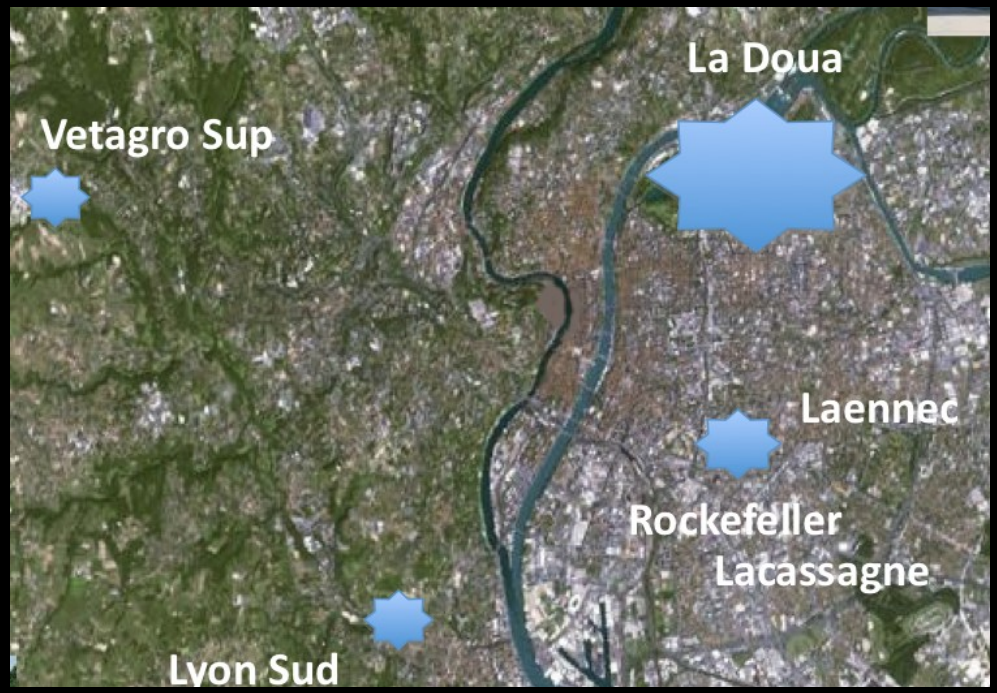
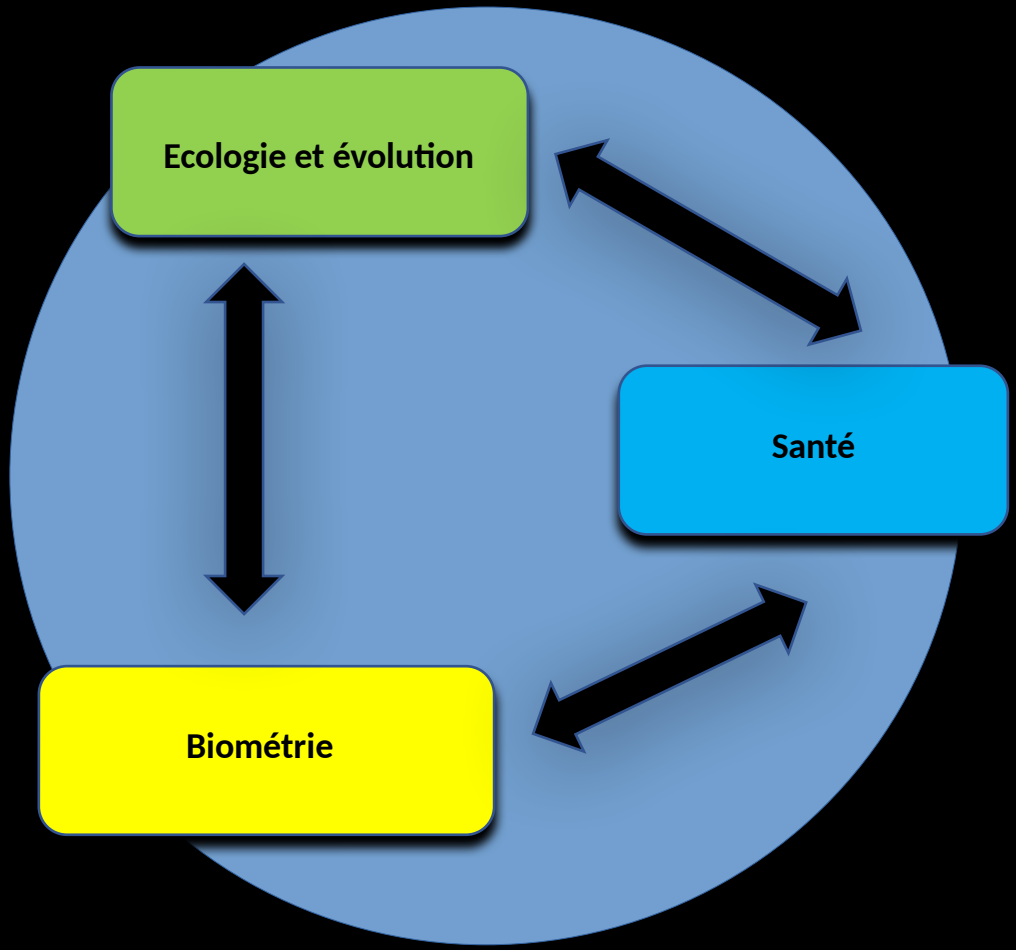
De l’écologie des communautés à la génomique : apport des méthodes d’analyses multivariées

