

BECNM – Analyses de Données Multivariées

Couplage de tableaux

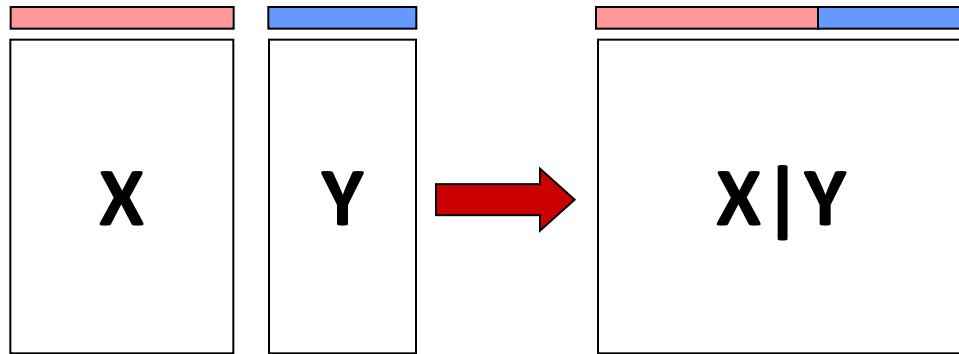
**Analyse Canonique des Correspondances
Analyse de Co-Inertie**

Gaël Grenouillet

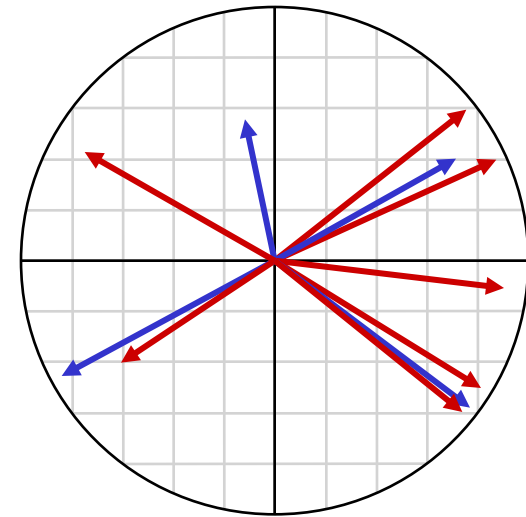
gael.grenouillet@univ-tlse3.fr

Premières approches

Juxtaposition



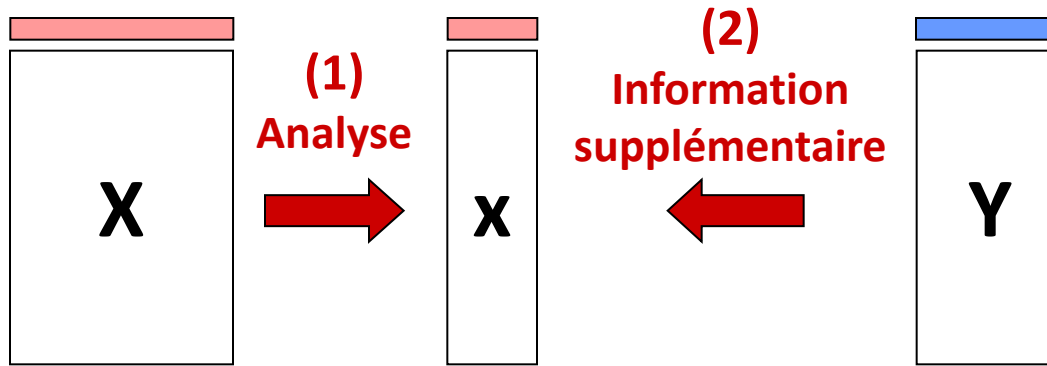
Deux tableaux ayant les mêmes lignes sont accolés
→ nouveau tableau
→ analyse simple (ex : ACP normée)



Cette approche ne fonctionne que si les inerties des deux tableaux sont comparables (sinon, l'un des deux tableaux impose son point de vue)

Premières approches

Illustration



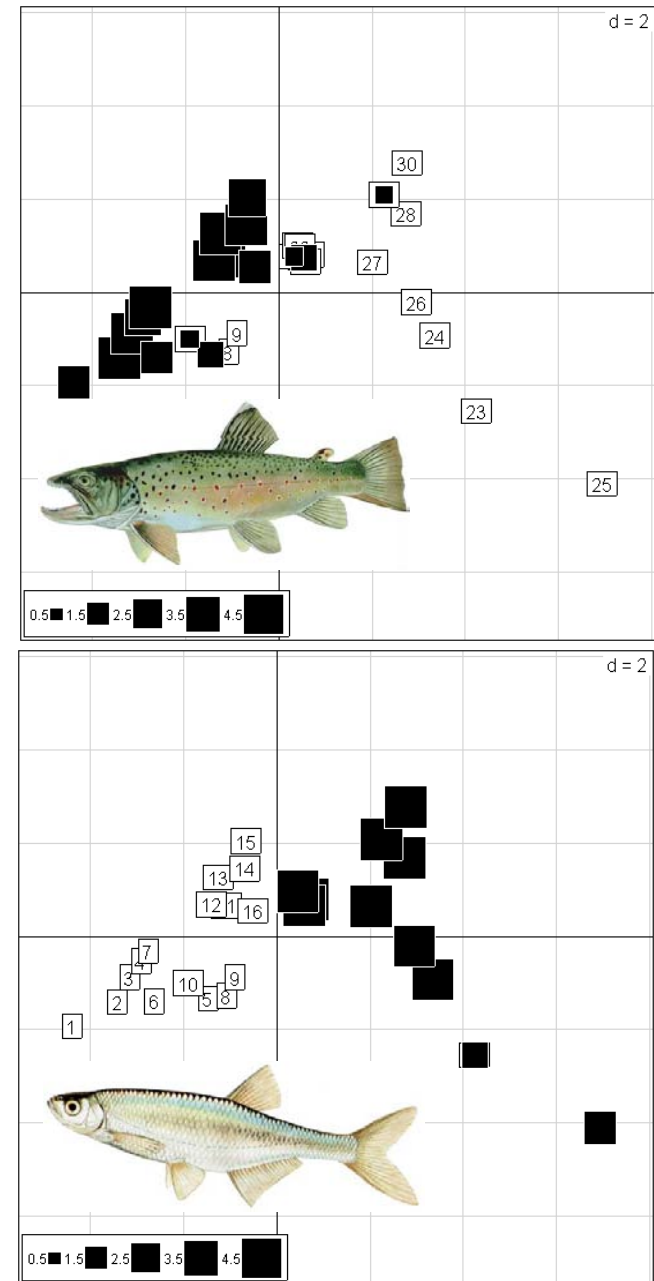
- (1) On pratique l'analyse d'un des tableaux
- (2) On introduit dans l'interprétation des éléments issus de l'autre tableau

