



**HAL**  
open science

## Conséquences de la non prise en compte de la détectabilité des espèces dans les analyses ACP et AFC.

Bonsacquet Lionel

### ► To cite this version:

Bonsacquet Lionel. Conséquences de la non prise en compte de la détectabilité des espèces dans les analyses ACP et AFC.. Sciences de l'environnement. 2017. hal-01558502

**HAL Id: hal-01558502**

**<https://ephe.hal.science/hal-01558502>**

Submitted on 7 Jul 2017

**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.

**MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE**

**ÉCOLE PRATIQUE DES HAUTES ÉTUDES**

**Science de la Vie et de la Terre**

**MÉMOIRE**

**Présenté  
Par**

*Bonsacquet Lionel*

Pour l'obtention du Diplôme de l'École Pratique des Hautes Études

**CONSÉQUENCES DE LA NON PRISE EN COMPTE DE LA DÉTECTABILITÉ  
DES ESPÈCES DANS LES ANALYSES ACP ET AFC**

Soutenue le 26 juin 2017

devant le jury suivant :

M. Bruno Delesalle – Président  
M. Aurélien Besnard – Tuteur scientifique  
M. Roger Prodon – Tuteur pédagogique  
M. Stéphane Dray – Rapporteur  
M. Gérard Balent – Examineur  
M. Jean-dominique Lebreton – Examineur

Mémoire préparé sous la direction de :

M. Aurélien Besnard  
Laboratoire: BIOGÉOGRAPHIE ET D'ÉCOLOGIE DES VERTÉBRÉS  
UMR 5175 - Centre d'Écologie Fonctionnelle et Évolutive (CEFE)  
1919 route de Mende - 34293 Montpellier Cedex 5  
Directeur : M. Claude MIAUD  
EPHE (Sciences et Vie de la Terre)

Et

M. Roger Prodon  
Laboratoire: BIOGÉOGRAPHIE ET D'ÉCOLOGIE DES VERTÉBRÉS  
UMR 5175 - Centre d'Écologie Fonctionnelle et Évolutive (CEFE)  
1919 route de Mende - 34293 Montpellier Cedex 5  
Directeur : M. Claude MIAUD  
EPHE (Sciences et Vie de la Terre)

## Remerciements

Mes remerciements vont en premier lieu à mes tuteurs Roger Prodon et Aurélien Besnard. Ils ont été des guides et des soutiens sans faille. Je suis attaché au terme de guide, car il évoque des personnes ouvrant de nouveaux horizons aux néophytes, leurs évitant les écueils, tout en les laissant trébucher, car ils savent que personne n'avance sans chute. Aurélien Besnard est à l'origine de mes interrogations et de ma sensibilisation vis-à-vis des problèmes de détections des espèces. Il m'a ouvert les portes de l'EPHE et du monde de la recherche, et m'a fait confiance. Je lui en suis énormément reconnaissant. Roger Prodon a lui accepté d'être mon deuxième tuteur, et a joué un rôle important dans ma compréhension des analyses multivariées m'apportant articles et me poussant à me poser les bonnes questions. Tout deux ont toujours été à l'écoute, patients et stimulants. Je leur dois quatre ans de plaisir à chercher et d'évasion salvatrice dans le monde des mathématiques appliqués à l'écologie.

Je n'aurais cependant jamais pu mener à bien ce travail sans mon entourage. Tout d'abord mon entourage familial.

Toi isabelle je te dois tant. Tu as accepté mon quatre-vingt puis mon cinquante pour cent. Je ne peux pas lister tous les efforts que tu as dû produire pour que je puisse mener à bien cette aventure : tes corrections, ta patience, ton souci constant de me donner du temps, en prenant en charge bien plus qu'à ton tour la charge des enfants et du fonctionnement de la maison.

À toi mon frère, je ne te remercierai jamais assez pour les longs skypes malgré ta fatigue et ton emploi du temps. Ton opiniâtreté durant les relectures est exemplaire. À n'en pas douter, tu es un traceur d'avenir!

Élizabeth et Gérard, Chantal et Gérard, Corine et Cyriaque, Elsa, vos maisons et vous avez fait les frais de mes heures d'isolement et d'insomnie, ainsi que les insomnies de mes deux petits bouts. De plus, je n'ai pas été très disponible ces derniers temps. Coco tu m'as poussé à simplifier mon raisonnement, à le rendre audible sans le dénaturer. Chantal et Élizabeth vos relectures pour corriger mes fautes d'orthographe ont été très précieuses.

Annick, la famille est grande chez nous et tu en fais partie. Merci pour ton rôle répété à l'excès de nounou et de soutien lors des gardes d'Isabelle entre autres.

Mon entourage professionnel proche vient ensuite : les collègues du secteur de Modane mais aussi Thierry et Jérôme. Vous n'avez jamais été avares d'encouragement et de source d'inspiration. Malgré les conséquences pour vous, vous n'avez jamais eu un mot pour me reprocher mon choix du temps partiel, merci pour cette délicatesse.

Tous autant que vous êtes, je ne pourrai certainement pas vous remercier assez dans ces quelques lignes. Alors merci.

# Table des matières

Remerciements .....	I
Table des matières .....	1
Liste des symboles et des abréviations .....	3
Table des figures .....	4
Introduction .....	6
Chapitre 1 : La détectabilité, un enjeu majeur dans les études écologiques .....	7
1.1 Approche globale de la détectabilité .....	7
1.1.1 Définition de la détectabilité .....	7
1.1.2 Une probabilité variable .....	7
1.1.3 Impact de la PdD dans les études écologiques .....	9
1.1.4 Quelques solutions au problème de la PdD dans les mesures d'abondances .....	10
1.2 Présentation de deux analyses factorielles pour lesquelles les conséquences de la PdD sont encore inconnues .....	12
1.2.1 Présentation de l'ACP, de l'AFC .....	12
1.2.2 Utilisation et limites intrinsèques de l'ACP en écologie .....	16
1.2.3 Utilisation et limites intrinsèques de l'AFC en écologie .....	18
1.3 Une question légitime .....	19
Chapitre 2 : Matériels et méthode .....	22
2.1 Des données issues de simulations .....	22
2.2 Matériels et Méthode .....	22
2.2.1 Définition des gradients et des sites .....	23
2.2.2 Simulation des abondances brutes (non dégradées) .....	23
2.2.3 Simulation des PdD et des abondances dégradées .....	29
2.2.4 Des indicateurs de changements entre ACP ou AFC .....	32
Chapitre 3 : Présentation des résultats .....	35
3.1 Préambule à la présentation .....	35
3.2 Les résultats pour les ACP .....	37
3.2.1 Les SIMULATIONs sans gradient de détection .....	37
3.2.2 Les SIMULATIONs avec un gradient de détection linéaire .....	40
3.2.3 Les SIMULATIONs avec gradient de détection quadratique .....	49
3.3 Les résultats des AFC .....	56
3.3.1 Les SIMULATIONs sans gradient de détection .....	58
3.3.2 Les SIMULATIONs avec gradient de détection linéaires .....	60
3.3.3 Les SIMULATIONs avec gradient de détection quadratique « centré » .....	66
3.3.4 Les SIMULATIONs avec gradients de détection quadratiques « décalés » .....	73
Chapitre 4 : Analyse des résultats et discussion .....	81

4.1 Avant propos .....	81
4.2 Analyse des résultats des effets de la probabilité de détection dans les ACP .....	81
4.2.1 Les SIMULATIONS sans gradient de PdD .....	81
4.2.2 Les SIMULATIONs avec des gradients de détection linéaires .....	82
4.2.3 Les SIMULATIONs avec des gradients de détection quadratiques.....	84
4.2.4 Bilan pour les ACP .....	85
4.3 Analyse des résultats des effets de la probabilité de détection dans les AFC .....	85
4.3.1 Les SIMULATIONs sans gradient de PdD .....	85
4.3.2 Les SIMULATIONs avec des gradients de détection linéaires .....	86
4.3.3 Les simulations avec des gradients de détection quadratiques « centrés » .....	86
4.3.4 Les simulations avec des gradients de détection quadratiques « décalés » .....	88
4.4 Discussion. ....	90
4.4.1 Sur les résultats.....	90
4.4.2 Sur la méthode.....	92
4.4.3 Sur des solutions potentielles. ....	93
Conclusion.....	95
Bibliographie .....	97
Annexes A .....	I
Annexes B .....	VIII

## Listes des éléments présents dans les annexes :

### Dans l'annexe A :

Tableau I : les valeurs des variables  $g_1$  et  $g_2$  pour lesquelles chacune des 20 espèces trouve son optimum lors des SIMULATIONs des ACP

Tableau II : les valeurs des variables  $g_1$  et  $g_2$  sur les sites lors des SIMULATIONs des ACP

Tableau III : les valeurs des variables  $g_1$  et  $g_2$  pour lesquelles chacune des 20 espèces trouve son optimum lors des SIMULATIONs des AFC

Tableau IV : les valeurs des variables  $g_1$  et  $g_2$  sur les sites lors des SIMULATIONs des AFC

Tableau V : les valeurs des paramètres utilisées lors des simulations des ACP, avec des PdD sans gradient

Tableau VI : les valeurs des paramètres utilisées lors des simulations des ACP, avec des gradients de PdD linéaires

Tableau VII : les valeurs des paramètres utilisées lors des simulations des ACP, avec des gradients de PdD de formes quadratiques centrées

Tableau VIII : les valeurs des paramètres utilisées lors des simulations des AFC, avec des PdD sans gradient

Tableau IX : les valeurs des paramètres utilisées lors des simulations des AFC, avec des gradients de PdD linéaires

Tableau X : les valeurs des paramètres utilisées lors des simulations des AFC, avec des gradients de PdD de formes quadratiques centrées

Tableau XI : les valeurs des paramètres utilisées lors des simulations des AFC, avec des gradients de PdD de formes quadratiques « décalées »

### Dans l'annexe B

Page recto de la fiche d'aide à la lecture des résultats

Page verso de la fiche d'aide à la lecture des résultats

## Liste des symboles et des abréviations

$\alpha_j$ :	Ordonnée à l'origine de la droite modélisant la probabilité de détection linéaire de l'espèce j
$\beta_j$ :	Pente de la droite donnant la probabilité de détection linéaire de l'espèce j en fonction de $g_1$
$\beta_{max}$ :	Borne supérieure de la loi uniforme dans laquelle sont tirées les valeurs $\beta_j$
$\beta_{min}$ :	Borne inférieure de la loi uniforme dans laquelle sont tirées les valeurs $\beta_j$
$\gamma_j$ :	Coefficient de variation du gradient de détection de forme quadratique en fonction de $g_1$ pour l'espèces j
$\gamma_{max}$ :	Borne supérieure de la loi uniforme dans laquelle sont tirées les valeurs $\gamma_j$
$\gamma_{min}$ :	Borne inférieure de la loi uniforme dans laquelle sont tirées les valeurs $\gamma_j$
$\mu_j$ :	Valeur de la variable $g_1$ pour laquelle la probabilité de détection de l'espèce j vaut $d_{max.j}$
$v_{1j}$ :	Variance suivant $g_1$ de la loi multinormale donnant les abondances moyennes $N_{ij}$ de l'espèce j
$v_{1max}$ :	Borne supérieure de la loi uniforme dans laquelle sont tirées les valeurs
$v_{1min}$ :	Borne inférieure de la loi uniforme dans laquelle sont tirées les valeurs
$v_{2j}$ :	Variance suivant $g_2$ de la loi multinormale donnant les abondances moyennes $N_{ij}$ de l'espèce j
$v_{2max}$ :	Borne supérieure de la loi uniforme dans laquelle sont tirées les valeurs
$v_{2min}$ :	Borne inférieure de la loi uniforme dans laquelle sont tirées les valeur
$v_{detect}$ :	Écart type de la loi normale dans laquelle sont tirées les valeurs $d_{moy.j}$ ou $d_{max.j}$
$d_{ij}$ :	Probabilité de détection de l'espèce j sur le site i
$d_{max.j}$ :	Probabilité de détection maximale de l'espèce j lorsque cette PdD est de forme parabolique
$d_{max}$ :	Moyenne de la loi normale dans laquelle sont tirées les valeurs $d_{max.j}$
$d_{moy.j}$ :	Probabilité de détection de l'espèce j au centre de la portion de gradient de $g_1$ couverte par les sites lorsque le gradient de détection est linéaire
$d_{moy}$ :	Moyenne de la loi normale dans laquelle sont tirées les valeurs $d_{moy.j}$
G :	Centre de Gravité d'un nuage de points
$g_1$ :	Variable d'habitat formant le premier gradient expliquant la répartition de espèces
$g_{i1}$ :	Valeur de la variable $g_1$ sur le site i
$g_{1j}$ :	Valeur optimale de la variable $g_1$ pour l'espèce j
$g_2$ :	Variable d'habitat formant le deuxième gradient expliquant la répartition de espèces
$g_{i2}$ :	Valeur de la variable $g_2$ sur le site i
$g_{2j}$ :	Valeur optimale de la variable $g_2$ pour l'espèce j
$N_{ij}$ :	Abondance moyenne de l'espèce j sur le site i
$N_{max.j}$ :	Abondance moyenne de l'espèce j à son optimum
$N_{max}$ :	Borne supérieure de la loi uniforme dans laquelle sont tirées les valeurs $N_{max.j}$
$N_{min}$ :	Borne inférieure de la loi uniforme dans laquelle sont tirées les valeurs $N_{max.j}$
$n_{ij}$ :	Abondance de l'espèce j sur le site i dans la matrice de données non dégradée
$n_{ij}^D$ :	Abondance de l'espèce j sur le site i dans la matrice de données dégradée
$p_i$ :	Poids du point ou vecteur $x_i$
$x_i$ :	Point ou vecteur d'un espace quelconque muni d'une métrique
ACP :	Analyse en Composantes Principales
ACPd :	Résultats de l'ACP sur une matrice de données dégradée
ACPnd :	Résultats de l'ACP sur une matrice de données non dégradée
AFC :	Analyse Factorielle des Correspondances
AFCd :	Résultats de l'AFC sur une matrice d'abondance dégradée
AFCnd :	Résultats de l'AFC sur une matrice d'abondance non dégradée
IPA :	Indice Ponctuel d'Abondance
P1 :	Profil 1, la distribution des abondances moyennes des espèces ne ni de $g_1$ ni de $g_2$ .
P2 :	Profil 2, la distribution des abondances moyennes des espèces ne dépend que de $g_1$ .
P3 :	Profil 3, la distribution des abondances moyennes des espèces dépend de $g_1$ et de $g_2$ .
P4 :	Profil 4, la distribution des abondances moyennes des espèces ne dépend que de $g_2$ .
PdD :	Probabilité de détection

## Table des figures

Figure 1 : Schéma de la démarche employée pour déterminer si la probabilité de détection (PdD) a des conséquences sur les résultats des ACP et AFC.....	21
Figure 2 : coordonnées des sites et des optimums des 20 espèces en fonction des variables d'habitat $g_1$ et $g_2$ , (a) dans le cas des simulations des ACP, (b) dans le cas des simulations des AFC. ....	24
Figure 3 : schéma illustrant la simulation de l'abondance d'une espèce $j$ sur un site $i$ .....	25
Figure 4 : Figure en trois dimensions des quatre types de profils d'abondance moyenne des espèces, utilisés pour les SIMULATIONs.....	27
Figure 5 : Répartitions et numéros des SIMULATIONs en fonction des différentes hypothèses faites sur la PdD des espèces dans le cas des ACP.....	28
Figure 6 : Répartitions et numéros des SIMULATIONs en fonction des différentes hypothèses faites sur la PdD des espèces dans le cas des AFC.....	30
Figure 7 : Légendes des graphiques. ....	36
Figure 8 : graphique simple des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 1 de l'ACPnd, et les ordinations des espèces sur l'axe 1 et l'axe 2 de l'ACPd ; lors de SIMULATIONs sans gradient de PdD. ....	37
Figure 9 : graphique simple des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 2 de l'ACPnd, et les ordinations des espèces sur l'axe 1 et l'axe 2 de l'ACPd ; lors de SIMULATIONs sans gradient de PdD. ....	38
Figure 10 : graphique double des valeurs absolues des coordonnées du vecteur supplémentaire représentant $g_1$ , sur les axes 1 et 2 des ACPnd et des ACPd ; lors de SIMULATIONs sans gradient de PdD.....	38
Figure 11 : graphique double des valeurs absolues des coordonnées du vecteur supplémentaire représentant $g_2$ , sur les axes 1 et 2 des ACPnd et des ACPd ; lors de SIMULATIONs sans gradient de PdD.....	40
Figure 12 : graphique simple des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 1 de l'ACPnd, et les ordinations des espèces sur l'axe 1 et l'axe 2 de l'ACPd ; lors de SIMULATIONs avec un gradient de PdD linéaire. ....	41
Figure 13 : graphique simple des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 2 de l'ACPnd, et les ordinations des espèces sur l'axe 1 et l'axe 2 de l'ACPd ; lors de SIMULATIONs avec un gradient de PdD linéaire. ....	42
Figure 14 : graphique double des valeurs absolues des coordonnées du vecteur supplémentaire représentant $g_1$ , sur les axes 1 et 2 des ACPnd et des ACPd ; lors de SIMULATIONs avec un gradient de PdD linéaire. ....	42
Figure 15 : graphique double des valeurs absolues des coordonnées du vecteur supplémentaire représentant $g_2$ , sur les axes 1 et 2 des ACPnd et des ACPd ; lors de SIMULATIONs avec un gradient de PdD linéaire .....	43
Figure 16 : graphique double des valeurs absolues des coefficients de corrélation de Kendall entre les valeurs $d_{moy. j}$ et l'ordination des espèces sur les axes 1 et 2 des analyses ; lors de SIMULATIONs avec un gradient de PdD linéaire. ....	44
Figure 17 : illustrant les « séparations » des groupes d'espèces 1 (les 10 premières espèces) et 2 (les 10 dernières espèces), lors de SIMULATIONs avec un gradient de PdD linéaire. ....	45
Figure 18 : illustrant les « séparations » des groupes d'espèces A (les espèces paires) et B (espèces impaires), lors de SIMULATIONs avec un gradient de PdD linéaire. ....	46
Figure 19 : Graphique simple des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1 de l'ACPnd, et les ordinations des espèces sur les axes 1 et 2 de l'ACPd ; lors de SIMULATIONs avec un gradient de PdD quadratique.....	48
Figure 20 : graphique simple des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 2 de l'ACPnd, et les ordinations des espèces sur l'axe 1 et l'axe 2 de l'ACPd ; lors de SIMULATIONs avec un gradient de PdD quadratique .....	50
Figure 21 : graphique double des valeurs absolues des coordonnées du vecteur supplémentaire représentant $g_1$ , sur les axes 1 et 2 des ACPnd et des ACPd ; lors de SIMULATIONs avec un gradient de PdD quadratique.....	51
Figure 22: graphique double des valeurs absolues des coordonnées du vecteur supplémentaire représentant $g_2$ , sur les axes 1 et 2 des ACPnd et des ACPd ; lors de SIMULATIONs avec un gradient de PdD quadratique.....	51
Figure 23 : graphique double des valeurs absolues des coefficients de corrélation de Kendall entre les valeurs $d_{max. j}$ et l'ordination des espèces sur les axes 1 et 2 des analyses ; lors de SIMULATIONs avec un gradient de PdD quadratique. ....	52
Figure 24 : illustrant les « séparations » des groupes d'espèces 1 (les 10 premières espèces) et 2 (les 10 dernières espèces), lors de SIMULATIONs avec un gradient de PdD quadratique .....	53
Figure 25 : illustrant les « séparations » des groupes d'espèces A (les espèces paires) et B (espèces impaires), lors de SIMULATIONs avec un gradient de PdD quadratique.....	55
Figure 26 : graphique simple des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 1 de l'AFCnd, et les ordinations des espèces sur l'axe 1 et l'axe 2 de l'AFCd ; lors de SIMULATIONs sans gradient de PdD. ....	57

Figure 27 : graphique simple des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 2 de l'AFCnd, et les ordinations des espèces sur l'axe 1 et l'axe 2 de l'AFCd ; lors de SIMULATIONs sans gradient de PdD .....	57
Figure 28 : graphique simple des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1 de l'AFCnd, et les ordinations des espèces sur les axes 1 et 2 de l'AFCd ; lors de SIMULATIONs avec un gradient de PdD linéaire. ....	59
Figure 29 : graphique simple des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 2 de l'AFCnd, et les ordinations des espèces sur les axes 1 et 2 de l'AFCd ; lors de SIMULATIONs avec un gradient de PdD linéaire. ....	61
Figure 30 : graphique double des valeurs absolues des coefficients de corrélation de Kendall entre les valeurs dmoy. j et l'ordination des espèces sur les axes 1 et 2 des analyses ; lors de SIMULATIONs avec un gradient de PdD linéaire. ....	62
Figure 31 : illustrant les « séparations » des groupes d'espèces 1 (les 10 premières espèces) et 2 (les 10 dernières espèces), lors de SIMULATIONs avec un gradient de PdD linéaire .....	63
Figure 32 : illustrant les « séparations » des groupes d'espèces A (les espèces paires) et B (les espèces impaires), lors de SIMULATIONs avec un gradient de PdD linéaire .....	65
Figure 33 : graphique simple des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1 de l'AFCnd, et les ordinations des espèces sur les axes 1 et 2 de l'AFCd ; lors de SIMULATIONs avec un gradient de PdD quadratique centré. ....	67
Figure 34 : graphique simple des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 2 de l'AFCnd, et les ordinations des espèces sur les axes 1 et 2 de l'AFCd ; lors de SIMULATIONs avec un gradient de PdD quadratique centré. ....	68
Figure 35 : graphique double des valeurs absolues des coefficients de corrélation de Kendall entre les valeurs dmax. j et l'ordination des espèces sur les axes 1 et 2 des analyses ; lors de SIMULATIONs avec un gradient de PdD quadratique centré. ....	70
Figure 36 : illustrant les « séparations » des groupes d'espèces 1 (les 10 premières espèces) et 2 (les 10 dernières espèces), lors de SIMULATIONs avec un gradient de PdD quadratique centré .....	71
Figure 37 : illustrant les « séparations » des groupes d'espèces A (les espèces paires) et 2 (les espèces impaires), lors de SIMULATIONs avec un gradient de PdD quadratique centré. ....	72
Figure 38 : graphique simple des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1 de l'AFCnd, et les ordinations des espèces sur les axes 1 et 2 de l'AFCd ; lors de SIMULATIONs avec un gradient de PdD quadratique décalé. ....	74
Figure 39 : graphique simple des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 2 de l'AFCnd, et les ordinations des espèces sur les axes 1 et 2 de l'AFCd ; lors de SIMULATIONs avec un gradient de PdD quadratique décalé. ....	75
Figure 40 : graphique double des valeurs absolues des coefficients de corrélation de Kendall entre les valeurs dmax. j et l'ordination des espèces sur les axes 1 et 2 des analyses ; lors de SIMULATIONs avec un gradient de PdD quadratique décalé. ....	76
Figure 41 : illustrant les « séparations » des groupes d'espèces 1 (les 10 premières espèces) et 2 (les 10 dernières espèces), lors de SIMULATIONs avec un gradient de PdD quadratique décalé. ....	77
Figure 42 : illustrant les « séparations » des groupes d'espèces A (les espèces paires) et 2 (les espèces impaires), lors de SIMULATIONs avec un gradient de PdD quadratique décalé. ....	79
Figure 43 : graphiques des résultats de la millièmes simulation de la SIMULATION 3001. À gauche (a) le plan de projection de l'ACPnd, et à droite (b) le plan de projection de l'ACPd. ....	82
Figure 44 : graphiques des résultats de la millièmes simulation de la SIMULATION 3293. À gauche (a) le plan de projection de l'AFCnd, et à droite (b) le plan de projection de l'AFCd. ....	89
Figure 45 : graphiques des résultats de la millièmes simulation de la SIMULATION 3305. À gauche (a) le plan de projection de l'AFCnd, et à droite (b) le plan de projection de l'AFCd. ....	89

## Introduction

Les questions écologiques peuvent se poser à différentes échelles : celle de l'individu, de la population, de la communauté, de l'écosystème, ou encore celle du paysage (Fischesser et Dupuis-Tate 2002). Certaines études écologiques nécessitent une estimation de la taille des populations étudiées - c'est-à-dire le nombre d'individus constituant chacune des populations - pour savoir par exemple si ces dernières augmentent ou diminuent au cours des années. D'autres études s'intéressent à la richesse d'une communauté (le nombre d'espèces différentes), pour connaître l'état de la biodiversité d'un site par exemple. D'autres encore se concentrent sur la présence ou l'absence d'une ou plusieurs espèces sur un ensemble de sites.

Pour toutes ces études, une question apparaît rapidement lors de l'estimation des paramètres associés à ces populations qu'elles soient animales ou végétales : quelle est la différence entre ce qui est détecté par les observateurs et la réalité. Autrement dit, quelle est la détectabilité des espèces ou des individus sur les sites étudiés. Cette question est traitée depuis plus de 40 ans par de nombreux auteurs en écologie du fait de son importance pour obtenir des estimations d'abondance, de présence-absence ou de richesse non biaisées par la plus ou moins grande efficacité de l'échantillonnage (Pollock et al. 2002; Royle et Nichols 2003; Dorazio et Royle 2005).

Pour approfondir cette question, mon étude s'attachera à évaluer l'effet de la non prise en compte de la détectabilité dans les Analyses en Composantes Principales (ACP) et les Analyses Factorielles des Correspondances (AFC).

Dans le premier chapitre, je rappellerai ce qu'est la détectabilité et ses conséquences, et je regarderai comment celle-ci est prise ou non en compte dans différents types d'études écologiques. Je conclurai ce premier chapitre en précisant la question qui m'intéresse dans ce travail et les raisons de ce questionnement.

Le deuxième chapitre présentera les outils et méthodes d'analyses mis en place.

Le troisième chapitre présentera les résultats bruts de mon étude.

Le dernier chapitre concernera l'analyse et la discussion des résultats, et proposera quelques perspectives de recherche à relativement court terme.

# Chapitre 1 : La détectabilité, un enjeu majeur dans les études écologiques

## 1.1 Approche globale de la détectabilité

### 1.1.1 Définition de la détectabilité

Il est aisé de comprendre que lorsque l'on visite un site sur le terrain, tous les individus d'une espèce animale donnée ne sont pas détectés par l'observateur (Pollock et al. 2002). Il en va de même lors de relevés de richesse spécifique, toutes les espèces ne sont pas vues (Yamaura et al. 2011) ou même lors de relevés en présence-absence (Gu et Swihart 2004). Pour ce dernier cas, si la présence ne fait pas de doute (sauf erreur de détermination), une absence reste ambiguë : est-ce une vraie absence de l'espèce ou une non détection par l'observateur malgré la présence bien réelle de l'espèce sur le site ?

La détectabilité peut se décrire donc plus précisément comme une probabilité de détection, c'est-à-dire la probabilité de détecter une espèce ou un individu sur un site lors d'une visite sachant qu'elle/il est présent sur ce site lors de cette visite. Elle est comprise entre zéro et un, zéro signifiant que l'on ne détecte jamais rien, et un signifiant la certitude d'une détection exhaustive. Dès lors, pour la suite de cette étude, j'utiliserai le terme « probabilité de détection » -abrégé par PdD- en lieu et place de « détectabilité ».

Il me faut ici apporter une précision importante sur la nature de la PdD, et pour ce faire différencier deux cas : celui des mesures d'abondances et celui des mesures de présences-absence. Dans le premier cas (abondance), la PdD est la probabilité de détecter un individu donné de l'espèce suivie présent sur le site étudié, je la noterai  $PdD_{ind}$ . Dans le deuxième cas (présence/absence) la PdD est la probabilité de détecter au moins un individu de l'espèce en sachant que cette dernière est présente sur le site (MacKenzie et al. 2002), je la noterai  $PdD_{sp}$ . Dans la majorité des cas plus une espèce est abondante sur un site, plus la  $PdD_{sp}$  (présence de l'espèce) sera forte, alors que la  $PdD_{ind}$  (un individu particulier) ne sera, elle, pas impactée par l'abondance locale (Royle et Nichols 2003).

Au travers de très nombreuses publications, il apparaît que la PdD est souvent, pour ne pas dire toujours, inférieure à un (MacKenzie et al. 2002; Pollock et al. 2002; Royle et Nichols 2003; Chen et al. 2013). Par ailleurs il est maintenant bien admis que ce problème concerne aussi bien les oiseaux (Kéry, Royle, et Schmid 2005), les mammifères (Gu et Swihart 2004; Burton et al. 2012), les amphibiens (MacKenzie et al. 2002; Bailey, Simons, et Pollock 2004b), les squamates (Vacher-Vallas, Bonnet, et Naulleau 1999), les insectes (Kéry, Gardner, et Monnerat 2010), mais aussi les plantes (Kéry et al. 2006; Chen et al. 2009; Chen et al. 2013).

### 1.1.2 Une probabilité variable

La PdD, d'un individu ( $PdD_{ind}$ ) ou d'une espèce ( $PdD_{sp}$ ), toujours inférieure ou égale à un, est aussi dépendante de nombreux facteurs qui la rendent variable dans le temps et dans l'espace. Tous les facteurs l'affectant ne peuvent pas être cités ici. Cependant, ils peuvent être répartis en trois niveaux selon les principes de la communication : un niveau « émetteur », un niveau « environnement » qui est le canal de communication, et un niveau « récepteur » (Iknayan et al. 2014).

Au **niveau émetteur**, il faut distinguer le niveau individu et le niveau espèce. Les facteurs du niveau individu induisent une PdD différente entre deux individus d'une même espèce, alors qu'ils sont dans les mêmes conditions d'observations (même milieu et même protocole de suivi). On peut

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

citer la taille, le sexe et l'âge comme facteurs de ce niveau individuel (Iknayan et al. 2014). Par exemple, la PdD sera plus forte pour un individu adulte d'une espèce d'arbre que pour une plantule de la même espèce (Chen et al. 2013). Le comportement est aussi variable entre les individus d'une même espèce et peut impacter leur PdD. Par exemple, les mâles de vipères aspic ne cherchent pas à se reproduire tous les ans. Durant la période de reproduction, les mâles ayant un comportement reproducteur, plus actifs, sont de ce fait plus détectables que les mâles n'ayant pas ce comportement cette année-là (Vacher-Vallas, Bonnet, et Naulleau 1999). Au niveau des espèces toutes n'ont pas la même PdD. La taille, la couleur, le comportement de l'espèce sont des facteurs connus pour influencer cette probabilité (Iknayan et al. 2014). Les plantes herbacées sont moins détectables que les sous-arbrisseaux, arbustes et arbres (Chen et al. 2013). Il y a également des différences entre différentes espèces d'arbrisseaux qui n'ont pas la même morphologie : l'espèce *Myrica rubra* est plus détectable que *Neolitsea aurata* var. *chekiangensis* (Chen et al. 2009). Les espèces animales présentent les mêmes cas de figure. Lors de comptages au chant sur un point fixe et durant 10 minutes, les individus de certaines espèces d'oiseaux sont plus détectables car chantant plus fréquemment, comme le troglodyte (*Troglodytes troglodytes*) avec une probabilité proche de 0,90, par rapport aux individus des espèces chantant moins fréquemment comme le grand pic (*Dryocopus pileatus*) qui ont une PdD<sub>ind</sub> aux alentours de 0,36 (Farnsworth et al. 2002).

Les facteurs du niveau **environnement** quant à eux peuvent être vus comme des filtres qui détériorent plus ou moins le signal de l'émetteur. En premier lieu on trouve le facteur habitat. L'habitat des sites suivis facilite ou complique la détection visuelle ou auditive. S'il y a hétérogénéité dans les habitats des sites suivis, il peut par conséquent y avoir une différence de PdD entre ces sites pour une même espèce (Iknayan et al. 2014). Par exemple, la densité du feuillage peut impacter la capacité de l'observateur à entendre ou identifier le chant des oiseaux (Richards 1981 dans Farnsworth et al. 2002). La visibilité des oiseaux, et donc leur PdD, augmente avec l'augmentation de l'ouverture du milieu (Popy, Bordignon, et Prodon 2009). L'altitude est également un facteur connu. Certaines espèces de plantes montrent une PdD variable (en présence-absence) le long d'un gradient altitudinal, car au même moment elles ne sont pas au même stade de développement en fonction de l'altitude. À certains stades, comme le stade en fleur, l'espèce est plus détectable qu'à d'autres (stades rosette ou faner) (Chen et al. 2013).

Les conditions météorologiques peuvent aussi affecter la visibilité (brouillard, pluie) ou jouer sur le comportement des animaux. Les oiseaux sont connus pour moins chanter en période de vent par exemple (Iknayan et al. 2014). Il a aussi été montré que l'augmentation de la température avait pour conséquence une augmentation de la PdD de la présence de *Bufo americanus* (MacKenzie et al. 2002). Ou encore, que c'est entre 27°C et 38°C et par une journée sans vent que la PdD visuelle le long d'un parcours des individus de lézards ocellés est la plus haute (Chabanier 2011).

Un facteur du niveau environnement, plus complexe à prendre en compte et jouant sur la PdD d'une espèce, est le rôle des relations interspécifiques. La détection de la présence de la chouette tachetée du nord (*Strix occidentalis caulinata*) est, par exemple, plus faible lorsque la chouette barrée (*Strix valia*) est présente sur le même site (Bailey et al. 2009) ; certainement parce que les deux espèces sont en compétition pour les proies et parce qu'occasionnellement *Strix valia* tue *Strix occidentalis caulinata*.

Enfin, la variabilité de la PdD peut être due au **récepteur**. Ce niveau de variabilité peut aussi bien être lié à l'observateur lui-même qu'à son protocole d'observation. En effet, tous les observateurs n'ont pas les mêmes capacités de détection (Farnsworth et al. 2002). L'expérience des observateurs peut être une raison de cela. Plus un observateur a de l'expérience, plus il détectera la ou les espèces suivies. Mais l'expérience des observateurs n'est pas le seul facteur : au sein d'un groupe d'observateurs compétent, tous ne détectent pas mieux ou moins bien les mêmes chants d'oiseaux (Alldredge, Simons, et Pollock 2007). On peut parler de sensibilité des observateurs.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Également, du point de vue du protocole, la période de prospection peut influencer la PdD. Il est démontré que plusieurs espèces d'oiseaux ont des PdD de présence (lors de point de comptage de 7 minutes) différentes si les suivis sont faits en période de reproduction ou non (Tingley et Beissinger 2013). Ainsi, considérant que la période de prospection est définie par le protocole, une méconnaissance de la phénologie de reproduction de l'espèce peut induire un protocole inadapté, source de variabilité par le récepteur.

Enfin les différentes techniques d'échantillonnage varient dans leur efficacité de la détection des espèces (Bailey, Simons, et Pollock 2004a). Ainsi la probabilité de capturer des salamandres de l'espèce *Plethodon jordani* est plus élevée si on utilise la méthode qui consiste à les rechercher sous des abris naturels le long d'un parcours de 50 mètres par trois mètres par deux observateurs, que lorsqu'on relève les individus de cette espèce présents sous des couverts artificiels de différentes tailles placés le long d'un parcours de même surface (Hyde et Simons 2001).

Enfin et surtout, l'effort de prospection, défini par le protocole, est un des principaux facteurs influençant la PdD. Celle de la présence de l'écureuil roux (*tamiasciurus hudsonicus*) par capture augmente par exemple avec le nombre de jours de captures sur le site (Gu et Swihart 2004).

Ainsi, on constate que la PdD a de nombreuses sources de variabilités, aussi bien liées aux individus et aux espèces observées (niveau émetteur), qu'à l'environnement du suivi (niveau environnement), qu'aux observateurs et aux protocoles qu'ils utilisent (niveau récepteur).

### 1.1.3 Impact de la PdD dans les études écologiques

Ces difficultés de détection ont des conséquences considérables dans l'étude de l'écologie des populations. Elles affectent les mesures de variables qui permettent de suivre les populations ou les communautés, donnant ainsi des images de la réalité faussées et pouvant amener à de fausses interprétations.

Hormis les rares cas où la PdD serait de 1, si cette PdD n'est pas prise en compte, trois paramètres sont systématiquement sous-estimés.

- l'abondance des espèces,
- le taux d'occupation des espèces,
- la richesse spécifique d'un site.

Ainsi, sans prise en compte de la PdD, le nombre d'individus relevés sur le terrain ne représente en général qu'une proportion inconnue de la population cible, et il en résulte une estimation biaisée négativement de la taille de la population (Williams, Nichols, et Conroy 2002). De même les données d'occupation (proportion de surface où l'espèce est présente) faites à partir d'une ou deux occasions d'échantillonnage sont systématiquement des sous-estimations de la surface réellement occupée par une espèce (Bailey, Simons, et Pollock 2004a). Enfin, pour ce qui est de la richesse en espèces d'une communauté, du fait d'une PdD des espèces inférieures à un, l'énumération complète de toutes les espèces présentes est rarement faisable (Dorazio et al. 2006).

Les résultats de nombreuses études illustrent ces différentes sous-estimations. La prise en compte de la PdD dans le calcul de l'abondance de huit espèces d'oiseaux montre, que sur le site étudié, les densités (nombre d'individus par  $km^2$ ) passent de 0.22 à 1.95 pour le gros bec (*Coccothraustes coccothraustes*) et de 0.41 à 0.43 pour le colvert (*Anas platyrhynchos*) par exemple (Kéry, Royle, et Schmid 2005). Après correction des effets de la PdD, la probabilité de présence de l'Aeshne bleue (*Aeshna cyanea*), un odonate commun en Suisse, augmente de 60% sur le territoire Suisse (Kéry, Gardner, et Monnerat 2010). En effet, si les fausses absences ne sont pas corrigées, elles sont comptabilisées comme de vraies absences, d'où une augmentation artificielle de la proportion de territoire non occupé. Pour la richesse, l'étude de la PdD sur une communauté d'oiseaux où 75 espèces différentes ont été détectées, montre que la probabilité que le nombre d'espèces soit en fait exactement de 75 est presque nulle ; une estimation corrigée de la richesse de cette communauté affiche une médiane de 90 et une moyenne de 93 espèces (Dorazio et al. 2006).

Si les paramètres abondance, richesse ou présence-absence sont biaisés, toutes les déductions faites à partir de ces paramètres le seront également, et notamment les inférences faites à partir de modèles linéaires généralisés. La variable à expliquer ayant des valeurs biaisées, les relations avec les variables explicatives ne pourront être justes. Ainsi les données de présence-absence sont utilisées dans les analyses des relations espèces-habitats, qui relient la probabilité de présence de l'espèce à des co-variables d'habitat. La non prise en compte de la PdD provoque dans tous les cas un biais dans les résultats de ces analyses. Ce biais peut être positif ou négatif, plus ou moins important. Son importance dépend de plusieurs facteurs : l'importance des co-variables d'habitat dans la distribution de l'espèce, le fait que la PdD soit elle aussi en relation ou non avec les co-variables d'habitat, le fait que l'espèce soit rare, commune ou abondante (Gu et Swihart 2004). Les mêmes liens peuvent être établis dans les relevés d'abondances ou de richesse spécifiques (Reino et al. 2009), et la PdD peut donc y induire les mêmes biais.

De plus, certaines études se basent sur la comparaison de paramètres dans le temps ou l'espace. Si une PdD inférieure à un a pour effet les sous-estimations des trois paramètres présentés ci-dessus, sa variabilité empêche quant à elle les comparaisons dans le temps et l'espace.

Quand la PdD n'est pas estimée, on ne peut savoir si la comparaison porte sur l'évolution du paramètre d'intérêt, ou sur l'évolution de la PdD elle-même, ou un mélange des deux (MacKenzie et al. 2002 ; MacKenzie et Nichols 2004).

Pour exemple, la comparaison de l'abondance d'une espèce sur un même site à deux périodes différentes informe théoriquement sur l'évolution de la taille de la population sur ce site. Cette comparaison dans le cas de suivi d'une seule espèce, sans correction des abondances mesurées par la PdD, implique l'hypothèse implicite que la valeur de la PdD est constante dans le temps (Pollock et al. 2002, Bailey, Simons et Pollock 2004a, Couturier et al. 2013). Cette hypothèse est pour le moins discutable (pour de très nombreuses raisons comme des effets observateurs, météo, modification du milieu etc.) et doit être évaluée (Kéry, Royle et Schmid 2005).

De même, la variabilité de la PdD fausse la comparaison des données de présence-absence, et ce pour la même raison que pour les comparaisons des indices d'abondance (MacKenzie et Nichols 2004). Or, la comparaison de données de présence-absence d'une espèce sur un même site, à deux périodes différentes, permet par exemple de calculer la probabilité d'extinction de l'espèce (Kéry et al. 2006). Un exemple de biais : une prise en compte de la PdD sur la mesure de la probabilité d'extinction de 11 espèces végétales ramène cette probabilité de 0.49 en moyenne (de 0.27 à 0.67) à 0.36 (de 0.09 à 0.61) (Kéry et al. 2006).

### 1.1.4 Quelques solutions au problème de la PdD dans les mesures d'abondances

Étant donné l'importance des biais qu'occasionnent les problèmes de détection pour l'étude des populations, les biostatisticiens ont été amenés à développer depuis les années 1960 des méthodes permettant de gérer ces difficultés, notamment en proposant des méthodes qui visent à estimer les PdD. Parmi ces méthodes je vais en citer trois : la méthode dite de Capture-Marquage-Recapture (CMR) (Pollock 1980), Le « *distance sampling* » (Buckland et al. 1993) et Le *N-mixture* (Royle 2004).

Les méthodes dites CMR sont les plus connues des méthodes qui gèrent les problèmes de détection des individus et celles qui ont connu le plus de développement. On s'intéresse ici aux modèles qui s'appliquent à une population fermée, c'est-à-dire une population qui durant la période de suivi ne présente ni naissance ni mort, ni phénomène migratoire (émigration/immigration). Ces modèles ont pour objectif une estimation de la taille de la population et donnent également une estimation de la PdD des individus, qui est alors une probabilité de capture.

Dans le cas le plus général, le déroulement est le suivant : à chaque prospection d'un site, que sont les sessions de captures, tous les individus capturés pour la première fois sont marqués de manière à ce que chaque individu soit identifiable. La prospection doit être faite de manière

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

aléatoire sur l'ensemble du site d'étude. Toutes les captures, de la première à la dernière, de tous les individus sont enregistrées. Tous les individus sont relâchés après chaque session. On constitue alors pour chaque individu identifié son histoire de capture. À partir des histoires de capture, on obtient la taille de la population et la PdD à l'aide de modèles statistiques tels que décrit par Pledger (2000) par exemple.

Les hypothèses sont que (1) la population est fermée entre la première et la dernière session, (2) qu'aucun individu marqué ne perd sa marque, (3) que l'enregistrement des histoires de capture soit sans erreur, (4) que les individus sont capturés de manière indépendante.

Le modèle de base  $M_0$  fait les hypothèses (a) que tous les individus ont une même probabilité de capture lors d'une session de capture, (b) que la probabilité de capture ne change pas entre les sessions. Des développements ont permis de relâcher un peu ces hypothèses et ont donné de nouveaux modèles, comme par exemple le modèle  $M_t$  qui ne prend pas en compte l'hypothèse (b).

La CMR est considérée comme non biaisée dès lors que les hypothèses du modèle sont respectées (Otis et al. 1978 dans Couturier et al. 2013). Otis et al. (1978) donne une estimation du nombre minimum de sessions de capture et de la valeur de la probabilité de capture la plus basse qu'accepte un modèle pour qu'il ne soit pas biaisé.

Cependant une perte des marques aura pour effet une surestimation de la population (Otis et al. 1978) ; et pour vérifier l'hypothèse d'une population close on est souvent contraint d'organiser l'ensemble des sessions de captures sur de courtes périodes (Pollock 1980).

La CMR est une méthode coûteuse en termes d'effort de prospection, en temps, et peut demander un investissement lourd pour les captures. C'est pourquoi elle est difficile à appliquer à de grandes échelles (MacKenzie et al. 2002; Bailey, Simons, et Pollock 2004b), ou pour plusieurs espèces en même temps. De surcroît, le marquage est parfois difficile. Par ailleurs, si les populations sont très petites ou si les individus ont des probabilités de capture très basses, il est difficile d'obtenir suffisamment de données pour assurer des estimations avec une précision raisonnable de la taille de ces populations (Royle 2004).

Le « *distance sampling* » est une autre méthode qui permet de déterminer l'abondance sans nécessiter le marquage des individus ni de répétition dans le temps. Elle répond donc à certains problèmes posés par la CMR. Cette méthode repose sur le fait que, plus un individu est loin de l'observateur, plus sa PdD diminue (Buckland et al. 1993; Borchers, Buckland, et Zucchini 2004).

Concrètement, sur un site à suivre, les observateurs pratiquent un sondage depuis des lignes ou des points de comptages. Ces points ou lignes sont disposés de manière aléatoire. Ils ne couvrent donc pas toute la superficie mais seulement une portion. Pour chaque observation d'individus ou groupe d'individus, les observateurs mesurent la distance perpendiculaire qui sépare l'objet détecté de la ligne de comptage ou la distance radiale dans le cas d'un point de comptage. À partir de ces informations recueillies sur le terrain, il est possible de déterminer l'abondance de une ou des espèces suivies en prenant en compte la PdD à l'aide de calculs détaillés dans Buckland et al. (1993).

On retrouve les hypothèses les plus importantes de cette méthode à respecter pour que les résultats ne soient pas biaisés dans Pollock et al. (2002) ou Marques et al. (2011) : (1) La PdD des individus sur le parcours de l'observateur (noté  $g(0)$ ) est de un. (2) Les animaux n'ont pas de mouvement avant leur détection ou bougent moins vite que l'observateur. (3) Les mesures des distances sont faites sans erreur. (4) La détection d'un individu est indépendante de la détection d'un autre.

Marques et al. (2011) rappellent quelques conséquences si on ne respecte pas ces hypothèses. Une sous-estimation de la densité de la population est constatée dans les cas où :

- $g(0) < 1$ ,
- les distances sont surestimées,
- il y a mouvement de fuite des individus avant la détection.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Inversement, il y a surestimation de la taille de la population dans les cas où on sous-estime les distances, ou s'il y a un mouvement de rapprochement des individus avant leurs détections, par exemple.

L'hypothèse que  $g(0)=1$  est parfois difficile à obtenir, dans le cas par exemple (mais pas uniquement) des espèces cryptiques (Couturier et al. 2013). L'hypothèse que l'individu se déplace lentement par rapport à l'observateur exclut *de facto* certaines espèces, par exemple les odonates qui ont un vol rapide.

Le *N-mixture* (Royle 2004) est une méthode plus récente permettant d'estimer les abondances et la PdD des individus. Elle part du constat que les tableaux de comptage d'une espèce sur plusieurs sites se caractérisent par un grand nombre de zéros et des indices de comptage de petites valeurs (Royle 2004) ; et que le fait d'utiliser directement ces comptages, sans correction des effets de la PdD dans les analyses, biaise les résultats. Cette méthode nécessite des répétitions de sessions de comptage sur tous les sites suivis mais ne nécessite pas de marquage des individus, ni mesure de distance. Elle peut donc être utilisée à grande échelle.

Ce modèle, développé pour le suivi d'une seule espèce a été étendu depuis aux cas de suivis multi-espèces (Dorazio et Connor 2014; Dorazio, Connor, et Askins 2015). On ne présente ici que le modèle de départ, c'est-à-dire à une espèce.

L'auteur fait les hypothèses (1) d'une population fermée sur chaque site, (2) de l'indépendance des individus, (3) de l'indépendance des sites, et (4) de l'indépendance des sessions de comptages.

On dispose donc sur plusieurs sites de données de comptages répétés dans le temps d'une espèce. À partir de ces données, il est possible d'estimer l'abondance de l'espèce et la PdD sur chaque site suivi à l'aide de calculs détaillés dans Royle (2004). Cette méthode permet, de plus, de faire varier la PdD en fonction de co-variables de temps ou d'habitat et l'abondance en fonction de co-variables d'habitat.

On retrouve l'emploi de cette méthode entre autres dans Royle, Nichols et Kéry (2005), Kéry, Royle et Schmid (2005) et Couturier et al. (2013).

On note toutefois que lors d'une comparaison des trois méthodes (CMR, distance sampling et N-mixture), dans le cadre d'un suivi d'une population de tortue d'hermann, le N-mixture apparaît comme potentiellement fortement instable pour des valeurs de PdD inférieures à 0.5 (Couturier et al. 2013).

Globalement on peut noter que, comme le « *distance sampling* » et le « *N-mixture* », des méthodes peuvent prendre en compte la problématique de la PdD pour des suivis multi-espèces et multi-sites. Cependant certaines de ces études multi-espèces et multi-sites n'intègrent pas cette problématique. Parmi celles-ci on trouve celles qui utilisent les ACP ou les AFC.

## 1.2 Présentation de deux analyses factorielles pour lesquelles les conséquences de la PdD sont encore inconnues

Si les biais que provoque la PdD dans certaines analyses de données écologiques sont connus (cf. partie 1.1.3), la question reste entière pour les ACP (Analyse en Composantes Principales) et les AFC (Analyse Factorielle des Correspondances). Avant d'aborder cette question il me faut présenter ces deux types d'analyses.

### 1.2.1 Présentation de l'ACP, de l'AFC

Il est courant en écologie (des communautés) d'avoir à analyser des tableaux regroupant l'abondance ou la présence-absence de plusieurs espèces sur plusieurs sites. Les sites sont généralement présentés en lignes et les espèces en colonnes. Il existe également des tableaux regroupant les mesures de plusieurs variables, sur différents sites (variables d'habitat comme le pH,

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

l'ouverture du milieu, etc.) ou pour différentes espèces (mesures de morphométries par exemple). Dans ce cas, les variables sont présentées en colonne.

A partir d'un de ces tableaux, dit matrice, on peut chercher quelles sont les ressemblances entre les sites, entre les espèces ou entre les variables descriptives (Legendre et Legendre 1998). Par exemple, pour une matrice d'abondance, on peut se demander quels sont les sites qui ont une même composition en espèces, et pour les espèces quelles sont celles qui utilisent les mêmes sites. On peut aussi chercher à mettre en évidence une ou des variables environnementales sous-jacentes à l'organisation des données de la matrice (des sites ou des espèces). Ces variables environnementales hypothétiques sont dites variables latentes. Les analyses multidimensionnelles, sont des analyses descriptives qui permettent de traiter ce type de tableaux et de répondre à ces questions.

Le terme d'analyses multidimensionnelles regroupe un nombre important de techniques. On peut d'ores et déjà séparer ces techniques en deux groupes : les classifications et les analyses factorielles. Une classification a pour but de regrouper en plusieurs classes les individus statistiques de telle sorte que les individus d'une même classe soient le plus semblables possible et que les classes soient le plus distinctes possible. Une analyse factorielle produit, elle, des ordinations d'individus, de variables ou de modalité de variable qualitative, susceptibles de révéler des facteurs explicatifs.

**Les similarités ACP et AFC.** Les ACP et AFC sont des analyses factorielles. Je ne détaillerai donc pas les classifications. L'ordination des individus, des variables ou des modalités le long d'un axe que proposent les analyses factorielles permet de synthétiser la structure de la matrice de données. Appliquées en écologie sur des tableaux d'abondance, ces méthodes « établissent une typologie des relevés selon leur composition spécifique, et une typologie des espèces en fonction de leur répartition dans les relevés. L'objectif de ces approches consiste à révéler l'existence d'une structure mais également à déterminer quels facteurs environnementaux principaux en sont la cause » (Dray 1999). L'ordination permet donc, entre autres, l'identification des variables latentes.

