}$, on a que $\text{signe}(X_2 - X_1) = \text{signe}(Y_2 - Y_1)$), n_d le nombre de paires discordantes (c'est-à-dire, pour $\{X_1, X_2\}$ et $\{Y_1, Y_2\}$, on a que $\text{signe}(X_2 - X_1) = -\text{signe}(Y_2 - Y_1)$), et n le nombre d'observations.

2.2. L'analyse factorielle des correspondances (AFC)

2.2.1. Généralités

L'analyse factorielle des correspondances (AFC ou CA pour « correspondence analysis » en anglais), mise au point par Jean-Paul Benzecri durant la période 1970 – 1990 (alors que la première formalisation mathématique de l'AFC date de 1935 et est due à Hirschfeld), est une méthode d'ordination particulière spécifiquement adaptée aux tableaux de contingence. Comme son nom l'indique, l'objectif est de mettre en correspondance les lignes avec les colonnes d'un tableau de fréquence. Cette analyse est très souvent employée pour résumer la structure d'un jeu de données de distribution d'espèces et mettre en relation les conditions écologiques des stations avec les espèces qui en sont caractéristiques.

L'analyse repose sur une pondération de chaque cellule du tableau par les totaux marginaux de la ligne d'une part et, par les totaux marginaux de la colonne correspondante d'autre part. C'est sur ces deux matrices pondérées que l'analyse recherche ensuite les axes principaux en veillant à conserver les distances euclidiennes entre les lignes d'une part et les colonnes d'autre part. Cette manière de procéder met rapidement en lumière les espèces qui ont le plus de relations avec des stations ou des groupes de stations. On peut visualiser facilement les relations entre les lignes et les colonnes en calculant pour chaque colonne le produit des probabilités conditionnelles calculées pour les lignes et pour les colonnes. (Carpentier, 2005 ; Périnel, 2009 ; Ramousse, Le Berre & Le Guelte, 1996).

L'AFC est particulièrement adaptée à l'étude de gradients écologiques car elle tend à produire des axes principaux qui expliquent au mieux les amplitudes écologiques des espèces en opposant le plus possible les optimums. (Chessel, Dufour, Dray, 2009 ; Dufrêne, 2003).

Soient deux variables X et Y, comportant respectivement p et q modalités. On a observé les valeurs de ces variables sur une population et on dispose d'un tableau de contingence K à I lignes et J colonnes donnant les effectifs conjoints c'est-à-dire les effectifs observés pour chaque combinaison d'une modalité i de X et d'une modalité j de Y.

Pour la suite, fixons les notations suivantes :

$$\begin{aligned}n_{ij} &: \text{effectif de la cellule (i,j)} & f_{ij} &: \text{fréquence de la cellule (i,j)} \\n_{i\bullet} = \sum_{j=1}^J n_{ij} &: \text{effectif total de la ligne i} & f_{i\bullet} = \frac{n_{i\bullet}}{n_{\bullet\bullet}} &: \text{fréquence de la ligne i} \\n_{\bullet j} = \sum_{i=1}^I n_{ij} &: \text{effectif total de la colonne j} & f_{\bullet j} = \frac{n_{\bullet j}}{n_{\bullet\bullet}} &: \text{fréquence de la colonne j} \\N = n_{\bullet\bullet} = \sum_{i=1}^I \sum_{j=1}^J n_{ij} &: \text{effectif total}\end{aligned}$$

2.2.2. Analyse des profils lignes et colonnes

		Variable 2			Σ
		1	j	J	
Variable 1	1				
	i				
	I				
	Σ				N

Tableau 2 : tableau de contingence

L'AFC peut être définie comme une double ACP (Analyse par Composantes Principales) : ACP du tableau des profils-lignes et ACP du tableau des profils-colonnes. À partir du tableau de contingence K (tableau 2), on passe au tableau des profils en ligne (respectivement en colonnes) (tableaux 2 et 3), c'est à dire des fréquences conditionnelles, obtenu en divisant chaque élément de la ligne i (respectivement j) par son total $n_{i\bullet}$ (respectivement $n_{\bullet j}$). Lignes et colonnes jouent un rôle symétrique. Le terme de profil fait référence à l'aspect visuel des diagrammes en bâtons fréquemment utilisés pour représenter les distributions, ici celles des fréquences conditionnelles. Les fréquences lignes (ou coordonnées des profils lignes) sont calculées par :

$$f_{ij}^l = \frac{n_{ij}}{n_{i\bullet}} = \frac{f_{ij}}{f_{i\bullet}}$$

		1	j	J	Σ
1					1
i	—	—	—		1
I					1

Tableau 3 : tableau des profils lignes

Les fréquences colonnes (ou coordonnées des profils colonnes) sont calculées par :

$$f_{ij}^c = \frac{n_{ij}}{n_{\bullet j}} = \frac{f_{ij}}{f_{\bullet j}}$$

		1	j	J
1				
i				
I				
Σ		1	1	1

Tableau 4 : tableau des profils colonnes

2.2.3. Distance du χ^2

La distance du χ^2 (Chi deux) est la distance utilisée pour l'analyse factorielle des correspondances. La distance du χ^2 entre la ligne i et la ligne i' est donnée par la formule :

$$d_{\chi^2}^2(i, i') = \sum_{j=1}^J \frac{(f_{ij}^l - f_{i'j}^l)^2}{f_{\bullet j}} = \sum_{j=1}^J \frac{N}{n_{\bullet j}} \left(\frac{n_{ij}}{n_{i\bullet}} - \frac{n_{i'j}}{n_{i'\bullet}} \right)^2$$

et la distance du χ^2 entre la colonne j et la colonne j' est donnée par :

$$d_{\chi^2}^2(j, j') = \sum_{i=1}^I \frac{(f_{ij}^c - f_{ij'}^c)^2}{f_{i\bullet}} = \sum_{i=1}^I \frac{N}{n_{i\bullet}} \left(\frac{n_{ij}}{n_{\bullet j}} - \frac{n_{ij'}}{n_{\bullet j'}} \right)^2$$

La distance du χ^2 est essentiellement utilisée pour deux raisons :

- contrairement à la distance euclidienne, la distance du χ^2 entre deux lignes ne dépend pas des poids respectifs des colonnes et vice-versa ;
- elle possède la propriété d'*équivalence distributionnelle* : si on regroupe deux modalités lignes, les distances entre les profils-colonnes ou entre les autres profils-lignes restent inchangées.

2.2.4. Valeurs propres et inertie

Le nombre de valeurs propres produites par la recherche des facteurs principaux est égal au minimum du nombre de lignes et du nombre de colonnes du tableau de contingence. Cependant, la première valeur propre est systématiquement égale à 1, et n'est pas mentionnée dans les résultats. Les autres valeurs propres sont des nombres positifs inférieurs à 1 et leur somme est égale au coefficient Phi-2 (ϕ^2) :

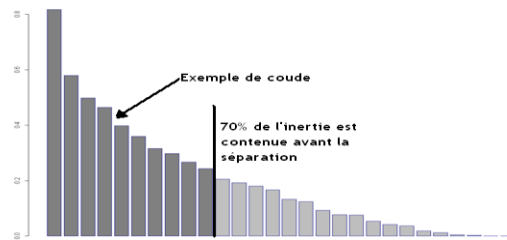
$$\phi^2 = \sum_{j=1}^J \frac{(f_{ij} - f_{i\bullet} \cdot f_{\bullet j})^2}{f_{i\bullet} \cdot f_{\bullet j}} = \frac{\chi^2}{N}$$

La somme des valeurs propres est toujours égale à l'inertie totale du nuage des individus ou du nuage des modalités. On caractérise ainsi chaque axe par le pourcentage d'inertie qu'il permet d'expliquer. La valeur de l'inertie totale est donc un indicateur de la dispersion du nuage et mesure la liaison entre deux variables.

On ne retient donc que les axes avec les plus fortes valeurs propres. Le choix des axes retenus est un peu délicat. On peut donner quelques règles :

- **Règle du coude** (Catell) : On observe souvent de fortes valeurs propres au départ, puis de faibles valeurs avec un décrochage dans le diagramme. On retient les axes avant le décrochage.

- **Règle de l'inertie minimale** : On sélectionne les premiers axes afin d'atteindre un pourcentage donné d'inertie expliquée (70% par exemple).
- **Règle du bon sens** : On analyse les plans et axes et on ne retient que ceux interprétables.



2.2.5. Contributions et distances

2.2.5.a. Distance au carré d'un individu à un axe

La contribution de l'axe (k) à l'individu (i) représente la qualité de la représentation de cet individu (i) par le sous-espace qui ajuste le nuage initial. Elle se mesure par le cosinus carré de l'angle formé par la droite reliant l'origine à l'individu (i) et l'axe factoriel (k) : si ce cosinus carré est proche de 1, l'individu (i) est bien représenté sur l'axe (k). On appelle **contribution relative du sous-espace** généré par les k premiers axes factoriels, la quantité :

$$QLT(i, k) = \sum_{q=1}^k \cos^2(i, q) = \sum_{q=1}^k \frac{\varphi_q^2}{d^2(i, G)}$$

où φ_k^2 et $d^2(i, G)$ désignent respectivement la coordonnée factorielle de l'axe (k) au carré et la distance entre l'individu i et le centre de gravité G du nuage au carré.

2.2.5.b. Contribution d'un individu à un axe

L'AFC présente une particularité du fait que les individus sont affectés de poids différents : un individu peut donc avoir plus ou moins d'influence dans le lien entre les variables ; on dispose pour mesurer cette influence d'un indice appelé "contribution relative d'un individu à un axe" donné par la relation suivante :

$$CTR_{axe(k)}(i) = \frac{m_i * c_{ik}^2}{\lambda_k}$$

avec : $m_i = \frac{n_i}{N}$ le poids de l'individu (i), λ_k l'inertie de l'axe (k)

c_{ik}^2 la coordonnée de l'individu (i) sur l'axe (k)

La somme de toutes les contributions vaut 1 et un profil- ligne (respectivement profil- colonne) est d'autant plus important dans la construction du k^{ème} axe factoriel que sa contribution est élevée.

2.2.6. AFC inter et intra classes

Soit N un tableau à r lignes (stations) et c colonnes (espèces) de terme général n_{ij} (nous devons donc utiliser la transposée de notre tableau initial).

L'analyse factorielle des correspondances ordinaire de N donne un triplet (F, G, D_λ) où F est le tableau des coordonnées factorielles des lignes, G celui des coordonnées factorielles des colonnes, et D_λ la matrice diagonale des valeurs propres. Les stations, au nombre de r , sont rangées en b blocs (sites) comptant respectivement r_1, \dots, r_b relevés. Pour $1 \leq i \leq r$, on note $b(i)$ le bloc auquel appartient le relevé i ; l'assertion $b(i) = k$ signifie que le relevé i est placé dans le bloc k ($1 \leq k \leq b$).

(Chessel et Dolédec, 1989 ; Cazes, 2004 ; Lobry, 2009)

2.2.6.a. AFC inter-classes

L'addition de toutes les lignes associées au même bloc (site) définit un tableau N^+ qui compte b lignes (sites) et c colonnes (espèces) de terme général :

$$n_{kj}^+ = \sum_{i|b(i)=k} n_{ij}$$

dont les sommes marginales par colonnes et par lignes sont :

$$n_{\bullet j}^+ = n_{\bullet j}$$

$$n_{k\bullet}^+ = \sum_{j=1}^c n_{kj}^+$$

la somme totale restant inchangée, soit : $n_{\bullet\bullet}^+ = n_{\bullet\bullet}$

L'AFC inter-groupes est par définition l'AFC ordinaire de N^+ accompagnée de la projection en individus supplémentaires de toutes les lignes de N . L'AFC de N^+ donne un triplet (F^+, G^+, D_λ^+) et on note F_s^+ le tableau associé des coordonnées des individus supplémentaires.

L'AFC inter-classes a pour intérêt de prendre en compte l'effet station et l'effet date.

2.2.6.b. AFC intra-classes

A l'analyse de N^+ on peut associer l'analyse de N^- , qui possède r lignes et c colonnes, de terme général :

$$n_{ij}^- = n_{ij} - \frac{n_{b(i)j}^+ n_{i\bullet}}{n_{b(i)\bullet}^+} + \frac{n_{i\bullet} n_{\bullet j}}{n_{\bullet\bullet}}$$

L'AFC intra-groupes est par définition l'AFC ordinaire de N^- qui donne un triplet (F^-, G^-, D_λ^-) . Les marges du tableau N^- sont :

$$n_{i\bullet}^- = \sum_{j=1}^c n_{ij}^- = n_{i\bullet} \quad \text{et} \quad n_{\bullet j}^- = \sum_{i=1}^r n_{ij}^- = n_{\bullet j}$$

la somme totale restant inchangée, soit : $n_{\bullet\bullet}^+ = n_{\bullet\bullet}$

L'AFC intra-classes a pour intérêt d'éliminer l'effet station et l'effet date.

2.2.6.c. Inertie

Si I_t est l'inertie totale de l'AFC de N , I_t^+ celle de N^+ et I_t^- celle de N^- , alors :

$$I_t = I_t^+ + I_t^-$$

ceci en prenant la précaution d'enlever de l'inertie totale la première valeur propre λ_0 qui vaut toujours l'unité. De plus, $\lambda_k \geq \lambda_k^+$ et $\lambda_k \geq \lambda_k^-$ pour tout k .

Les coordonnées factorielles des lignes ou stations (respectivement colonnes ou espèces) dans F , F^+ et F^- (respectivement G , G^+ et G^-) sont centrées (pour la pondération marginale) et de variances respectives λ_k , λ_k^+ et λ_k^- .

L'analyse de N maximise la dispersion totale, l'analyse de N^+ maximise la dispersion des centres de gravité des groupes et l'analyse de N^- maximise la moyenne de la dispersion à l'intérieur de chaque groupe.

2.3. Classification ascendante hiérarchique (CAH)

La classification ascendante hiérarchique (CAH) consiste à agréger de proche en proche des individus entre eux, puis des classes d'individus entre elles, jusqu'à obtenir une classe englobant l'ensemble des individus (par agglomération). Cette classification a pour intérêt de tenir compte des proximités établies sur plusieurs axes factoriels et de fournir une description synthétique des classes, notamment à l'aide de dendrogrammes. En pratique, l'analyse factorielle et la classification ascendante hiérarchique se complètent. (Legendre & Legendre, 1998 ; Chessel, Dufour et Lobry, 2006 ; Labatte, 2008)

2.3.1. Indice de distance

De nombreuses mesures de la distance entre individus ont été proposées. Le choix d'une (ou plusieurs) d'entre elles dépend des données étudiées.

- **Distance Euclidienne** : c'est le type de distance le plus couramment utilisé. Il s'agit simplement d'une distance géométrique dans un espace multidimensionnel.

$$d(I_i, I_j) = \sqrt{\sum_k (x_{ik} - x_{jk})^2}$$

- **Distance Euclidienne au carré** : on peut élever la distance euclidienne standard au carré afin de "surpondérer" les objets atypiques (éloignés).

$$d(I_i, I_j) = \sum_k (x_{ik} - x_{jk})^2$$

- **Distance du City-block ou de Manhattan** :

$$d(I_i, I_j) = \sum_k |x_{ik} - x_{jk}|$$

- **Distance de Tchebychev** :

$$d(I_i, I_j) = \max_k |x_{ik} - x_{jk}|$$

- **Distance à la puissance** :

$$d(I_i, I_j) = \left(\sum_k |x_{ik} - x_{jk}|^p \right)^{1/r}$$

2.3.2. Indice d'agrégation

L'application de la méthode suppose également que nous faisons le choix d'une "distance" entre classes. Là encore, de nombreuses solutions existent. Il faut noter que ces solutions permettent toutes de calculer la distance entre deux classes quelconques sans avoir à recalculer celles qui existent entre les individus composant chaque classe.

- **Distance entre centres de gravité :**

$$\delta_{CG}(C_1, C_2) = d(G_1, G_2)$$

avec G_1 et G_2 les centres de gravité des classes C_1 et C_2 .

- **Saut minimum** ou "**single linkage**" (**indice du lien minimum**). Les distances entre classes sont déterminées par la plus petite distance existant entre deux objets de classes différentes (c'est-à-dire les "voisins les plus proches") :

$$\delta_{\min}(C_1, C_2) = \min_{I \in C_1} \min_{J \in C_2} \{d(I, J)\}$$

- **Diamètre** ou "**complete linkage**" (**indice du lien maximum**). Dans cette méthode, les distances entre classes sont déterminées par la plus grande distance existant entre deux objets de classes différentes (c'est-à-dire les "voisins les plus éloignés") :

$$\delta_{\max}(C_1, C_2) = \max_{I \in C_1} \max_{J \in C_2} \{d(I, J)\}$$

- **Moyenne non pondérée des groupes associés.** Ici, la distance entre deux classes est calculée comme la moyenne des distances entre tous les objets pris dans l'une et l'autre des deux classes différentes :

$$\delta_{\text{moy}}(C_1, C_2) = \frac{1}{\text{card}(C_1) * \text{card}(C_2)} \sum_{I \in C_1, J \in C_2} d(I, J)$$

- **L'indice d'agrégation de Ward.** Cette méthode se distingue de toutes les autres en ce sens qu'elle utilise une analyse de la variance approchée afin d'évaluer les distances entre classes. En résumé, cette méthode tente de minimiser la Somme des Carrés de tous les couples de classes pouvant être formés à chaque étape. Les indices d'agrégation sont recalculés à chaque étape à l'aide de la règle suivante : si une classe M est obtenue en regroupant les classes K et L, sa distance à la classe J est donnée par

$$\delta_{\text{WARD}}(C_1, C_2) = \frac{m_1 m_2}{m_1 + m_2} d^2(G_1, G_2)$$

avec G_1 et G_2 les centres de gravité des classes C_1 et C_2 , m_1 et m_2 les poids des classes.

La méthode de Ward se justifie bien lorsque la "distance" entre les individus est le carré de la distance euclidienne. Choisir de regrouper les deux individus les plus proches revient alors à choisir la paire de points dont l'agrégation entraîne la diminution minimale de l'inertie du nuage. Le calcul des nouveaux indices entre la paire regroupée et les points restants revient alors à remplacer les deux points formant la paire par leur point moyen, affecté du poids 2.

2.3.3. Dendrogrammes

Un dendrogramme (du Grec *dendron* "arbre", *gramma* "dessiner") est un diagramme utilisé pour illustrer l'arrangement de groupes (*clusters* en anglais) générés par un regroupement hiérarchique, comme par exemple dans une classification hiérarchique ascendante.

La hauteur d'agrégation entre chaque pallier traduit le niveau de dissemblance entre les éléments agrégés : plus le pallier rassemblant deux individus ou groupes d'individus est haut, moins ceux-ci présentent de ressemblance.

2.3.4. Qualité de partition

Une partition est de bonne qualité si les classes sont homogènes et bien séparées, ou, en terme d'inertie, si l'inertie intra-classes est faible et l'inertie interclasses élevée.

Théorème de Huygens : l'inertie totale est égale à la somme des inerties intra et inter-classes, c'est-à-dire :

$$Inertie_{totale} = \sum_{I \in C_k} m_I d^2(I, G) = \sum_{k=1}^K \sum_{I \in C_k} m_I d^2(I, G_k) + \sum_{k=1}^K m_k d^2(G_k, G)$$

avec G le centre de gravité de l'ensemble des classes, G_k le centre de gravité de la classe C_k , et m_I le poids associé à l'individu I .

On dispose alors d'un indicateur de qualité d'une partition, défini par le ratio :

L'indice de Ward a pour propriété l'équivalence entre minimiser l'augmentation intra-classes et minimiser la perte de l'inertie interclasses, en d'autres mots, moins on a de classes (c'est-à-dire plus on agrège), plus l'inertie intra-classes augmente et plus l'inertie inter-classes diminue (l'inertie totale restant la même).

2.4. Le bootstrap (non paramétrique)

2.4.1. Généralités

Les techniques de bootstrap, introduites par Efron en 1979, sont des méthodes d'inférence statistique modernes, datant de la fin des années 70, et requérant des calculs informatiques intensifs. L'objectif est de connaître certaines indications sur une statistique : son estimation bien sûr, mais aussi la dispersion (variance ou écart type) dans le but d'établir des intervalles de confiance. Cette méthode est basée sur des simulations, comme les simulations de Monte-Carlo, à la différence près que le bootstrap ne nécessite pas d'information supplémentaire que celle disponible dans l'échantillon. En général, il est basé sur de « nouveaux échantillons » obtenus par tirage avec remise à partir de l'échantillon initial (on parle alors de technique de *rééchantillonnage* ou *resampling technic* en anglais). (Hinkley, 1988 ; Johns, 1988 ; Efron & Tibshirani, 1993 ; Davison & Hinkley, 1997)

L'aspect autocentré et itératif de la méthode a inspiré sa désignation anglaise : en effet, le bootstrap désigne le fait de « se hisser en tirant sur ses propres lacets » ou plus précisément sur ses « *bootstraps* », qui désignent en anglais les anneaux de cuir ou tissu cousus au rebord des bottes pour y passer les doigts afin de les enfiler plus facilement.

2.4.2. Principe théorique du bootstrap

Soient X une variable aléatoire de fonction de répartition F (inconnue), (X_1, \dots, X_n) un échantillon indépendant identiquement distribué comme X , θ un paramètre inconnu que l'on veut estimer et $T(X_1, \dots, X_n)$ une statistique utilisée pour estimer θ , satisfaisant :

$\theta = T(X_1, \dots, X_n)$. La technique de bootstrap permet de remédier à l'absence de connaissance sur la loi G_n de .

L'idée est d'approcher F la loi des X_i par la fonction de répartition empirique F_n donnée

$$\text{par : } F_n(t) = \frac{1}{n} \sum_{i=1}^n I(X_i < t)$$

Simuler des observations issues de la loi F_n revient à effectuer des tirages avec remise dans l'échantillon initial (X_1, \dots, X_n) . Ceci fournit l'algorithme de bootstrap non-paramétrique :

Pour b allant de 1 à B (fixé) :

- tirer un échantillon $(X_1^{*(b)}, \dots, X_n^{*(b)})$ avec remise dans (X_1, \dots, X_n)
- calculer $\theta^{*(b)} = T(X_1^{*(b)}, \dots, X_n^{*(b)})$

On a donc à disposition un échantillon $(\theta^{*\{1\}}, \dots, \theta^{*\{B\}})$ distribué comme θ^* que l'on peut utiliser pour construire la fonction de répartition empirique $G_{n,B}^*$ de :

$$G_{n,B}^*(t) = \frac{1}{B} \sum_{b=1}^B I(\theta^{*\{b\}} < t)$$

2.4.3. Intervalles de confiance

Supposons que l'on choisisse un niveau de confiance α de 5% et 1000 répliques ($B=1000$). L'échantillon $(\theta^{*\{1\}}, \dots, \theta^{*\{1000\}})$ est tout d'abord trié par ordre croissant. Le nouvel échantillon trié s'écrit alors $(\theta^{**\{1\}}, \dots, \theta^{**\{1000\}})$, avec $\theta^{**\{1\}} < \dots < \theta^{**\{1000\}}$.

La technique du bootstrap consiste à prendre 95% (ce qui correspond à $1 - \alpha$) des valeurs de l'échantillon dans le but d'établir un intervalle de confiance pour θ . La distribution de l'échantillon $(\theta^{*\{1\}}, \dots, \theta^{*\{1000\}})$ étant supposée symétrique (figure 7), les 95% sont pris entre la 25^{ème} (-) et la 975^{ème} (-) valeur de l'échantillon trié. On obtient alors un intervalle de confiance pour θ avec un risque α (ici 5%) de se tromper, intervalle donné par :

$$\theta \in \left[\theta^{**\{25\}}, \theta^{**\{975\}} \right]$$

avec $\theta^{**\{25\}}$ la 25^{ème} valeur de l'échantillon $(\theta^{**\{1\}}, \dots, \theta^{**\{1000\}})$ trié et $\theta^{**\{975\}}$ la 975^{ème} (pour 1000 répliques).

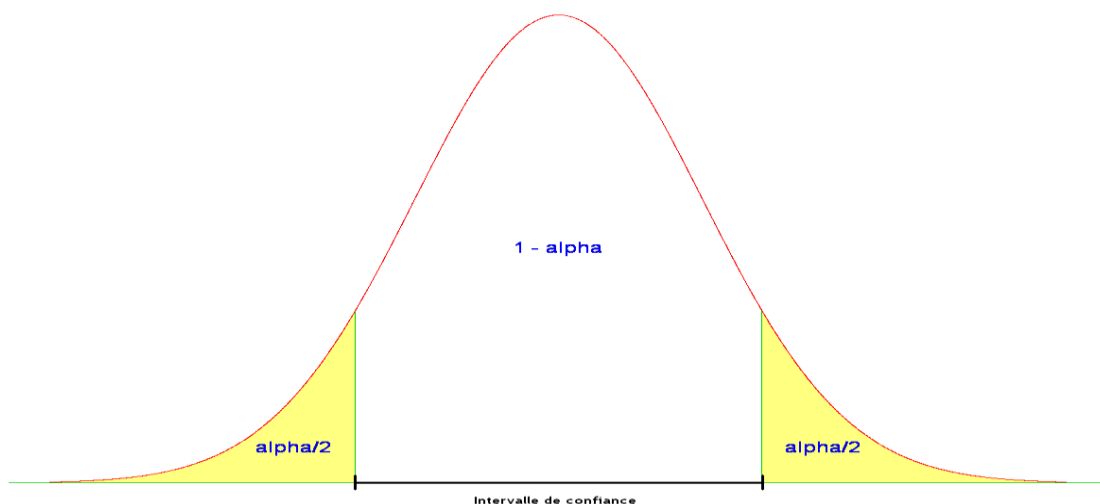


Figure 1 : intervalle de confiance pour une distribution symétrique

2.4.4. Optimisation du bootstrap

Le but de cette optimisation est de minimiser la largeur des intervalles de confiance. Pour cela, on se base sur le fait que la distribution de l'échantillon $(\theta^{*(1)}, \dots, \theta^{*(1000)})$ n'est pas forcément symétrique (figure 8).

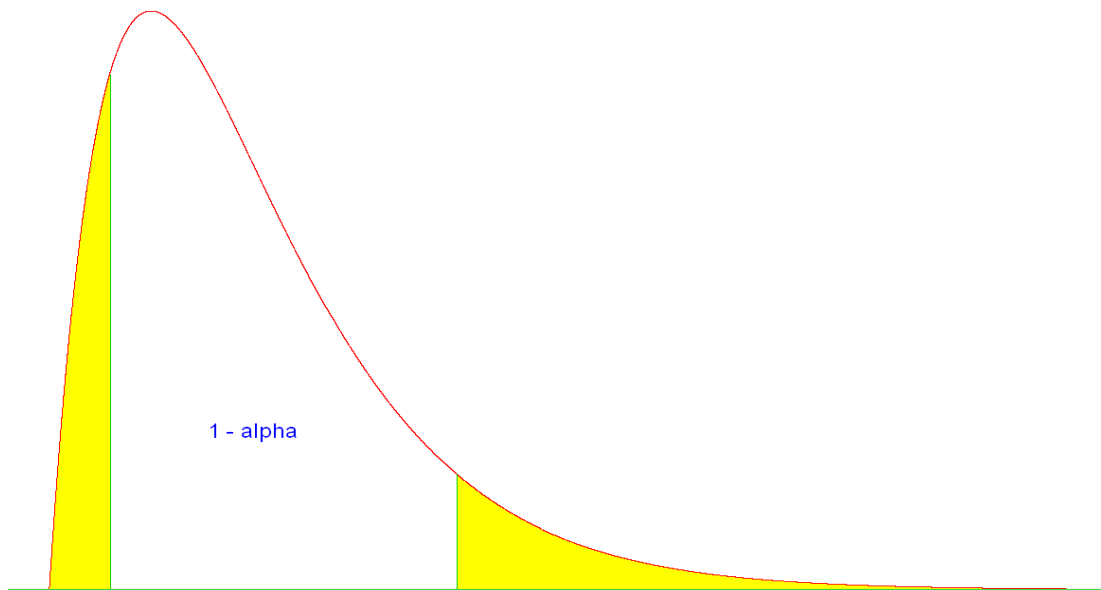


Figure 2 : intervalle de confiance pour une distribution asymétrique

Par exemple, si la 25^{ème} valeur de l'échantillon trié $(\theta^{*(1)}, \dots, \theta^{*(1000)})$ est nulle, alors toutes les valeurs la précédant sont elles aussi nulles (vu que l'échantillon est trié) : on prend alors comme bornes de l'intervalle de confiance la 1^{ère} valeur et la 951^{ème} valeur de l'échantillon trié. Le nouvel intervalle de confiance ainsi établi sera moins large que celui calculé dans la section précédente, tout en conservant 95% des valeurs (pour un niveau de confiance α de 5% et 1000 répliques).

C'est sur ce principe que se base l'optimisation : on compare à l'intervalle de confiance classique (c'est-à-dire avec comme bornes la 25^{ème} valeur et la 975^{ème} valeur de l'échantillon trié) les intervalles "décalés" à gauche (avec comme borne inférieure la 24^{ème} valeur de l'échantillon trié et la 974^{ème} valeur comme borne supérieure, et ainsi de suite) et à droite, et on garde le plus petit intervalle trouvé (qui, soit dit en passant, conserve 95% des valeurs de l'échantillon).

3. Contextes d'application

3.1. Banque de diaspores et végétation

Nous disposons de deux tableaux de contingence, un premier qui contient une estimation de la banque de diaspores par station (annexe I.1), et un second qui contient une estimation du recouvrement de la végétation (annexe I.2) sur chaque station à différentes dates.

3.1.1. Banque de diaspores

3.1.1.a. Définition

"Toutes les graines viables présentes sur ou dans le sol ou associées à la litière constituent la banque de graines du sol" (Simpson et al. 1989). Elle se forme par accumulation dans ou sur le sol de graines qui restent viables, le plus souvent en état de dormance (certains auteurs – Cavers, 1995 – incluent également dans la banque de graines les graines encore attachées à la plante mère). La banque de graines constitue un véritable compartiment de la végétation. Par extension, la banque de diaspores représente la banque qui contient les graines et les autres types de diaspores (rhizomes, bourgeons...) (Combroux, 2002).

Dans la suite de ce rapport, nous utiliserons indifféremment les termes de banque de diaspores ou banque de graines pour désigner le même compartiment.

3.1.1.b. Protocole de mesure de la banque de graines

Les diaspores dans la banque ont été quantifiés par la méthode de la "levée des plantules" (Thompson et Grime, 1979 ; Combroux et al., 2002). Cinq stations (uniformément réparties) ont été échantillonnées sur chaque site. L'échantillonnage a été effectué après la stratification hivernale et avant la germination printanière. Dix carottes de sédiments (de 4 cm de diamètre) ont été recueillies sur chaque station. Seule les couches supérieures (de 5 cm) ont été utilisées, car cette profondeur est généralement considérée comme la profondeur maximale en dessous de laquelle la germination ne peut avoir lieu (Bonis et Lepart, 1994). Les dix carottes ont été soigneusement mélangées et divisées en deux sous-échantillons, vu que deux types de germination ont été effectués :

- par inondation : des échantillons de sédiments ont été inondés en permanence dans moins de 10 cm d'eau,
- par exposition : des échantillons de sédiment ont été maintenus au point de saturation en eau.

L'émergence des semis a été suivie dans une serre (photopériode naturelle ; température de l'air maintenue au-dessus de 15 ° C). Les échantillons ont été répartis dans des bacs en plastique (épaisseur des sédiments inférieure à 2 cm). La couche d'enracinement a été augmentée par la diffusion de 2 cm de sable stérile. Un brassage d'air maintient une forte teneur en oxygène (90 à 100%) et des mouvements d'eau lents. L'émergence des semis a été suivie d'avril 2007 à Juin 2008. Les semis ont été régulièrement identifiés, attribués à un type de diaspore (graine ou oospore, bourgeon d'hiver, rhizome, fragment non spécialisé), et retirés des bacs pour éviter la concurrence et l'allélopathie (ensemble de plusieurs interactions biochimiques, directes ou indirectes, positives ou négatives, d'une plante sur une autre). Pour chaque station, la valeur maximale des deux traitements a été utilisée pour estimer l'abondance de chaque espèce et de chaque type de propagule dans la banque.