- **Ordination directe :**

Analyse du tableau de milieu et étude de la répartition des espèces

- **Ordination indirecte :**

Analyse du tableau des espèces et étude de la répartition des variables

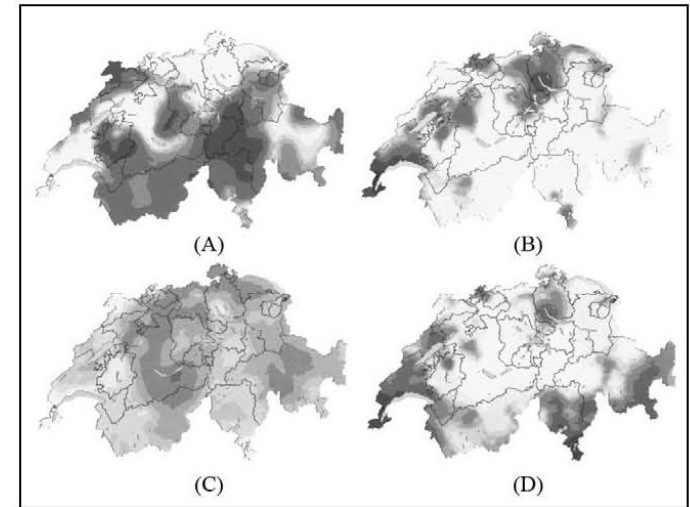


Ex : Ordination directe

Introduction

En écologie, un problème récurrent...

- **corrélations** entre variables
- **autocorrélation spatiale**

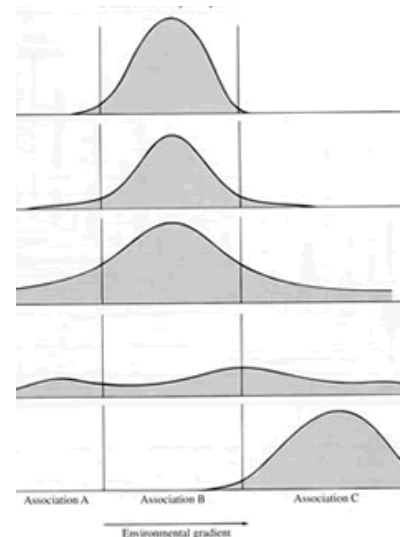


→ **Notion fondamentale : gradients environnementaux**

- Synthétisent de nombreuses variables corrélées
- Influencent fortement la distribution des espèces