Chaque analyse factorielle se base sur une ou des mesures de dissimilarité. Une mesure de dissimilarité permet d'évaluer la ressemblance ou la différence entre deux sites, deux espèces ou deux variables. L'ensemble de ces mesures est détaillé, entre autres, dans Legendre et Legendre (1998) et Zuur, Ieno, et Smith (2007). Cependant, les ACP considèrent les lignes de la matrice de données (les sites en général) comme des individus statistiques, et les colonnes comme des variables, alors que l'AFC considère respectivement les lignes et les colonnes de la matrice comme les modalités de deux variables qualitatives distinctes. Or, on ne peut utiliser la même mesure de dissimilarité si la question porte sur des individus statistiques, des variables, ou des modalités de variables qualitatives. Les ACP utilisent donc deux mesures de dissimilarité, une pour les sites et une pour les variables, et les AFC une mesure de dissimilarité encore différente de celles des ACP. Je présenterai les mesures de dissimilarités propres à chaque analyse (ACP et AFC) dans les paragraphes qui les concernent spécifiquement.

En plus de toutes se servir de mesures de dissimilarité, les analyses factorielles ont en commun l'utilisation de deux notions : le centre de gravité et l'inertie. Pour comprendre ces deux notions, on se place dans un espace muni d'une norme (ou métrique). Une mesure de la distance entre deux points de cet espace est donc possible. On considère alors  $I$  points ou vecteurs (chaque point représentant l'extrémité d'un vecteur) de cet espace noté  $x_i$  ( $i=1, \dots, I$ ). Chacun de ces points est affecté d'un poids  $p_i$  ( $i=1, \dots, I$ ), avec  $p_i > 0$  et  $\sum_i^I p_i = 1$ . Le centre de gravité de cet ensemble de points, noté  $G$ , est alors défini par la somme vectorielle suivante :

$$G = \sum_i^I p_i \cdot x_i$$

Le centre de gravité peut s'interpréter comme une généralisation de la notion de moyenne (Cailliez et Pages 1976).

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Quant à l'inertie d'un ensemble (ou nuage) de points, elle peut être calculée par rapport à un point, un axe, un plan ou un hyperplan. On note  $D^i$  la distance du point  $x_i$  au point, axe, plan ou hyperplan noté  $u$ . L'inertie du nuage par rapport à  $u$  a pour formule :

$$I_u = \sum_i^I p_i (D^i)^2$$

On appelle inertie totale du nuage de points, l'inertie du nuage calculé par rapport à son centre de gravité. La décomposition du calcul de l'inertie totale à l'aide du théorème de Pythagore permet de montrer qu'elle est également la somme des inerties du nuage par rapport aux axes deux à deux orthogonaux et passant par le centre de gravité. L'inertie peut s'interpréter comme une généralisation de la notion de variance (Cailliez et Pages 1976).

La pondération des points d'un nuage a donc son importance dans les calculs du centre de gravité et de l'inertie. Les ACP et les AFC n'appliquent pas les mêmes pondérations. Je présenterai les pondérations propres à chaque analyse (ACP et AFC) dans les paragraphes qui les concernent spécifiquement.

Ainsi l'ACP et l'AFC ont en commun de construire, à partir de la matrice de données, deux nuages de points. Un nuage sera composé de points représentatifs des lignes de la matrice de données (les sites en général). Ce nuage se trouve dans un espace ayant un nombre de dimensions égal, dans le cas de l'ACP, au nombre de colonnes de la matrice. L'autre nuage sera composé de points représentatifs des colonnes de la matrice de données (les espèces en général). Ce deuxième nuage se trouve dans un espace ayant un nombre de dimensions égal au nombre de lignes de la matrice. Ces nuages ayant de nombreuses dimensions qui ne peuvent toutes être représentées sur un plan, il importe d'en simplifier la représentation tout en conservant un maximum d'informations. L'idée est de projeter ces nuages de points sur des plans, produits par les deux axes orthogonaux les plus représentatifs des ressemblances entre points-lignes et des ressemblances entre points-colonnes.

L'ACP et l'AFC consistent en la recherche d'une suite d'axes deux à deux orthogonaux et qui, une fois les points projetés sur ces axes, conservent au mieux les distances entre les points ou les angles entre les vecteurs. L'inertie sert de mesure de proximité entre un axe et les points du nuage. Cailliez et Page (1976) détaillent la méthode d'obtention des axes de projections.

L'ACP et l'AFC décomposent donc l'inertie totale du nuage, en la répartissant sur des axes deux à deux orthogonaux. Chaque axe porte donc une portion de l'inertie totale du nuage que l'on appelle la part d'inertie expliquée par l'axe. Plus cette dernière est importante, plus l'axe donne une représentation proche du nuage d'origine.

Des relations mathématiques appelées relations de transitions peuvent permettre de retrouver les résultats de l'analyse d'un des nuages à partir des résultats de l'autre (Escofier et Pagès 2008).

**Les spécificités de l'ACP.** En revanche l'ACP s'applique sur un tableau regroupant les mesures de plusieurs variables sur plusieurs individus statistiques, et l'AFC s'applique, elle, sur un tableau de contingence. Par conséquent elles diffèrent par trois points :

- la mesure de dissimilarité, (cf pages précédentes),
- la méthode de construction des nuages de points,
- et le poids des points dans les nuages.

Ces différences font que ces deux analyses n'apportent pas les mêmes informations sur les ressemblances et les relations entre les différents éléments de la matrice de départ.

Dans le cas de l'ACP d'une matrice d'abondance, on regarde les sites comme des individus statistiques et les espèces comme des variables. Une question porte alors sur les individus : quelles sont leurs ressemblances ; et une question porte sur les variables : quelles sont leurs relations (Dray

2014). La mesure de la proximité entre les sites est la distance euclidienne (somme des différences entre deux lignes pour toutes les variables). Plus cette distance est grande entre deux sites, plus ceux-ci diffèrent. La mesure des relations entre les espèces est le coefficient de corrélation linéaire (corrélation entre deux colonnes pour tous les sites). Plus ce dernier est élevé entre deux espèces, plus celles-ci varient ensemble (co-varient).

Pour la suite de mon étude, je fais le choix de me placer dans le cas des ACP centrées et réduites. « Centrer » consiste à mettre le centre de gravité du nuage des individus à l'origine du repère, point où l'inertie est minimum (Cailliez et Pages 1976). L'origine du repère peut s'interpréter comme le site moyen (site théorique où toutes les espèces sont à leurs abondances moyennes). « Réduire » (ou « normer ») consiste à donner à toutes les variables une norme identique, égale à un, donc la même importance. Dans le cas des tableaux d'abondance ou toutes les variables ont une unité commune, l'ACP peut ne pas être normée, mais il est souvent préférable de normer quand même afin que les espèces à forte variance ne s'imposent pas trop dans les résultats. A contrario (dans une ACP normée), une forte proportion d'espèces rares peut influencer indûment l'analyse et celles-ci auront tendance à dominer les résultats (Jongman, Ter Braak, et Van Tongeren 1995). Du fait de la réduction, l'inertie totale de chacun des nuages est égale au nombre de variables. Pour ce qui concerne les pondérations dans une ACP centrée réduite, tous les sites ont un poids égal à l'inverse du nombre total de site, et toutes les variables ont un poids égal à un.

À la fin de l'analyse j'obtiens pour le nuage des individus des axes de projections appelés axes principaux. Ces axes sont ordonnés par ordre décroissant d'inertie expliquée. Le plan formé par les axes principaux d'ordre un et deux est le plan qui a la plus grande inertie projetée possible. Il est le plan de projection qui respecte le plus les distances entre les individus. **Ce plan permet d'avoir la meilleure image possible des proximités des différents sites du point de vue des abondances des espèces.**

Pour le nuage des variables, les axes de projections sont appelés composantes principales. Ces composantes principales sont ordonnées par ordre décroissant d'inertie expliquée. Les composantes principales sont des combinaisons linéaires des variables (espèces). Ce sont donc également des variables. La composante principale d'ordre un est celle qui a la meilleure corrélation avec l'ensemble des variables de la matrice de départ. Le plan de projection constitué des composantes principales d'ordre un et deux **donne l'image la moins déformée des corrélations linéaires entre les variables.** Seules les variables qui ont une norme assez proche de 1 dans le plan de projection sont bien représentées. La coordonnée de la projection d'une variable sur une composante principale est égale au coefficient de corrélation entre cette variable et la composante principale.

De plus, dans le cas des ACP « lorsque l'on regarde simultanément les deux graphiques, un individu est du côté des variables pour lesquelles il a de fortes valeurs et à l'opposé des variables pour lesquelles il a de faibles valeurs » (Escofier et Pagès 2008). Ceci permet la mise en évidence des relations site-espèces, en plus des ordinations spécifiques aux sites et aux espèces.

**Les spécificités de l'AFC.** À la différence de l'ACP qui s'applique à des variables quantitatives relevées sur différents individus, l'AFC est conçue pour s'appliquer à un tableau de contingence, c'est-à-dire au croisement de deux variables qualitatives. Dans un tableau de contingence, chaque ligne correspond à une modalité d'une des variables qualitatives, et chaque colonne correspond à une modalité de l'autre variable. Dans chaque case du tableau, on indique le nombre d'individus statistiques qui ont à la fois la modalité qui correspond à la ligne et la modalité qui correspond à la colonne.

Dans le cas des tableaux d'abondances, l'AFC travaille sur les mêmes données que l'ACP, mais en les considérant comme issues d'un tableau de contingence. Chaque site (ligne) est alors une modalité d'une variable « site », et chaque espèce (colonne) une modalité de l'autre variable « espèce ». Les individus statistiques coïncident avec les individus biologiques. Les questions portent alors sur la ressemblance entre deux modalités d'une même variable, et comment s'associe la modalité d'une variable avec une modalité de l'autre variable (Escofier et Pagès 2008). Ainsi à la

différence de l'ACP, en AFC les colonnes et les lignes jouent des rôles symétriques et représentent des objets similaires : des sous-populations de la population étudiée. Le traitement et la lecture des résultats sont donc complètement similaires entre les deux nuages de points, l'un issu des lignes et l'autre des colonnes du tableau de contingence. On ne décrira donc ici que la démarche pour le nuage issu des lignes du tableau.

Afin d'obtenir ce nuage de points, l'AFC nécessite en premier lieu une transformation des données brutes. Plusieurs approches et donc plusieurs transformations des données brutes sont possibles, mais toutes conduisent au même résultat. Une transformation possible consiste à transformer les données en proportion ; c'est-à-dire que l'on calcule pour chaque espèce et sur chaque site la proportion que l'espèce représente de la population totale du site. Escofier et Pagès (2008) et Dray (2014) détaillent ces transformations.

Chaque point du nuage représente donc un site, et donc une modalité de la variable qualitative « site ». La mesure de dissimilarité alors utilisée entre les sites est la distance du  $\chi^2$ . Pour ce qui est des pondérations, chaque site a un poids égal à la probabilité pour un individu statistique d'être dans ce site. Ce poids est imposé en AFC et est donc dépendant des effectifs de chaque ligne.

Le centre de gravité du nuage est confondu avec l'origine du repère. Ce centre de gravité représente l'hypothèse d'indépendance des deux variables qualitatives. Deux points sont donc d'autant plus dépendants l'un de l'autre, que ces points s'écartent de l'origine du repère dans la même direction et avec la même norme. À partir d'un tableau d'abondance, ceci revient à dire que **deux points représentant deux sites sont d'autant plus proches que la distribution en proportion des espèces est la même sur les deux sites.**

L'AFC fournit alors, pour le nuage issu des lignes, des axes de projections qui sont ordonnés par ordre décroissant d'inertie expliquée. Le plan formé par les axes d'ordre un et deux est le plan qui a la plus grande inertie projetée possible. Il est le plan de projection qui respecte le plus les distances entre les points. Dans le cas d'une matrice d'abondance, ce plan permet d'avoir **une image des proximités des différents sites du point de vue de la répartition en proportion des espèces**, proximité donnée alors par une distance euclidienne.

La même démarche est adoptée avec le nuage des points issus des colonnes. Dans le cas d'une matrice d'abondance, les plans de projections permettent d'avoir **une image des proximités des différentes espèces du point de vue de leurs répartitions en proportions sur les différents sites**. Le plan formé par les axes d'ordre un et deux est le plan de projection qui respecte le plus les distances entre les points.

Les relations de transitions justifient la présentation superposée des résultats de l'analyse du nuage des lignes et du nuage des colonnes sur des plans engendrés par des axes de même ordre.

## 1.2.2 Utilisation et limites intrinsèques de l'ACP en écologie

En écologie, l'ACP peut être utilisée de deux manières différentes. Premièrement, comme une méthode de réduction du nombre de variables en une ou des variables synthétiques. L'ACP met alors en évidence la corrélation entre les variables. Les composantes principales étant des combinaisons linéaires des variables de la matrice de données et étant orthogonales entre elles, donc non corrélées, elles peuvent être utilisées comme nouvelles variables dans d'autres analyses. C'est dans cet objectif par exemple que Reino et al. 2009 utilisent l'ACP. Ils résument 16 variables environnementales en 5 variables orthogonales qu'ils utilisent par la suite dans un modèle linéaire généralisé (glm).

Deuxièmement, l'ACP est utilisée comme méthode pour obtenir une typologie des individus statistiques. Ces individus sont souvent des sites de comptage. On regarde les individus qui se ressemblent et ceux qui diffèrent. On peut regarder ces ressemblances à l'aide des projections des points sur les plans principaux ou se contenter de l'ordination des individus sur les axes de projections. On doit également être en mesure de comprendre la signification écologique (si elle

existe) des axes de projections qui permettent cette ordination. Si ces axes ont un sens, on peut alors parler de variables latentes (variables environnementales hypothétiques).

L'ordination des relevés par une ACP faite sur un tableau « sites × espèces » est pratiquée par exemple en hydrobiologie (Lebreton et al. 1988). Elle permet également de déterminer quelles espèces sont corrélées entre elles. Un test de l'ACP sur des relevés d'abondance d'invertébrés aquatiques sur 20 stations permet de mettre en évidence une ordination des stations qui se succèdent le long d'une structure (non linéaire) corrélée à un gradient de la qualité de l'eau (Dupiereux, Feytmans, et Micha 1983). L'ACP, malgré la non linéarité des relations entre les abondances des espèces de cette étude, apparaît comme un bon outil de classement des stations écologiques dans un espace réduit (Dupiereux, Feytmans, et Micha 1983).

En effet, les ACP donnent une ordination des individus avec la meilleure combinaison linéaire des variables. Il faut donc que cette relation linéaire (des variables) ait un sens écologique. Ainsi il est possible d'utiliser l'ACP (en vue de déterminer une variable latente) pour les tableaux « sites × espèces » à la condition que ce soit sur un gradient court (Prodon et Lebreton 1994). En effet, sur un gradient court sur lequel les espèces présentent des gradients d'abondance sinon linéaires du moins monotones, il y a de grandes chances que l'optimum de l'espèce soit en dehors ou en limite de la zone prospectée. Ainsi les variables-espèces sont soit croissantes soit décroissantes et respectent la notion de relation linéaire.

Il importe maintenant de prendre en compte plusieurs effets indésirables lors de l'utilisation de l'ACP. Un des effets est lié à l'utilisation de la distance euclidienne. Cette distance donne la même importance à toutes les valeurs d'abondance et entre autres aux valeurs nulles (zéro). Si deux sites ont chacun une valeur nulle d'abondance pour une même espèce, ces deux zéros ne font pas augmenter la distance entre les deux sites (la dissimilarité). Si ces deux zéros ont les mêmes raisons écologiques, tout va pour le mieux. Si ce n'est pas le cas (un zéro dû à un mauvais comptage conséquence d'une mauvaise météo, et un zéro dû à un habitat très défavorable pour cette espèce par exemple) la similitude est trompeuse voir fausse. C'est le problème des doubles zéros (Legendre et Legendre 1998). L'ACP est sensible aux doubles zéros (Prodon et Lebreton 1994). Avec un grand nombre de zéros dans la matrice d'abondance, deux sites de la matrice même bien différents écologiquement auront des zéros en commun, ce qui aura pour conséquence de donner une distance faible entre ces deux sites qui se retrouveront alors proches. L'ACP est donc mal adaptée aux matrices contenant beaucoup de zéros (Prodon et Jacquet 2014).

L'effet connu sous le vocable de l'effet « fer à cheval » est une expression de ce phénomène. Sur un gradient au long duquel il y a un changement important des espèces, les sites du début du gradient et ceux de la fin du gradient ont beaucoup de zéros en commun. Les sites en positions extrêmes sur le gradient se retrouvent côte à côte dans le premier plan de l'ACP. Dans ce cas, les sites sont disposés en forme de fer à cheval. Ceci renforce l'importance de travailler sur des gradients courts pour l'ACP sur des tableaux d'abondances.

Dans l'ACP, un dernier effet connu est appelé « l'effet taille ». Si un site de comptage n'a que peu d'individus pour chaque espèce (c'est-à-dire en dessous de la moyenne pour chaque espèce) et un autre de nombreux individus pour toutes les espèces, l'ACP séparera ces deux sites. Si cela ne se produit pas pour deux sites mais pour deux groupes de sites, les sites restant ayant un comportement intermédiaire, alors le premier axe principal sépare ces deux groupes et la première composante principale montre une forte corrélation positive de toutes les variables (les espèces). Cela peut traduire la différence entre milieux pauvres et milieux riches. Cela peut aussi être la conséquence de comptages erronés sur certains sites du fait de mauvaises conditions météorologiques. Lorsque cet effet apparaît, il est possible de regarder le plan formé par les axes 2 et 3 de l'analyse. Ce plan peut être vu comme l'analyse de ce qui n'est pas expliqué par le premier axe (analyse des résidus). Ce plan 2-3 est directement une analyse de la forme (Chessel, Dufour et Thioulouse 2003). Une autre solution est de faire l'analyse en centrant également le nuage des variables : le tableau de départ est alors doublement centré. Ceci a pour conséquence la suppression de la corrélation positive de toutes les espèces et donc de l'effet taille.

En définitive, les méthodes faisant l'hypothèse de relations linéaires entre les variables, dont l'ACP, n'ont trouvé que des applications limitées en écologie, car généralement les espèces ont des réponses non linéaires ou non monotones aux variables environnementales (Ter Braak et Prentice 2004). Des critiques ont également été adressées à l'encontre de l'utilisation de la distance euclidienne sur les données d'abondances à cause du « comportement » de cette distance face aux doubles zéros (Legendre et Legendre 1998). Il en résulte que l'utilisation de l'ACP sur les tableaux d'abondances diminue au cours du temps depuis les années 1990 dans les études sur la végétation, alors que cette technique est restée utilisée de manière prépondérante pour la réduction du nombre de variables environnementales (Von Wehrden et al. 2009). On peut supposer que cette tendance est généralisable pour toutes les espèces autres que végétales. Cependant, avec les précautions nécessaires, l'utilisation de l'ACP sur des matrices d'abondances reste tout à fait justifiable. Une autre méthode multidimensionnelle répond aux critiques de l'hypothèse des relations linéaires entre les variables et du traitement des doubles zéros : c'est l'AFC.

### 1.2.3 Utilisation et limites intrinsèques de l'AFC en écologie

Dans le traitement de tableaux « sites × espèces », l'AFC peut être utilisée de deux manières. Et dans les deux cas, comme pour les ACP, si les axes d'ordinations ont un sens écologique, on peut alors parler de la mise en évidence de variables latentes.

Dans le premier cas, l'AFC peut être vue comme une méthode qui réalise une ordination des espèces et des relevés de façon simultanée, permettant de placer une espèce à la moyenne des relevés qu'elle occupe, et un relevé à la moyenne des espèces qu'il contient (Dray et Chessel 2008). Cette information nous est donnée lors de la présentation superposée des résultats de l'analyse du nuage des sites et du nuage des espèces sur des plans engendrés par des axes de même ordre.

Notons cependant « que la moyenne ne constitue un résumé satisfaisant de la position de l'espèce sur le gradient que si la distribution de cette dernière est uni-modale » (Prodon et Lebreton 1994). L'AFC et la représentation simultanée des sites et des espèces sont donc adaptées au traitement de tableaux « sites × espèces » sur de longs gradients pour lesquels les distributions des espèces sont uni-modales (Prodon et Lebreton 1994; Dray et Chessel 2008). En effet, les espèces montrent communément une courbe en forme de cloche en réponse à un gradient environnemental lorsque celui-ci est très fort (Jongman, Ter Braak, et Van Tongeren 1995).

Dans le cas de distributions uni-modales, « l'AFC est à l'intersection des notions d'optimum écologique et d'espèce indicatrice et elle évite les ordinations a priori en donnant la meilleure ordination réciproque des espèces et des relevés possibles » (Prodon et Lebreton 1994). Pour mettre en évidence les notions d'optimum écologique et d'espèce indicatrice dans les AFC, les auteurs utilisent le graphique canonique. Ce graphique est obtenu par le croisement de l'axe de projection d'ordre un, issu de l'analyse des colonnes, et de l'axe de projection d'ordre un, issu de l'analyse des lignes de la matrice de données.

Dans le deuxième cas cependant, l'AFC pouvant être présentée comme une variante de l'ACP, on retrouve la notion de relation linéaire. En d'autres termes, l'AFC donne la meilleure combinaison linéaire des lignes, qui maximise la séparation des colonnes et inversement (Dray 2014). Il n'est plus question d'une distribution uni-modale sous-jacente. Cependant, dans ces conditions, on ne regarde que les projections des lignes (des sites) ou que celles des colonnes (les espèces) sur le plan de projection. L'information est dans la position relative d'un site (ou espèce) par rapport à l'origine du repère comme dans une ACP.

Pour conclure sur la différence de ces deux approches possibles de l'AFC citons Oksanen 1987 : « using specifically an underlying unimodal response model for species along gradients. This means that we can, in principle, extract more information from the joint plots than the previous linear response model allows : not only the direction and distance in respect to the origin are important, but also the exact position and proximities between points ».

Donnons un exemple de l'utilisation de L'AFC. Cette analyse est utilisée dans l'étude des tableaux de contingence d'espèces d'oiseaux dans les forêts Rhénanes Alsaciennes (Denis 2009). Une AFC est réalisée sur les résultats d'IPA (Indice Ponctuel d'Abondance) desquels on a retiré les oiseaux d'eau au sens strict car leur distribution est trop différente de celle des oiseaux forestiers (ce qui est trivial). Cette analyse permet de conclure que : le principal gradient avifaunistique en forêt rhénane oppose la présence de milieux humides (bras en eau ou roselières) aux milieux plus secs. ; et l'ouverture des peuplements forestiers apparaît comme le second facteur de variabilité. Cet exemple illustre la mise en évidence de deux variables latentes.

Malgré tout, l'AFC peut également présenter des effets indésirables.

Si l'AFC, en utilisant la distance du  $\chi^2$ , n'est pas sensible au problème des doubles zéros, elle peut cependant présenter une disposition des sites en arc de cercle dans le premier plan de projection qui rappelle l'effet fer à cheval (des ACP), mais dont l'origine est différente. On l'observe lorsque les données sont ordonnées sur un seul ou très prédominant gradient. C'est l'effet Guttman. Le deuxième axe, qui est un artéfact, rapproche les points qui sont à l'extrémité du premier axe et oppose les valeurs moyennes aux valeurs extrêmes. Tout se passe comme si l'AFC, recherchant des axes non corrélés linéairement entre eux, fournit des axes orthogonaux au premier en pliant celui-ci en deux pour avoir le deuxième, en trois pour le troisième etc... Dans le cas d'un effet Guttman avéré, il est donc conseillé de ne présenter que l'ordination sur le premier axe.

Cependant, toutes les analyses dont les graphes prennent cette forme parabolique ne sont pas victimes d'un effet Guttman. Aussi, un deuxième gradient peut également exister et ne pas être mis en évidence par le deuxième axe et être relégué sur des axes ultérieurs (Prodon et Lebreton 1994). C'est le cas si ce deuxième gradient porte moins d'inertie projetée que les axes artéfactuels. Il est possible, pour tenter de faire ressortir un deuxième gradient ayant un sens écologique, de réitérer l'analyse sur les résidus des régressions linéaires des abondances des espèces sur la variable prépondérante du premier gradient (Jongman, Ter Braak, et Van Tongeren 1995).

L'AFC est également sensible aux espèces rares si celles-ci sont présentes sur des sites pauvres en espèces, et place alors ces sites et espèces à l'extrémité du premier axe (Gauch, 1982 dans Ter Braak, 1987 ; Jongman, Ter Braak, et Van Tongeren 1995 ; Zuur, Ieno, et Smith 2007). Enfin l'AFC présente une tendance à donner une image comprimée des distances entre les points qui sont aux extrémités des gradients, par rapport à ceux en position centrale (Ter Braak 1985).

### 1.3 Une question légitime

Les analyses AFC et ACP sont basées sur des mesures de dissimilarités, la distance du  $\chi^2$  pour les AFC et la distance euclidienne pour les ACP, donc sur des comparaisons. J'ai montré dans la partie 1.1.3 que la non prise en compte de la variabilité de la PdD compromettrait les résultats des comparaisons. Il est donc légitime de poser la question de l'impact que peut avoir une PdD inférieure à 1 lors de l'utilisation de ces techniques ACP et AFC pour analyser des tableaux de relevés d'abondances ou de présences-absences de plusieurs espèces sur plusieurs sites.

Or, je n'ai constaté que deux types de démarches vis-à-vis de la PdD, lors de l'utilisation de l'AFC ou de l'ACP. Soit la question n'est tout simplement pas abordée, soit les auteurs suppriment, avant analyse, les données des espèces suspectées d'une PdD « très » faible, sans autres considérations.

Pour le premier cas on peut citer par exemple l'étude sur la résilience à long terme des communautés végétales herbacées de pelouse calcicole (Forey 2003). Dans cette étude, les relevés en présence-absence de la végétation sont faits sur une friche et un témoin sur trois sites différents. Une AFC est effectuée sur le tableau des résultats, qui croise 60 relevés et 64 espèces. Cette AFC met en évidence un gradient de fermeture du milieu expliquant la distribution des espèces.

C'est le cas également dans l'étude expliquant les possibles utilisations de l'ACP sur les échantillons d'invertébrés benthiques en eaux douce (Dupiereux, Feytmans, et Micha 1983). Dans

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

cet exemple, l'ACP est appliquée sur un tableau croisant les abondances de 51 taxons dans 60 échantillons (soit 3 bourriches de substrats artificiels dans 20 stations).

Le fait que la PdD ne soit pas évoquée conduit aux hypothèses suivantes : soit la PdD est supposée égale à un pour toutes les espèces et sur tous les points d'échantillonnage, hypothèse forte qui demande une justification. Soit on considère que les résultats des analyses AFC ou ACP ne sont pas impactés ni par une PdD inférieure à un, ni par une potentielle variabilité spatio-temporelle de la PdD.

Le deuxième cas est rencontré dans une étude de l'élévation altitudinale de la zone de reproduction des oiseaux, dans une vallée alpine italienne, à partir de données de présence-absence tirées de deux atlas effectués à 11 ans d'intervalle (Popy, Bordignon, et Prodon 2009). Dans cette étude, les auteurs utilisent une AFC. Ils conservent les espèces dont ils supposent que les distributions issues des deux atlas correspondent bien à la distribution des espèces lors de la période de reproduction. Néanmoins, ils excluent le tichodrome échelette (*Tichodroma muraria*) des données avant l'AFC car ils estiment que cette espèce a « une PdD particulièrement faible ». Les auteurs évoquent par ailleurs que la PdD de la présence des espèces d'oiseaux peut varier en fonction de la fermeture du milieu. On en déduit qu'ils font l'hypothèse que seules les espèces ayant une très faible PdD sont susceptibles d'influencer de manière significative les résultats de l'AFC. Cependant quelle valeur limite donner à une « très faible PdD » ?

Ainsi, à ma connaissance, et aussi surprenant que cela puisse paraître au vu de l'immense littérature sur la PdD en écologie, aucune étude sur les potentiels biais que la PdD peut provoquer dans les analyses AFC et ACP n'a été faite à ce jour. Il me semble essentiel de répondre à cette question.

Dans la suite de mon travail, je me limiterai aux tableaux d'abondances analysés par ACP ou AFC. Je décomposerai la question des conséquences de la PdD dans ces analyses comme suit :

- Les sous-estimations des abondances, induites par une PdD inférieure à un, sans aucune variabilité (ni entre espèces, ni entre sites), provoquent-elles une altération des résultats des analyses ?
- Si oui, une variabilité dans la PdD peut-elle amplifier l'altération, voire en changer même le « sens » ?
- Si la seule baisse uniforme de la PdD ne provoque pas une altération, sa variabilité peut-elle être l'origine d'une altération dans les résultats des analyses ?

Je me limiterai à des cas où la variabilité de la PdD entre les sites correspond à un gradient.

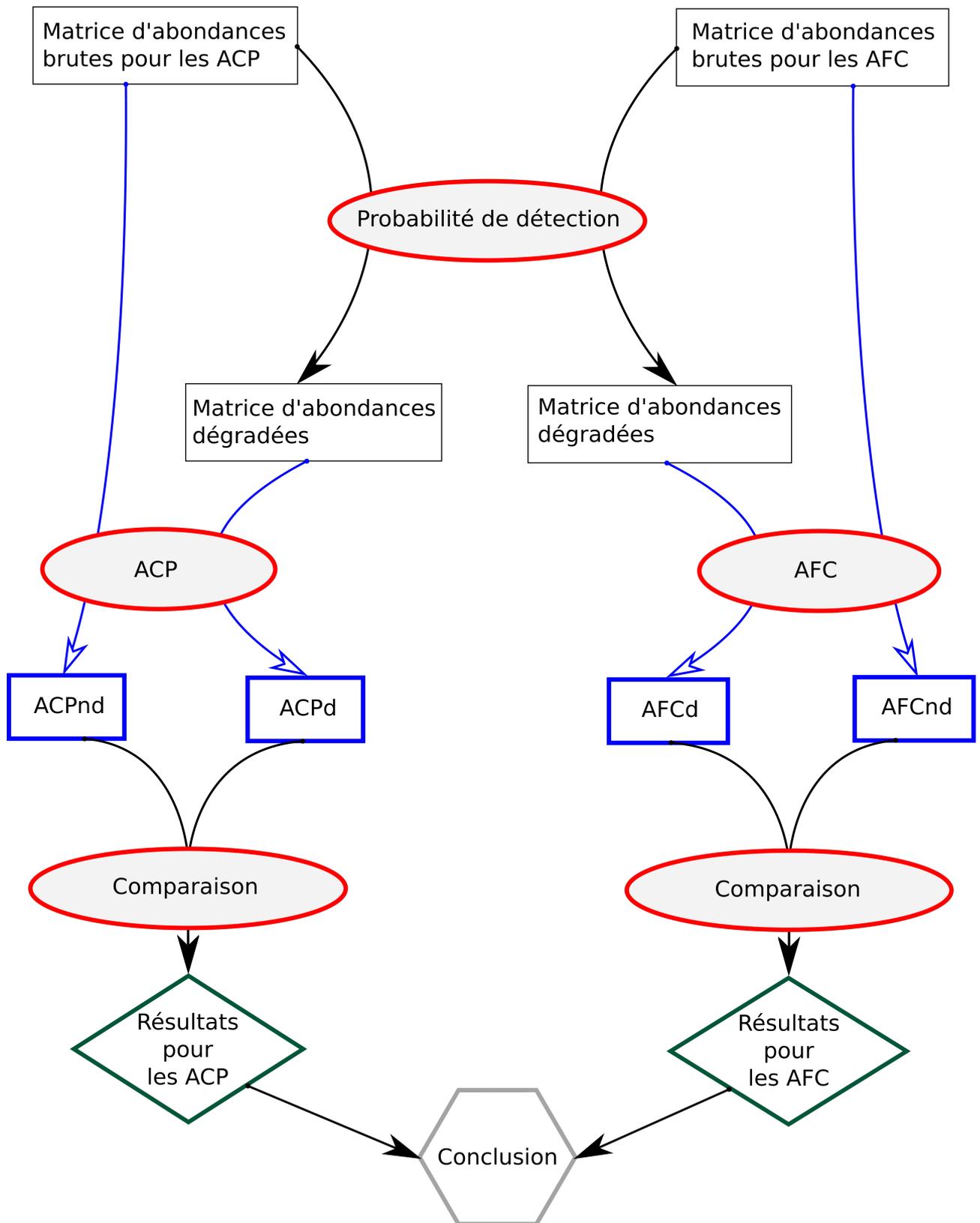


Figure 1 : Schéma de la démarche employée pour déterminer si la probabilité de détection (PdD) a des conséquences sur les résultats des ACP et AFC. Le principe est de comparer les résultats de l'analyse issus d'une matrice d'abondance brute (où la PdD est de 1 pour toutes les espèces) aux résultats issus d'une matrice dégradée, elle-même déduite de la matrice brute, mais où la PdD des espèces est différente de 1.

## Chapitre 2 : Matériels et méthode

### 2.1 Des données issues de simulations

Pour pouvoir répondre aux questions présentées à la fin du chapitre précédent, il me faudrait connaître l'abondance réelle de plusieurs espèces sur plusieurs sites et également la PdD de ces différentes espèces sur ces sites. La question portant sur les effets ou non de la variabilité de la PdD sur les résultats des analyses, je devrais alors être en mesure de comparer les conséquences de différentes formes de variabilité sur ceux-ci.

La maîtrise des facteurs conditionnant les valeurs d'abondances et les valeurs de la PdD est donc un pré-requis que seules des données simulées peuvent m'apporter. Les conditions d'emplois des méthodes ACP et AFC étant différentes, les données devront également remplir ces conditions. Je tenterai donc de traiter les questions posées à l'aide de données créées avec le logiciel R version 3.2.4 (R Core team 2016). Pour ce faire, je me suis appuyé sur les bibliothèques: `ade4` (Chessel, Dufour et Thioulouse 2004), `stats` (R Core team 2016), `grDevices` (R Core team 2016) et `scatterplot3d` (Ligges et Mächler 2003).

### 2.2 Matériels et Méthode

La démarche générale utilisée pour déterminer si la PdD a des conséquences sur les résultats des ACP et AFC, schématisée ci-contre par la figure 1, est la suivante :

je simule, sur plusieurs sites, les abondances de 20 espèces différentes ; 20 espèces étant un compromis entre d'une part un ordre de grandeur cohérent avec celui de tableaux d'abondances que l'on peut trouver dans certaines études (une tableau floristique de 60 placettes et de 30 espèces pour Forey (2003), 20 stations d'échantillonnage et 51 espèces pour Duperrioux (1983)), et d'autre part la facilité d'interprétation des résultats. Ces abondances sont issues de la loi de Poisson (cf. Partie 2.2.2), communément utilisée pour modéliser la distribution des abondances des organismes (Royle 2004). De là, je forme le tableau des abondances par site et par espèce (les sites en ligne et les espèces en colonne). Ces abondances sont celles qui seraient mesurées en cas de PdD égale à un. C'est-à-dire qu'il n'y a aucune sous-estimation des abondances. Ce tableau est nommé « matrice d'abondances brutes », avec  $n_{ij}$  l'abondance de l'espèce  $j$  sur le site  $i$ .

J'applique une AFC ou une ACP sur cette matrice d'abondances brutes. J'appelle les résultats de l'ACP ou de l'AFC sur cette matrice respectivement « ACPnd » et « AFCnd » (« nd » pour « non dégradée »).

Dans un deuxième temps, je déduis de la matrice d'abondances brutes un autre tableau d'abondances, appelé « matrice d'abondances dégradées ». Cette matrice d'abondances dégradées est une des matrices des abondances qui serait « possiblement » obtenue dans le cas de PdD inférieures à un. J'utilise le terme « possiblement » car les abondances de cette matrice sont obtenues à l'aide de lois binomiales, donc avec une part aléatoire analogue à celle de la PdD. Je me placerai toujours dans le cas où tous les individus d'une même espèce et sur un même site ont la même PdD. Cette dernière ne dépendant que de l'espèce et du site sera notée  $d_{ij}$ . Je nomme  $n_{ij}^D$  l'abondance de l'espèce  $j$  sur le site  $i$  dans la matrice d'abondances dégradées. Chaque valeur  $n_{ij}^D$  est obtenue à partir de la loi binomiale ayant pour caractéristiques un nombre de tirages égal à l'abondance  $n_{ij}$  et une probabilité  $d_{ij}$ . Je détaillerai dans la partie 2.2.3 la méthode d'obtention des valeurs de la PdD.

Une ACP ou une AFC est alors appliquée sur la matrice d'abondances dégradées. J'appelle les résultats de l'ACP ou de l'AFC sur cette matrice respectivement « ACPd » et « AFCd » (« d » pour « dégradée »).

Enfin, je compare les ACPd aux ACPnd et les AFCd aux AFCnd. Pour certaines comparaisons j'ai besoin d'outils statistiques. Je présenterai ces comparaisons et outils statistiques qui m'informent sur les effets de la PdD sur les analyses AFC ou ACP dans la partie 2.2.4.

Étant donnée la part aléatoire dans l'obtention des matrices d'abondances dégradées, mais également dans celle des matrices d'abondances brutes du fait de l'utilisation de lois de Poisson, il me faut, pour mettre en évidence une tendance et non un cas particulier, réitérer plusieurs fois une simulation avec des paramètres de même valeur. Je choisis un nombre de répétitions égal à 1000. Pour des raisons de simplicité, on appelle SIMULATION un ensemble de 1000 simulations dont les paramètres des populations (abondance attendue sur chaque site pour chaque espèce) et des PdD sont les mêmes. Ces paramètres sont décrits dans la suite du document. Chaque SIMULATION est numérotée et ces numéros seront utilisés lors de la présentation des résultats. On retrouve en annexe A, les 7 tableaux listant les SIMULATIONs et les valeurs des paramètres pour chacune d'entre elles (tableau de A.5 à A.11).

Les conditions d'emplois des méthodes ACP et AFC étant différentes, j'insisterai au sein de chacune des parties suivantes sur les précautions que j'ai prises pour les respecter.

### 2.2.1 Définition des gradients et des sites

Les sites où sont mesurées les abondances sont définis à l'aide de deux variables indépendantes. Ces deux variables forment deux gradients hypothétiques d'habitat (par exemple l'altitude ou l'exposition au soleil),  $g_1$  et  $g_2$ , sur lesquels les sites sont répartis. Pour les valeurs de  $g_1$  je me suis inspiré d'un gradient altitudinal et pour les valeurs de  $g_2$  d'un gradient d'orientation mesuré en degrés, mais sans tenir compte de l'aspect circulaire des degrés, c'est-à-dire qu'ici 0 et 360 ne sont pas égaux mais se placent chacun à une extrémité du gradient (hypothétique). Les valeurs prises par la variable  $g_1$  sur l'ensemble des sites sont comprises dans l'intervalle [0, 1000] pour les simulations qui concernent les AFC, ce qui permet de simuler un gradient long, et dans l'intervalle [500, 1000] pour les simulations destinées aux ACP, afin de simuler un gradient court. Les valeurs prises par la variable  $g_2$  sur l'ensemble des sites sont comprises dans l'intervalle [0, 360] (gradient long) pour les simulations qui concernent les AFC et dans l'intervalle [90, 270] (gradient court) pour les simulations destinées aux ACP. Le nombre de sites et leur répartition sur ces deux gradients sont identiques entre toutes les simulations qui concernent un même type d'analyse (AFC ou ACP), mais différent entre les simulations faites pour les ACP et celles faites pour les AFC. Pour les ACP j'ai simulé les données sur 50 sites notés de 1 à 50. Il importe en effet pour les ACP d'avoir bien plus de sites que d'espèces (Prodon et Jacquet 2014). C'est pourquoi j'ai mis un rapport de 2.5 entre le nombre de sites et le nombre d'espèces. Pour les AFC, ayant doublé la portion de gradient que couvrent les sites par rapport au ACP, et pour conserver une même pression d'échantillonnage, j'ai donc simulé les données sur 100 sites notés de 1 à 100. Les valeurs de  $g_1$  et  $g_2$ , fixées pour chacun des sites sont présentées en annexe A dans les tableaux A.2 et A.4. Dans les deux cas, avec ces valeurs, on a une simulation d'un échantillonnage régulier, comme le montre, la figure 2. Par commodité, on note les valeurs des variables  $g_1$  et  $g_2$  sur le site  $i$ , respectivement  $g_{i1}$  et  $g_{i2}$ .

### 2.2.2 Simulation des abondances brutes (non dégradées)

En préalable à la simulation des abondances brutes des 20 espèces sur chacun des sites, que l'on numérote de 1 à 20, je commence par fixer les valeurs, sur les gradients  $g_1$  et  $g_2$  d'un unique optimum pour chacune des espèces. Ce point simule l'optimum écologique de l'espèce par rapport aux deux variables  $g_1$  et  $g_2$ . Je note les valeurs des variables  $g_1$  et  $g_2$  à l'optimum de l'espèce  $j$ , respectivement  $g_{1j}$  et  $g_{2j}$ .

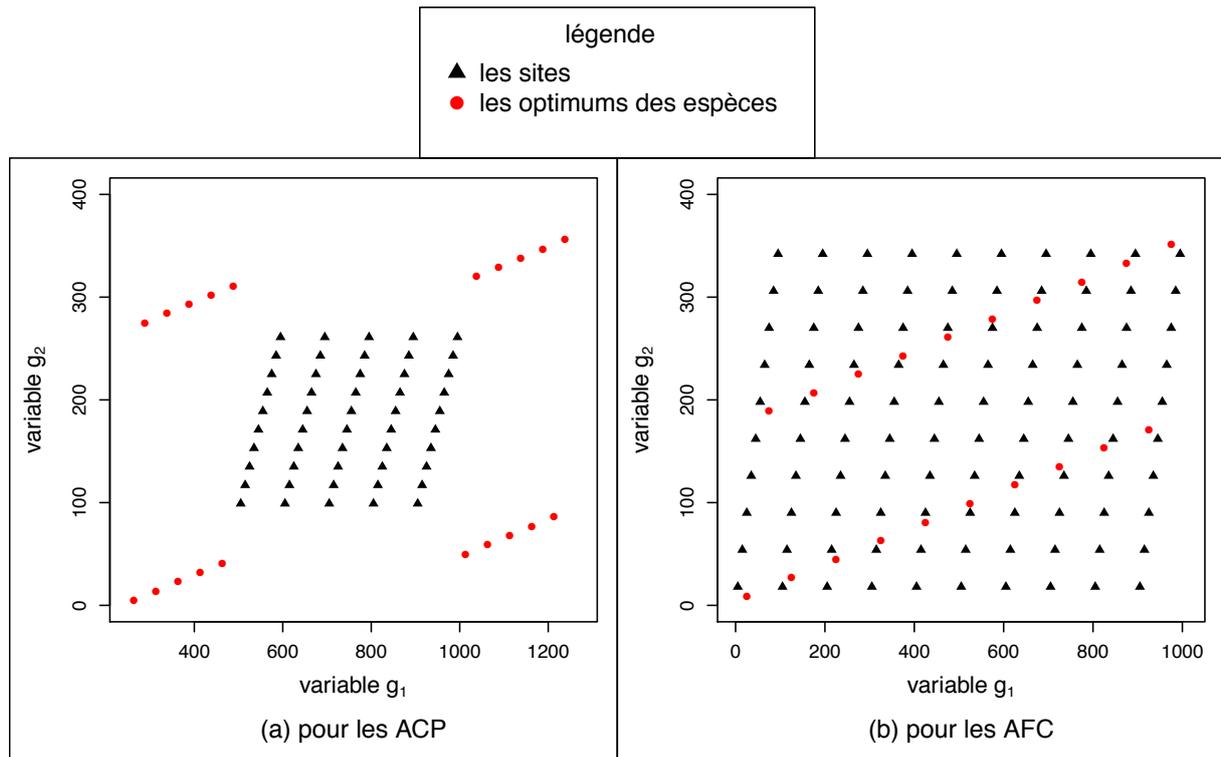


Figure 2 : coordonnées des sites et des optimums des 20 espèces en fonction des variables d'habitat  $g_1$  et  $g_2$ , (a) dans le cas des simulations des ACP, (b) dans le cas des simulations des AFC. Chaque site  $i$  (triangle noir) est caractérisé par le couple de valeurs  $(g_{i1}, g_{i2})$  ; et l'optimum d'une espèce  $j$  (cercle rouge) est donnée par le couple de valeurs  $(g_{1j}, g_{2j})$ . J'ai ainsi une image de la proximité entre sites et optimum des espèces.

La figure 2 (a) montre que, pour les populations dont les données d'abondances sont soumises à une ACP, les valeurs des optimums sont choisies de manière à ce que l'optimum de chaque espèce soit à l'extérieur de la portion de gradient que couvrent les sites sur les gradients  $g_1$  et  $g_2$ . Les optimums des 10 premières espèces ont des valeurs  $g_{1j}$  inférieures aux valeurs  $g_{i1}$  des sites, et les optimums des 10 dernières espèces ont des valeurs de  $g_{1j}$  supérieures aux valeurs  $g_{i1}$ . De plus les optimums des 10 espèces impaires ont des valeurs  $g_{2j}$  inférieures aux valeurs  $g_{i2}$  des sites, et les optimums des 10 espèces paires des valeurs de  $g_{2j}$  supérieures aux valeurs  $g_{i2}$ . Ainsi, sur cet intervalle de gradient, toutes les abondances moyennes des populations seront juste croissantes ou décroissantes (si elles ne sont pas constantes), ce pour quoi l'ACP est adaptée (cf. chapitre 1). Précisément les abondances des 10 premières espèces et des 10 dernières sont, respectivement, décroissantes et croissantes en fonction de  $g_1$  sur la portion de gradient que couvrent les sites. Et les abondances des 10 espèces impaires et des 10 paires sont, respectivement, décroissantes et croissantes en fonction de  $g_2$  sur la portion de gradient que couvrent les sites.

La figure 2 (b) montre que pour les populations dont les données d'abondances sont soumises à une AFC, les optimums des espèces sont placés dans la portion de gradient que couvrent les sites sur les gradients  $g_1$  et  $g_2$ . Ainsi je peux simuler une réponse unimodale de l'abondance des espèces le long des gradients. De plus, les 10 premières espèces et les 10 dernières ont respectivement des valeurs  $g_{1j}$  inférieures et supérieures à 500 (centre de la portion de gradient que couvrent les sites sur le gradient  $g_1$ ) ; et les espèces paires et impaires ont respectivement des valeurs  $g_{2j}$  inférieures et supérieures à 180 (centre de la portion de gradient que couvrent les sites sur le gradients  $g_2$ ).

Les valeurs des coordonnées des optimums des 20 espèces dans les cas des ACP et des AFC sont respectivement données dans les tableaux A.1 et A.3 de l'annexe A.

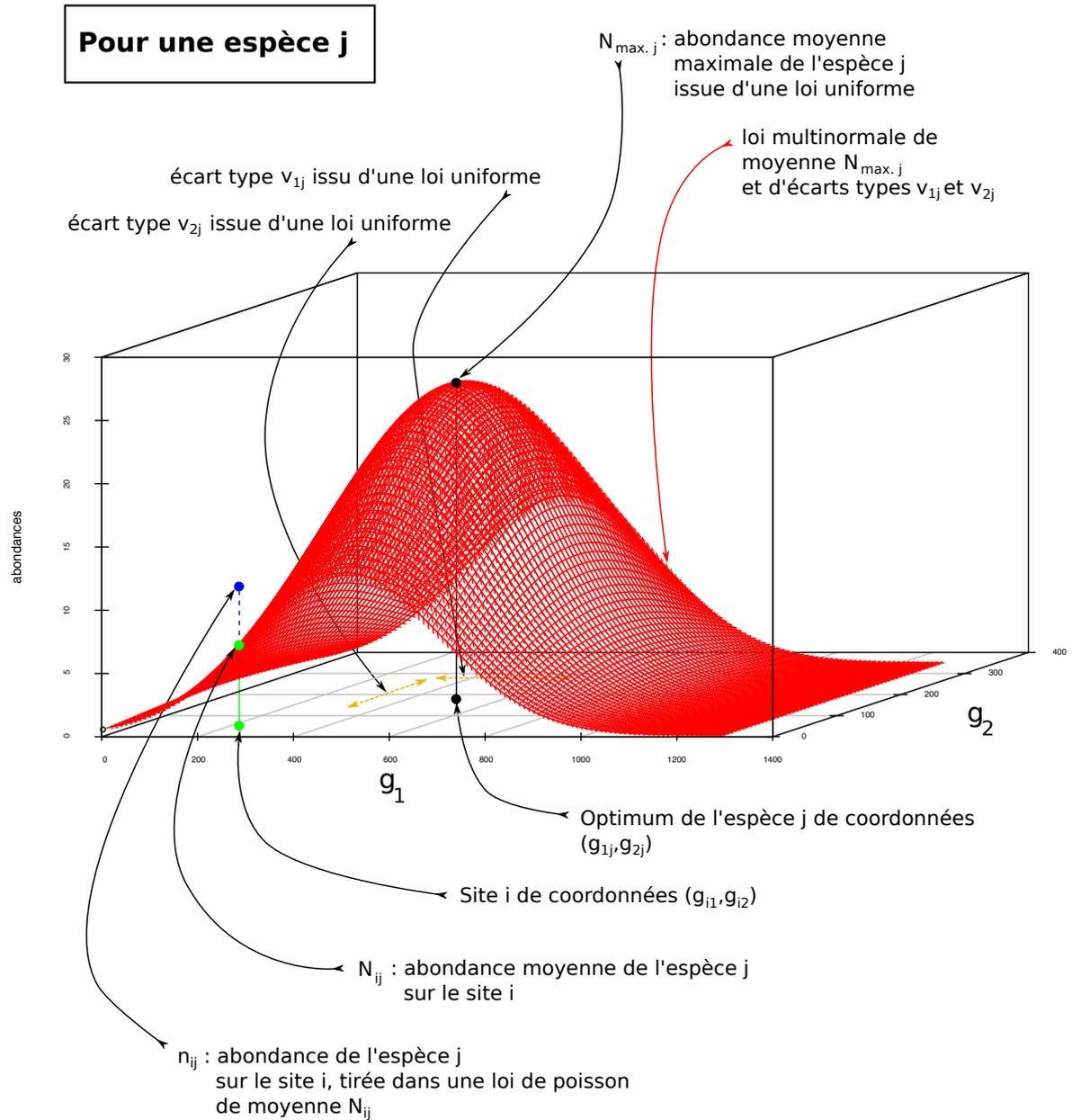


Figure 3 : schéma illustrant la simulation de l'abondance d'une espèce j sur un site i.

Une fois ces paramètres fixés pour les sites et les espèces, je peux simuler les abondances des espèces sur chacun des sites. La figure 3 page précédente résume l'ensemble des critères pris en compte pour déterminer l'abondance d'une espèce sur un site, et rappelle les appellations des paramètres.

Je commence par déterminer pour chaque espèce son abondance moyenne à son optimum. Ce sera donc l'abondance moyenne la plus élevée de l'espèce. À chaque simulation, ces 20 valeurs (une pour chaque espèce), que je note  $N_{\max.j}$  pour l'espèce  $j$ , sont tirées de manière indépendante et aléatoire dans une même loi uniforme définie par ses deux valeurs extrêmes  $N_{\min} = 5$  et  $N_{\max} = 50$  (valeurs fixes pour toutes les simulations et communes à toutes les espèces). Ainsi, à chaque simulation, la communauté d'espèces simulée contient un nombre aléatoire d'espèces rares, communes et abondantes.