L'échantillonnage de base et la méthode d'émergence ne permettent habituellement pas de détecter les propagules qui sont profondément enfouies dans le sol (rhizomes principalement). En conséquence, l'abondance des rhizomes dans la banque de propagules a été estimée à mi-printemps (mai 2007) en comptant les pousses de germination à partir de méristèmes des rhizomes dans trois quadrats (30 cm x 30 cm) sur chaque station. Quand une espèce de rhizome a été détectée, soit par la méthode d'émergence, soit par l'échantillonnage par quadrats, l'estimation de l'abondance des rhizomes la plus importante a été utilisée (Combroux et al., 2001).

3.1.2. La végétation

La végétation aquatique a été étudiée au cours des printemps, étés et automnes 2007 et 2008 en utilisant le pourcentage de couverture, estimé à l'œil nu pour chaque espèce rencontrée.

3.2. Anciens bras morts reconnectés au Rhin

3.2.1. Contexte historique :

les aménagements du Rhin et les projets de restauration

Pour comprendre l'intérêt porté aux sites étudiés dans ce rapport, il est nécessaire de bien appréhender l'histoire du Rhin, de son aménagement à sa restauration (Meyer, 2009).

3.2.1.a. Le Rhin dans le fossé rhénan : généralités

Le Rhin est le plus grand fleuve d'Europe occidentale. Il se forme en Suisse par la réunion de deux rivières alpines, l'Hinterrhein et le Vorderrhein. Sa longueur est de 1325 km pour un bassin versant de 185 000 km². Le débit peut varier entre un minimum de 200 m³/s jusqu'à un maximum de 5700 m³/s, avec un module annuel d'environ 1000 m³/s à Strasbourg. Le Rhin est un fleuve avec un régime nival, c'est-à-dire avec un débit maximal, et donc des crues fréquentes, en juin/juillet.

3.2.1.b. Les aménagements du Rhin

Le Rhin a maintes fois été aménagé par l'Homme durant les deux derniers siècles :

- le projet Tulla (1842 - 1876) a consisté à endiguer le fleuve. Son cours a ainsi été raccourci de 30km ;
- le cours du Rhin a été régularisé entre 1906 et 1956 par la construction d'épis transversaux et de seuils de fond ;
- l'aménagement hydroélectrique français du fleuve entre 1928 et 1959 a été effectué grâce à la construction du Grand Canal d'Alsace entre Kembs et Vogelgrun ;
- à partir de 1959 des aménagements en feston et en ligne ont permis la construction de nouvelles centrales hydroélectriques franco-allemandes entre Vogelgrun et Lauterbourg.

Ces aménagements ont eu comme conséquence la déconnection de la plaine alluviale du cours principal. Cette suppression de la dynamique fluviale empêche les anciens bras d'échanger Flore et Faune avec le cours principal.

3.2.1.c. L'amélioration de l'état du Rhin et sa restauration

Le Programme Action Rhin (PAR, 1987-2000) et le programme Rhin 2020

Après la seconde Guerre Mondiale, les Pays-Bas ont eu beaucoup de difficultés à assurer la potabilité de leur eau à cause de la pollution importante des eaux du Rhin par des phénols rejetés par les industries pétrolières et du sel en provenance des mines de potasse d'Alsace. Les Pays-Bas ont alors pris l'initiative de créer, avec la collaboration de la France, de l'Allemagne, du Luxembourg et de la Suisse, une *Commission Internationale Pour le Rhin* (CIPR). Cette commission fut instituée à Bâle le 11 juillet 1950 avec pour but le traitement en commun des problèmes de protection des eaux du Rhin. Durant les trois décennies suivantes la CIPR permit la signature de plusieurs accords, protocoles et/ou conventions dans le but de protéger le Rhin de divers polluants comme les chlorures. Mais le choc de l'incendie du dépôt *Sandoz* (firme pharmaceutique suisse) à Bâle le 1er novembre 1986, et surtout la pollution des eaux du Rhin par des dérivés organiques de mercures et des esters phosphoriques, ainsi que l'hécatombe piscicole qui suivit, provoquèrent une remise en question générale. Les gouvernements des pays riverains du Rhin chargèrent alors la CIPR de mettre en place un plan de « sauvetage » du Rhin. Et c'est ainsi que le *Programme d'Action Rhin* (PAR) vit le jour en 1987. L'objectif principal du PAR était l'assainissement progressif du Rhin pour l'an 2000 au travers de la réimplantation dans le Rhin d'espèces piscicoles disparues tel le saumon, de la potabilité de l'eau du Rhin et d'une pollution minimale des sédiments fluviaux. Le bilan établi après la fin du PAR est positif et a mis en avant sa réussite globale, avec une amélioration de la qualité des eaux du Rhin, une baisse du nombre d'accidents industriels et un regain de "vitalité" dans le Rhin. Ce bilan a aussi été l'occasion de démontrer la nécessité d'un nouveau programme dans la continuité du PAR. La CIPR a ainsi lancé dès l'an 2000 un nouveau programme nommé *Rhin 2020* qui se concentre en priorité sur l'écologie, la protection de la nature, la prévention des crues et la protection des eaux souterraines, sans toutefois oublier la surveillance et l'amélioration de la qualité de l'eau (CIPR, 2003).

3.2.1.d. Un enjeu européen

Le programme Life-Nature Rhin Vivant et le site Natura 2000 Rhin-Ried-Bruche

La Région Alsace et ses partenaires, tel le *Conservatoire des Sites Alsaciens* (CSA), ont ainsi profité de l'existence du programme *Life-Nature* pour créer le programme *Life-Nature Rhin Vivant*, dans le but d'assurer la pérennité du site *Natura 2000* « Rhin-Ried-Bruche », qui, avec une surface de 34 434 ha (il s'étend sur les zones de la bande rhénane, du Ried Centre Alsace et du Bruch de l'Andlau), est le plus grand site *Natura 2000* d'Alsace. Sa grande taille a malheureusement pour conséquence de multiplier les menaces à contrer dans le but de maintenir son intégrité. Les pressions anthropiques agricoles, industrielles et urbaines sont des menaces non négligeables pour les habitats du site, mais le principal problème reste l'alimentation insuffisante en eau du Rhin des habitats alluviaux, les mettant en danger eux et leur fonctionnalité (CSA, 2005a).

3.2.2. Les sites

Six sites ont été étudiés. Ces sites sont localisés sur de petites rivières ayant comme caractéristique commune un apport d'eau permanent du Rhin ou du Vieux Rhin. Parmi ces six sites, quatre ont été récemment restaurés et reconnectés au Rhin, et deux, dits de référence, n'ont pas été déconnectés du Rhin par les aménagements (ils sont restés connectés au Vieux Rhin après la canalisation). Chaque site est considéré comme une zone "homogène", c'est-à-dire comme un compartiment fonctionnel qui bénéficie d'apports limités de nutriments des berges et de l'absence d'un ou plusieurs affluents. Ainsi les 100 premiers mètres après la connexion ont été étudiés pour chaque cours d'eau, lorsque cela était possible.

Chaque site a été divisé en 5 transects, chiffrés selon leur position sur le gradient amont-aval, c'est-à-dire de 1 à 5. Les transects sont séparés par une distance moyenne comprise entre 10 et 20 m, cette longueur variant selon la longueur totale du site étudié.

3.2.2.a. Plan d'expérience

Les six sites ont été visités à 3 reprises (au printemps, en été et en automne 2007) excepté pour les sites du *Schollengiessen* et du *Fahrgiessen* qui, suite à des crues trop importantes, n'ont pas été visités en été. Un échantillon de terre a été prélevé au printemps dans le but de quantifier la banque de graines, et, à chaque saison, la végétation présente sur chaque site a été soigneusement répertoriée grâce à la méthode précédemment citée.

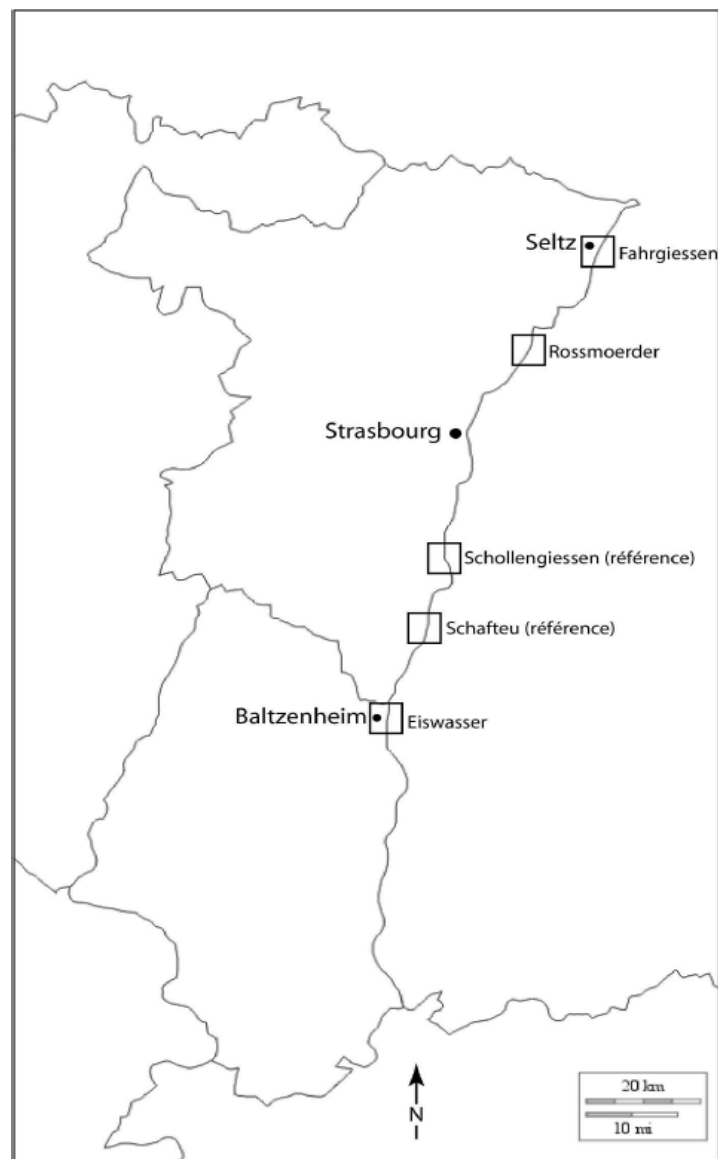


Figure 3 : situation des sites

3.2.2.b. Sites de référence

Le Schaftheu (Scha)

Le *Schaftheu* est situé sur l'île de Rhinau, qui est issue d'un aménagement en feston du Rhin à proximité de Rhinau en 1963. Cette grande île (10,5km de long et jusqu'à 600m de large) est isolée entre le Vieux Rhin et sa dérivation canalisée et elle s'étire entre la commune de Rhinau au nord et celle de Schoenau au sud. Une partie de l'île de Rhinau, 311ha, est classée en réserve naturelle nationale depuis 1991.

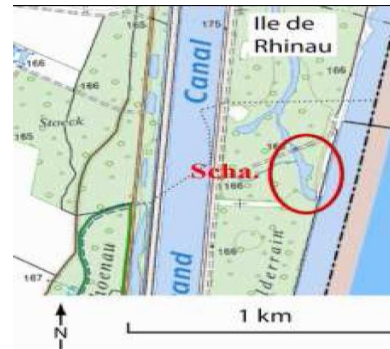


Figure 4 : emplacement du Schaftheu

Le Schollengiessen (Schol)

Le *Schollengiessen* est situé sur l'île de Gerstheim, qui provient de l'aménagement en feston du Rhin à proximité de Gerstheim en 1967. Elle est occupée par une forêt alluviale essentiellement au stade pionnier (saulaie-peupleraie) parcourue par un ancien chenal de tressage.

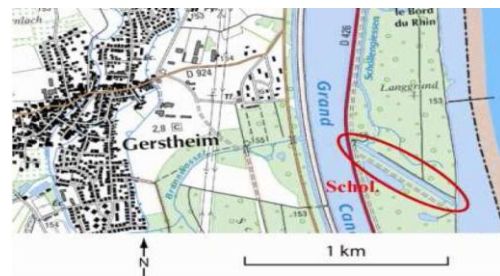


Figure 5 : emplacement du Schollengiessen

3.2.2.c. Sites restaurés

Le Fahrgiessen (Fahr)

Le *Fahrgiessen* est situé le long du massif forestier situé entre Seltz et le Rhin. Il a bénéficié en 2006 d'une remise en eau, d'un désenvasement et de coupes sélectives d'arbres sur la rive. Cette restauration a eu lieu dans le cadre du programme « Life Nature Rhin Vivant ». Ce tronçon rejoint l'ancien cours du Fahrgiessen situé dans la réserve naturelle de Seltz – Munchhausen créée en 1995.



Figure 6 : emplacement du Fahrgiessen

L'Eiswasser (EisM en amont et EisV en aval)

L'Eiswasser est une rivière traversant le massif forestier de Kunheim-Baltzenheim au sud de Marckolsheim. Cette rivière a été entièrement reconstituée par re-creusement d'un ancien chenal à sec et reconnecté en 2002 dans le cadre des programmes Interreg IIc et IRMA. Ce site est divisé en deux pseudo-réplicats, la zone restaurée étant longue (1,1km), ce qui permet une certaine indépendance des deux pseudo-réplicats, et car ces deux sites possèdent une couverture végétale correcte, avec un recouvrement supérieur à 20%. L'Eiswasser Amont est situé directement après la prise d'eau et l'Eiswasser Aval se situe plus en aval, à environ 1km du premier site.

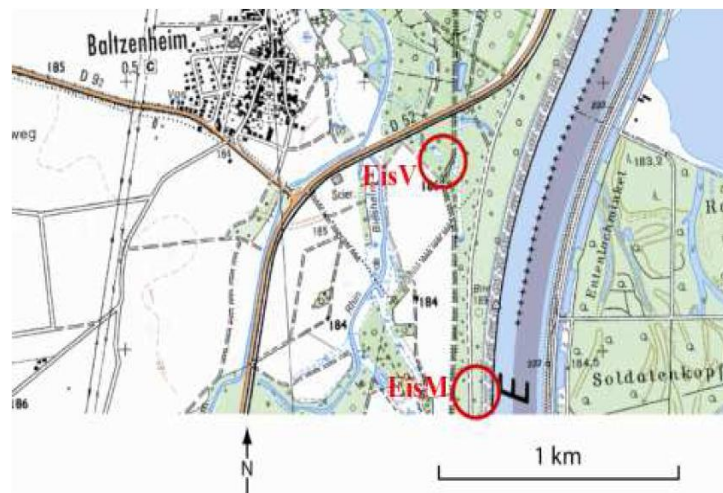


Figure 7 : emplacement de l'Eiswasser

Le Rossmoerder (Ross)

Le Rossmoerder se situe dans le massif alluvial d'Offendorf, au nord de Strasbourg. Ce cours d'eau a bénéficié d'une remise en eau et d'un désenvasement complétés par un entretien de la flore au bord des cours d'eau. Cette restauration a eu lieu en 2004 dans le cadre du programme Life Nature Rhin Vivant. L'ouvrage de prise d'eau a été conçu pour permettre un débit de 3 à 7 m³/s avec une maîtrise de ce débit grâce à une vanne pouvant réguler le débit et l'interrompre en cas, par exemple, de pollution de l'eau. Cette partie restaurée rejoint le cours principal du Rossmoerder, qui fait l'interface entre une réserve biologique à l'est et la réserve naturelle



Figure 8 : emplacement du Rossmoerder

3.3. Bassin d'orage du Waldweg à Rouffach

3.3.1. Schéma et photographie du site du bassin d'orage

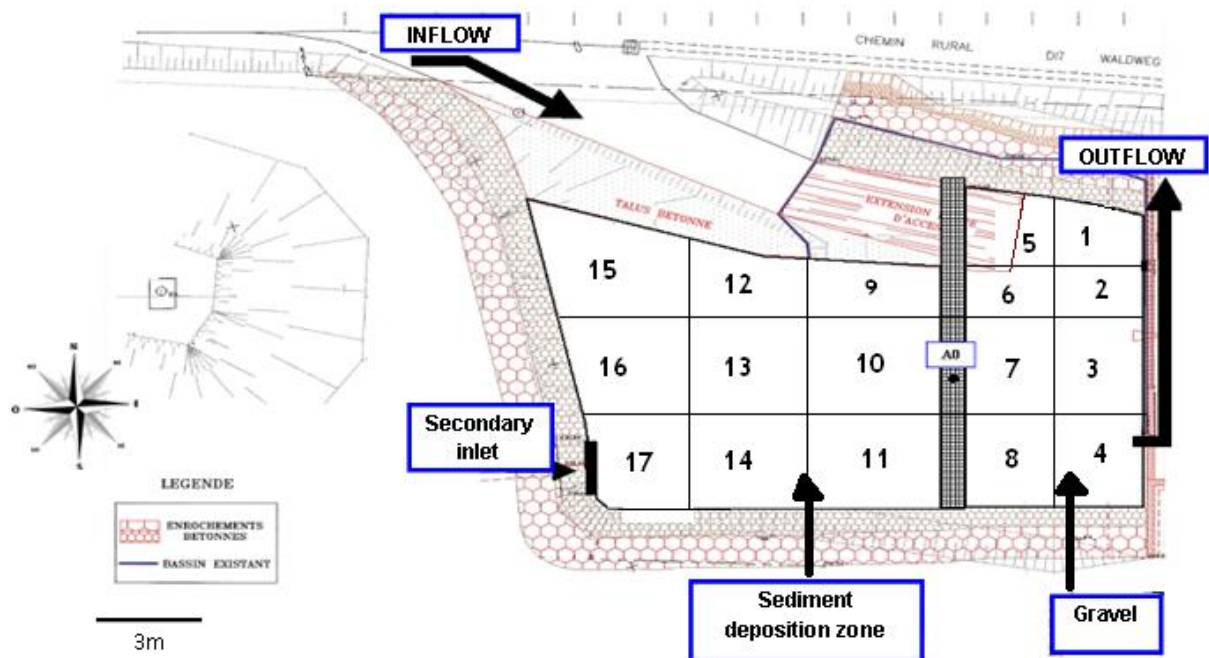


Figure 10 : bassin d'orage du Waldweg à Rouffach (photographie ArtWET)

3.3.2. Généralités et description du bassin d'orage

Suite à de violentes coulées de boues et inondations, de nombreux bassins de rétention des eaux pluviales ont été construits en amont des villages, à l'interface entre les agro-systèmes et les zones urbaines. Tout comme les zones humides naturelles, ces zones humides artificielles possèdent un potentiel de bio-remédiation intrinsèque. Fort de ces constatations et des exigences européennes concernant le maintien de la qualité de la ressource en eau, le projet interdisciplinaire ArtWET (*Artificial Wetland*) prévoit, entre autres, d'étudier l'opportunité de l'utilisation de ces ouvrages pour la remédiation des phytosanitaires. Il s'agit en particulier d'optimiser le fonctionnement du quadruplet eau / sédiment / végétaux / bactéries afin d'améliorer ce potentiel et ainsi d'utiliser les bassins d'orage non seulement à des fins hydrauliques mais aussi à des fins de remédiation des phytosanitaires.

Le bassin d'orage (dit du Waldweg) de Rouffach (Haut-Rhin, France), transformé en station de dégradation des résidus de pesticides par les plantes et les micro-organismes, sert actuellement de prototype. Ce bassin de rétention de 1 500 m³ recueille les eaux ruisselant sur 42 ha de coteaux essentiellement plantés de vignes. Lorsque la pluie tombe sur le bassin versant (appelé bassin versant amont du Hohrain) et que le sol n'est plus capable d'infiltrer l'eau (saturation du sol), l'eau ruisselle sur les parcelles et sur le réseau de chemins et de routes jusqu'à leur collecte dans le bassin d'orage en aval.

Des lits filtrants de sable et de gravier ont été installés dans le but de créer deux zones, l'une de décantation, l'autre de filtration. La végétation s'y est installée, avec son cortège de micro-organismes qui ont commencé leur travail de captation et de dégradation des pesticides. Cette végétation a été soigneusement répertoriée au printemps et en été 2010 grâce à la méthode précédemment citée, ainsi que la banque de diaspores présente sur le site, dans le but d'étudier la relation entre ces deux derniers. Le site du bassin d'orage de Rouffach a été divisé en 17 "stations" de tailles à peu près égales (figure 9) et numérotées de 1 à 17 pour étudier cette relation selon la proximité de l'entrée ou de la sortie d'eaux ou selon la zone (de décantation ou de filtration).

Une soixantaine de bassins ressemblant au bassin d'orage expérimental de Rouffach existent sur le territoire alsacien. La volonté de ce projet pluridisciplinaire est d'étendre la mise en place d'ouvrages de rétention et de remédiation à faible coût à l'aval sur d'autres bassins versants viticoles du piémont alsacien pour limiter le transfert des résidus de pesticides et ainsi limiter les besoins de traitement dans une perspective d'alimentation en eau potable.

(sources : artWET et L'Alsace.fr)

Plan d'expérience

Des carottes de terre ont été prélevées sur chaque station et la végétation a été répertoriée au printemps et en été 2009.

3.4. Quelques exemples d'espèces

Voici quelques exemples de plantes aquatiques étudiées :

Equisetum sylvaticum ou "prêle des bois" :
famille des équisétacées

(photographie de R. Prelli)



Potamogeton pectinatus ou "potamot
pectiné" : famille des potamogetonacées

(photographie de Franck Le Driant)



Lemna minor ou "lentille d'eau" :
famille des lemnacées

(photographie de Hanno Schäfer)



Chara vulgaris ou "Charagne vulgaire" :
famille des characées (une sorte d'algue)

(photographie de Geoffrey Hall)



Figure 11 : exemples d'espèces

4. La réponse par la réalisation du programme

4.1. Les données

Le programme, écrit en langage R, prend en entrée deux tableaux *Excel* : un premier tableau contenant la base de données "banque de graines" (annexe I.1), et un deuxième contenant la base de données "végétation" (annexe I.2). Ces deux tableaux sont constitués de 72 lignes, les individus (en fait les espèces de végétaux étudiées) et de 32 colonnes pour le premier (une colonne "type de végétaux", à savoir terrestre, hydrophyte ou héliophyte – c'est-à-dire que les racines sont sous l'eau, mais les tiges, fleurs et feuilles sont aériennes - , une colonne "diaspore", à savoir graine, bourgeon dormant, rhizome, fragment ou bulbe, et de 30 colonnes pour les stations – 6 sites composés de 5 stations chacun). Le deuxième tableau quant à lui comporte 92 colonnes (les deux premières identiques à celles de la banque de graines, puis 90 colonnes contenant 3 saisons – printemps, été et automne – pour les 30 stations).

4.2. L'AFC et la CAH

Les banques de graines des six sites et des trente stations, ainsi que les végétations des différentes stations, sont comparées grâce à une analyse factorielle des correspondances et une classification ascendante hiérarchique. Celles-ci nous informent sur les similarités des différents sites et stations en terme de banque de graines et de végétation.

4.3. Les indices

Deux indices de similarité ont été retenus : un indice binaire (de présence / absence), l'indice de Jaccard, et un indice quantitatif, l'indice de Ruzicka. Ces indices ont été choisis car ils fournissent des intervalles de confiance moins larges que les autres indices étudiés, ce qui n'est pas négligeable vu que les indices varient entre 0 et 1.

4.4. Le bootstrap

Des intervalles de confiance bootstrap (optimisés) ont été calculés autour des indices qui comparent la similarité entre banque de graines des différentes stations et végétation (sur la même station ou sur d'autres stations du même site).

5. Discussions

La discussion ci-après concerne les bras morts reconnectés au Rhin, le bassin d'orage étant traité en annexe.

5.1. L'analyse factorielle des correspondances

Seules les analyses factorielles des correspondances de la banque de graines des stations et des sites (AFC inter-station) seront détaillées dans cette partie (les détails de l'AFC de la végétation des stations sont donnés en annexe). On cherche donc ici à comparer les différentes stations en terme de banque de graines. La commande "R" pour réaliser l'AFC est la suivante (le jeu de données s'appelle "station_banque") :

```
AFC_station <- dudi.coa (station_banque, scannf=TRUE)
```

Le package "ade4" est nécessaire au bon fonctionnement de la fonction "dudi.coa", il doit être installé et chargé avant toute utilisation. L'option "scannf" permet, en étant en position "TRUE" (vrai), de tracer le graphe en barre des valeurs propres (figure 12) dans le but de choisir le nombre d'axes à retenir.

5.1.1. Les valeurs propres

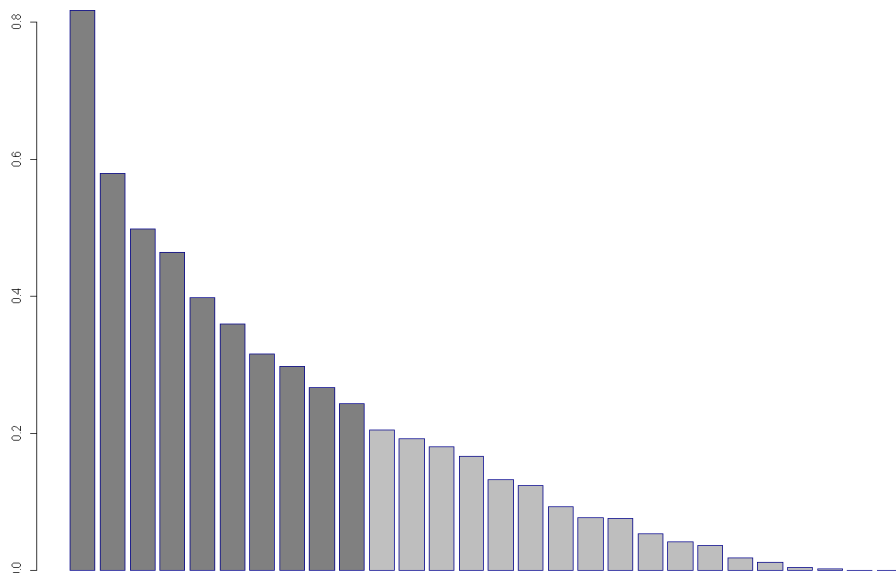


Figure 12 : graphe en barres des valeurs propres de l'AFC de la banque de graines par stations

Axe	Valeur propre	Pourcentage de variation cumulée
1	0,817330984	14,46%
2	0,579561059	24,71%
3	0,498329317	33,53%
4	0,463853611	41,73%
5	0,398391994	48,78%
6	0,359971365	55,15%
7	0,315996945	60,74%
8	0,298062922	66,01%
9	0,266483434	70,73%
10	0,242796277	75,02%
11	0,204795324	78,64%
12	0,192192043	82,04%
13	0,180607731	85,24%
14	0,16595229	88,17%
15	0,13181893	90,51%
16	0,123294214	92,69%
17	0,093195134	94,34%
18	0,076655769	95,69%
19	0,075922523	97,03%
20	0,053229306	97,98%
21	0,041188575	98,70%
22	0,036038283	99,34%
23	0,017685725	99,66%
24	0,011675238	99,86%
25	0,0047107	99,95%
26	0,002509872	99,99%
27	0,000356014	99,99%
28	0,000230516	100,00%

Tableau 5 : valeurs propres et pourcentage de variation cumulée de l'AFC de la banque par stations

Ces valeurs propres ont été calculées grâce aux fonctions "R" suivantes :

```
inertie_station <- inertia.dudi ( AFC_station, row.inertia=TRUE, col.inertia=TRUE )
```

```
inertie_station$TOT
```

Les 10 premiers axes, qui représentent 75% de l'inertie totale, ont été conservés.

5.1.2. Distances aux carrés des individus aux axes

La distance au carré d'un individu à un axe se mesure à l'aide des cosinus carrés. Les commandes du logiciel "R" pour calculer ces cosinus carrés sont les suivantes :

inertie_station\$col.rel pour les distances des colonnes ou stations aux axes (tableau 6)

inertie_station\$row.rel pour les distances des lignes ou individus aux axes (tableau 7)

Les cosinus carrés sont donnés en 10 000^{ème}, et donc les individus dont les cosinus carrés sont supérieurs à 5 000 en valeur absolue (en rouge) sont bien représentés sur l'axe correspondant.