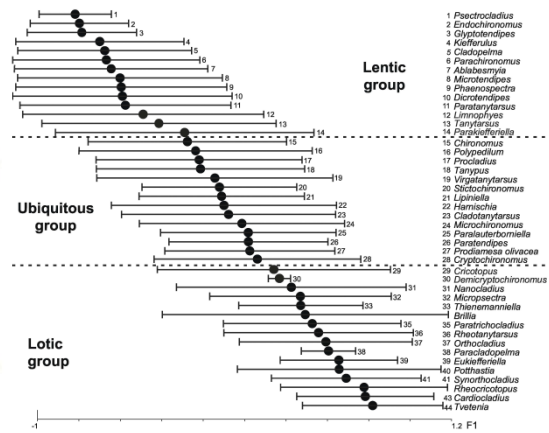
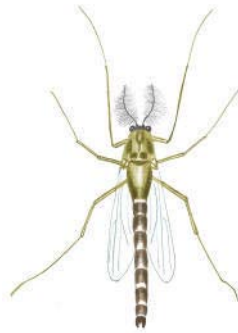
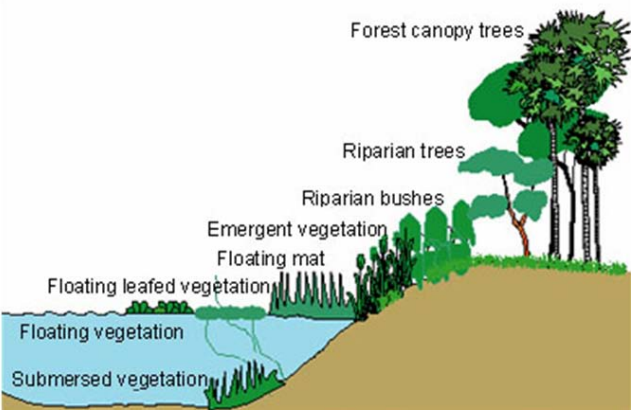
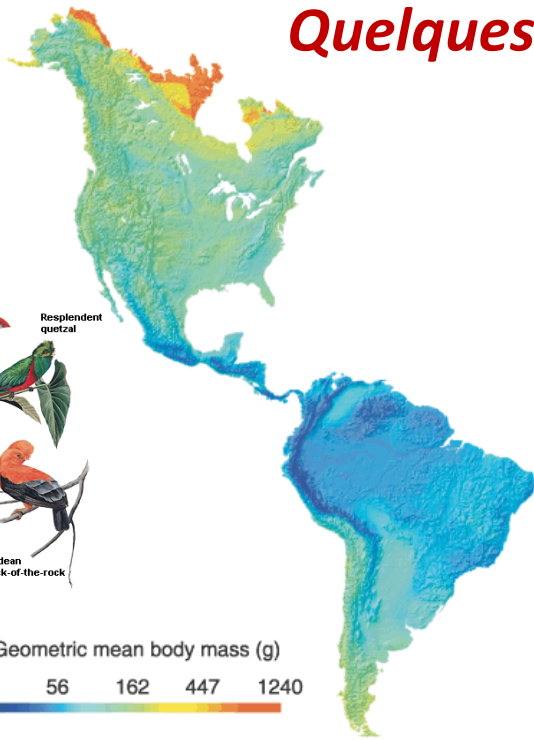
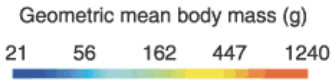
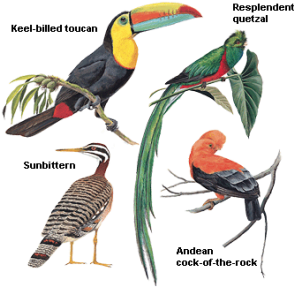
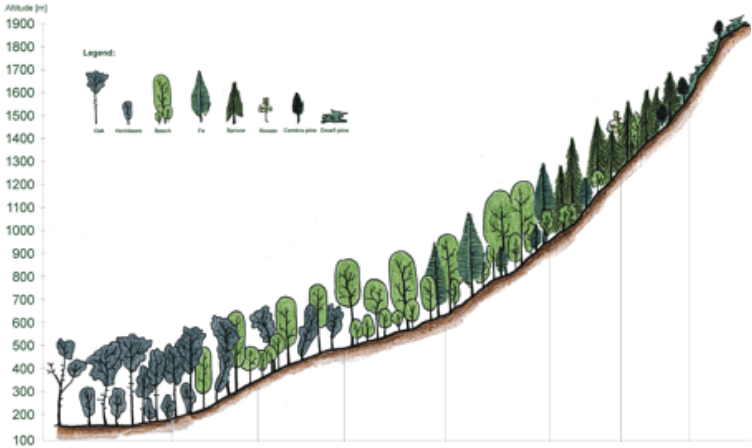
→ Nécessité de relier des patrons de distribution d'espèces (structure des communautés) à des patrons de distribution de variables (gradients environnementaux)