Puis, à chaque espèce  $j$  et à chaque simulation, j'associe une loi multi-normale définie sur les variables  $g_1$  et  $g_2$  et ayant son mode aux valeurs  $(g_{1j}, g_{2j})$  de l'optimum de l'espèce. Les deux écarts types de la loi, que l'on note  $v_{1j}$  et  $v_{2j}$ , respectivement sur la variable  $g_1$  et sur la variable  $g_2$ , sont tirés à chaque simulation aléatoirement, chacun dans une loi uniforme. Ces deux lois uniformes sont définies par leurs valeurs minimum et maximum respectives :  $v_{1\min}, v_{1\max}$ , et  $v_{2\min}, v_{2\max}$ . Ces quatre dernières valeurs sont communes à toutes les espèces et fixes pour une SIMULATION. On peut alors calculer, à chaque simulation, la valeur moyenne d'abondance de l'espèce  $j$  sur le site  $i$  que l'on note  $N_{ij}$  à l'aide de la formule suivante :

$$N_{ij} = N_{\max.j} * e^{-\frac{1}{2} \left( \left( \frac{g_{i1} - g_{1j}}{v_{1j}} \right)^2 + \left( \frac{g_{i2} - g_{2j}}{v_{2j}} \right)^2 \right)}$$

Cette formule permet d'obtenir les valeurs moyennes d'abondance pour chacune des 20 espèces, variant entre les sites en fonction de deux variables caractéristiques des sites, avec une forme de loi multi-normale.

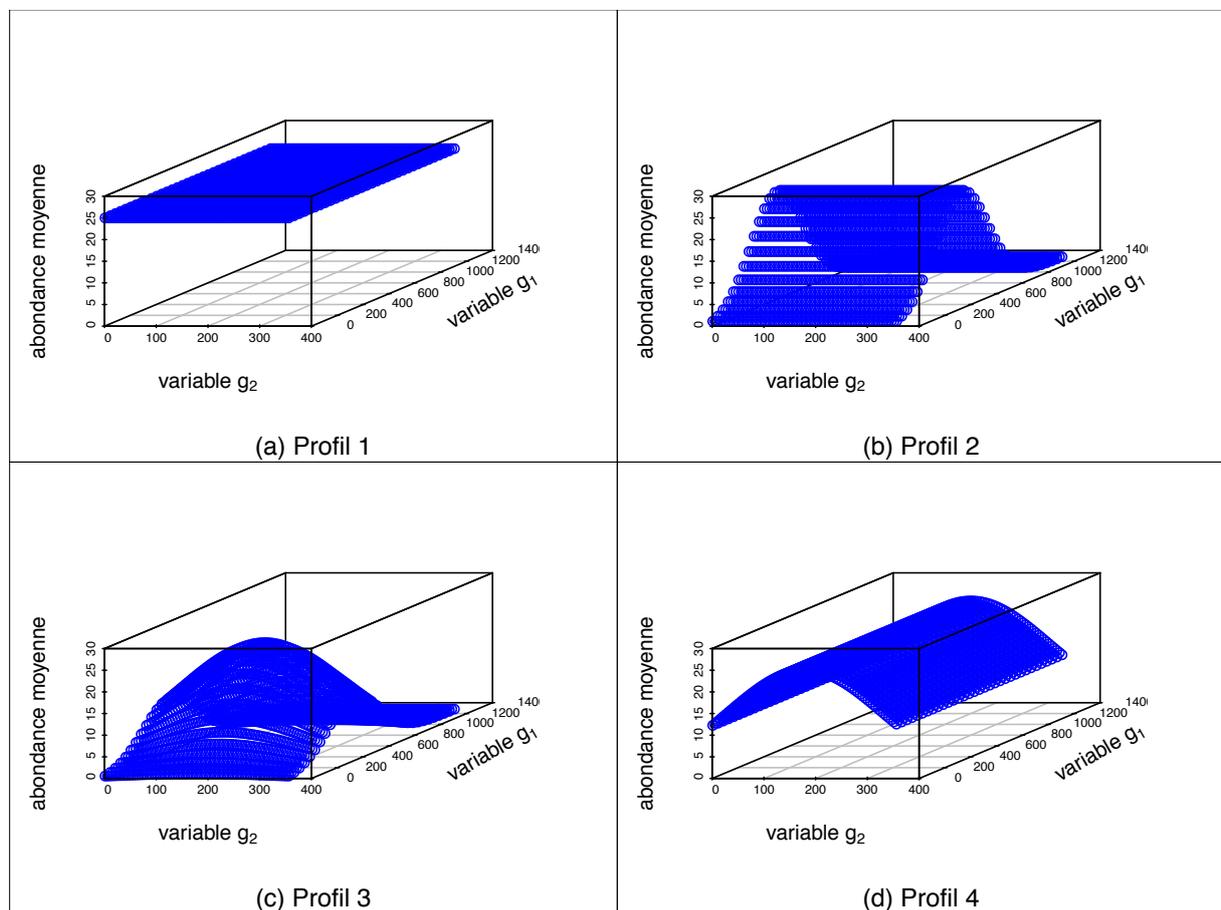
À partir des quatre paramètres  $v_{1\min}, v_{1\max}, v_{2\min}$  et  $v_{2\max}$ , on définit quatre types de profils de moyenne d'abondance vis à vis des variables d'habitat  $g_1$  et  $g_2$ . Lors d'une simulation on impose à toutes les espèces le même type de profils. Ces profils, qu'illustre la figure 4 page suivante, sont :

**Profil 1 :** Pour chaque espèce, la moyenne d'abondance ne varie pas ni en fonction de  $g_1$  ni de  $g_2$ . Pour obtenir ce profil on fixe, aussi bien pour les ACP que pour les AFC,  $v_{1\min} = v_{1\max} = 100000$  et  $v_{2\min} = v_{2\max} = 50000$ . Ces SIMULATIONs sont numérotées de 1001 à 1309.

**Profil 2 :** Toutes les espèces ont une moyenne d'abondance qui varie uniquement en fonction de  $g_1$ . L'indépendance des abondances vis-à-vis de  $g_2$  est obtenue en posant  $v_{2\min} = v_{2\max} = 50000$ . Les valeurs de  $v_{1\min}$  et de  $v_{1\max}$  doivent permettre d'avoir à chaque simulation, et pour me rapprocher de la réalité de terrain, des espèces avec des valences écologiques allant de larges (les espèces généralistes) à faibles (les espèces spécialistes). Cependant, il me faut tenir compte de trois éléments : (1) les optimums des espèces sont à l'extérieur de la portion de gradient que couvrent les sites pour les ACP et à l'intérieur pour les AFC ; (2) les portions de gradients couvertes entre ACP et AFC sont différentes ; (3) Enfin et surtout, la PdD lors de la création de la matrice d'abondances dégradées ne provoque pas, pour aucune espèce et quelle que soit la valeur de PdD simulée, des abondances nulles sur tous les sites. Toutes ces considérations m'amènent à poser les valeurs suivantes : pour les ACP  $v_{1\min} = 250$  et  $v_{1\max} = 1000$ , et pour les AFC  $v_{1\min} = 200$  et  $v_{1\max} = 1000$ . Ces SIMULATIONs sont numérotées de 2001 à 2309.

**Profil 3 :** Toutes les espèces ont une moyenne d'abondance qui varie en fonction de  $g_1$  et  $g_2$ . Je garde la démarche qui me permet d'obtenir aussi bien des espèces généralistes que spécialistes, vis à vis des deux variables  $g_1$  et  $g_2$ . Les mêmes considérations sur les différences des simulations pour les ACP et les AFC (positions des optimums des espèces, portions de gradients couvertes par les sites) et la nécessité de ne pas avoir de colonne de matrices d'abondances dégradées remplie uniquement de zéros, m'amènent à poser les valeurs suivantes : pour les ACP  $v_{1min} = 250$ ,  $v_{1max} = 1000$ ,  $v_{2min} = 200$  et  $v_{2max} = 400$ , et pour les AFC  $v_{1min} = 200$ ,  $v_{1max} = 1000$ ,  $v_{2min} = 150$  et  $v_{2max} = 400$ . Ces SIMULATIONS sont numérotées de 3001 à 3309.

**Profil 4 :** Toutes les espèces ont une moyenne d'abondance qui varie uniquement en fonction  $g_2$ . L'indépendance des abondances vis à vis de  $g_1$  est obtenue en posant  $v_{1min} = v_{1max} = 100000$  (aussi bien pour les ACP que pour les AFC). Toujours avec les mêmes considérations énoncées pour établir les valeurs des paramètres des profils 2 et 3, je donne les valeurs suivantes: pour les ACP  $v_{2min} = 150$  et  $v_{2max} = 400$ , et pour les AFC  $v_{2min} = 100$  et  $v_{2max} = 400$ . Ces SIMULATIONS sont numérotées de 4001 à 4309



*Figure 4 : Figure en trois dimensions des quatre types de profils d'abondance moyenne des espèces, utilisés pour les SIMULATIONS. Les variables  $g_1$  et  $g_2$  sont des variables d'habitat hypothétiques qui caractérisent les sites où sont mesurées les abondances. Chaque site est défini par une paire de valeurs  $(g_1, g_2)$ . Un rond bleu donne la valeur d'abondance moyenne de l'espèce pour la paire de valeurs  $(g_1, g_2)$  correspondante en abscisse.*

Les SIMULATIONS dans le cas des ACP

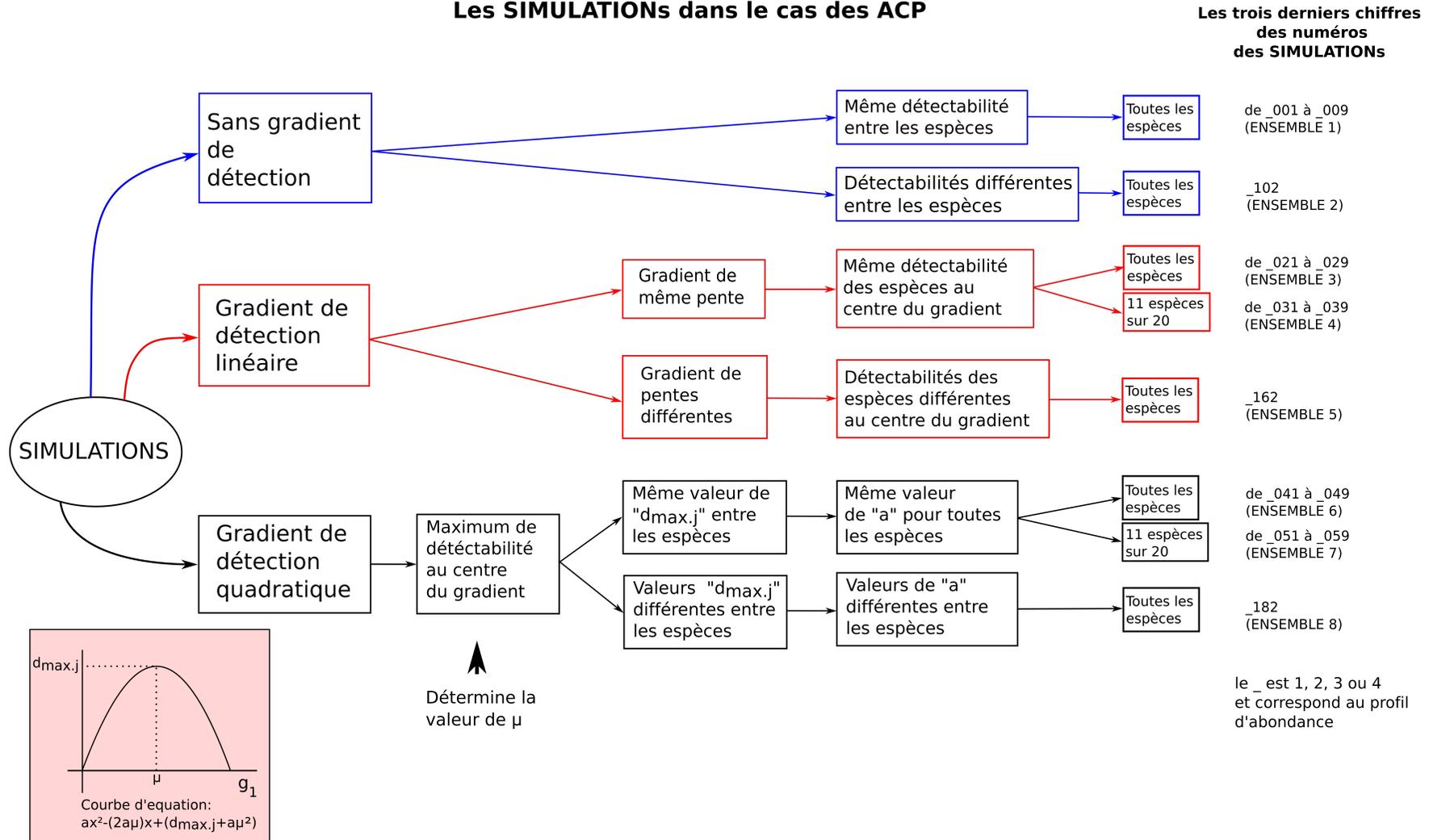


Figure 5 : Répartitions et numéros des SIMULATIONS en fonction des différentes hypothèses faites sur la PdD des espèces dans le cas des ACP. Je décline les trois formes de gradients (sans gradient, linéaire et quadratique) en fonction des paramètres qui leur sont propres. Les trois chiffres donnés dans la colonne de droite sont les chiffres terminant les numéros des SIMULATIONS. Les SIMULATIONS sont regroupées au sein d'ENSEMBLES (également numérotés) utilisés lors de la présentation des résultats. Le chiffre 1, 2, 3 ou 4 qui commence tout numéro de SIMULATIONS correspond aux quatre profils d'abondance des espèces utilisés.

### 2.2.3 Simulation des PdD et des abondances dégradées

Pour obtenir la matrice d'abondances dégradées à partir de la loi binomiale, il reste à définir la valeur  $d_{ij}$  de la PdD des individus de l'espèce  $j$  sur le site  $i$ . Je ne peux pas essayer toutes les valeurs et les variabilités possibles de la PdD. Je procède donc au choix des différentes formes de variabilité de la PdD que j'étudie. Pour cela je m'appuie sur les formes de variabilités de la PdD évoquées dans des articles tels que Royle (2004) ou Yamaura et al. (2011). Ces choix, détaillés ci-après, sont présentés pour les ACP par la figure 5 page précédente, et pour les AFC par la figure 6 page suivante.

**Sans gradient de détection** : La première hypothèse est que la PdD des individus ne varie pas entre les sites. Autrement dit, la PdD n'est pas fonction des variables  $g_1$  et  $g_2$  : il n'y a pas de gradient de détection.

Dans cette hypothèse, et afin de simuler des variabilités de PdD plus ou moins complexes entre les espèces, j'étudierai deux cas :

(i) Dans le premier cas la PdD est identique entre toutes les espèces. Afin de connaître les effets dus à la valeur de la PdD, une SIMULATION est faite pour chacune des valeurs de PdD suivantes : 0.1, 0.3, 0.5, 0.7 et 0.9.

(ii) Dans le deuxième cas, les PdD sont différentes pour chaque espèce. Chaque espèce voit sa PdD tirée à chaque simulation dans une loi normale, dont on fixe la valeur moyenne ( $d_{moy} = 0.5$ ) et l'écart type ( $v_{detect} = 0.2$ ). Toutes les valeurs inférieures à 0.1 sont ramenées à 0.1, et toutes les valeurs supérieures à 0.9 sont ramenées à 0.9. Ainsi, bien qu'hétérogènes, les valeurs de PdD ne sont pas toutes concentrées sur les valeurs extrêmes (0.1 et 0.9).

**Avec un gradient de détection linéaire** : j'impose ensuite le fait que la PdD dépende de la variable  $g_1$  (et non de  $g_2$ ) avec la relation suivante :

$$\text{logit}(d_{ij}) = \alpha_j + \beta_j g_{i1}$$

Cette forme de gradient de détection illustre, par exemple, l'hypothèse que plus on monte en altitude, plus le milieu est ouvert et plus la PdD des oiseaux augmente (Popy, Bordignon, et Prodon 2009). Ce gradient est déterminé par une pente  $\beta_j$ , et une ordonnée à l'origine  $\alpha_j$ , qui sont propres à chaque espèce, et calculées à chaque simulation. En ce sens, la PdD peut aussi varier entre espèces.

Les pentes  $\beta_j$  sont toutes positives. Elles sont tirées aléatoirement à chaque simulation dans une loi uniforme dont les valeurs extrêmes ( $\beta_{min}$  et  $\beta_{max}$ ) sont fixes pour une SIMULATION. On calcule  $\alpha_j$  de manière à ce que la valeur de la PdD de l'espèce soit égale à une valeur prédéterminée ( $d_{moy,j}$ ) au milieu du gradient  $g_1$  couvert par les sites ( $g_1$  égale 750 pour les ACP et 500 pour les AFC). Les valeurs  $d_{moy,j}$ , sont tirées à chaque simulation dans une loi normale, dont on fixe la valeur moyenne ( $d_{moy}$ ) et l'écart type ( $v_{detect}$ ). Moyenne et écart type de cette loi normale sont fixes pour une SIMULATION. Les valeurs sortant de l'intervalle [0.1, 0.9] sont, comme précédemment, modifiées pour appartenir à cet intervalle.

Les **SIMULATIONS** dans le cas des AFC

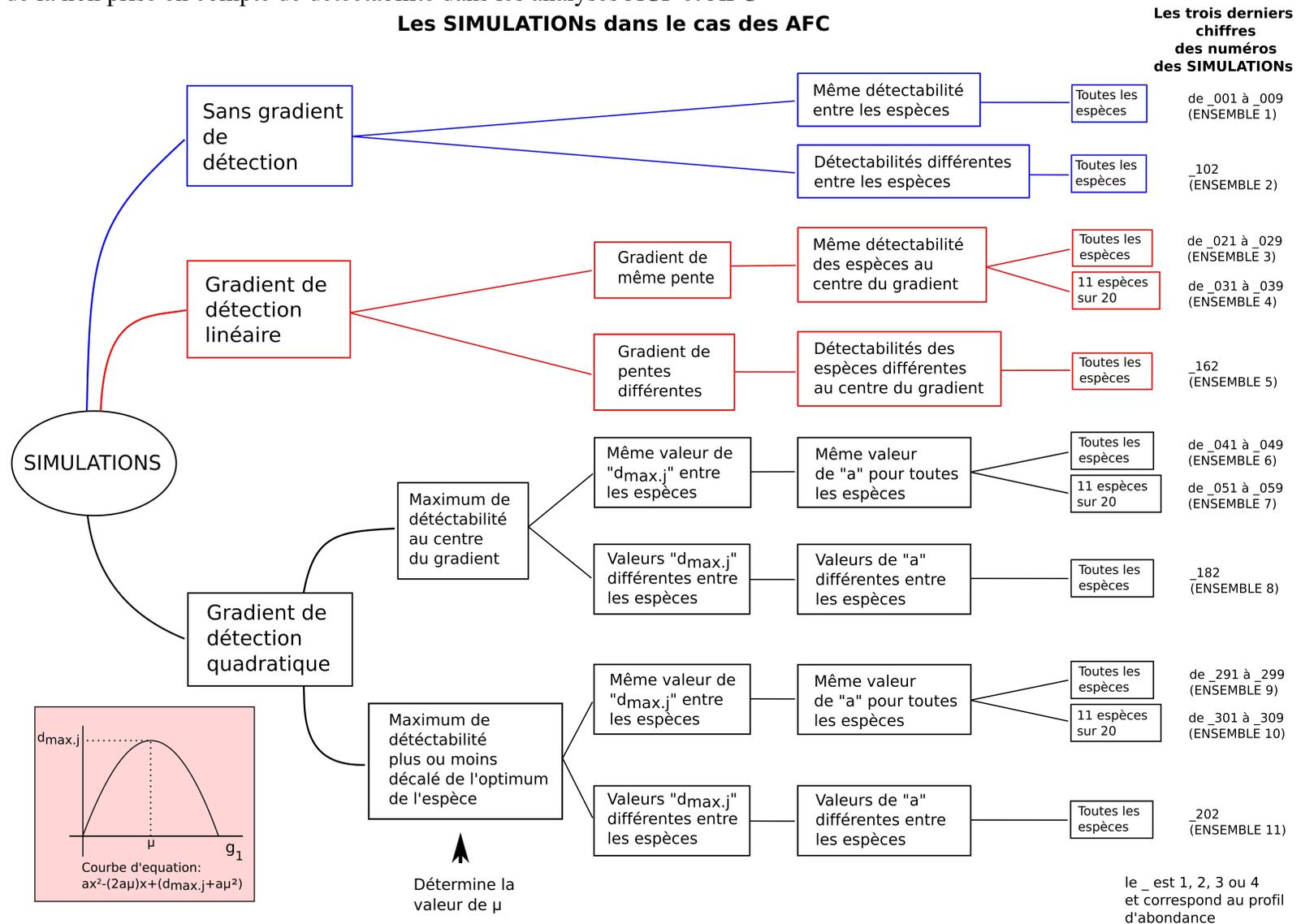


Figure 6 : Répartitions et numéros des SIMULATIONS en fonction des différentes hypothèses faites sur la PdD des espèces dans le cas des AFC. Je décline les quatre formes de gradients (sans gradient, linéaire, quadratique « centré » et quadratique « décalé ») en fonction des paramètres qui leur sont propres. Les trois chiffres donnés dans la colonne de droite sont les chiffres terminant les numéros des SIMULATIONS. Les SIMULATIONS sont regroupées au sein d'ENSEMBLES (également numérotés) utilisés lors de la présentation des résultats. Le chiffre 1, 2, 3 ou 4 qui commence tout numéro de SIMULATIONS correspond aux quatre profils d'abondance des espèces utilisées.

Dès lors, afin de simuler des variabilités plus ou moins complexes, j'étudierai les deux hypothèses suivantes :

(i) une PdD identique entre toutes les espèces avec  $v_{detect} = 0$  et  $\beta_{min} = \beta_{max}$ . Dans cette hypothèse, afin de se placer sur des gradients moyens, j'impose une pente suffisante (pour que le gradient se constate après dégradation des abondances par la PdD), mais ne faisant pas passer les PdD des espèces de 0 à 1 sur la portion de gradient couverte par les sites. Les mêmes considérations que lors des simulations des profils d'abondance sont à prendre en compte pour établir les valeurs des pentes. Je ne donne donc pas les mêmes valeurs à  $\beta_{min}$  et  $\beta_{max}$  dans le cas des simulations pour les ACP ou pour les AFC. Pour les ACP, je pose  $\beta_{min} = \beta_{max} = 0.008$ , et pour les AFC  $\beta_{min} = \beta_{max} = 0.005$ . Enfin une SIMULATION est faite pour chacune des valeurs de  $d_{moy}$  suivantes : 0.1, 0.3, 0.5, 0.7 et 0.9.

(ii) une PdD différente pour chaque espèce avec d'une part des probabilités  $d_{moy.j}$  différentes pour chaque espèce en posant  $d_{moy} = 0.5$  et  $v_{detect} = 0.2$ , et d'autre part des  $\beta_j$  différents entre espèces. Je choisis de pouvoir avoir des pentes faisant varier la PdD de 0 à 1 sur la portion de gradients couverte par les sites. Les valeurs des pentes étant tirées dans une loi uniforme, je prends garde à ne pas avoir trop de valeurs permettant cette forte variation de la PdD dans l'intervalle des valeurs extrêmes de ma loi. Encore une fois les valeurs sont différentes entre les simulations pour les ACP et celles pour les AFC. Pour les ACP  $\beta_{min} = 0$  et  $\beta_{max} = 0.03$ , et pour les AFC  $\beta_{min} = 0$  et  $\beta_{max} = 0.01$ .

**Avec un gradient de détection quadratique** : Enfin, je donne une forme quadratique à la PdD le long de la variable  $g_1$ . Ceci illustre le constat que les individus des espèces végétales, à certains stades de leur développement, ont une plus forte PdD qu'à d'autres -Par exemple le stade en fleur versus les stades fanés et végétatifs (Chen et al. 2013). Appliqué le long d'un gradient altitudinal, on peut supposer que pour une espèce végétale les individus à basse altitude sont déjà fanés, ceux de moyenne altitude en fleur, et ceux de haute altitude encore au stade végétatif. On utilise donc la relation suivante :

$$\text{logit}(d_{ij}) = \gamma_j g_{i1}^2 - (2\gamma_j \mu_j) g_{i1} + \left( \ln \left( \frac{d_{max.j}}{1 - d_{max.j}} \right) + \gamma_j \mu_j^2 \right)$$

Dans cette hypothèse je ferai varier la PdD :

- (i) Entre site en fonction de leur position sur  $g_1$ ,
- (ii) Entre espèce, en fonction d'une PdD maximale ( $d_{max.j}$ ), de la position de ce maximum sur  $g_1$  ( $\mu_j$ ), et par un facteur de diminution de la PdD ( $\gamma_j$ ); chacune de ces trois variables étant propre à chaque espèce et calculée à chaque simulation.

Les valeurs  $\gamma_j$  sont tirées à chaque simulation aléatoirement dans une loi uniforme définie par ses valeurs extrêmes ( $\gamma_{min}$  et  $\gamma_{max}$ ).

Les valeurs  $d_{max.j}$  sont tirées dans une loi normale, dont on fixe la valeur moyenne ( $d_{max}$ ) et l'écart type ( $v_{detect}$ ). Les valeurs sortant de l'intervalle [0.1, 0.9] sont, comme précédemment, modifiées pour appartenir à cet intervalle. Le choix d'une loi normale permet de limiter la probabilité d'avoir des valeurs très écartées de la valeur moyenne en comparaison à une loi uniforme. Moyenne et écart type de cette loi normale, tout comme  $\gamma_{min}$  et  $\gamma_{max}$  de la loi uniforme, sont fixes pour une SIMULATION. Toutes ces valeurs évolueront suivant deux hypothèses :

(i) une PdD identique entre toutes les espèces avec  $v_{detect} = 0$  et  $\gamma_{min} = \gamma_{max}$ . Dans cette hypothèse, afin de se placer sur des gradients moyens, j'impose une valeur de  $\gamma_j$  suffisante (pour que le gradient se constate après dégradation des abondances par la PdD), mais ne faisant pas passer les PdD des espèces de 0 à 1 sur la portion de gradients couverte par les sites. Aussi, les mêmes considérations que lors des simulations des profils d'abondance sont à prendre en compte

pour établir les valeurs  $\gamma_j$ . Je ne donne donc pas les mêmes valeurs à  $\gamma_{min}$  et  $\gamma_{max}$  dans le cas des simulations pour les ACP ou pour les AFC. Pour les ACP, je pose  $\gamma_{min} = \gamma_{max} = -0.00004$ , et pour les AFC  $\gamma_{min} = \gamma_{max} = -0.00001$ . Enfin une SIMULATION est faite pour chacune des valeurs de  $d_{max}$  suivantes : 0.1, 0.3, 0.5, 0.7 et 0.9.

(ii) une PdD différente pour chaque espèce avec d'une part des probabilités  $d_{max,j}$  différentes pour chaque espèce en posant  $d_{max} = 0.5$  et  $v_{detect} = 0.2$ , et d'autre part des  $\beta_j$  différents entre espèces. Encore une fois les valeurs sont différentes entre les simulations pour les ACP et celles pour les AFC. Pour les ACP  $\gamma_{min} = 0$  et  $\gamma_{max} = -0.00004$ , et pour les AFC  $\gamma_{min} = 0$  et  $\gamma_{max} = -0.00001$ .

Les valeurs  $\mu_j$  évolueront suivant deux hypothèses :

(i) Toutes les espèces ont une PdD maximale au centre du gradient de  $g_1$  couvert par les sites, ( $\mu_j = 500$  pour les AFC et 750 pour les ACP) –illustrant une situation pour laquelle toutes les espèces végétales ont une même phénologie et sont au même stade de développement à une altitude donnée.

(ii) Toutes les espèces ont une PdD maximale plus ou moins décalée par rapport à leur abondance moyenne maximale (optimum écologique) –comme si la prospection se faisait en avance sur la floraison à l'optimum écologique. Le décalage est aléatoire et tiré dans une loi uniforme ayant pour bornes 0 et 300.

Il semble fort probable que ces deux hypothèses sur  $\mu_j$  induisent des courbes d'abondances unimodales. Suivant ce qui a été dit au premier chapitre dans les parties 1.2.2 et 1.2.3, les courbes unimodales sont impropres à l'analyse par ACP. C'est pourquoi, voyant déjà une limite à l'utilisation de ces hypothèses pour les ACP, seule la première hypothèse sera étudiée par le biais des ACP.

Au delà de ces trois gradients et afin d'explorer les effets de la PdD sur les résultats des analyses, j'ai choisi d'appliquer ces gradients à l'ensemble des espèces, puis dans un second temps seulement à 11 espèces sur 20. Ces 11 espèces sont fixées et sont les espèces numérotées 1, 2, 3, 4, 5, 13, 14, 15, 16, 17 et 18. Le choix de 11 espèces sur 20 permet d'avoir deux groupes d'espèces qui peuvent contenir équitablement (sur 1000 simulations) des espèces rares et abondantes, généralistes et spécialistes. Ceci dans le but d'identifier si possible une différenciation des deux groupes par l'analyse sur les tableaux d'abondances dégradées.

Au final, je simule plusieurs « formes » de variation de la PdD, que j'applique toutes sur les quatre profils d'abondance évoqués dans la partie 2.2.2. J'obtiens ainsi l'ensemble de mes SIMULATIONS.

## 2.2.4 Des indicateurs de changements entre ACP ou AFC

A l'issue de chaque simulation, j'ai les résultats d'une AFC ou d'une ACP sur la matrice d'abondances brutes, et les résultats d'une AFC ou d'une ACP sur la matrice d'abondances dégradées.

Pour estimer les effets de la PdD sur ces analyses, il me reste à comparer les AFCnd aux AFCd et les ACPnd aux ACPd. Étant donné que j'ai construit les populations à partir d'au plus deux variables indépendantes, les AFCnd et les ACPnd ne peuvent au mieux faire ressortir que deux variables latentes. Je ne regarde donc que les axes d'ordre 1 et 2 de chaque analyse (nommé axe 1 et axe 2) et le plan formé par ces deux axes.

Commençons par les comparaisons et les indicateurs utilisés dans le cas des ACP.

Dans les ACP les lignes des tableaux analysés sont des individus statistiques et les colonnes des variables. Les graphiques représentant l'ordination des sites sur les axes principaux et ceux représentant les espèces (les variables) sur les composantes principales ne se lisent pas de la même manière (cf. partie 1.2.1). Je choisis néanmoins de ne faire porter mes comparaisons que sur les résultats qui concernent les espèces. En effet, en tant qu'écologue, ce sont les ordinations des espèces qui m'intéressent ; de plus une dégradation de l'ordination des espèces dans les résultats des analyses implique également une dégradation de l'ordination des sites.

Un des intérêts des ACP est, nous l'avons vu, de fournir des ordinations des espèces le long des composantes principales. Je comparerai donc l'ordination des espèces sur chaque composante principale de l'ACPnd à celle l'ACPd. Pour faire cette comparaison, je calculerai le coefficient de corrélation de rangs de Kendall entre ces deux **ordinations**. Le coefficient de corrélation de rangs de Kendall donne une valeur de corrélation entre deux variables quantitatives mesurées sur les mêmes individus. Ce coefficient est obtenu en comparant le rang des individus sur une variable au rang des mêmes individus sur l'autre variable. Ce test est non paramétrique. On trouve une définition et les formules de calcul de ce coefficient de corrélation dans Millot (2011). Ce coefficient de corrélation prend des valeurs entre -1 et 1 et s'interprète comme tout coefficient de corrélation. Le sens des axes pouvant être inversé d'une ACP à une autre, je regarde les valeurs absolues de ce coefficient. Par la suite, je regarderai et comparerai toujours les valeurs absolues des coefficients de corrélation de rang de Kendall.

Puisque les populations sont simulées pour être dépendantes de l'une ou des deux variables  $g_1$  et  $g_2$ , on s'attend à ce que l'une ou les deux premières composantes principales de l'ACPnd correspondent à ces variables. Les coordonnées des espèces sur ces composantes principales permettent de calculer les coefficients de corrélation linéaire entre les variables espèces et la composante principale. Par construction, les abondances des espèces sont corrélées positivement ou négativement à  $g_1$  et  $g_2$  en fonction de leurs valeurs d'optimums. Je rappelle que les 10 espèces numérotées de un à dix sont corrélées négativement à  $g_1$ , et les 10 autres espèces sont positivement corrélées à  $g_1$ . Les espèces impaires sont corrélées négativement à  $g_2$  et les espèces paires, sont positivement corrélées à  $g_2$  (cf. partie 2.2.2).

Par conséquent, pour chaque ACPnd nous savons si les espèces de un à dix (qu'on appellera le groupe 1) sont séparées des espèces onze à vingt (qu'on appellera le groupe 2) sur une composante principale ; et/ou les espèces paires (qu'on appellera le groupe A) doivent être séparées des espèces impaires (qu'on appellera le groupe B) sur l'autre composante principale.

À partir de là, je m'attacherai à évaluer **la séparation de ces groupes**, entre eux (1 et 2 ou A et B) dans une même ACP, entre ACPnd et ACPd, et ce pour chacune des composantes principales.

Dans un premier temps, je dois orienter toutes les ACP dans le même sens. Pour ce faire, je calcule à chaque simulation la médiane des coordonnées des groupes 1 et A, sur les composantes principales. Si une médiane est négative, on inverse le sens de l'axe correspondant en multipliant par -1 les coordonnées de toutes les espèces sur cet axe. Une fois regroupées par SIMULATION je peux évaluer la variabilité de la position des espèces d'un groupe sur chaque axe. Cette évaluation graphique se fera par le biais de représentations en boxplots. Je pourrai alors évaluer :

- (i) Les répartitions de chaque groupe sur chaque axe,
- (ii) La séparations des groupes 1 et 2 ou A et B dans une même ACP,
- (iii) La similitude des séparations des groupes entre ACPnd et ACPd.

J'étudierai également à l'aide du coefficient de corrélation de Kendall le lien potentiel entre les ordinations des espèces et les valeurs de PdD de chacune d'elles -  $d_{moy,j}$  dans le cas des gradients de détection linéaire,  $d_{max,j}$  pour les gradients de détection quadratique. Ce lien sera mis en évidence en comparant le coefficient de corrélation de Kendall calculé avec les ordinations issu de l'ACPnd et celui calculé avec les ordinations issu de l'ACPd.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Un dernier indicateur dans le cadre des ACP permettant d'évaluer la dégradation potentielle des résultats entre ACPnd et ACPd, utilise la possibilité d'ajouter **des colonnes supplémentaires** à la matrice de données de départ. Il s'agit de variables qui apparaissent dans les résultats sans qu'elles participent à la construction des composantes principales, mais dont les corrélations avec celles-ci sont explicitées (Prodon et Jacquet 2014). J'intègre donc des variables supplémentaires non dégradées par la PdD, dont on connaît le comportement dans l'ACPnd. Il s'agira alors de comparer ce comportement à celui observé dans l'ACPd. Le modèle établi permettant d'affirmer que les composantes principales doivent révéler à bon escient les variables  $g_1$  et  $g_2$ , j'utiliserai ces deux variables comme variables supplémentaires (les valeurs  $g_{i1}$  et  $g_{i2}$ ). Concrètement, l'analyse se fera par la comparaison des coordonnées des vecteurs supplémentaires sur les composantes principales dans l'ACPnd, aux coordonnées de ces mêmes vecteurs dans les ACPd.

Hormis l'utilisation des variables supplémentaires, j'utilise toutes les méthodes de comparaisons des ACPnd aux ACPd, décrites ci-dessus, pour comparer les AFCnd aux AFCd. Je ne regarde également que les résultats vis-à-vis des espèces. Rappelons ici la partie 2 du premier chapitre, qui me permet d'affirmer que les comparateurs utilisés pour les ACP sont également valables pour les AFC.

## Chapitre 3 : Présentation des résultats

### 3.1 Préambule à la présentation

Les résultats sont présentés à l'aide de graphiques. Afin d'en simplifier la lecture, je donne quelques indications sur l'organisation de ces graphiques et sur la terminologie utilisée. Une fiche d'aide à la lecture détachable reprend les éléments suivants et est disponible à la fin du document en annexe B.

Une figure présentant des résultats se compose de plusieurs graphiques qui concernent un même indicateur. Chaque graphique présente les résultats de plusieurs SIMULATIONs qui forme alors un ENSEMBLE de SIMULATIONs. Pour former ces ENSEMBLES, je regroupe toutes les SIMULATIONs dans lesquelles l'hétérogénéité de PdD est similaire. Ainsi, on différencie trois types d'ENSEMBLE :

- L'un où dans chaque SIMULATIONs, les espèces ont toutes le même gradient de PdD ;
- Le second où dans chaque SIMULATIONs, 11 espèces sur 20 ont toutes le même gradient de PdD, les 9 restantes ayant une PdD constante et égale à 1 ;
- Le dernier où dans chaque SIMULATIONs, les espèces ont un gradient de PdD différents.

Chacun de ces trois types d'ENSEMBLES se retrouve dans les différentes formes de gradient de PdD (sans gradient, linéaire, etc.).

Aussi, j'ai indiqué que chaque forme de variabilité de la PdD est appliquée sur les quatre profils d'abondance moyenne des espèces (ne dépendant ni de  $g_1$  ni de  $g_2$ , que de  $g_1$ , de  $g_1$  et de  $g_2$ , que de  $g_2$  ; cf. chapitre 2). C'est pourquoi on retrouve dans chaque graphique les SIMULATIONs regroupées par profils d'abondances moyennes (spécifiés en abscisse par P1, P2, P3 ou P4).

Enfin, les SIMULATIONs, au sein d'un même profil, seront alors ordonnées de façon croissante en fonction des valeurs  $d_{moy.j}$  ou  $d_{max.j}$ . Ainsi je révélerai dans chacun de ces graphiques les effets de l'intensité de la PdD à travers les variations des valeurs de  $d_{moy.j}$  ou  $d_{max.j}$  dans un même profil d'abondances moyennes.

Je rappelle que les numéros des SIMULATIONs apportent des informations complémentaires sur les valeurs des paramètres sous-jacents (cf. les figures 5 et 6 ainsi que la feuille détachable).

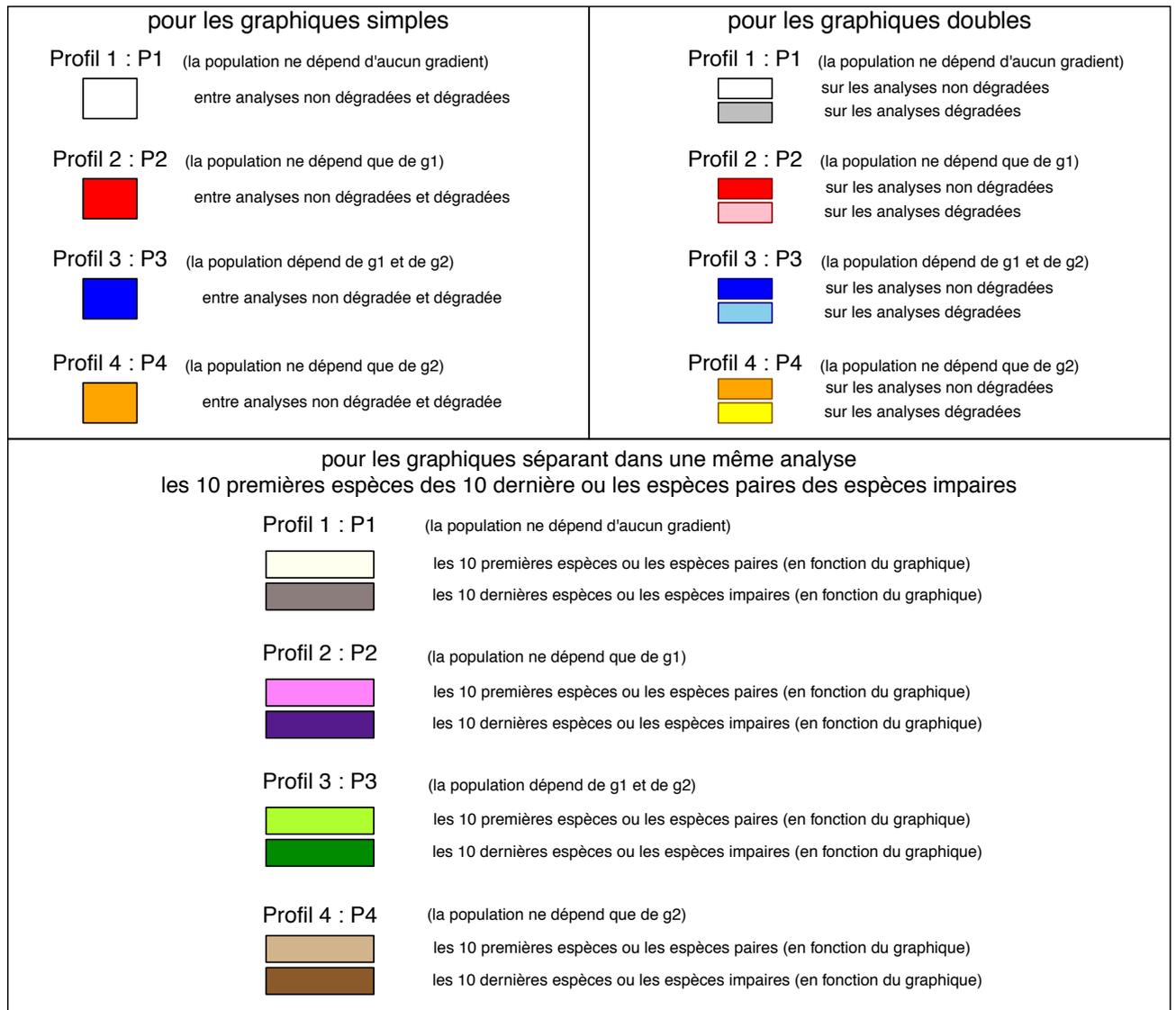
Dans ce chapitre je présente uniquement pour chaque ENSEMBLES de SIMULATIONs les résultats des indicateurs qui apportent de nouvelles informations et qui répondent à ma problématique. Les résultats non présentés peuvent être fournis sur demande.

En ce qui concerne les formes et légendes des graphiques, celles-ci sont fonction des indicateurs (cf. figure 7 page suivante). Certains indicateurs présentés dans le chapitre précédent sont des mesures directes de la différence entre les analyses non dégradées et les analyses dégradées. Les résultats de ces mesures sont alors présentés dans des graphiques dits « simples ». Autrement dit, il n'y a qu'un seul boxplot par SIMULATIONs. C'est le cas, par exemple, pour le coefficient de corrélation de Kendall entre les ordinations issues de l'analyse non dégradée et de l'analyse dégradée.

Dans d'autres cas, une mesure faite sur l'analyse non dégradée est comparée à la même mesure faite sur l'analyse dégradée. Je présente alors simultanément ces deux mesures dans un même graphique dit « double ». Il y a donc deux boxplots par SIMULATIONs. Cette forme est notamment utilisée pour les valeurs absolues des coordonnées des vecteurs supplémentaires.

Reste le cas des graphiques des coordonnées des groupes d'espèces 1 et 2, ou A et B. On présente alors, dans une même figure les quatre graphiques des boxplots tels que décrits dans la partie 2.2.3.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



*Figure 7 : Légendes des graphiques. Trois types de graphiques sont utilisés pour la présentation des résultats. Le premier, les « graphiques simples » qui présentent pour chaque SIMULATION un boxplot des valeurs données par un indicateur comparant directement les résultats de l'analyse non dégradée à ceux de l'analyse dégradée. Le deuxième, « les graphiques doubles » présentent pour chaque SIMULATION deux boxplots que l'on peut comparer, un pour les valeurs données par un indicateur appliqué sur les résultats de l'analyse non dégradée, et un pour les valeurs du même indicateur appliqué sur les résultats de l'analyse dégradée. Enfin le dernier, le graphique des boxplots des coordonnées des groupes d'espèces 1 et 2 ou des groupes d'espèces A et B. Dans chaque graphique je regroupe les SIMULATIONS par profil d'abondances moyennes des espèces utilisé. Je définis des couleurs propres pour chaque profil (P1, P2, P3, et P4).*

Toujours pour simplifier la lecture, la légende couleur utilisée pour chaque type de graphique (simple, double, etc.) restera strictement identique, et telle que décrite dans la figure 7 qui est reprise dans la fiche de lecture.

Une dernière précision : j'utilise souvent comme indicateur des coefficients de corrélations. N'ayant pas besoin de les décrire avec une grande précision, j'utiliserai les termes de corrélations : faibles pour celles comprises entre 0 et 0.4, moyennes entre 0.4 et 0.7, et fortes entre 0.7 et 1.

## 3.2 Les résultats pour les ACP

### 3.2.1 Les SIMULATIONs sans gradient de détection

Je présente en premier les SIMULATIONs qui traitent des cas **sans gradient de détection**. Il y a deux ENSEMBLEs de SIMULATIONs dans ce cas. Le premier ENSEMBLE (ENSEMBLE 1) regroupe les SIMULATIONs où dans chacune d'elles les espèces ont toutes la même PdD et dont les trois derniers chiffres sont compris entre 001 et 009. Le deuxième ENSEMBLE (ENSEMBLE 2) regroupe les SIMULATIONs où dans chacune d'elles les espèces ont des PdD différentes et dont les derniers chiffres sont égaux à 102.

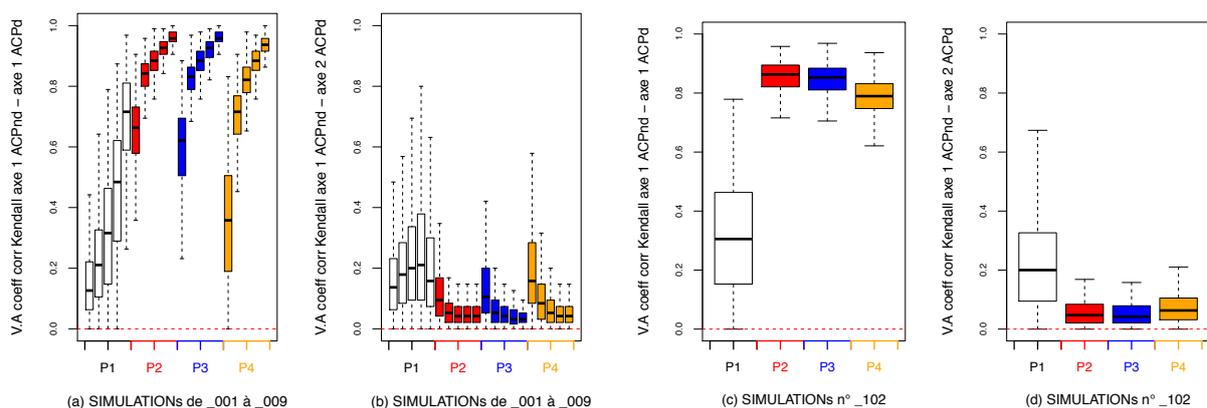
Je commence par les résultats des valeurs absolues des **coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1 de l'ACPnd**, et l'ordination des espèces sur les axes 1 et 2 de l'ACPd. Nous les retrouvons dans la figure 8.

**Conseils de lecture :** On juxtapose dans la figure les deux graphiques de chaque ENSEMBLE de SIMULATIONs. Le premier graphique de chaque ENSEMBLE (graphique (a) et (c) de la figure) correspond au coefficient de corrélation avec l'axe 1 de l'ACPd ; et le deuxième de chaque ENSEMBLE (graphique (b) et (d)) avec l'axe 2 de l'ACPd.

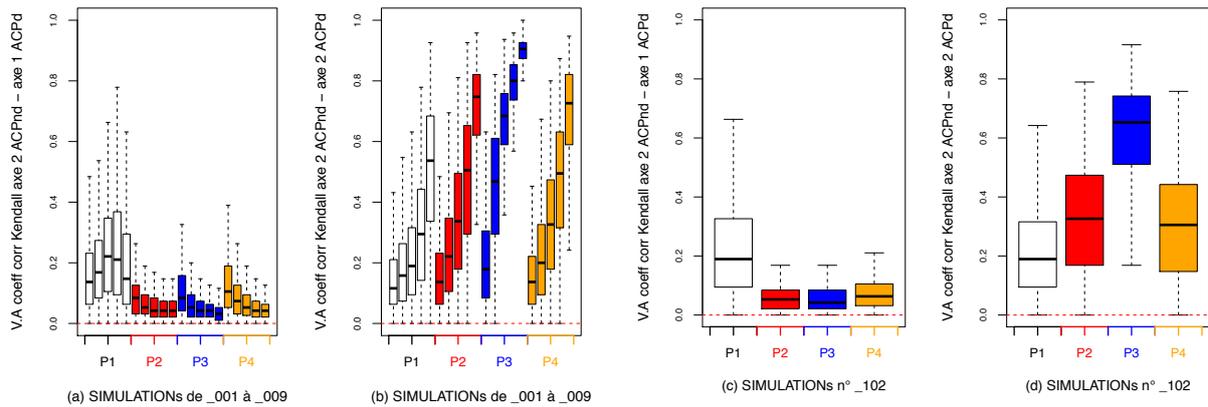
**Observations :** Les premiers graphiques de chaque ENSEMBLE (graphiques (a) et (c)) donnent des résultats comparables. D'une part les corrélations sont globalement élevées dans les profils d'abondance 2, 3 et 4, où l'abondance dépend au moins d'une des deux variables d'habitat ( $g_1$  ou  $g_2$ ), et les corrélations dans le profil 1 (où les abondances ne dépendent d'aucune des deux variables d'habitat) sont nettement inférieures à celles des autres profils. D'autre part, les coefficients de corrélation des SIMULATIONs s'ordonnent en fonction du profil (toute chose étant égale par ailleurs) comme suit :  $P2 > P3 > P4 > P1$ .

Le graphique (a) montre que dans chaque profil, la corrélation augmente avec l'augmentation de la PdD  $d_{moy,j}$ . De plus la SIMULATIONs des profils 2, 3 et 4, pour lesquels  $d_{moy,j}$  égale 0.1, est nettement décrochée des quatre suivantes.

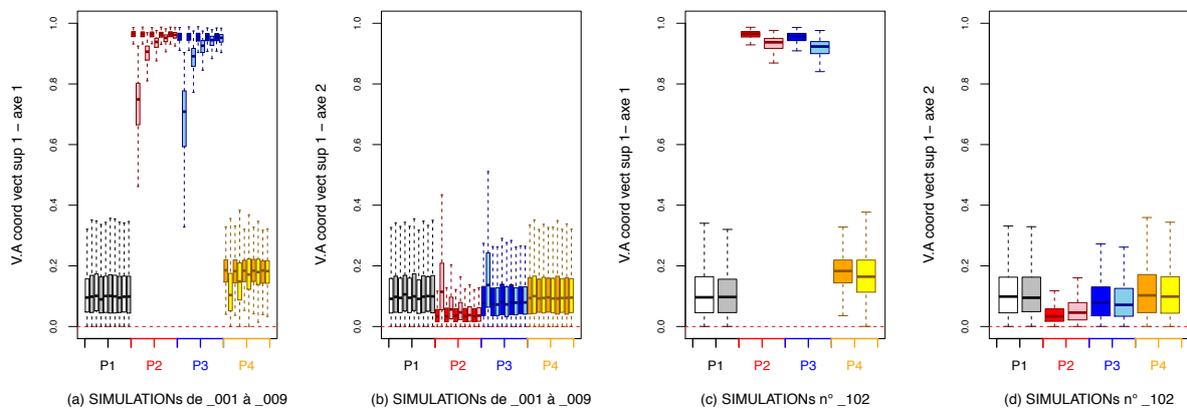
Dans les graphiques (b) et (c) (concernant l'axe 2), on note de faibles coefficients de corrélations dans tous les cas, mais avec des corrélations plus importantes et plus variables dans le profil 1 (P1) que dans les trois autres profils (P2, P3 et P4).