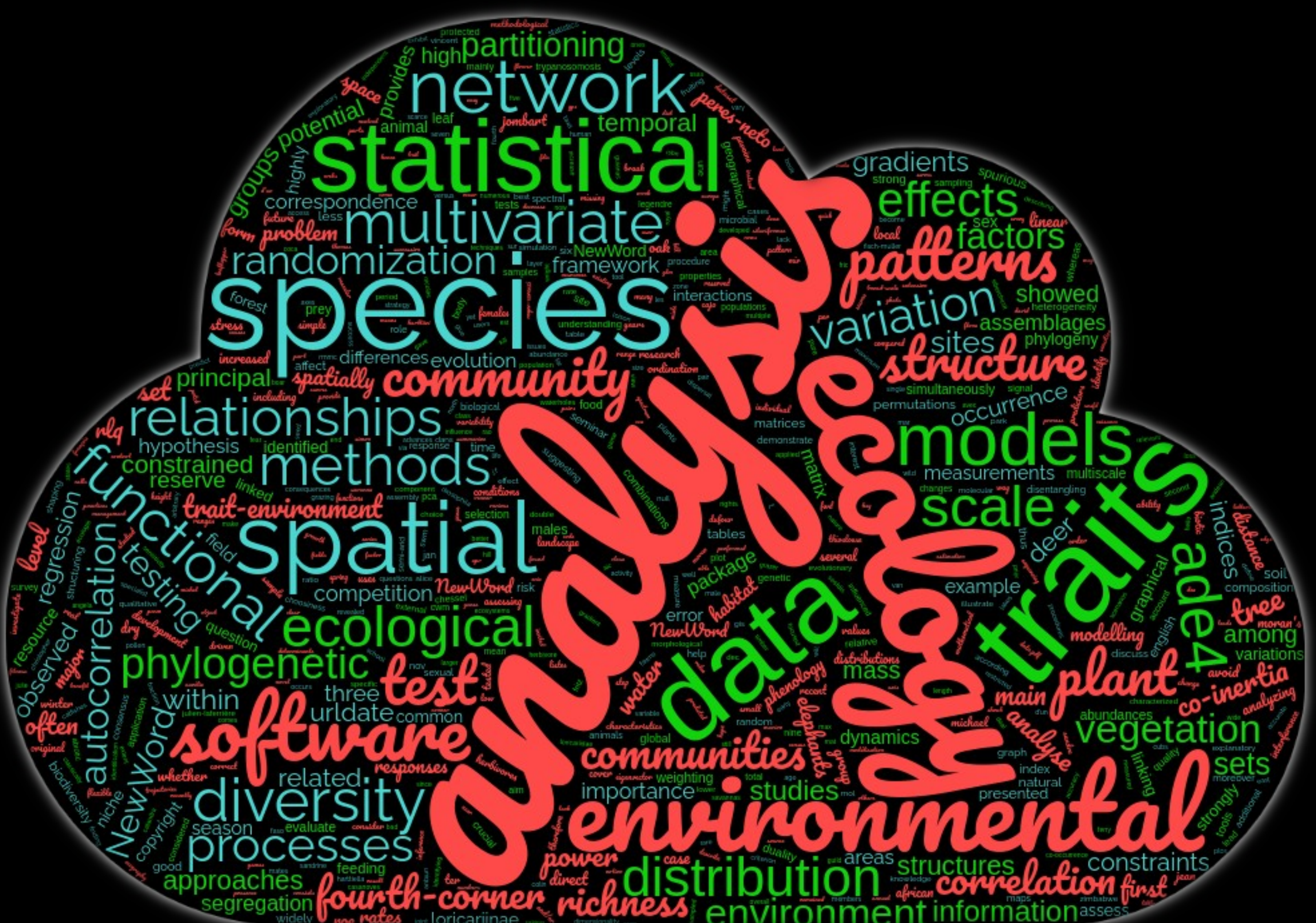
Stéphane Dray

30 mars 2023
Journée StatOmique « Analyses Factorielles »

LBBE


La Doua
Vetagro Sup
Laennec
Rockefeller
Lacassagne
Lyon Sud


Ecologie et évolution
Santé
Biométrie



analysis
ecology
traits
environmental
species
statistical
network
spatial
data
multivariate
models
software
partitioning
randomization
patterns
effects
factors
variation
structure
community
relationships
methods
scale
functional
testing
ecological
phylogenetic
test
plant
ade4
diversity
processes
distribution
environment
communities
vegetation
fourth-correlation
fourth-corner
richness
autocorrelation
co-inertia
gradients
sites
showed
assemblages
trait-environment
principal
constrained
hypothesis
competition
within
urldate
importance
studies
structures
constraints
information
approaches
segregation

Thèse : Eléments d'interface entre analyses multivariées, systèmes d'information géographique et observations écologiques