→ **Méthodes de couplages de tableaux**



Introduction

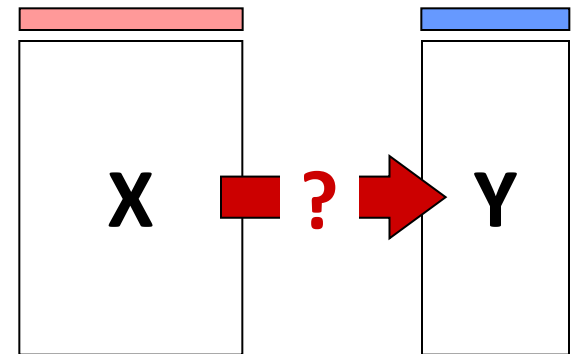
Quelques exemples



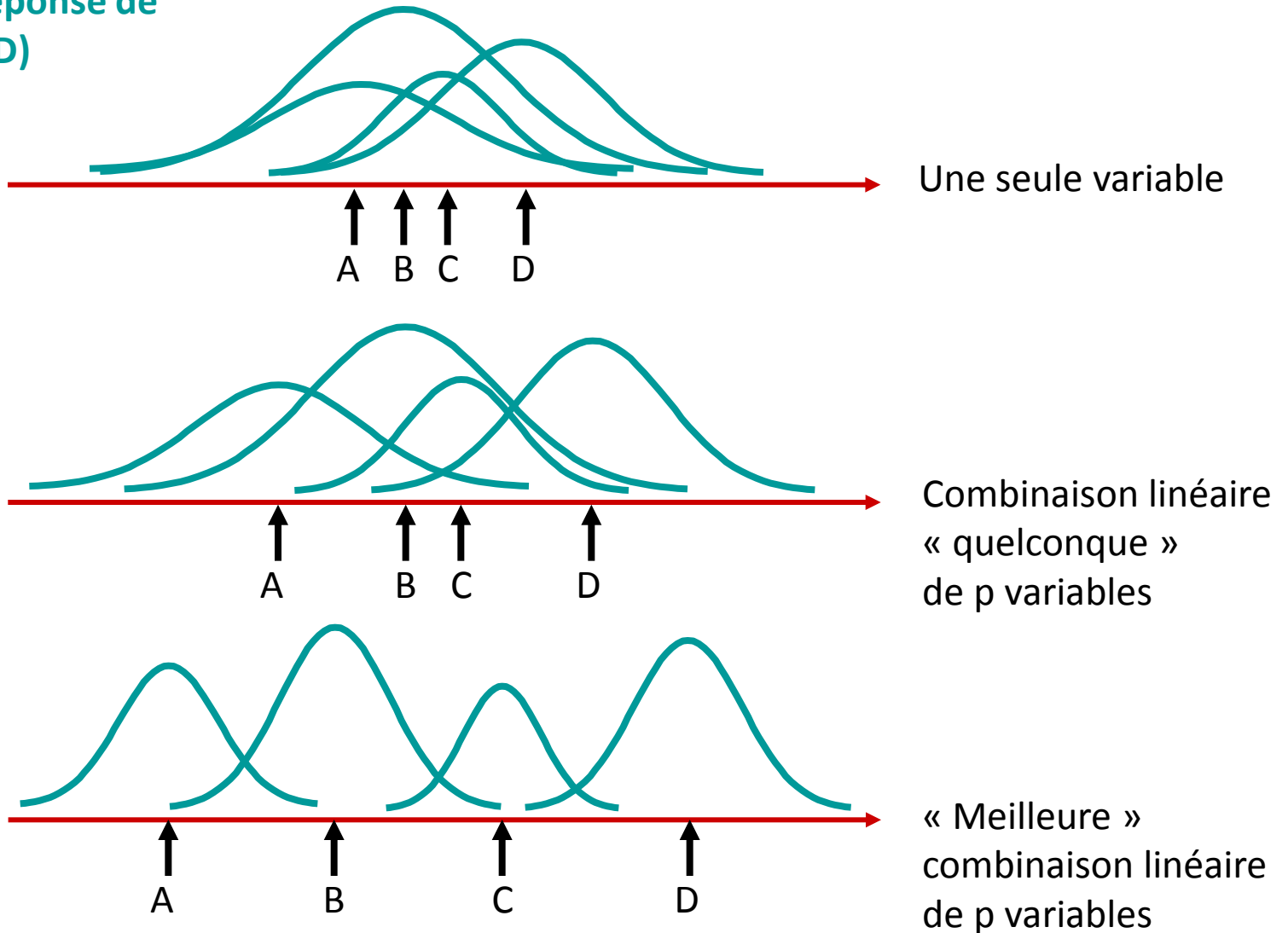
Analyse Canonique des Correspondances

- *Canonical Correspondence Analysis*
- **1986** Ter Braak

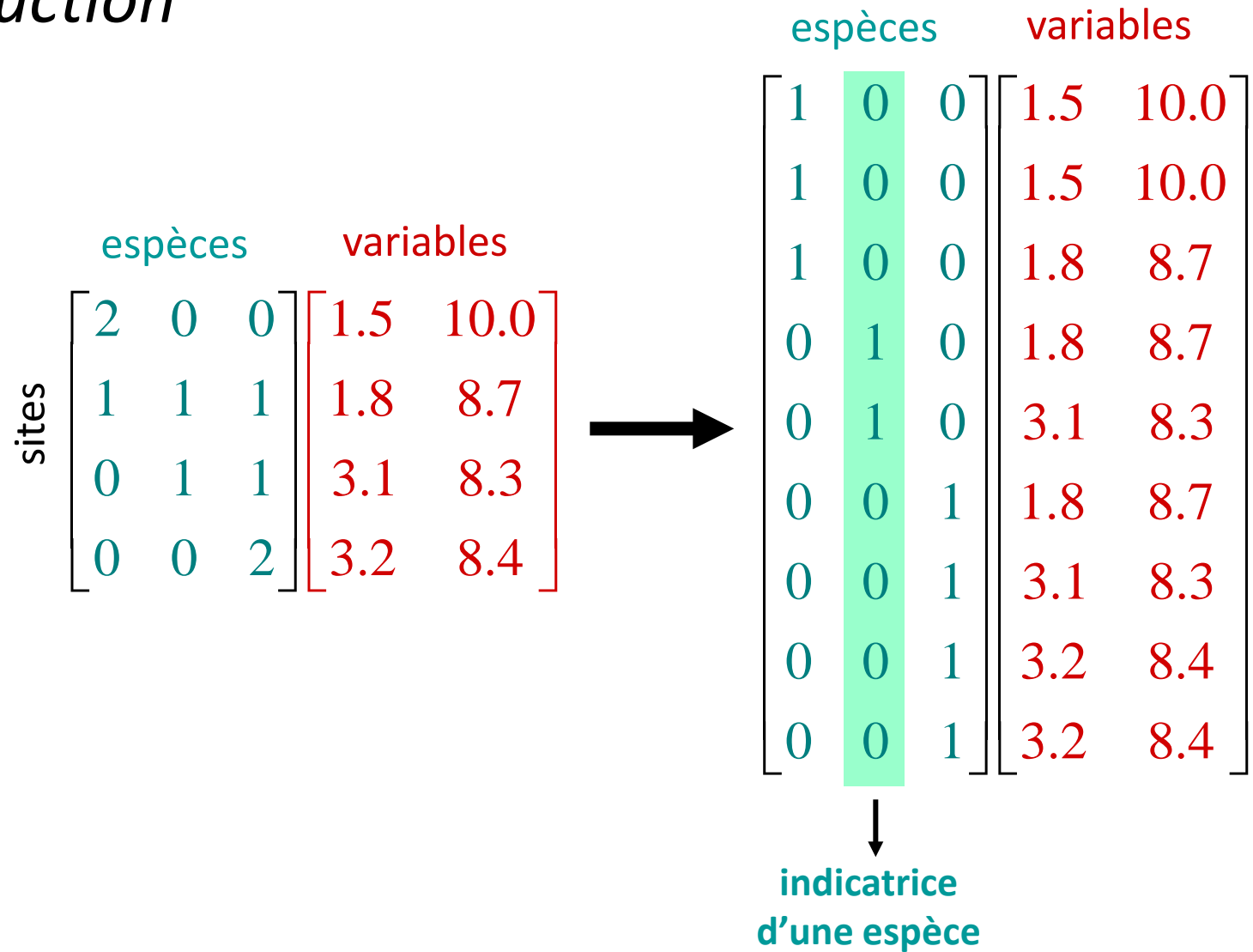
- Méthode de couplage la plus populaire en écologie
- Fondement : **AFC** du tableau faune
ACP normée du tableau de milieu
Ordination sous contrainte