Station	Axe1	Axe2	Axe3	Axe4	Axe5	Axe6	Axe7	Axe8	Axe9	Axe10	Autres
EISM1	-1009	322	559	-519	-21	-1045	63	-194	387	115	52
EISM2	-793	557	-29	-511	207	352	394	-123	-34	1085	327
EISM3	-204	338	90	-264	-83	-196	-97	47	-1384	1557	393
EISM4	-1098	1391	187	-2638	-115	148	-0	-863	-8	-32	261
EISM5	-571	1126	110	-2922	-133	727	-3	-2298	-65	-499	495
EISV1	-218	550	22	1163	-1055	-585	-659	433	-1799	6	307
EISV2	-625	1365	31	1922	-1306	28	-798	739	-1153	-231	36
EISV3	-409	296	-1	5073	-632	159	-77	135	27	-694	342
EISV4	-243	49	-136	-1514	-1152	283	1031	3720	1022	55	461
EISV5	-482	522	-0	-1551	-101	263	293	630	160	323	213
SCHA1	-204	-11	472	53	-6	-1908	89	-186	682	23	42
SCHA2	-893	-66	-111	-252	4819	-302	-1	357	-6	-420	244
SCHA3	-915	77	-19	-349	5600	-1070	-50	635	0	-645	477
SCHA4	-342	-180	139	-13	-1	302	-4350	8	1865	86	397
SCHA5	-603	134	268	1492	-323	360	4	-302	864	-729	335
SCHOL1	-813	-67	-1042	2157	1171	1133	1245	-395	-0	1261	553
SCHOL2	-678	-112	-1071	168	2	255	325	8	-554	-1638	304
SCHOL3	-384	-353	-5	2	-279	-5922	244	-1077	799	293	525
SCHOL4	-651	513	264	43	-255	-478	-14	-1	-3	89	304
SCHOL5	-899	653	11	366	-2	-96	-291	223	-1491	784	67
ROSS1	-264	-4934	-2529	-414	-829	-26	26	-6	-328	-218	729
ROSS2	-161	-3149	6113	0	15	91	297	17	-123	-1	875
ROSS3	-129	-2956	-317	-105	112	529	-3046	-0	48	680	456
ROSS4	-221	910	60	-4	-332	-200	-601	595	-3283	268	182
ROSS5	-1194	636	236	1711	-812	-28	-33	1	598	-634	27
FAHR1	-133	32	74	307	-662	-118	329	970	1719	-45	126
FAHR2	5516	-0	-2	-100	6	2	165	194	1	82	566
FAHR3	6471	-7	57	29	17	-173	8	-260	144	-49	101
FAHR4	8478	-33	-11	-45	60	17	-2	-144	-0	-15	665
FAHR5	7526	111	39	285	-53	-89	-87	2	-239	-33	139

Tableau 6 : distances aux carrés des colonnes aux axes

Espèces	Axe1	Axe2	Axe3	Axe4	Axe5	Axe6	Axe7	Axe8	Axe9	Axe10	Autres
Acer_pseudoplatanus	-362	538	167	-319	-253	797	-0	-2176	90	-1224	106
Acorus_calamus	-144	160	96	17	-115	-240	-8	-0	-2	66	55
Callitriche_obtusangula	-331	-39	-696	1548	978	1047	1311	-441	-0	1728	68
Carex_elata	-56	131	41	-128	-47	-123	-69	36	-1172	1447	87
Ceratophyllum_demersum	-144	160	96	17	-115	-240	-8	-0	-2	66	88
Chara_vulgaris	-486	355	-2	5608	-951	385	-148	122	117	-955	476
Dicot_herbacée	-906	-35	1279	135	-13	-3735	154	-302	988	30	295
Dicot_terrestres	-182	-571	-864	-1268	-1509	108	855	2540	323	-5	260
Dicot_phanérophyte	11	24	-26	-464	-779	80	1023	3445	1472	46	250
Echinocloa_crus_galli	-148	131	-26	-1137	-424	261	665	2161	649	227	238
Elodea_nuttallii	-332	-1774	4973	110	-8	491	-89	-1	370	-64	134
Epilobium_hirsutum	4441	-25	-9	-41	64	20	-2	-207	-0	-26	62
Equisetum_silvaticum	-2740	2397	75	-3776	-21	174	-1	-7	-39	93	433
Hypericum_perforatum	-136	-3595	-2143	-377	-878	-31	35	-9	-519	-379	173
Iris_pseudacorus	-185	183	-11	-210	99	186	238	-79	-24	852	176
Juncus_articulatus	-382	-6206	-85	-361	11	231	-1524	15	-19	51	350
Juncus_effusus	-693	-220	-785	712	2739	539	48	-53	-6	1522	265
Juncus_sp_(petit)	-144	-187	-3	1	-215	-5047	237	-1108	920	370	252
Lemna_minor	9419	-10	-0	-28	48	1	0	-69	0	-5	929
Lythrum_salicaria	-490	-2506	-3032	9	-135	174	591	-53	-868	-577	355
Phalaris_arundinacea_(G)	-147	46	107	640	-161	199	3	-202	645	-597	172
Phalaris_arundinacea_(R)	-342	851	104	-2580	-136	662	-2	-2722	-76	-611	395
Plantago_major	-99	-2733	6172	0	19	127	473	29	-233	-2	745
Poacées	18	-4251	-1421	-358	-64	192	-1430	-31	-55	93	176
Polygonum_aviculare	-144	160	96	17	-115	-240	-8	-0	-2	66	88
Populus_alba	-137	-1019	-10	-54	31	576	-5877	2	1177	660	234
Potamogeton_berchtoldii_(B)	-98	-73	66	-7	-0	197	-3235	7	1645	83	93
Potamogeton_berchtoldii_(G)	-318	38	-11	-214	3997	-845	-45	606	0	-756	78
Potamogeton_berchtoldii_(R)	-555	-59	8	-119	1194	-1466	-1465	17	1708	-12	93
Potamogeton_pectinatus	-1196	3	-815	990	1698	446	1600	-263	-0	1102	273
Potamogeton_perfoliatus_(G)	-98	-73	66	-7	-0	197	-3235	7	1645	83	93
Potamogeton_perfoliatus_(R)	-565	3	-79	-257	5858	-812	-15	792	-14	-1348	510
Robinia_pseudacacia	-147	-34	-381	64	1	126	183	5	-369	-1197	133
Salix_sp.	-46	834	66	531	-526	-574	-862	898	-3762	178	375
Solanum_dulcamara	-56	131	41	-128	-47	-123	-69	36	-1172	1447	276
Solidago_gigantea	-286	-3054	-1363	-203	-1058	-2612	216	-586	6	-5	290
Sparganium_emersum	-684	15	-152	2111	220	671	853	-582	114	711	63
Spirodella_polyrhiza_(F)	3442	-0	-1	-68	7	1	159	187	-0	97	358
Spirodella_polyrhiza_(G)	4757	-3	-4	-110	25	8	157	95	1	77	93
Urtica_dioica	-499	-5	-337	2291	465	1297	1101	-670	153	618	113
Veronica_beccabunga	-139	31	3	388	-699	-2312	-108	-0	-308	114	199
Zanichellia_palustris	-206	369	58	-874	-44	63	-0	-445	-5	-20	102

Tableau 7 : distances aux carrés des lignes aux axes

On voit par exemple que l'espèce *Lemna minor* (lentille d'eau) est très bien représentée sur le premier axe factoriel (son cosinus carré vaut quasiment 1, et donc l'angle entre la droite reliant l'origine à l'individu *Lemna minor* et le premier axe factoriel est quasi-nul).

5.1.3. Contribution des axes aux individus

La contribution des individus aux axes se calculent à l'aide des commandes suivantes :

inertie_station\$col.abs pour les contributions des colonnes ou stations aux axes (tableau 8)

inertie_station\$row.abs pour les contributions des lignes ou individus aux axes (tableau 9)

Station	Axe1	Axe2	Axe3	Axe4	Axe5	Axe6	Axe7	Axe8	Axe9	Axe10
EISM1	36	16	33	33	2	85	6	19	42	14
EISM2	179	178	11	204	96	181	231	77	24	826
EISM3	55	129	40	126	46	121	68	35	1154	1425
EISM4	198	354	55	838	43	61	0	427	4	20
EISM5	195	544	62	1762	93	565	2	2157	68	575
EISV1	46	165	8	434	459	282	361	251	1170	4
EISV2	15	47	1	83	66	2	51	50	87	19
EISV3	97	99	0	2116	307	85	47	87	20	553
EISV4	78	22	71	852	755	205	852	3256	1000	59
EISV5	71	108	0	402	31	88	111	255	72	160
SCHA1	6	0	23	3	0	127	7	15	61	2
SCHA2	150	16	31	75	1666	116	0	165	3	238
SCHA3	302	36	10	203	3787	801	43	574	0	716
SCHA4	94	70	63	7	0	188	3091	6	1571	79
SCHA5	140	44	102	609	153	190	3	192	614	569
SCHOL1	311	36	654	1455	920	985	1232	414	0	1625
SCHOL2	143	33	369	62	1	122	177	4	357	1159
SCHOL3	139	181	3	1	208	4880	229	1071	889	358
SCHOL4	137	152	91	16	110	228	8	0	2	63
SCHOL5	42	43	1	30	0	10	35	28	212	122
ROSS1	133	3507	2090	368	857	30	34	8	507	370
ROSS2	98	2687	6066	0	19	125	465	28	229	2
ROSS3	41	1314	164	58	73	379	2484	0	47	722
ROSS4	28	162	12	1	86	57	196	206	1270	114
ROSS5	22	17	7	57	31	1	2	0	34	40
FAHR1	12	4	11	47	118	23	74	232	459	13
FAHR2	2159	0	1	69	5	2	167	208	1	109
FAHR3	450	1	7	4	2	27	2	50	31	12
FAHR4	3899	22	8	36	57	17	2	181	0	23
FAHR5	724	15	6	48	10	19	22	1	71	11
Total (%)	73,32	75,08	81,56	76,32	79,85	72,31	76,59	74,85	70,54	81,7

Tableau 8 : contribution des colonnes aux axes

Les contributions sont elles-aussi exprimées en 10 000^{ème}. On voit par exemple que les stations FAHR4, FAHR2, FAHR5 et FAHR3 ont fortement contribué à la construction du premier axe factoriel (respectivement 39%, 21,6%, 7.2% et 4.5%, pour un total de 73.32% de l'inertie totale du premier axe factoriel).

Espèces	Axe1	Axe2	Axe3	Axe4	Axe5	Axe6	Axe7	Axe8	Axe9	Axe10
Acer_pseudoplatanus	26	55	20	41	38	132	0	436	20	301
Acorus_calamus	5	9	6	1	9	21	1	0	0	8
Callitriche_obtusangula	16	3	53	128	94	111	159	57	0	273
Carex_elata	3	11	4	14	6	17	11	6	215	292
Ceratophyllum_demersum	9	14	10	2	14	33	1	0	0	13
Chara_vulgaris	160	165	1	3251	642	287	126	110	118	1058
Dicot_herbacée	185	10	428	48	5	1731	81	169	619	21
Dicot_terrestres	33	145	255	402	558	44	398	1255	179	3
Dicot_phanérophyte	2	6	7	141	276	31	457	1631	780	27
Echinocloa_crus_galli	24	30	7	330	143	98	283	976	328	126
Elodea_nuttallii	31	231	754	18	1	103	21	0	105	20
Epilobium_hirsutum	191	1	1	3	6	2	0	24	0	4
Equisetum_silvaticum	820	1011	37	1990	13	118	1	6	36	94
Hypericum_perforatum	16	605	419	79	215	8	11	3	190	152
Iris_pseudacorus	23	31	2	45	25	51	75	26	9	349
Juncus_articulatus	93	2119	34	154	6	127	954	10	14	41
Juncus_effusus	127	57	236	230	1031	224	23	26	3	940
Juncus_sp_(petit)	25	46	1	0	77	1998	107	530	492	217
Lemna_minor	6052	9	0	32	64	1	1	121	0	11
Lythrum_salicaria	120	868	1221	4	68	97	375	35	654	477
Phalaris_arundinacea_(G)	17	8	21	134	39	54	1	66	235	239
Phalaris_arundinacea_(R)	93	328	47	1242	76	410	2	2039	64	562
Plantago_major	51	1987	5217	0	20	149	630	41	368	3
Poacées	2	729	283	77	16	53	450	10	21	38
Polygonum_aviculare	9	14	10	2	14	33	1	0	0	13
Populus_alba	22	233	3	15	10	212	2462	1	584	360
Potamogeton_berchtoldii_(B)	6	7	7	1	0	29	536	1	323	18
Potamogeton_berchtoldii_(G)	17	3	1	20	442	104	6	90	0	137
Potamogeton_berchtoldii_(R)	36	5	1	13	157	214	243	3	337	3
Potamogeton_pectinatus	226	1	252	329	657	191	781	136	0	700
Potamogeton_perfoliatus_(G)	6	7	7	1	0	29	536	1	323	18
Potamogeton_perfoliatus_(R)	199	1	46	160	4237	650	13	766	15	1600
Robinia_pseudacacia	14	4	58	10	0	26	43	1	104	371
Salix_sp.	12	305	28	242	280	338	578	638	2992	156
Solanum_dulcamara	11	35	13	43	18	53	34	19	686	929
Solidago_gigantea	57	864	448	72	435	1189	112	322	4	4
Sparganium_emersum	30	1	11	162	20	66	96	70	15	104
Spirodella_polyrhiza_(F)	852	0	0	30	4	1	102	127	0	81
Spirodella_polyrhiza_(G)	307	0	0	13	3	1	26	17	0	17
Urtica_dioica	39	1	43	316	75	230	223	144	37	163
Veronica_beccabunga	19	6	1	94	197	722	39	0	130	53
Zanichellia_palustris	15	37	7	109	6	10	0	86	1	5
Total (%)	77,24	75,78	71,92	75,44	73,45	73,25	77,82	78,35	71,75	76,47

Tableau 9 : contribution des lignes aux axes

On voit par exemple que l'espèce *Lemna minor* (lentille d'eau) a très fortement contribué à la construction du premier axe factoriel (60.5%).

5.1.4. Premier plan factoriel

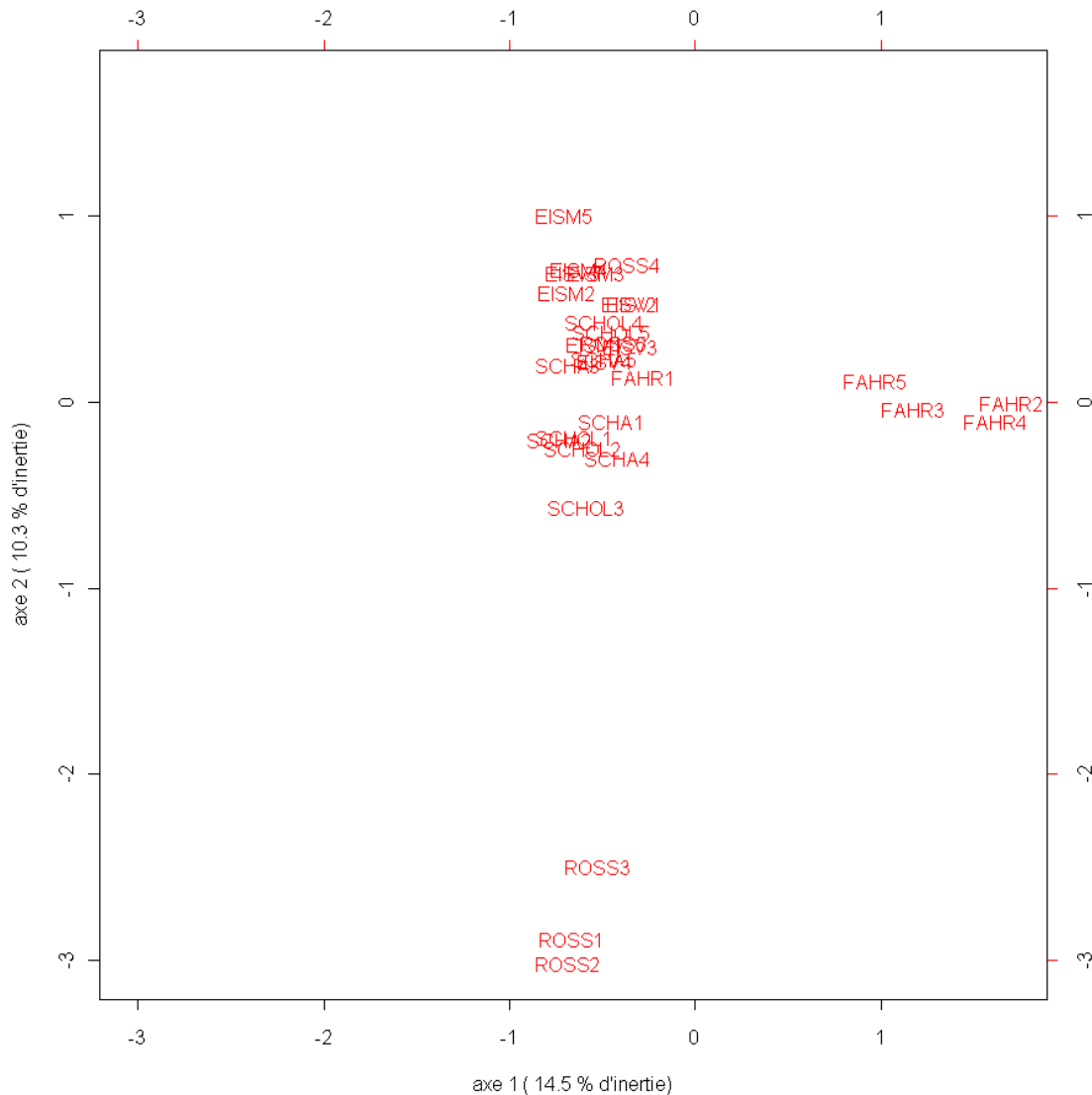


Figure 13 : premier plan factoriel de l'AFC de la banque de graines des stations

Le site du *Fahrgiessen* ressemble en fait à un bassin de décantation : l'eau y est plus ou moins stagnante (excepté sur la première station FAHR1 où se fait l'entrée d'eau) et le sédiment beaucoup plus riche en minéraux.

Le site du *Rossmoerder* est quant à lui très encaissé avec des berges très abruptes, quasiment verticales, et peu large par rapport aux autres sites, du moins sur les trois premiers transects. Les quatrième et cinquième stations (ROSS4 et ROSS5) du site du *Rossmoerder* sont en effet moins encaissées, légèrement plus larges et avec des pentes de berges plus douces, et de ce fait ressemblent aux autres sites. Le courant y est plus faible, et le fond est constitué de sable et de limon, contrairement aux trois premières stations qui sont composées essentiellement de galets.

5.1.5. AFC inter-stations

Les commandes "R" pour effectuer l'analyse factorielle des correspondances inter-stations sont les suivantes (la transposée est due au fait que l'AFC inter-stations s'utilise sur un tableau avec les espèces en colonnes et les stations en lignes) :

```
sites <- c ( )  
for ( i in 1 : nb_sites ) {  
  sites <- c ( sites, rep ( noms [i], nb_stations ) ) # "noms" est un vecteur qui contient  
  } # les noms des sites  
sites <- as.factor ( sites )  
inter_stations <- t ( between ( t ( AFC_stations ), scannf=TRUE, fac=sites ) )  
inertie_inter_stations <- inertia.dudi ( inter_stations, row.inertia=TRUE, col.inertia=TRUE )
```

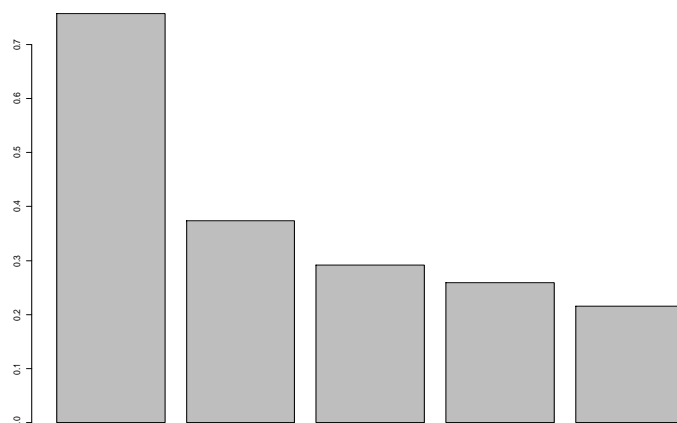


Figure 14 : graphe en barres des valeurs propres de l'AFC inter-stations de la banque de graines

Les valeurs propres, ainsi que le pourcentage de variation (tableau 10), sont données par :

inertie_inter_stations\$TOT

Axe	Valeur propre	Pourcentage de variation cumulée
1	0.7571423	39,89%
2	0.3741775	59,61%
3	0.2918817	74,99%
4	0.2589181	88,63%
5	0.2156448	100%

Tableau 10 : valeurs propres et pourcentage de variation cumulée de l'AFC inter-stations

Deux axes ont été retenus, représentant 60% de variation cumulée.

Les distances aux carrés (cosinus carrés) et les contributions des sites (colonnes) aux axes pour l'AFC inter-stations se mesurent à l'aide des commandes "R" respectives suivantes :

inertie_inter_stations\$col.rel pour les cosinus carrés des sites aux axes (tableau 11a)
inertie_inter_stations\$col.abs pour les contributions des sites aux axes (tableau 11b)

Site	Cosinus carrés			Sites	Contribution	
	Axe 1	Axe 2	Autres		Axe 1	Axe 2
EISM	1901	5148	1733	EISM	826	4526
EISV	898	-4	1221	EISV	275	3
SCHA	2193	5	1305	SCHA	717	3
SCHOL	2643	-248	1199	SCHOL	794	151
ROSS	740	-6115	1710	ROSS	317	5304
FAHR	-9962	9	1221	FAHR	7071	14
				Total (%)	70,71	98,3

Tableau 11 : cosinus carrés et contributions des colonnes (sites) de l'AFC inter-stations de la banque de graines

Le site FAHR (Fahrgiessen) est très bien représenté sur l'axe 1, il contribue d'ailleurs à 70% de la construction de cet axe, alors que les sites EISM et ROSS (Eiswasser amont et Rossmoerder) sont bien représentés sur l'axe 2 et contribuent respectivement à 45% et 53% de la construction de celui-ci.

Les distances aux carrés (cosinus carrés) et les contributions des espèces (lignes) aux axes pour l'AFC inter-stations se mesurent à l'aide des commandes "R" respectives suivantes :

inertie_inter_stations\$row.rel pour les cosinus carrés des espèces aux axes (tableau 10a)
inertie_inter_stations\$row.abs pour les contributions des espèces aux axes (tableau 10b)

On voit que les espèces *Epilobium hirsutum* (épilobe à grandes fleurs), *Lemna minor* (lentille d'eau), qui contribue d'ailleurs à plus de 60% à la construction du premier axe (tableau 12b), et *Spirodella polyrhiza* (lentille d'eau géante), sont, avec des cosinus carrés très proches de 1, très bien représentés sur l'axe 1 (tableau 12a).

De même, les espèces *Carex elata* (laiche élevée), *Hypericum perforatum* (millepertuis), *Iris pseudacorus* (iris des marais), *Juncus articulatus* (jonc articulé), ..., sont entre autres bien représentés sur le deuxième axe (tableau 12a).

Les espèces *Equisetum silvaticum* (prêle des bois), *Juncus articulatus*, *Phalaris arundinacea* (baldingère faux-roseau) en rhizome et *Plantago major* (Grand plantain) sont les espèces qui ont le plus fortement contribué à la construction du deuxième axe factoriel de l'AFC inter-stations avec respectivement 18,15,12 et 13% de contribution (tableau 12b).

Espèces	Cosinus carrés			Espèces	Contributions	
	Axe 1	Axe 2	Autres		Axe 1	Axe 2
Acer pseudoplatanus	2173	3823	63	Acer pseudoplatanus	34	122
Acorus calamus	979	-186	35	Acorus calamus	9	3
Callitriche obtusangula	979	-186	55	Callitriche obtusangula	14	5
Carex elata	946	5183	28	Carex elata	7	73
Ceratophyllum demersum	979	-186	55	Ceratophyllum demersum	14	5
Chara vulgaris	830	-155	626	Chara vulgaris	130	49
Dicot herbacée	3468	-38	132	Dicot herbacée	115	3
Dicot terrestres	728	-2307	131	Dicot terrestres	24	153
Dicot phanérophte	-2082	1	81	Dicot phanérophte	42	0
Echinocloa crus galli	322	-3	148	Echinocloa crus galli	12	0
Elodea nuttallii	1378	-2287	115	Elodea nuttallii	40	134
Epilobium hirsutum	-9900	19	64	Epilobium hirsutum	160	1
Equisetum silvaticum	4462	4044	883	Equisetum silvaticum	988	1811
Hypericum perforatum	345	-5769	147	Hypericum perforatum	13	430
Iris pseudacorus	946	5183	111	Iris pseudacorus	26	293
Juncus articulatus	691	-5703	523	Juncus articulatus	91	1514
Juncus effusus	2323	-1198	208	Juncus effusus	121	127
Juncus sp. (petit)	979	-186	111	Juncus sp. (petit)	27	10
Lemna minor	-9894	20	2485	Lemna minor	6162	25
Lythrum salicaria	1458	-4606	290	Lythrum salicaria	106	677
Phalaris arundinacea (G)	865	4	125	Phalaris arundinacea (G)	27	0
Phalaris arundinacea (R)	946	5183	473	Phalaris arundinacea (R)	112	1244
Plantago major	345	-5769	440	Plantago major	38	1289
Poacées	-23	-5805	189	Poacées	1	557
Polygonum aviculare	979	-186	55	Polygonum aviculare	14	5
Populus alba	1378	-2287	115	Populus alba	40	134
Potamogeton berchtoldii (B)	865	4	63	Potamogeton berchtoldii (B)	14	0
Potamogeton berchtoldii (G)	865	4	63	Potamogeton berchtoldii (G)	14	0
Potamogeton berchtoldii (R)	1418	-2	120	Potamogeton berchtoldii (R)	43	0
Potamogeton pectinatus	4299	459	197	Potamogeton pectinatus	213	46
Potamogeton perfoliatus (G)	865	4	63	Potamogeton perfoliatus (G)	14	0
Potamogeton perfoliatus (R)	1061	1	578	Potamogeton perfoliatus (R)	154	0
Robinia pseudacacia	979	-186	55	Robinia pseudacacia	14	5
Salix sp.	457	-3006	177	Salix sp.	20	269
Solanum dulcamara	946	5183	89	Solanum dulcamara	21	233
Solidago gigantea	1188	-5918	176	Solidago gigantea	53	529
Sparganium emersum	1793	-173	76	Sparganium emersum	34	7
Spirodella polyrhiza (F)	-9900	19	289	Spirodella polyrhiza (F)	718	3
Spirodella polyrhiza (G)	-9900	19	96	Spirodella polyrhiza (G)	239	1
Urtica dioica	2097	-152	78	Urtica dioica	41	6
Veronica beccabunga	977	-58	100	Veronica beccabunga	25	3
Zanichellia palustris	946	5183	89	Zanichellia palustris	21	233
				Total (%)	71,5	70,64

Tableau 12 : cosinus carrés et contributions des lignes de l'AFC inter-stations de la banque de graines

Le premier plan factoriel de l'analyse factorielle des correspondances inter-stations (ou de l'AFC des sites) est tracé (figure 15) à l'aide de la commande suivante :

```
biplot ( inter_stations$co[,1:2], inter_stations$co[,1:2], var.axes=F, xlab=paste( "axe1 ( ",
ceiling ( inter_stations$eig[1]/sum(inter_stations$eig ) *1000 )/10,"% d'inertie ) " ),
ylab=paste ( "axe 2 ( ", ceiling ( inter_stations$eig[2]/sum ( inter_stations$eig )
*1000 )/10, "% d'inertie ) " ), main="AFC inter stations" )
```

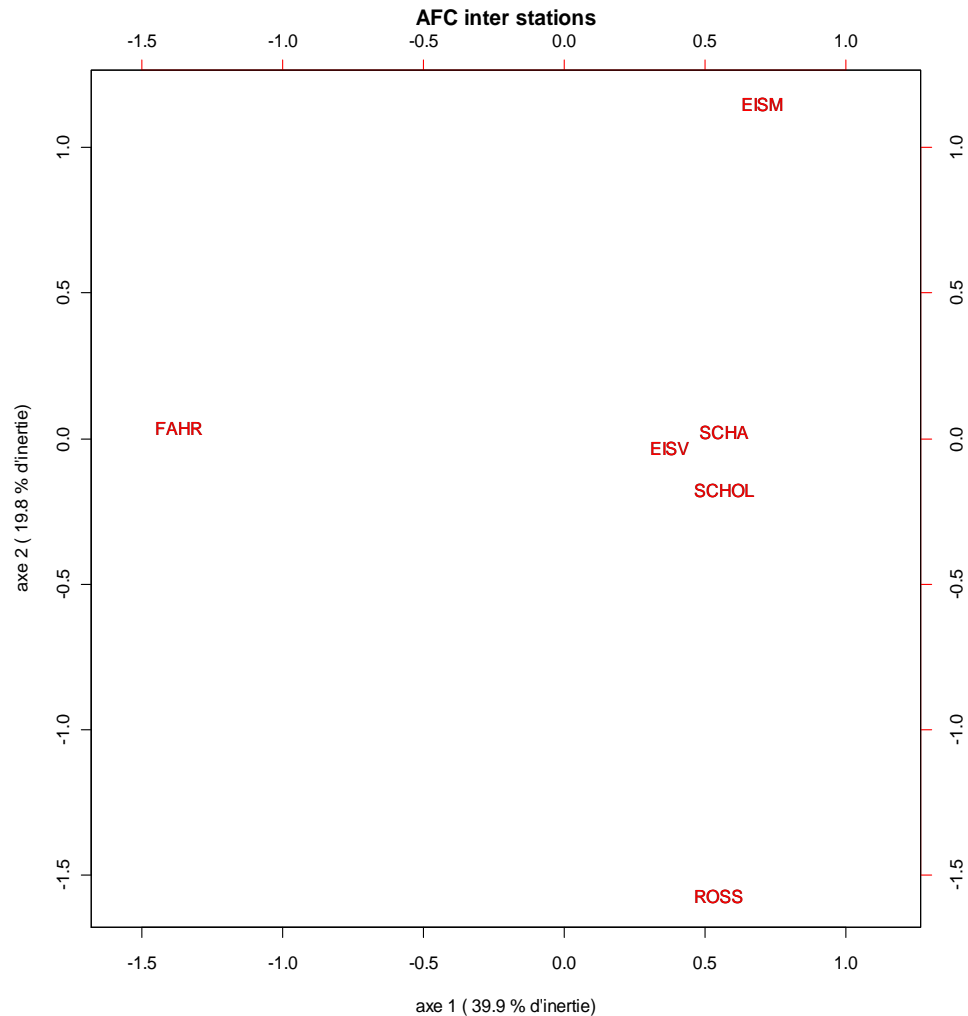


Figure 15 : premier plan factoriel de l'AFC inter-stations de la banque de graines

Comme dit précédemment, le courant sur le site du *Fahrgiessen* est très faible comparé aux autres sites, ce qui ralentit le déplacement des différentes diaspores.

L'axe 2 traduit l'encaissement du site : en effet, les berges du site du *Rossmoerder* sont très abruptes alors que celles de l'*Eiswasser Amont* sont en pente beaucoup plus douces. Les autres sites ont un encaissement moyen par rapport aux deux sites précédemment cités.

5.1.6. AFC intra-stations et intra-dates

5.1.6.a. AFC de la banque intra-stations

Avec une partition du tableau de la banque de graines suivant les sites, nous obtenons une collection des structures stationnelles sur chaque site (figure 16).

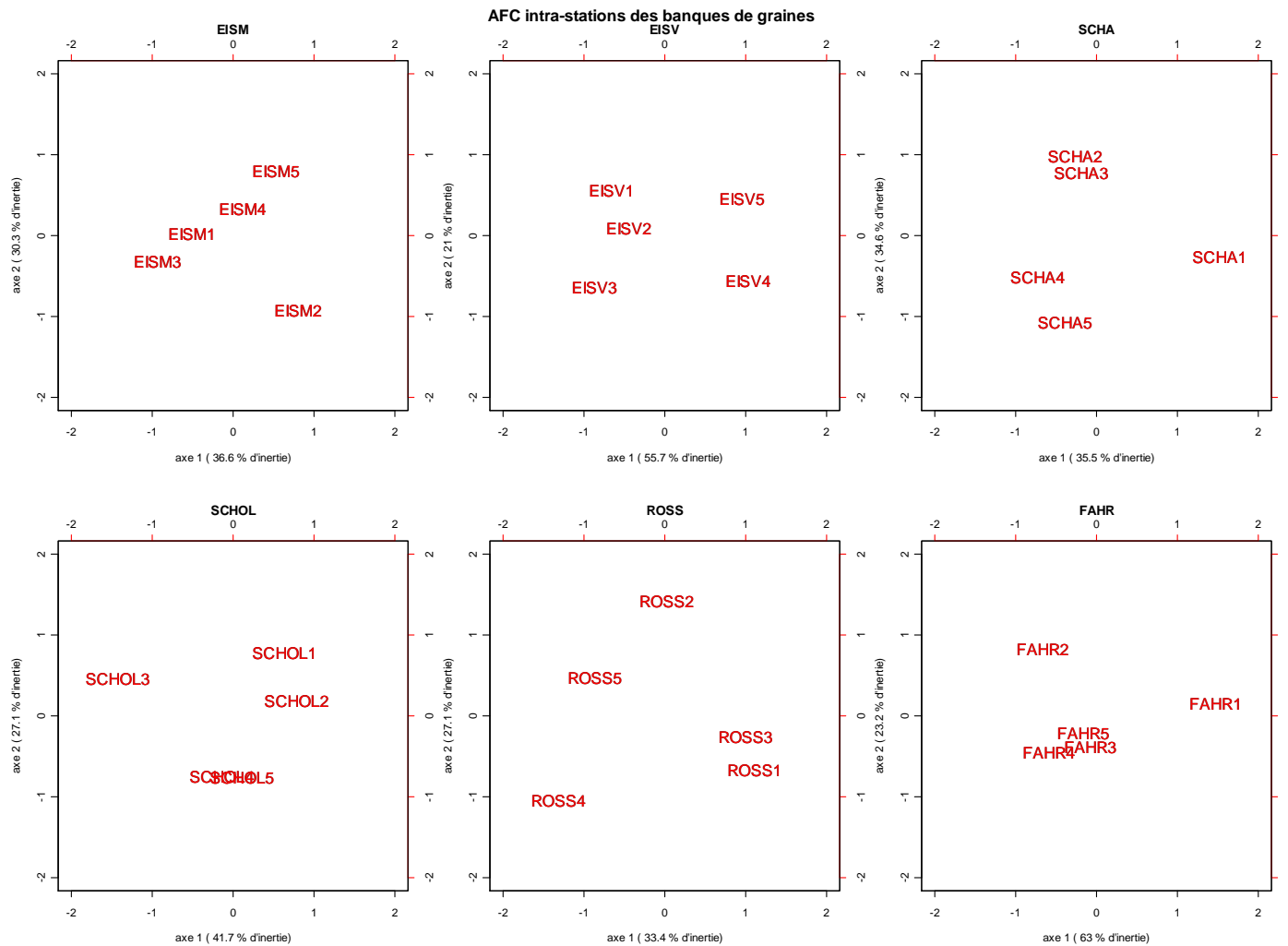


Figure 16 : AFC de la banque de graines intra-stations

La dispersion des stations est quasiment identique sur chaque site, excepté sur le site du *Fahrgiessen*, qui est le site le plus récemment connecté.