**Figure 8 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 1 de l'ACPnd, et les ordinations des espèces sur l'axe 1 (graphiques (a) et (c)) et l'axe 2 (graphiques (b) et (d)) de l'ACPd ; lors de SIMULATIONs sans gradient de PdD. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 1 où dans chacune des SIMULATIONs toutes les espèces ont la même PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 2 où dans chacune des SIMULATIONs les espèces ont des PdD différentes. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.



**Figure 9 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 2 de l'ACPnd, et les ordinations des espèces sur l'axe 1 (graphiques (a) et (c)) et l'axe 2 (graphiques (b) et (d)) de l'ACPd ; lors de SIMULATIONs sans gradient de PdD. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 1 où dans chacune des SIMULATIONs toutes les espèces ont la même PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 2 où dans chacune des SIMULATIONs les espèces ont des PdD différentes. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.



**Figure 10 :** graphique double (deux boxplots par SIMULATION, un pour l'ACPnd et un pour l'ACPd) des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_1$ , sur les axes 1 (graphique (a) et (c)) et 2 (graphique (b) et (d)) des ACPnd et des ACPd, lors de SIMULATIONs sans gradient de PdD. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 1 où dans chacune des SIMULATIONs toutes les espèces ont la même PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 2 où dans chacune des SIMULATIONs les espèces ont des PdD différentes. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

La figure 9 (ci-contre), présente les résultats des valeurs absolues des **coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 2 de l'ACPnd**, et l'ordination des espèces sur les axes 1 et 2 de l'ACPd

La lecture de cette figure est similaire à celle de la figure précédente.

**Observations :** Les graphiques (a) et (c) (concernant l'axe 1), montrent de faibles corrélations quel que soit le profil.

Les graphiques (b) et (d) (concernant l'axe 2), montrent une corrélation nettement plus importante dans le profil 3 (P3) (où les abondances moyennes dépendent des deux variables d'habitat). Ce profil présente des corrélations de moyennes à fortes dans l'ENSEMBLE 2 (PdD différentes entre espèces) et dans les quatre dernières SIMULATIONs dans l'ENSEMBLE 1 (même PdD entre espèces).

En regardant uniquement le graphique (b) qui représente les résultats de l'ENSEMBLE 1 (même PdD entre espèces) vis-à-vis de l'axe 2, on constate également une croissance des corrélations avec l'augmentation de  $d_{moy.j}$ . Cette croissance est forte. En effet, les corrélations passent de faibles à moyennes pour le profil 1, de faibles à fortes pour les profils 2, 3 et 4 et  $P3 > P2 \geq P4$ .

Présentons maintenant, dans la figure 10 (ci-contre), les résultats **des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_1$** , sur les axes 1 et 2 des ACPnd et des ACPd.

**Conseils de lecture :** Le mode de lecture ne change toujours pas, cependant les graphiques sont « doubles ». On s'intéresse donc à la position de chaque boxplot, mais également aux écarts entre les boxplots correspondant aux ACPnd et à ceux correspondant aux ACPd.

**Observations :** Les deux ENSEMBLES (1 et 2) (même PdD entre les espèces, et PdD différentes entre les espèces) donnent des résultats comparables, aussi bien sur les axes 1 que sur les axes 2.

Les graphiques (a) (pour l'ENSEMBLE 1, axe 1) et (c) (ENSEMBLE 2, axe 1) affichent comme attendus pour les ACPnd, une faible corrélation du vecteur supplémentaire avec l'axe 1 dans les profils 1 et 4 (où les abondances moyennes ne dépendent pas de  $g_1$ ), et une forte corrélation dans les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ).

Pour les ACPd, on retrouve les faibles corrélations des profils 1 et 4, mais si les corrélations dans les profils 2 et 3 restent fortes, ces dernières sont systématiquement inférieures à celles obtenues lors des ACPnd. La différence est d'autant plus forte que la valeur  $d_{moy.j}$  diminue.

Enfin les graphiques (b) (ENSEMBLE 1, coordonnées sur l'axe 2) et (c) (ENSEMBLE 2, coordonnées sur l'axe 2) montrent une faible corrélation de l'axe 2 avec le vecteur supplémentaire quel que soit le profil ou l'analyse (d ou nd).

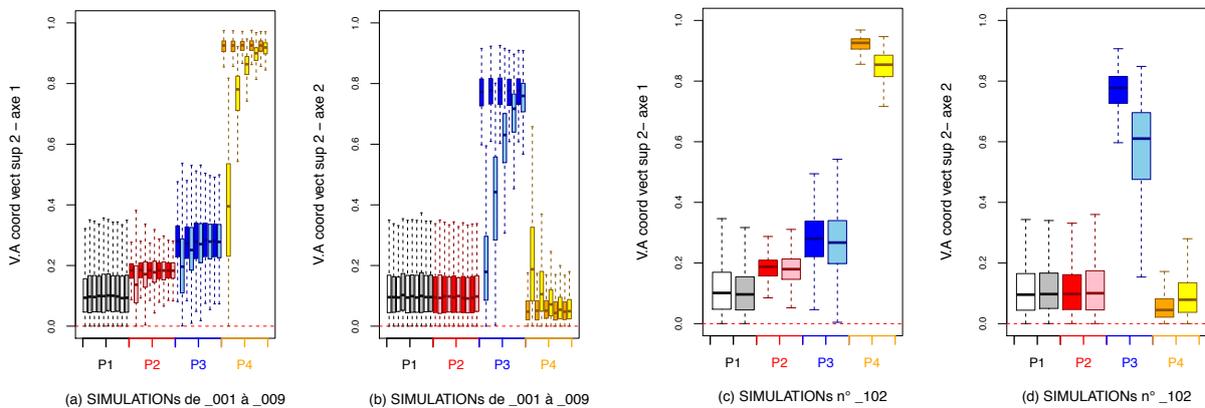
Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

La figure 11 regroupe les résultats des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_2$ , sur les axes 1 et 2 des ACPnd et des ACPd.

La lecture de cette figure est similaire à celle de la figure précédente.

**Observations :** On note les fortes corrélations dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) sur l'axe 1 (graphique (a) et (c)) et dans le profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) sur l'axe 2 (graphique (b) et (d)) lors des ACPnd. Dans le cas des ACPd, ces corrélations sont systématiquement plus faibles. De plus ces différences augmentent et deviennent importantes lorsque la valeur  $d_{moy,j}$  diminue. C'est le profil 3 sur l'axe 2 qui présente les écarts les plus marqués.

Dans les autres cas on ne relève au mieux que de légères différences entre les corrélations des ACPnd et des ACPd. Ces différences vont en diminuant lorsque  $d_{moy,j}$  augmente.



**Figure 11 :** graphique double (deux boxplots par SIMULATION, un pour l'ACPnd et un pour l'ACPd) des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_2$ , sur les axes 1 (graphique (a) et (c)) et 2 (graphique (b) et (d)) des ACPnd et des ACPd ; lors de SIMULATIONs sans gradient de PdD. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 1 où dans chacune des SIMULATIONs toutes les espèces ont la même PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 2 où dans chacune des SIMULATIONs les espèces ont des PdD différentes. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

Je retiens des résultats concernant une PdD sans gradient (pas de variation entre les sites), que l'ordination des axes 1 et 2 des ACPnd est dans tous les cas modifiée, donc dégradée par la PdD. Cette dégradation est d'autant plus importante que la PdD est faible et est plus importante sur les axes d'ordre 2. Enfin, le fait d'imposer une PdD différente entre les espèces tend à faire légèrement augmenter les dégradations.

### 3.2.2 Les SIMULATIONs avec un gradient de détection linéaire.

Je présente maintenant les SIMULATIONs avec un gradient de détection linéaire. Je m'appuie sur trois ENSEMBLES de SIMULATIONs. Le premier ENSEMBLE (ENSEMBLE 3) regroupe les SIMULATIONs où dans chacune d'elles les espèces ont toutes le même gradient de PdD et dont les trois derniers chiffres sont compris entre 021 et 029. Le deuxième ENSEMBLE (ENSEMBLE 4) regroupe les SIMULATIONs où dans chacune d'elles, 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1) et dont les trois derniers chiffres sont compris entre 031 et 039. Le dernier ENSEMBLE (ENSEMBLE 5) regroupe les SIMULATIONs où dans chacune d'elles les espèces ont des gradients de PdD différents et dont les derniers chiffres sont égaux à 162.

Regardons maintenant dans la figure 12 les résultats des valeurs absolues des **coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1 de l'ACPnd**, et l'ordination des espèces sur les axes 1 et 2 de l'ACPd. Elle regroupe les résultats des ENSEMBLES 4 (même gradient de PdD pour 11 espèces sur 20) et 5 (gradients de PdD différents entre espèces). En effet, les résultats de l'ENSEMBLE 3 (même gradient de PdD entre les espèces) diffèrent de ceux de l'ENSEMBLE 4 sur des points qui ne portent pas à conséquence pour mes conclusions.

**Conseil de lecture :** On retrouve la même présentation et donc le même mode de lecture que pour les figures similaires des ACP sans gradient de détection.

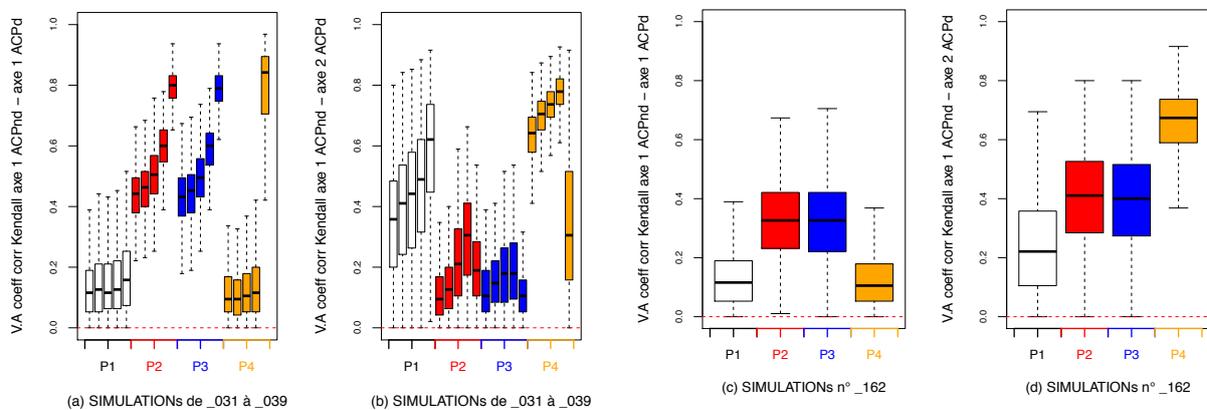
**Observations :** Dans le premier graphique (a) (concernant l'axe 1 pour l'ENSEMBLE 4), les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ) présentent des corrélations croissantes, de moyennes à fortes, avec l'augmentation de  $d_{moy.j}$ .

Ces mêmes profils (2 et 3) présentent dans le graphique (c) (concernant l'axe 1 pour l'ENSEMBLE 5) des corrélations globalement faibles.

Par contre, les profils 1 et 4 (où les abondances moyennes ne dépendent pas de  $g_1$ ) sur les axes 1, montrent quant à eux de faibles corrélations, à l'exception de la dernière SIMULATION ( $d_{moy.j} = 0.9$ ) du profil 4 du graphique (a) qui elle est forte.

Sur les graphiques (b) et (d) (concernant l'axe 2) je note pour les profils 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) des corrélations:

- De valeur moyenne et supérieure à celles des autres profils dans le graphique (d) (ENSEMBLE 5).
- Croissantes de moyenne à forte, jusqu'à ce que  $d_{moy.j}$  atteigne la valeur 0.7 dans le graphique (b) (ENSEMBLE 4). Au delà, lorsque les  $d_{moy.j}$  atteignent la valeur 0.9, les corrélations chutent.

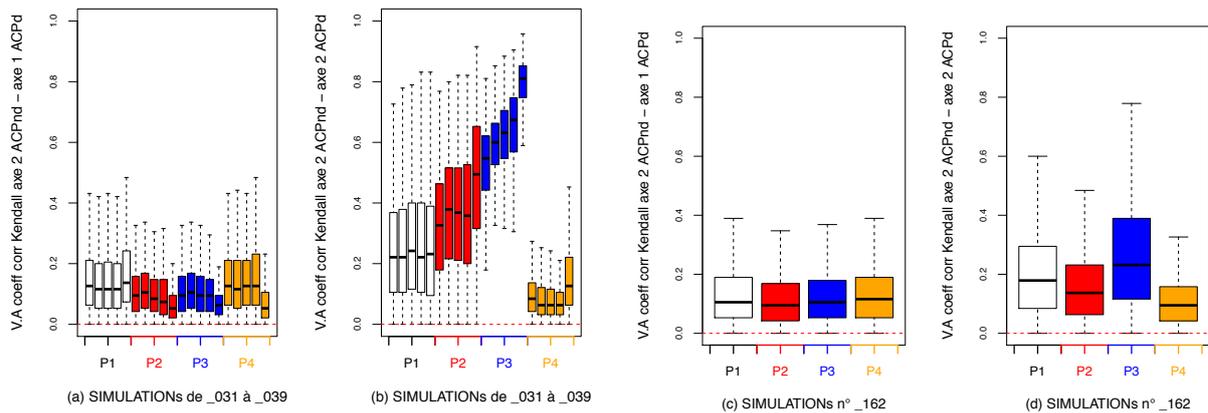


**Figure 12 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 1 de l'ACPnd, et les ordinations des espèces sur l'axe 1 (graphiques (a) et (c)) et l'axe 2 (graphiques (b) et (d)) de l'ACPd ; lors de SIMULATIONs avec un gradient de PdD linéaire. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 4 où dans chacune des SIMULATIONs 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 5 où dans chacune des SIMULATIONs les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

La figure 13 (page suivante) représente les valeurs absolues des **coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 2 de l'ACPnd**, et leurs ordinations sur les axes 1 et 2 de l'ACPd. Seuls les ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20) et 5 (gradients de PdD différents entre les espèces) sont représentés.

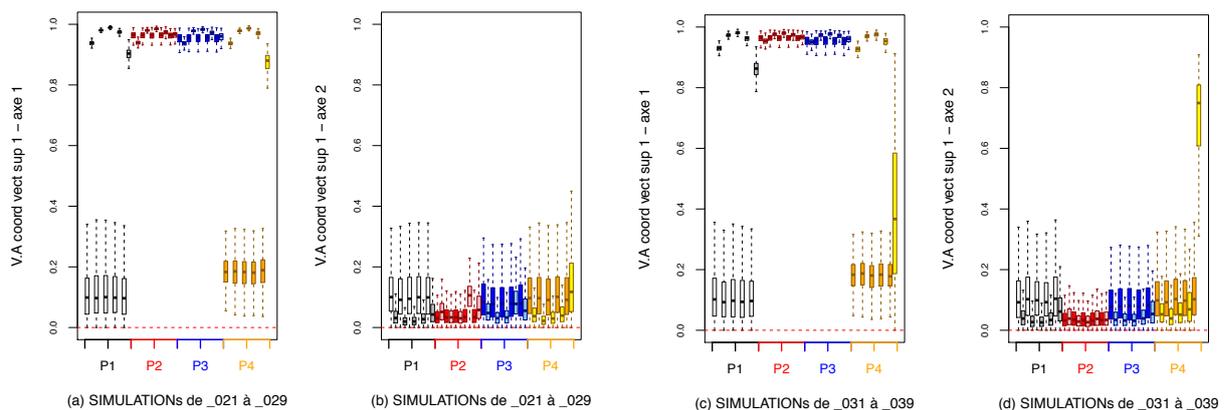
**Observations** : je remarque deux éléments :

- (i) D'une part, dans le graphique (b) de cette figure (axe 2 de l'ENSEMBLE 4), le profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) présente les plus fortes corrélations, croissantes de moyennes à fortes avec l'augmentation de  $d_{moy,j}$ .
- (ii) D'autre part, toutes les SIMULATIONs des graphiques (a), (c), (d), ainsi que les SIMULATIONs des profils 1 et 4 du graphique (b), présentent des corrélations faibles et avec peu de changement d'un profil à l'autre. J'insiste sur les faibles taux de corrélation dans le cas du profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) sur l'axe2 de l'ENSEMBLE 5 (graphique (d)).



**Figure 13** : graphique simple (un boxplot par SIMULATION) des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 2 de l'ACPnd, et les ordinations des espèces sur l'axe 1 (graphiques (a) et (c)) et l'axe 2 (graphiques (b) et (d)) de l'ACPd ; lors de SIMULATIONs avec un gradient de PdD linéaire. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 4 où dans chacune des SIMULATIONs 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 5 où dans chacune des SIMULATIONs les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

L'indicateur suivant est illustré par les graphiques doubles des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_1$ . La figure 14 regroupe les résultats des ENSEMBLES 3 (même gradient de PdD entre les espèces) et 4 (même gradient de PdD pour 11 espèces sur 20).



**Figure 14** : graphique double (deux boxplots par SIMULATION, un pour l'ACPnd et un pour l'ACPd) des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_1$ , sur les axes 1 (graphique (a) et (c)) et 2 (graphique (b) et (d)) des ACPnd et des ACPd ; lors de SIMULATIONs avec un gradient de PdD linéaire. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 3 où dans chacune des SIMULATIONs toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 4 où dans chacune des SIMULATIONs 11 espèces sur 20 ont un même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

**Observations :** On constate pour les deux ENSEMBLES (3 et 4), que le vecteur supplémentaire est fortement corrélé aux axes 1 des ACPnd (graphique (a) et (c)) pour les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ), et est peu corrélé pour les profils 1 et 4 (où les abondances moyennes ne dépendent pas de  $g_1$ ).

A contrario dans les ACPd, le vecteur supplémentaire est fortement corrélé aux axes 1 pour toutes les SIMULATIONs dans les profils, sauf lorsque que  $d_{moy.j}$  atteint 0.9 dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) de l'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20).

Sur les axes 2 (graphiques (b) et (d)), les corrélations sont faibles, à l'exception de la dernière SIMULATION ( $d_{moy.j} = 0.9$ ) du profil 4 de l'ENSEMBLE 4 (graphique (d)), dont les corrélations varient de moyennes à fortes.

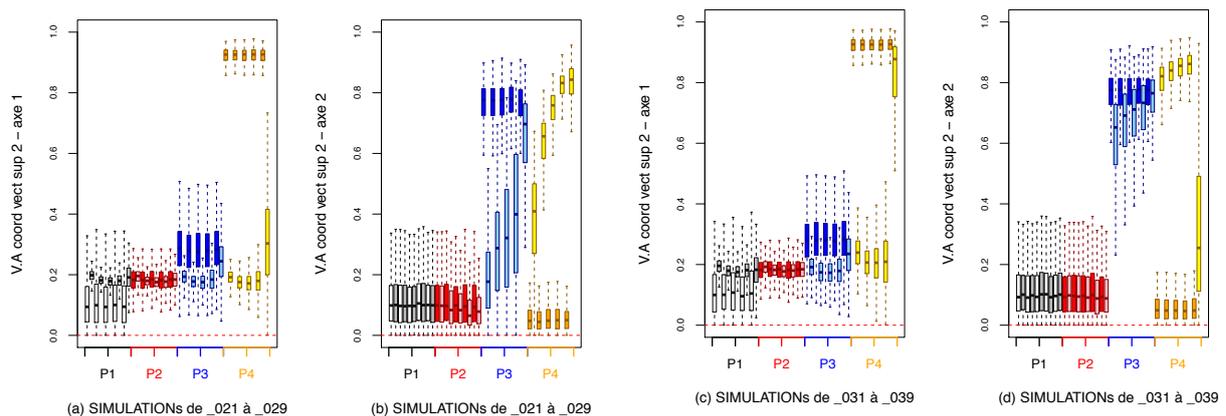
La figure 15 regroupe les résultats des ENSEMBLES 3 (même gradient de PdD entre les espèces) et 4 (même gradient de PdD pour 11 espèces sur 20) des **valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_2$** .

**Observations :** On note de fortes corrélations dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) sur l'axe 1 (graphique (a) et (c)) et dans le profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) sur l'axe 2 (graphique (b) et (d)) lors des ACPnd. Les corrélations, lors de toutes les autres AFCnd, sont faibles.

En revanche les ACPd présentent sur les axes 1 pour tous les profils des corrélations faibles et quasi constantes, à l'exception du profil 4 de l'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20) lorsque  $d_{moy.j}$  atteint 0.9.

Sur les axes 2 des ACPd, on remarque pour les profils 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) des corrélations croissantes avec  $d_{moy.j}$ , se rapprochant de celles des ACPnd. On constate également que ces écarts entre les corrélations des ACPd et celles des ACPnd sont nettement moins importantes dans l'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20).

D'autre part, pour les profils 4 (où les abondances moyennes ne dépendent que de  $g_2$ ), les corrélations lors des ACPd sont plus importantes que lors des ACPnd, et augmentent avec  $d_{moy.j}$ , passant de moyennes à fortes.



**Figure 15 :** graphique double (deux boxplots par SIMULATION, un pour l'ACPnd et un pour l'ACPd) des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_2$ , sur les axes 1 (graphique (a) et (c)) et 2 (graphique (b) et (d)) des ACPnd et des ACPd ; lors de SIMULATIONs avec un gradient de PdD linéaire. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 3 où dans chacune des SIMULATIONs toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 4 où dans chacune des SIMULATIONs 11 espèces sur 20 ont un même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

La figure 16 présente les graphiques doubles des **valeurs absolues des coefficients de corrélation de Kendall entre les valeurs  $d_{moy,j}$**  et l'ordination des espèces sur les axes 1 et 2 des analyses.

**Conseil de lecture :** La figure se lit de façon identique aux figures présentant les résultats des vecteurs supplémentaires. Elle regroupe les résultats des ENSEMBLES 4 (même gradient de PdD pour 11 espèces sur 20) et 5 (gradients de PdD différents entre les espèces).

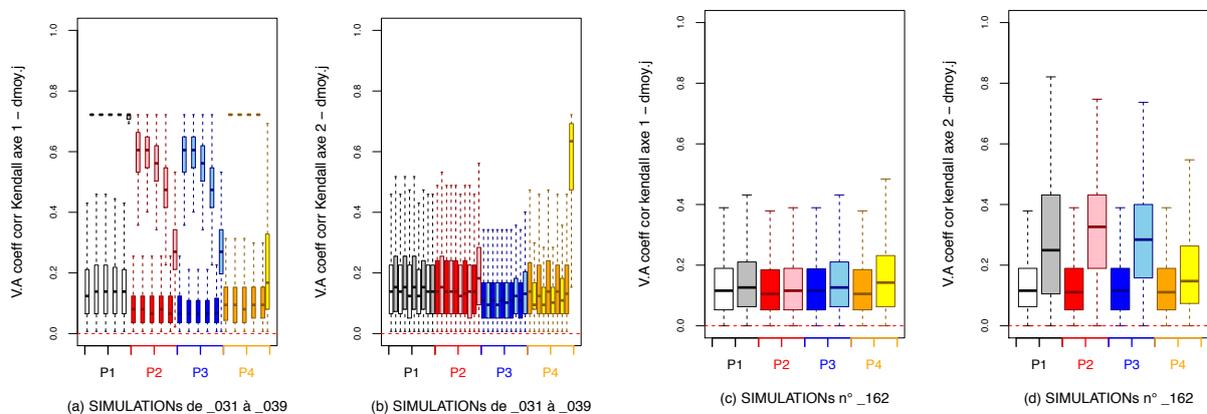
**Observations :** les corrélations sont constamment faibles lors des ACPnd dans tous les graphiques.

En revanche, pour les ACPd, on remarque que les résultats sur les axes 1 de l'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20) (graphique (a)) donnent des corrélations clairement plus importantes que lors des ACPnd correspondantes. Cette remarque vaut aussi pour les axes 2 de l'ENSEMBLE 5 (gradients de PdD différents entre espèces) avec cependant des écarts entre les coefficients de corrélations issus des ACPnd et ceux issus des ACPd nettement moins importants.

Le graphique (a) montre également que ces corrélations sont fortes et constantes dans tout le profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ) et dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) tant que  $d_{moy,j} \leq 0.7$ . Elles décroissent avec l'augmentation des valeurs de  $d_{moy,j}$  dans les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ), restant globalement moyenne.

On remarque aussi que si la corrélation décroît fortement lorsque  $d_{moy,j}$  égale 0.9 dans le profil 4 du premier graphique, celle ci augmente fortement sur l'axe 2 correspondant (graphique (b)).

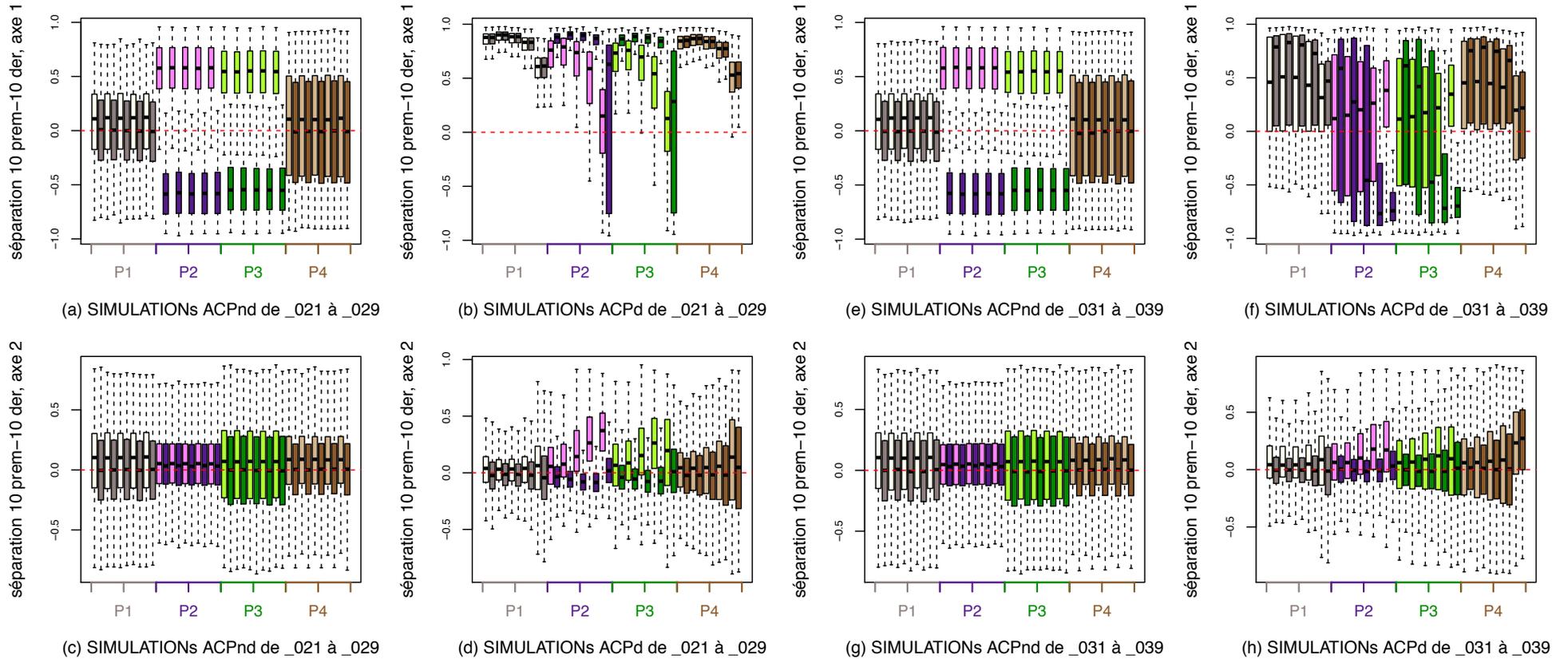
Enfin, sur le graphique (d), les corrélations sont faibles et décroissantes du profil 2 au profil 4.



**Figure 16 :** graphique double (deux boxplots par SIMULATION, un pour l'ACPnd et un pour l'ACPd) des valeurs absolues des coefficients de corrélation de Kendall entre les valeurs  $d_{moy,j}$  et l'ordination des espèces sur les axes 1 (graphique (a) et (c)) et 2 (graphique (b) et (d)) des analyses ; lors de SIMULATIONS avec un gradient de PdD linéaire. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 4 où dans chacune des SIMULATIONS 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 5 où dans chacune des SIMULATIONS les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

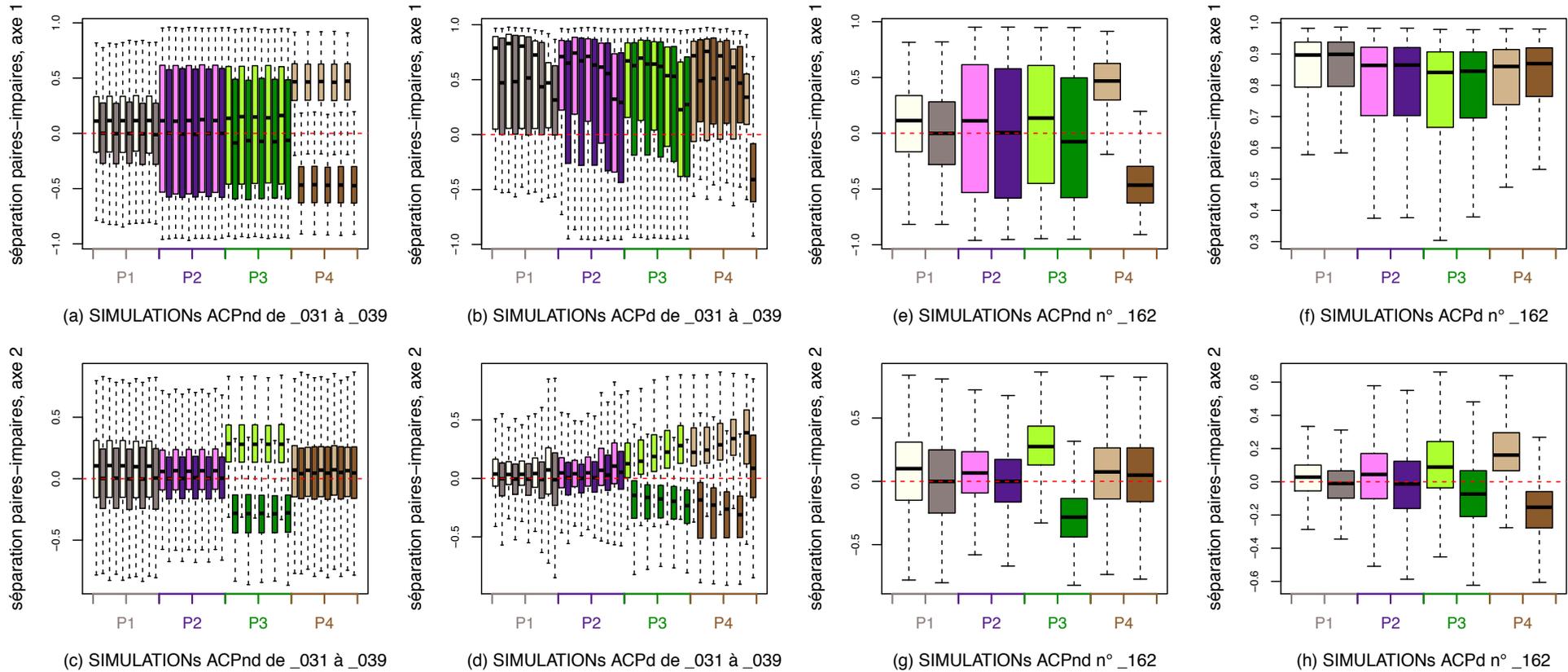
La figure 17 (page suivante) illustre les séparations **des groupes d'espèces 1 et 2**. On présente les résultats des ENSEMBLE 3 (même gradient de PdD pour toutes les espèces) et 4 (même gradient de PdD pour 11 espèces sur 20).

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 17 :** illustrant les « séparations » des groupes d'espèces 1 (les 10 premières espèces) et 2 (les 10 dernières espèces), lors de SIMULATIONs avec un gradient de PdD linéaire. Elle présente deux boxplots par SIMULATION, un pour chaque groupe d'espèces. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE 3 où dans chacune des SIMULATIONs toutes les espèces ont le même gradient de PdD. Les quatre derniers graphiques (e), (f), (g) et (h) illustrent les résultats de l'ENSEMBLE 4 où dans chacune des SIMULATIONs 11 espèces sur 20 ont un même gradient de PdD, les 9 autres ayant une PdD constamment égale à 1. Les première et troisième colonnes de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des ACPnd, et les deuxième et quatrième colonnes de graphiques ((b), (d) et (f), (h)) lors des ACPd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 18** : illustrant les « séparations » des groupes d'espèces A (les espèces paires) et B (espèces impaires), lors de SIMULATIONS avec un gradient de PdD linéaire. Elle présente deux boxplots par SIMULATION, un pour chaque groupe d'espèces. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE 4 où dans chacune des SIMULATIONS 11 espèces sur 20 ont un même gradient de PdD, les 9 autres ayant une PdD constamment égale à 1. Les quatre derniers graphiques (e), (f), (g) et (h) illustrent les résultats de l'ENSEMBLE 5 où dans chacune des SIMULATIONS toutes les espèces ont des gradients de PdD différents. Les première et troisième colonne de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des ACPnd, et les deuxième et quatrième colonne de graphiques ((b), (d) et (f), (h)) lors des ACPd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

**Conseil de lecture :** nous avons dans cette figure huit graphiques sur quatre colonnes. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE 3, et les quatre suivants (e), (f), (g) et (h) les résultats de l'ENSEMBLE 4. Les première et troisième colonnes de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des ACPnd, et les deuxième et quatrième colonnes de graphiques ((b), (d) et (f), (h)) lors des ACPd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses.

**Observations :** On retrouve bien lors des ACPnd les séparations attendues des groupes dans les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ) sur les axes 1 (graphique (a) et (e)). On a également dans tous les autres cas, une répartition des espèces de chaque groupe quasi équitable de chaque côté de l'origine de l'axe (médiante centrée sur l'axe).

Dans les ACPd, la séparation des groupes disparaît dans les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ) sur l'axe 1 totalement dans l'ENSEMBLE 3 (même gradient de PdD pour toutes les espèces) (graphique (b)), et jusqu'à ce que  $d_{moy,j} = 0.7$  dans l'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20) (graphique (f)).

On remarque aussi une tendance au regroupement des espèces du même côté de l'axe 1. Cette tendance est visible, bien que moins marquée dans les SIMULATIONs de l'ENSEMBLE 4 (graphique (f)) que dans celles de l'ENSEMBLE 3 (graphique (b)).

Enfin, dans l'ENSEMBLE 3 (les quatre premiers graphiques), la séparation des groupes, perdue sur l'axe 1 des ACPd (graphique (b)), se retrouve, partiellement sur l'axe 2 (graphique (d)).

Enfin, la figure 18 (page précédente), illustre la séparation **des groupes d'espèces A et B**. On affiche les résultats des ENSEMBLES 4 (même gradient de PdD pour 11 espèces sur 20) et 5 (gradients de PdD différents entre espèces).

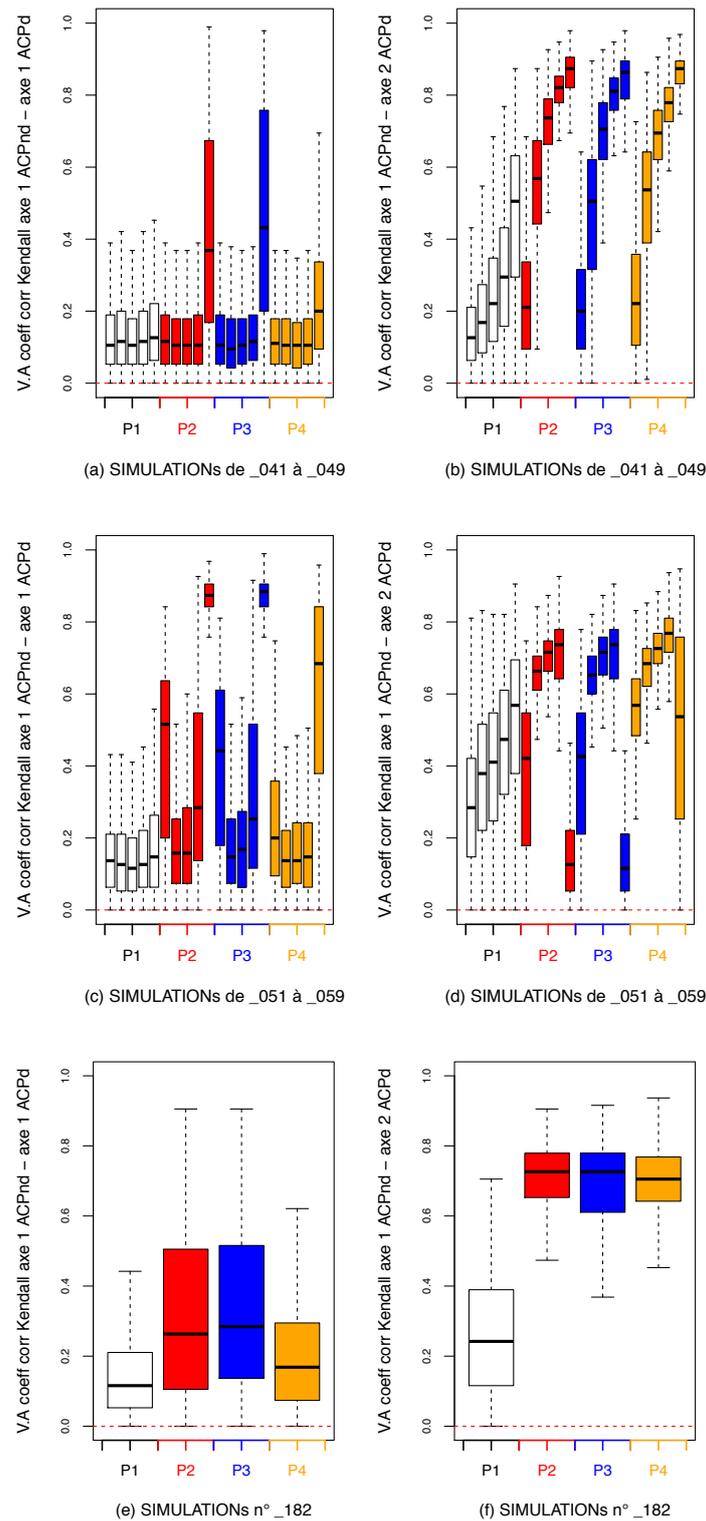
**Conseil de lecture :** les quatre premiers graphiques (a), (b), (c) et (d) correspondent à l'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20) et les quatre autres graphiques (e), (f), (g) et (h) à l'ENSEMBLE 5 (gradients de PdD différents entre espèces). Mise à part cette particularité, le mode de lecture de cette figure, est similaire à celui de la figure précédente.

**Observations :** Les séparations attendues, dans les ACPnd des groupes A et B dans les profils 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) sur les axes 1 (graphique (a) et (e)), et dans les profils 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) sur les axes 2 (graphiques (c) et (g)), sont bien visibles. Les autres SIMULATIONs donnent une répartition des espèces des deux groupes quasi équitable de chaque côté de l'origine de l'axe.

Pour les ACPd (graphique (b), (d), (f) et (h)) ces séparations sont respectivement :

- Perdues sur l'axe 1 (graphiques (b) et (f)) et on retrouve, sur l'ensemble des profils, la tendance à regrouper les espèces du même côté de l'axe.
- Perdue sur l'axe 2 dans l'ENSEMBLE 5 (gradients de PdD différents entre les espèces) (graphique (h)).
- Visibles mais dégradées, pour l'ENSEMBLE 4 (Même gradient de PdD pour 11 espèces sur 20), avec une dégradation d'autant plus forte que  $d_{moy,j}$  diminue (graphique (d)).
- Enfin, dans les deux ENSEMBLES, la séparation des groupes, perdue sur l'axe 1 des ACPd, se retrouve, bien que dégradée sur l'axe 2 de ces mêmes ACPd (graphiques (d) et (h)).

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 19 :** Graphique simple (un boxplot par SIMULATION) des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1 de l'ACPnd, et les ordinations des espèces sur les axes 1 (graphique (a), (c) et (e)) et 2 (graphique (b), (d) et (f)) de l'ACPd ; lors de SIMULATIONS avec un gradient de PdD quadratique. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 6 où dans chacune des SIMULATIONS toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 7 où dans chacune des SIMULATIONS 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1) Les graphiques (e) et (f) illustrent les résultats de l'ENSEMBLE 8 où dans chacune des SIMULATIONS les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartiennent les SIMULATIONS.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Je retiens des résultats concernant un gradient linéaire de PdD plusieurs points :

- (i) Que les ordinations des axes 1 et 2 des ACPnd sont dans tous les cas modifiées, donc dégradées par la PdD. Ces dégradations sont d'autant plus importantes que la PdD est faible
- (ii) Qu'un certain nombre d'espèces changent de côté vis-à-vis de l'origine de l'axe 1 entre ACPnd et ACPd. Ceci provoque l'apparition d'un effet taille lors des ACPd dans les ENSEMBLES 3 et 5 (où dans chaque SIMULATIONs les espèces ont toutes un gradient de PdD, identique entre les espèces pour l'ENSEMBLE 3 et différent pour l'ENSEMBLE 5).
- (iii) Que je retrouve sur l'axe 2 de certaines ACPd l'ordination, bien que dégradée, présente sur l'axe 1 lors des ACPnd correspondantes lorsque cette ordination met en évidence la variable  $g_2$  (profil 4).
- (iv) La mise en évidence d'une relation entre les ordinations sur les axes 2 lors des ACPd de l'ENSEMBLE 5 et valeur moyenne de la PdD des espèces.

Je reviendrai sur ces points dans le chapitre 4.

### 3.2.3 Les SIMULATIONs avec gradient de détection quadratique

Je présente maintenant les SIMULATIONs qui traitent des cas **avec un gradient de détection quadratique**. Je m'appuie sur trois ENSEMBLES de SIMULATIONs. Le premier ENSEMBLE (ENSEMBLE 6) regroupe les SIMULATIONs où, dans chacune d'elles, les espèces ont toutes le même gradient de PdD et dont les trois derniers chiffres sont compris entre 041 et 049. Le deuxième ENSEMBLE (ENSEMBLE 7) regroupe les SIMULATIONs où dans chacune d'elles, 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1) et dont les trois derniers chiffres sont compris entre 051 et 059. Le dernier ENSEMBLE (ENSEMBLE 8) regroupe les SIMULATIONs où dans chacune d'elles les espèces ont des gradients de PdD différents et dont les derniers chiffres sont égaux à 182.

On retrouve les indicateurs dans le même ordre que pour le gradient de détection linéaire.

La figure 19 (ci-contre) représente les valeurs absolues des **coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1** des ACPnd et les ordinations sur les axes 1 et 2 des ACPd.

**Conseil de lecture** : Dans cette figure on superpose les résultats des trois ENSEMBLES. L'ENSEMBLE 6 sur la première ligne (graphique (a) et (b)), l'ENSEMBLE 7 sur la deuxième (graphique (c) et (d)), et l'ENSEMBLE 8 sur la troisième (graphique (e) et (f)). Les graphiques de la colonne gauche illustrent les coefficients de corrélation de Kendall entre les ordonnées sur l'axe 1 des ACPnd et celles de l'axe 1 des ACPd, et la colonne de droite ceux entre les ordonnées de l'axe 1 des ACPnd et celles de l'axe 2 des ACPd.

**Observations** : Chaque ENSEMBLE donne des résultats différents.

Les corrélations, pour l'ENSEMBLE 6 (même gradient de PdD entre les espèces), sur l'axe 1 des ACPd (graphique (a)), sont faibles et peu variables dans tout le profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ) et dans les profils 2, 3 et 4 (où l'abondance dépend au moins d'une des deux variables d'habitat ( $g_1$  ou  $g_2$ )), tant que  $d_{max,j} \leq 0.7$ .

Sur l'axe 2 des ACPd (graphique (b)), les 4 profils donnent des corrélations qui augmentent avec  $d_{max,j}$  ; augmentation qui fait passer la corrélation de faible à moyenne dans le profil 1 et de faible à forte dans les profils 2, 3 et 4.

Pour l'ENSEMBLE 7 (même gradient de PdD pour 11 espèces sur 20). Sur l'axe 1 (graphique (c)) :

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

- Dans le profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ), les corrélations sont faibles et peu variables.
- Pour les profils 2, 3 et 4 (où l'abondance dépend au moins d'une des deux variables d'habitat ( $g_1$  ou  $g_2$ )) les corrélations passent de moyennes lorsque  $d_{max,j}$  égale 0.1, à faibles pour  $d_{max,j}$  comprise entre 0.3 et 0.7, à fortes lorsque  $d_{max,j}$  égale à 0.9 ; décrivant ainsi une forme parabolique concave.

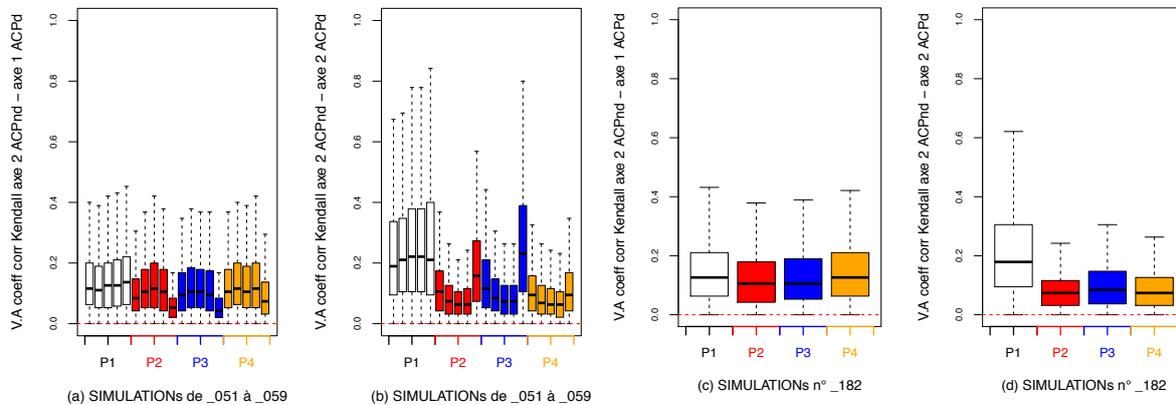
Sur l'axe 2 (graphique (d)), le profil 1 affiche une corrélation moyenne mais croissante avec  $d_{max,j}$ . Et les profils 2, 3 et 4, présentent une évolution symétrique et donc convexe à ceux de l'axe 1 ; avec les coefficients de corrélation s'ordonnant en fonction du profil (toute chose étant égale par ailleurs) comme suit :  $P4 > P2 \geq P3$ .

L'ENSEMBLE 8 (gradient de PdD différents entre espèces) présente sur l'axe 1 (graphique (e)) des corrélations globalement faibles pour les profils 1 et 4 (où les abondances moyennes ne dépendent pas de  $g_1$ ), et fortement variables allant de faible à moyennes pour les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ).

Les résultats sur l'axe 2 (graphique (f)) présentent, pour les profils 2, 3 et 4, des corrélations fortes (à la limite de moyennes), moins variables et supérieures à celles du profil 1.

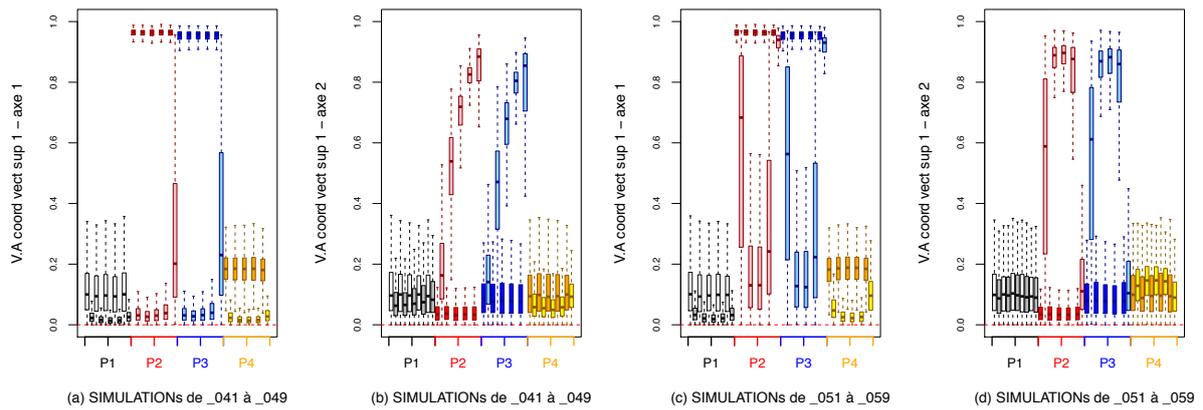
La figure 20, des valeurs absolues des **coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 2** des ACPnd et les ordinations sur les axes 1 et 2 des ACPd reprend les résultats des ENSEMBLES 7 (même gradient de PdD pour 11 espèces sur 20) et 8 (gradient de PdD différents entre espèces).

**Observations :** On retient de cet indicateur, que dans tous les ENSEMBLES de SIMULATIONs les corrélations obtenues sont globalement faibles et peu variables pour tous les profils

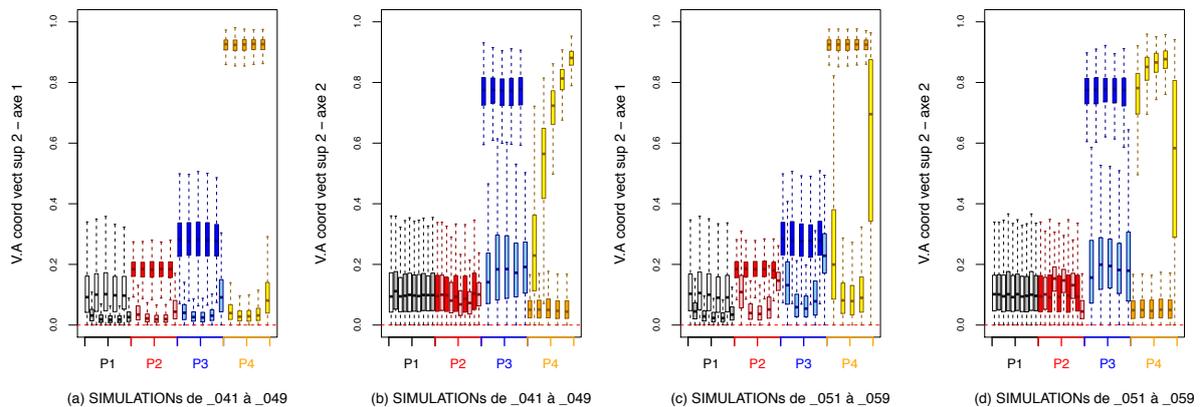


**Figure 20 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 2 de l'ACPnd, et les ordinations des espèces sur l'axe 1 (graphiques (a) et (c)) et l'axe 2 (graphiques (b) et (d)) de l'ACPd ; lors de SIMULATIONs avec un gradient de PdD quadratique. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 7 où dans chacune des SIMULATIONs 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 8 où dans chacune des SIMULATIONs les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 21 :** graphique double (deux boxplots par SIMULATION, un pour l'ACPnd et un pour l'ACPd) des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_1$ , sur les axes 1 (graphique (a) et (c)) et 2 (graphique (b) et (d)) des ACPnd et des ACPd; lors de SIMULATIONS avec un gradient de PdD quadratique. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 6 où dans chacune des SIMULATIONS toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 7 où dans chacune des SIMULATIONS 11 espèces sur 20 ont un même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.