Courbes de réponse de
4 espèces (A-D)



Introduction



➡ Le couple de tableaux appariés par les sites devient un couple de tableaux appariés par les occurrences

Introduction

espèces			variables	
1	0	0	1.5	10.0
1	0	0	1.5	10.0
1	0	0	1.8	8.7
0	1	0	1.8	8.7
0	1	0	3.1	8.3
0	0	1	1.8	8.7
0	0	1	3.1	8.3
0	0	1	3.2	8.4
0	0	1	3.2	8.4

Analyse du **tableau des espèces** → **AFC**

Analyse du **tableau des variables** → **ACP normée**

Pour que le couplage fonctionne,
les poids doivent être cohérents

→ **Centrer et normer le tableau de milieu avec les
poids issus du tableau faunistique**

Position moyenne de l'espèce 1 sur la variable 1 :
 $(1.5+1.5+1.8)/3$

Introduction

Le milieu décrit dans un relevé est d'autant plus important que ce relevé contient beaucoup d'individus (un relevé vide n'intervient pas dans l'analyse)

Les facteurs limitants communs aux différentes espèces sont minimisés
Les facteurs de séparation de niches écologiques sont maximisés



Vision de la niche écologique des espèces et des gradients environnementaux sur lesquels se séparent les niches des espèces

Représente les variables du 1er tableau (abondance des espèces) sur des axes qui sont des combinaisons des variables du 2ème tableau (variables de milieu)

L'Analyse Canonique des Correspondances (**ACC**) recherche les combinaisons linéaires des variables de milieu qui **maximise la dispersion des espèces**

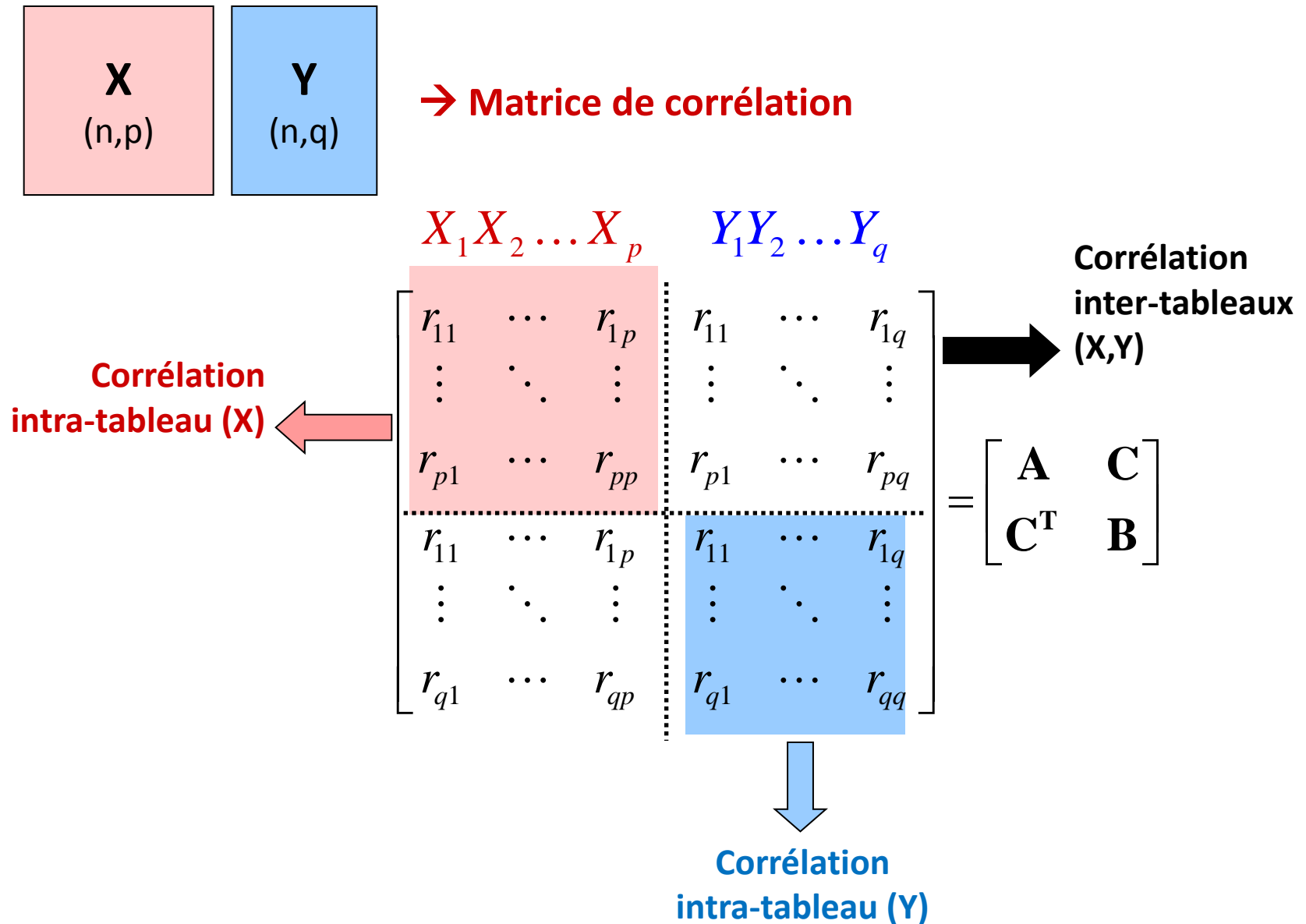
Remarque

L'AFC aussi maximise la dispersion des espèces mais indépendamment des variables de milieu