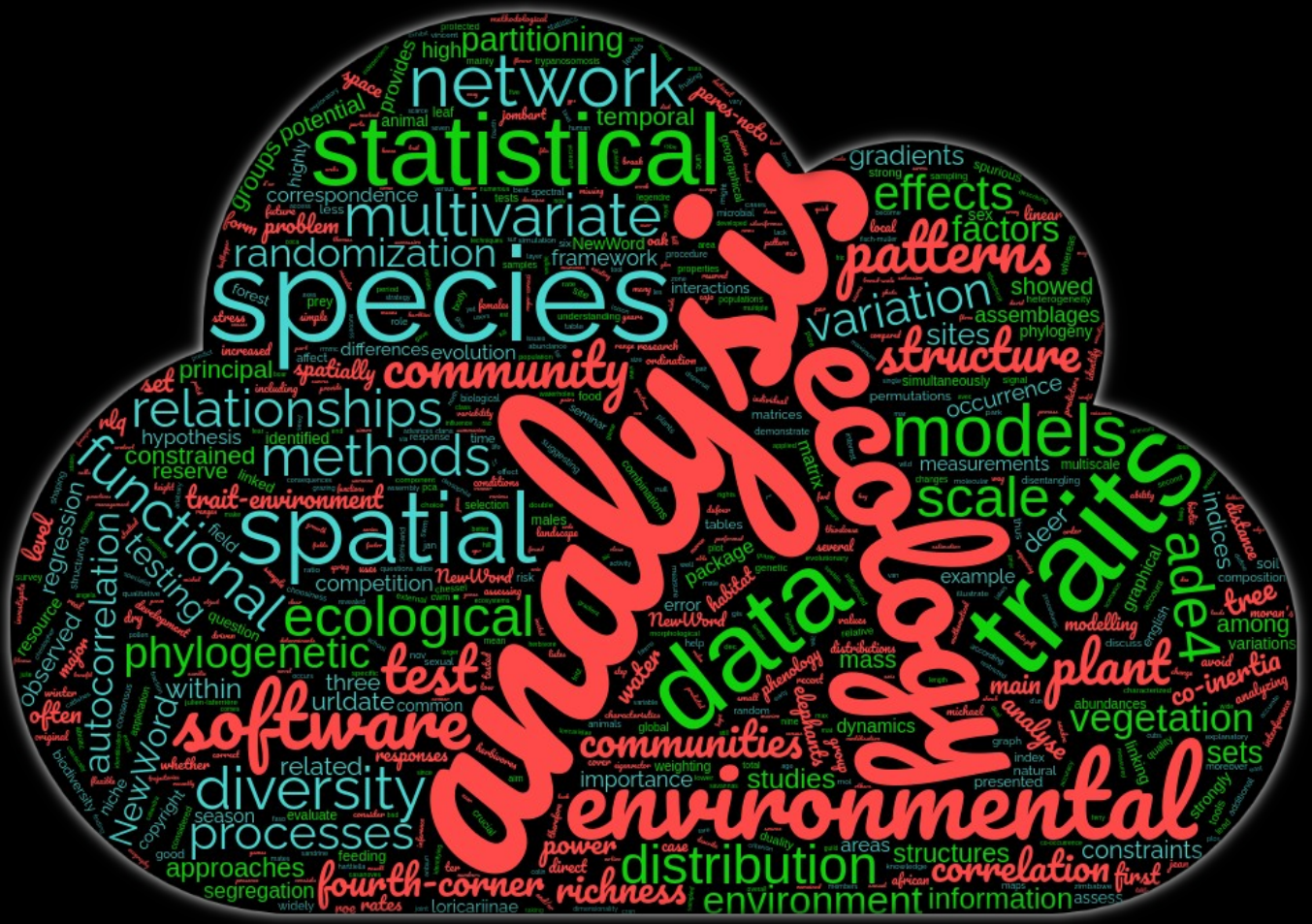
2003

Jury :
M. Antoine GUISAN
M. **Francis LALOË**
M. **Daniel CHESSEL**
M. Jean-Dominique LEBRETON
M. Alain PAVE
Mme Dominique PONTIER

2003

Post-Doc

2005


analysis
ecology
data
traits
environmental
statistical
network
species
multivariate
spatial
ecological
phylogenetic
functional
models
scale
patterns
structure
community
relationships
methods
testing
software
diversity
processes
distribution
test
plant
vegetation
communities
effects
variation
partitioning
randomization
correlation
richness
fourth-corner
trait-environment
co-inertia
ade4
LBBE
cnrs
2003
2005

2003

2005

2008

Thibaut Jombart

Thèse : Analyses multivariées de marqueurs génétiques : développements méthodologiques, applications et extensions

Thibaut Jombart

Thèse : Analyses multivariées de marqueurs génétiques : développements méthodologiques, applications et extensions