**Figure 22:** graphique double (deux boxplots par SIMULATION, un pour l'ACPnd et un pour l'ACPd) des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_2$ , sur les axes 1 (graphique (a) et (c)) et 2 (graphique (b) et (d)) des ACPnd et des ACPd; lors de SIMULATIONS avec un gradient de PdD quadratique. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 6 où dans chacune des SIMULATIONS toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 7 où dans chacune des SIMULATIONS 11 espèces sur 20 ont un même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

La figure 21 (ci-contre), **des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_1$**  regroupe les résultats des ENSEMBLES 6 (même gradient de PdD entre les espèces) et 7 (même gradient de PdD pour 11 espèces sur 20).

**Observations :** les ACPnd affichent une forte et constante corrélation du vecteur supplémentaire avec l'axe1 (graphique (a) et (c)) dans les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ). Dans les autres cas (aussi bien sur l'axe 1 que sur l'axe 2) les corrélations sont faibles.

Les résultats des ACPd montrent d'importantes différences avec ceux des ACPnd dans les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ).

Là encore les résultats des deux ENSEMBLES diffèrent. Cependant, les descriptions faites de la figure ACP-GQ-K1, pour les ENSEMBLES 6 et 7, sont valables mots pour mots pour décrire les corrélations des profils 2 et 3 dans les ACPd.

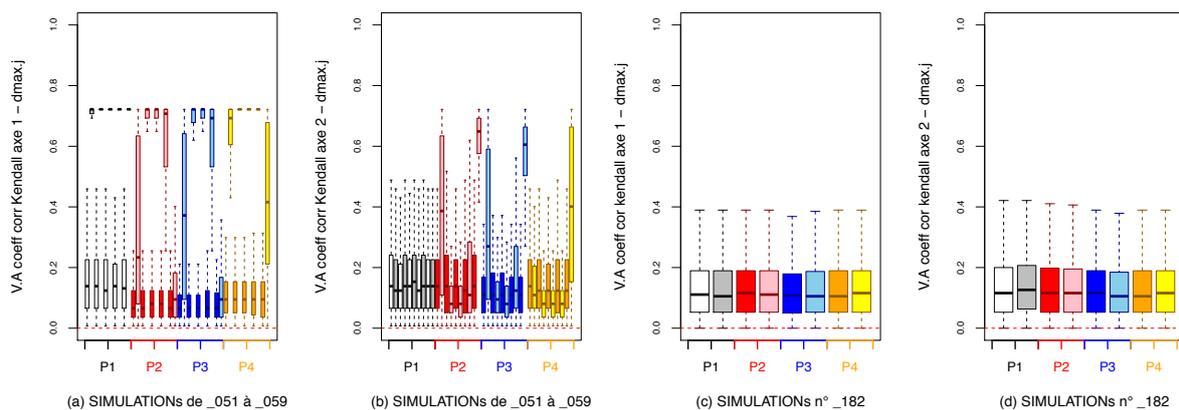
La figure 22 (ci-contre), **des valeurs absolues des coordonnées du vecteur supplémentaire représentant  $g_2$**  regroupe les résultats des ENSEMBLES 6 (même gradient de PdD entre les espèces) et 7 (même gradient de PdD pour 11 espèces sur 20).

**Observations :** Les résultats des ACPnd, sont, là encore, sans surprise : une forte corrélation entre le vecteur supplémentaire et l'axe 1 (graphique (a) et (c)) dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ), et avec l'axe 2 (graphique (b) et (d)) dans le profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ).

Pour les ACPd :

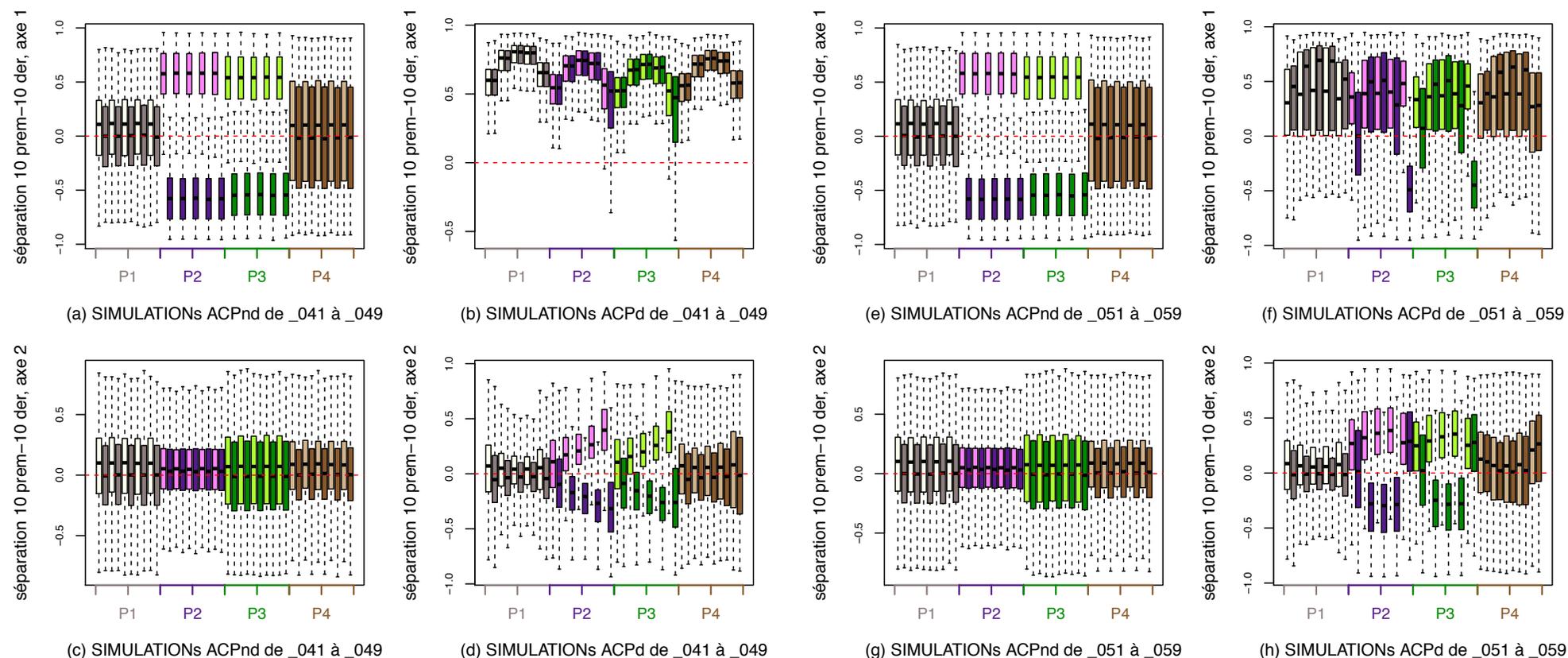
- D'une part, les corrélations dans le profil 3, sont toujours faibles et bien inférieures à celles observées dans les ACPnd en particulier sur les axes 2.
- D'autre part, les descriptions faites de la figure ACP-GQ-K1, pour les ENSEMBLES 6 et 7, sont valables mot pour mot pour décrire les corrélations du profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) dans les ACPd.

La figure 23, présente les **valeurs absolues des coefficients de corrélation de Kendall entre les valeurs  $d_{moy,j}$**  et l'ordination des espèces sur l'axe 1 et l'axe 2 des analyses. On y retrouve les résultats des ENSEMBLES 7 (même gradient de PdD pour 11 espèces sur 20) et 8 (gradient de PdD différents entre espèces).



**Figure 23 :** graphique double (deux boîtes par SIMULATION, un pour l'ACPnd et un pour l'ACPd) des valeurs absolues des coefficients de corrélation de Kendall entre les valeurs  $d_{max,j}$  et l'ordination des espèces sur les axes 1 (graphique (a) et (c)) et 2 (graphique (b) et (d)) des analyses ; lors de SIMULATIONS avec un gradient de PdD quadratique. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 7 où dans chacune des SIMULATIONS 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 8 où dans chacune des SIMULATIONS les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 24 :** illustrant les « séparations » des groupes d'espèces 1 (les 10 premières espèces) et 2 (les 10 dernières espèces), lors de SIMULATIONS avec un gradient de PdD quadratique Elle présente deux boxplots par SIMULATION, un pour chaque groupe d'espèces. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE 6 où dans chacune des SIMULATIONS toutes les espèces ont le même gradient de PdD. Les quatre derniers graphiques (e), (f), (g) et (h) illustrent les résultats de l'ENSEMBLE 7 où dans chacune des SIMULATIONS 11 espèces sur 20 ont un même gradient de PdD, les 9 autres ayant une PdD constamment égale à 1. La première et la troisième colonne de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des ACPnd, et la deuxième et la quatrième colonne de graphiques ((b), (d) et (f), (h)) lors des ACPd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

**Observations :** L'ENSEMBLE 8 (gradients de PdD différents entre espèces), (les graphiques (c) et (d)) ne présente que de faibles corrélations pour toutes les analyses. En revanche pour l'ENSEMBLE 7 (même gradient de PdD pour 11 espèces sur 20), seules les corrélations des ACPnd restent constamment faibles sur les deux axes. Les corrélations des ACPd sur l'axe 1 (graphique (a)) leur sont toujours supérieures. En particulier pour le profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ) et pour les valeurs de  $d_{max.j}$  comprises entre 0.3 et 0.7 pour les profils 2, 3 et 4 (où l'abondance dépend au moins d'une des deux variables d'habitat ( $g_1$  ou  $g_2$ )), les corrélations sont fortes.

Sur l'axe 2 (graphique (b)) des ACPd, quelques simulations ne présentent pas de faibles corrélations. On a des corrélations relativement fortes pour les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ) lorsque  $d_{max.j}$  égale 0.9, et des corrélations moyennes mais surtout très variables quand  $d_{max.j}$  égale 0.1 dans les profils 2 et 3, et quand  $d_{max.j}$  égale 0.9 dans le profil 4.

La figure 24 (ci-contre), des séparations **des groupes d'espèces 1 et 2** regroupe les résultats des ENSEMBLES 6 (même gradient de PdD entre les espèces) et 7 (même gradient de PdD pour 11 espèces sur 20)..

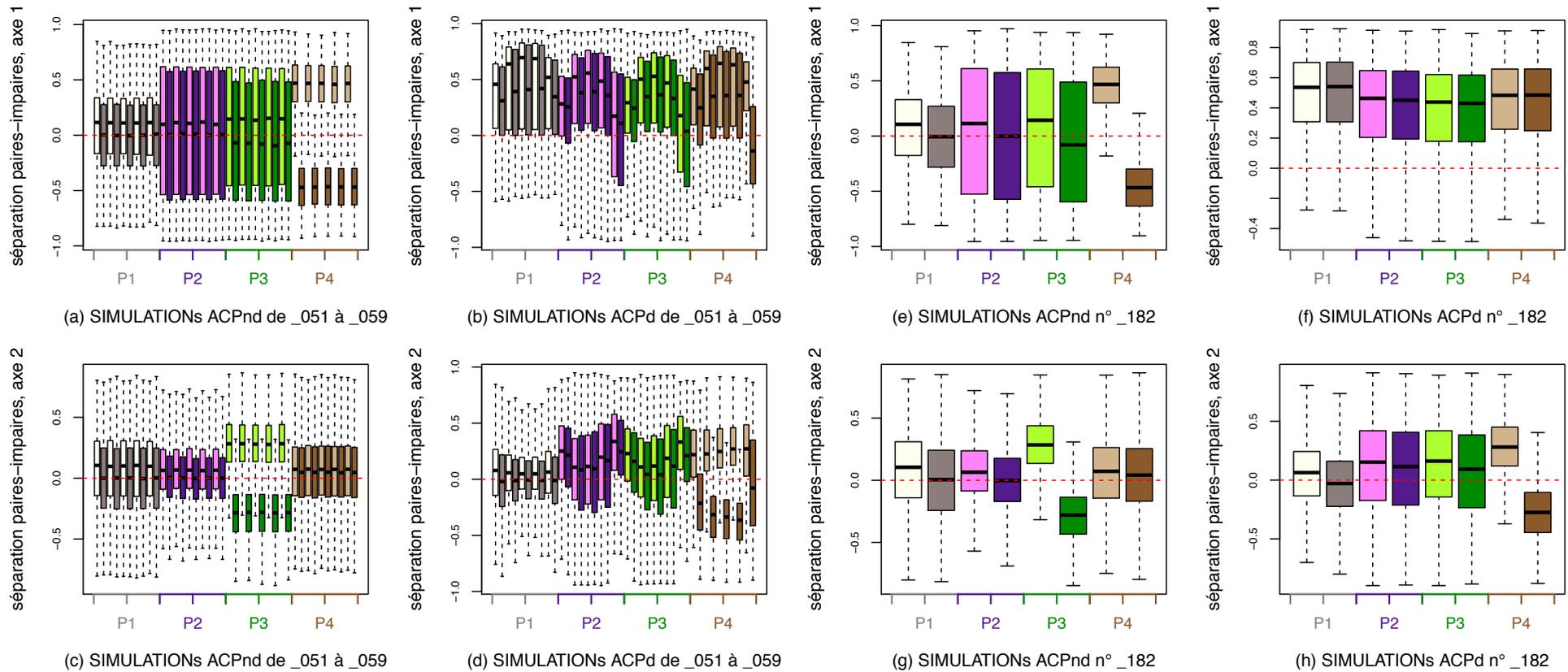
**Observations :** Dans les cas des ACPnd, on retrouve bien les séparations attendues des groupes, et seulement celles attendues.

Cependant, sur les axes 1 de l'ENSEMBLE 7 (même gradient de PdD pour 11 espèces sur 20) des ACPd (graphique (f)), si  $d_{max.j}$  est inférieure à 0.9, on ne retrouve pas les séparations des groupes 1 et 2.

Les ACPd (graphiques (b) et (f)) ont également une forte tendance à mettre toutes les espèces du même côté de l'axe 1, et ce, sur tous les profils. Aussi cet effet augmente jusqu'à ce que  $d_{max.j} = 0.5$ , puis diminue.

Pour les deux ENSEMBLES de simulations, on retrouve parfois sur l'axe 2 des ACPd (graphiques (d) et (h)), les séparations des groupes 1 et 2 présentes sur l'axe 1 des ACPnd. Pour l'ENSEMBLE 6 on les retrouve, dans les profils 2 et 3 (où l'abondance moyenne des espèces dépend au moins de  $g_1$ ) lorsque  $d_{max.j} \geq 0.3$ , et cette séparation est d'autant moins dégradée que  $d_{max.j}$  augmente. Pour l'ENSEMBLE 7 on les retrouve, dans les profils 2 et 3 lorsque  $d_{max.j}$  est supérieures à 0.1 et inférieur à 0.9.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 25 :** illustrant les « séparations » des groupes d'espèces A (les espèces paires) et B (espèces impaires), lors de SIMULATIONS avec un gradient de PdD quadratique. Elle présente deux boxplots par SIMULATION, un pour chaque groupe d'espèces. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE 7 où dans chacune des SIMULATIONS 11 espèces sur 20 ont un même gradient de PdD, les 9 autres ayant une PdD constamment égale à 1. Les quatre derniers graphiques (e), (f), (g) et (h) illustrent les résultats de l'ENSEMBLE 8 où dans chacune des SIMULATIONS toutes les espèces ont des gradients de PdD différents. La première et la troisième colonne de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des ACPnd, et la deuxième et la quatrième colonne de graphiques ((b), (d) et (f), (h)) lors des ACPd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

Enfin la figure 25 (ci-contre), présente les séparations **des groupes d'espèces A et B** et regroupe les ENSEMBLES 7 (même gradient de PdD pour 11 espèces sur 20) et 8 (gradients de PdD différents entre espèces).

**Observations** : Les ACPnd permettent toujours la séparation des groupes A et B, sur l'axe 1 dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ), et sur l'axe 2 dans le profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ).

Là encore ces séparations sont détériorées lors des ACPd (graphique (b), (d), (f) et (h)). D'une part on retrouve dans les deux ENSEMBLES de SIMULATIONs, l'effet de regroupement de toutes les espèces du même côté de l'axe 1 (graphique (b), (d)). Ceci a pour conséquence, en plus du faux regroupement, la perte de la séparation des groupes A et B dans le profil 4. D'autre part, sur les axes 2 des ACPd (graphique (f), (h)), les profils 2 et 3 montrent également une tendance, parfois légère, au regroupement des espèces du même côté de l'axe ; tendance plus forte lorsque que  $d_{max,j} = 0.1$  et  $d_{max,j} = 0.9$ . On a donc une perte de la séparation des groupes A et B dans le profil 3.

Enfin, pour les deux ENSEMBLES, on retrouve dans le profil 4 sur l'axe 2 des ACPd et bien que dégradée, la séparation des groupes A et B présente dans le profil 4 sur l'axe 1 lors des ACPnd. Pour l'ENSEMBLE 7 on note que la séparation est clairement visible lorsque  $0.1 < d_{max,j} < 0.9$ .

Je retiens des résultats concernant un gradient linéaire de PdD plusieurs points :

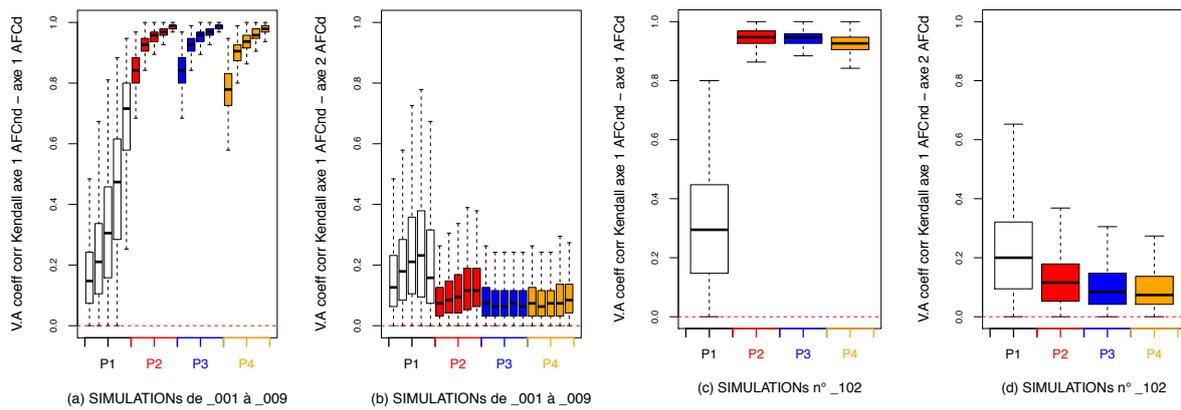
- (i) Que les ordinations des axes 1 et 2 des ACPnd, sont dans tous les cas modifiées, donc dégradées par la PdD. Ces dégradations sont d'autant plus importantes que la PdD est faible et que l'ordination se trouve sur l'axe 2.
- (ii) Le changement de côté vis-à-vis de l'origine de l'axe 1 d'un certain nombre d'espèces entre ACPnd et ACPd.
- (iii) Le fait de retrouver sur l'axe 2 de certaines ACPd l'ordination, bien que dégradée, présente sur l'axe 1 lors des ACPnd correspondantes (que cette ordination mette en évidence la variable  $g_1$  ou  $g_2$ ). Ceci va de paire avec la perte dans le premier plan de l'ACPd de l'ordination présente sur l'axe 2 de l'ACPnd correspondantes.

Je reviendrai sur ces points dans le chapitre 4.

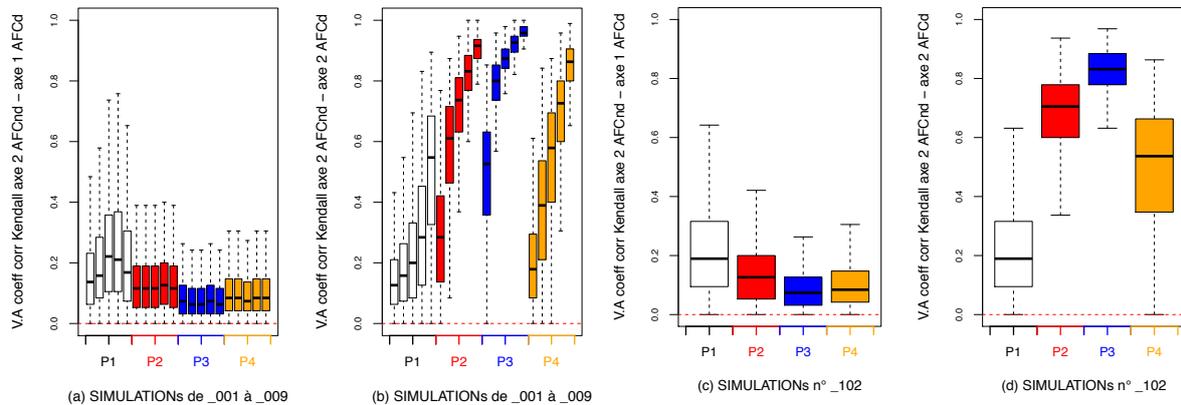
### 3.3 Les résultats des AFC

J'aurai la même conduite pour présenter les résultats concernant les AFC que pour la présentation des résultats des ACP. C'est-à-dire que je les présente en fonction du gradient de détection que j'impose lors des SIMULATIONs.

Aussi, dans cette partie lorsque des résultats de SIMULATIONs ne sont pas présentés, ceci indique qu'ils n'apportent pas d'informations supplémentaires qui aideraient à répondre à mes questions.



**Figure 26 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 1 de l'AFCnd, et les ordinations des espèces sur l'axe 1 (graphiques (a) et (c)) et l'axe 2 (graphiques (b) et (d)) de l'AFCd ; lors de SIMULATIONs sans gradient de PdD. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 1 où dans chacune des SIMULATIONs toutes les espèces ont la même PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 2 où dans chacune des SIMULATIONs les espèces ont des PdD différentes. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.



**Figure 27 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues du taux de Kendall entre l'ordination des espèces sur l'axe 2 de l'AFCnd, et les ordinations des espèces sur l'axe 1 (graphiques (a) et (c)) et l'axe 2 (graphiques (b) et (d)) de l'AFCd ; lors de SIMULATIONs sans gradient de PdD. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 1 où dans chacune des SIMULATIONs toutes les espèces ont la même PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 2 où dans chacune des SIMULATIONs les espèces ont des PdD différentes. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

### 3.3.1 Les SIMULATIONs sans gradient de détection

Commençons par les SIMULATIONs **sans gradient de détection**. Nous avons deux ENSEMBLES de SIMULATIONs. Le premier ENSEMBLE (ENSEMBLE 1) regroupe les SIMULATIONs où dans chacune d'elles, les espèces ont toutes la même PdD et dont les trois derniers chiffres sont compris entre 001 et 009. Le deuxième ENSEMBLE (ENSEMBLE 2) regroupe les SIMULATIONs où dans chacune d'elles, les espèces ont des PdD différentes et dont les derniers chiffres sont égaux à 102.

La figure 26 (ci-contre), présente les résultats des **valeurs absolues des coefficients de corrélations de Kendall entre les ordinations des espèces sur l'axe1 des AFCnd** et les ordinations sur les axes 1 et 2 des AFCd.

**Conseil de lecture** : La figure se lit comme celle du même indicateur pour les ACP.

**Observations** : Sur l'axe 1 de chaque ENSEMBLE (graphique (a) et (c)), les corrélations sont fortes et peu variables dans les profils 2, 3 et 4 (où l'abondance dépend au moins d'une des deux variables d'habitat ( $g_1$  ou  $g_2$ )), et celles du profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ) vont de faibles à fortes et sont toujours inférieures à celles des autres profils.

Les résultats de l'ENSEMBLE 1 (graphique (a)) montrent une corrélation systématiquement croissante avec l'augmentation de  $d_{moy.j}$ .

L'axe 2 de chaque ENSEMBLE (graphiques (b) et (d)) quant à lui ne montre que des corrélations faibles.

La figure 27 (ci-contre) illustre des résultats des **valeurs absolues des coefficients de corrélations de Kendall entre les ordinations des espèces sur l'axe2 des AFCnd** et les ordinations sur les axes 1 et 2 des AFCd.

**Observations** : Sur les axes 1 des deux ENSEMBLES de SIMULATIONs (graphiques (a) et (c)), on ne trouve que des corrélations faibles.

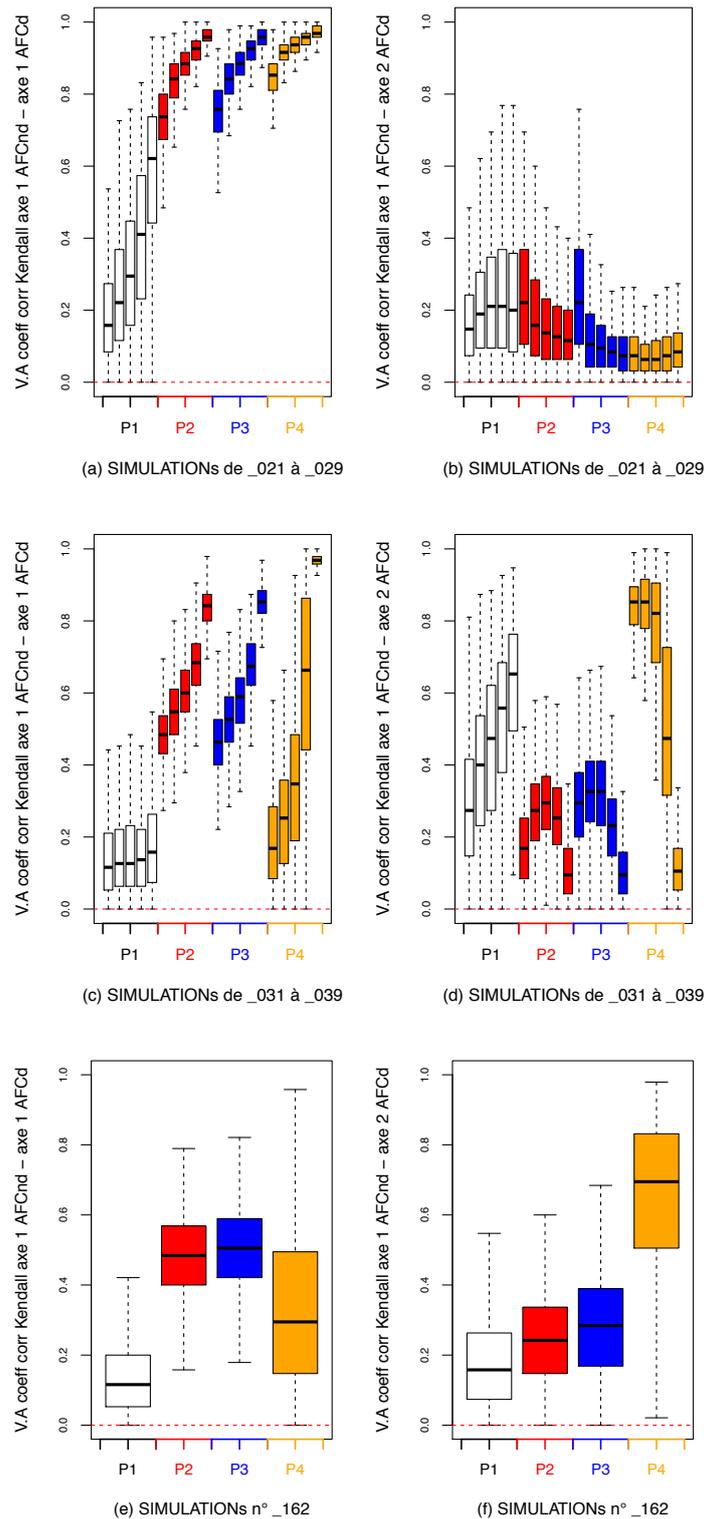
Sur les axes 2 (graphiques (b) et (d)), on remarque que le profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) présente les corrélations les plus élevées (toute chose étant égale par ailleurs).

Dans le cas de l'ENSEMBLE 1 (même PdD entre les espèces) (graphique (b)), les corrélations de chaque profil sont croissantes avec  $d_{moy.j}$ . Elles passent de moyennes à fortes pour les profils 2, 3 et 4, et de faibles à moyennes pour le profil 1.

Pour l'ENSEMBLE 2 (PdD différents entre espèces), les corrélations sont faibles pour les profils 1, fortes pour les profils 2 et 3, et moyennes pour le profil 4.

Pour le cas des SIMULATIONs sans gradient de détection, les autres indicateurs ne sont pas présentés. En effet, à l'image de la figure 26 (page 57), ces derniers ne mettent en évidence que de faibles différences entre AFCnd et AFCd. Je retiens tout de même que les ordinations des axes 1 et 2 des AFCnd, sont dans tous les cas modifiées, donc dégradées par la PdD. Que si ces dégradations sont faibles, elles sont néanmoins d'autant plus importantes que la PdD est faible, et plus importantes sur les axes 2 que sur les axes 1.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 28 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1 de l'AFCnd, et les ordinations des espèces sur les axes 1 (graphique (a), (c) et (e)) et 2 (graphique (b), (d) et (f)) de l'AFCd ; lors de SIMULATIONs avec un gradient de PdD linéaire. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 3 où dans chacune des SIMULATIONs toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 4 où dans chacune des SIMULATIONs 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (e) et (f) illustrent les résultats de l'ENSEMBLE 5 où dans chacune des SIMULATIONs les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

### 3.3.2 Les SIMULATIONs avec gradient de détection linéaires

Pour la présentation des résultats des SIMULATIONs avec gradient de détection linéaire, j'ai trois ENSEMBLEs de SIMULATIONs. Le premier ENSEMBLE (ENSEMBLE 3) regroupe les SIMULATIONs où dans chacune d'elles, les espèces ont toutes le même gradient de PdD et dont les trois derniers chiffres sont compris entre 021 et 029. Le deuxième ENSEMBLE (ENSEMBLE 4) regroupe les SIMULATIONs où dans chacune d'elles, 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1) et dont les trois derniers chiffres sont compris entre 031 et 039. Le dernier ENSEMBLE (ENSEMBLE 5) regroupe les SIMULATIONs où dans chacune d'elles, les espèces ont des gradients de PdD différents et dont les derniers chiffres sont égaux à 162.

Je commence par **valeurs absolues des coefficients de corrélations de Kendall entre les ordinations des espèces sur l'axe1 des AFCnd** et celles sur les axes 1 et 2 des AFCd. La figure 28 (ci-contre), présente les résultats des trois ENSEMBLEs de SIMULATIONs.

**Observations :** Chaque ENSEMBLE (un par ligne de graphique), donne des résultats différents.

L'ENSEMBLE 3 (même gradient de PdD entre les espèces) (graphique (a) et (b)), montre sur l'axe 1 des AFCd (graphique (a)) que pour tous les profils, les corrélations sont croissantes avec l'augmentation de  $d_{moy.j}$ . On a cependant de fortes corrélations dans les profils 2, 3 et 4 (où l'abondance dépend au moins d'une des deux variables d'habitat ( $g_1$  ou  $g_2$ )) quelle que soit la valeur de  $d_{moy.j}$ , et des corrélations passant de faibles à moyennes pour le profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ). Sur l'axe 2 des AFCd (graphique (b)), les corrélations sont toutes globalement faibles.

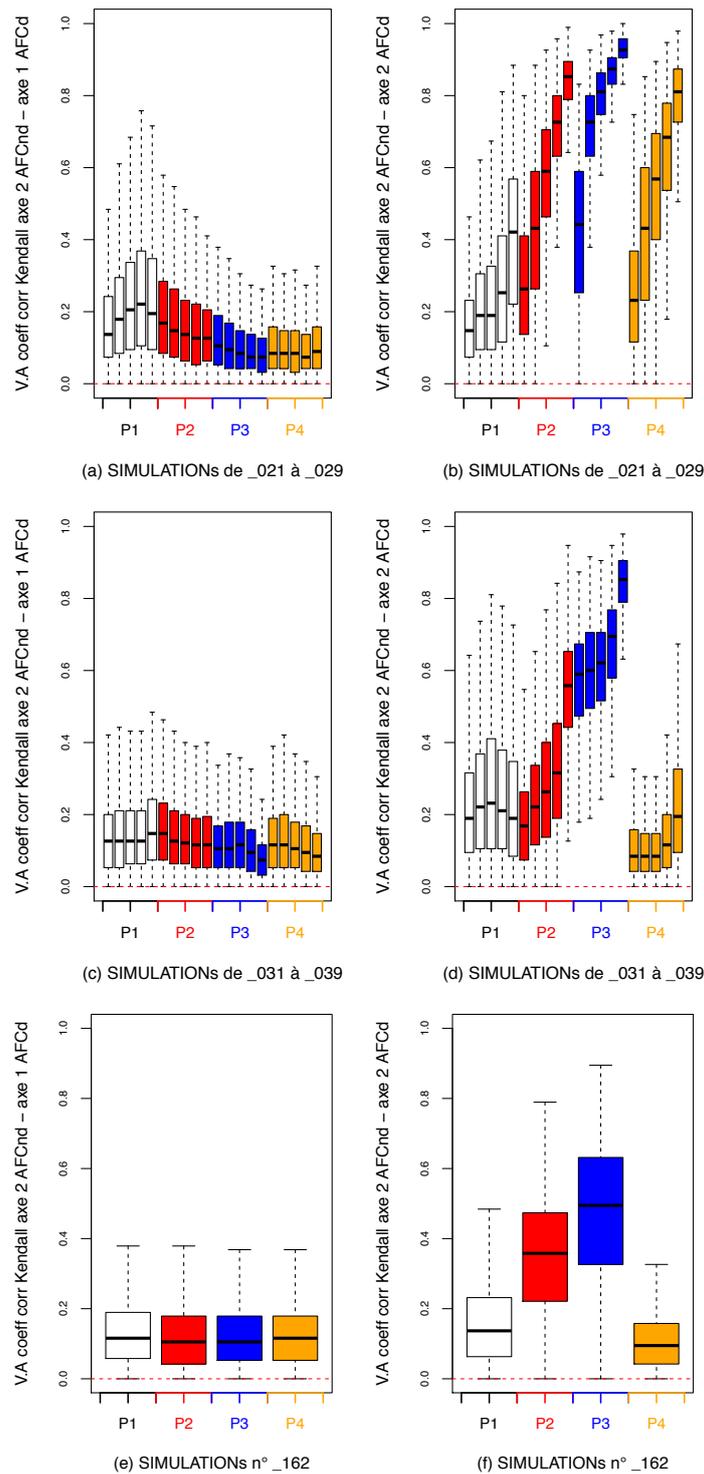
L'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20), montre les mêmes résultats pour les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ), aussi bien sur l'axe 1 que sur l'axe 2. L'axe 1, montre également des corrélations croissantes avec l'augmentation de  $d_{moy.j}$  dans tous les profils.

Cependant, cette corrélation reste faible pour le profil 1, passe de moyenne à forte pour les profils 2 et 3, et de faible à forte pour le profil 4.

Sur l'axe 2 (graphique (d)), le profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ) montre une corrélation croissante et moyenne, les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ) des corrélations faibles ; et le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) des corrélations décroissantes passant de fortes à faibles en fonction de  $d_{moy.j}$ .

L'ENSEMBLE 5 (gradient de PdD différents entre espèces), n'affiche au mieux, sur l'axe 1 (graphique (e)), que des corrélations moyennes. Sur l'axe 2 (graphique (f)), ces corrélations sont toutes faibles, dans les profils 1, 2 et 3, et évoluent de moyennes à fortes dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ).

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 29 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 2 de l'AFCnd, et les ordinations des espèces sur les axes 1 (graphique (a), (c) et (e)) et 2 (graphique (b), (d) et (f)) de l'AFCd ; lors de SIMULATIONs avec un gradient de PdD linéaire. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 3 où dans chacune des SIMULATIONs toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 4 où dans chacune des SIMULATIONs 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (e) et (f) illustrent les résultats de l'ENSEMBLE 5 où dans chacune des SIMULATIONs les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

La figure 29 (ci-contre), illustre les **valeurs absolues des coefficients de corrélations de Kendall entre les ordinations des espèces sur l'axe 2 des AFCnd** et celles des axes 1 et 2 des AFCd.

**Observations :** Les graphiques de la première colonne ((a), (c) et (e)), (axe 1 des AFCd) affichent pour les trois ENSEMBLES, des corrélations faibles dans tous les cas de figures.

Les graphiques de la deuxième colonne ((b), (d) et (f)) (axes 2 des AFCd) ont une allure comparable.

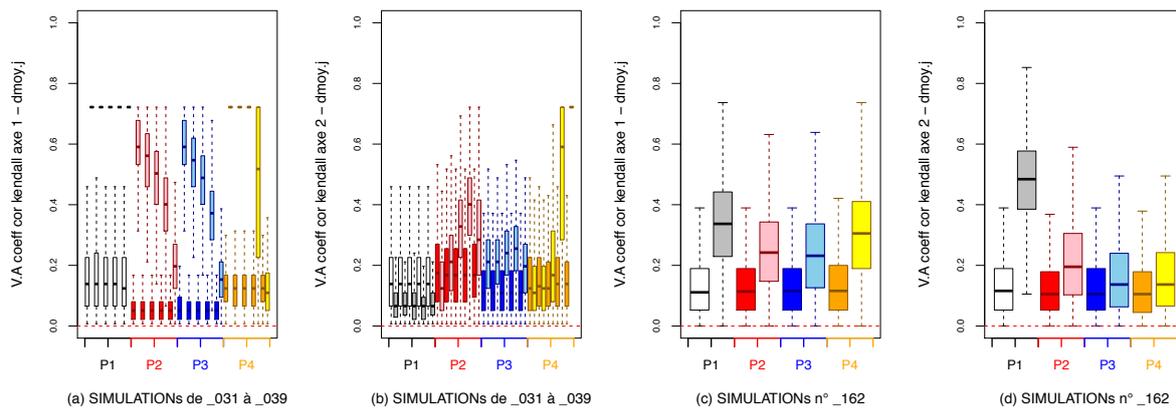
Les profils 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ) ont des corrélations faibles, et légèrement croissantes dans le cas de l'ENSEMBLE 3.

Les profils 2, (où les abondances moyennes ne dépendent que de  $g_1$ ) ont des corrélations plus importantes et croissantes ; de faibles à fortes pour l'ENSEMBLE 3, de faibles à moyennes pour l'ENSEMBLE 4 et faible pour l'ENSEMBLE 5.

Idem pour le profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) avec des valeurs encore plus fortes ; de moyennes à fortes pour l'ENSEMBLE 3, idem mais moins variables pour l'ENSEMBLE 4 et moyennes pour l'ENSEMBLE 5.

Enfin, le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) montre des corrélations plus faibles que le profil 1, sauf pour l'ENSEMBLE 3, où ces valeurs sont similaires à celles du profil 2.

Les **valeurs absolues des coefficients de corrélation de Kendall entre les valeurs  $d_{moy,j}$  et l'ordination des espèces sur l'axe 1 et l'axe 2 des analyses**, sont présentées dans la figure 30.

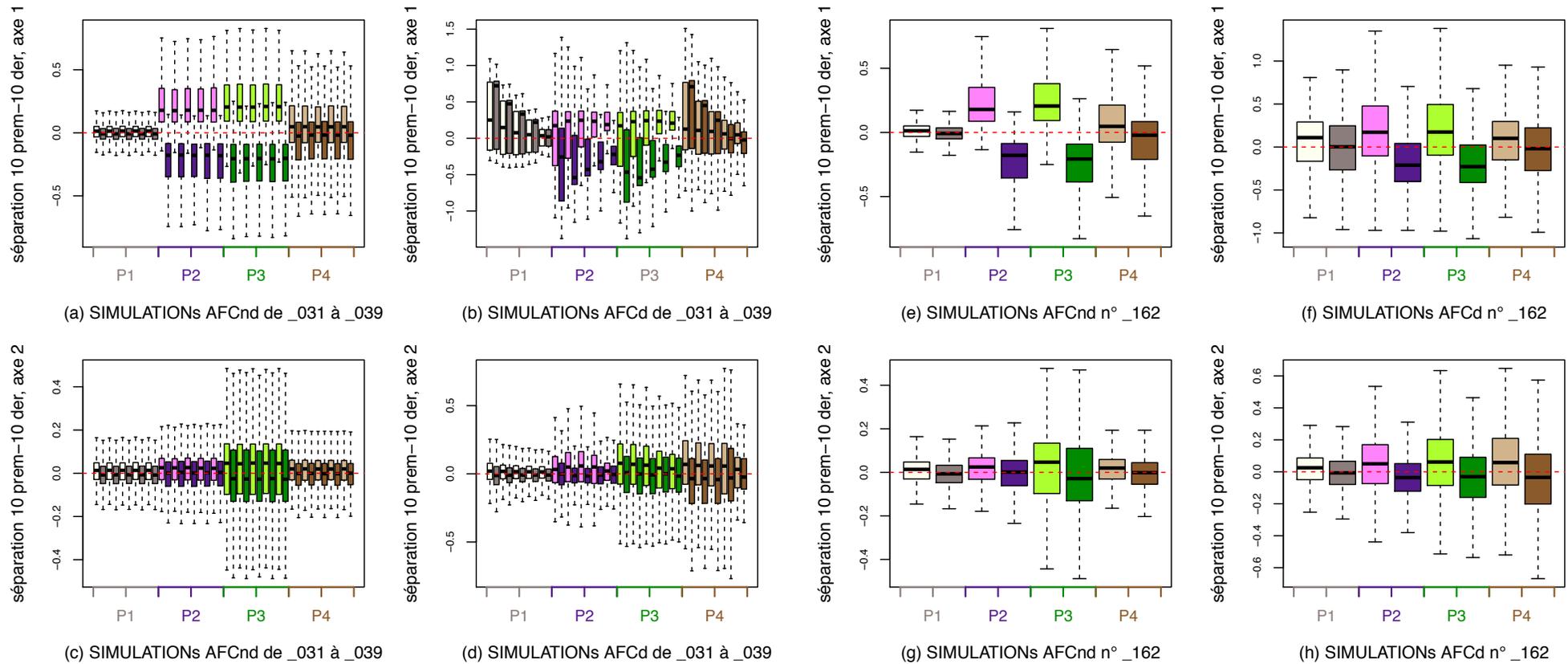


**Figure 30 :** graphique double (deux boxplots par SIMULATION, un pour l'AFCnd et un pour l'AFCd) des valeurs absolues des coefficients de corrélation de Kendall entre les valeurs  $d_{moy,j}$  et l'ordination des espèces sur les axes 1 (graphique (a) et (c)) et 2 (graphique (b) et (d)) des analyses ; lors de SIMULATIONs avec un gradient de PdD linéaire. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 4 où dans chacune des SIMULATIONs 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 5 où dans chacune des SIMULATIONs les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

**Observations :** On a, pour tous les graphiques et toutes les SIMULATIONs, des corrélations faibles dans le cas des AFCnd. En revanche les corrélations obtenues lors des AFCd sont toutes autres et on note principalement :

- Pour l'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20), sur l'axe 1 (graphique (a)), le profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ) affiche des corrélations fortes et stables. Les profils 2, 3 et 4 (où l'abondance dépend au moins d'une des deux variables d'habitat ( $g_1$  ou  $g_2$ )) présentent, eux, des corrélations décroissantes allant de fortes à faible avec l'augmentation de  $d_{moy,j}$ .

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 31 :** illustrant les « séparations » des groupes d'espèces 1 (les 10 premières espèces) et 2 (les 10 dernières espèces), lors de SIMULATIONS avec un gradient de PdD linéaire. Elle présente deux boxplots par SIMULATION, un pour chaque groupe d'espèces. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE 4 où dans chacune des SIMULATIONS 11 espèces sur 20 ont un même gradient de PdD, les 9 autres ayant une PdD constamment égale à 1. Les quatre derniers graphiques (e), (f), (g) et (h) illustrent les résultats de l'ENSEMBLE 5 où dans chacune des SIMULATIONS toutes les espèces ont des gradients de PdD différents. La première et la troisième colonnes de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des AFCnd, et la deuxième et la quatrième colonnes de graphiques ((b), (d) et (f), (h)) lors des AFCd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

- La chute des valeurs des corrélations dans les deux dernières SIMULATIONS du profil 4 sur l'axe1, voit symétriquement une augmentation des corrélations lors des mêmes SIMULATIONS sur l'axe 2
- Pour l'ENSEMBLE 5 (gradient de PdD différents entre espèces), les corrélations sont faibles mais clairement supérieures à celles obtenues lors des AFCnd, dans tous les profils sur l'axe 1 et sur le profil 1 sur l'axe2.

La figure 31 (ci-contre) illustre les séparations **des groupes d'espèces 1 et 2** sur les axes des AFCnd et des AFCd pour les ENSEMBLES de SIMULATIONS 4 ((même gradient de PdD pour 11 espèces sur 20) et 5 (gradient de PdD différents entre espèces). En effet pour cet indicateur il n'est pas nécessaire de présenter les résultats de l'ENSEMBLE 3 (ces derniers sont consultables dans l'annexe B) pour répondre aux questions posées.

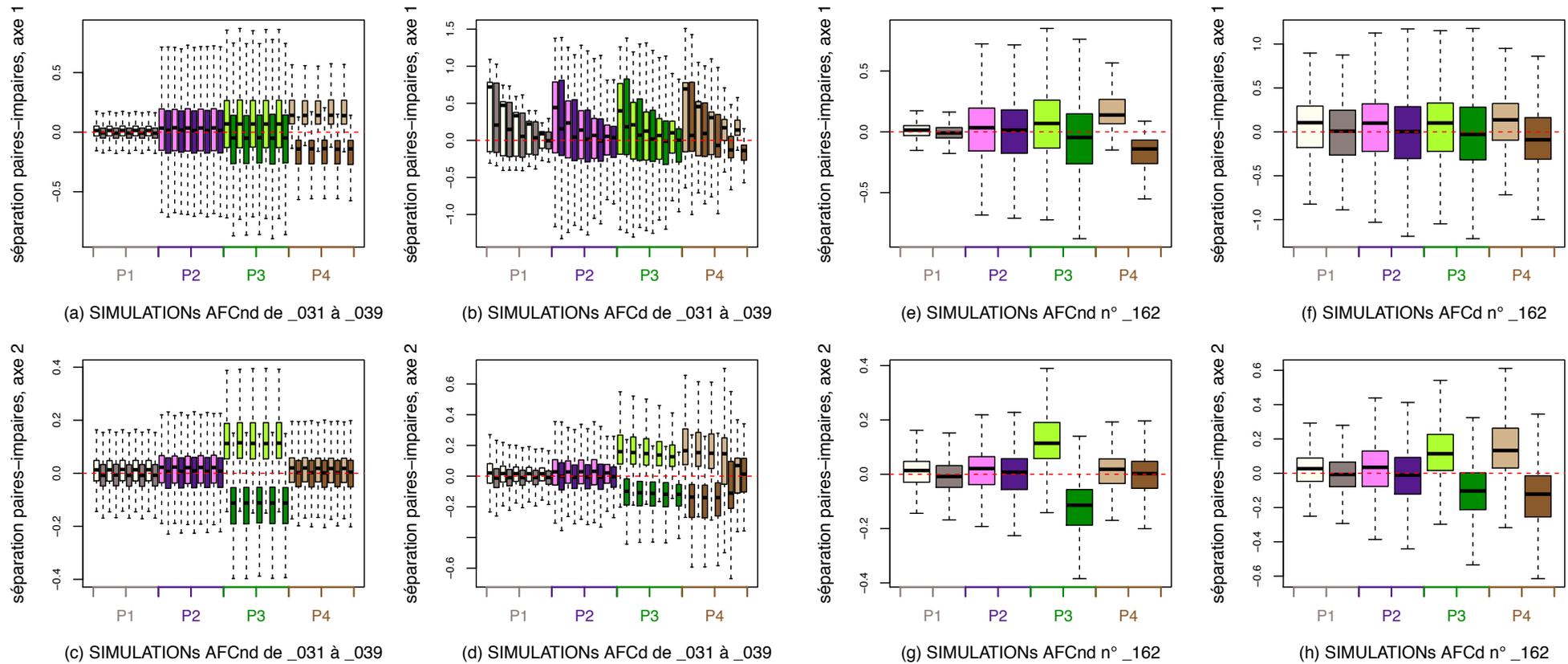
**Conseil de lecture** : les quatre premiers graphiques (a), (b), (c) et (d) correspondent à l'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20) et les quatre autres graphiques (e), (f), (g) et (h) à l'ENSEMBLE 5 (gradients de PdD différents entre espèces). Le mode de lecture de cette figure, est similaire à celui de la figure 3.11.

**Observations** : Dans les deux ENSEMBLES de SIMULATIONS nous retrouvons les séparations attendues lors des AFCnd, c'est-à-dire sur l'axe 1 (graphique (a) et (e)), dans les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ).

Je regarde les résultats des AFCd, dans chaque ENSEMBLE de SIMULATIONS:

- Dans l'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20), sur l'axe 1 (graphique (b)) on observe une dégradation des séparations des groupes, toujours augmentant avec la diminution de  $d_{moy.j}$ . Ceci est tel que la séparation est clairement perdue dans les deux premières SIMULATIONS des profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ). En plus de la dégradation des séparations existantes lors des AFCnd, on observe une tendance au regroupement de toutes les espèces du même côté de l'axe 1.
- Enfin l'ENSEMBLE 5 (gradient de PdD différents entre espèces), montre également une perte d'information sur la séparation des groupes sur l'axe 1 (profil 2 et 3) (graphique (f)).
- Pour tous les ENSEMBLES, les résultats sur l'axe 2 (graphiques (c), (d), (g) et (h)) ne présentent pas de différence remarquable entre AFCnd et AFCd.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 32 :** illustrant les « séparations » des groupes d'espèces A (les espèces paires) et B (les espèces impaires), lors de SIMULATIONs avec un gradient de Pdd linéaire Elle présente deux boxplots par SIMULATION, un pour chaque groupe d'espèces. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE 4 où dans chacune des SIMULATIONs 11 espèces sur 20 ont un même gradient de Pdd, les 9 autres ayant une Pdd constamment égale à 1. Les quatre derniers graphiques (e), (f), (g) et (h) illustrent les résultats de l'ENSEMBLE 5 où dans chacune des SIMULATIONs toutes les espèces ont des gradients de Pdd différents. La première et la troisième colonnes de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des AFCnd, et la deuxième et la quatrième colonnes de graphiques ((b), (d) et (f), (h)) lors des AFCd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

Enfin, la figure 32 (ci-contre), compare la séparation **des groupes d'espèces A et B** entre les axes des AFCnd et les axes des AFCd. Cette figure ne présente que les résultats des ENSEMBLES 4 et 5. En effet, pour cet indicateur, les résultats de l'ENSEMBLE 3 ne sont pas indispensables pour répondre aux questions posées.

**Observations :** Dans les graphiques concernant les AFCnd ((a), (c), (e) et (g)), les séparations des groupes A et B sont sans surprise : les séparations existent uniquement dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) sur l'axe 1, et dans le profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) sur l'axe 2.

Les résultats pour les AFCd (graphiques (b), (d), (f) et (h)), montrent sur les axes 1 de l'ENSEMBLE 4 (même gradient de PdD pour 11 espèces sur 20) (graphique (b)), une tendance à regrouper les espèces du même côté de l'axe. Cet effet s'accroît avec la diminution des valeurs de  $d_{moy.j}$ .

On observe, toujours sur les axes 1, la disparition de la séparation des groupes A et B, dans le profil 4 de l'ENSEMBLE 5 (graphique (f)), et jusqu'à ce que  $d_{moy.j}$  égale 0.5 dans l'ENSEMBLE 4 (graphique (b)).

Sur les axes 2 des deux ENSEMBLES (graphique (d) et (h)), malgré une dégradation des séparations dans les profils 3, celles-ci restent visibles, et il apparaît des séparations des deux groupes dans le profil 4.