Originalité de l'ACC

- Méthode hybride entre **ordination** et **régression multiple**
- Les axes ne sont pas les gradients dominants issus d'analyses séparées pour chaque tableau, mais sont des combinaisons linéaires des variables indépendantes (milieu) qui **expliquent le plus de variance** des variables dépendantes (espèces)

Principe



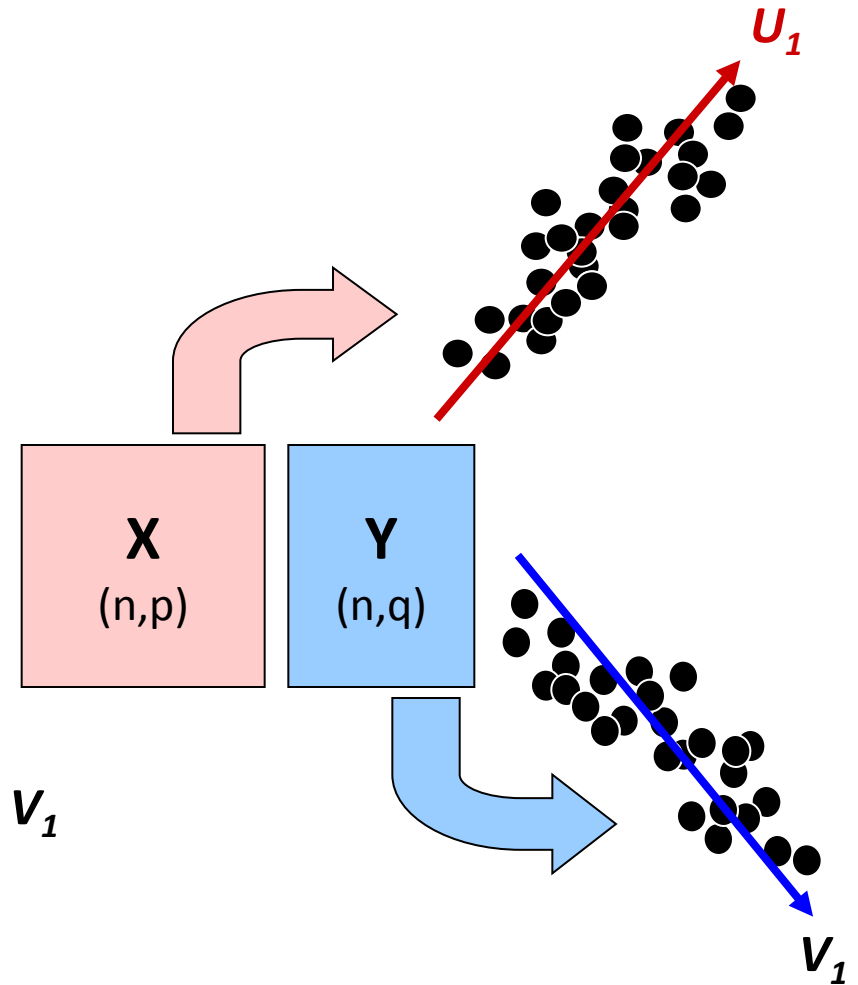
Principe

Variables canoniques

La première variable canonique est obtenue en trouvant les coefficients des fonctions linéaires

$$U_1 = \sum_{j=1}^p a_{1j} X_j \quad V_1 = \sum_{j=1}^q b_{1j} Y_j$$

qui maximisent la corrélation entre U_1 et V_1



Au final, l'analyse fournit

- **r variables canoniques** avec $r = \min(p,q)$
- **r valeurs propres** (carré de corrélation canonique)

Une combinaison linéaire des variables de milieu est une somme pondérée de la forme :

$$x_i = c_0 + c_1 V_{1i} + c_2 V_{2i} + \dots + c_q V_{qi}$$

V_{ji} = valeur de la variable environnementale j pour le site i

c_j = poids de chaque variable

x_i = score environnemental pour le site i

Coefficients de corrélation canonique : $\rho_i = \sqrt{\lambda_i}$

Le coefficient de régression multiple de la régression finale est appelé **corrélation espèce-environnement**