2003

2005

2008

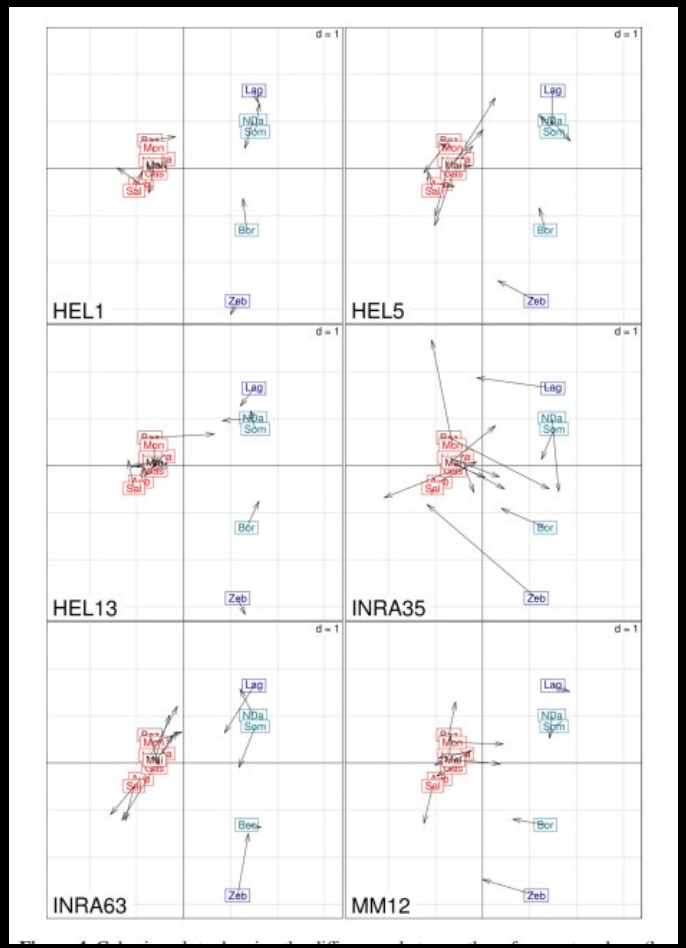
Genet. Sel. Evol. 39 (2007) 545–567
© INRA, EDP Sciences, 2007
DOI: 10.1051/gse:2007021

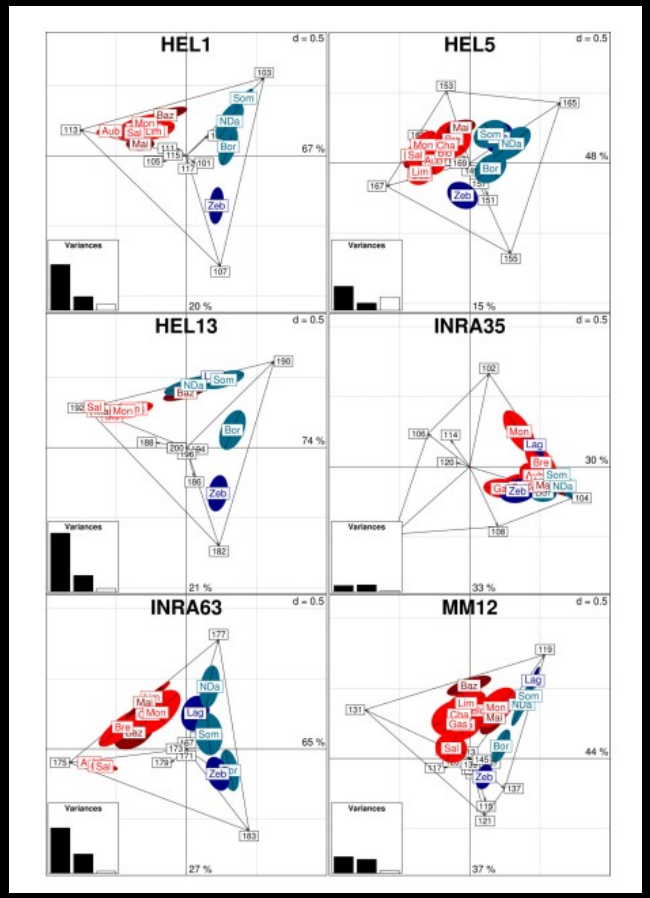
Available online at:
www.gse-journal.org

Original article

Consensus genetic structuring and typological value of markers using multiple co-inertia analysis

Denis LALOË[a*], Thibaut JOMBART[b], Anne-Béatrice DUFOUR[b], Katayoun MOAZAMI-GOUDARZI[c]





2003

2005

2008

Genet. Sel. Evol. 39 (2007) 545–567
© INRA, EDP Sciences, 2007
DOI: 10.1051/gse:2007021