Je retiens des résultats concernant un gradient linéaire de PdD lors des AFC, plusieurs points :

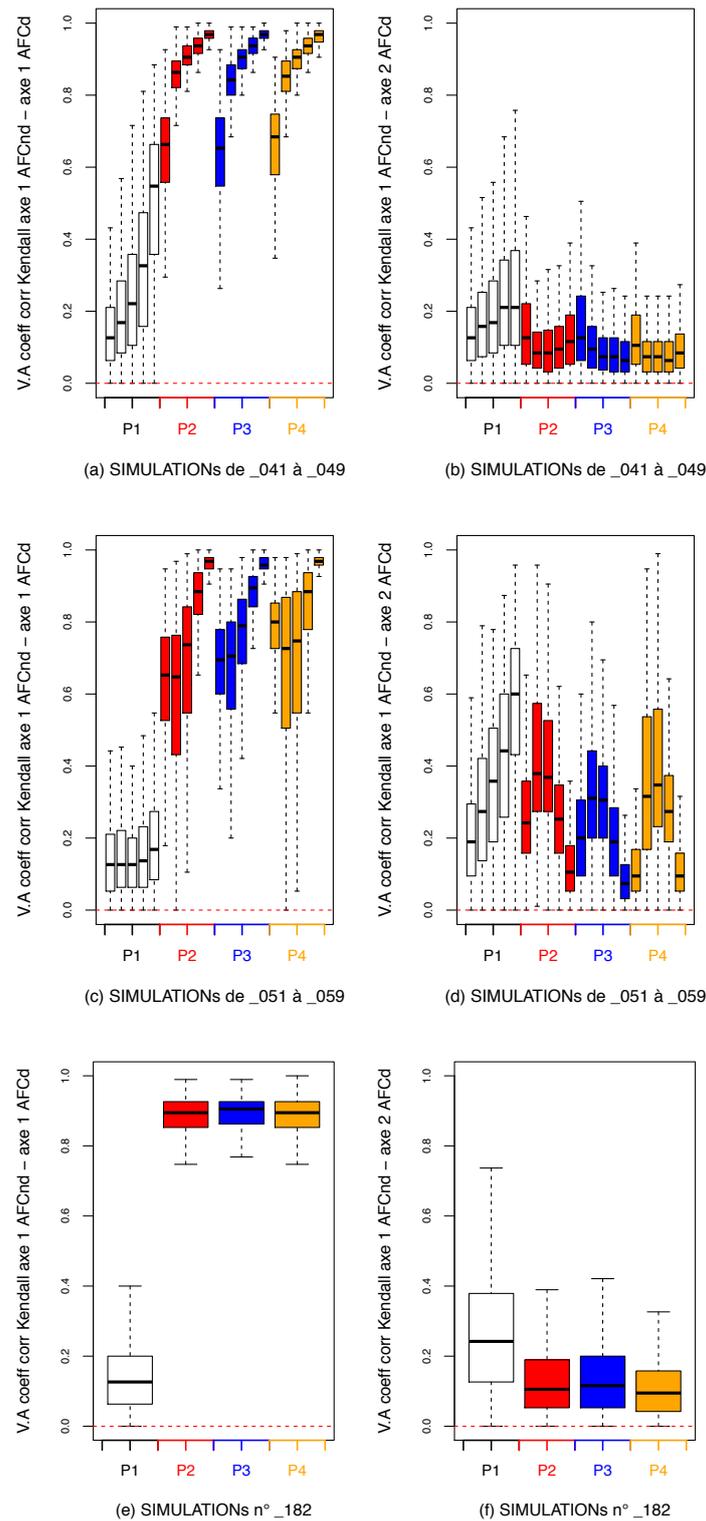
- (i) Que les ordinations des axes 1 et 2 des AFCnd, sont dans tous les cas modifiées, donc dégradées par la PdD. Ces dégradations sont d'autant plus importantes que la PdD est faible et que l'ordination se trouve sur l'axe 2.
- (ii) Que les dégradations des ordinations des axes 1, quand elles (les ordinations) correspondent à une des deux variables  $g_1$  ou  $g_2$ , ne sont pas très importantes lorsque toutes les espèces ont le même gradient de PdD.
- (iii) Que les dégradations sont plus importantes dès lors que j'introduis en plus du gradient (hétérogénéité de PdD entre les sites) une hétérogénéité de PdD entre les espèces.
- (iv) Que je retrouve régulièrement sur l'axe 2 des AFCd l'ordination, bien que dégradée, présente sur l'axe 1 lors des ACPnd correspondantes, lorsque cette ordination met en évidence la variable  $g_2$  (profil 4).
- (v) Que l'on met en évidence une relation entre certaines ordinations lors des AFCd, dans les ENSEMBLES 4 et 5, et valeur moyenne de la PdD des espèces.

Je reviendrai sur ces points dans le chapitre 4.

### 3.3.3 Les SIMULATIONs avec gradient de détection quadratique « centré ».

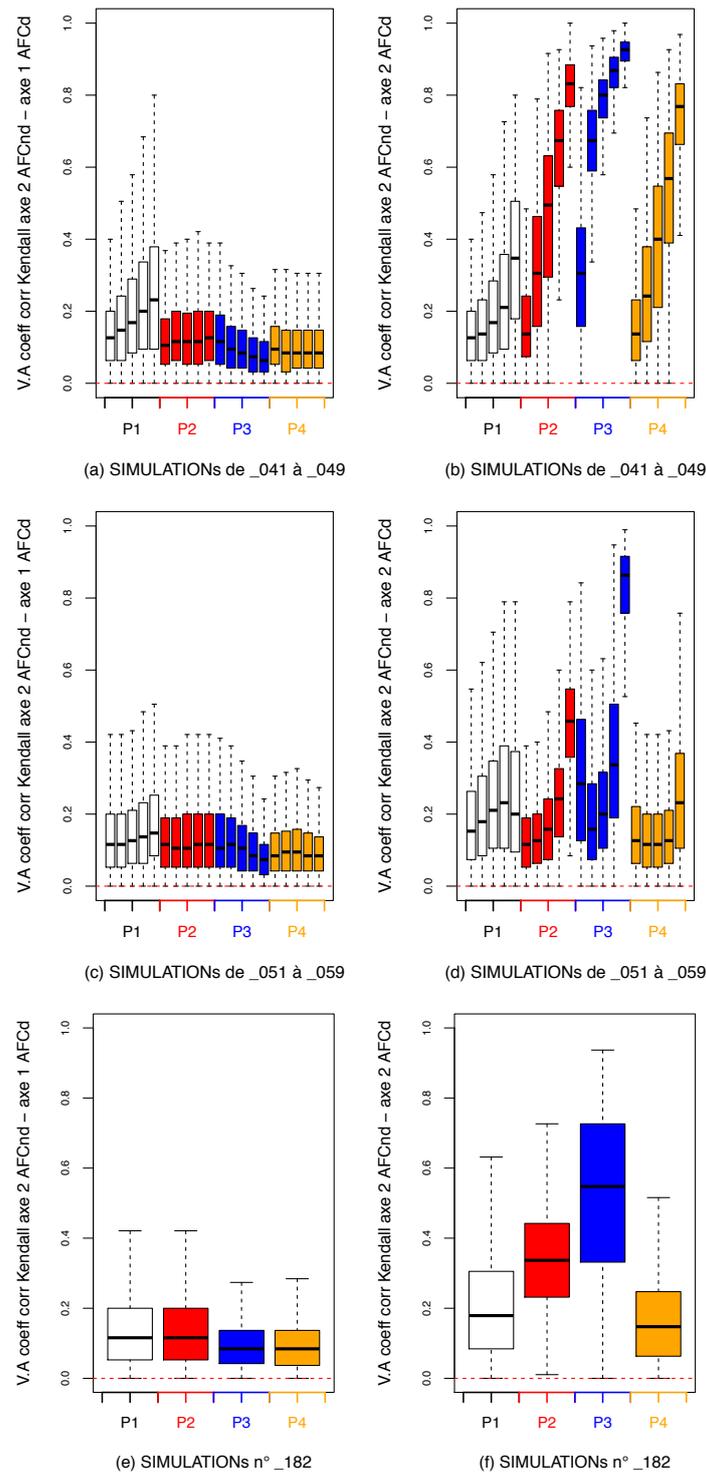
Après les gradients de détection linéaires, on s'intéresse aux conséquences du premier type de gradient quadratique. Je rappelle que dans les SIMULATIONs de cette partie, la PdD de toutes les espèces atteint son maximum lorsque  $g_1 = 500$ . C'est pourquoi je parlerai du gradient de détection centré.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 33 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1 de l'AFCnd, et les ordinations des espèces sur les axes 1 (graphique (a), (c) et (e)) et 2 (graphique (b), (d) et (f)) de l'AFCd ; lors de SIMULATIONs avec un gradient de PdD quadratique centré. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 6 où dans chacune des SIMULATIONs toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 7 où dans chacune des SIMULATIONs 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (e) et (f) illustrent les résultats de l'ENSEMBLE 8 où dans chacune des SIMULATIONs les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 34 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 2 de l'AFCnd, et les ordinations des espèces sur les axes 1 (graphique (a), (c) et (e)) et 2 (graphique (b), (d) et (f)) de l'AFCd ; lors de SIMULATIONS avec un gradient de PdD quadratique centré. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 6 où dans chacune des SIMULATIONS toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 7 où dans chacune des SIMULATIONS 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (e) et (f) illustrent les résultats de l'ENSEMBLE 8 où dans chacune des SIMULATIONS les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Pour la présentation des résultats des SIMULATIONs avec gradient de détection quadratique centré, j'ai trois ENSEMBLEs de SIMULATIONs. Le premier ENSEMBLE (ENSEMBLE 6) regroupe les SIMULATIONs où dans chacune d'elles, les espèces ont toutes le même gradient de PdD et dont les trois derniers chiffres sont compris entre 041 et 049. Le deuxième ENSEMBLE (ENSEMBLE 7) regroupe les SIMULATIONs où dans chacune d'elles, 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1) et dont les trois derniers chiffres sont compris entre 051 et 059. Le dernier ENSEMBLE (ENSEMBLE 8) regroupe les SIMULATIONs où dans chacune d'elles, les espèces ont des gradients de PdD différents et dont les derniers chiffres sont égaux à 182.

Commençons avec les **valeurs absolues des coefficients de corrélations de Kendall entre les ordinations des espèces sur l'axe1 des AFCnd** et les ordinations sur les axes 1 et 2 des AFCd. La figure 33 (page 67) présente les résultats des trois ENSEMBLEs de SIMULATIONs.

**Observations** : sur les axes 1 (graphique (a), (c) et (e)), les profils 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ) se détachent par des coefficients de corrélations toujours inférieurs à ceux des autres profils. Ces derniers sont d'ailleurs faibles dans les profils 1 des ENSEMBLEs 7 (même gradient de PdD pour 11 espèces sur 20) (graphique (c)) et 8 (gradient de PdD différents entre espèces) (graphique (e)) et dans l'ENSEMBLE 6 (même gradient de PdD entre les espèces) (graphique (a)) tant que  $d_{max,j} < 0.9$ .

Toujours sur l'axe 1, dans les profils 2, 3 et 4 (où l'abondance dépend au moins d'une des deux variables d'habitat ( $g_1$  ou  $g_2$ )), des corrélations fortes et croissantes avec  $d_{max,j}$  apparaissent dans tous les ENSEMBLEs. L'ENSEMBLE 7 présente néanmoins des corrélations moins importantes que celles de l'ENSEMBLE 6 toute chose étant égale par ailleurs (profil et valeur de  $d_{max,j}$ ).

Pour les axes 2 (graphique (b), (d) et (f)), les ENSEMBLEs 6 et 8 font apparaître des corrélations faibles et stables quel que soit le profil. Seul l'ENSEMBLE 7 affiche un profil 1 croissant de faible à moyen, et des profils 2, 3 et 4 globalement faibles, avec un aspect parabolique convexe.

La figure 34 (page 68), illustre les **valeurs absolues des coefficients de corrélations de Kendall entre les ordinations des espèces sur l'axe2 des AFCnd** et les ordinations sur les axes 1 et 2 des AFCd.

**Observations** : Dans tous les cas, sur les axes 1 (graphique (a), (c) et (e)), toutes les corrélations sont faibles.

Les graphiques (b), (d) et (f) représentent les axes 2 des AFCd.

L'ENSEMBLE 6 (même gradient de PdD entre les espèces) (graphique (b)), affiche, des corrélations croissantes pour chacun des profils, de faibles à moyennes pour le profil 1, et de faibles à fortes pour les autres profils.

L'ENSEMBLE 7 (même gradient de PdD pour 11 espèces sur 20) (graphique (d)), pour les profils 1, 2 et 4 présente des corrélations faibles légèrement croissantes, jusqu'à moyennes pour le profil 2 (où les abondances moyennes ne dépendent que de  $g_1$ ) . Seul le profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) présente une évolution des corrélations parabolique concave, toutes faibles, mais fortes lorsque  $d_{max,j}$  égale 0.9.

Pour l'ENSEMBLE 8 (gradient de PdD différents entre espèces) (graphique (f)), on note que les corrélations sont au mieux moyennes dans le profil 3 et faibles dans les autres.

La figure 35 expose les **valeurs absolues des coefficients de corrélation de Kendall** entre les **valeurs  $d_{max,j}$**  et l'ordination des espèces sur l'axe 1 et l'axe 2 des analyses.

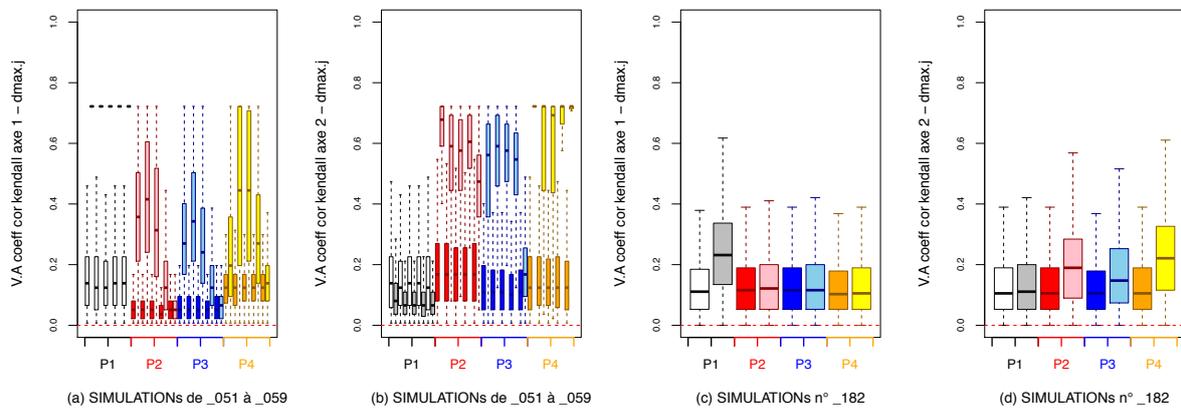
**Observations** : On retrouve à nouveau pour cet indicateur, des corrélations constamment faibles lors des AFCnd.

En revanche, les AFCd affichent à une exception près (ENSEMBLE 7, axe 2, profil 1 (graphique (b))) des corrélations systématiquement supérieures ou égales à celles obtenues lors des AFCnd.

Pour les deux ENSEMBLES, le profil 1 sur l'axe 1, présente toujours les corrélations les plus importantes. Ces dernières sont fortes dans l'ENSEMBLE 7 (même gradient de PdD pour 11 espèces sur 20) (graphique (a)), et faibles dans l'ENSEMBLE 8 (gradient de PdD différents entre espèces) (graphique (c)).

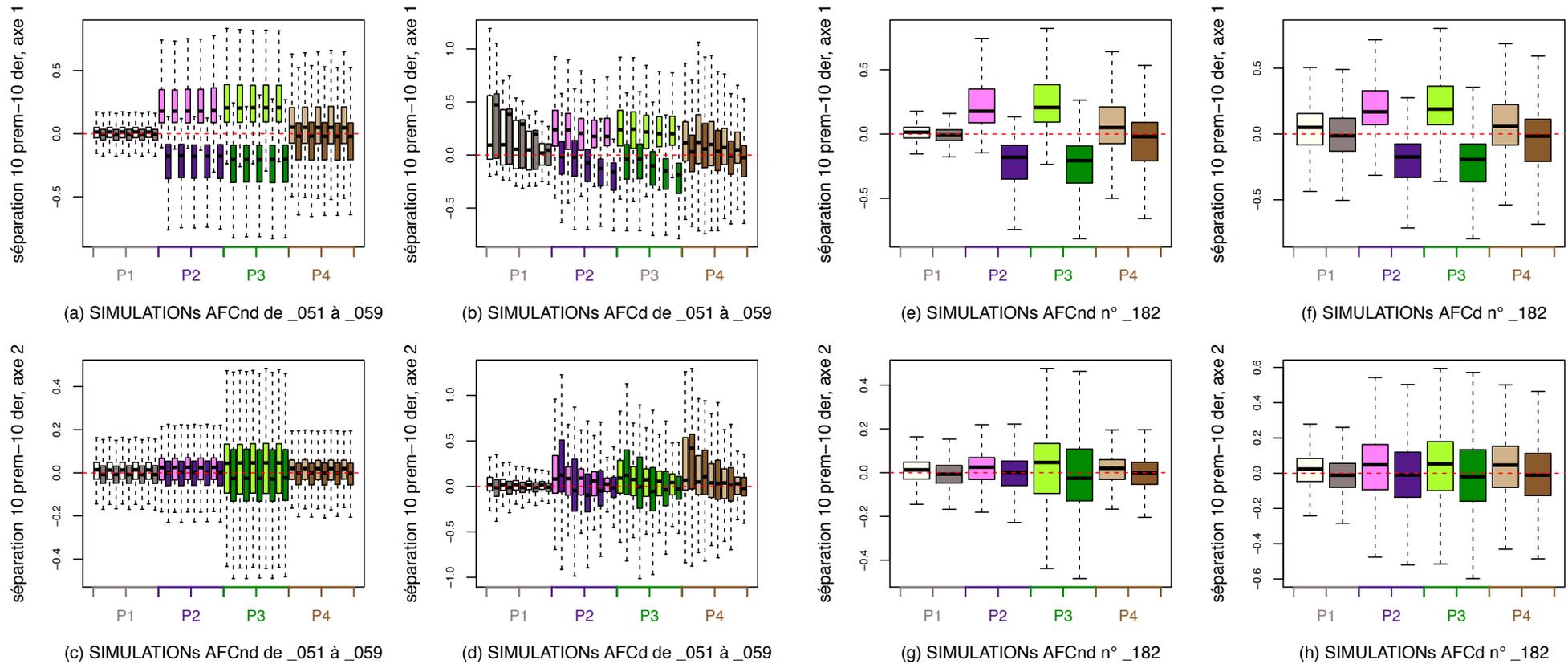
Les profils 2, 3 et 4 sur les axes 1 et 2, affichent également des corrélations supérieures à celles obtenues lors des AFCnd.

Les corrélations les plus importantes après celles du profil 1 sur l'axe 1 sont celles de l'axe 2 ; et s'organisent de façon décroissantes entre les profils 4, puis 2, puis 3 ; de fortes à moyenne pour l'ENSEMBLE 7 (graphique (b)), et constamment faibles pour l'ENSEMBLE 8 (graphique (d)).



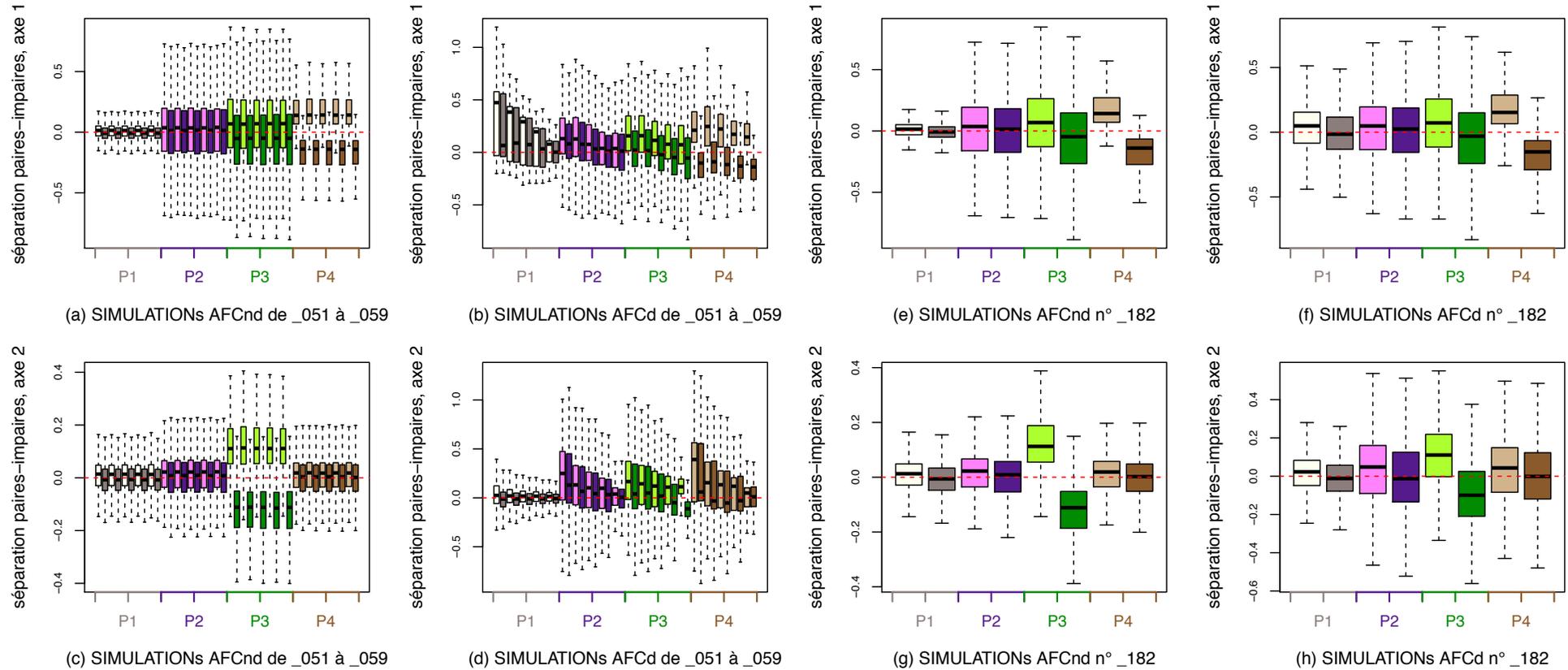
**Figure 35** : graphique double (deux boxplots par SIMULATION, un pour l'AFCnd et un pour l'AFCd) des valeurs absolues des coefficients de corrélation de Kendall entre les valeurs  $d_{max,j}$  et l'ordination des espèces sur les axes 1 (graphiques (a) et (c)) et 2 (graphiques (b) et (d)) des analyses ; lors de SIMULATIONs avec un gradient de PdD quadratique centré. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 7 où dans chacune des SIMULATIONs 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 8 où dans chacune des SIMULATIONs les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 36 :** illustrant les « séparations » des groupes d'espèces 1 (les 10 premières espèces) et 2 (les 10 dernières espèces), lors de SIMULATIONS avec un gradient de PdD quadratique centré. Elle présente deux boxplots par SIMULATION, un pour chaque groupe d'espèces. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE 7 où dans chacune des SIMULATIONS 11 espèces sur 20 ont un même gradient de PdD, les 9 autres ayant une PdD constamment égale à 1. Les quatre derniers graphiques (e), (f), (g) et (h) illustrent les résultats de l'ENSEMBLE 8 où dans chacune des SIMULATIONS toutes les espèces ont des gradients de PdD différents. La première et la troisième colonnes de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des AFCnd, et la deuxième et la quatrième colonnes de graphiques ((b), (d) et (f), (h)) lors des AFCd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 37 :** illustrant les « séparations » des groupes d'espèces A (les espèces paires) et 2 (les espèces impaires), lors de SIMULATIONS avec un gradient de PdD quadratique centré. Elle présente deux boxplots par SIMULATION, un pour chaque groupe d'espèces. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE 7 où dans chacune des 11 espèces sur 20 ont un même gradient de PdD, les 9 autres ayant une PdD constamment égale à 1. Les quatre derniers graphiques (e), (f), (g) et (h) illustrent les résultats de l'ENSEMBLE 8 où dans chacune des SIMULATIONS toutes les espèces ont des gradients de PdD différents. La première et la troisième colonnes de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des AFCnd, et la deuxième et la quatrième colonnes de graphiques ((b), (d) et (f), (h)) lors des AFCd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

La figure 36 (page 71), présente les séparations **des groupes d'espèces 1 et 2** sur les axes des AFCnd et sur les axes des AFCd.

**Conseil de lecture** : la figure 36, présente les résultats des ENSEMBLES 7 (même gradient de PdD pour 11 espèces sur 20) et 8 (gradient de PdD différents entre espèces). La lecture est similaire à celle de la figure 31 (page 63).

**Observations** : Dans les deux ENSEMBLES nous retrouvons les séparations attendues lors des AFCnd, c'est-à-dire sur l'axe 1 (graphiques (a) et (e)), dans les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ).

Lors des AFCd, on retrouve dans l'ENSEMBLE 8, une dégradation des séparations existantes ; dégradations légèrement accentuées lorsque  $d_{max,j}$  égale 0.1. Cependant les séparations sont toujours visibles.

Dans l'ENSEMBLE 7, non seulement la dégradation des séparations est toujours présente, mais elle est aussi plus prononcée. De plus, le profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ) présente sur l'axe 1 une tendance au regroupement des espèces du même côté de l'axe.

La figure 37 (page 72) permet la comparaison des séparations **des groupes d'espèces A et B**. Cette figure présente les résultats des ENSEMBLE 7 (même gradient de PdD pour 11 espèces sur 20), et 8 (gradients de PdD différents entre espèces).

**Observations** : Les résultats lors des AFCnd sont conformes aux attentes.

Pour les AFCd (colonne de droite) le constat le plus important est la forte dégradation dans l'ENSEMBLE 8 (graphique (h)), et la perte totale dans l'ENSEMBLE 7 (graphique (d)) (exception faite lorsque  $d_{max,j} = 0.9$ ), de la séparation des groupes A et B dans le profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) sur l'axe2.

On retrouve également la tendance au regroupement des espèces du même côté de l'axe 1 dans le profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ) de l'ENSEMBLE 7 (graphique (b)).

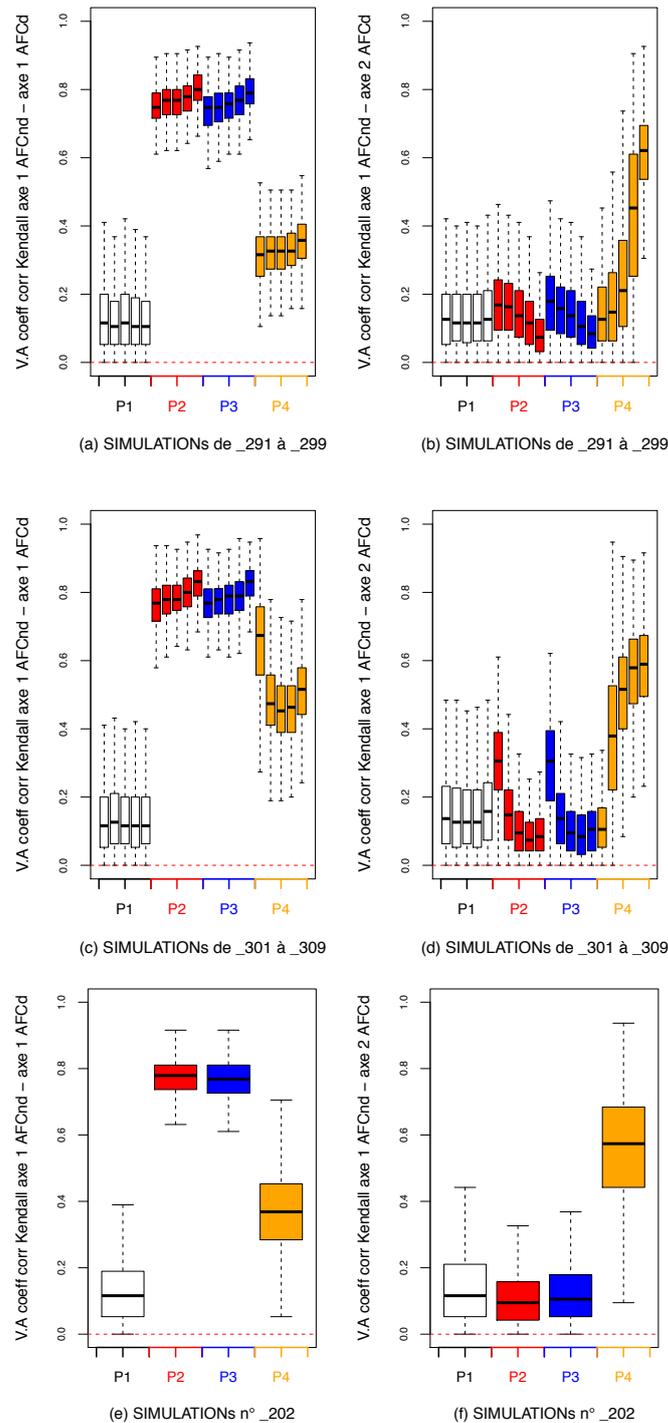
En revanche, les séparations des groupes A et B, dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ), ne semblent pas altérées dans l'ENSEMBLE 8 (graphique (f)) et seulement dégradées lorsque  $d_{max,j}$  égale 0.3 ou 0.5 dans l'ENSEMBLE 7 (graphique (b)).

### 3.3.4 Les SIMULATIONs avec gradients de détection quadratiques « décalés ».

Passons maintenant aux résultats correspondant au deuxième type de gradient quadratique. On rappelle que les SIMULATIONs de cette partie sont alors faites avec une PdD qui, pour chaque espèce atteint son maximum lorsque  $g_1$  prend une valeur plus ou moins inférieure à  $g_{1j}$  (valeur de  $g_1$  à l'optimum de l'espèce). Le plus grand écart possible entre ces deux valeurs est de 300. C'est pourquoi on parlera du gradient de détection décalé.

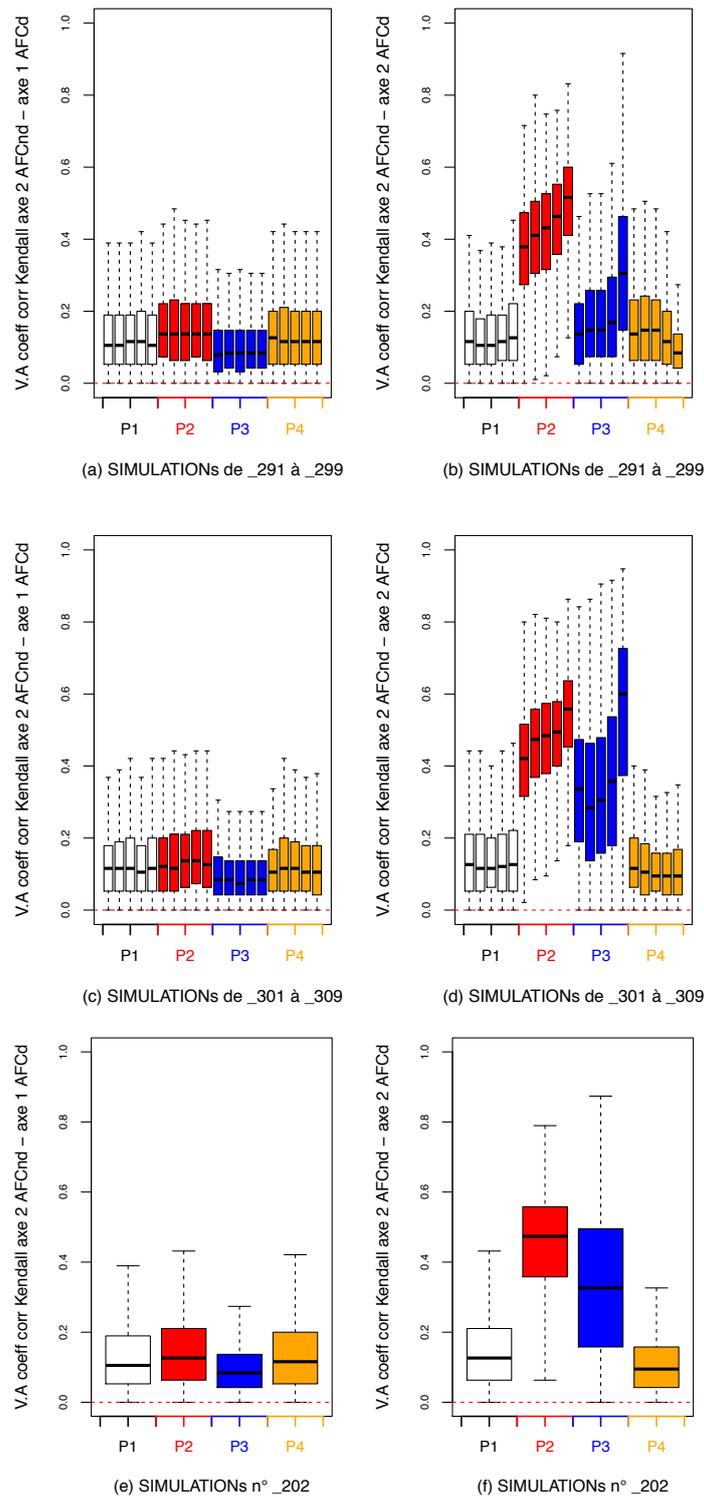
Pour la présentation des résultats des SIMULATIONs avec gradients de détection quadratiques décalés, j'ai trois ENSEMBLES de SIMULATIONs. Le premier ENSEMBLE (ENSEMBLE 9) regroupe les SIMULATIONs où dans chacune d'elles, les espèces ont toutes le même gradient de PdD et dont les trois derniers chiffres sont compris entre 291 et 299. Le deuxième ENSEMBLE (ENSEMBLE 10) regroupe les SIMULATIONs où dans chacune d'elles, 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1) et dont les trois derniers chiffres sont compris entre 301 et 309. Le dernier ENSEMBLE (ENSEMBLE 11) regroupe les SIMULATIONs où dans chacune d'elles, les espèces ont des gradients de PdD différents et dont les derniers chiffres sont égaux à 202.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 38** : graphique simple (un boxplot par SIMULATION) des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 1 de l'AFCnd, et les ordinations des espèces sur les axes 1 (graphique (a), (c) et (e)) et 2 (graphique (b), (d) et (f)) de l'AFCd ; lors de SIMULATIONs avec un gradient de PdD quadratique décalé. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 9 où dans chacune des SIMULATIONs toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 10 où dans chacune des SIMULATIONs 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (e) et (f) illustrent les résultats de l'ENSEMBLE 11 où dans chacune des SIMULATIONs les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 39 :** graphique simple (un boxplot par SIMULATION) des valeurs absolues des coefficients de corrélation de Kendall entre l'ordination des espèces sur l'axe 2 de l'AFCnd, et les ordinations des espèces sur les axes 1 (graphique (a), (c) et (e)) et 2 (graphique (b), (d) et (f)) de l'AFCd ; lors de SIMULATIONS avec un gradient de PdD quadratique décalé. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 9 où dans chacune des SIMULATIONS toutes les espèces ont le même gradient de PdD. Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 10 où dans chacune des SIMULATIONS 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (e) et (f) illustrent les résultats de l'ENSEMBLE 11 où dans chacune des SIMULATIONS les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Les valeurs absolues des coefficients de corrélations de Kendall entre les ordinations des espèces sur l'axe1 des AFCnd et les ordinations sur les axes 1 et 2 des AFCd sont présentées dans la figure 38 (page 74).

**Observations :** sur les axes 1 (graphiques (a), (c) et (e)), on retrouve pour tous les ENSEMBLES, des corrélations faibles dans les profils 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ), fortes et augmentant peu avec l'augmentation de  $d_{max,j}$  dans les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ), et moyennes dans les profils 4 (où les abondances moyennes ne dépendent que de  $g_2$ ).

Sur les axes 2 (graphiques (b), (d) et (f)), les profils 1 montrent tous des corrélations faibles et constantes.

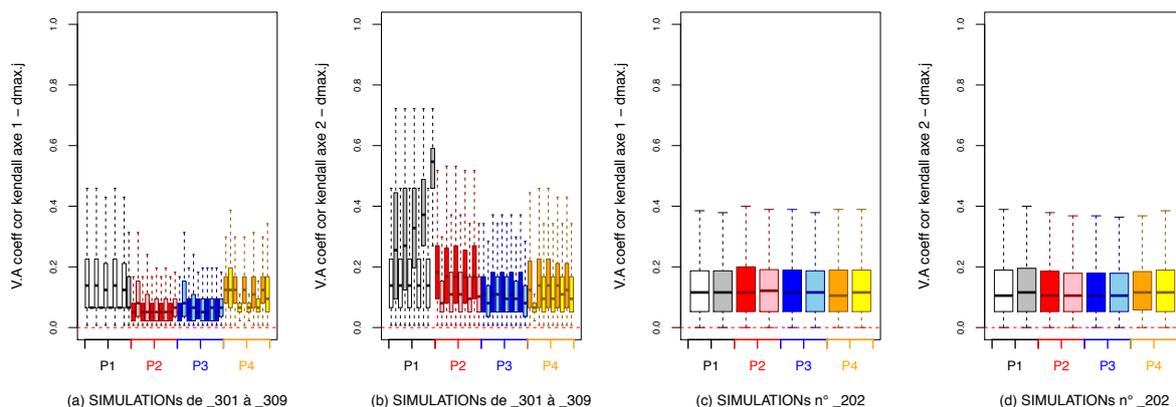
Alors que les corrélations diminuent avec l'augmentation de  $d_{max,j}$  dans les profils 2 et 3 (où les abondances moyennes dépendent entre autre de  $g_1$ ), celles-ci augmentent dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ). De plus les corrélations du profil 4, tout en restant moyennes, atteignent, voir dépassent, les valeurs obtenues sur l'axe 1.

La figure 39 (page 75), présente les valeurs absolues des coefficients de corrélations de Kendall entre les ordinations des espèces sur l'axe2 des AFCnd et les ordinations sur les axes 1 et 2 des AFCd.

**Observations :** on pourrait presque se contenter de ne retenir que l'absence de corrélations fortes.

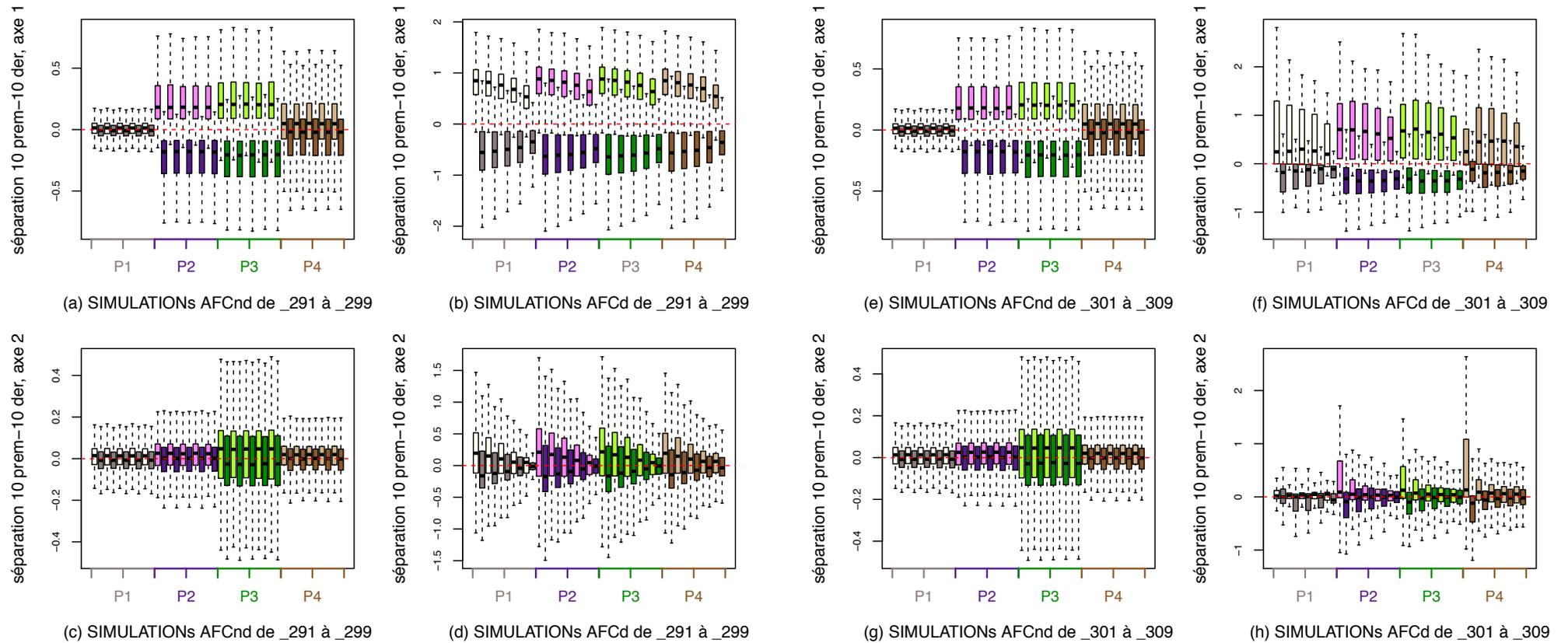
On précisera cependant que:

- Sur l'axe 1 (graphique de gauche, (a), (c) et (e)) les corrélations sont faibles pour tous les ENSEMBLES et tous les profils.
- De même sur l'axe 2 (graphique de droite, (b), (d) et (f)) pour les profils 1 et 4 (où les abondances moyennes ne dépendent pas de  $g_1$ ).
- Toujours sur l'axe 2, les profils 2 (où les abondances moyennes ne dépendent que de  $g_1$ ) des trois ENSEMBLES présentent des corrélations faiblement croissantes avec  $d_{max,j}$ , et toujours moyennes.
- Enfin, les corrélations du profil 3 (où les abondances moyennes dépendent de  $g_1$  et de  $g_2$ ) évoluent entre faibles et moyennes.



**Figure 40 :** graphique double (deux boxplots par SIMULATION, un pour l'AFCnd et un pour l'AFCd) des valeurs absolues des coefficients de corrélation de Kendall entre les valeurs  $d_{max,j}$  et l'ordination des espèces sur les axes 1 (graphiques (a) et (c)) et 2 (graphiques (b) et (d)) des analyses ; lors de SIMULATIONs avec un gradient de PdD quadratique décalé. Les graphiques (a) et (b) illustrent les résultats de l'ENSEMBLE 10 où dans chacune des SIMULATIONs 11 espèces sur 20 ont le même gradient de PdD (les autres espèces ayant une PdD constante et égale à 1). Les graphiques (c) et (d) illustrent les résultats de l'ENSEMBLE 11 où dans chacune des SIMULATIONs les espèces ont des gradients de PdD différents. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONs.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 41 :** illustrant les « séparations » des groupes d'espèces 1 (les 10 premières espèces) et 2 (les 10 dernières espèces), lors de SIMULATIONS avec un gradient de PdD quadratique décalé. Elle présente deux boxplots par SIMULATION, un pour chaque groupe d'espèces. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE 9 où dans chacune des SIMULATIONS toutes les espèces ont le même gradient de PdD. Les quatre derniers graphiques (e), (f), (g) et (h) illustrent les résultats de l'ENSEMBLE 10 où dans chacune des SIMULATIONS 11 espèces sur 20 ont un même gradient de PdD, les 9 autres ayant une PdD constamment égale à 1. La première et la troisième colonnes de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des AFCnd, et la deuxième et la quatrième colonnes de graphiques ((b), (d) et (f), (h)) lors des AFCd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

La figure 40 (page 76) donne les **valeurs absolues des coefficients de corrélation de Kendall entre les valeurs  $d_{max,j}$**  et l'ordination des espèces sur l'axe 1 et l'axe 2 des analyses.

**Observations** : Les coefficients de corrélations sont faibles lors des AFCnd pour toutes les SIMULATIONs.

Dans les AFCd, seules les SIMULATIONs de l'ENSEMBLE 10 (même gradient de PdD pour 11 espèces sur 20), sur l'axe 2 (graphique (b)) et dans le profil 1 (où les abondances ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ), diffèrent de celles des AFCnd. Les valeurs de corrélations augmentent de faibles à moyennes avec  $d_{max,j}$ .

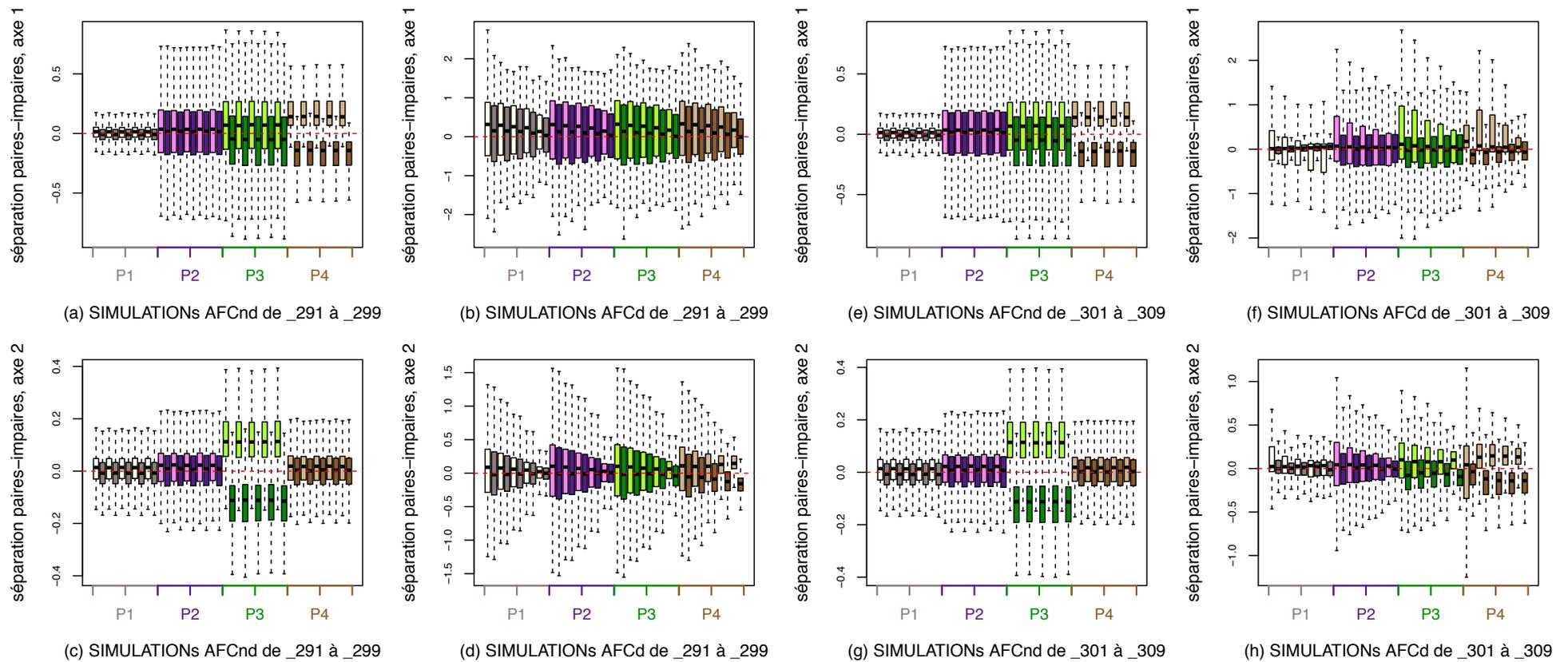
La figure 41 (ci-contre) présente les séparations **des groupes d'espèces 1 et 2** sur les axes des AFCnd et sur les axes des AFCd, pour les ENSEMBLES 9 (même gradient de PdD entre espèces) et 10 (même gradient de PdD pour 11 espèces sur 20).

**Observations** : Dans les deux ENSEMBLES de SIMULATIONs, nous retrouvons les séparations attendues lors des AFCnd ; sur l'axe 1 (graphiques (a) et (e)) dans les profils 2 et 3.

Or, lors des AFCd, j'observe (graphiques (b), (d), (f) et (h)):

- sur les axes 1 (graphiques (b), (f)), l'apparition de la séparation des groupes 1 et 2 dans tous les profils et pour toutes les SIMULATIONs. Ces séparations sont plus marquées dans l'ENSEMBLE 9 (même gradient de PdD entre les espèces) que dans l'ENSEMBLE 10 (même gradient de PdD pour 11 espèces sur 20).
- Sur les axes 2 (graphiques (d), (h)), la même tendance de séparation des groupes apparaît dans tous les profils, décroissantes avec  $d_{max,j}$ , mais peu marquée bien que plus visible pour l'ENSEMBLE 9 (graphiques (d)) que pour l'ENSEMBLE 10 (graphiques (h)).

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



**Figure 42 :** illustrant les « séparations » des groupes d'espèces A (les espèces paires) et 2 (les espèces impaires), lors de SIMULATIONS avec un gradient de PdD quadratique décalé. Elle présente deux boxplots par SIMULATION, un pour chaque groupe d'espèces. Les quatre premiers graphiques (a), (b), (c) et (d) illustrent les résultats pour l'ENSEMBLE où dans chacune des SIMULATIONS toutes les espèces ont le même gradient de PdD. Les quatre derniers graphiques (e), (f), (g) et (h) illustrent les résultats de l'ENSEMBLE 10 où dans chacune des SIMULATIONS 11 espèces sur 20 ont un même gradient de PdD, les 9 autres ayant une PdD constamment égale à 1. La première et la troisième colonnes de graphiques ((a), (c) et (e), (g)) donnent les résultats lors des AFCnd, et la deuxième et la quatrième colonnes de graphiques ((b), (d) et (f), (h)) lors des AFCd. La première ligne de graphiques ((a), (b) et (e), (f)) correspond aux résultats de l'axe 1 des analyses. La deuxième ligne de graphiques ((c), (d) et (g), (h)) correspond aux résultats de l'axe 2 des analyses. Enfin P1, P2, P3 et P4 correspondent au profil d'abondance auquel appartient les SIMULATIONS.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Enfin, la figure 42 (ci-contre) permet de comparer les séparations **des groupes d'espèces A et B** sur les axes des AFCnd aux séparations des mêmes groupes sur les axes des AFCd. Cette figure ne présente que les résultats des ENSEMBLES 9 (même gradient de PdD entre espèces) et 10 (même gradient de PdD pour 11 espèces sur 20)..

**Observations** : les résultats lors des AFCnd sont conformes aux attentes à savoir une séparation des groupes dans le profil 4 sur l'axe 1 (graphiques (a) et (e)) et dans le profil 3 sur l'axe 2 (graphiques (c) et (g)). En revanche pour les AFCd (graphiques (b), (d), (f) et (h)) on observe :

- Pour l'ENSEMBLE 9 (même gradient de PdD entre les espèces) (graphiques (b) et (d)) une perte totale de la séparation dans le profil 3 quel que soit l'axe. Une légère différenciation dans le profil 4 sur l'axe 2 (graphique (d)), lorsque  $d_{max.j}$  est supérieur ou égal à 0.7 ; séparation absente de l'axe 1. On observe aussi une similitude pour les quatre profils d'abondances sur l'axe 1 (graphique (b)).
- Pour l'ENSEMBLE 10 (même gradient de PdD pour 11 espèce sur 20) (graphique (f) et (h)), l'analyse est identique, seule la différenciation dans le profil 4 (où les abondances moyennes ne dépendent que de  $g_2$ ) sur l'axe 2 (graphique (h)) apparaît un peu plus tôt avec  $d_{max.j}$  égale à 0.3.