On constate qu'il n'y a pas de gradient (ou d'organisation) amont-aval, c'est-à-dire que les stations ne sont pas représentées dans l'ordre amont-aval (de 1 à 5) sur ce premier plan factoriel, mais aléatoirement.

5.1.6.b. AFC de la végétation intra-stations

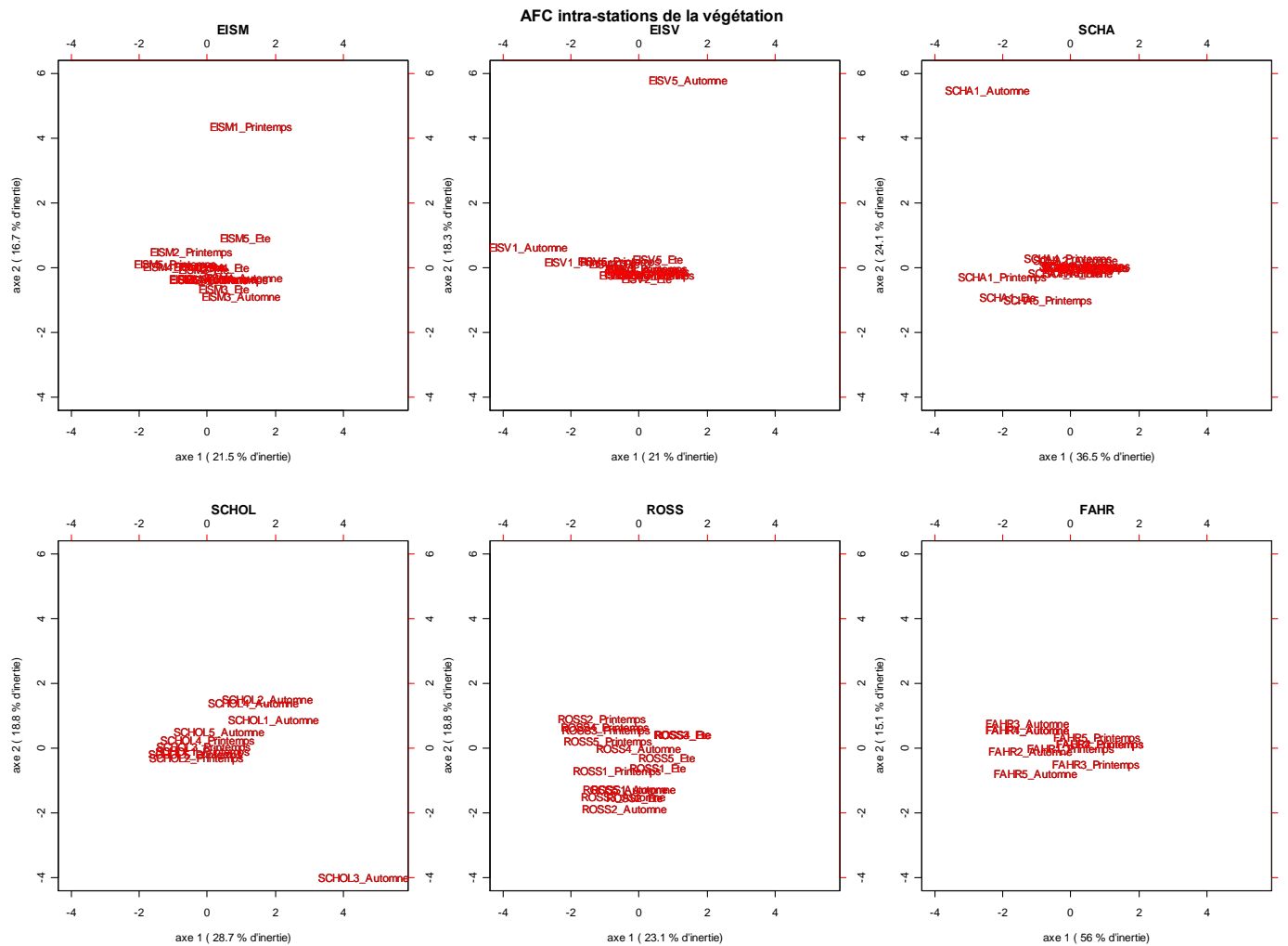


Figure 17 : AFC de la végétation intra-stations

Nous remarquons que les différents sites sont assez ressemblants en terme de végétation, excepté pour la première station de *l'Eiswasser Amont* au printemps (EISM1) et du *Schafftheu* en automne (SCHA1), et la cinquième de *l'Eiswasser Aval* en automne (EISV5), où les espèces se font très rares. La troisième station du *Schollengiessen* (SCHOL3) est, par contre, est envahie en automne par *l'elodea nutallii* (élodée à feuilles étroites ou du Canada).

5.2. La classification ascendante hiérarchique

5.2.1. Classification ascendante hiérarchique de la banque des stations

Les données de la banque de graines étudiées étant des proportions de densités, la distance utilisée ici est la distance du Chi2 et l'indice d'agrégation l'indice de Ward, qui consiste à regrouper les classes pour lesquelles la perte d'inertie est la plus faible.

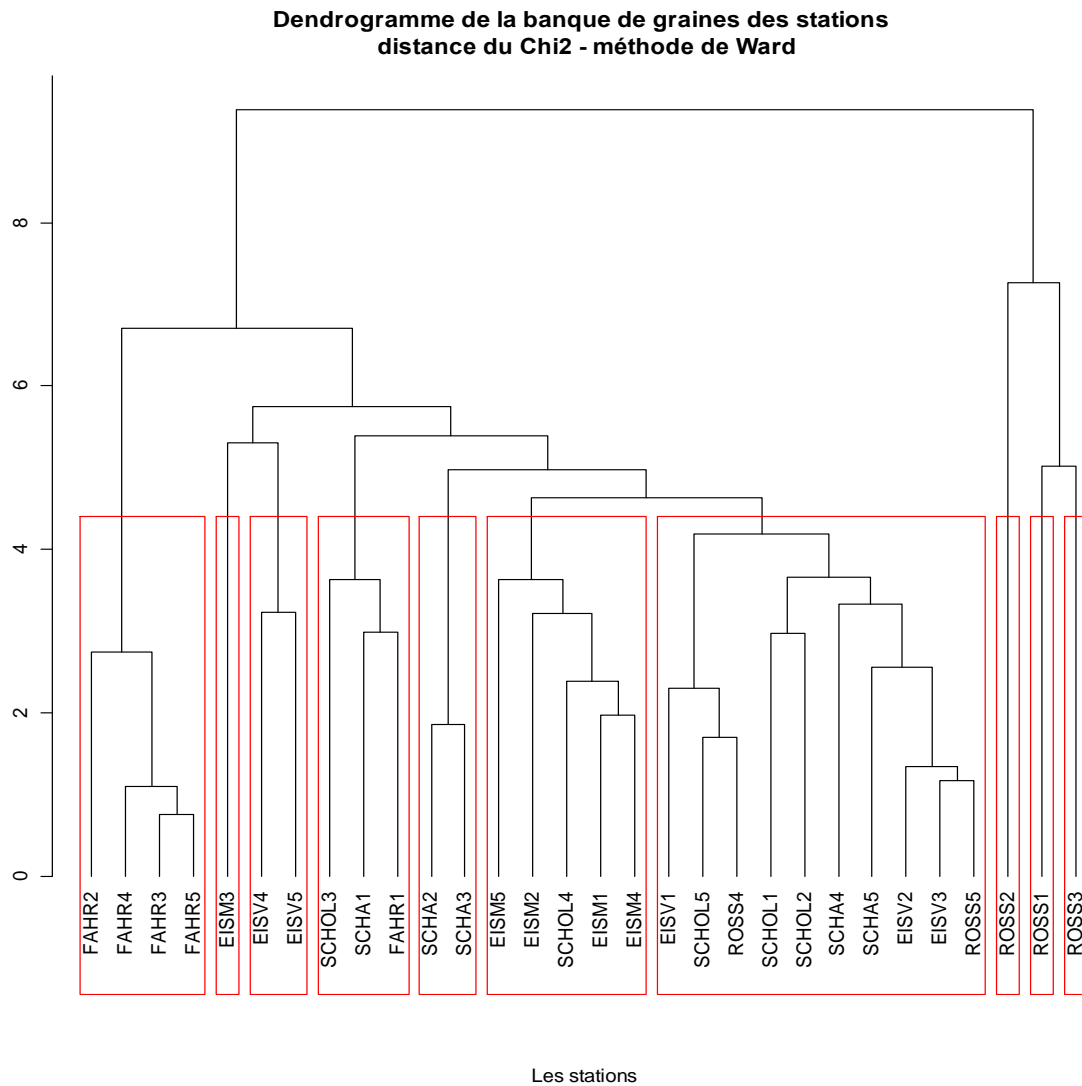


Figure 19 : dendrogramme de la banque de graines des stations

Nous pouvons y voir que les stations du *Rossmoerder* et du *Fahrgiessen*, sites restaurés plus récemment, sont plus éloignés des stations des autres sites en terme de banque de graines.

5.2.2. Classification ascendante hiérarchique de la banque des sites

La distance utilisée ici est, pour les mêmes raisons que précédemment, à nouveau la distance du Chi2, et l'indice d'agrégation l'indice de Ward.

Les commandes "R" pour tracer ce dendrogramme sont les suivantes :

```
dbanque_site <- dist.dudi ( t ( inter_stations ) )
cah_site = hclust ( dbanque_site^2, method="ward" )
plot ( cah_site, main = "Dendrogramme de la banque de graines des sites \n distance du Chi2
méthode de Ward", xlab = "Les sites", hang = -1)
```

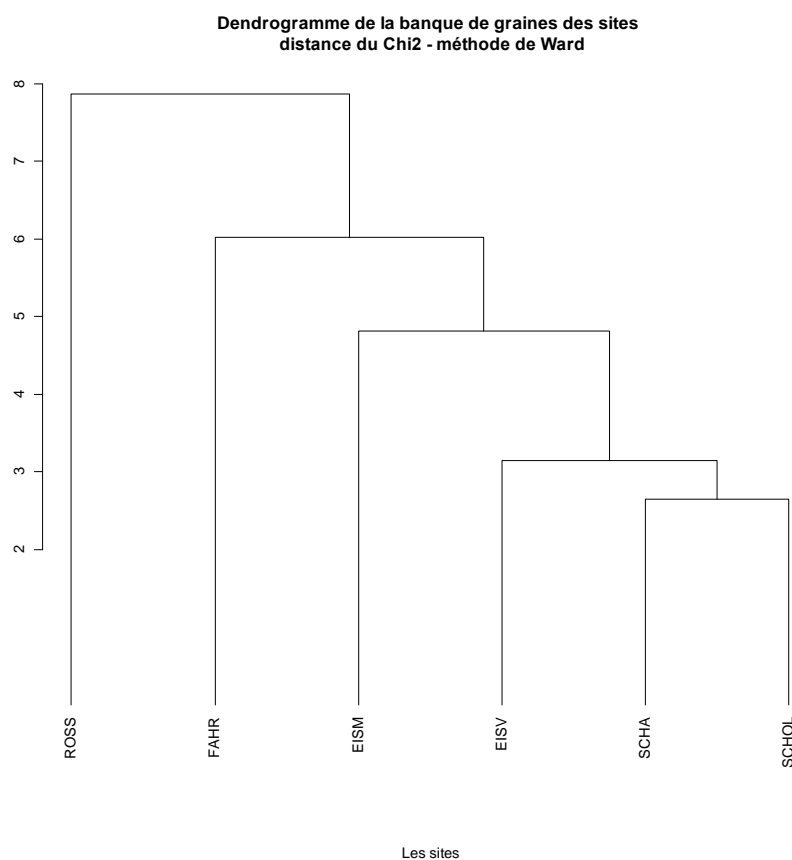


Figure 20 : dendrogramme de la banque de graines des sites (inter-stations)

Ce dendrogramme nous montre qu'en terme de banque de graines, les deux sites de références (le *Schafteu* et le *Scholengiessen*) sont les plus proches, suivis de près par le site restauré il y a quelques années (l'*Eiswasser*), et enfin les sites restaurés plus récemment (le *Rossmoerder* et le *Fahrgiessen*), ce que nous avons déjà pu conclure de l'analyse factorielle des correspondances inter-stations.

5.2.3. Classification ascendante hiérarchique de la végétation des stations

La distance utilisée ici est, pour les mêmes raisons que précédemment, à nouveau la distance du Chi2, et l'indice d'agrégation l'indice de Ward.

Les commandes "R" pour tracer ce dendrogramme sont les suivantes :

```
AFC_vegetation <- dudi.coa ( t ( stations_vegetation ), scannf=FALSE )
dbanque_vegetation <- dist.dudi ( t ( AFC_vegetation ) )
cah_vegetation = hclust ( dbanque_vegetation^2, method="ward" )
plot ( cah_vegetation, main = "Dendrogramme de la végétation moyenne des stations \n
      methode de Ward", xlab = "Les stations", hang=-1)
rect.hclust ( cah_vegetation, k= 6, border = "red")
```

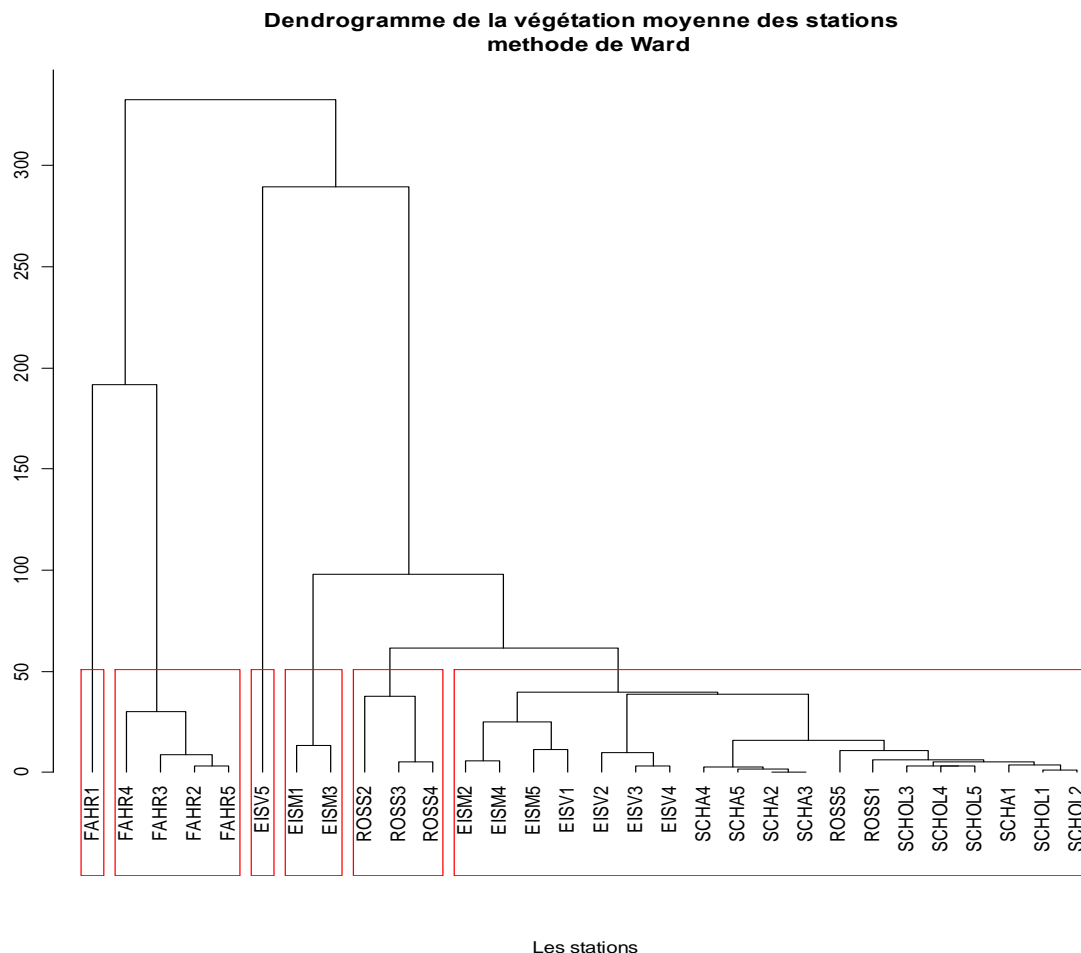


Figure 21 : dendrogramme de la végétation moyenne des stations

Comme pour la banque de graines, les sites de références ainsi que les sites moins récemment restaurés sont placés dans la même classe, ce qui signifie que les végétations observées sur ces stations se ressemblent.

5.3. Les indices et intervalles de confiance bootstrap associés

Des indices de similitude de Jaccard (et de Ruzicka) et des intervalles de confiance bootstrap autour de ces indices ont été calculés et tracés (figure 11) dans le but de comparer banque de graines et végétation correspondante.

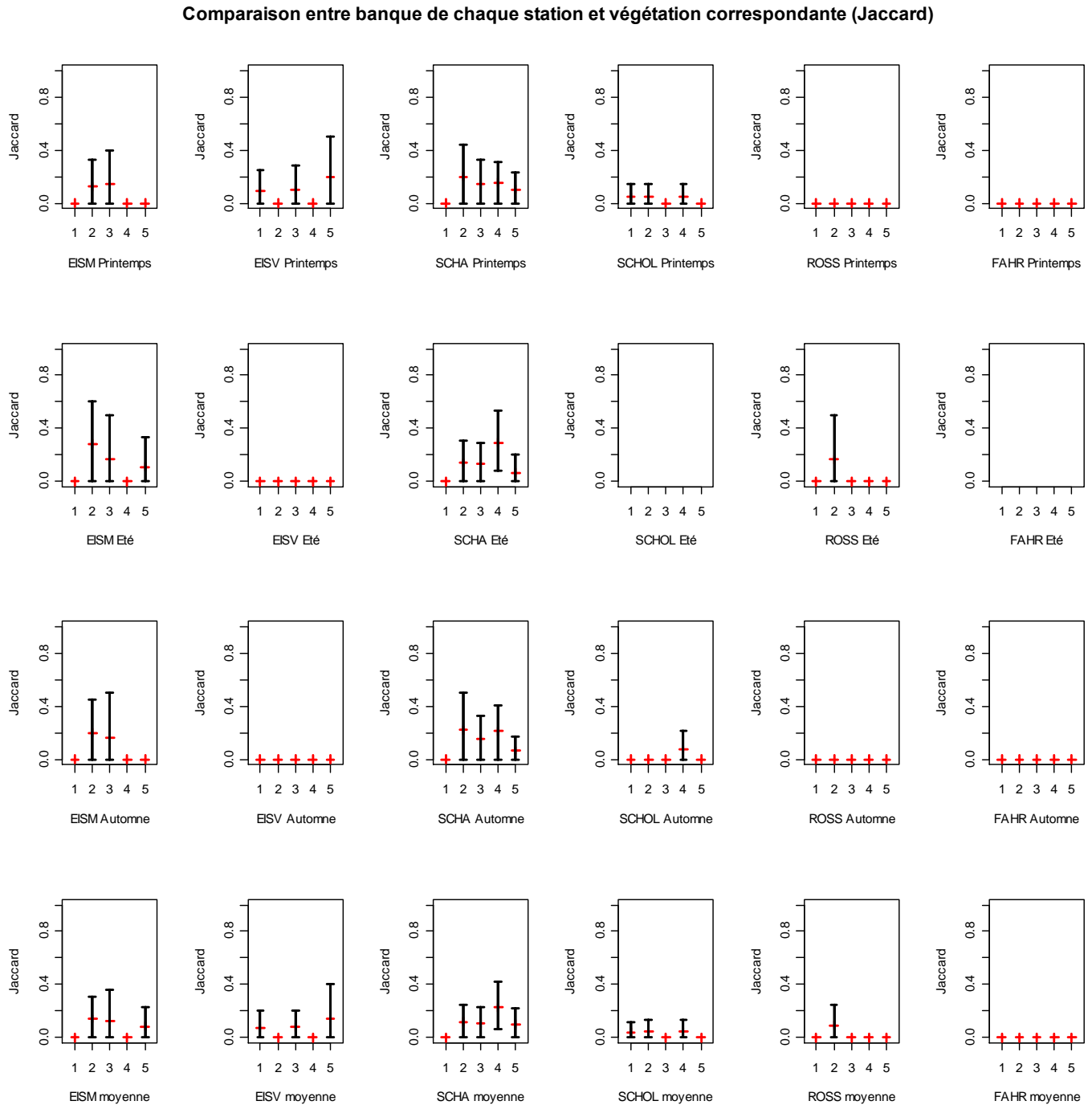


Figure 22 : indices de Jaccard et intervalles de confiance de la comparaison banque de graines / végétation

Chaque cadre correspond à un site pour une saison (printemps pour la première ligne, été pour la deuxième, automne pour la troisième et enfin, la moyenne des trois saisons dans la dernière ligne) : la banque de graines des 5 stations de chaque site est donc comparée à la végétation correspondante pour chaque saison. La croix rouge (+) désigne l'indice (ici de Jaccard) qui compare banque de graines et végétation, et la bande noire (I) correspond à l'intervalle de confiance bootstrap associé. Lorsque l'indice est nul, l'intervalle de confiance bootstrap est lui aussi nul, et seul l'indice est alors tracé sur le graphe (par exemple les 5 stations du site Rossmoerder au printemps). Les graphes vierges correspondent aux données manquantes (pour les sites du Schollengiessen et du Fahrgiessen en été).

Les graphes se lisent de la manière suivante : par exemple, le premier cadre représente le site *Eiswasser amont* (EISM) au printemps. Le premier indice et son intervalle de confiance bootstrap associé dans ce cadre représente la comparaison entre la banque de graines de la station EISM1 avec la végétation de la même station (EISM1) au printemps ; le deuxième représente la comparaison entre la banque de graines de la station EISM2 avec la végétation de la même station (EISM2) au printemps, et ainsi de suite.

Les sites du *Schafteu* (SCHA) et du *Schollengiessen* (SCHOL) sont les sites de référence, c'est-à-dire qui n'ont pas été déconnectés du Rhin ; ceux de l'*Eiswasser* (EISM et EISV) sont ceux qui ont été les premiers reconnectés (en 2002), c'est pourquoi ils présentent une légère similarité entre banque de graines et végétation, et enfin, les sites du *Rossmoerder* (ROSS) et du *Fahrgiessen* (FAHR) ont été les derniers à être reconnectés, respectivement en 2004 et 2006, ce qui explique en grande partie pourquoi ces deux derniers ne présentent quasi-aucune similarité entre banque de graines et végétation.

Excepté sur les sites de l'*Eiswasser aval* (EISV), qui ne se situe pas sur l'embouchure du Rhin, et du *Schollengiessen* (SCHOL), l'indice comparant banque de graines et végétation sur le premier transect est toujours nul. Ceci peut être en partie expliqué par le fort courant en amont, à l'endroit où l'eau entre sur le site, qui peut alors déplacer les graines vers les transects suivants, d'où l'étude de la comparaison entre banque de graines et végétation moyenne de chaque station sur le même site qui suit.

Comparaison entre banque de chaque station et végétation moyenne de chaque station sur le même site (Jaccard)

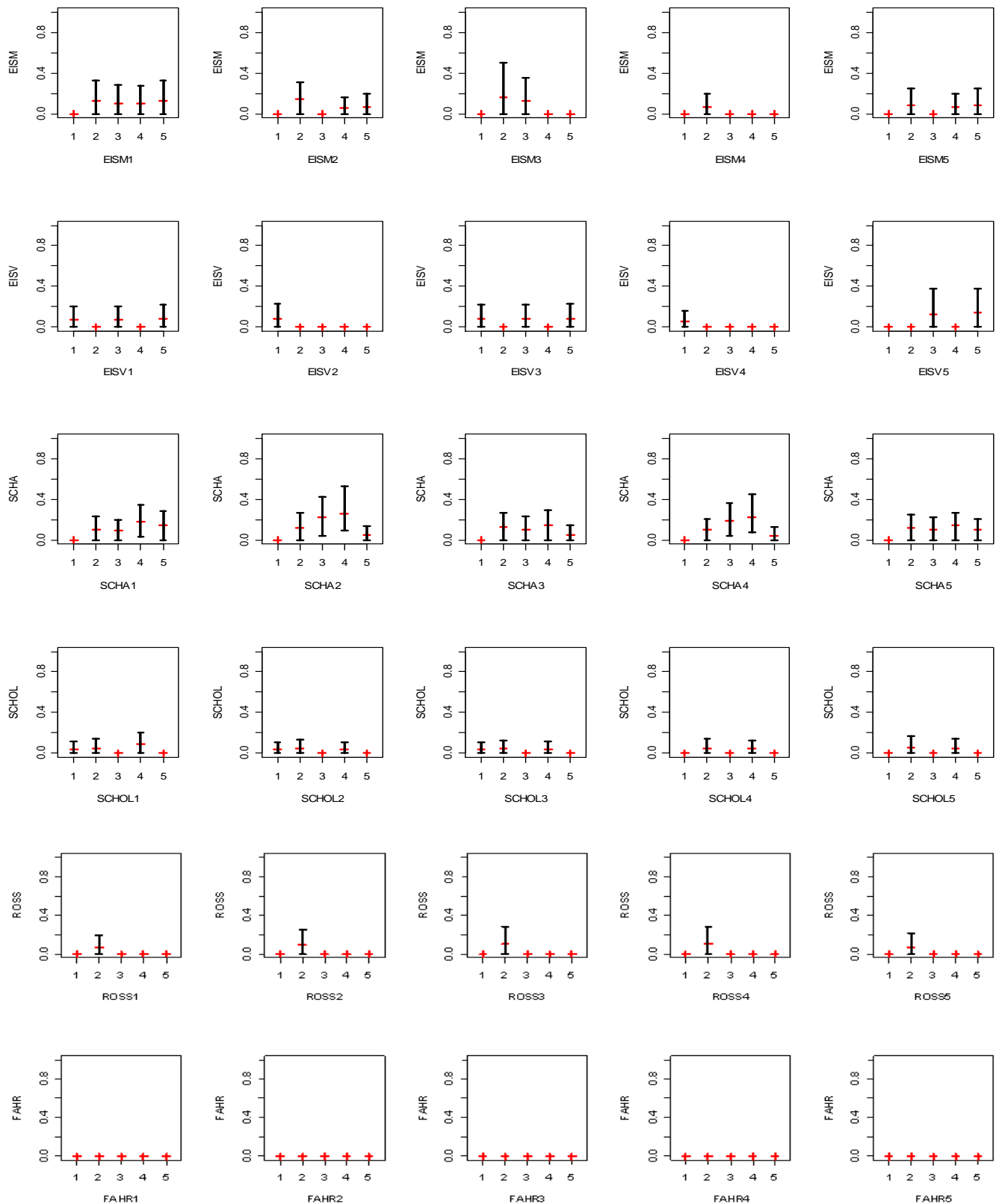


Figure 23 : indices de Jaccard et intervalles de confiance de la comparaison banque de graines / végétation moyenne de chaque station sur le même site que la banque de graines

Chaque ligne représente un site : la première représente le site *Eiswasser amont*, la deuxième le site *Eiswasser aval*, et ainsi de suite. Chaque ligne contient 5 graphiques, un par station. Chaque graphe compare la banque de graines de la station en question (par exemple EISM1 pour le premier cadre) avec la végétation de chaque station située sur le même site que la station sur laquelle on compare la banque de graines. Par exemple, dans le premier cadre sont comparées les banque de graines et végétation du premier transect du site *Eiswasser amont* (EISM1), puis toujours la banque de graines de la première station du site *Eiswasser amont* (EISM1) avec la végétation de la deuxième station du même site (EISM2), et ainsi de suite.

Les mêmes conclusions que précédemment peuvent être tirées, à savoir que le site le plus récemment restauré, c'est-à-dire le site du *Fahrgiessen* (FAHR) restauré en 2006, ne présente aucune similarité entre banque de graines et végétation, que le site du *Rossmoerder* (ROSS), restauré en 2004, ne présente qu'une légère similarité entre banque de graines et végétation, et que les sites restaurés il y a quelques années, à savoir les deux sites de l'*Eiswasser* (EISM et EISV), restaurés en 2002, s'approchent des sites de références, le *Schafteu* (SCHA) et le *Schollengiessen* (SCHOL).

6. Adaptation du programme

Le programme a été écrit pour pouvoir être adapté à d'autres situations, comme par exemple avec un nombre de sites, de stations ou de saisons différent de celui-ci-dessus, ou encore pour d'autres jeux de données, comme par exemple pour l'étude des invertébrés.

7. Conclusions

7.1. Conclusions

L'objectif de ce stage était d'établir une liaison entre banque de diaspores et végétation, par le calcul d'indices de similarité et d'intervalles de confiance. Que ce soient les analyses factorielles des correspondances, les classifications ascendantes hiérarchiques ou les calculs d'indices de similarité et d'intervalles de confiance bootstrap, toutes les techniques statistiques utilisées lors de ce stage nous ont mené à la même conclusion : conformément à ce à quoi nous nous attendions, les sites récemment restaurés sont plus différents des sites de référence que les sites restaurés il y a plus longtemps.

Bien entendu, le pourcentage de végétation présente sur chaque site étant évalué à l'œil nu, les résultats restent approximatifs.

Le programme sera régulièrement utilisé pour suivre l'évolution des différents sites restaurés, jusqu'à ce que l'état de ces sites soit similaire à celui des sites de référence. C'est à ce moment seulement que tout le travail entrepris ces dernières années pour restaurer ces sites portera enfin ses fruits.

7.2. Conclusions personnelles

Ce stage a été très formateur pour moi, tant au point de vue statistique (analyses multivariées, classifications, techniques du bootstrap), qu'écologique (notamment grâce aux sorties sur le terrain). Il m'a permis de me familiariser un peu plus avec le logiciel de programmation statistique R, surtout vis-à-vis du bootstrap.

J'ai beaucoup appris durant ce stage, par exemple qu'une absence de réponse ou d'explication était aussi importante qu'une présence, ou encore que le plus important était de comprendre ce que l'on attendait réellement de nous, afin de pouvoir répondre à ces attentes par la proposition d'outils statistiques appropriés.

Ce stage m'a en tout cas conforté dans mon idée d'allier statistiques et écologie pour en faire mon métier.