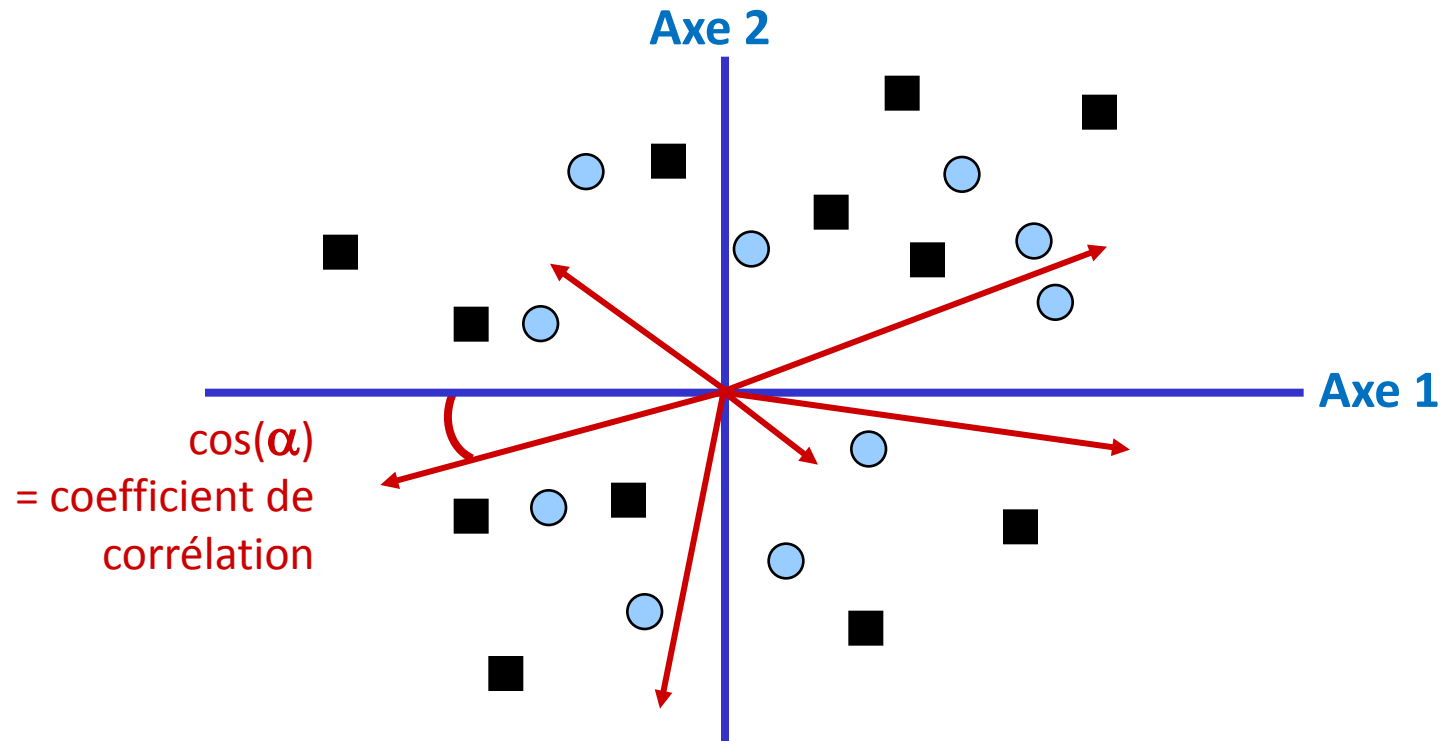
→ Mesure la part de variation explicable par les variables de milieu

Représentation graphique

« Triplot »

Représentation simultanée :

- des relevés (sites)
- des espèces
- des variables de milieu



Position d'un point-relevé

= moyenne des espèces qu'il contient

L'interprétation par rapport aux variables de milieu révèle les conditions environnementales du relevé

Position d'un point-espèce

= moyenne des relevés qui contiennent cette espèce

= optimum écologique de l'espèce

(l'abondance ou la probabilité de présence de l'espèce diminue lorsqu'on s'éloigne de ce point)

Cette représentation indique pour chaque relevé les espèces susceptibles d'être présentes

Interprétations

- **Coefficients des variables canoniques**

Les variables avec les plus fort coefficients
(positifs ou négatifs) permettent d'interpréter les gradients

$$U_1 = -0.09X_1 - 0.29X_2 + 0.48X_3 + 0.29X_4$$

représente une opposition entre X_3 et X_4 d'un côté, et X_2 de l'autre

- **Corrélations entre variables initiales et variables canoniques**

Les variables avec les plus fortes corrélations sont les plus importantes pour une variable canonique donnée

- **Quelle confiance accorder aux résultats ?**

S'assurer que le nombre d'observations est important par rapport au nombre de variables ($N/p+q$ grand)

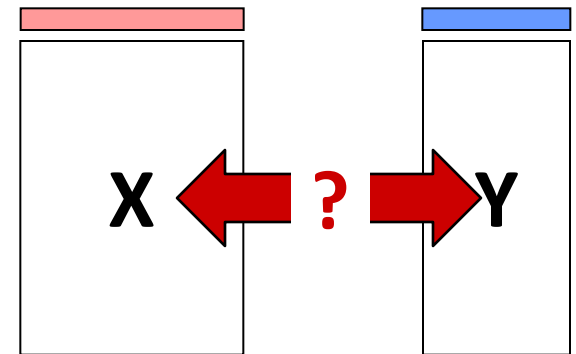
- Ne suppose pas de relations linéaires entre descripteurs
- La multicolinéarité (corrélation entre descripteurs de milieu) n'est pas un obstacle
- L'ordination est facilement interprétable : les axes fournissent une représentation des associations entre variables indépendantes qui expliquent le mieux la distribution des variables dépendantes
- L'ACC est peu affectée par des distributions asymétriques pour les espèces, mais il est préférable de transformer les variables de milieu si leurs distributions sont fortement asymétriques

Analyse de Co-Inertie

- *Co-inertia Analysis*
- **1994** Dolédec & Chessel

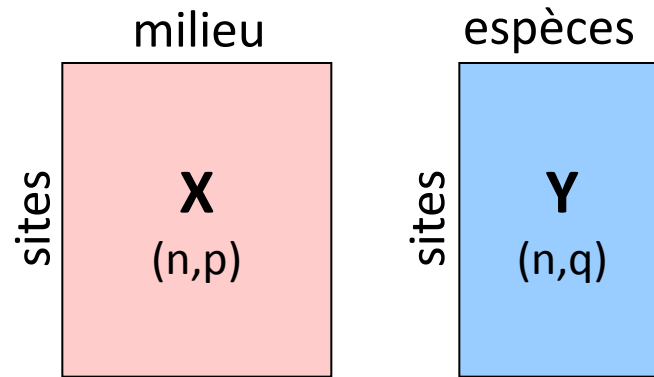
Applicable à tout couple d'analyses
(utilisant la même pondération des individus)

Seule méthode utilisable si les variables de
milieu sont qualitatives



- Réalise une **double analyse d'inertie** des tableaux
- Garantit que les deux systèmes de coordonnées sont les plus cohérents possibles