Je retiens des résultats concernant un gradient quadratique décalé de PdD lors des AFC, plusieurs points :

- (i) Que les ordinations des axes 1 et 2 des AFCnd, sont dans tous les cas modifiées, donc dégradées par la PdD. Ces dégradations sont d'autant plus importantes que la PdD est faible et que l'ordination se trouve sur l'axe 2.
- (ii) Qu'il apparaît une séparation des groupes d'espèces 1 et 2 sur les axes 1 des AFCd dans les profils où les abondances des espèces ne dépendent pas de  $g_1$  (profil 1 et 4). Que ces séparations présentes lors des AFCnd, sont modifiées mais existantes dans les AFCd.
- (iii) Que les ordinations présentes sur les axes 2 des AFCnd, lorsqu'elles correspondent à une variable latente ( $g_2$ , dans le profil 3), sont perdues dans le premier plan de projection de AFCd.
- (iv) Que les ordinations présentes sur l'axe 1 de l'AFCnd et correspondant à la variable  $g_2$  (profil 4), sont perdues sur l'axe 1 de l'AFCd, mais bien que dégradées, parfois retrouvées sur les axes 2 de ces dernières (lorsque la valeur de  $d_{max.j}$  est suffisamment grande).
- (v) Que pour l'ENSEMBLE 10 (11 espèces sur 20 ont un gradient de PdD identique) une relation est mise en évidence entre la valeur maximale de la PdD des espèces et les ordinations sur l'axe 2 lors des AFCd faites sur le profil 1 (où les abondances des espèces ne dépendent d'aucune des deux variables  $g_1$  et  $g_2$ ).

Je reviendrai sur ces points dans le chapitre 4.

## Chapitre 4 : Analyse des résultats et discussion

### 4.1 Avant propos

Avant de commencer l'analyse des résultats, j'apporte quelques précisions. On attend des ACP et des AFC qu'elles ordonnent les espèces (et les sites) et mettent ainsi en évidence une ou des variables latentes (les facteurs environnementaux qui sont la cause des ordinations). Ainsi, la démarche pour interpréter les résultats sera toujours la même. On comparera les ordinations des analyses non dégradées avec les ordinations des analyses dégradées pour déterminer, d'une part s'il y a perte d'informations données par les ACPnd et les AFCnd ; d'autre part s'il y apparaît de nouvelles informations erronées dans les ACPd et les AFCd.

Connaissant les axes porteurs d'informations et les ordinations correspondantes dans les analyses de matrices non dégradées, tout changement dans les ordinations des matrices affectées par une  $PdD < 1$  signalera une altération de l'information donnée par les ACP et les AFC. Dès lors j'appelle dégradation d'une ordination toute différence entre l'ordination issue de l'analyse de la matrice non dégradée et l'ordination issue de l'analyse de la matrice dégradée.

Je considère qu'il y a deux niveaux de dégradations : (1) la dégradation partielle lorsqu'une ordination des espèces est modifiée sans pour autant perdre la mise en évidence de la variable latente ; (2) la dégradation totale d'une ordination des espèces lorsque cette dégradation a pour conséquence la disparition de la variable latente.

Un autre type d'altération de l'information d'une ACP et d'une AFC également possible à observer, est le passage d'une ordination des espèces sur un axe d'ordre inférieur.

Décrivons maintenant les effets observés de la PdD pour chaque type d'analyse (ACP ou AFC), et chaque type de gradient de PdD imposé (sans gradient, gradient linéaire, etc.).

### 4.2 Analyse des résultats des effets de la probabilité de détection dans les ACP

#### 4.2.1 Les SIMULATIONS sans gradient de PdD

Lors des simulations **sans gradient de PdD**, tous les indicateurs montrent, et cela sur les deux axes de projection, une baisse de la qualité de l'information donnée par les ACPnd qui est d'autant plus forte que la PdD est faible. On remarque également que plus l'inertie expliquée par un axe des ACPnd est faible, plus la détérioration de l'ordination sur cet axe est forte. De plus, le fait d'imposer une PdD différente entre les espèces a tendance à faire légèrement augmenter les dégradations.

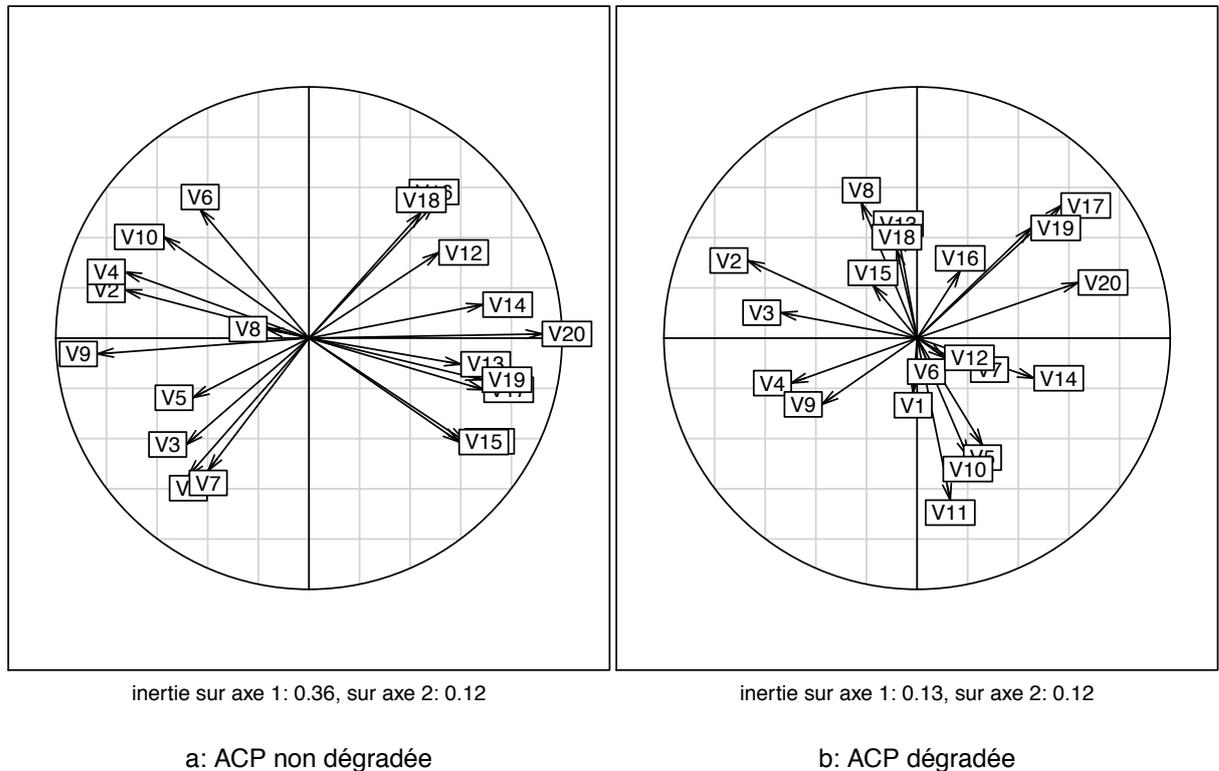
On distingue dès lors les 2 niveaux de dégradations des ordinations:

- (i) Une dégradation totale des ordinations (perte de la variable latente) lorsqu'il y a association entre faible probabilité de détection (valeur proche de 0.3) et faible inertie expliquée par l'axe de l'ACPnd (dans mon cas une inertie proche de 0.12 cf. figure 43); ce que nous montre le graphique (b) de la figure 9 (page 38).
- (ii) Une dégradation partielle dans les autres cas (on retrouve la variable latente  $g_1$  ou  $g_2$ ). En effet, les coordonnées des vecteurs supplémentaires lors des AFCd restent proches de celles des mêmes vecteurs lors des AFCnd (cf. les figures 10 (page 38) et 11 (page 40)). Cependant, les coefficients de corrélation de Kendall entre les ordinations des ACPnd et des ACPd montrent que ces ordinations ne sont pas identiques. On a donc conservé la

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

variable latente mise en évidence lors de l'ACPnd, mais dans le meilleur des cas, la position relative d'au moins une espèce par rapport à une autre est altérée.

La difficulté est alors de définir une limite acceptable à la dégradation partielle des ordinations. Je reviendrai sur cette question lors de la discussion des résultats.



*Figure 43 : graphiques des résultats de la millième simulation de la SIMULATION 3001. Dans cette simulation, toutes les espèces ont une abondance moyenne dépendante de deux variables environnementales ( $g_1$  et  $g_2$ ) ; et dans le cas de l'ACP dégradée (b), toutes les espèces ont une PdD constamment égale à 0.1. À gauche (a) le plan de projection de l'ACPnd, et à droite (b) le plan de projection de l'ACPd. Dans chaque plan de projection on retrouve les 20 espèces (numérotées de 1 à 20), qui sont pour les ACP 20 variables, ce qui explique le V présent devant le numéro de l'espèce. Dans le cas de l'ACPnd (a), la première composante principale correspond à  $g_1$  et la deuxième composante principale à  $g_2$ . On constate de forts changements entre les ordinations données par l'ACPnd et l'ACPd.*

Il y a bien entendu tous les intermédiaires entre ces deux niveaux de dégradation, comme par exemple sur l'axe d'ordre 2 du profil 3 dans l'ENSEMBLE 2 (cf. figure 11 page 40).

Une première conclusion est, d'une part que la seule sous-estimation des abondances, même minimale, a des conséquences sur l'organisation des espèces sur les axes de l'analyse ; d'autre part, même en se plaçant dans les conditions « idéales » (même PdD pour toutes les espèces sur tous les sites), une PdD faible (proche de 0,1) modifie fortement les ordinations, donc les résultats des ACP. La figure 43 illustre ce propos. Enfin sous l'hypothèse d'une PdD homogène entre les sites, il n'y a pas d'apparition d'ordination artificielle.

#### 4.2.2 Les SIMULATIONS avec des gradients de détection linéaires

Pour les simulations avec des **gradients de détection linéaires**, les indicateurs montrent, sur les deux axes de projection de chaque analyse, des dégradations des ordinations, quelle que soit la PdD. Ces dégradations augmentent avec la diminution de la PdD moyenne des espèces ( $d_{moy.j}$ ) et vont de partielles à totales. Là aussi, les cas où les espèces ont, entre elles, des PdD moyennes différentes, montrent des dégradations d'ordinations plus importantes et plus variables.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Pour expliquer ces dégradations, on identifie deux effets et leurs conséquences :

Le premier effet est dû à la variation de la PdD suivant un gradient. Cette variation fait que, pour chaque espèce, la PdD tend à donner une image de l'évolution de l'abondance en fonction de  $g_1$  identique à la sienne. Je résume cet effet sous le nom **d'effet d'attraction** du gradient de la PdD. Lorsque, comme c'est le cas dans mes simulations pour les ACP, le gradient de la PdD est linéaire et croissant en fonction de  $g_1$  et que les abondances moyennes des espèces sont monotones sur  $g_1$  (uniquement croissantes, invariantes ou décroissantes), il y a trois possibilités. D'une part si l'espèce a une abondance croissante le gradient de PdD tend à augmenter cette croissance (en terme de pente). D'autre part, si les abondances de l'espèce sont invariantes, le gradient de PdD tend à donner des abondances croissantes en fonction de  $g_1$ . Enfin, si l'abondance de l'espèce est décroissante, le gradient de PdD atténue la décroissance (en terme de pente), et parfois même inverse les évolutions des abondances et donne une image d'abondances croissantes en fonction de  $g_1$ . Cet effet tend ainsi à imposer aux espèces soumises au gradient de la PdD, une corrélation positive entre elles, ainsi qu'à la variable  $g_1$ .

Les conséquences de cet effet sont les suivantes :

- (i) Lorsque dans les versions non dégradées, **aucune variable latente** n'est corrélée avec le gradient de la PdD ( $g_1$ ), il en apparaît une dans les versions dégradées. On a donc apparition d'une variable latente artéfactuelle. Si aucune information (variable latente) n'existe dans l'ACPnd, le gradient de PdD (la variable latente artéfactuelle) se retrouve sur le premier axe de l'ACPd (d'ordre 1). Si une variable latente non corrélée avec le gradient de la PdD existe dans l'ACPnd (je l'appelle variable latente initiale), le gradient de PdD pourra se lire sur l'axe d'ordre 1 ou 2 selon la part d'inertie relative expliquée par chacune des deux variables latentes (l'initiale et l'artéfactuelle). Dans tous les cas, l'ordination de l'ACPnd est dégradée.
- (ii) Si dans les ACPnd il y a **déjà une variable latente** corrélée au gradient de PdD ( $g_1$ ), alors cette variable latente réapparaît dans les ACPd, mais l'ordination qui lui correspond cumule l'ordination de l'ACPnd et l'effet d'attraction de la PdD. L'ordination d'origine est donc dégradée et la variable latente retrouvée, mais en partie pour de mauvaises raisons. Si en plus de cette première variable latente, l'ACPnd met en évidence une deuxième variable latente, qui n'est donc pas corrélée avec le gradient de PdD, l'ordination correspondant à cette deuxième variable est également dégradée. Une ordination mettant en évidence une variable latente indépendante du gradient de PdD est donc également impactée par ce dernier. Néanmoins, pour des parts d'inertie expliquées par les axes porteurs des ordinations équivalentes (lors des ACPnd), une telle ordination semble moins dégradée qu'une ordination mettant en évidence une variable latente corrélée avec le gradient de la PdD.  
Enfin, toujours à propos de l'effet d'attraction, on note dans les cas particuliers où toutes les espèces ont exactement le même gradient, l'apparition d'un effet taille. Plus la probabilité moyenne de détection augmente, plus on retrouve au sein du regroupement des espèces une partie de l'ordination des espèces d'origine.

Un deuxième effet, moins important, est lié à la variabilité de la PdD entre les espèces (visible sur la figure 16 page 44). Les ACPd, lorsque les espèces ont toutes des gradients de PdD différents, tendent à séparer les espèces à forte PdD moyenne, de celles à faible PdD moyenne. Je suppose que la faible PdD de ces espèces induit une faible abondance sur tous les sites. Les abondances de ces espèces ne peuvent donc plus ou peu varier entre les sites et tendent ainsi à covarier. Je résume cet effet sous le nom **d'effet de séparation** (qui est indépendant de l'effet d'attraction).

En résumé, les effets des gradients de détection linéaires dégradent toujours et parfois fortement les résultats des analyses. Les dégradations des ordinations sont variables et fonction de différents facteurs, comme la valeur de l'inertie expliquée par l'axe porteur de l'ordination lors de l'ACPnd,

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

l'hétérogénéité des gradients de PdD entre les espèces, les valeurs moyennes de la PdD pour chaque espèce, et l'indépendance ou non de la variable latente vis-à-vis du gradient de PdD. Je reviendrai plus précisément sur ces points dans la partie discussion (cf. partie 4.4), car le fait que les dégradations des ordinations soit négligeable ou non dépendra fortement de l'information recherchée par la personne qui réalise les analyses.

### 4.2.3 Les SIMULATIONS avec des gradients de détection quadratiques

Enfin, pour le cas des ACP, les simulations avec des **gradients de détection quadratiques** centrés nous montrent également des dégradations des informations sur les deux axes de projections des analyses.

On retrouve l'effet d'attraction du gradient de PdD. Ce dernier tend toujours à donner une image de l'évolution des abondances en fonction de  $g_1$  identique à l'évolution des PdD des espèces en fonction de  $g_1$ . C'est-à-dire, ici, une forme parabolique convexe.

Ceci se traduit dans les ACPd, d'une part par la tendance, pour toutes les espèces qui subissent un gradient de PdD, à être corrélées positivement entre elles après la dégradation (donc tend à l'apparition d'un effet taille); et d'autre part, par l'absence de forte corrélation linéaire entre les axes d'ordre 1 et les deux variables  $g_1$  et  $g_2$ . On voit donc ici les conséquences du fait que les gradients de détection ont tous leur maximum à la même abscisse ( $g_1 = 750$ ).

On note que cet effet est plus important pour certaines valeurs de PdD maximales ( $d_{max.j}$ ) égales à 0.3, 0.5 et 0.7. Ces valeurs correspondent, en effet, aux simulations où les variations du gradient de la PdD (pentes) sont les plus fortes. C'est une conséquence directe de la méthodologie appliquée pour établir les gradients quadratiques. Je reviendrai sur ce point lors de la discussion.

Cet effet d'attraction a pour conséquence la création d'une nouvelle ordination artéfactuelle dans les ACPd, prenant régulièrement la place de l'ordination portée par les axes d'ordre 1 des ACPnd.

On trouve alors dans les ACPd deux possibilités en fonction de l'importance de l'effet d'attraction : (i) si l'effet est très marqué, l'ordination artéfactuelle prend la place de l'ordination de l'axe d'ordre 1 de l'ACPnd et celle-ci (si elle existe) est reléguée et dégradée sur l'axe d'ordre 2. Cette dégradation est d'autant plus forte que les probabilités de détection maximales sont petites ; et va jusqu'à la dégradation totale (perte de la variable latente) pour une probabilité maximale de détection proche de 0.1. (ii) Si l'effet est peu marqué (lorsque  $d_{max.j}$  égale 0.9 lorsque seulement 11 espèces sur 20 ont un gradient de PdD, les autres ayant une PdD constamment égale à 1 (l'ENSEMBLE 7)), l'information de l'axe d'ordre 1 de l'ACPnd est juste dégradée sur l'axe d'ordre 1 de l'ACPd, et la tendance au regroupement de toutes les espèces du même côté de l'origine se retrouve sur l'axe d'ordre 2.

Ainsi, les axes d'ordre 1 et 2 des ACPd sont dans tous les cas « occupés » soit par l'effet d'attraction, soit par l'ordination dégradée des axes d'ordre 1 des ACPnd. Par conséquent, l'ordination (donc la variable latente) des axes d'ordre 2 des ACPnd n'est jamais retrouvée dans le premier plan de projection. Les indicateurs n'ayant pas été utilisés sur les résultats des axes d'ordre 3, je ne peux dire si les ordinations des axes d'ordre 2 des ACPnd sont reléguées à un axe d'ordre inférieur ou non. Du fait de ce potentiel changement d'axe, elles sont au minimum dégradées. Quant à l'effet de séparation des espèces à forte PdD maximale d'avec les espèces à faible PdD maximale, déjà évoqué pour les gradients de PdD linéaire, on ne le retrouve pas sur les deux premiers axes de projection. Peut-être s'exprime-t'il, là encore, sur des axes d'ordre inférieur, mais je ne peux le dire ici. C'est pourquoi je ne peux m'exprimer sur l'effet d'augmentation ou de diminution des dégradations dû à l'hétérogénéité de la PdD entre les espèces dans le cas d'un gradient de détection quadratique. Il apparaît néanmoins que dans mes simulations cette hétérogénéité joue un rôle modéré sur les dégradations des ordinations des deux premiers axes de projections.

#### 4.2.4 Bilan pour les ACP

En plus de la sous-estimation des abondances que provoque une  $PdD < 1$ , je mets en évidence deux effets de la  $PdD$  sur les abondances des espèces qui ont des conséquences sur les résultats des ACP.

- (i) Un effet d'attraction qui, lorsque l'hétérogénéité de la  $PdD$  entre les sites s'exprime sous la forme d'un gradient, tend à faire covarier les abondances d'une espèce avec le gradient de  $PdD$  qui lui est propre.
- (ii) Un effet de séparation qui en diminuant fortement les abondances des espèces ayant une  $PdD$  faible, sépare dans les résultats de l'ACP les espèces qui ont une  $PdD$  moyenne faible de celle ayant une  $PdD$  moyenne forte

Dans les ACP, en fonction de divers facteurs, la  $PdD$  peut alors provoquer : (1) une simple dégradation des ordinations des espèces sans perdre la variable latente qui y est associée, (2) une dégradation des ordinations totale, c'est-à-dire provoquant la perte de la variable latente, et (3) la création d'une nouvelle ordination pouvant avoir un sens écologique et donc une variable latente artificielle. Les deux premiers axes de chaque analyse sont impactés par ces dégradations.

Je note également que l'hétérogénéité de la  $PdD$  entre les espèces est parfois un facteur d'augmentation de ces dégradations. Que les dégradations sont plus importantes lorsqu'une ordination est le résultat du cumul d'une ordination « réelle » et de l'ordination artificielle provoquée par l'effet d'attraction. Enfin, que plus les probabilités de détection des espèces et les parts d'inertie des axes porteurs des ordinations d'origine sont faibles, plus les dégradations sont importantes.

Je reviens sur l'ensemble de ces points dans la partie discussion (cf. partie 4.4)

### 4.3 Analyse des résultats des effets de la probabilité de détection dans les AFC

#### 4.3.1 Les SIMULATIONs sans gradient de $PdD$

Dans les SIMULATIONs **sans gradient de  $PdD$** , on remarque que les dégradations des ordinations des AFCnd apparaissent sur les deux axes de projection et sont d'autant plus importantes que la  $PdD$  est faible. Elles restent cependant limitées hormis sur les axes d'ordre 2 lorsque la  $PdD$  est proche de 0.1. En effet les coefficients de corrélation de Kendall entre les ordinations sur les axes d'ordre 1 des AFCnd et des AFCd affichés dans la figure 26 (page 57) sont élevés, même si dans tous les cas une espèce au moins change de position par rapport à une autre.

De plus ces dégradations sont plus importantes pour les ordinations portées par les axes d'ordre 2.

Bien que réduite, ces dégradations sont elles pour autant négligeables ? Je reviendrai sur cette question dans la partie discussion.

On note également que l'hétérogénéité créée en imposant une  $PdD$  différente à chaque espèce (ENSEMBLE 2 de SIMULATIONs) est un facteur qui seul, aggrave peu les biais.

Je retiens donc, comme première information, que la seule sous-estimation des abondances a des conséquences souvent minimales sur l'ensemble des axes de l'analyse. Ainsi, lorsqu'une espèce a la même  $PdD$  sur tous les sites, sa répartition en proportion sur ces sites change peu. Ainsi, une hétérogénéité de la  $PdD$  uniquement entre les espèces a peu d'impact sur les résultats des AFC. Bien entendu, le cumul d'une faible inertie expliquée par l'axe porteur de l'ordination lors l'AFCnd avec soit une très forte hétérogénéité de  $PdD$  entre les espèces, soit avec une  $PdD$  moyenne des espèces très faible (proche de 0.1) remettrait en cause l'idée d'impacts limités sur les ordinations.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

### 4.3.2 Les SIMULATIONs avec des gradients de détection linéaires

Les SIMULATIONs avec **des gradients de détection linéaires** présentent des dégradations variables et pouvant être importantes. Elles portent toujours sur les deux axes de projection de chaque analyse et sont d'autant plus importantes que la PdD moyenne ( $d_{moy,j}$ ) est faible. J'identifie principalement l'effet d'attraction du gradient de la PdD comme responsable de ces dégradations.

Le fait de simuler un gradient de PdD provoque l'apparition de l'effet d'attraction. Dans le cas de distributions des espèces uni-modales, le gradient de PdD linéaire et croissant en fonction de  $g_1$  tend, quand l'espèce possède un optimum sur  $g_1$ , à décaler cet optimum vers de plus fortes valeurs. Quand l'espèce n'a pas d'optimum sur cette variable, l'effet d'attraction lui en crée un à l'extrémité du gradient.

Les conséquences sur les résultats changent si on applique ou non le même gradient de PdD linéaire à toutes les espèces. Lorsque toutes les espèces ont exactement le même gradient de PdD, les dégradations des ordinations sont un peu plus importantes et évoluent de manière identique à celles observées lors des simulations sans gradient de PdD. On en déduit que si, sur un site toutes les espèces ont une même PdD, les proportions de chaque espèce sur le site changent peu, donnant ainsi des dégradations limitées (propriété de la distance du  $\chi^2$ ). La différence de PdD, entre les sites, augmente cependant légèrement les dégradations. Le décalage de l'optimum est pour ainsi dire presque le même pour toutes les espèces, ce qui permet d'avoir une ordination dégradée proche de l'initiale, dès lors que la PdD moyenne et la part d'inertie expliquée de l'axe porteur de l'ordination ne sont pas trop faibles.

Cependant, dès lors que l'on introduit en plus (dans la PdD) une hétérogénéité entre les espèces, soit en n'appliquant le gradient de PdD qu'à 11 espèces sur 20 (ENSEMBLE 4), soit en donnant à chaque espèce un gradient différent (ENSEMBLE 5), les dégradations augmentent, allant parfois jusqu'à la dégradation totale de l'ordination et donc la perte de la variable latente. L'expression simultanée des hétérogénéités de la PdD entre les espèces et entre les sites est donc un facteur important d'aggravation des biais. Puisque les gradients sont différents entre les espèces, les déplacements de leurs optimums sur  $g_1$  le sont également. De nouvelles ressemblances entre les espèces vis-à-vis de leur répartition « altérée » en proportions sur les sites apparaissent alors dans les AFCd. Ce phénomène est mis en évidence dans la figure 30 (page 62) : les AFCd tendent à séparer les espèces en fonction de leur PdD moyenne (forte ou faible) et non plus en fonction de la position réelle de leur optimum. Les valeurs de PdD moyenne ne sont pas les seuls facteurs jouant sur l'importance du déplacement de l'optimum des espèces (abondances initiales des espèces, pentes des gradients etc.) ; ceci explique que les valeurs de corrélations entre les ordinations des espèces sur les axes des AFCd et les valeurs de PdD moyennes des espèces ne sont jamais importantes dans la figure 30.

Cette nouvelle ordination (issue du déplacement des optimums) a les mêmes conséquences sur les résultats que ceux décrits dans le paragraphe 4.2.1 (gradient de PdD linéaire pour les ACP), en dehors de l'effet taille qui est uniquement observé lors des simulations pour les ACP. Je note donc qu'à part d'inertie expliquée par les axes porteurs des ordinations équivalente (lors des AFCnd), une ordination mettant en évidence une variable latente non corrélée avec le gradient de la PdD semble moins dégradée qu'une ordination mettant en évidence une variable latente corrélée avec le gradient de la PdD.

### 4.3.3 Les simulations avec des gradients de détection quadratiques « centrés »

Les SIMULATIONs des gradients **de détection quadratiques centrés** affichent des dégradations des résultats encore différentes mais concernant toujours les deux axes de projection de chaque analyse.

Les simulations où les espèces ont toutes le même gradient de PdD (ENSEMBLE 6) nous montrent des dégradations croissantes avec la diminution de la PdD maximale et plus importantes sur les axes d'ordre 2. Ces dégradations restent faibles sur les axes d'ordre 1 quand la PdD maximale est supérieure à 0.1. Comme lors des SIMULATIONS où toutes les espèces ont le même gradient de PdD linéaire (ENSEMBLE 3), le fait d'avoir sur chaque site la même PdD pour toutes les espèces limite les dégradations. On remarque néanmoins que les dégradations sont plus importantes lorsque les espèces ont toutes le même gradient de PdD quadratique que lorsqu'elles ont toutes le même gradient de PdD linéaire. Ceci montre que la forme du gradient de la PdD n'est pas sans conséquence ; ceci parce deux formes différentes impliquent des déplacements des optimums des espèces différents. En effet si un gradient de PdD linéaire tend à déplacer les optimums des espèces dans le même sens (vers de plus fortes valeurs de  $g_1$ ), un gradient de détection quadratique centré tend à rapprocher ces optimums de la valeur de  $g_1$  pour laquelle la PdD est maximale (au centre du gradient).

Là encore, les dégradations d'une ordination sont d'autant plus fortes que la part d'inertie expliquée par l'axe porteur de l'ordination dans l'AFCnd est faible.

Lorsque l'on applique des gradients différents aux espèces, mes indicateurs mettent en évidence un effet de « spécialisation ». La forme parabolique convexe du gradient de la PdD a tendance à « spécialiser » les espèces (réduit la valence), en réduisant fortement les abondances sur les sites qui sont en dehors des zones de détection optimale. Ce phénomène est mis en évidence dans la figure 35 (page 70): les AFCd tendent à séparer les espèces en fonction de leur PdD maximales (forte ou faible).

Si on se concentre sur les SIMULATIONS où seulement 11 espèces sur 20 ont un gradient de PdD (ENSEMBLE 7), ces séparations des espèces ne peuvent plus être expliquées par le déplacement des optimums des espèces. Les espèces non soumises au gradient de PdD semblent être plus « généralistes » que les autres. Elles tendent alors à se rassembler sur l'axe car leurs répartitions en proportion sur les différents sites se ressemblent.

De plus, cette séparation des espèces « généralistes » des espèces plus « spécialistes » est une nouvelle ordination indépendante de l'ordination des espèces selon leurs optimums sur  $g_1$ . Cette nouvelle ordination, artificielle, s'exprime donc dans les AFCd en fonction de la part d'inertie expliquée par l'axe qui la porte. On la retrouve sur les axes d'ordre 1 lorsque ces derniers n'expriment pas de variables latentes lors des AFCnd (profil 1), et principalement sur les axes d'ordre 2 dans les autres cas. L'apparition de cette ordination artificielle sur un axe des AFCd provoque la relégation sur un axe d'ordre inférieur et la dégradation de l'ordination initialement présente sur cet axe dans l'AFCnd.

Lorsque toutes les espèces ont un gradient de PdD différent (ENSEMBLE 8), le même mécanisme est à l'œuvre. Dans mes simulations cependant, cette ordination artificielle est portée par un axe ayant une part d'inertie expliquée très variable et assez faible (paramètres de simulations). Ceci donne des dégradations variables mais pouvant faire perdre complètement les ordinations des axes d'ordre 2 des AFCnd dans le premier plan de projection des AFCd.

Par ailleurs, on s'attendrait à retrouver l'effet d'attraction obtenu dès lors que la PdD s'exprime sous la forme d'un gradient et qui tendrait sur  $g_1$ , à déplacer l'optimum de l'espèce quand il existe, ou à en créer un, au centre du gradient. Cependant les indicateurs que j'ai choisis ne permettent pas de mettre en évidence ce fait. Un indicateur spécifique à cette question compléterait notre compréhension des dégradations constatées et des coefficients de corrélations affichés dans la figure 33 (page 67).

En conclusion, le changement de forme de gradient de PdD (de linéaire à quadratique convexe centré), me permet de dire que l'expression simultanée des hétérogénéités de la PdD entre les espèces et les sites est là encore le facteur qui provoque le plus de biais dans les AFC ; et que ces dégradations sont dépendantes de la forme donnée au gradient de PdD.

#### 4.3.4 Les simulations avec des gradients de détection quadratiques « décalés »

Ces dernières simulations nous montrent avant tout que les gradients de la PdD peuvent induire une variable latente prépondérante.

Nous retrouvons l'effet d'attraction, qui avec ces gradients quadratiques décalés tendent à donner des abondances uni-modales, centrées sur les valeurs de  $g_1$  pour lesquelles les espèces ont leurs PdD maximales. De plus, les PdD maximales sont ordonnées sur  $g_1$ , et fonction de l'optimum écologique de l'espèce quand il existe. C'est entre autre cette nouvelle ordination que les AFCd ont tendance à faire ressortir.

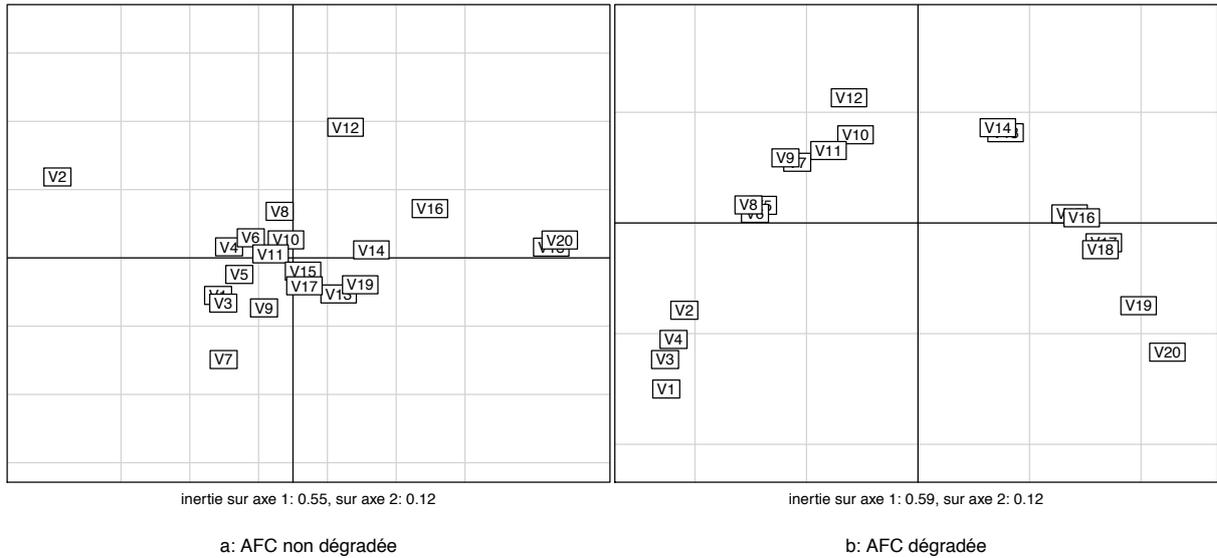
Si, dans les versions non dégradées, **aucune variable latente** ne correspond à la variable  $g_1$ , il en apparaît une dans les versions dégradées. On a donc apparition d'une variable latente artificielle. Si aucune variable latente n'existe dans l'AFCnd (cas des SIMULATION où j'ai simulé des espèces sans optimum (profil 1)), la variable latente artificielle se place dans l'AFCd sur l'axe d'ordre 1. Si une variable latente existe dans l'AFCnd (qui correspond donc à la variable  $g_2$ ), la variable latente artificielle se placera sur l'axe d'ordre 1 ou 2 en fonction de la part d'inertie relative expliquée par chacune des variables. Dans tous les cas l'ordination de l'AFCnd est dégradée. Dans la majorité des cas, elle est reléguée sur l'axe d'ordre 2 (lorsque la variable latente artificielle prend sa place). Ceci signifie que l'axe porteur de la nouvelle ordination possède une grande part d'inertie expliquée.

Si dans les AFCnd il y a déjà **une variable latente** qui correspond à la variable  $g_1$ , alors cette variable latente réapparaît dans les AFCd. Cependant l'ordination obtenue cumule l'ordination de l'AFCnd et l'ordination créée par le gradient de la PdD. L'ordination d'origine est donc dégradée. Dans les cas où en plus de la variable latente qui correspond à la variable  $g_1$ , l'AFCnd met en évidence une seconde variable latente (correspondant à la variable  $g_2$ ), cette seconde variable latente est dans la majorité des cas perdue dans le premier plan de projection de l'AFCd sans que je puisse dire si cette variable est reportée ou non sur les axes d'ordre inférieur.

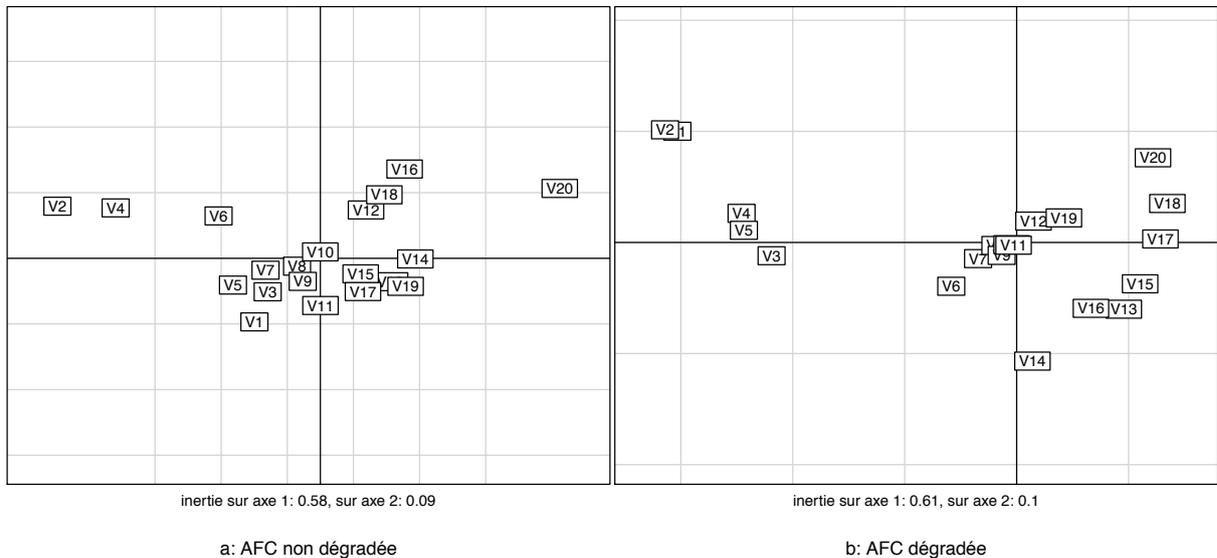
La spécificité de ces gradients de détection décalés, est l'expression importante de l'effet de « spécialisation » (déjà évoqué dans le cas des gradients quadratiques centrés). Cet effet est révélé par les résultats de la répartition des groupes des 10 premières et des 10 dernières espèces lors des SIMULATIONs où dans chacune d'elles toutes les espèces ont le même gradient de PdD (figure 41 page 77). Dans cette figure, les quatre profils d'abondances présentent des résultats similaires pour les axes d'ordre 2 des AFCd (graphique (d)). Cela indique que dans les AFCd, la variable latente identifiée par l'ordination des espèces sur l'axe d'ordre 1 ( $g_1$ ), impose également une ordination des espèces sur l'axe d'ordre 2. Cette variable latente, devenue prépondérante, induit un effet Guttman (cf. paragraphe 1.2.3 page 18), comme nous le montre la figure 44 page suivante.

Enfin, l'hétérogénéité de la PdD lorsque seulement 11 espèces sur 20 ont un gradient de PdD (les 9 restantes ayant une PdD constamment égale à 1) (ENSEMBLE 10) apporte un biais supplémentaire. Il s'observe dans les résultats pour le profil 1 d'abondance de cet ENSEMBLE de SIMULATIONs, donnant le coefficient de corrélation de Kendall entre les ordinations des espèces sur les axes des AFCd et la PdD maximale des espèces (graphique (b) de la figure 40 page 76). Il se traduit par la séparation, sur l'axe d'ordre 2, des espèces en fonction de leur PdD maximale. C'est une autre expression de l'effet de « spécialisation ». On retrouve la séparation des espèces « spécialistes » des espèces « généralistes ». Il est alors surprenant de ne pas l'identifier dans les autres profils d'abondances. Je le suppose cependant présent dans toutes les simulations portant ce type d'hétérogénéité ; ainsi que la figure 45 tend à le montrer. Il me faudrait un nouvel indicateur pour confirmer cette hypothèse tel que le calcul du taux de corrélation de Kendall entre les valeurs moyennes de PdD des espèces et l'ordination des espèces en fonction de leurs distances à l'origine du plan de projection.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC



*Figure 44 : graphiques des résultats de la millièmes simulation de la SIMULATION 3293. Dans cette simulation, toutes les espèces ont une abondance moyenne dépendante de deux variables environnementales ( $g_1$  et  $g_2$ ) ; et dans le cas de l'AFC dégradée (b), toutes les espèces ont un gradient de PdD quadratique « décalé » dont la PdD maximale est égale à 0.3. À gauche (a) le plan de projection de l'AFCnd, et à droite (b) le plan de projection de l'AFCd. Dans chaque plan de projection on retrouve les 20 espèces (numérotées de V1 à V20). Dans le cas de l'AFCnd (a), la première composante principale correspond à  $g_1$  et la deuxième composante principale à  $g_2$ . L'AFCd présente clairement un effet Guttman montrant la prépondérance de la variable latente artificielle de l'axe d'ordre 1 dû au gradient de la PdD.*



*Figure 45 : graphiques des résultats de la millièmes simulation de la SIMULATION 3305, Dans cette simulation, toutes les espèces ont une abondance moyenne dépendante de deux variables environnementales ( $g_1$  et  $g_2$ ) ; et dans le cas de l'AFC dégradée (b), 11 espèces sur 20 ont un gradient de PdD quadratique « décalé » dont la PdD maximale est égale à 0.5. Les 9 autres espèces ont une PdD constante et égale à 1. À gauche (a) le plan de projection de l'AFCnd, et à droite (b) le plan de projection de l'AFCd. Dans chaque plan de projection on retrouve les 20 espèces (numérotées de V1 à V20). Dans le cas de l'AFCnd (a), la première composante principale correspond à  $g_1$  et la deuxième composante principale à  $g_2$ . Dans l'AFCd, les espèces 6, 7, 8, 9, 10, 11, 12, 19 et 20, sont regroupées et correspondent aux espèces qui ont une PdD constamment égale à 1.*

## 4.4 Discussion.

### 4.4.1 Sur les résultats.

Je rappelle ma problématique. J'avais centré mon travail autour des questions suivantes :

- (i) Une probabilité de détection inférieure à 1 (PdD) mais constante entre espèces et sites modifie-t-elle les ordinations dans les ACP et ou les AFC ?
- (ii) Si oui, une PdD hétérogène entre les sites ou/et entre les espèces est-elle un facteur d'aggravation?
- (iii) Si une PdD constante entre espèces et entre sites ne modifie pas les ordinations dans les analyses, une hétérogénéité de la PdD entre sites ou/et entre espèces est-elle à elle seule un facteur de dégradation des ordinations ?

La méthodologie mise en place et les simulations effectuées à partir de celle-ci, me permettent de répondre en grande partie aux questions posées :

- (1) Une PdD inférieure à 1, constante entre espèces et sur tous les sites, provoque des dégradations dans les résultats des ACP et des AFC.
- (2) L'hétérogénéité de la PdD (entre sites ou espèce) ne diminue jamais les dégradations et les augmente dans la grande majorité des cas.

Identifions maintenant **les facteurs de dégradations**. Si les simulations présentent toutes des dégradations, elles nous montrent également que ces dégradations sont très variables. On peut alors essayer de faire ressortir quelques facteurs pouvant expliquer cette variabilité.

Tout d'abord, une ordination initiale (je réutilise cette dénomination par la suite), obtenue sans problème de détection des espèces (PdD=1), est d'autant plus dégradée par une PdD<1 que la part d'inertie expliquée par l'axe portant l'ordination initiale est petite.

Par ailleurs, le facteur de dégradation des ordinations qui apparaît lorsque la PdD est constante entre les espèces et les sites, est la valeur de la PdD elle-même. Plus celle-ci est faible, plus les dégradations des ordinations sont importantes. En effet, plus les effectifs réels sont sous-estimés sur le terrain, moins les évolutions des abondances sont visibles dans les données recueillies, et moins les analyses AFC et ACP peuvent faire ressortir les ordinations initiales. Cependant, on peut remarquer que les AFC sont moins sensibles à ce facteur que les ACP. Ceci sans doute parce que les AFC comparent des proportions et parce que lorsque la détection est constante ces proportions sont globalement conservées (mis à part une PdD très faible).

Dès l'introduction d'une hétérogénéité entre les espèces ou les sites, de nouveaux facteurs de dégradation apparaissent. Organiser l'hétérogénéité de la PdD entre les sites suivant un gradient, tout comme introduire une hétérogénéité entre les espèces, génère une ou des ordinations artificielles. Ce sont ces nouvelles ordinations que les analyses tendent à faire plus ou moins ressortir en fonction de deux facteurs principaux:

d'une part l'ordination artificielle est-elle ou non corrélée à une variable latente réellement explicative de la répartition des espèces ? Si oui, l'ordination liée à cette variable latente n'en sera que plus dégradée. Les variables latentes réelles non corrélées à l'ordination artificielle seront moins impactées.

d'autre part, quelle est la part d'inertie expliquée par l'axe portant l'ordination artificielle ? Plus celle-ci est grande, plus cette ordination artificielle devient prépondérante. Je ne peux formuler que des pistes pour expliquer cette part plus ou moins importante d'inertie expliquée des axes portant les ordinations artificielles : pente des gradients, PdD moyenne, interaction entre les deux.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

**Ces biais sont-ils acceptables ?** La réponse à cette question dépend bien évidemment de l'importance des dégradations mais également de la précision de l'information recherchée par la personne qui réalise l'analyse. En effet, toutes les ACP ou les AFC n'ont pas toutes le même objectif en termes de niveaux de précision. La simple mise en évidence de variables écologiques expliquant la répartition des espèces tel que présenté dans le travail de Denis (2009) requiert le plus bas niveau de précision. L'information qui demande le plus haut niveau de précision s'attache aux coordonnées des sites sur les variables latentes comme dans l'article de Popy, Bordignon, et Prodon (2009). J'analyserai l'importance des dégradations au regard de ces deux extrêmes.

Pour la simple mise en évidence de variables écologiques, on distingue trois sources d'erreurs possibles.

(1) la PdD peut engendrer la perte d'une variable latente. Sans prise en compte de la PdD, il est possible de méconnaître des variables expliquant la distribution des espèces.

(2) les variables latentes apparaissant dans les analyses dégradées peuvent être totalement ou très grandement dues aux variations de la PdD elle-même. Dans ce cas, la variable latente est artificielle, n'explique pas la répartition des espèces et conduit à de fausses interprétations.

(3) Enfin, la PdD peut modifier la part d'inertie relative des variables latentes. Ceci conduit à mal estimer l'importance relative des variables latentes.

Néanmoins, en l'absence d'apparition de variables artificielles dans les analyses dégradées, les ordinations présentes dans ces analyses, lorsqu'elles sont encore interprétables, mettent en évidence des variables expliquant réellement la répartition des espèces. Dès lors, l'absence, réelle ou hypothétique, d'une hétérogénéité de la PdD entre espèces pour les ACP, et entre espèces ou entre sites pour les AFC permet de dire qu'une variable environnementale mise en évidence par l'analyse existe réellement.

Dans les autres cas il est impossible de faire la part entre vraie et fausse variable latente, sans connaître la PdD et ses variations spatiales et entre espèces ; informations rarement disponibles, même si l'apparition d'un effet taille (ACP) ou d'un effet Guttman (AFC) peut alerter sur la présence d'un biais. Dès lors, si la PdD n'est pas prise en compte, j'exprime un doute sur toute variable latente issue d'une ACP ou d'une AFC.

Pour la mise en évidence de la position relative et précise des sites ou des espèces sur les variables latentes mises en évidence, je constate que **toutes les simulations** ont montré des changements - parfois minimes - entre les ordinations des analyses non dégradées et dégradées. Ces changements d'ordinations ont lieu sur les deux axes du premier plan de projection. Dans mes simulations, des changements d'ordinations semblent négligeables, mais d'autres ne le sont pas ; aussi bien dans les ACP que dans les AFC. Ainsi prendre en compte la PdD apparaît obligatoire pour toute ACP ou AFC, sauf à démontrer ou supposer que ces effets sont négligeables. Dans tous les cas, cette démonstration en nécessite la connaissance.

Mon étude ne me permet pas de faire la liste des conditions à réunir pour que la PdD soit considérée négligeable. De plus, une telle liste serait prématurée au regard du nombre de facteurs en jeu et de leurs interactions possibles.

Néanmoins, les simulations ont permis de mettre en évidence les effets d'attraction (cf. partie 4.2.2, 4.2.3, 4.3.2 et 4.3.4) et de spécialisation (cf. partie 4.3.3 et 4.3.4) dus à une hétérogénéité de la PdD entre sites, ainsi que l'effet de séparation (cf. partie 4.2.2) dû à une hétérogénéité de la PdD entre espèces. Elles ont également permis de montrer que l'hétérogénéité de la PdD entre espèces impliquait des effets d'attractions et de spécialisations différents pour chaque espèce, ceci aggravant les dégradations des ordinations. Ainsi, plus ces effets seront limités, plus les dégradations seront négligeables. Les éléments qui limitent ces effets sont de deux ordres : les éléments propres à la distribution des espèces, et les éléments propres à la PdD.

Pour ce qui concerne la distribution des espèces, plus les espèces seront spécialisées (faible valence), moins les effets d'attractions et de spécialisations pourront s'exprimer.

Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Pour ce qui concerne la PdD, plus l'hétérogénéité entre espèces et/ou entre sites sera faible (gradient à faible pente) plus les dégradations seront négligeables. Ces deux conditions devraient être suffisantes pour limiter les dégradations, car en cas de très faible PdD, les analyses ne devraient plus fournir des ordinations interprétables.

Je ne suis cependant pas en mesure de donner des valeurs pour définir « faible valence » et « faible hétérogénéité ». D'autant plus que ces valeurs sont interdépendantes.

Dans la pratique, les études démontrent dans la quasi-totalité des cas une hétérogénéité de la PdD entre espèces et sites (cf. chapitre 1). Or, ainsi que démontré, cette hétérogénéité est un facteur important d'aggravation des dégradations. En conséquence, les conditions dans lesquelles la PdD serait négligeable dans les analyses AFC et ACP sont probablement rarement réunies, au moins dans le cas des ACP. En effet les ACP sont plus sensibles que les AFC aux effets de la PdD et demanderont donc le respect de conditions plus contraignantes. En l'état actuel des choses, les AFC doivent donc être privilégiées pour analyser des matrices d'abondance ou de présence absence.

Pour synthétiser mes résultats, je peux évoquer MacKenzie et al. (2002) et MacKenzie et Nichols (2004) qui parlent du dilemme face à l'évolution de l'abondance d'une espèce pour laquelle la PdD n'a pas été prise en compte. Finalement, nous sommes ici en présence de la même question : les résultats reflètent-ils les évolutions des abondances, les évolutions des probabilités de détection, ou encore un mélange des deux ? Manifestement toute ACP ou AFC nécessite une approche de la PdD pour lever cette ambiguïté.

Au-delà de ces analyses, assurément sensibles à la PdD, la question se pose désormais légitimement pour les autres analyses multidimensionnelles. En effet, toutes reposent sur des approches similaires.

#### 4.4.2 Sur la méthode.

La démarche que j'ai utilisée, ainsi que les indicateurs choisis, permettent de mettre en évidence l'existence de la dégradation des résultats des ACP et des AFC causée par la non prise en compte de la PdD. Cependant mes simulations ne concernent que certaines formes de gradients de PdD. Ceci pose la question des effets des autres formes de gradients. Il est impossible de tester toutes les formes imaginables. Il me semble cependant que la forme qui fait le plus défaut dans mon étude est celle où la PdD serait totalement aléatoire. Une telle PdD pourrait être rencontrée par exemple lorsque plusieurs observateurs, chacun prospectant des sites différents, n'ont pas la même habileté pour détecter certaines espèces. En effet, sans gradient, la PdD ne structurera pas les données d'abondance comme nous l'avons vu. Je suppose donc que les analyses ne pourront pas produire d'ordinations artéfactuelles ; mais cependant, que les sous-estimations aléatoires des abondances des différentes espèces ne seront pas sans conséquence sur les ordinations existantes. Je justifie mon hypothèse par le fait qu'une PdD constante entre les sites et les espèces implique déjà des dégradations, et que les hétérogénéités testées (entre sites ou/et entre espèces) les augmentent. De plus cette PdD implique une hétérogénéité entre sites et entre espèces.

Néanmoins j'ai choisi des formes de gradient pouvant avoir des raisons écologiques et pouvant donc être rencontrées sur le terrain. Aucune des formes que j'ai testées n'est sans conséquence. De plus, certaines formes de variabilité de PdD me semblent être potentiellement encore plus néfastes que celles que j'ai testées. C'est le cas par exemple pour une probabilité de détection quadratique concave, ou un mélange de gradients convexe et concave. On retrouve par exemple ces formes pour la PdD des oiseaux en fonction des jours de prospections lors de leur période de reproduction (Annexe A dans Tingley et Beissinger (2013)).