ANNEXES

I. Les données des bras reconnectés au Rhin

I.1. La banque de graines

Espèces	Type	Diaspore	EISM1	EISM2	EISM3	EISM4	EISM5	EISV1	EISV2
Acer_pseudoplatanus	Ter	G	0	0	0	0	35,3682	0	0
Acorus_calamus	He	R	0	0	0	0	0	0	0
Agrostis_stolonifera	Ter	0	0	0	0	0	0	0	0
Algues_filamenteuse	Hy	0	0	0	0	0	0	0	0
Azolla	Hy	0	0	0	0	0	0	0	0
Berula_erecta	Hy	0	0	0	0	0	0	0	0
Bryophyte	Hy	0	0	0	0	0	0	0	0
Callitriche_obtusangula	Hy	G	0	0	0	0	0	0	0
Carex_elata	He	R	0	0	11,111	0	0	0	0
Ceratophyllum_demersum	Hy	F	0	0	0	0	0	0	0
Chara_vulgaris	Hy	G	0	0	0	35,3682	0	176,841	70,7364
Cladophora	Hy	0	0	0	0	0	0	0	0
Dicot_herbacée	Ter	G	70,7364	0	35,3682	70,7364	0	0	0
Dicot_terrestres	Ter	F	0	0	0	0	0	0	0
Dicot_phanérophyte	Ter	G	0	0	0	0	0	0	0
Echinocloa_crus_galli	Ter	G	0	0	0	0	0	0	0
Elodea_nuttallii	Hy	F	0	0	0	0	0	0	0
Enteromorpha	Hy	0	0	0	0	0	0	0	0
Epilobium_hirsutum	Ter	G	0	0	0	0	0	0	0
Equisetum_silvaticum	He	G	70,7364	282,945	106,105	389,05	282,945	70,7364	35,3682
Fontinallis	Hy	0	0	0	0	0	0	0	0
Glyceria_fluitans	He	0	0	0	0	0	0	0	0
Graminées	Ter	0	0	0	0	0	0	0	0
Hildenbrandia	Hy	0	0	0	0	0	0	0	0
Hydrodiction	Hy	0	0	0	0	0	0	0	0
Hypericum_perforatum	Ter	G	0	0	0	0	0	0	0
Iris_pseudacorus	He	R	0	44,444	0	0	0	0	0
Juncus_articulatus	He	G	0	0	0	0	0	0	0
Juncus_effusus	He	G	0	0	0	0	0	0	0
Juncus_sp_(petit)	He	G	0	0	0	0	0	0	0
Lemna_minor	Hy	F	0	0	0	0	0	0	0
Lemna_minuta	Hy	0	0	0	0	0	0	0	0
Lythrum_salicaria	He	G	0	0	0	0	0	0	0
Melosira	Hy	0	0	0	0	0	0	0	0
Mentha_aquatica	He	0	0	0	0	0	0	0	0
Myosotis_palustris	He	0	0	0	0	0	0	0	0

Myriophyllum_spicatum	Hy	0	0	0	0	0	0	0	0	0
Myriophyllum_verticillatum	Hy	0	0	0	0	0	0	0	0	0
Nastutium_officinale	He	0	0	0	0	0	0	0	0	0
Nitella	Hy	0	0	0	0	0	0	0	0	0
Phalaris_arundinacea	He	G	0	0	0	0	0	0	0	0
Phalaris_arundinacea	He	R	0	0	0	55,556	133,33	0	0	0
Phragmites	He	0	0	0	0	0	0	0	0	0
Plantago_major	Ter	G	0	0	0	0	0	0	0	0
Poacées	Ter	G	0	0	0	0	0	0	0	0
Polygonum_aviculare	Ter	G	0	0	0	0	0	0	0	0
Populus_alba	Ter	G	0	0	0	0	0	0	0	0
Potamogeton_berchtdii	Hy	B	0	0	0	0	0	0	0	0
Potamogeton_berchtdii	Hy	G	0	0	0	0	0	0	0	0
Potamogeton_berchtdii	Hy	R	0	0	0	0	0	0	0	0
Potamogeton_crispus	Hy	0	0	0	0	0	0	0	0	0
Potamogeton_lucens	Hy	0	0	0	0	0	0	0	0	0
Potamogeton_nodosus	Hy	0	0	0	0	0	0	0	0	0
Potamogeton_pectinatus	Hy	R	0	106,105	0	0	0	0	0	0
Potamogeton_perfoliatus	Hy	G	0	0	0	0	0	0	0	0
Potamogeton_perfoliatus	Hy	R	0	0	0	0	0	0	0	0
Ranunculus_trichophyllus	Hy	0	0	0	0	0	0	0	0	0
Ranunculus_fluitans	Hy	0	0	0	0	0	0	0	0	0
Robinia_pseudacacia	Ter	G	0	0	0	0	0	0	0	0
Salix_sp.	Ter	G	0	0	35,3682	0	0	176,841	35,3682	0
Scirpus	He	0	0	0	0	0	0	0	0	0
Solanum_dulcamara	Ter	G	0	0	35,3682	0	0	0	0	0
Solidago_gigantea	Ter	G	0	0	0	0	0	0	0	0
Sparganium_emersum	Hy	R	0	0	0	0	0	0	0	0
Spirodella_polyrhiza	Hy	F	0	0	0	0	0	0	0	0
Spirodella_polyrhiza	Hy	G	0	0	0	0	0	0	0	0
Spirogyra	Hy	0	0	0	0	0	0	0	0	0
Urtica_dioica	Ter	G	0	0	0	0	0	0	0	0
Vaucheria	Hy	0	0	0	0	0	0	0	0	0
Veronica_anagallis	He	0	0	0	0	0	0	0	0	0
Veronica_beccabunga	He	G	0	0	0	0	0	70,7364	0	0
Zanichellia_palustris	Hy	F	0	0	0	35,3682	0	0	0	0

EISV3	EISV4	EISV5	SCHA1	SCHA2	SCHA3	SCHA4	SCHA5	SCHOL1	SCHOL2	SCHOL3	SCHOL4
0	0	0	0	0	0	0	35,3682	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	22,222
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	35,3682	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	35,3682	0	0
70,7364	0	0	0	0	35,3682	0	0	0	0	0	70,7364
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	70,7364	0
0	0	0	0	0	0	0	22,222	44,444	0	0	22,222
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	35,3682	70,7364	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	35,3682	0
0	0	0	0	0	0	0	0	0	0	0	0

SCHOL5	ROSS1	ROSS2	ROSS3	ROSS4	ROSS5	FAHR1	FAHR2	FAHR3	FAHR4	FAHR5
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
35,3682	0	0	0	0	70,7364	70,7364	0	35,3682	70,7364	141,473
0	0	0	0	0	0	0	0	0	0	0
35,3682	0	70,7364	0	0	35,3682	70,7364	0	70,7364	0	70,7364
0	35,3682	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	35,3682	35,3682	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	35,3682	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
70,7364	0	0	0	106,105	35,3682	0	0	0	70,7364	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	35,3682	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	70,7364	35,3682	70,7364	0	0	0	0	0	0	0
35,3682	0	0	35,3682	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0

I.2. La végétation

Espèces	Type	Diaspore	EISM1			EISM2		
			Printemps	Eté	Automne	Printemps	Eté	Automne
Acer_pseudoplatanus	Ter	G	0	0	0	0	0	0
Acorus_calamus	He	R	0	0	0	0	0	0
Agrostis_stolonifera	Ter	0	0	0	0	0	0	1
Algues_filamenteuse	Hy	0	0	0	0	0	0	0
Azolla	Hy	0	0	0	0	0	0	0
Berula_erecta	Hy	0	0	0	0	0	0	0
Bryophyte	Hy	0	0	0	0	0	0	0
Callitriche_obtusangula	Hy	G	0	0	0	0	0	0
Carex_elata	He	R	0	0	1	0	0	0
Ceratophyllum_demersum	Hy	F	0	0	0	0	0	0
Chara_vulgaris	Hy	G	0	0	0	0	0	0
Cladophora	Hy	0	0,1	0,5	1	1	1	3
Dicot_herbacée	Ter	G	0	0	0	0	0	0
Dicot_terrestres	Ter	F	0	0	0	0	0	0
Dicot_phanérophyte	Ter	G	0	0	0	0	0	0
Echinocloa_crus_galli	Ter	G	0	0	0	0	0	0
Elodea_nuttallii	Hy	F	0	0	0	0	0	0,5
Enteromorpha	Hy	0	0	0	0	0	0	0
Epilobium_hirsutum	Ter	G	0	0	0	0	0	0
Equisetum_silvaticum	He	G	0	0	0	0	0	0
Fontinallis	Hy	0	0,5	1	1	1	1	2
Glyceria_fluitans	He	0	0	0	0	0	0	0
Graminées	Ter	0	0	0	0	1	0	0
Hildenbrandia	Hy	0	0	0,5	0	0	0,5	1
Hydrodictyon	Hy	0	0	0	0	0	0	0
Hypericum_perforatum	Ter	G	0	0	0	0	0	0
Iris_pseudacorus	He	R	0	0	0	0,1	1	1
Juncus_articulatus	He	G	0	0	0	0	0	0
Juncus_effusus	He	G	0	0	0	0	0	0
Juncus_sp._(petit)	He	G	0	0	0	0	0	0
Lemna_minor	Hy	F	0	0	0	0	0	0
Lemna_minuta	Hy	0	0	0	0	0	0	0
Lythrum_salicaria	He	G	0	0	0	0	0	0
Melosira	Hy	0	0	0	0	0	0,5	0
Mentha_aquatica	He	0	0	0	0	0	0	0
Myosotis_palustris	He	0	0	0	0	0	0	0
Myriophyllum_spicatum	Hy	0	0	0	0	0	0	1
Myriophyllum_verticillatum	Hy	0	0	0	0	0	0	0
Nastutium_officinale	He	0	0	0	0	0	0	0
Nitella_	Hy	0	0	0	0	0	0	0
Phalaris_arundinacea	He	G	0	0	0	0	0	0
Phalaris_arundinacea	He	R	1	0	0	0,1	0	0
Phragmites	He	0	0	0	0	0	0	0

Plantago_major	Ter	G	0	0	0	0	0	0
Poacées	Ter	G	0	0	0	0	0	0
Polygonum_aviculare	Ter	G	0	0	0	0	0	0
Populus_alba	Ter	G	0	0	0	0	0	0
Potamogeton_berchtoldii	Hy	B	0	0	0	0	0	0
Potamogeton_berchtoldii	Hy	G	0	0	0	0	0	0
Potamogeton_berchtoldii	Hy	R	0	0	0	0	0	0
Potamogeton_crispus	Hy	0	0	0	0	0	0	0
Potamogeton_lucens	Hy	0	0	0	0	0	0	0
Potamogeton_nodosus	Hy	0	0	0	0	0	0	0
Potamogeton_pectinatus	Hy	R	0	0,5	0	0	3	1
Potamogeton_perfoliatus	Hy	G	0	0	0	0	0	0
Potamogeton_perfoliatus	Hy	R	0	0	0	0	0	0
Ranunculus_trichophyllus	Hy	0	0	0	0	0	0	0
Ranunculus_fluitans	Hy	0	0	0	0	0	0	5
Robinia_pseudacacia	Ter	G	0	0	0	0	0	0
Salix_sp.	Ter	G	0	0	0	0	0	0
Scirpus	He	0	0	0	0	0	0	0
Solanum_dulcamara	Ter	G	0	0	0	0	0	0
Solidago_gigantea	Ter	G	0	0	0	0	0	0
Sparganium_emersum	Hy	R	0	0	0	0	0	0
Spirodella_polyrhiza	Hy	F	0	0	0	0	0	0
Spirodella_polyrhiza	Hy	G	0	0	0	0	0	0
Spirogyra	Hy	0	0	0	0	0	0	0
Urtica_dioica	Ter	G	0	0	0	0	0	0
Vaucheria	Hy	0	0	0	0	0	0	0
Veronica_anagallis	He	0	0	0	0	0	0	0
Veronica_beccabunga	He	G	0	0	0	0,1	0	0
Zanichellia_palustris	Hy	F	0	0	0	0	0	0

EISM3			EISM4			EISM5			EISV1		
Printemps	Été	Automne	Printemps	Été	Automne	Printemps	Été	Automne	Printemps	Été	Automne
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0,5	0,5	0,5	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	7	0,5	1	15	0,5	0,5	1	0,5	0,5
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0,5	0	0,5	0	0	0	0,5
0	0	0	0	0	0,5	0	0,5	0	0	0	0,5
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0,01	0	0
0	0	0	0	0	0	0	0	0	0	0	0

EISV2			EISV3			EISV4			EISV5		
Printemps	Eté	Automne	Printemps	Eté	Automne	Printemps	Eté	Automne	Printemps	Eté	Automne
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0,5
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0,5	0	0	0	0,5	0	0	0
0,01	0,5	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0,5	0,5	1	1	1	0,01	0,5	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
1	0	0,5	2	0,5	2	1	0,5	0,5	0,02	0,5	0,5
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0,01	0	0	0	0	0	0	0,5	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0,5	0,5	0,5	0	0,5	0,5	0,01	0,5	0,5	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0,5	0
0	0	0	0	0	0	0,02	0	0	0	0	0

0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
1	NA	0	5	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0,5
7	NA	0	15	NA	0	0	0	0	0	0	0	0
7	NA	0	15	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	1	0,5	0,5	0,1	0	3	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0,5	NA	0	0	0	0	0	0	0	0
0	NA	0,5	0	NA	0	0	0	0	0	0	0	0
0,01	NA	0	0,01	NA	0	1	0	0	25	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0
0	NA	0	0	NA	0	0	0,5	0	0	0	0	0
0	NA	0	0	NA	0	0	0,5	0	0	0	0	0
0	NA	0	0	NA	0	0	0	0	0	0	0	0

ROSS3			ROSS4			ROSS5			FAHR1		
Printemps	Eté	Automne	Printemps	Eté	Automne	Printemps	Eté	Automne	Printemps	Eté	Automne
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0,5	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0,5	0	0	0	3	0	0,5	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
2	0	0,5	0,1	0	0	0,2	0,5	10	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0

10	0	0	2	0	0	2	0	0	0	NA	0
0	0	0	0	0	0	0	0,5	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0,5
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
0	0	0	0	0	0	0	0	0	0	NA	0
FAHR2			FAHR3			FAHR4			FAHR5		
Printemps	Été	Automne	Printemps	Été	Automne	Printemps	Été	Automne	Printemps	Été	Automne
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0,5	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0,1	NA	1
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0,5	0	NA	0,5	0	NA	0	0,1	NA	1
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0

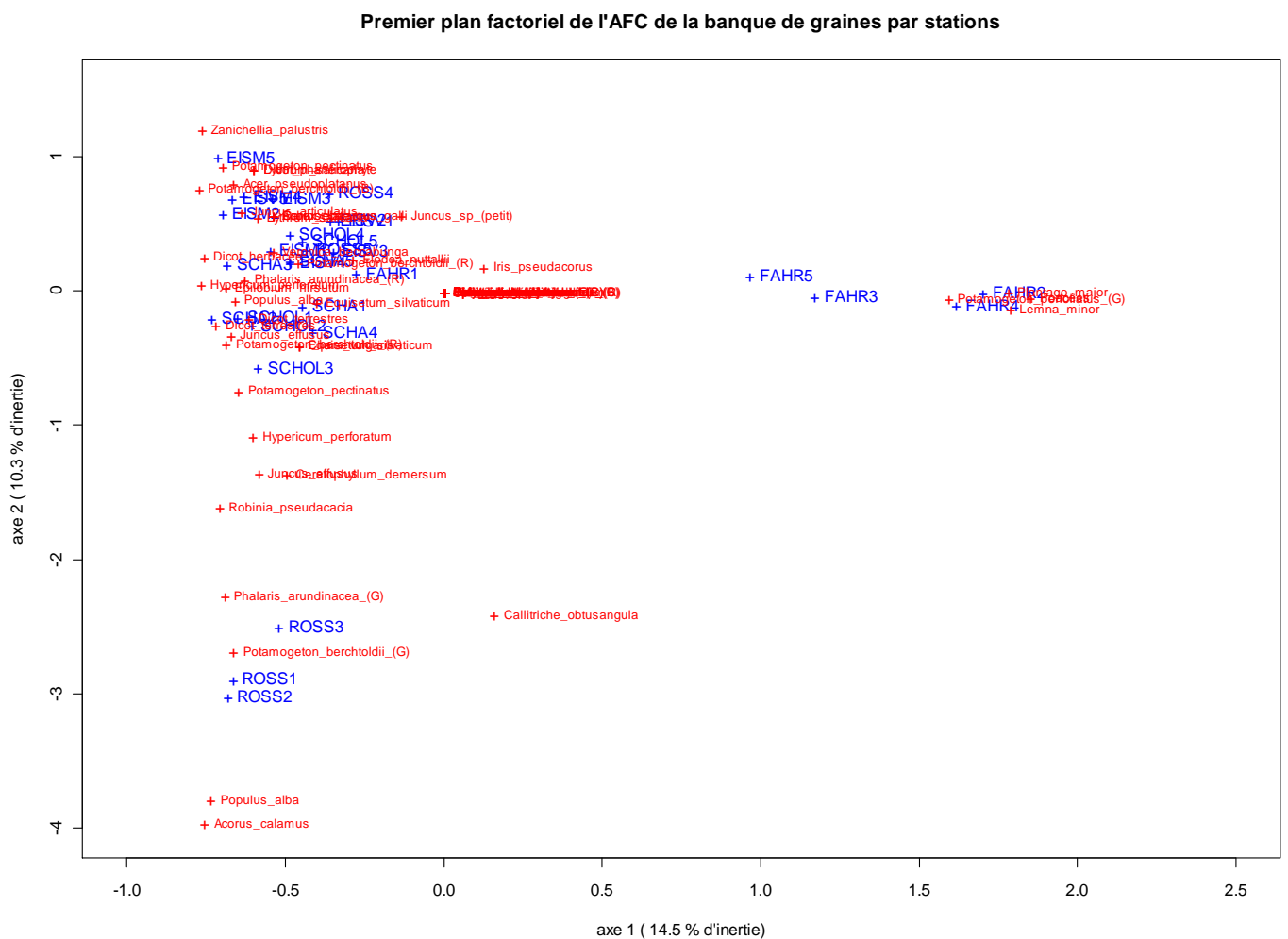
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0,2	NA	0	1	NA	0	2	NA	0	1	NA	0
0,2	NA	0	1	NA	0	2	NA	0	1	NA	0
0,2	NA	0	1	NA	0	2	NA	0	1	NA	0
0	NA	0	0,1	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0,5	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0,5	NA	0	0	NA	0	0	NA	0,5
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0,5	0	NA	2	0	NA	1	0	NA	2
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0
0	NA	0	0	NA	0	0	NA	0	0	NA	0

II. Les analyses factorielles des correspondances

II.1. Premier plan factoriel de l'AFC de la banque de graines par stations

Les commandes R pour tracer ce graphe sont les suivantes :

```
plot ( AFC_stations$co[,1], AFC_stations$co[,2], main="Premier plan factoriel de l'AFC de
la banque de graines par stations", pch="+", xlab=paste( "axe 1 ( ", ceiling (
inertie_stations$TOT[1,3]*1000)/10, "% d'inertie ) " ), ylab=paste ( "axe 2 ( ",
ceiling ( ( inertie_stations$TOT[2,3] - inertie_stations$TOT[1,3])*1000)/10, "%
d'inertie ) " ), xlim=c ( -1,2.5), ylim=c ( -4,1.5 ), col="blue" )
text ( AFC_stations$co[,1], AFC_stations$co[,2], names ( banque[-1] ), pos=4, col="blue" )
points ( AFC_stations$li[,1], AFC_stations$li[,2], pch="+", col="red" )
text ( AFC_stations$li[,1], AFC_stations$li[,2], banque[,1], pos=4, cex=0.75, col="red" )
```



III. Indices de Ruzicka et intervalles de confiance bootstrap

Comparaison entre banque de chaque station et végétation correspondante (Ruzicka)

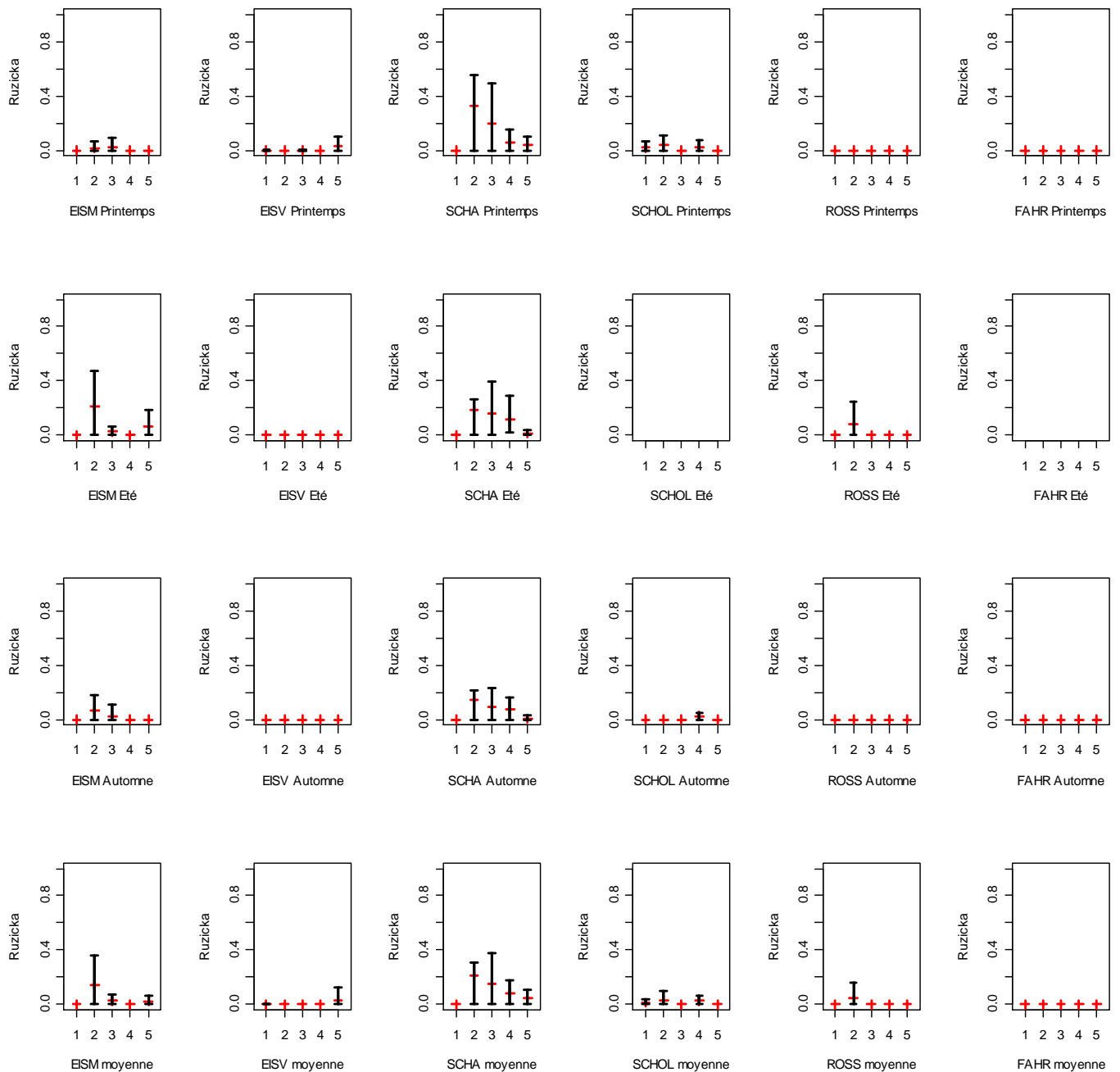


Figure 24 : indices de Ruzicka et intervalles de confiance pour la comparaison entre banque de graines de chaque station et végétation correspondante

IV. Bassin d'orage

IV.1. Banque de graines

Espèce	Diaspore	Station 1	Station 2	Station 3	Station 4	Station 5
Dicot	G	23,5849057	0	0	0	0
Echinochloa_crus-galli	G	0	0	0	47,1698113	117,924528
Epilobium_hirsutum	G	0	0	0	0	0
Equisetum	G	23,5849057	0	0	0	0
Glyceria_fluitans	G	0	0	0	0	23,5849057
Lemna_minor	G	0	0	0	0	0
Matricaria_discoidea	G	0	0	0	0	0
Medicago_sativa	G	0	0	0	0	0
Melilotus_albus	G	0	0	0	0	0
Plantago_major	G	0	0	0	0	0
Poa_trivialis	G	0	0	0	0	0
Poacée	G	0	0	0	0	0
Polygonum_aviculare	G	0	0	0	0	0
Populus_sp	G	0	0	0	0	0
Ranunculus_repens	G	0	23,5849057	0	0	47,1698113
Sparganium_erectum	G	0	0	0	0	0
Epilobium_hirsutum	R	0	0	0	0	0
Equisetum_arvensis	R	0	0	0	0	0
Phragmites_australis	R	181,481481	122,222222	181,481481	133,333333	140,740741
Algue_filamenteuse	0	0	0	0	0	0
Hedera_helix	0	0	0	0	0	0
cornus_sp.	0	0	0	0	0	0
salix_alba	0	0	0	0	0	0
Urtica_dioica	0	0	0	0	0	0
Eupatorium_cannabinum	0	0	0	0	0	0
Gallium_aparine	0	0	0	0	0	0
Apiaceae	0	0	0	0	0	0
Triticum_aestivum	0	0	0	0	0	0
Rubus_sp.	0	0	0	0	0	0
Artemisia_vulgaris	0	0	0	0	0	0
Rosa_canina	0	0	0	0	0	0
Asteracea	0	0	0	0	0	0
Medicago_lupulina	0	0	0	0	0	0
Torrilis	0	0	0	0	0	0

Station 6	Station 7	Station 8	Station 9	Station 10	Station 11	Station 12	Station 13
0	0	0	0	0	0	23,5849057	23,5849057
0	0	23,5849057	70,754717	23,5849057	23,5849057	23,5849057	0
0	0	0	23,5849057	0	0	0	0
23,5849057	0	0	0	0	0	0	0
0	0	0	70,754717	0	0	23,5849057	0
0	0	0	0	0	0	23,5849057	0
0	0	0	0	0	0	94,3396226	47,1698113

IV.2. Végétation

Espèce	Diaspore	Station 1		Station 2		Station 3	
		Printemps	Eté	Printemps	Eté	Printemps	Eté
Dicot	G	0	0	0	0	0	0
Echinochloa_crus-galli	G	0	0	0	0	0	0
Epilobium_hirsutum	G	0	0	0	0	0	0
Equisetum	G	0	0	0	0	0	0
Glyceria_fluitans	G	0	0	0	0	0	0
Lemna_minor	G	0	0	0	0	0	0
Matricaria_discoidea	G	0	0	0	0	0	0
Medicago_sativa	G	0	0	0	0	0	0
Melilotus_albus	G	0	0	0	0	0	0
Plantago_major	G	0	0	0	0	0	0
Poa_trivialis	G	0	0	0	0	0	0
Poacée	G	0	0	0	0	0	0
Polygonum_aviculare	G	0	0	0	0	0	0
Populus_sp	G	0	0	0	0	0	0
Ranunculus_repens	G	0	0	0	0	0	0
Sparganium_erectum	G	0	0	0	0	0	0
Epilobium_hirsutum	R	0	0	0	0	0	0
Equisetum_arvensis	R	0	0	0	0	0	0
Phragmites_australis	R	100	100	100	95	80	99
Algue_filamenteuse	0	5	0	5	1	25	1
Hedera_helix	0	0	0	0	0	0	0
cornus_sp.	0	0	0	0	0	0	0
salix_alba	0	0	0	0	0	0	0
Urtica_dioica	0	0	0	0	0	0	0
Eupatorium_canbinum	0	0	0	0	0	0	0
Gallium_aparine	0	0	0	0	0	0	0
Apiaceae	0	0	0	0	0	0	0
Triticum_aestivum	0	0	0	0	0	0	0
Rubus_sp.	0	0	0	0	0	0	0
Artemisia_vulgaris	0	0	0	0	0	0	0
Rosa_cani	0	0	0	0	0	0	0
Asteracea	0	0	0	0	0	0	0
Medicago_lupili	0	0	0	0	0	0	0
Torrilis	0	0	0	0	0	0	0

Station 4		Station 5		Station 6		Station 7		Station 8	
Printemps	Eté	Printemps	Eté	Printemps	Eté	Printemps	Eté	Printemps	Eté
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0,1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0,1	0	2	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0,1	0,1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
100	99	100	95	100	100	95	99,5	95	99,5
5	0,5	0	0	1	0	0	0,5	0	0,5
0	0	0	0	0,1	0	0	0	0	0
0	0	0	0	0,1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0,1	0	0	0	0	0	0	0	0
0	0	0	0,1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Station 9		Station 10		Station 11		Station 12		Station 13	
Printemps	Eté	Printemps	Eté	Printemps	Eté	Printemps	Eté	Printemps	Eté
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
2	2	0	0	0	0	2	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0,5	0,1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0,5	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
100	97	100	100	100	100	100	98	100	100
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0,1	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0,5	0,5	0	0

0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Station 14		Station 15		Station 16		Station 17	
Printemps	Eté	Printemps	Eté	Printemps	Eté	Printemps	Eté
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0,1	0,5	0,5	0,5	0,5	0,5
0	0	0	0,1	0	0	0,5	0,5
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	2	0	0,5	0	0,5	0
0	0	1	0	0	0,5	0	0,5
0	0	0,5	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
100	100	95	97	100	99	100	99
0	0	0	0	0	0	0,5	0,1
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	1	0,1	0	0	0	0
0	0	5	2	0	0	0	0
0	0	0,5	1	0	0	0	0
0	0	3	0	0	0	0,1	0
0	0	0,5	0	0	0	0	0
0	0	0,5	0	0	0	0	0
0	0	0,1	0	0	0	0	0
0	0	0	0	0,5	0	0	0
0	0	0	0	0	0	0,1	0,1
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0,5	0	0	0	0

IV.3. AFC de la banque de graines du bassin d'orage

IV.3.1. Valeurs propres

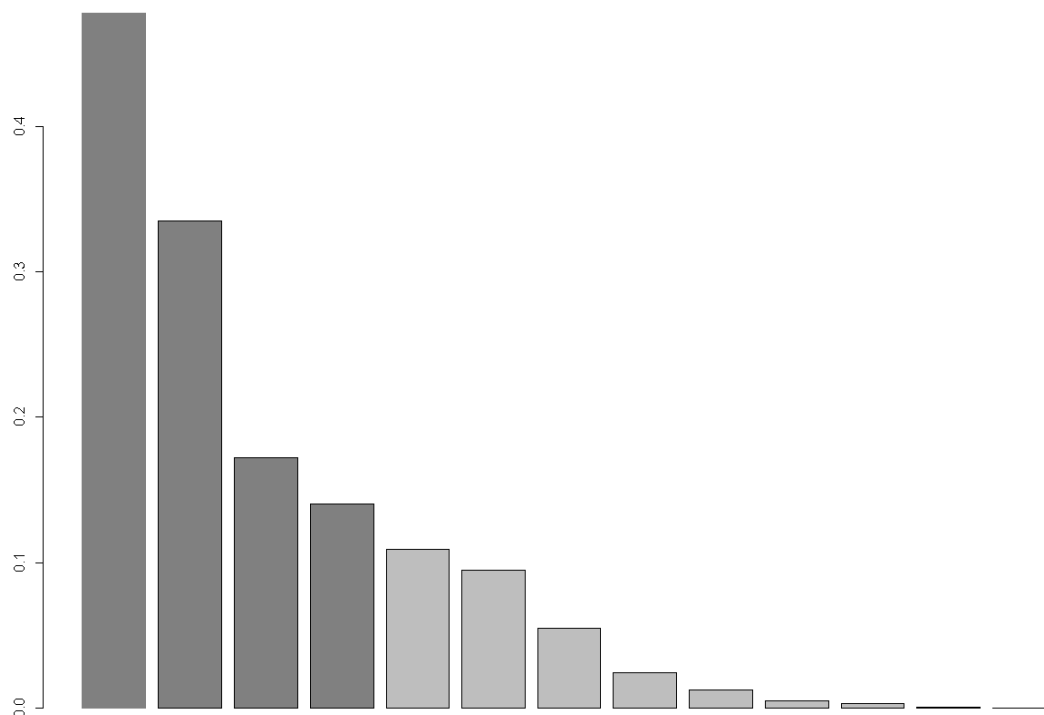


Figure 25 : graphe en barres des valeurs propres de l'AFC de la banque de graines du bassin d'orage

Axes	Valeur propre	Pourcentage de variation cumulée
1	0,477680239	33,42%
2	0,335202242	56,87%
3	0,172339338	68,93%
4	0,140314512	78,74%
5	0,108830855	86,36%
6	0,094891428	93%
7	0,054836669	96,83%
8	0,02410945	98,52%
9	0,012403022	99,39%
10	0,00482524	99,73%
11	0,002927164	99,93%
12	0,000820763	99,99%
13	0,000124689	100%

Tableau 13 : valeurs propres de l'AFC de la banque de graines du bassin d'orage

IV.3.2. Distances des colonnes aux axes

	Axe 1	Axe 2	Axe 3	Axe 4	Autres
Station 1	-3993	-2325	-288	67	617
Station 2	-5941	-851	-666	21	206
Station 3	-5898	-2093	-374	121	356
Station 4	-5335	-654	-26	-1303	276
Station 5	-2545	152	118	-4139	536
Station 6	-5091	-2323	-710	282	692
Station 7	-5898	-2093	-374	121	174
Station 8	-7631	-246	-223	-27	180
Station 9	-922	3166	1183	-21	494
Station 10	-4203	97	52	1109	219
Station 11	-6426	-3	-352	-94	198
Station 12	20	3087	2007	4011	907
Station 13	370	-2027	409	1369	79
Station 14	-972	-803	2033	60	876
Station 15	3027	-234	2255	-2624	916
Station 16	178	6321	-2706	-89	1359
Station 17	6426	-2488	-470	174	1916