Available online at:
www.gse-journal.org

Original article

Consensus genetic structuring and typological value of markers using multiple co-inertia analysis

Denis LALOË[a*], Thibaut JOMBART[b], Anne-Béatrice DUFOUR[b], Katayoun MOAZAMI-GOUDARZI[c]

BIOINFORMATICS **APPLICATIONS NOTE** *Vol. 24 no. 11 2008, pages 1403–1405*
doi:10.1093/bioinformatics/btn129

Genetics and population analysis

adegenet: a R package for the multivariate analysis of genetic markers

Thibaut Jombart*

ACKNOWLEDGEMENTS

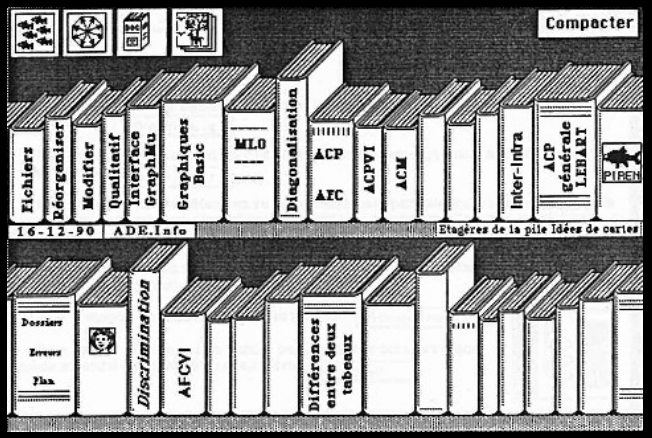
The author is grateful to R-Forge for hosting *adegenet*, to P. Sólymos for his contribution and to A.-B. Dufour, S. Devillard, D. Laloë and D. Pontier for their constructive comments.

2003

2005

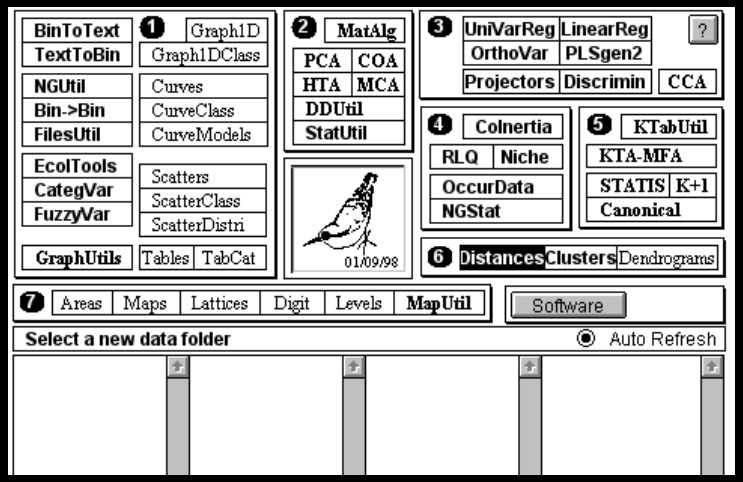
2008

grand "merci" à Stéphane Dray, Sébastien Devillard, Denis Laloë, Katayoun Moazami-Goudarzi et Sandrine Pavoine, avec qui j'ai eu le plaisir de collaborer. En particulier, merci à Stéphane pour m'avoir fait participer au SEDAR, et à Sébastien et Denis pour avoir servi de beta-testeurs à adegenet. Merci également à Hilmar Lapp et aux organisateurs et aux participants du 'R