Une autre critique de ma méthode porte sur les indicateurs utilisés pour caractériser les différences entre analyses dégradées et non dégradées. Les indicateurs que j'ai choisis présentent des limites. Chacun ne s'appuie que sur une caractéristique particulière de l'analyse. Aucun indicateur ne

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

permet de couvrir toutes les lectures possibles de ces analyses. Je me suis concentré sur la perte ou non des ordinations que fournissent les analyses non dégradées et sur l'importance de la probabilité moyenne ou maximale de détection. Ce choix permet de mettre en évidence l'existence des dégradations des ordinations mais ne permet pas la totale compréhension des mécanismes sous-jacents à l'émergence de ces dégradations. D'autres indicateurs pourraient apporter des informations complémentaires. Ainsi un indicateur se concentrant sur les espèces proches les unes des autres dans le plan de projection, et sur les raisons de cette proximité, aurait été utile pour mettre en évidence les effets de « spécialisation » des espèces. Un exemple d'indicateur serait de faire une classification des espèces à partir des coordonnées des espèces dans le premier plan de projection de chaque analyse (voir à partir des trois premières axes), puis relever le nombre de fois (sur les 1000 d'une SIMULATION) où la régression logistique entre l'appartenance des espèces à une classe et l'interaction de la valeur moyenne de la PdD avec la pente du gradient de la PdD est significative par rapport au modèle nul.

De plus, j'ai constaté qu'il serait intéressant de regarder au moins l'axe d'ordre 3 de chaque analyse. En effet, ceci permettrait de voir si les messages présents sur les axes d'ordre 1 ou 2 se retrouvent parfois relégués sur celui-ci après dégradation.

Si une limite de ma méthode se trouve dans les indicateurs utilisés, une autre se trouve dans la manière de simuler les gradients de PdD. Si les gradients linéaires peuvent être simulés en jouant indépendamment sur la pente et sur les coordonnées à l'origine, je n'ai pas choisi cette possibilité pour les gradients de détection quadratiques. Même si ceci ne fait pas partie des questions que je me suis posées au départ de cette étude, en imposant le changement de la pente du gradient en cas de changement du maximum de PdD d'une espèce, je me suis interdit la possibilité de tester l'importance respective de ces deux facteurs ou de leur interaction dans l'expression des dégradations. Je suppose que cette interaction joue un rôle important. Ceci ne m'empêche pas de répondre à mes questions, mais imposera une approche différente des simulations pour qui voudrait approfondir la question des facteurs de dégradations.

Enfin, il importerait de mieux définir, si possible, les cas où la PdD serait négligeable dans les analyses AFC et ACP. Cela imposerait de donner des critères et des limites claires à une dégradation acceptable, chose qui ne m'était pas nécessaire pour répondre à ma problématique. De plus, l'évaluation de l'acceptabilité de la dégradation des ordinations ne peut se faire qu'en regard de l'utilisation des résultats de l'analyse (cf. partie 4.4.1).

### 4.4.3 Sur des solutions potentielles.

En dehors des cas évoqués où la PdD provoque des dégradations des ordinations acceptables aux vues de l'information recherchée par la personne qui réalise l'analyse, existe-t-il des solutions pour obtenir des ordinations sans dégradations ?

J'ai montré dans mes simulations l'apparition de l'effet taille dans certaines ACP. L'effet taille peut être corrigé par un double centrage de la matrice des abondances. Dans le cas des ACP, uniquement lorsque les espèces présentent un même gradient (aucune hétérogénéité entre les espèces), cette correction fera que les variables latentes (mise en évidence par les ordinations) correspondront à des variables écologiques existant réellement. On retrouve alors la possibilité d'utiliser les résultats de l'analyse si la personne se contente de mettre en évidence une variable écologique expliquant la répartition des espèces. Ainsi, la correction apportée par cette méthode n'a de sens que pour des cas précis et limités, et les ordinations sur les axes restant dégradées après cette correction, certaines variables latentes peuvent ne pas être identifiées.

Dès lors, la correction des abondances en fonction de la PdD, par des méthodes existantes (cf. chapitre 1) me semble être une, voire LA solution. Etant sur des suivis multi-sites et multi espèces,

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

les techniques de CMR sont inenvisageables (trop coûteuses en effort et budgétairement). Restent le distance-sampling (Buckland et al. 1993) et les N-mixture multi-espèces (Yamaura et al. 2011). La démarche serait alors d'estimer les abondances de l'ensemble des espèces sur tous les sites, donc à avoir des estimations corrigées du point de vue de la PdD à l'aide de ces techniques, puis de pratiquer l'analyse (ACP ou AFC) sur ces estimations corrigées.

Dans le cas où les espèces peuvent être suivies par cette méthode du distance-sampling, cette méthode semble la plus efficace pour corriger les effets de la PdD. L'estimation de la densité de 18 espèces de passereaux communs dans la réserve naturelle du val d'Allier par l'utilisation du distance-sampling le long de neuf line-transects entre le mois de décembre 2015 et le mois de janvier 2016 (Guélin 2016), montre que les abondances corrigées de plusieurs espèces sur plusieurs sites peuvent être obtenues à l'aide de cette technique.

Pour le N-mixture, l'intérêt réside dans la modélisation de la PdD à l'aide de covariables. Ceci permettant d'évaluer la PdD à partir d'un simple échantillonnage. Cependant les travaux de Couturier et al. (2013), mais aussi d'autres travaux récents (Dennis, Morgan, et Ridout 2015) mettent en évidence que ces méthodes sont pour le moment à manier avec précaution notamment pour des PdD inférieures à 0,5. Pour Royle (2004), la PdD limite serait proche de 0,3.

Une autre piste serait d'explorer les méthodes d'ordination non contraintes, basées sur des modèles (« Model-based ») (, Hui et al. 2015, Hui 2015). Ces approches sont des extensions des Modèles Linéaires Généralisés (GLMs) pour les données multi-espèces et multi-sites. Il doit être possible d'intégrer dans ces approches le modèle de N-mixture (Royle 2004) qui se base sur une loi de vraisemblance, donnant la probabilité d'avoir un nombre donné d'individus d'une espèce sur un site, prenant en compte la probabilité de détection.

Théoriquement, toutes les études utilisant une ACP ou une AFC nécessiteraient une prise en compte de la PdD. Ceci peut se faire actuellement soit par le biais de méthodes estimant la PdD (N-mixture) soit par des méthodes qui corrigent ces effets (distance-sampling). Reste le possible développement de nouvelles méthodes. Dans tous les cas, il semble indispensable d'approcher d'une manière ou d'une autre la PdD en tant que telle dans les protocoles de comptage afin que les ACP et les AFC soient plus fiables.

## Conclusion

Jusqu'à présent, et dans le meilleurs des cas (à ma connaissance), seules les données d'abondances des espèces ayant de très faible probabilité de détection (PdD) sont retirées des tableaux d'abondances avant d'y appliquer une ACP ou une AFC. Ainsi les ordinations fournies par ces analyses sont supposées exactes sans prendre en compte les potentiels effets de la PdD sur les résultats des abondances des espèces conservées et donc sur les résultats de ces analyses. Or, des simulations de dégradations des données de tableaux d'abondances avant l'application des méthodes ACP et AFC, ont permis de démontrer que ces méthodes sont sensibles aux effets de la probabilité de détection. Cette sensibilité se traduit par des dégradations ou des biais pouvant être très importants, allant jusqu'à masquer les variables latentes réelles des résultats, parfois en les remplaçant par des variables artéfactuelles. Les autres dégradations sont des changements dans les ordinations pouvant conduire à des conclusions erronées sur les milieux et les espèces suivies.

Les dégradations engendrées par la PdD dans les résultats de ces analyses sont variables et dépendent de nombreux facteurs. Mes résultats font écho à ceux présentés par Gu et Swihart (2004) sur les biais engendrés par la probabilité de détection sur les analyses de richesse. **Il est donc difficile, pour ne pas dire impossible, de prédire ces dégradations et de les corriger sans avoir estimé la probabilité de détection de toutes les espèces sur tous les sites prospectés en amont.**

En conséquence, sans l'examen des effets de la PdD des espèces sur les résultats d'une ACP ou d'une AFC, des hypothèses fortes sur la probabilité de détection sont actuellement nécessaires pour pouvoir affirmer qu'une variable environnementale mise en évidence par l'analyse existe réellement (la position des espèces dans l'ordination restant malgré tout incertaine). Dans le cas des ACP, l'hypothèse est qu'il n'y a aucune hétérogénéité de PdD entre les espèces et nécessite en plus un double centrage de la matrice d'abondance. Dans le cas des AFC, l'hypothèse est qu'il n'y a soit aucune hétérogénéité de la PdD entre les espèces, soit aucune hétérogénéité de la PdD entre les sites. De futures études devraient permettre de relâcher quelques unes de ces hypothèses.

De plus, l'hétérogénéité de la probabilité de détection entre sites ou entre espèces est un facteur d'aggravation des dégradations des ordinations ; en particulier si elle (l'hétérogénéité) possède une structure (spatiale par exemple) pouvant être mise en évidence par une ordination lors des analyses. Ainsi, les cas où les probabilités de détection diffèrent entre les sites et les espèces affichent des dégradations parmi les plus importantes. Or, dans les milieux naturels, une telle hétérogénéité est des plus probables (par exemple entre sites du fait du degré d'ouverture du milieu, et entre espèces du fait de couleurs de robes différentes ou de chants plus ou moins audibles).

Dès lors, sans prise en compte de la probabilité de détection, il est toujours possible de se demander si les résultats reflètent les évolutions des abondances, des probabilités de détection, ou un mélange des deux.

Cependant prendre en compte la PdD de plusieurs espèces sur plusieurs sites, dans plusieurs habitats et à plusieurs dates, ne va pas sans poser de problème de faisabilité, tant sur l'élaboration de protocoles adaptés que sur leur mise en œuvre et leur coût. Cette prise en compte apparaît néanmoins souhaitable lors de l'utilisation de ces analyses multidimensionnelles (ACP et AFC) sur des tableaux d'abondance. De plus, au vu des études conduites sur la PdD, aucun groupe d'espèces ne peut faire exception. Si cette prise en compte ne peut se faire, les hypothèses faites sur la PdD doivent être affichées, et la potentielle fragilité des résultats énoncée et discutée.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

Cette sensibilité à la PdD des ACP et des AFC appliquées à des tableaux d'abondances interroge sur celle de ces mêmes analyses appliquées à des tableaux de présences, mais aussi sur celle des autres analyses multidimensionnelles comme les analyses de coinertie. Elle ouvre donc la porte à de nouveaux travaux, dans un futur proche, pour continuer à explorer son impact dans des situations diverses, mais surtout pour développer les outils et méthodes permettant de corriger les dégradations induites.

## Bibliographie

- Allredge, M. W., Simons, T. R., & Pollock, K. H. (2007). Factors affecting aural detections of songbirds. *Ecological Applications*, 17 (3), 948-955.
- Bailey, L. L., Simons, T. R., & Pollock, K. H. (2004a). Estimating detection probability parameters for Plethodon salamanders using the robust capture-recapture design. *Journal of Wildlife Management*, 68 (1), 1-13.
- Bailey, L. L., Simons, T. R., & Pollock, K. H. (2004b). Estimating site occupancy and species detection probability parameters for terrestrial salamanders. *Ecological Applications*, 14 (3), 692-702.
- Bailey, L. L., Reid, J. A., Forsman, E. D., & Nichols, J. D. (2009). Modeling co-occurrence of northern spotted and barred owls: accounting for detection probability differences. *Biological Conservation*, 142 (12), 2983-2989.
- Borchers, D. L., Buckland, S. T., & Zucchini, W. (2004). Estimating animal abundance. Closed Populations (2<sup>ème</sup> éd.). Springer-verlag, London. 314 p.
- Buckland, S. T., Anderson, D. R., Burnham, K. P., & Laake, J. L. (1993). Distance Sampling: Estimating Abundance of Biological Populations. Chapman and Hall, London. 446 p.
- Burton, A. C., Sam, M. K., Balangtaa, C., & Brashares, J. S. (2012). Hierarchical Multi-Species Modeling of Carnivore Responses to Hunting, Habitat and Prey in a West African Protected Area. *PLoS ONE*, 7 (5), 14 p.
- Cailliez, F., & Pages, J. P. (1976). Introduction à l'analyse des données. 75019 Paris: Société de mathématiques appliquées et de sciences humaines (SMASH). 616 p.
- Chabanier, O. (2011). Suivis des tendances de population de Lézards ocellés (*Timon lepidus*) dans la steppe de Crau : estimation de la probabilité de détection par radio-téléométrie. Master Professionnel. Université Henry Poincaré, Nancy, 38 p
- Chen, G., Kéry, M., Zhang, J., & Ma, K. (2009). Factors affecting detection probability in plant distribution studies. *Journal of Ecology*, 97 (6), 1383-1389.
- Chen, G., Kéry, M., Plattner, M., Ma, K., & Gardner, B. (2013). Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology*, 101 (1), 183-191.
- Chessel, D., Dufour, A. B., & Thioulouse, J. (2003). Tableaux homogènes. Fiche de biostatistique: BS81 [Cours en ligne]. Consulté à l'adresse <http://pbil.univ-lyon1.fr/R/cours/bs81.pdf>. 23 p.
- Chessel, D., Dufour, A. B., & Thioulouse, J. (2004): The ade4 package-I- One-table methods. R News. 4: 5-10.
- Couturier, T., Cheylan, M., Bertolero, A., Astruc, G., & Besnard, A. (2013). Estimating abundance and population trends when detection is low and highly variable: A comparison of three methods for the Hermann's tortoise: Three Methods for Estimating *T. hermanni* Abundance. *The Journal of Wildlife Management*, 77 (3), 454-462.
- Denis, P. (2009). L'avifaune nicheuse des forêts rhénanes alsaciennes : relations avec le milieu et mise en perspective géographique et historique. Mémoire EPHE, Laboratoire d'Ecologie et Biogéographie des Vertébrés, EPHE, 98 p.
- Dennis, E. B., Morgan, B. J. T., & Ridout, M. S. (2015). Computational aspects of N-mixture models: Computational Aspects of N-Mixture Models. *Biometrics*, 71 (1), 237-246.
- Dorazio, R. M., & Royle, J. A. (2005). Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, 100 (470), 389-398.
- Dorazio, R. M., Royle, J. A., Söderström, B., & Glimskar, A. (2006). Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, 87 (4), 842-854.
- Dorazio, R. M., & Connor, E. F. (2014). Estimating Abundances of Interacting Species Using Morphological Traits, Foraging Guilds, and Habitat. *PLoS ONE*, 9 (4), 9 p.
- Dorazio, R. M., Connor, E. F., & Askins, R. A. (2015). Estimating the Effects of Habitat and Biological Interactions in an Avian Community. *PLoS ONE*, 10 (8), 16 p.
- Dray, S. (1999). Utilisation des listes d'occurrences spécifiques spatialisées en écologie et en biogéographie (D.E.A). Université Claude Bernard Lyon I, Lyon. 30 p.
- Dray, S., & Chessel, D. (2008). L'ordination des tableaux écologiques. Fiches TD avec le logiciel R: Ter 2 [Cours en ligne]. Consulté à l'adresse <http://pbil.univ-lyon1.fr/R/pdf/ter2.pdf>. 26 p.

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

- Dray, S. (2014). Analysing a pair of tables: coinertia analysis and duality diagrams. In *Visualization and verbalization of data* (CRC Press.). J. Blasius and M. Greenacre. 289-300
- Dupiereux, E., Feytmans, E., & Micha, J. C. (1983). Utilisation critique de l'analyse en composante principale et du cluster analysis pour la détermination d'échantillons d'invertébrés benthiques en eau douce. *Oikos*, 40 (1), p. 81-94.
- Escofier, B., & Pagès, J. (2008). Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation, 4<sup>ème</sup> ed., Dunod, Vottem.
- Farnsworth, G. L., Pollock, K. H., Nichols, J. D., Simons, T. R., Hines, J. E., & Sauer, J. R. (2002). A removal model for estimating detection probabilities from point-count surveys. *The Auk*, 119 (2), 414-425.
- Fischesser, B., & Dupuis-Tate, M. F. (2002). Le guide illustré de l'écologie. Édition de la Martinière-Cemagref Édition.
- Forey, E. (2003). Résilience à long terme des communautés végétales herbacées : application à la gestion conservatoire de la réserve nationale des vallées de la Grande-Pierre et de Vitain (Loir et Cher). (D.E.A). Université d'Aix-Marseille III. Institut Méditerranéen d'Écologie et de Paléoécologie, Marseille. 35 p.
- Guélin, F. (2016). Premiers résultats de dénombrement par Distance Sampling des populations hivernantes de passereaux communs dans la réserve Naturelle Nationale du Val d'Allier (03). *Le Grand-Duc*, (84), 86-103.
- Gu, W., & Swihart, R. K. (2004). Absent or undetected ? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, 116 (2), 195-203.
- Hui, F. K. C. (2015). boral-Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in R. *Methods in Ecology and Evolution*, 17 p.
- Hui, F. K., Taskinen, S., Pledger, S., Foster, S. D., & Warton, D. I. (2015). Model-based approaches to unconstrained ordination. *Methods in Ecology and Evolution*, 6 (4), 399-411.
- Hyde, E. J., & Simons, T. R. (2001). Sampling plethodontid salamanders: sources of variability. *Journal of Wildlife Management*, 65 (4), 624-632.
- Iknayan, K. J., Tingley, M. W., Furnas, B. J., & Beissinger, S. R. (2014). Detecting diversity: emerging methods to estimate species diversity. *Trends in Ecology & Evolution*, 29 (2), 97-106.
- Jongman, R. H., Ter Braak, C. J. F., & Van Tongeren, O. F. R. (Éd.). (1995). Data analysis in community and landscape ecology (New ed., with corr). Cambridge University Press, New York.
- Kéry, M., Royle, J. A., & Schmid, H. (2005). Modeling avian abundance from replicated counts using binomial mixture models. *Ecological applications*, 15 (4), 1450-1461.
- Kéry, M., Spillmann, J. H., Truong, C., & Holderegger, R. (2006). How biased are estimates of extinction probability in revisitation studies? *Journal of Ecology*, 94 (5), 980-986.
- Kéry, M., Gardner, B., & Monnerat, C. (2010). Predicting species distributions from checklist data using site-occupancy models: Distribution modelling from checklists. *Journal of Biogeography*, (37), 1851-1862.
- Lebreton, J. D., Chessel, D., Prodon, R., & Yoccoz, N. (1988). L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. I: Variables de milieu quantitatives. *Acta oecologica. Oecologia generalis*, 9 (1), 53-67.
- Legendre, P., & Legendre, L. (1998). *Numerical Ecology*, 2<sup>e</sup> éd., Elsevier, Amsterdam.
- Ligges, U. & Mächler, M. (2003). Scatterplot3d - an R Package for Visualizing Multivariate Data. *Journal of Statistical Software* 8(11), 1-20.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83 (8), 2248-2255.
- MacKenzie, D. I., & Nichols, J. D. (2004). Occupancy as a surrogate for abundance estimation. *Animal biodiversity and conservation*, 27 (1), 461-467.
- Marques, T. A., Buckland, S. T., Borchers, D. L., Rexstad, E. G., & Thomas, L. (2011). Distance Sampling. In « *International Encyclopedia of Statistical Science* ». M. Lovric (éd.), Springer Berlin Heidelberg. p. 398-400.
- Millot, G. (2011). Comprendre et réaliser les tests statistiques à l'aide de R. 2<sup>e</sup> éd., De Boeck, Bruxelles.
- Oksanen, J. (1987). Problems of joint display of species and site scores in correspondence analysis. *Vegetatio*, 72(1), 51-57.
- Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). Statistical Inference from Capture Data on

## Conséquence de la non prise en compte de détectabilité dans les analyses ACP et AFC

- Closed Animal Populations. *The Wildlife Society*, (62), 3-135.
- Pledger, S. (2000). Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics*, 56 (2), 434-442.
- Pollock, K. H. (1980). Capture-recapture models: a review of current methods, assumptions and experimental design. *Citeseer*. 30 p.
- Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L., & Sauer, J. R. (2002). Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics*, 13 (2), 105-119.
- Popy, S., Bordignon, L., & Prodon, R. (2009). A weak upward elevational shift in the distributions of breeding birds in the Italian Alps: Elevational shift in Alpine bird distributions. *Journal of Biogeography*, 37 (1), 57-67.
- Prodon, R., & Lebreton, J. D. (1994). Analyses multivariées des relations espèces-milieu structure et interprétation écologique. *Vie et milieu*, 44 (1), 69-91.
- Prodon, R., & Jacquet, K. (2014). Analyses multivariées avec ADE4 dans R [Cours en ligne]. Consulté à l'adresse <https://pbil.univ-lyon1.fr/R/pdf/jacquet-prodon.pdf>. 32 p.
- R Core Team (2016). R: A language and environment for statistical computing. R. Foundation for Statistical Computing, Vienna, Austria. URL. <https://www.R-project.org/>.
- Reino, L., Beja, P., Osborne, P. E., Morgado, R., Fabião, A., & Rotenberry, J. T. (2009). Distance to edges, edge contrast and landscape fragmentation: Interactions affecting farmland birds around forest plantations. *Biological Conservation*, 142 (4), 824-838.
- Royle, J. A., & Nichols, J. D. (2003). Estimating abundance from repeated presence-absence data or point counts. *Ecology*, 84 (3), 777-790.
- Royle, J. A. (2004). N-Mixture Models for Estimating Population Size from Spatially Replicated Counts. *Biometrics*, 60 (1), 108-115.
- Royle, J. A., Nichols, J. D., & Kéry, M. (2005). Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, 110 (2), 353-359.
- Ter Braak, C. J. F. (1985). Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics*, 41 (4), 859-873.
- Ter Braak, C. J. F. (1987). The analysis of vegetation–environment relationships by canonical correspondence analysis. In « *Theory and models in vegetation science* ». 69-77.
- Ter Braak, C. J. F., & Prentice, I. C. (2004). A Theory of Gradient Analysis. In « *Advances in Ecological Research* » Vol. 34. Elsevier. 235-282.
- Tingley, M. W., & Beissinger, S. R. (2013). Cryptic loss of montane avian richness and high community turnover over 100 years. *Ecology*, 94 (3), 598–609.
- Vacher-Vallas, M., Bonnet, X., & Naulleau, G. (1999). Relations entre les comportements sexuels et les accouplements chez *Vipera aspis*: étude en milieu naturel. *Revue d'Écologie*, 54 (4), 375-391.
- Von Wehrden, H., Hanspach, J., Bruelheide, H., & Wesche, K. (2009). Pluralism and diversity: trends in the use and application of ordination methods 1990-2007. *Journal of Vegetation Science*, 20 (4), 695-705.
- Williams, B. K., Nichols, J. D., & Conroy, M. J. (2002). Analysis and management of animal populations: modeling, estimation, and decision making. San Diego: *Academic Press*. 817 p.
- Yamaura, Y., Royle, J. A., Kuboi, K., Tada, T., Ikeno, S., & Makino, S. (2011). Modelling community dynamics based on species-level abundance models from detection/nondetection data: Multi-species abundance model. *Journal of Applied Ecology*, 48 (1), 67-75.
- Zuur, A. F., Ieno, E. N., & Smith, G. M. (2007). Analysing ecological data. Springer. 672 p. + XXVI.

## Annexes A

Cette annexe présente les tableaux des valeurs des paramètres choisis pour les SIMULATIONs.

- ❖ En premier les lieux les valeurs des variables  $g_1$  et  $g_2$  sur les sites et aux optimums de chaque espèce, lors des ACP puis lors des AFC.
  - Le tableau I donne les valeurs de  $g_1$  et  $g_2$ , pour lesquelles chacune des 20 espèces trouve son optimum lors des SIMULATIONs des ACP.
  - Le tableau II donne les valeurs des variables d'habitat  $g_1$  et  $g_2$  sur les 50 sites lors des Simulations des ACP.
  - Le tableau III donne les valeurs de  $g_1$  et  $g_2$ , pour lesquelles chacune des 20 espèces trouve son optimum lors des SIMULATIONs des AFC.
  - Le tableau IV donne les valeurs des variables d'habitat  $g_1$  et  $g_2$  sur les 100 sites lors des Simulations des AFC.
  
- ❖ Puis les tables donnant les valeurs des paramètres utilisées lors des SIMULATIONs des ACP.
  - Le tableau V présente les valeurs utilisées lorsqu'il n'y a pas de gradient de PdD.
  - Le tableau VI présente les valeurs utilisées lorsque les gradients sont linéaires.
  - Le tableau VII présente les valeurs utilisées lorsque les gradients sont quadratiques « centrés ».
  
- ❖ Enfin les tables donnant les valeurs des paramètres utilisées lors des SIMULATIONs des AFC.
  - Le tableau VIII présente les valeurs utilisées lorsqu'il n'y a pas de gradient de PdD.
  - Le tableau IX présente les valeurs utilisées lorsque les gradients sont linéaires.
  - Le tableau X présente les valeurs utilisées lorsque les gradients sont quadratiques « centrés ».
  - Le tableau XI présente les valeurs utilisées lorsque les gradients sont quadratiques « décalés ».

Tableau I : les valeurs des variables  $g_1$  et  $g_2$  pour lesquelles chacune des 20 espèces trouve son optimum lors des SIMULATIONs des ACP

Espèce	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
sur $g_1$	263	288	313	338	363	388	413	438	463	488	1013	1038	1063	1088	1113	1138	1163	1188	1213	1238
sur $g_2$	5	275	14	284	23	293	32	302	41	311	50	320	59	329	68	338	77	347	86	356

Tableau II : les valeurs des variables  $g_1$  et  $g_2$  sur les sites lors des SIMULATIONs des ACP

Sites	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
sur $g_1$	505	515	525	535	545	555	565	575	585	595	605	615	625	635	645	655	665	675	685	695
sur $g_2$	99	117	135	153	171	189	207	225	243	261	99	117	135	153	171	189	207	225	243	261
Sites	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
sur $g_1$	705	715	725	735	745	755	765	775	785	795	805	815	825	835	845	855	865	875	885	895
sur $g_2$	99	117	135	153	171	189	207	225	243	261	99	117	135	153	171	189	207	225	243	261
Sites	41	42	43	44	45	46	47	48	49	50										
sur $g_1$	905	915	925	935	945	955	965	975	985	995										
sur $g_2$	99	117	135	153	171	189	207	225	243	261										

Tableau III : les valeurs des variables  $g_1$  et  $g_2$  pour lesquelles chacune des 20 espèces trouve son optimum lors des SIMULATIONs des AFC

Espèce	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
sur $g_1$	25	75	125	175	225	275	325	375	425	475	525	575	625	675	725	775	825	875	925	975
sur $g_2$	9	189	27	207	45	225	63	243	81	261	99	279	117	297	135	315	153	333	171	351

Tableau IV : les valeurs des variables  $g_1$  et  $g_2$  sur les sites lors des SIMULATIONs des AFC

Sites	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
sur $g_1$	5	15	25	35	45	55	65	75	85	95	105	115	125	135	145	155	165	175	185	195
sur $g_2$	18	54	90	126	162	198	234	270	306	342	18	54	90	126	162	198	234	270	306	342
Sites	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
sur $g_1$	205	215	225	235	245	255	265	275	285	295	305	315	325	335	345	355	365	375	385	395
sur $g_2$	18	54	90	126	162	198	234	270	306	342	18	54	90	126	162	198	234	270	306	342
Sites	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
sur $g_1$	405	415	425	435	445	455	465	475	485	495	505	515	525	535	545	555	565	575	585	595
sur $g_2$	18	54	90	126	162	198	234	270	306	342	18	54	90	126	162	198	234	270	306	342
Sites	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
sur $g_1$	605	615	625	635	645	655	665	675	685	695	705	715	725	735	745	755	765	775	785	795
sur $g_2$	18	54	90	126	162	198	234	270	306	342	18	54	90	126	162	198	234	270	306	342
Sites	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
sur $g_1$	805	815	825	835	845	855	865	875	885	895	905	915	925	935	945	955	965	975	985	995
sur $g_2$	18	54	90	126	162	198	234	270	306	342	18	54	90	126	162	198	234	270	306	342

Tableau V : les valeurs des paramètres utilisées lors des simulations des ACP, avec des PdD sans gradient

SIMULATION par pas de 2	Pour les abondances						Pour les PdD				
	N <sub>min</sub>	N <sub>max</sub>	$\nu$ 1min	$\nu$ 1max	$\nu$ 2min	$\nu$ 2max	Nbr espèces concernées	d <sub>moy</sub>	$\nu$ detect	$\beta$ min	$\beta$ max
de 1001 à 1009	5	50	100000	100000	50000	50000	Toutes	de 0.1 à 0.9	0	0	0
de 2001 à 2009			250	1000	50000	50000					
de 3001 à 3009			250	1000	200	400					
de 4001 à 4009			100000	100000	150	400					
1102	5	50	100000	100000	50000	50000	Toutes	0.5	0.2	0	0
2102			250	1000	50000	50000					
3102			250	1000	200	400					
4102			100000	100000	150	400					

Tableau VI : les valeurs des paramètres utilisées lors des simulations des ACP, avec des gradients de PdD linéaires

SIMULATION par pas de 2	Pour les abondances						Pour les PdD				
	N <sub>min</sub>	N <sub>max</sub>	$\nu$ 1min	$\nu$ 1max	$\nu$ 2min	$\nu$ 2max	Nbr espèces concernées	d <sub>moy</sub>	$\nu$ detect	$\beta$ min	$\beta$ max
de 1021 à 1029	5	50	100000	100000	50000	50000	Toutes	de 0.1 à 0.9	0	0.008	0.008
de 2021 à 2029			250	1000	50000	50000					
de 3021 à 3029			250	1000	200	400					
de 4021 à 4029			100000	100000	150	400					
de 1031 à 1039	5	50	100000	100000	50000	50000	11 sur 20	de 0.1 à 0.9	0	0.008	0.008
de 2031 à 2039			250	1000	50000	50000					
de 3031 à 3039			250	1000	200	400					
de 4031 à 4039			100000	100000	150	400					
1162	5	50	100000	100000	50000	50000	Toutes	0.5	0.2	0	0.03
2162			250	1000	50000	50000					
3162			250	1000	200	400					
4162			100000	100000	150	400					

Tableau VII : les valeurs des paramètres utilisées lors des simulations des ACP, avec des gradients de PdD de formes quadratiques centrées

SIMULATION	Pour les abondances						Pour les PdD					
par pas de 2	N <sub>min</sub>	N <sub>max</sub>	$\nu$ 1min	$\nu$ 1max	$\nu$ 2min	$\nu$ 2max	Nbr espèces concernées	d <sub>moy</sub>	$\nu$ detect	$\gamma$ min	$\gamma$ max	$\mu$
de 1041 à 1049	5	50	100000	100000	50000	50000	Toutes	de 0.1 à 0.9	0	0.00004	0.00004	centrée
de 2041 à 2049			250	1000	50000	50000						
de 3041 à 3049			250	1000	200	400						
de 4041 à 4049			100000	100000	150	400						
de 1051 à 1059	5	50	100000	100000	50000	50000	11 sur 20	de 0.1 à 0.9	0	0.00004	0.00004	centrée
de 2051 à 2059			250	1000	50000	50000						
de 3051 à 3059			250	1000	200	400						
de 4051 à 4059			100000	100000	150	400						
1182	5	50	100000	100000	50000	50000	Toutes	0.5	0.2	0	0.00004	centrée
2182			250	1000	50000	50000						
3182			250	1000	200	400						
4182			100000	100000	150	400						

Tableau VIII : les valeurs des paramètres utilisées lors des simulations des AFC, avec des PdD sans gradient

SIMULATION	Pour les abondances						Pour les PdD				
par pas de 2	N <sub>min</sub>	N <sub>max</sub>	$\nu$ 1min	$\nu$ 1max	$\nu$ 2min	$\nu$ 2max	Nbr espèces concernées	d <sub>moy</sub>	$\nu$ detect	$\beta$ min	$\beta$ max
de 1001 à 1009	5	50	100000	100000	50000	50000	Toutes	de 0.1 à 0.9	0	0	0
de 2001 à 2009			200	1000	50000	50000					
de 3001 à 3009			200	1000	150	400					
de 4001 à 4009			100000	100000	100	400					
1102	5	50	100000	100000	50000	50000	Toutes	0.5	0.2	0	0
2102			200	1000	50000	50000					
3102			200	1000	150	400					
4102			100000	100000	100	400					

Tableau IX : les valeurs des paramètres utilisées lors des simulations des AFC, avec des gradients de PdD linéaires

SIMULATION	Pour les abondances						Pour les PdD				
par pas de 2	N <sub>min</sub>	N <sub>max</sub>	$\nu$ 1min	$\nu$ 1max	$\nu$ 2min	$\nu$ 2max	Nbr espèces concernées	d <sub>moy</sub>	$\nu$ detect	$\beta$ min	$\beta$ max
de 1021 à 1029	5	50	100000	100000	50000	50000	Toutes	de 0.1 à 0.9	0	0.005	0.005
de 2021 à 2029			200	1000	50000	50000					
de 3021 à 3029			200	1000	150	400					
de 4021 à 4029			100000	100000	100	400					
de 1031 à 1039	5	50	100000	100000	50000	50000	11 sur 20	de 0.1 à 0.9	0	0.005	0.005
de 2031 à 2039			200	1000	50000	50000					
de 3031 à 3039			200	1000	150	400					
de 4031 à 4039			100000	100000	100	400					
1162	5	50	100000	100000	50000	50000	Toutes	0.5	0.2	0	0.01
2162			200	1000	50000	50000					
3162			200	1000	150	400					
4162			100000	100000	100	400					

Tableau X : les valeurs des paramètres utilisées lors des simulations des AFC, avec des gradients de PdD de formes quadratiques centrées

SIMULATION	Pour les abondances						Pour les PdD					
par pas de 2	N <sub>min</sub>	N <sub>max</sub>	$\nu$ 1min	$\nu$ 1max	$\nu$ 2min	$\nu$ 2max	Nbr espèces concernées	d <sub>moy</sub>	$\nu$ detect	$\gamma$ min	$\gamma$ max	$\mu$
de 1041 à 1049	5	50	100000	100000	50000	50000	Toutes	de 0.1 à 0.9	0	0.00001	0.00001	centrée
de 2041 à 2049			200	1000	50000	50000						
de 3041 à 3049			200	1000	150	400						
de 4041 à 4049			100000	100000	100	400						
de 1051 à 1059	5	50	100000	100000	50000	50000	11 sur 20	de 0.1 à 0.9	0	0.00001	0.00001	centrée
de 2051 à 2059			200	1000	50000	50000						
de 3051 à 3059			200	1000	150	400						
de 4051 à 4059			100000	100000	100	400						
1182	5	50	100000	100000	50000	50000	Toutes	0.5	0.2	0	0.00001	centrée
2182			200	1000	50000	50000						
3182			200	1000	150	400						
4182			100000	100000	100	400						

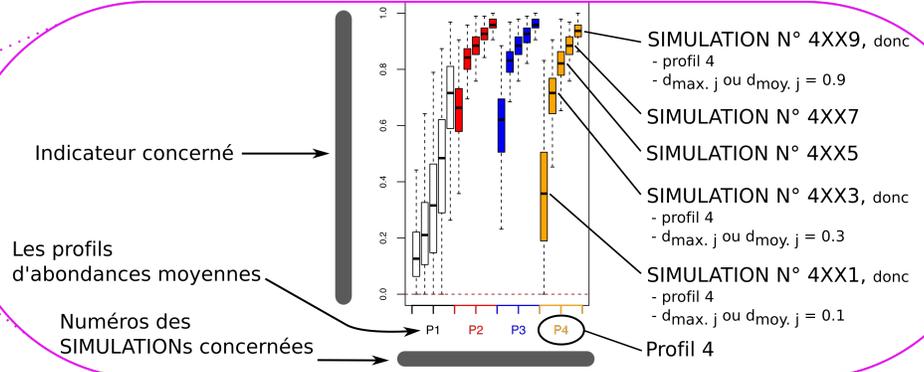
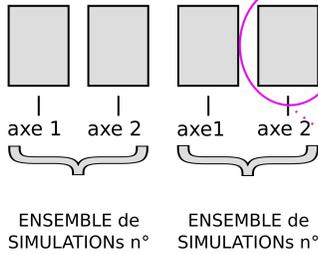
Tableau XI : les valeurs des paramètres utilisées lors des simulations des AFC, avec des gradients de PdD de formes quadratiques « décalées »

SIMULATION	Pour les abondances						Pour les PdD					
par pas de 2	N <sub>min</sub>	N <sub>max</sub>	$\nu$ 1min	$\nu$ 1max	$\nu$ 2min	$\nu$ 2max	Nbr espèces concernées	d <sub>moy</sub>	$\nu$ detect	$\gamma$ min	$\gamma$ max	$\mu$
de 1291 à 1299	5	50	100000	100000	50000	50000	Toutes	de 0.1 à 0.9	0	0.00001	0.00001	décalée
de 2291 à 2299			200	1000	50000	50000						
de 3291 à 3299			200	1000	150	400						
de 4291 à 4299			100000	100000	100	400						
de 1301 à 1309	5	50	100000	100000	50000	50000	11 sur 20	de 0.1 à 0.9	0	0.00001	0.00001	décalée
de 2301 à 2309			200	1000	50000	50000						
de 3301 à 3309			200	1000	150	400						
de 4301 à 4309			100000	100000	100	400						
1202	5	50	100000	100000	50000	50000	Toutes	0.5	0.2	0	0.00001	décalée
2202			200	1000	50000	50000						
3202			200	1000	150	400						
4202			100000	100000	100	400						

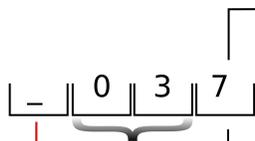
## **Annexes B**

Cette annexe est la fiche de lecture détachable, qui est sur les deux pages suivantes, qui sont détachables.

## Détails sur l'organisation dans un graphique d'une figure



## Code des numéros de simulations et légendes des graphiques :



### Dans les cas où $d_{max.j}$ ou $d_{moy.j}$ sont identiques entre les espèces

Ce chiffre indique la valeur de ces paramètres (probabilité de détection):

- =1 : cette probabilité est de 0.1
- =3 : cette probabilité est de 0.3
- =5 : cette probabilité est de 0.5
- =7 : cette probabilité est de 0.7
- =9 : cette probabilité est de 0.9

### Dans les cas où $d_{max.j}$ ou $d_{moy.j}$ sont différents entre les espèces

les valeurs de ces paramètres sont tirées dans une loi normale et:

- =2 : la moyenne de la loi est de 0.5 et son écart type est de 0.2

Se référer aux figures 5 et 6 des différentes simulations des probabilités de détection

Indique sous lequel des 4 profils d'abondances la SIMULATION est produite:

- **profil 1** : les abondances ne dépendent ni de  $g_1$  ni de  $g_2$
- **profil 2** : les abondances dépendent uniquement de  $g_1$
- **profil 3** : les abondances dépendent de  $g_1$  et de  $g_2$
- **profil 4** : les abondances dépendent uniquement de  $g_2$

### pour les graphiques simples

- Profil 1 : P1** (la population ne dépend d'aucun gradient)  
entre analyses non dégradées et dégradées
- Profil 2 : P2** (la population ne dépend que de  $g_1$ )  
entre analyses non dégradées et dégradées
- Profil 3 : P3** (la population dépend de  $g_1$  et de  $g_2$ )  
entre analyses non dégradée et dégradée
- Profil 4 : P4** (la population ne dépend que de  $g_2$ )  
entre analyses non dégradée et dégradée

### pour les graphiques doubles

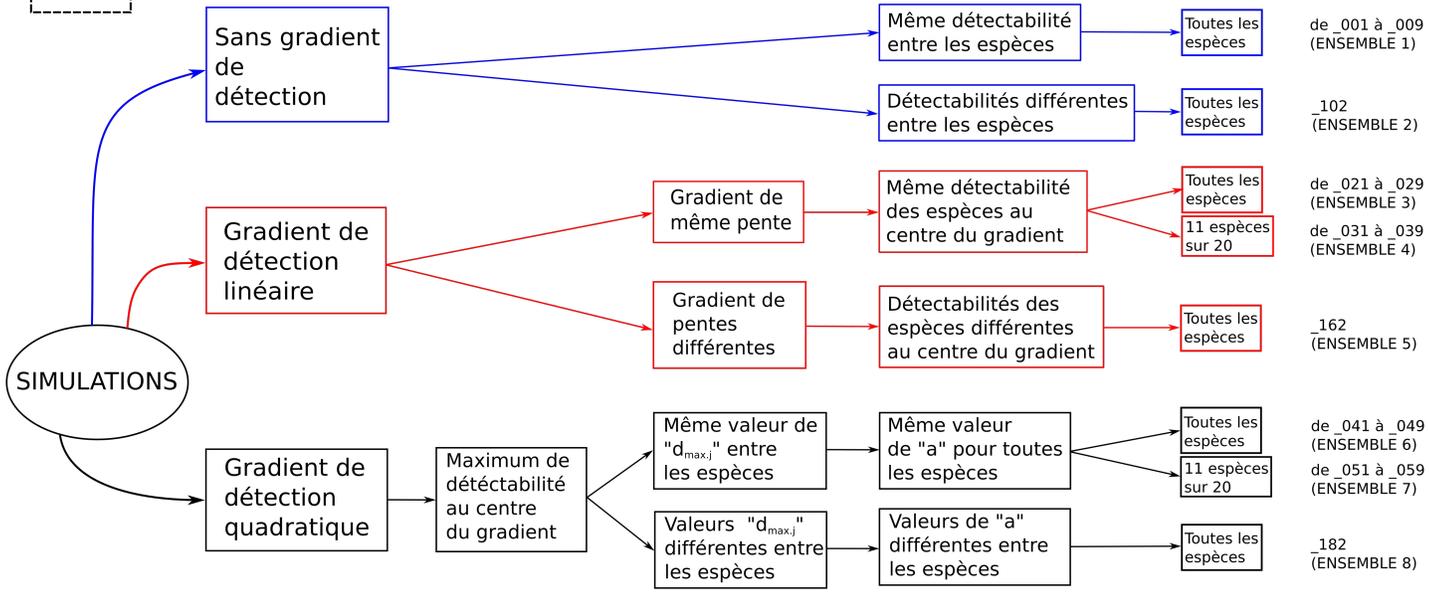
- Profil 1 : P1** (la population ne dépend d'aucun gradient)  
sur les analyses non dégradées  
sur les analyses dégradées
- Profil 2 : P2** (la population ne dépend que de  $g_1$ )  
sur les analyses non dégradées  
sur les analyses dégradées
- Profil 3 : P3** (la population dépend de  $g_1$  et de  $g_2$ )  
sur les analyses non dégradées  
sur les analyses dégradées
- Profil 4 : P4** (la population ne dépend que de  $g_2$ )  
sur les analyses non dégradées  
sur les analyses dégradées

### pour les graphiques séparant dans une même analyse les 10 premières espèces des 10 dernière ou les espèces paires des espèces impaires

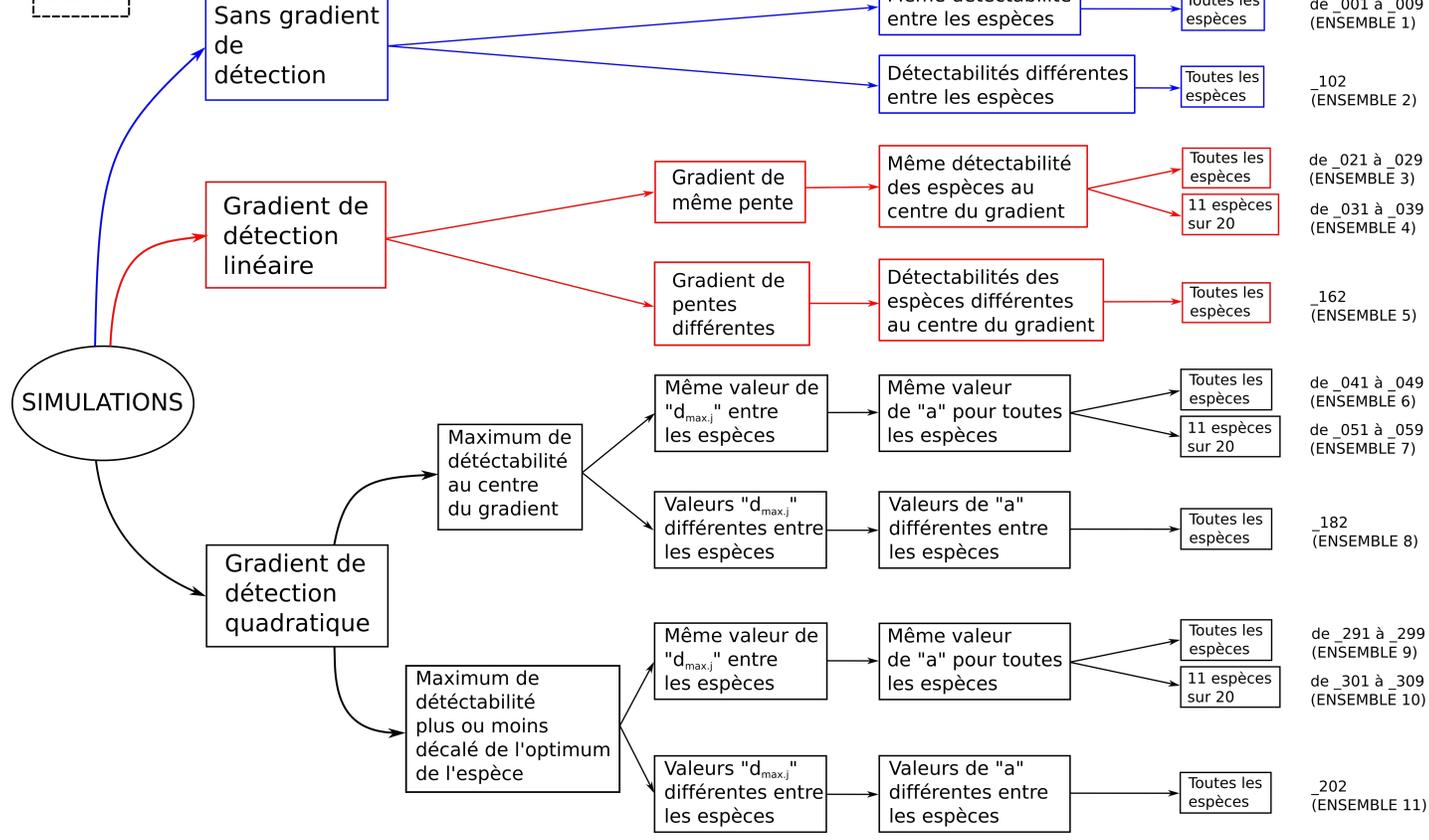
- Profil 1 : P1** (la population ne dépend d'aucun gradient)  
les 10 premières espèces ou les espèces paires (en fonction du graphique)  
les 10 dernières espèces ou les espèces impaires (en fonction du graphique)
- Profil 2 : P2** (la population ne dépend que de  $g_1$ )  
les 10 premières espèces ou les espèces paires (en fonction du graphique)  
les 10 dernières espèces ou les espèces impaires (en fonction du graphique)
- Profil 3 : P3** (la population dépend de  $g_1$  et de  $g_2$ )  
les 10 premières espèces ou les espèces paires (en fonction du graphique)  
les 10 dernières espèces ou les espèces impaires (en fonction du graphique)
- Profil 4 : P4** (la population ne dépend que de  $g_2$ )  
les 10 premières espèces ou les espèces paires (en fonction du graphique)  
les 10 dernières espèces ou les espèces impaires (en fonction du graphique)

Les trois derniers chiffres des numéros des SIMULATIONS dont on donne les résultats

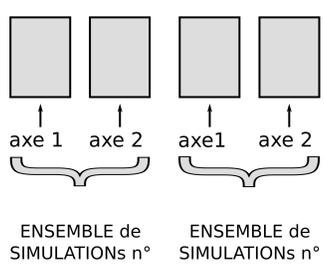
ACP



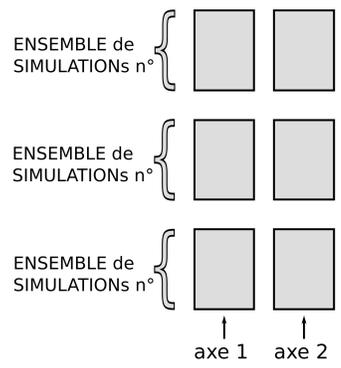
AFC



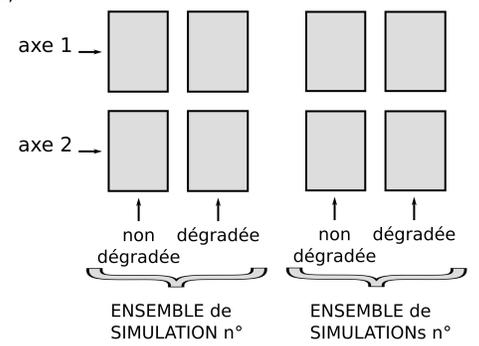
Les figures avec deux ensembles de SIMULATIONS



Les figures avec trois ensembles de SIMULATIONS



Les figures des séparations des groupes 1 et 2, et A et B



ÉCOLE PRATIQUE DES HAUTES ÉTUDES  
SCIENCE DE LA VIE ET DE LA TERRE

CONSÉQUENCES DE LA NON PRISE EN COMPTE DE LA DÉTECTABILITÉ  
DES ESPÈCES DANS LES ANALYSES ACP ET AFC

*Bonsacquet Lionel*

Le 26 juin 2017

## Résumé

Les analyses factorielles sont couramment utilisées en écologie pour analyser les données multivariées. Ces analyses fournissent des ordinations permettant de visualiser les principaux facteurs de variations au sein des données dans des graphiques en deux dimensions. Les tableaux regroupant les données d'abondances ou de présences de plusieurs espèces sur plusieurs sites peuvent être explorés à l'aide de ces analyses. Dans ces derniers cas cependant, la question des conséquences sur les résultats de ces analyses, de la non prise en compte du fait que la probabilité de détection des espèces sur le terrain est généralement inférieure à un lors de l'élaboration des tableaux de données, n'a jamais été clairement posée.

Pour aborder cette question, je me suis concentré sur deux analyses factorielles classiques que sont les Analyses en Composantes Principales (ACP) et les Analyses Factorielles des Correspondances (AFC). Pour chacune des analyses, j'ai comparé les résultats obtenus à partir d'un tableau d'abondances simulé avec une probabilité de détection égale à un, avec ceux obtenus à partir du même tableau simulé cette fois avec des probabilités de détection inférieures à un. Je me suis limité aux cas où la probabilité de détection s'organisait selon un gradient, plus ou moins complexe, entre sites et entre espèces. Mes simulations montrent que les résultats des analyses sont dégradés par la non prise en compte de la probabilité de détection. Ces dégradations sont multifactorielles et sont donc variées. Les ordinations proposées par les analyses peuvent être le reflet, partiel ou total, de l'hétérogénéité de la probabilité de détection entre les espèces ou/et les sites. Aussi, certains facteurs de variations entre sites et espèces existant réellement peuvent ne plus apparaître dans les résultats des analyses sur les tableaux impactés par des probabilités de détection inférieures à un. L'application de telles analyses sur des tableaux d'abondances qui ne prennent pas en compte la probabilité de détection, peut donc mener à des conclusions erronées. Ceci me conduit à recommander fortement de prendre en compte la probabilité de détection en amont de ces analyses sur des tableaux d'abondances, par exemple à travers l'utilisation de méthodes telles que le Distance sampling ou le N-mixture. Ces travaux appellent aussi au développement de méthodes spécifiques combinant l'estimation de ces probabilités de détection et l'ordination des sites et des espèces.

**Mots Clés :** ACP, AFC, analyses multidimensionnelles, biais, dégradations, gradient, ordination, probabilité de détection.