Tableau 14 : distance des colonnes aux axes

IV.3.3. Contributions des colonnes aux axes

	Axe 1	Axe 2	Axe 3	Axe 4
Station 1	737	612	147	42
Station 2	366	75	114	4
Station 3	628	318	111	44
Station 4	441	77	6	366
Station 5	408	35	53	2258
Station 6	1054	685	408	199
Station 7	308	156	54	22
Station 8	410	19	33	5
Station 9	136	666	484	10
Station 10	276	9	9	248
Station 11	381	0	58	19
Station 12	5	1194	1510	3707
Station 13	9	68	27	110
Station 14	255	300	1476	54
Station 15	830	92	1713	2448
Station 16	72	3662	3049	124
Station 17	3684	2033	747	340
Total (%)	69,33	68,89	77,48	84,13

Tableau 15 : contributions des colonnes aux axes

IV.3.4. Premier plan factoriel de l'AFC de la banque de graines

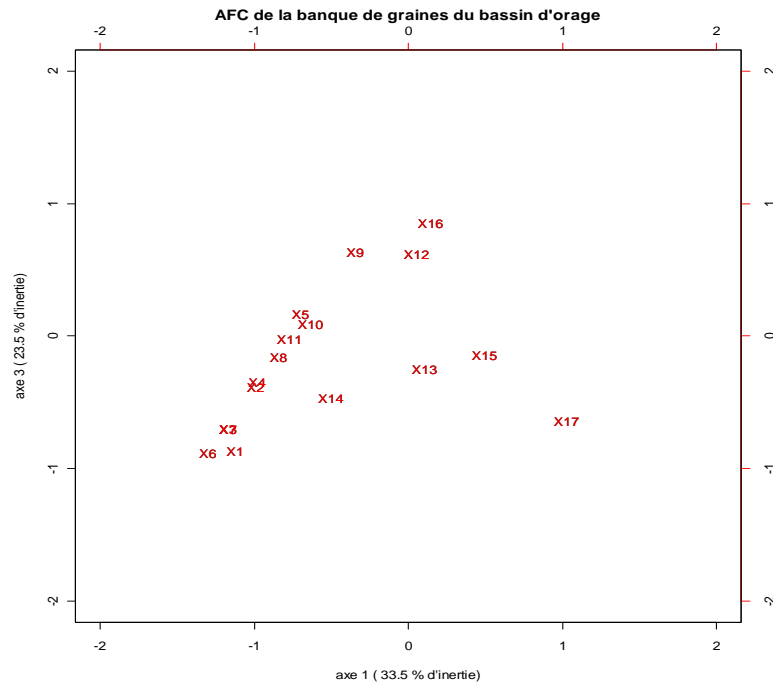


Figure 26 : Premier plan factoriel de l'AFC de la banque de graines du bassin d'orage

IV.4. Classification ascendante hiérarchique

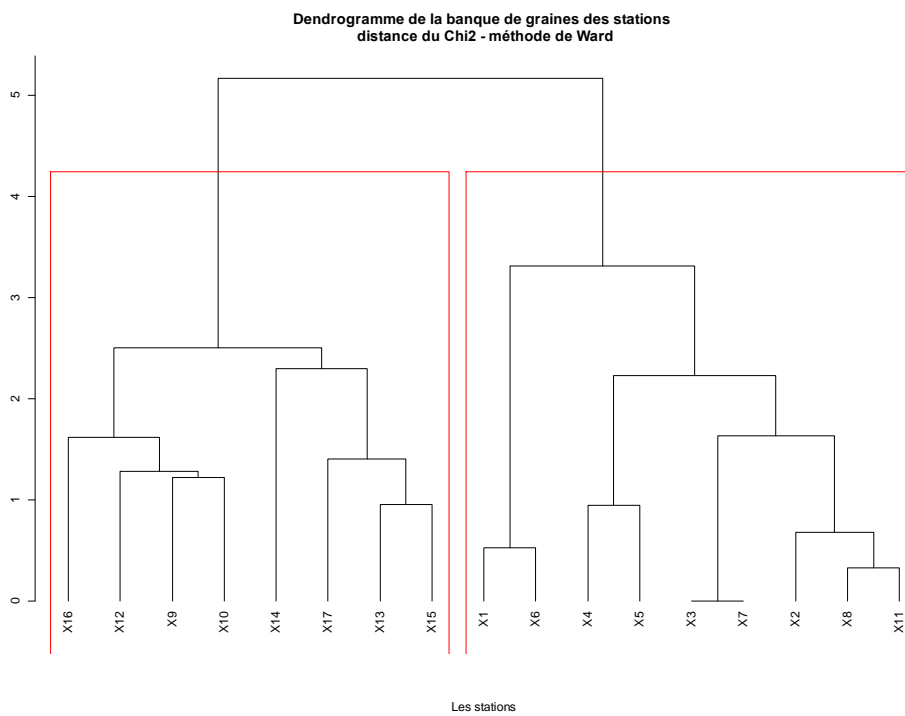


Figure 27 : dendrogramme de la banque de graines des stations du bassin d'orage

On remarque sur les deux figures précédentes une nette séparation entre les stations situées dans la zone de décantation (stations 9 à 17) et celles situées dans la zone de filtration (stations 1 à 8).

IV.4. Indices et intervalles de confiance bootstrap

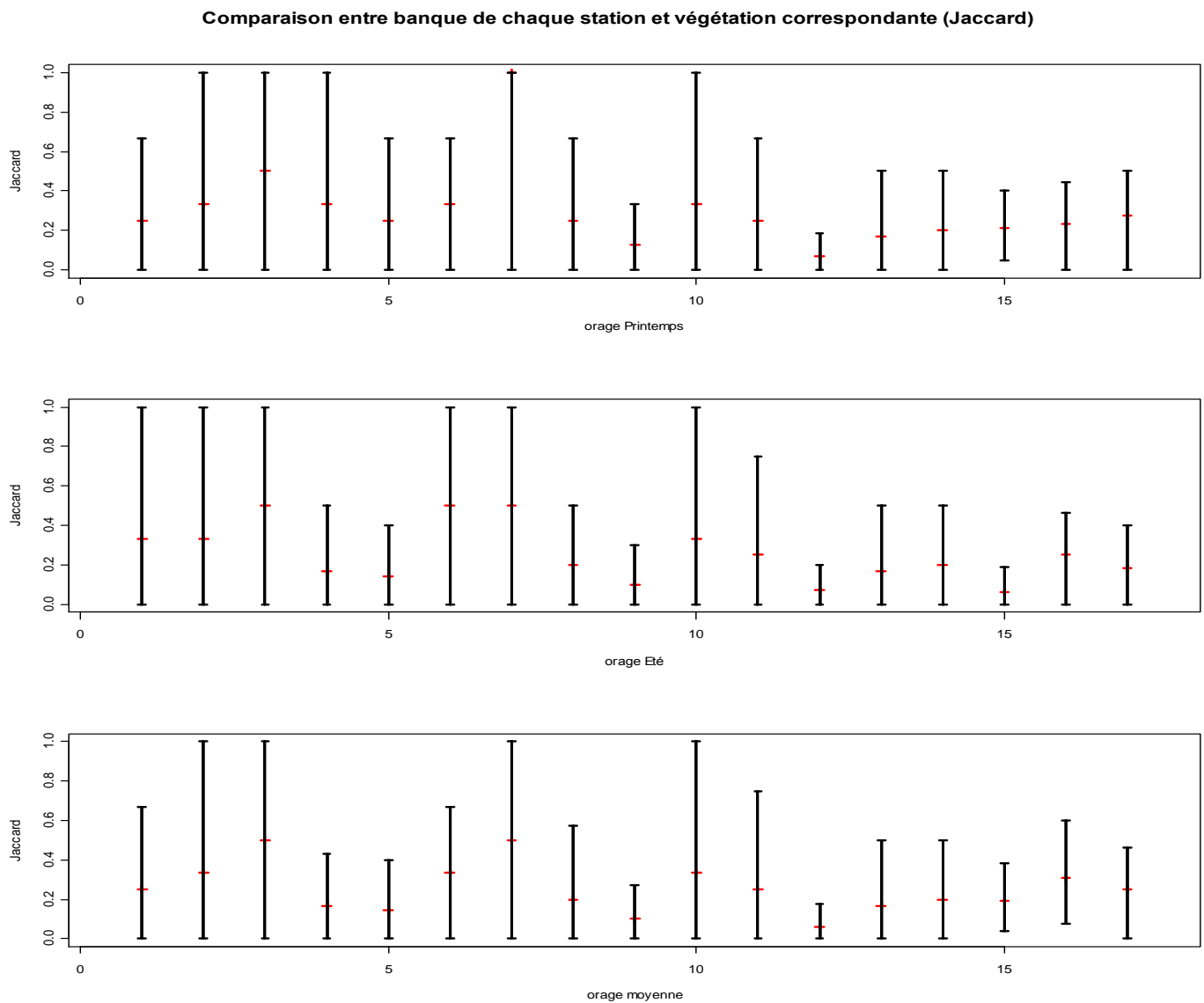


Figure 28 : indices de Jaccard et intervalles de confiance bootstrap associés

Les intervalles de confiances sont très larges à cause du manque d'espèces (19 espèces) présentes sur le site du bassin d'orage. La connexion de ce bassin étant récente, ce problème pourrait être réglé d'ici quelques années.

V. Le script du programme

```
library(boot)           # pour pouvoir utiliser la fonction "boot" pour le bootstrap
library(xlsReadWrite)  # pour pouvoir lire et enregistrer au format ".xls" (Excel)
library(ade4)          # pour les AFC et CAH

### Cibles pour l'enregistrement

dirJaccard="C:/Documents and Settings/étudiant/Bureau/Comparaisons banque végétation/Intervalles de
confiance/Jaccard"
dirRuzicka="C:/Documents and Settings/étudiant/Bureau/Comparaisons banque végétation/Intervalles de
confiance/Ruzicka"

### Importation des jeux de données

banque<-read.xls("C:/Documents and Settings/étudiant/Bureau/Comparaisons banque végétation/Banque et
végétation/banque.xls",colNames=TRUE,sheet=1)
vegetation<-read.xls("C:/Documents and Settings/étudiant/Bureau/Comparaisons banque végétation/Banque et
végétation/vegetation.xls",colNames=TRUE,sheet=1)

### Nombre de sites et nombre de stations par site

print("")
print("Entrer le nombre de sites : ");nb_sites<-as.numeric(readline())
print("Entrer le nombre de stations par site : ");nb_stations<-as.numeric(readline())

### Noms des sites

print("Entrer les noms des sites : ")
print("1 : EISM, EISV, SCHA, SCHOL, ROSS, FAHR (6 sites)")
print("2 : Bassin d'orage (1 site)")
print("3 : Autres");noms_sites<-as.numeric(readline())

if (noms_sites==1) {
  if (nb_sites!=6) {
    print("Erreur : le nombre de sites ne correspond pas!")
    break
  } else {
    noms<-c("EISM","EISV","SCHA","SCHOL","ROSS","FAHR")
  }
} else {
if (noms_sites==2) {
  if (nb_sites!=1) {
    print("Erreur : le nombre de sites ne correspond pas!")
    break
  } else {
    noms<-"orage"
  }
} else {
if (noms_sites==3) {
  noms<-c()
  print("Entrer le nom du 1er site : ");nom_sites<-as.character(readline())
  noms<-c(noms,nom_sites)
  for (i in 2:nb_sites) {
    print(paste("Entrer le nom du ",i,"ème site : "));nom_sites<-as.character(readline())
    noms<-c(noms,nom_sites)
  }
} else {
```



```

    print("Erreur : nom inconnu!")
    break;
  }
}

# Nombre et noms des saisons :

print("Entrer le nombre de saisons : ");nb_saison<-as.numeric(readline())

print("Entrer les noms des saisons : ")
print("1 : Printemps,Eté,Automne (3 saisons)")
print("2 : Printemps,Eté (2 saisons)")
print("3 : Autres");season<-as.numeric(readline())

if (season==1) {
  if (nb_saison!=3) {
    print("Erreur : le nombre de saisons ne correspond pas!")
    break
  } else {
    saison<-c("Printemps","Eté","Automne")
  }
}
if (season==2) {
  if (nb_saison!=2) {
    print("Erreur : le nombre de saisons ne correspond pas!")
    break
  } else {
    saison<-c("Printemps","Eté")
  }
}
if (season==3) {
  saison<-c()
  print("Entrer le nom de la 1ère saison : ");nom_saison<-as.character(readline())
  saison<-c(saison,nom_saison)
  for (i in 2:nb_saison) {
    print(paste("Entrer le nom de la ",i,"ème saison : "));nom_saison<-as.character(readline())
    saison<-c(saison,nom_saison)
  }
}

### Type (Terrestre, Hydrophyte ou Hélophyte)

print("Entrer le type : ")
print("Terrestre : Ter , Hydrophyte : Hy , Hélophyte : He , Total : all")
print("Terrestre et Hydrophyte : TerHy , Terrestre et Hélophyte : TerHe , Hydrophyte et Hélophyte :
HyHe");type<-as.character(readline())

if (type=="all") {
  } else {
if (type=="Ter") {
  banque<-banque[banque$Type=="Ter",]
  vegetation<-vegetation[vegetation$Type=="Ter",]
  } else {
if (type=="Hy") {
  banque<-banque[banque$Type=="Hy",]
  vegetation<-vegetation[vegetation$Type=="Hy",]
  } else {
if (type=="He") {

```

```

banque<-banque[banque$Type=="He",]
vegetation<-vegetation[vegetation$Type=="He",]
} else {
if (type=="TerHy") {
banque<-banque[banque$Type!="He",]
vegetation<-vegetation[vegetation$Type!="He",]
} else {
if (type=="TerHe") {
banque<-banque[banque$Type!="Hy",]
vegetation<-vegetation[vegetation$Type!="Hy",]
} else {
if (type=="HyHe") {
banque<-banque[banque$Type!="Ter",]
vegetation<-vegetation[vegetation$Type!="Ter",]
} else {
print("Erreur : type non reconnu!")
break
}}}}}}

```

Diaspore (Graine, Bourgeon dormant, Rhizome, fragment ou Bulbe)

```

print("Entrer la diaspore")
print("Graine : G ; Bourgeon dormant : Bd ; Rhizome : R ; Fragment : F ; Bulbe : Bl ; Total : all")
print("Graine et Bourgeon dormant : GBd ; Graine et Rhizome : GR ; Graine et Fragment : GF ; Graine et
Bulbe : GBBl")
print("Bourgeon dormant et Rhizome : BdR ; Bourgeon dormant et Fragment : BdF ; Bourgeon dormant et
Bulbe : BdBl")
print("Rhizome et Fragment : RF ; Rhizome et Bulbe : RBl ; Fragment et Bulbe : FBBl")
print("Graine, Bourgeon dormant et Rhizome : GBdR ; Graine, Bourgeon dormant et Fragment : GBdF")
print("Graine, Bourgeon dormant et Bulbe : GBdBl ; Graine, Rhizome et Fragment : GRF ; Graine, Rhizome
et Bulbe : GRBl")
print("Graine, Fragment et Bulbe : GFBl ; Bourgeon dormant, Rhizome et Fragment : BdRF ; Bourgeon
dormant, Rhizome et Bulbe : BdRBl")
print("Bourgeon dormant, Fragment et Bulbe : BdFBl ; Rhizome, Fragment et Bulbe : RFBl ; Graine,
Bourgeon dormant, Rhizome et Fragment : GBdRF")
print("Graine, Bourgeon dormant, Rhizome et Bulbe : GBdRBl ; Graine, Bourgeon dormant, Fragment et Bulbe
: GBdFBl")
print("Graine, Rhizome, Fragment et Bulbe : GRFBl ; Bourgeon dormant, Rhizome, Fragment et Bulbe :
BdRFBl");diasp<-as.character(readline())

if (diasp=="all") {
} else {
if (diasp=="G") {
banque<-banque[banque$Diaspore=="G",]
vegetation<-vegetation[vegetation$Diaspore=="G",]
} else {
if (diasp=="Bd") {
banque<-banque[banque$Diaspore=="Bd",]
vegetation<-vegetation[vegetation$Diaspore=="Bd",]
} else {
if (diasp=="R") {
banque<-banque[banque$Diaspore=="R",]
vegetation<-vegetation[vegetation$Diaspore=="R",]
} else {
if (diasp=="F") {
banque<-banque[banque$Diaspore=="F",]
vegetation<-vegetation[vegetation$Diaspore=="F",]
} else {
if (diasp=="Bl") {

```

```

banque<-banque[banque$Diaspore=="Bl",]
vegetation<-vegetation[vegetation$Diaspore=="Bl",]
} else {
if (diasp=="GBd") {
banque<-banque[banque$Diaspore=="G" & banque$diaspore=="Bd",]
vegetation<-vegetation[vegetation$Diaspore=="G" & vegetation$diaspore=="Bd",]
} else {
if (diasp=="GR") {
banque<-banque[banque$Diaspore=="G" & banque$diaspore=="R",]
vegetation<-vegetation[vegetation$Diaspore=="G" & vegetation$diaspore=="R",]
} else {
if (diasp=="GF") {
banque<-banque[banque$Diaspore=="G" & banque$diaspore=="F",]
vegetation<-vegetation[vegetation$Diaspore=="G" & vegetation$diaspore=="F",]
} else {
if (diasp=="GBI") {
banque<-banque[banque$Diaspore=="G" & banque$diaspore=="Bl",]
vegetation<-vegetation[vegetation$Diaspore=="G" & vegetation$diaspore=="Bl",]
} else {
if (diasp=="BdR") {
banque<-banque[banque$Diaspore=="Bd" & banque$diaspore=="R",]
vegetation<-vegetation[vegetation$Diaspore=="Bd" & vegetation$diaspore=="R",]
} else {
if (diasp=="BdF") {
banque<-banque[banque$Diaspore=="Bd" & banque$diaspore=="F",]
vegetation<-vegetation[vegetation$Diaspore=="Bd" & vegetation$diaspore=="F",]
} else {
if (diasp=="BdBl") {
banque<-banque[banque$Diaspore=="Bd" & banque$diaspore=="Bl",]
vegetation<-vegetation[vegetation$Diaspore=="Bd" & vegetation$diaspore=="B",]
} else {
if (diasp=="RF") {
banque<-banque[banque$Diaspore=="R" & banque$diaspore=="F",]
vegetation<-vegetation[vegetation$Diaspore=="R" & vegetation$diaspore=="F",]
} else {
if (diasp=="RBl") {
banque<-banque[banque$Diaspore=="R" & banque$diaspore=="Bl",]
vegetation<-vegetation[vegetation$Diaspore=="R" & vegetation$diaspore=="Bl",]
} else {
if (diasp=="FBl") {
banque<-banque[banque$Diaspore=="F" & banque$diaspore=="Bl",]
vegetation<-vegetation[vegetation$Diaspore=="F" & vegetation$diaspore=="Bl",]
} else {
if (diasp=="GBdR") {
banque<-banque[banque$Diaspore=="G" & banque$diaspore=="Bd" & banque$diaspore=="R",]
vegetation<-vegetation[vegetation$Diaspore=="G" & vegetation$diaspore=="Bd" &
vegetation$diaspore=="R",]
} else {
if (diasp=="GBdF") {
banque<-banque[banque$Diaspore=="G" & banque$diaspore=="Bd" & banque$diaspore=="F",]
vegetation<-vegetation[vegetation$Diaspore=="G" & vegetation$diaspore=="Bd" &
vegetation$diaspore=="F",]
} else {
if (diasp=="GBdBl") {
banque<-banque[banque$Diaspore=="G" & banque$diaspore=="Bd" & banque$diaspore=="Bl",]
vegetation<-vegetation[vegetation$Diaspore=="G" & vegetation$diaspore=="Bd" &
vegetation$diaspore=="Bl",]
} else {
if (diasp=="GRF") {

```



```

print("Entrer le nombre de réplifications bootstrap : ");B<-as.numeric(readline())
print("Entrer le risque alpha en pourcentage : ");alpha<-as.numeric(readline())

###

vect<-seq(1,2*(length(vegetation)),nb_saison+1)
vect2<-seq(4,4*(length(banque)),nb_saison+1)
vect3<-seq(1,2*(length(vegetation)),nb_saison)
vect4<-seq(1,3*nb_saison,3)

### Assemblage des données (banque + végétation -> donnees)

donnees<-banque
for (i in 1:(length(banque)-3)) {
  donnees[,vect[i]]<-banque[,3+i]
  colnames(donnees)[vect[i]]<-colnames(banque)[3+i]
  for (j in 1:nb_saison) {
    donnees[,vect[i]+j]<-vegetation[,vect3[i]+j+2]
    colnames(donnees)[vect[i]+j]<-colnames(vegetation)[vect3[i]+j+2]
  }
}

### Banque de graines des différents sites

site_banque<-matrix(0,length(donnees[,1]),nb_sites)
colnames(site_banque)<-noms
for (j in 1:nb_sites) {
  site_banque[,j]<-rowSums(banque[,(((nb_sites-1)*(j-1))+4):((1+(nb_sites-1)*(j-1))+
(nb_stations+2))],na.rm=TRUE)
}

stations_banque<-banque[,4:(nb_sites*nb_stations+3)]

stations_vegetation<-stations_banque
for (i in 1:(nb_sites*nb_stations)) {
  stations_vegetation[,i]<-rowMeans(vegetation[,:(vect3[i]+3):(vect3[i+1]+2)],na.rm=T)
}

### transformation des données (proportions)

for (j in 1:length(donnees[,1])) {
  if (is.na(donnees[,j])) {
  } else {
    somme<-sum(donnees[,j])
    for (i in 1:length(donnees[,1])) {
      donnees[i,j]<-ifelse(somme==0,0,donnees[i,j]/somme)
    }
  }
}

for (j in 4:length(banque)) {
  if (is.na(banque[,j])) {
  } else {
    somme<-sum(banque[,j])
    for (i in 1:length(banque[,1])) {
      banque[i,j]<-ifelse(somme==0,0,banque[i,j]/somme)
    }
  }
}

```

```

for (j in 4:length(vegetation)) {
  if (is.na(vegetation[1,j])) {
    } else {
    somme<-sum(vegetation[,j])
    for (i in 1:length(vegetation[,1])) {
      vegetation[i,j]<-ifelse(somme==0,0,vegetation[i,j]/somme)
    }
  }
}

```

```
lg<-(length(banque[1,])+length(vegetation[1,])-6)/nb_sites
```

```
### Extraction des sites (listes des stations sur chaque site)
```

```

banq<-list()
for (i in 1:nb_sites) {
  banq<-c(banq,list(donnees[,((i-1)*lg+1):(i*lg)]))
}

veget<-list()
for (i in 1:nb_sites) {
  veget<-c(veget,list(vegetation[,((4+(i-1)*nb_saison*nb_stations):(3+i*nb_saison*nb_stations)]))
}

```

```
### Indices
```

```

Jaccard<-function(x,y){
  a<-0
  b<-0
  d<-0
  l<-length(x)
  for(i in 1:l){
    if(x[i]!=0){
      a<-a+1
      if(y[i]!=0){
        b<-b+1
        d<-d+1
      }
    } else {
      if(y[i]!=0){
        b<-b+1
      }
    }
  }
  if (a+b-d==0) {
    Q<-0
  } else {
    Q<-d/(a+b-d)
  }
  return(Q)
}

```

```

Ruzicka<-function(x,y){
  a<-c()
  b<-c()
  l<-length(x)
  for(i in 1:l){
    a<-c(a,min(x[i],y[i]))
    b<-c(b,max(x[i],y[i]))
  }
}

```

```

    if (sum(b)==0) {
      Q<-0
    } else {
      Q<-sum(a)/sum(b)
    }
  }
  return(Q)
}

### AFC et dendogrammes

##### AFC stations

if (nb_stations>1) {
  if(nb_sites*nb_stations>7) {
    AFC_stations<-dudi.coa(stations_banque,scannf=FALSE,nf=10)
    inertie_stations<-inertia.dudi(AFC_stations)
    win.metafile("AFC banque stations.emf",width=10,height=10)
    par(mfrow=c(1,2))
    biplot(AFC_stations$co[,1:2],AFC_stations$co[,1:2],var.axes=F,xlab=paste("axe 1
      (" ,ceiling(inertie_stations$TOT[1,3]*1000)/10,"% d'inertie"),ylab=paste("axe 2
      (" ,ceiling((inertie_stations$TOT[2,3]-inertie_stations$TOT[1,3])*1000)/10,"% d'inertie"))
    biplot(AFC_stations$co[,3:4],AFC_stations$co[,3:4],var.axes=F,xlab=paste("axe 3
      (" ,ceiling((inertie_stations$TOT[3,3]-inertie_stations$TOT[2,3])*1000)/10,"%
      d'inertie"),ylab=paste("axe 4 (" ,ceiling((inertie_stations$TOT[4,3]-
      inertie_stations$TOT[3,3])*1000)/10,"% d'inertie"),xlim=c(-3,6.5))
    biplot(AFC_stations$co[,5:6],AFC_stations$co[,5:6],var.axes=F,xlab=paste("axe 5
      (" ,ceiling((inertie_stations$TOT[5,3]-inertie_stations$TOT[4,3])*1000)/10,"%
      d'inertie"),ylab=paste("axe 6 (" ,ceiling((inertie_stations$TOT[6,3]-
      inertie_stations$TOT[5,3])*1000)/10,"% d'inertie"),xlim=c(-2,3))
    biplot(AFC_stations$co[,7:8],AFC_stations$co[,7:8],var.axes=F,xlab=paste("axe 7
      (" ,ceiling((inertie_stations$TOT[7,3]-inertie_stations$TOT[6,3])*1000)/10,"%
      d'inertie"),ylab=paste("axe 8 (" ,ceiling((inertie_stations$TOT[8,3]-
      inertie_stations$TOT[7,3])*1000)/10,"% d'inertie"),xlim=c(-4.5,2))
    par(oma=c(1,1,2,1),font=2)
    mtext(outer=T,paste("AFC des banques de graines des stations sur 8 axes \n Inertie sur les 8 axes
      :",ceiling(inertie_stations$TOT[8,3]*1000)/10,"% de l'inertie totale"),side=3)
    dev.off()
  } else {
    AFC_stations<-dudi.coa(stations_banque,scannf=FALSE,nf=2)
    inertie_stations<-inertia.dudi(AFC_stations)
    win.metafile("AFC banque stations.emf",width=10,height=10)
    biplot(AFC_stations$co[,1:2],AFC_stations$co[,1:2],var.axes=F,xlab=paste("axe 1
      (" ,ceiling(inertie_stations$TOT[1,3]*1000)/10,"% d'inertie"),ylab=paste("axe 2
      (" ,ceiling((inertie_stations$TOT[2,3]-inertie_stations$TOT[1,3])*1000)/10,"% d'inertie"))
    par(oma=c(1,1,2,1),font=2)
    mtext(outer=T,paste("AFC des banques de graines des stations sur 2 axes \n Inertie sur les 2 axes
      :",ceiling(inertie_stations$TOT[2,3]*1000)/10,"% de l'inertie totale"),side=3)
    dev.off()
  }
}

##### AFC sites (AFC inter-stations)

sites<-c()
for (i in 1:nb_sites) {
  sites<-c(sites,rep(noms[i],nb_stations))
}

sites<-as.factor(sites)

```

```

if (nb_sites>1) {
  inter_stations<-t(between(t(AFC_stations),scannf=FALSE,nf=4,fac=sites))
  inertie_site<-inertia.dudi(inter_stations)
  if(nb_sites>3) {
    win.metafile("AFC banque sites.emf",width=10,height=10)
    par(mfrow=c(1,2))
    biplot(inter_stations$co[,1:2],inter_stations$co[,1:2],var.axes=F,xlab=paste("axe 1
      (",ceiling(inertie_site$TOT[1,3]*1000)/10,"% d'inertie"),ylab=paste("axe 2
      (",ceiling((inertie_site$TOT[2,3]-inertie_site$TOT[1,3])*1000)/10,"% d'inertie"))
    biplot(inter_stations$co[,3:4],inter_stations$co[,3:4],var.axes=F,xlab=paste("axe 3
      (",ceiling((inertie_site$TOT[3,3]-inertie_site$TOT[2,3])*1000)/10,"% d'inertie"),ylab=paste("axe 4
      (",ceiling((inertie_site$TOT[4,3]-inertie_site$TOT[3,3])*1000)/10,"% d'inertie"))
    par(oma=c(1,1,3,1),font=2)
    mtext(outer=T,paste("AFC des banques de graines des sites sur 4 axes \n Inertie sur les 4 axes :
      ",ceiling(inertie_site$TOT[4,3]*1000)/10,"% de l'inertie totale"),side=3)
    dev.off()
  } else {
    win.metafile("AFC banque sites.emf",width=10,height=10)
    biplot(inter_stations$co[,1:2],inter_stations$co[,1:2],var.axes=F,xlab=paste("axe 1
      (",ceiling(inertie_site$TOT[1,3]*1000)/10,"% d'inertie"),ylab=paste("axe 2
      (",ceiling((inertie_site$TOT[2,3]-inertie_site$TOT[1,3])*1000)/10,"% d'inertie"))
    par(oma=c(1,1,3,1),font=2)
    mtext(outer=T,paste("AFC des banques de graines des sites sur 4 axes \n Inertie sur les 4 axes :
      ",ceiling(inertie_site$TOT[4,3]*1000)/10,"% de l'inertie totale"),side=3)
    dev.off()
  }
}

```

AFC végétation

```

if (nb_stations>1) {
  if(nb_sites*nb_stations>7) {
    AFC_vegetation<-dudi.coa(stations_vegetation,scannf=FALSE,nf=8)
    inertie_vegetation<-inertia.dudi(AFC_vegetation)
    win.metafile("AFC végétation stations.emf",width=10,height=10)
    par(mfrow=c(2,2))
    biplot(AFC_vegetation$co[,1:2],AFC_vegetation$co[,1:2],var.axes=F,xlab=paste("axe 1
      (",ceiling(inertie_vegetation$TOT[1,3]*1000)/10,"% d'inertie"),ylab=paste("axe 2
      (",ceiling((inertie_vegetation$TOT[2,3]-inertie_vegetation$TOT[1,3])*1000)/10,"% d'inertie"))
    biplot(AFC_vegetation$co[,3:4],AFC_vegetation$co[,3:4],var.axes=F,xlab=paste("axe 3
      (",ceiling((inertie_vegetation$TOT[3,3]-inertie_vegetation$TOT[2,3])*1000)/10,"%
      d'inertie"),ylab=paste("axe 4 (",ceiling((inertie_vegetation$TOT[4,3]-
      inertie_vegetation$TOT[3,3])*1000)/10,"% d'inertie"))
    biplot(AFC_vegetation$co[,5:6],AFC_vegetation$co[,5:6],var.axes=F,xlab=paste("axe 5
      (",ceiling((inertie_vegetation$TOT[5,3]-inertie_vegetation$TOT[4,3])*1000)/10,"%
      d'inertie"),ylab=paste("axe 6 (",ceiling((inertie_vegetation$TOT[6,3]-
      inertie_vegetation$TOT[5,3])*1000)/10,"% d'inertie"))
    biplot(AFC_vegetation$co[,7:8],AFC_vegetation$co[,7:8],var.axes=F,xlab=paste("axe 7
      (",ceiling((inertie_vegetation$TOT[7,3]-inertie_vegetation$TOT[6,3])*1000)/10,"%
      d'inertie"),ylab=paste("axe 8 (",ceiling((inertie_vegetation$TOT[8,3]-
      inertie_vegetation$TOT[7,3])*1000)/10,"% d'inertie"))
    par(oma=c(1,1,2,1),font=2)
    mtext(outer=T,paste("AFC de la végétation des stations sur 8 axes \n Inertie sur les 8 axes
      : ",ceiling(inertie_vegetation$TOT[8,3]*1000)/10,"% de l'inertie totale"),side=3)
    dev.off()
  } else {
    AFC_vegetation<-dudi.coa(stations_vegetation,scannf=FALSE,nf=2)
    inertie_vegetation<-inertia.dudi(AFC_vegetation)
    win.metafile("AFC végétation stations.emf",width=10,height=10)
  }
}