Principe



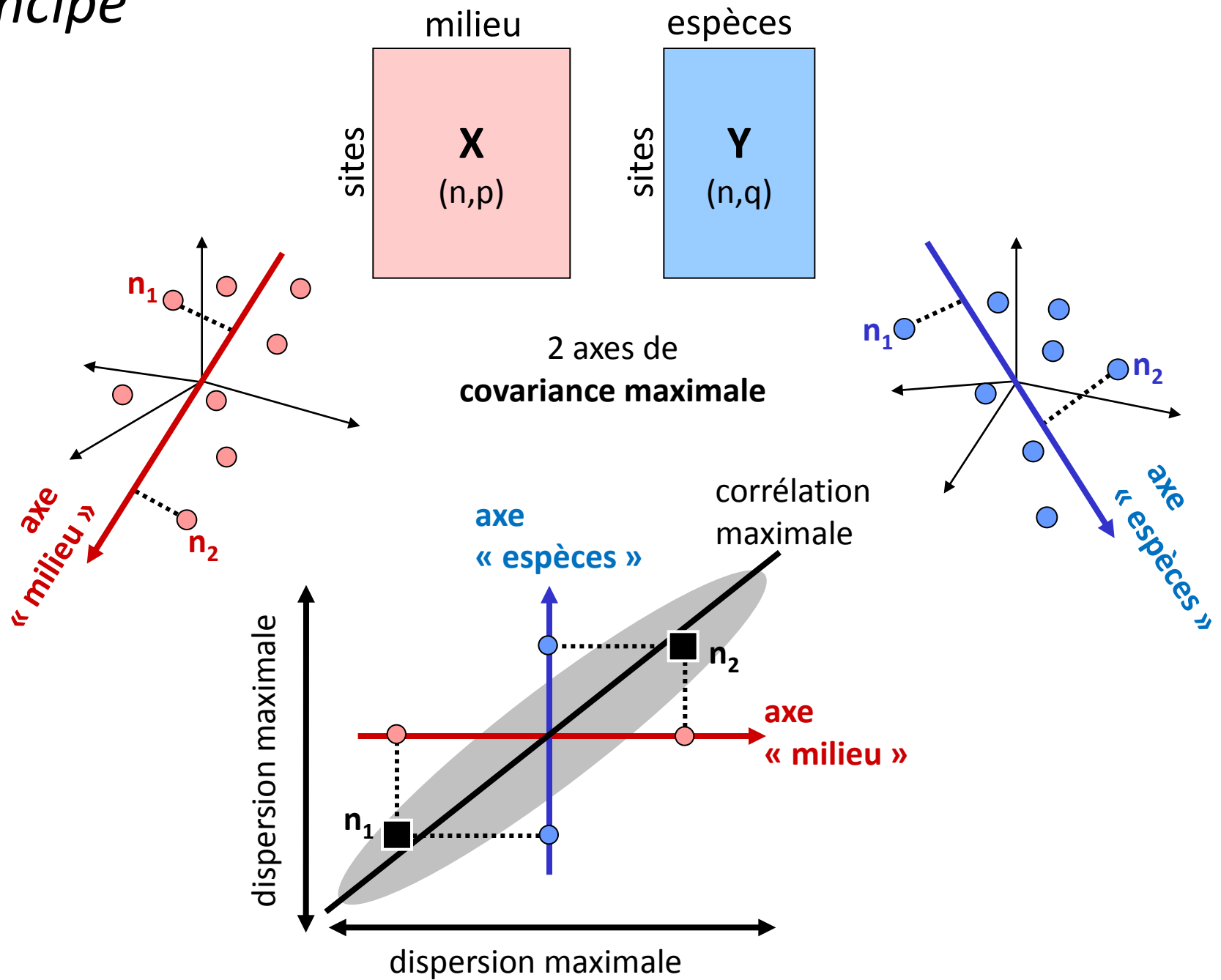
Deux tableaux définissent deux nuages de points

La méthode cherche un axe dans chaque espace et projette les nuages
→ **deux systèmes de coordonnées**

La **covariance** entre deux systèmes définit la **co-inertie** des deux axes

Trouver les axes de co-inertie des deux nuages, c'est trouver les deux axes qui maximisent la covariance des deux systèmes de coordonnées

Principe



Interprétations

Mise en évidence de la **structure commune** à deux tableaux

En général, les valeurs propres des analyses de co-inertie sont très structurées

Tests de permutation → test des co-structures observées

Hypothèse nulle : l'appariement des deux tableaux est aléatoire (tiré au hasard parmi les $n!$ permutations possibles des lignes de l'un, l'autre restant fixe)

Remarque :

Permet de coupler des tableaux ayant des nombres de lignes et de colonnes similaires (ce que ne permet pas l'ACC)

Univers assez complexe
Grande diversité d'expressions

→ Réflexion préalable à toute analyse pour définir les objectifs

Analyse canonique des correspondances

AFC – ACP normée

Maximise la variance des moyennes par espèces avec des combinaisons de variables de milieu normalisées

Analyse de co-inertie

Tout couple d'analyses

Maximise la covariance entre deux systèmes de coordonnées



☐ **Escoffier & Pagès.**

Analyses factorielles simples et multiples – Objectifs, méthodes et interprétation. Dunod

☐ **Lebart, Morineau & Piron.**

Statistique exploratoire multidimensionnelle. Dunod

☐ **Husson, Lê & Pagès.**

Analyse de données avec R - Presses Universitaires de Rennes

☐ **Manly.**

Multivariate statistical methods – A primer. Chapman & Hall

☐ **Jongman, Ter Braak & Van Tongeren.**

Data analysis in community and landscape ecology. Cambridge University Press

☐ **McGarigal, Cushman, Stafford.**

Multivariate statistics for wildlife and ecology research. Springer