```



```

biplot(AFC_vegetation$c1[,1:2],AFC_vegetation$c1[,1:2],var.axes=F,xlab=paste("axe 1
(",ceiling(inertia_vegetation$TOT[1,3]*1000)/10,"% d'inertie"),ylab=paste("axe 2
(",ceiling((inertia_vegetation$TOT[2,3]-inertia_vegetation$TOT[1,3])*1000)/10,"% d'inertie"))
dev.off()
}}

##### Classification ascendante hiérarchique (CAH) des sites (banque)

if (nb_sites>1) {
  dbanque_site<-dist.dudi(t(inter_stations))
  cah_site=hclust(dbanque_site,method="ward")
  win.metafile("Dendrogramme banque sites.emf",width=10,height=10)
  plot(cah_site,main = "Dendrogramme de la banque de graines des sites \n distance du Chi2 - méthode de
    Ward",xlab = "Les sites",ylab="", sub = "",hang=-1)
  rect.hclust(cah_site, k=nb_sites/2, border="red")
  dev.off()
}

##### Classification ascendante hiérarchique (CAH) des stations (banque)

if (nb_stations>1) {
  afc_stations <- dudi.coa(t(stations_banque),scan=F)
  dbanque_stations<-dist.dudi(afc_stations)
  cah_stations=hclust(dbanque_stations,method="ward")
  win.metafile("Dendrogramme banque stations.emf",width=10,height=10)
  plot(cah_stations,main = "Dendrogramme de la banque de graines des stations \n distance du Chi2 - méthode de
    Ward",xlab = "Les stations",ylab="", sub = "",hang=-1)
  rect.hclust(cah_stations, k=nb_sites*nb_stations/3, border="red")
  dev.off()
}

##### Classification ascendante hiérarchique (CAH) des stations (végétation)

afc_vegetation <- dudi.coa(t(stations_vegetation),scan=F)
dbanque_vegetation<-dist.dudi(t(AFC_vegetation))
cah_vegetation=hclust(dbanque_vegetation^2,method="ward")
win.metafile("Dendrogramme végétation stations.emf",width=10,height=10)
plot(cah_vegetation,main = "Dendrogramme de la végétation moyenne des stations \n methode de Ward",xlab =
  "Les stations",ylab="", sub = "",hang=-1)
rect.hclust(cah_vegetation, k=ifelse(nb_sites==1,3,nb_sites), border="red")
dev.off()

### Graphes : Comparaisons banque / végétation correspondante

##### Jaccard

# Enregistrement du graphe

win.metafile("J a . Comparaison banque \ végétation correspondante (Jaccard).emf",width=10,height=10)

# Gestion du nombre de graphes par fenêtre et du nombre de fenêtres

if (nb_saison<4 & nb_sites<7) {
  op <- par(mfrow=c(nb_saison+1,nb_sites))
  n<-nb_sites
} else {
  if (nb_sites<7) {
    op <- par(mfrow=c(4,nb_sites))
    n<-nb_sites

```

```

    } else {
      op <- par(mfrow=c(4,6))
      n<-6
    }
  }
cpt<-(-1)
m<-1

for (g in 1:nb_saison) {
  for (z in 1:nb_sites) {

# Création d'une matrice résultat nulle

res_jaccard<-matrix(0,nb_stations,3*nb_saison)
noms_col<-c()

for (k in 1:nb_saison) {
  noms_col<-c(noms_col,saison[k],"Indice",saison[k])
}

colnames(res_jaccard)<-noms_col
rownames(res_jaccard)<-names(banq[[z]])[vect[1:nb_stations]]

# Gestion du nombre de graphes par fenêtre et du nombre de fenêtres

cpt<-1+cpt
if (cpt>4*n-1) {
  par(oma=c(1,1,1,1),font=2)
  mtext(outer=T,"Comparaison entre banque de chaque station et végétation correspondante (Jaccard)",side=3)
  dev.off()
  m<-m+1
  win.metafile(paste("J",letters[m],". Comparaison banque \ végétation correspondante
(Jaccard).emf"),width=10,height=10)
  op <- par(mfrow=c(4,n))
  cpt<-0
}

# Tracé des graphes vierges

plot(0,0,xlim=c(0.5,(nb_stations+0.5)),ylim=c(0,1),type="n",xlab=paste(noms[z],saison[g]),ylab="Jaccard")

for (i in 1:nb_stations) {

# Gestion des valeurs manquantes

if (is.na(banq[[z]][1,vect[i]+g])) {
  res_jaccard[i,vect4[g]]<-NA
  res_jaccard[i,vect4[g]+2]<-NA
}
else {

# Création de la statistique

mystat_jaccard<-function(x,idx)
{
  boot.sample.x<-x[idx,vect[i]]
  boot.sample.y<-x[idx,vect[i]+g]
  Jaccard(boot.sample.x,boot.sample.y)
}

```

```

# Indices

res_jaccard[i,vect4[g]+1]<-Jaccard(banq[[z]][,vect[i]],banq[[z]][,vect[i]+g])

# Bootstrap (optimisé) et tracé des indices et intervalles de confiance correspondants

boot.results<-boot(data=banq[[z]],statistic=mystat_jaccard,R=B)
replic<-sort(boot.results$t)
if (replic[alpha*B/200]==0) {
  res_jaccard[i,vect4[g]]<-replic[1]
  res_jaccard[i,vect4[g]+2]<-replic[B-alpha*B/100]
  if (replic[B-alpha*B/100]<0.001) {
    points(i,0,pch=3,col=2,lwd=2)
  } else {
    points(i,Jaccard(banq[[z]][,vect[i]],banq[[z]][,vect[i]+g]),pch=3,col=2,lwd=2)
    segments(i,replic[1],i,replic[B-alpha*B/100],lwd=2)
    arrows(i,replic[1],i,replic[B-alpha*B/100],angle=90,code=2,length=0.03,lwd=2)
    arrows(i,replic[B-alpha*B/100],i,replic[1],angle=90,code=2,length=0.03,lwd=2)
  } else {
    points(i,Jaccard(banq[[z]][,vect[i]],banq[[z]][,vect[i]+g]),pch=3,col=2,lwd=2)
    dif<-replic[B-alpha*B/200]-replic[alpha*B/200]
    dif2<-dif
    for (l in 1:(alpha*B/200-1)) {
      if (dif2<=dif) {
        dif2<-replic[B-alpha*B/200-l]-replic[alpha*B/200-l]
        ind_l<-l
      }
    }
    dif3<-dif
    for (k in 1:alpha*B/200) {
      if (dif3<=dif) {
        dif3<-replic[B-alpha*B/200+k]-replic[alpha*B/200+k]
        ind_k<-k
      }
    }
    if (dif2<=dif3) {
      res_jaccard[i,vect4[g]]<-replic[alpha*B/200-ind_l]
      res_jaccard[i,vect4[g]+2]<-replic[B-alpha*B/200-ind_l]
      segments(i,replic[alpha*B/200-ind_l],i,replic[B-alpha*B/200-ind_l],lwd=2)
      arrows(i,replic[alpha*B/200-ind_l],i,replic[B-alpha*B/200-ind_l],angle=90,code=2,length=0.03,lwd=2)
      arrows(i,replic[B-alpha*B/200-ind_l],i,replic[alpha*B/200-ind_l],angle=90,code=2,length=0.03,lwd=2)
    } else {
      res_jaccard[i,vect4[j]]<-replic[alpha*B/200+ind_k]
      res_jaccard[i,vect4[j]+2]<-replic[B-alpha*B/200+ind_k]
      segments(i,replic[alpha*B/200+ind_k],i,replic[B-alpha*B/200+ind_k],lwd=2)
      arrows(i,replic[alpha*B/200+ind_k],i,replic[B-alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
      arrows(i,replic[B-alpha*B/200+ind_k],i,replic[alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
    }
  }
}

# Impression dans un fichier Excel

ch=paste(dirJaccard,"/IC_",noms[z],"_Jaccard.xls",sep="")
write.xls(res_jaccard,ch,rowNames=T)
}}
#### Printemps + Eté + Automne

veg<-list()
# Création d'une matrice résultat nulle

res_jaccard_moy<-matrix(0,nb_stations*nb_sites,3)

```

```

colnames(res_jaccard_moy)<-c("IC_Moyenne","Indice","IC_Moyenne")
rownames(res_jaccard_moy)<-names(stations_banque)

for (z in 1:nb_sites) {

# Tracé des graphes vierges

plot(0,0,xlim=c(0.5,(nb_stations+0.5)),ylim=c(0,1),xlab=paste(noms[z],"moyenne"),type="n",ylab="Jaccard")

### Moyenne des saisons

veg[[z]]<-matrix(0,length(vegetation[1]),nb_stations)
for (j in 1:nb_stations) {
  veg[[z]][,j]<-rowMeans(veget[[z]][,(1+(j-1)*nb_saison):(j*nb_saison)],na.rm=TRUE)
}

banq[[z]]<-cbind(banq[[z]],veg[[z]])

for (i in 1:nb_stations) {

#### Graphes (moyenne des trois saisons)

# Création de la statistique

mystat_jaccard<-function(x,idx)
{
  boot.sample.x<-x[idx,vect[i]]
  boot.sample.y<-x[idx,lg+i]
  Jaccard(boot.sample.x,boot.sample.y)
}

# Indices

res_jaccard_moy[i+5*(z-1),2]<-Jaccard(banq[[z]][,vect[i]],banq[[z]][,lg+i])

# Bootstrap (optimisé) et tracé des indices et intervalles de confiance correspondants

boot.results<-boot(data=banq[[z]],statistic=mystat_jaccard,R=B)
replic<-sort(boot.results$t)
if (replic[alpha*B/200]==0) {
  res_jaccard_moy[i+5*(z-1),1]<-replic[1]
  res_jaccard_moy[i+5*(z-1),3]<-replic[B-alpha*B/100]
  if (replic[B-alpha*B/100]<0.001) {
    points(i,0,pch=3,col=2,lwd=2)
  } else {
    points(i,Jaccard(banq[[z]][,vect[i]],banq[[z]][,lg+i]),pch=3,col=2,lwd=2)
    segments(i,replic[1],i,replic[B-alpha*B/100],lwd=2)
    arrows(i,replic[1],i,replic[B-alpha*B/100],angle=90,code=2,length=0.03,lwd=2)
    arrows(i,replic[B-alpha*B/100],i,replic[1],angle=90,code=2,length=0.03,lwd=2)
  } else {
    points(i,Jaccard(banq[[z]][,vect[i]],banq[[z]][,lg+i]),pch=3,col=2,lwd=2)
    dif<-replic[B-alpha*B/200]-replic[alpha*B/200]
    dif2<-dif
    for (l in 1:(alpha*B/200-1)) {
      if (dif2<=dif) {
        dif2<-replic[B-alpha*B/200-l]-replic[alpha*B/200-l]
        ind_l<-l
      }
    }
  }
}

```

```

dif3<-dif
for (k in 1:alpha*B/200) {
  if (dif3<=dif) {
    dif3<-replic[B-alpha*B/200+k]-replic[alpha*B/200+k]
    ind_k<-k
  }
}
if (dif2<=dif3) {
  res_jaccard_moy[i+5*(z-1),1]<-replic[alpha*B/200-ind_l]
  res_jaccard_moy[i+5*(z-1),3]<-replic[B-alpha*B/200-ind_l]
  segments(i,replic[alpha*B/200-ind_l],i,replic[B-alpha*B/200-ind_l],lwd=2)
  arrows(i,replic[alpha*B/200-ind_l],i,replic[B-alpha*B/200-ind_l],angle=90,code=2,length=0.03,lwd=2)
  arrows(i,replic[B-alpha*B/200-ind_l],i,replic[alpha*B/200-ind_l],angle=90,code=2,length=0.03,lwd=2)
} else {
  res_jaccard_moy[i+5*(z-1),1]<-replic[alpha*B/200+ind_k]
  res_jaccard_moy[i+5*(z-1),3]<-replic[B-alpha*B/200+ind_k]
  segments(i,replic[alpha*B/200+ind_k],i,replic[B-alpha*B/200+ind_k],lwd=2)
  arrows(i,replic[alpha*B/200+ind_k],i,replic[B-alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
  arrows(i,replic[B-alpha*B/200+ind_k],i,replic[alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
}}
}

ch=paste(dirJaccard,"/IC_moyenne_végétation_Jaccard.xls",sep="")
write.xls(res_jaccard_moy,ch,rowNames=T)
}

par(oma=c(1,1,1,1),font=2)
mtext(outer=T,"Comparaison entre banque de chaque station et végétation correspondante (Jaccard)",side=3)
dev.off()

# Fin du graphe

### Comparaison entre banque de chaque station et végétation moyenne de chaque station sur le même site

##### Jaccard

# Enregistrement du graphe

win.metafile("J 1 . Comparaison banque \ végétation moyenne de chaque station sur le même site
(Jaccard).emf",width=10,height=10)

# Gestion du nombre de graphes par fenêtre et du nombre de fenêtres

if (nb_stations<7) {
  op <- par(mfrow=c(4,nb_stations))
  n<-nb_stations
} else {
}
if (nb_sites<7) {
  op <- par(mfrow=c(4,nb_sites))
  n<-nb_sites
} else {
  op <- par(mfrow=c(4,6))
  n<-6
}
cpt<-(-1)
m<-1

for (z in 1:(nb_sites)) {
  for (i in 1:nb_stations) {

```

```

# Gestion des valeurs manquantes

if (is.na(banq[[z]][1,lg+i])) {
  }
  else {

# Gestion du nombre de graphes par fenêtre et du nombre de fenêtres

cpt<-1+cpt
if (cpt>4*n-1) {
  par(oma=c(1,1,1,1),font=2)
  mtext(outer=T,"Comparaison entre banque de chaque station et végétation moyenne de chaque station sur le
même site (Jaccard)",side=3)
  dev.off()
  m<-m+1
  win.metafile(paste("J",m,". Comparaison banque \ végétation moyenne de chaque station sur le même site
(Jaccard).emf"),width=10,height=10)
  op <- par(mfrow=c(4,n))
  cpt<-0
  }

# Tracé des graphes vierges

plot(0,0,xlim=c(0.5,(nb_stations+0.5)),ylim=c(0,1),xlab=colnames(banq[[z]][vect[i]],type="n",ylab=noms[z])

  for (j in 1:nb_stations) {

# Création de la statistique

mystat_jaccard<-function(x,idx)
{
  boot.sample.x<-x[idx,vect[j]]
  boot.sample.y<-x[idx,lg+i]
  Jaccard(boot.sample.x,boot.sample.y)
  }

# Bootstrap (optimisé) et tracé des indices et intervalles de confiance correspondants

boot.results<-boot(data=banq[[z]],statistic=mystat_jaccard,R=B)
replic<-sort(boot.results$t)
if (replic[alpha*B/200]==0) {
  if (replic[B-alpha*B/100]<0.001) {
    points(j,0,pch=3,col=2,lwd=2)
  } else {
    points(j,Jaccard(banq[[z]][vect[j]],banq[[z]][,lg+i]),pch=3,col=2,lwd=2)
    segments(j,replic[1],j,replic[B-alpha*B/100],lwd=2)
    arrows(j,replic[1],j,replic[B-alpha*B/100],angle=90,code=2,length=0.03,lwd=2)
    arrows(j,replic[B-alpha*B/100],j,replic[1],angle=90,code=2,length=0.03,lwd=2)
  } else {
    points(j,Jaccard(banq[[z]][vect[j]],banq[[z]][,lg+i]),pch=3,col=2,lwd=2)
    dif<-replic[B-alpha*B/200]-replic[alpha*B/200]
    dif2<-dif
    for (l in 1:(alpha*B/200-1)) {
      if (dif2<=dif) {
        dif2<-replic[B-alpha*B/200-l]-replic[alpha*B/200-l]
        ind_l<-l
      }
    }
  }
}

```

```

dif3<-dif
for (k in 1:alpha*B/200) {
  if (dif3<=dif) {
    dif3<-replc[B-alpha*B/200+k]-replc[alpha*B/200+k]
    ind_k<-k
  }
}
if (dif2<=dif3) {
  segments(j,replc[alpha*B/200-ind_1],j,replc[B-alpha*B/200-ind_1],lwd=2)
  arrows(j,replc[alpha*B/200-ind_1],j,replc[B-alpha*B/200-ind_1],angle=90,code=2,length=0.03,lwd=2)
  arrows(j,replc[B-alpha*B/200-ind_1],j,replc[alpha*B/200-ind_1],angle=90,code=2,length=0.03,lwd=2)
} else {
  segments(j,replc[alpha*B/200+ind_k],j,replc[B-alpha*B/200+ind_k],lwd=2)
  arrows(j,replc[alpha*B/200+ind_k],j,replc[B-alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
  arrows(j,replc[B-alpha*B/200+ind_k],j,replc[alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
}}}}}}

par(oma=c(1,1,1,1),font=2)
mtext(outer=T,"Comparaison entre banque de chaque station et végétation moyenne de chaque station sur le
même site (Jaccard)",side=3)
dev.off()

# Fin du graphe

### Graphes : Comparaisons banque / végétation correspondante

##### Ruzicka

# Enregistrement du graphe

win.metafile("R a . Comparaison banque \ végétation correspondante (Ruzicka).emf",width=10,height=10)

# Gestion du nombre de graphes par fenêtre et du nombre de fenêtres

if (nb_saison<4 & nb_sites<7) {
  op <- par(mfrow=c(nb_saison+1,nb_sites))
  n<-nb_sites
} else {
if (nb_sites<7) {
  op <- par(mfrow=c(4,nb_sites))
  n<-nb_sites
} else {
  op <- par(mfrow=c(4,6))
  n<-6
}
}
cpt<-(-1)
m<-1

for (g in 1:nb_saison) {
for (z in 1:nb_sites) {

# Création d'une matrice résultat nulle

res_ruzicka<-matrix(0,nb_stations,3*nb_saison)
noms_col<-c()

for (k in 1:nb_saison) {
  noms_col<-c(noms_col,saison[k],"Indice",saison[k])
}
}
}

```

```

}

colnames(res_ruzicka)<-noms_col
rownames(res_ruzicka)<-names(banq[[z]][vect[1:nb_stations]])

# Gestion du nombre de graphes par fenêtre et du nombre de fenêtres

cpt<-1+cpt
if (cpt>4*n-1) {
  par(oma=c(1,1,1,1),font=2)
  mtext(outer=T,"Comparaison entre banque de chaque station et végétation correspondante (Ruzicka)",side=3)
  dev.off()
  m<-m+1
  win.metafile(paste("R",letters[m],". Comparaison banque \ végétation correspondante
(Ruzicka).emf"),width=10,height=10)
  op <- par(mfrow=c(4,n))
  cpt<-0
}

# Tracé des graphes vierges

plot(0,0,xlim=c(0.5,(nb_stations+0.5)),ylim=c(0,1),type="n",xlab=paste(noms[z],saison[g]),ylab="Ruzicka")

for (i in 1:nb_stations) {

# Gestion des valeurs manquantes

if (is.na(banq[[z]][1,vect[i]+g])) {
  res_ruzicka[i,vect4[g]]<-NA
  res_ruzicka[i,vect4[g]+2]<-NA
}
else {

# Création de la statistique

mystat_ruzicka<-function(x,idx)
{
  boot.sample.x<-x[idx,vect[i]]
  boot.sample.y<-x[idx,vect[i]+g]
  Ruzicka(boot.sample.x,boot.sample.y)
}

# Indices

res_ruzicka[i,vect4[g]+1]<-Ruzicka(banq[[z]][vect[i]],banq[[z]][vect[i]+g])

# Bootstrap (optimisé) et tracé des indices et intervalles de confiance correspondants

boot.results<-boot(data=banq[[z]],statistic=mystat_ruzicka,R=B)
replic<-sort(boot.results$t)
if (replic[alpha*B/200]==0) {
  res_ruzicka[i,vect4[g]]<-replic[1]
  res_ruzicka[i,vect4[g]+2]<-replic[B-alpha*B/100]
  if (replic[B-alpha*B/100]<0.001) {
    points(i,0,pch=3,col=2,lwd=2)
  } else {
    points(i,Ruzicka(banq[[z]][vect[i]],banq[[z]][vect[i]+g]),pch=3,col=2,lwd=2)
    segments(i,replic[1],i,replic[B-alpha*B/100],lwd=2)
    arrows(i,replic[1],i,replic[B-alpha*B/100],angle=90,code=2,length=0.03,lwd=2)

```



```

    arrows(i,replic[B-alpha*B/100],i,replic[1],angle=90,code=2,length=0.03,lwd=2)
  } } else {
points(i,Ruzicka(banq[[z]][,vect[i]],banq[[z]][,vect[i]+g]),pch=3,col=2,lwd=2)
dif<-replic[B-alpha*B/200]-replic[alpha*B/200]
dif2<-dif
for (l in 1:(alpha*B/200-1)) {
  if (dif2<=dif) {
    dif2<-replic[B-alpha*B/200-l]-replic[alpha*B/200-l]
    ind_l<-l
  }
}
dif3<-dif
for (k in 1:alpha*B/200) {
  if (dif3<=dif) {
    dif3<-replic[B-alpha*B/200+k]-replic[alpha*B/200+k]
    ind_k<-k
  }
}
if (dif2<=dif3) {
  res_ruzicka[i,vect4[g]]<-replic[alpha*B/200-ind_l]
  res_ruzicka[i,vect4[g]+2]<-replic[B-alpha*B/200-ind_l]
  segments(i,replic[alpha*B/200-ind_l],i,replic[B-alpha*B/200-ind_l],lwd=2)
  arrows(i,replic[alpha*B/200-ind_l],i,replic[B-alpha*B/200-ind_l],angle=90,code=2,length=0.03,lwd=2)
  arrows(i,replic[B-alpha*B/200-ind_l],i,replic[alpha*B/200-ind_l],angle=90,code=2,length=0.03,lwd=2)
} else {
  res_ruzicka[i,vect4[j]]<-replic[alpha*B/200+ind_k]
  res_ruzicka[i,vect4[j]+2]<-replic[B-alpha*B/200+ind_k]
  segments(i,replic[alpha*B/200+ind_k],i,replic[B-alpha*B/200+ind_k],lwd=2)
  arrows(i,replic[alpha*B/200+ind_k],i,replic[B-alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
  arrows(i,replic[B-alpha*B/200+ind_k],i,replic[alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
}}}}

# Impression dans un fichier Excel

ch=paste(dirRuzicka,"/IC_",noms[z],"_Ruzicka.xls",sep="")
write.xls(res_ruzicka,ch,rowNames=T)
}}

#### Printemps + Eté + Automne

# Création d'une matrice résultat nulle

res_ruzicka_moy<-matrix(0,nb_stations*nb_sites,3)
colnames(res_ruzicka_moy)<-c("IC_Moyenne","Indice","IC_Moyenne")
rownames(res_ruzicka_moy)<-names(stations_banque)

for (z in 1:nb_sites) {

# Tracé des graphes vierges

plot(0,0,xlim=c(0.5,(nb_stations+0.5)),ylim=c(0,1),xlab=paste(noms[z],"moyenne"),type="n",ylab="Ruzicka")

  for (i in 1:nb_stations) {

#### Graphes (moyenne des trois saisons)

# Création de la statistique

mystat_ruzicka<-function(x,idx)

```

```

{
boot.sample.x<-x[idx,vect[i]]
boot.sample.y<-x[idx,lg+i]
Ruzicka(boot.sample.x,boot.sample.y)
}

# Indices

res_ruzicka_moy[i+5*(z-1),2]<-Ruzicka(banq[[z]][,vect[i]],banq[[z]][,lg+i])

# Bootstrap (optimisé) et tracé des indices et intervalles de confiance correspondants

boot.results<-boot(data=banq[[z]],statistic=mystat_ruzicka,R=B)
replic<-sort(boot.results$t)
if (replic[alpha*B/200]==0) {
  res_ruzicka_moy[i+5*(z-1),1]<-replic[1]
  res_ruzicka_moy[i+5*(z-1),3]<-replic[B-alpha*B/100]
  if (replic[B-alpha*B/100]<0.001) {
    points(i,0,pch=3,col=2,lwd=2)
  } else {
    points(i,Ruzicka(banq[[z]][,vect[i]],banq[[z]][,lg+i]),pch=3,col=2,lwd=2)
    segments(i,replic[1],i,replic[B-alpha*B/100],lwd=2)
    arrows(i,replic[1],i,replic[B-alpha*B/100],angle=90,code=2,length=0.03,lwd=2)
    arrows(i,replic[B-alpha*B/100],i,replic[1],angle=90,code=2,length=0.03,lwd=2)
  } } else {
  points(i,Ruzicka(banq[[z]][,vect[i]],banq[[z]][,lg+i]),pch=3,col=2,lwd=2)
  dif<-replic[B-alpha*B/200]-replic[alpha*B/200]
  dif2<-dif
  for (l in 1:(alpha*B/200-1)) {
    if (dif2<=dif) {
      dif2<-replic[B-alpha*B/200-l]-replic[alpha*B/200-l]
      ind_l<-l
    }
  }
  dif3<-dif
  for (k in 1:alpha*B/200) {
    if (dif3<=dif) {
      dif3<-replic[B-alpha*B/200+k]-replic[alpha*B/200+k]
      ind_k<-k
    }
  }
  if (dif2<=dif3) {
    res_ruzicka_moy[i+5*(z-1),1]<-replic[alpha*B/200-ind_l]
    res_ruzicka_moy[i+5*(z-1),3]<-replic[B-alpha*B/200-ind_l]
    segments(i,replic[alpha*B/200-ind_l],i,replic[B-alpha*B/200-ind_l],lwd=2)
    arrows(i,replic[alpha*B/200-ind_l],i,replic[B-alpha*B/200-ind_l],angle=90,code=2,length=0.03,lwd=2)
    arrows(i,replic[B-alpha*B/200-ind_l],i,replic[alpha*B/200-ind_l],angle=90,code=2,length=0.03,lwd=2)
  } else {
    res_ruzicka_moy[i+5*(z-1),1]<-replic[alpha*B/200+ind_k]
    res_ruzicka_moy[i+5*(z-1),3]<-replic[B-alpha*B/200+ind_k]
    segments(i,replic[alpha*B/200+ind_k],i,replic[B-alpha*B/200+ind_k],lwd=2)
    arrows(i,replic[alpha*B/200+ind_k],i,replic[B-alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
    arrows(i,replic[B-alpha*B/200+ind_k],i,replic[alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
  } } }

ch=paste(dirRuzicka,"/IC_moyenne_végétation_Ruzicka.xls",sep="")
write.xls(res_ruzicka_moy,ch,rowNames=T)
}

```

```

par(oma=c(1,1,1,1),font=2)
mtext(outer=T,"Comparaison entre banque de chaque station et végétation correspondante (Ruzicka)",side=3)
dev.off()

# Fin du graphe

### Comparaison entre banque de chaque station et végétation moyenne de chaque station sur le même site

##### Ruzicka

# Enregistrement du graphe

win.metafile("R 1 . Comparaison banque \ végétation moyenne de chaque station sur le même site
(Ruzicka).emf",width=10,height=10)

# Gestion du nombre de graphes par fenêtre et du nombre de fenêtres

if (nb_stations<7) {
  op <- par(mfrow=c(4,nb_stations))
  n<-nb_stations
} else
if (nb_sites<7) {
  op <- par(mfrow=c(4,nb_sites))
  n<-nb_sites
} else {
  op <- par(mfrow=c(4,6))
  n<-6
}
cpt<-(-1)
m<-1

for (z in 1:(nb_sites)) {
  for (i in 1:nb_stations) {

# Gestion des valeurs manquantes

if (is.na(banq[[z]][1,lg+i])) {
  }
  else {

# Gestion du nombre de graphes par fenêtre et du nombre de fenêtres

cpt<-1+cpt
if (cpt>4*n-1) {
  par(oma=c(1,1,1,1),font=2)
  mtext(outer=T,"Comparaison entre banque de chaque station et végétation moyenne de chaque station sur le
même site (Ruzicka)",side=3)
  dev.off()
  m<-m+1
  win.metafile(paste("R",m,". Comparaison banque \ végétation moyenne de chaque station sur le même site
(Ruzicka).emf"),width=10,height=10)
  op <- par(mfrow=c(4,n))
  cpt<-0
}

# Tracé des graphes vierges

plot(0,0,xlim=c(0.5,(nb_stations+0.5)),ylim=c(0,1),xlab=colnames(banq[[z]])[vect[i]],type="n",ylab=noms[z])

```

```

    for (j in 1:nb_stations) {

# Création de la statistique

mystat_ruzicka<-function(x,idx)
{
  boot.sample.x<-x[idx,vect[j]]
  boot.sample.y<-x[idx,lg+i]
  Ruzicka(boot.sample.x,boot.sample.y)
}

# Bootstrap (optimisé) et tracé des indices et intervalles de confiance correspondants

boot.results<-boot(data=banq[[z]],statistic=mystat_ruzicka,R=B)
repl<-sort(boot.results$t)
if (repl[alpha*B/200]==0) {
  if (repl[B-alpha*B/100]<0.001) {
    points(j,0,pch=3,col=2,lwd=2)
  } else {
    points(j,Ruzicka(banq[[z]][vect[j]],banq[[z]][lg+i]),pch=3,col=2,lwd=2)
    segments(j,repl[1],j,repl[B-alpha*B/100],lwd=2)
    arrows(j,repl[1],j,repl[B-alpha*B/100],angle=90,code=2,length=0.03,lwd=2)
    arrows(j,repl[B-alpha*B/100],j,repl[1],angle=90,code=2,length=0.03,lwd=2)
  } else {
    points(j,Ruzicka(banq[[z]][vect[j]],banq[[z]][lg+i]),pch=3,col=2,lwd=2)
    dif<-repl[B-alpha*B/200]-repl[alpha*B/200]
    dif2<-dif
    for (l in 1:(alpha*B/200-1)) {
      if (dif2<=dif) {
        dif2<-repl[B-alpha*B/200-l]-repl[alpha*B/200-l]
        ind_l<-l
      }
    }
    dif3<-dif
    for (k in 1:alpha*B/200) {
      if (dif3<=dif) {
        dif3<-repl[B-alpha*B/200+k]-repl[alpha*B/200+k]
        ind_k<-k
      }
    }
    if (dif2<=dif3) {
      segments(j,repl[alpha*B/200-ind_l],j,repl[B-alpha*B/200-ind_l],lwd=2)
      arrows(j,repl[alpha*B/200-ind_l],j,repl[B-alpha*B/200-ind_l],angle=90,code=2,length=0.03,lwd=2)
      arrows(j,repl[B-alpha*B/200-ind_l],j,repl[alpha*B/200-ind_l],angle=90,code=2,length=0.03,lwd=2)
    } else {
      segments(j,repl[alpha*B/200+ind_k],j,repl[B-alpha*B/200+ind_k],lwd=2)
      arrows(j,repl[alpha*B/200+ind_k],j,repl[B-alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
      arrows(j,repl[B-alpha*B/200+ind_k],j,repl[alpha*B/200+ind_k],angle=90,code=2,length=0.03,lwd=2)
    }
  }
}

par(oma=c(1,1,1,1),font=2)
mtext(outer=T,"Comparaison entre banque de chaque station et végétation moyenne de chaque station sur le
même site (Ruzicka)",side=3)
dev.off()

# Fin du graphe

print("FIN DU PROGRAMME")

```

Mode d'emploi du programme

Pour toute question, merci de me contacter à l'adresse suivante :
th.wilhelm@laposte.net

1. Introduction

1.1. Objectif

L'objectif de ce programme est d'établir la liaison entre deux jeux de données, basée sur le calcul d'indices de similarité. La liaison entre banque de diaspores et végétation a été prise ici comme exemple, mais le programme peut être utilisé pour d'autres types de données.

1.2. Contexte

Un site est une unité géographique que l'on a divisée en plusieurs stations (5 dans notre exemple).

Le type représente le type de milieu dans lequel évolue une espèce : il peut être terrestre, hydrophyte ou héliophyte.

La diaspore représente le type de reproduction d'une espèce, qui peut être une graine, un bourgeon dormant, un rhizome, un fragment ou un bulbe

1.3. Résultats

Le programme va tout d'abord comparer les différents sites et stations entre eux en terme de banque de graines et de végétation à l'aide d'analyses factorielles des correspondances (AFC) et de dendrogrammes issus de classifications ascendantes hiérarchiques (CAH). Des indices et des intervalles de confiance bootstrap autour de ces indices vont ensuite être calculés, puis tracés sur des graphes, dans le but de comparer les banques de graines et la végétation correspondante sur chaque station dans un premier temps, puis les banques de graines de chaque station avec la végétation des stations situées sur le même site que la station sur laquelle on étudie la banque de graines.

2. Les données (tableaux Excel)

Les données utilisées par ce programme doivent être entrées sous format Excel. La version utilisée dans ce programme est la version « Microsoft Office Excel 97-2003 » (au format .xls).

Les données sont rangées dans 2 classeurs différents (classeurs à 1 seule feuille chacun), classeurs qui doivent être placés dans le sous dossier « Banque et végétation » du dossier « Comparaisons banque végétation » :

- Un classeur « banque » qui contient les données des banques de graines de chaque station nommé « banque.xls » (en minuscule)
- Un classeur « végétation » qui contient les données de la végétation pour chaque station (stations ordonnées comme dans le classeur « banque ») et pour chaque saison, classeur nommé « vegetation.xls » (sans accent, ni majuscule)

Chaque site contient plusieurs stations (ou transects).

Les deux classeurs doivent avoir le même nombre de lignes, avec les mêmes espèces : pour les données manquantes, c'est-à-dire pour une espèce présente dans un tableau et pas dans l'autre, remplir l'autre tableau avec des 0 (zéro).

Les classeurs sont ordonnés comme suit (attention : respecter les majuscules et la syntaxe pour « Espèces », « Type » et « Diaspore », ne pas mettre d'espace (_ underscore ou rien) dans les noms d'espèces ou les noms de stations) : (voir les exemples)

- Le classeur banque :

Espèces	Type	Diaspore	Site_1_Station_1	Site_1_Station_2	...	Site_2_Station_1	...
<i>espèce_1</i>							
<i>espèce_2</i>							
...							

- Le classeur végétation :

Espèces	Type	Diaspore	Site_1_Station_1 _saison_1	Site_1_Station_1 _saison_2	...	Site_1_Station_2 _saison_1	...
<i>espèce_1</i>							
<i>espèce_2</i>							
...							

Les type et diaspore se remplissent avec les abréviations suivantes (respecter les majuscules) :

- Type : Terrestre (Ter), Hydrophyte (Hy) ou Hélophyte (He)
- Diaspore : Graine (G), bourgeon dormant (Bd), Rhizome (R), Fragment (F) ou Bulbe (Bl) (si la diaspore est inconnue, la case "Diaspore" est remplie avec un 0 (zéro))

Les nombres décimaux s'écrivent avec des virgules et pas des points.

Les saisons pour lesquelles il n'y a pas de données restent vierges (toutes les saisons présentes sur une station doivent obligatoirement l'être sur les autres stations)

3. Le logiciel R

Le logiciel de programmation statistique « R » est disponible sur le site du Cran à l'adresse suivante :

<http://cran.r-project.org/>

en cliquant sur « Windows », puis sur « base » et enfin sur « Download R for Windows »

Installer le logiciel « R » sur l'ordinateur (« Exécuter »)

- Lors de la première ouverture du logiciel R, copier-coller les lignes suivantes :

```
install.packages("ade4", dependencies=TRUE)
```

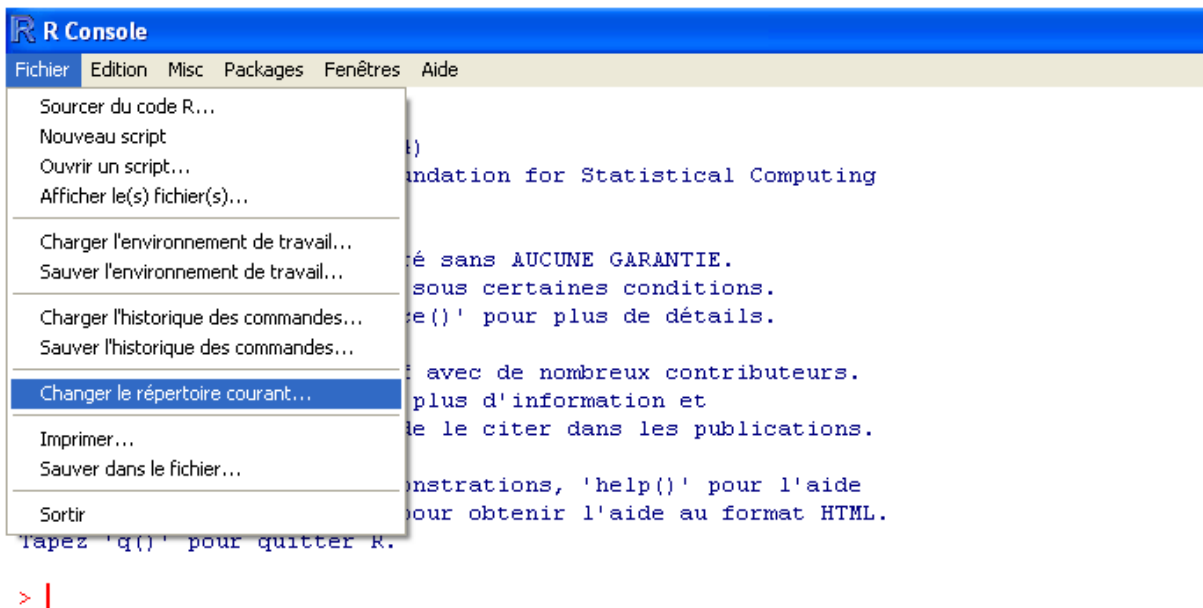
```
install.packages("boot", dependencies=TRUE)
```

```
install.packages("xlsReadWrite", dependencies=TRUE)
```

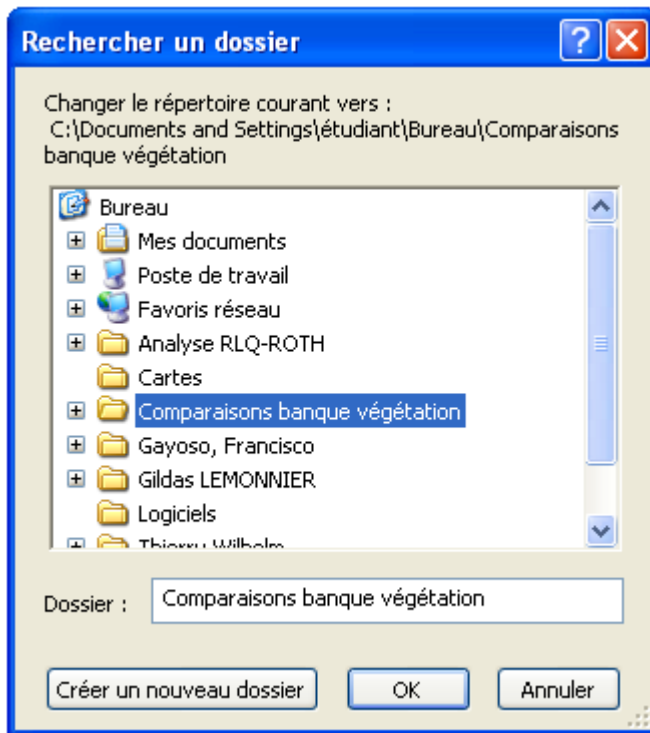
(à l'ouverture de la fenêtre “Cran mirror”, choisir un site miroir en France)

- A chaque utilisation de R, vérifier que le répertoire courant est le bon :

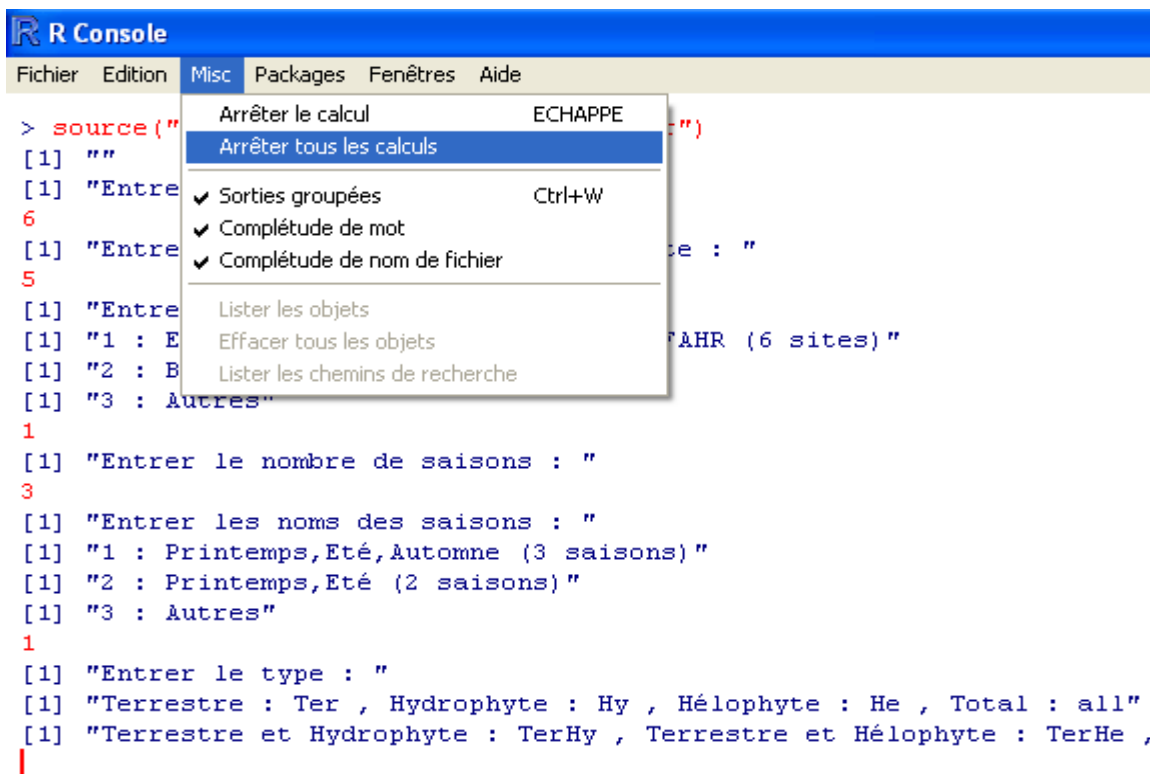
Cliquer sur “Fichier”, puis sur “Changer le répertoire courant...”



A l'ouverture de la fenêtre « Rechercher un dossier », cliquer sur le dossier « Comparaisons banque végétation », situé sur le bureau (préalablement enregistré, contient le programme) :



En cas d'erreur de saisie ou pour arrêter le programme en cours d'exécution, cliquer sur « Misc » puis sur « Arrêter tous les calculs » :



4. Le programme

Pour lancer le programme, copier-coller la ligne suivante (en rouge) dans le logiciel R (validation avec la touche "Entrée") :

```
source("Comparaison banque végétation.r")
```

Le programme demande alors d'entrer le nombre de sites, puis le nombre de stations par site :

```
R R Console
Fichier Edition Misc Packages Fenêtres Aide

> source("Comparaison banque végétation.r")
[1] ""
[1] "Entrer le nombre de sites : "
```

```
R R Console
Fichier Edition Misc Packages Fenêtres Aide

> source("Comparaison banque végétation.r")
[1] ""
[1] "Entrer le nombre de sites : "
6
[1] "Entrer le nombre de stations par site : "
```

L'étape suivante consiste à entrer les noms des sites : 2 jeux de noms sont préenregistrés : il suffit de taper « 1 » pour « EISM », « EISV », ... (c'est-à-dire pour 6 sites) ou « 2 » pour bassin d'orage (1 seul site).

```
R R Console
Fichier Edition Misc Packages Fenêtres Aide

> source("Comparaison banque végétation.r")
[1] ""
[1] "Entrer le nombre de sites : "
6
[1] "Entrer le nombre de stations par site : "
5
[1] "Entrer les noms des sites : "
[1] "1 : EISM, EISV, SCHA, SCHOL, ROSS, FAHR (6 sites)"
[1] "2 : Bassin d'orage (1 site)"
[1] "3 : Autres"
|
```

En tapant « 3 », on peut entrer les noms des sites à la main (les sites doivent être dans le même ordre que dans le tableau Excel) :

```
R R Console
Fichier Edition Misc Packages Fenêtres Aide

> source("Comparaison banque végétation.r")
[1] ""
[1] "Entrer le nombre de sites : "
6
[1] "Entrer le nombre de stations par site : "
5
[1] "Entrer les noms des sites : "
[1] "1 : EISM, EISV, SCHA, SCHOL, ROSS, FAHR (6 sites)"
[1] "2 : Bassin d'orage (1 site)"
[1] "3 : Autres"
3
[1] "Entrer le nom du 1er site : "
site_1
[1] "Entrer le nom du 2 ème site : "
|
```

L'étape suivante consiste à entrer le nombre et les noms des saisons. 2 jeux de saisons sont préenregistrés : il suffit de taper « 1 » pour « Printemps », « Eté » et « Automne » (c'est-à-dire pour 3 saisons) ou « 2 » pour « Printemps » et « Eté » (2 saisons). En tapant « 3 », on peut, comme pour les noms des sites, entrer les noms des saisons à la main :

```

> source("Comparaison banque végétation.r")
[1] ""
[1] "Entrer le nombre de sites : "
6
[1] "Entrer le nombre de stations par site : "
5
[1] "Entrer les noms des sites : "
[1] "1 : EISM, EISV, SCHA, SCHOL, ROSS, FAHR (6 sites)"
[1] "2 : Bassin d'orage (1 site)"
[1] "3 : Autres"
1
[1] "Entrer le nombre de saisons : "
3
[1] "Entrer les noms des saisons : "
[1] "1 : Printemps,Eté,Automne (3 saisons)"
[1] "2 : Printemps,Eté (2 saisons)"
[1] "3 : Autres"
3
[1] "Entrer le nom de la 1ère saison : "
saison1
[1] "Entrer le nom de la 2 ème saison : "
|

```

L'étape suivante consiste à entrer le type (Terrestre, Hydrophyte ou Hélophyte) des espèces que l'on souhaite analyser. Il suffit d'entrer l'abréviation correspondante (en respectant les majuscules) : « Ter » pour Terrestre, « Hy » pour Hydrophyte, « He » pour Hélophyte, « TerHy » pour Terrestre et Hydrophyte, « TerHe » pour Terrestre et Hélophyte, « HyHe » pour Hydrophyte et Hélophyte, ou « all » pour tout les types réunis :

```

> source("Comparaison banque végétation.r")
[1] ""
[1] "Entrer le nombre de sites : "
6
[1] "Entrer le nombre de stations par site : "
5
[1] "Entrer les noms des sites : "
[1] "1 : EISM, EISV, SCHA, SCHOL, ROSS, FAHR (6 sites)"
[1] "2 : Bassin d'orage (1 site)"
[1] "3 : Autres"
1
[1] "Entrer le nombre de saisons : "
3
[1] "Entrer les noms des saisons : "
[1] "1 : Printemps,Eté,Automne (3 saisons)"
[1] "2 : Printemps,Eté (2 saisons)"
[1] "3 : Autres"
1
[1] "Entrer le type : "
[1] "Terrestre : Ter , Hydrophyte : Hy , Hélophyte : He , Total : all"
[1] "Terrestre et Hydrophyte : TerHy , Terrestre et Hélophyte : TerHe , Hydrophyte et Hélophyte : HyHe"
|

```

On fait ensuite de même pour les diaspores (« G » pour Graine, « Bd » pour Bourgeon dormant, « R » pour Rhizome, « F » pour Fragment, « Bl » pour Bulbe ou « all » pour toutes les diaspores). Pour les combinaisons de diaspores, se reporter au script affiché pendant l'exécution du programme (respecter les majuscules) :

```

R Console
Fichier Edition Misc Packages Fenêtres Aide
> source("Comparaison banque végétation.r")
[1] ""
[1] "Entrer le nombre de sites : "
6
[1] "Entrer le nombre de stations par site : "
5
[1] "Entrer les noms des sites : "
[1] "1 : EISM, EISV, SCHA, SCHOL, ROSS, FAHR (6 sites)"
[1] "2 : Bassin d'orage (1 site)"
[1] "3 : Autres"
1
[1] "Entrer le nombre de saisons : "
3
[1] "Entrer les noms des saisons : "
[1] "1 : Printemps,Eté,Automne (3 saisons)"
[1] "2 : Printemps,Eté (2 saisons)"
[1] "3 : Autres"
1
[1] "Entrer le type : "
[1] "Terrestre : Ter , Hydrophyte : Hy , Héliophyte : He , Total : all"
[1] "Terrestre et Hydrophyte : TerHy , Terrestre et Héliophyte : TerHe , Hydrophyte et Héliophyte : HyHe"
all
[1] "Entrer la diaspore"
[1] "Graine : G ; Bourgeon dormant : Bd ; Rhizome : R ; Fragmite : F ; Bulbe : Bl ; Total : all"
[1] "Graine et Bourgeon dormant : GBd ; Graine et Rhizome : GR ; Graine et Fragmite : GF ; Graine et Bulbe : GB1"
[1] "Bourgeon dormant et Rhizome : BdR ; Bourgeon dormant et Fragmite : BdF ; Bourgeon dormant et Bulbe : BdBl"
[1] "Rhizome et Fragmite : RF ; Rhizome et Bulbe : RBl ; Fragmite et Bulbe : FB1"
[1] "Graine, Bourgeon dormant et Rhizome : GBdR ; Graine, Bourgeon dormant et Fragmite : GBdF"
[1] "Graine, Bourgeon dormant et Bulbe : GBdBl ; Graine, Rhizome et Fragmite : GRF ; Graine, Rhizome et Bulbe : GRBl"
[1] "Graine, Fragmite et Bulbe : GFBl ; Bourgeon dormant, Rhizome et Fragmite : BdRF ; Bourgeon dormant, Rhizome et Bulbe : BdRBl"
[1] "Bourgeon dormant, Fragmite et Bulbe : BdFB1 ; Rhizome, Fragmite et Bulbe : RFB1 ; Graine, Bourgeon dormant, Rhizome et Fragmite : GBdRF"
[1] "Graine, Bourgeon dormant, Rhizome et Bulbe : GBdRBl ; Graine, Bourgeon dormant, Fragmite et Bulbe : GBdFB1"
[1] "Graine, Rhizome, Fragmite et Bulbe : GRFB1 ; Bourgeon dormant, Rhizome, Fragmite et Bulbe : BdRFB1"

```

La dernière étape consiste à entrer le nombre de réplifications bootstrap que l'on souhaite effectuer pour établir les intervalles de confiance (au minimum 200) et le risque alpha associé (entrer par exemple 5 pour 5%) en pourcentage. Plus le risque est grand, plus les intervalles de confiance seront petits et vice-versa. Un risque de 5% signifie que 95% des valeurs seront comprises dans l'intervalle de confiance.

Attention, plus le nombre de réplifications est élevé, plus le programme est long à exécuter, mais plus les intervalles de confiance sont précis (jusqu'à un certain nombre de réplifications ; en général, 1000 réplifications suffisent)

(compter entre 5 et 10 minutes environ par tranche de 1000 réplifications, le temps d'exécution du programme variant aussi selon la taille des données)

```

> source("Comparaison banque végétation.r")
[1] ""
[1] "Entrer le nombre de sites : "
6
[1] "Entrer le nombre de stations par site : "
5
[1] "Entrer les noms des sites : "
[1] "1 : EISM, EISV, SCHA, SCHOL, ROSS, FAHR (6 sites)"
[1] "2 : Bassin d'orage (1 site)"
[1] "3 : Autres"
1
[1] "Entrer le nombre de saisons : "
3
[1] "Entrer les noms des saisons : "
[1] "1 : Printemps,Eté,Automne (3 saisons)"
[1] "2 : Printemps,Eté (2 saisons)"
[1] "3 : Autres"
1
[1] "Entrer le type : "
[1] "Terrestre : Ter , Hydrophyte : Hy , Héliophyte : He , Total : all"
[1] "Terrestre et Hydrophyte : TerHy , Terrestre et Héliophyte : TerHe ,
all
[1] "Entrer la diaspore"
[1] "Graine : G ; Bourgeon dormant : Bd ; Rhizome : R ; Fragmite
[1] "Graine et Bourgeon dormant : GBd ; Graine et Rhizome : GR ; C
[1] "Bourgeon dormant et Rhizome : BdR ; Bourgeon dormant et Fragmite
[1] "Rhizome et Fragmite : RF ; Rhizome et Bulbe : RB1 ; Fragmite e
[1] "Graine, Bourgeon dormant et Rhizome : GBdR ; Graine, Bourgeon dc
[1] "Graine, Bourgeon dormant et Bulbe : GBdB1 ; Graine, Rhizome et F
[1] "Graine, Fragmite et Bulbe : GFb1 ; Bourgeon dormant, Rhizome et
[1] "Bourgeon dormant, Fragmite et Bulbe : BdFB1 ; Rhizome, Fragmite
[1] "Graine, Bourgeon dormant, Rhizome et Bulbe : GBdRB1 ; Graine, Bc
[1] "Graine, Rhizome, Fragmite et Bulbe : GRFB1 ; Bourgeon dormant, F
all
[1] "Entrer le nombre de réplifications bootstrap : "
1000
[1] "Entrer le risque alpha en pourcentage : "
5

```

5. Les sorties (intervalles de confiance et graphes)

Une fois le programme exécuté, un message « FIN DU PROGRAMME » s'affiche dans la fenêtre de R (ne pas essayer d'ouvrir les graphes avant la fin de l'exécution).

Les différents graphes (à savoir les AFC, les dendrogrammes et les intervalles de confiance) sont directement enregistrés dans le dossier "Comparaisons banque végétation".

Les intervalles de confiance sont enregistrés dans un classeur Excel (un classeur par station) dans le sous-dossier "Intervalles de confiance" du dossier "Comparaisons banque végétation".

Attention : Les graphes et les intervalles de confiance sont automatiquement effacés si le programme est relancé. Veiller à en faire une copie.

5.1. Les classeurs Excel

Plusieurs classeurs Excel sont créés :

- Comparaison des sites : les sites sont comparés à l'aide du calcul d'un indice, qui est compris entre 0 (pas de ressemblance) et 1 (identiques).
- Comparaison des stations : idem
- IC moyenne végétation : contient les indices (et intervalles de confiance autour de ces indices) entre banque de graines sur chaque station et végétation moyenne correspondante.
- Pour chaque site, un classeur Excel est créé et contient les indices (et intervalles de confiance autour de ces indices) entre banque de graines sur chaque station et végétation correspondante pour chaque saison.

5.2. Les graphes

5.2.1. Les AFC

Sont comparées les banques de graines des sites entre elles, puis celles des stations entre elles. Attention, les proximités peuvent être trompeuses, il faut faire une synthèse de tous les axes de l'AFC représentés.

5.2.2. Les dendrogrammes

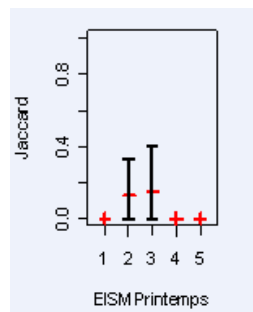
Sont comparées les banques de graines des sites entre elles, puis celles des stations entre elles, et enfin les végétations moyennes de chaque station entre elles. Les dendrogrammes se lisent de bas en haut : plus les ponts reliant deux stations sont bas, plus ces stations sont proches. Les cadres rouges représentent les groupes des stations qui se ressemblent le plus.

5.2.3. Les comparaisons banque / végétation

La croix rouge (+) représente l'indice calculé entre banque de graine et végétation, alors que la bande noire représente l'intervalle de confiance autour de cet indice. Si l'indice est nul, l'intervalle de confiance n'est pas tracé (car lui aussi est nul), et si rien n'est tracé dans la fenêtre, cela signifie qu'il n'y a pas de données (données manquantes) pour cette comparaison.

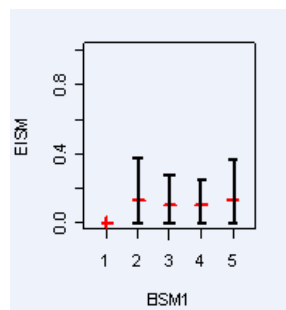
La première lettre du nom des graphes (J ou R) représente l'indice utilisé (Jaccard ou Ruzicka).

- Si cette lettre est suivie d'une autre lettre (a, b,...), le graphe en question compare la banque de graines à la végétation correspondante (en d'autres mots, le graphe compare la banque de graines et la végétation du même site et de la même station) :



Le graphe se lit alors de la manière suivante : ici on compare la banque de graines de la station EISM1 avec la végétation de EISM1 pour la saison Printemps, puis la banque de graines de la station EISM2 avec la végétation de EISM2 pour la saison Printemps, et ainsi de suite...

- Si, par contre, la première lettre est suivie d'un chiffre, le graphe en question compare la banque de graines de chaque station à la végétation de chacune des autres stations du même site :



Le graphe se lit alors de la manière suivante : ici on compare la banque de graines de la station EISM1 avec la végétation de EISM1 pour la moyenne des saisons, puis toujours la banque de graines de la station EISM1 avec la végétation de EISM2, et ainsi de suite...

Liste des figures et tableaux

Tableau 1 : occurrences pour données binaires	7
Tableau 2 : tableau de contingence	13
Tableau 3 : tableau des profils lignes	13
Tableau 4 : tableau des profils colonnes	13
Tableau 5 : valeurs propres et pourcentage de variation cumulée de l'AFC de la banque par stations	36
Tableau 6 : distances aux carrés des colonnes aux axes	37
Tableau 7 : distances aux carrés des lignes aux axes	38
Tableau 8 : contribution des colonnes aux axes	39
Tableau 9 : contribution des lignes aux axes	40
Tableau 10 : valeurs propres et pourcentage de variation cumulée de l'AFC inter-stations	42
Tableau 11 : cosinus carrés et contributions des colonnes de l'AFC inter-stations de la banque	43
Tableau 12 : cosinus carrés et contributions des lignes de l'AFC inter-stations de la banque	44
Tableau 13 : valeurs propres de l'AFC de la banque de graines du bassin d'orage	83
Tableau 14 : distance des colonnes aux axes	84
Tableau 15 : contributions des colonnes aux axes	84
Figure 1 : intervalle de confiance pour une distribution symétrique	22
Figure 2 : intervalle de confiance pour une distribution asymétrique	23
Figure 3 : situation des sites	28
Figure 4 : emplacement du Schaftheu	29
Figure 5 : emplacement du Schollengiessen	29
Figure 6 : emplacement du Fahrgiessen	29
Figure 7 : emplacement de l'Eiswasser	30
Figure 8 : emplacement du Rossmoerder	30
Figure 9 : schéma du bassin d'orage du Waldweg à Rouffach	31
Figure 10 : bassin d'orage du Waldweg à Rouffach (photographie ArtWET)	31
Figure 11 : exemples d'espèces	33
Figure 12 : graphe en barres des valeurs propres de l'AFC de la banque de graines par stations	35
Figure 13 : premier plan factoriel de l'AFC de la banque de graines des stations	41
Figure 14 : graphe en barres des valeurs propres de l'AFC inter-stations de la banque	42
Figure 15 : premier plan factoriel de l'AFC inter-stations de la banque de graines	45
Figure 16 : AFC de la banque intra-stations	46
Figure 17 : AFC de la végétation intra-stations	47
Figure 18 : AFC de la végétation intra-dates	48
Figure 19 : dendrogramme de la banque de graines des stations	49
Figure 20 : dendrogramme de la banque de graines des sites (inter-stations)	50
Figure 21 : dendrogramme de la végétation moyenne des stations	51
Figure 22 : indices de Jaccard et intervalles de confiance de la comparaison banque de graines / végétation	52
Figure 23 : indices de Jaccard et intervalles de confiance de la comparaison banque de graines / végétation moyenne de chaque station sur le même site que la banque de graines	54
Figure 24 : indices de Ruzicka et intervalles de confiance pour la comparaison entre banque de graines de chaque station et végétation correspondante	77
Figure 25 : graphe des valeurs propres de l'AFC de la banque de graines du bassin d'orage	83
Figure 26 : Premier plan factoriel de l'AFC de la banque de graines du bassin d'orage	85
Figure 27 : dendrogramme de la banque de graines des stations du bassin d'orage	85
Figure 28 : indices de Jaccard et intervalles de confiance bootstrap associés	86

Bibliographie

- Benzécri J.-P. & al. (1973a)** : *L'Analyse des Données*, tome I : *La taxinomie* - Dunod.
- Borcard D. (2009)** : *Introduction à l'analyse multidimensionnelle*.
- Boyce R.-L. (2001)** : *Choosing the Best Similarity Index when Performing Fuzzy Set Ordination on Abundance Data*.
- Carpentier F.-G. (2005)** : *Introduction aux analyses multidimensionnelles*.
- Cazes P. (2004)** : *Quelques méthodes d'analyse factorielle d'une série de tableaux de données* – La revue de Modulad.
- Chessel D. et Dolédec S. (1989)** : *Rythmes saisonniers et composantes stationnelles en milieu aquatique* - Acta (Ecologica - Ecologia Generalis - volume10.
- Chessel D., Dufour A.-B., Dray S. (2009)** : *Analysis of Ecological Data : Exploratory and Euclidean methods in Environmental sciences*. (ADE-4)
- Chessel D., Dufour A.-B., Lobry J. (2006)** : *Dendrogrammes* - Fiche TD avec le logiciel R.
- Combroux I. (2002)** : *Influence de la nature et de la fréquence des perturbations sur les banques de diaspores dans les communautés végétales aquatiques* – Mémoire pour l'obtention du diplôme de doctorat.
- Dalgaard P. (2002)** : *Statistics and computing : Introductory statistics with R* - Springer.
- Davison A.C. and Hinkley D.V. (1997)** : *Bootstrap Methods and Their Application*.
- Dufrêne M. (2003)** : *Méthodes d'analyse des données écologiques et biogéographiques*.
- Efron B. and Tibshirani R. (1993)** : *An Introduction to the Bootstrap* - Chapman & Hall.
- Gayoso F. (2009)** : *Etude des liens éventuels entre les pressions subies par les stations d'Alsace et les traits de vie des macrophytes* - Mémoire pour l'obtention du diplôme de Master statistiques.
- Hinkley D.-V. (1988)** : *Bootstrap methods* - Journal of the Royal Statistical Society.
- Johns M.V. (1988)** : *Importance sampling for bootstrap confidence intervals*.
- Labatte J.-M. (2008)** : *Biostatistiques* - Rappel de cours et travaux dirigés.
- Legendre P. & Legendre L. (1998)** : *Numerical ecology*.
- Lobry J.-R. (2009)** : *Pratique de l'AFC inter-intra sur données génomiques* - Fiche TD avec le logiciel R.
- Meyer A. (2009)** : *Rôle du phosphore dans la dynamique de recolonisation par la végétation de bras reconnectés au Rhin* - Mémoire pour l'obtention du diplôme de Master PCB.
- Périnel E. (2009)** : *Analyse de données* – Notes de cours et travaux dirigés.
- Pielou E. C. (1984)** : *The interpretation of ecological data - A primer on classification and ordination* - John Wiley and sons.
- Ramousse R., Le Berre M. & Le Guelte L. (1996)** : *Introduction aux statistiques*.
- Venables W.-N. et Smith D.-M. (2007)** : *An introduction to R*.