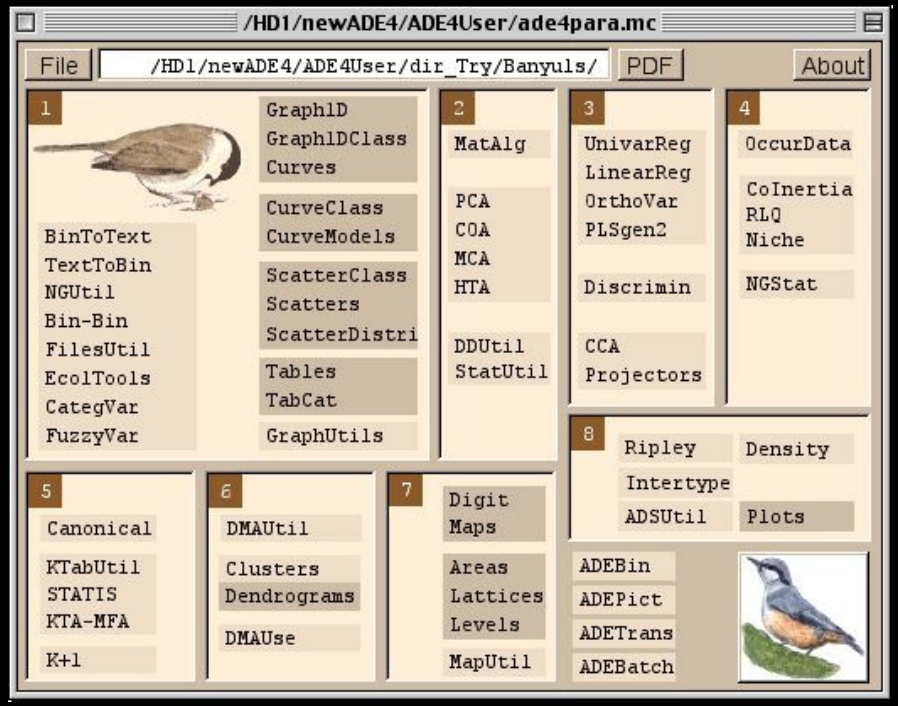
ADECO (1989)

- modules in Microsoft QuickBasic
- Hypercard interface



ADE-4 (1995)

- modules in C
- Hypercard and Winplus interfaces



ADE-4 (2000)

- Metacard interface
- batch mode


ade4
2002

The ade4 Package — II: Two-table and *K*-table Methods

by Stéphane Dray, Anne B. Dufour and Daniel Chessel

Journal of Statistical Software
August 2018, Volume 86, Issue 1. *doi: 10.18637/jss.v086.i01*

Supervised Multiblock Analysis in R with the ade4 Package

The ade4 package - I : One-table methods

by Daniel Chessel, Anne B Dufour and Jean Thioulouse

For example, if **X** is a table containing normalized quantitative variables, if **Q** is the identity matrix $\mathbf{I}_p$

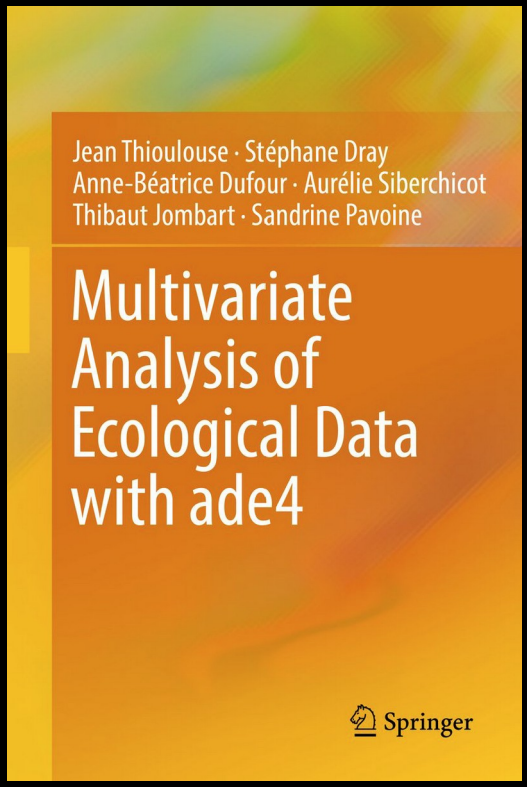
Journal of Statistical Software
September 2007, Volume 22, Issue 5. *http://www.jstatsoft.org/*

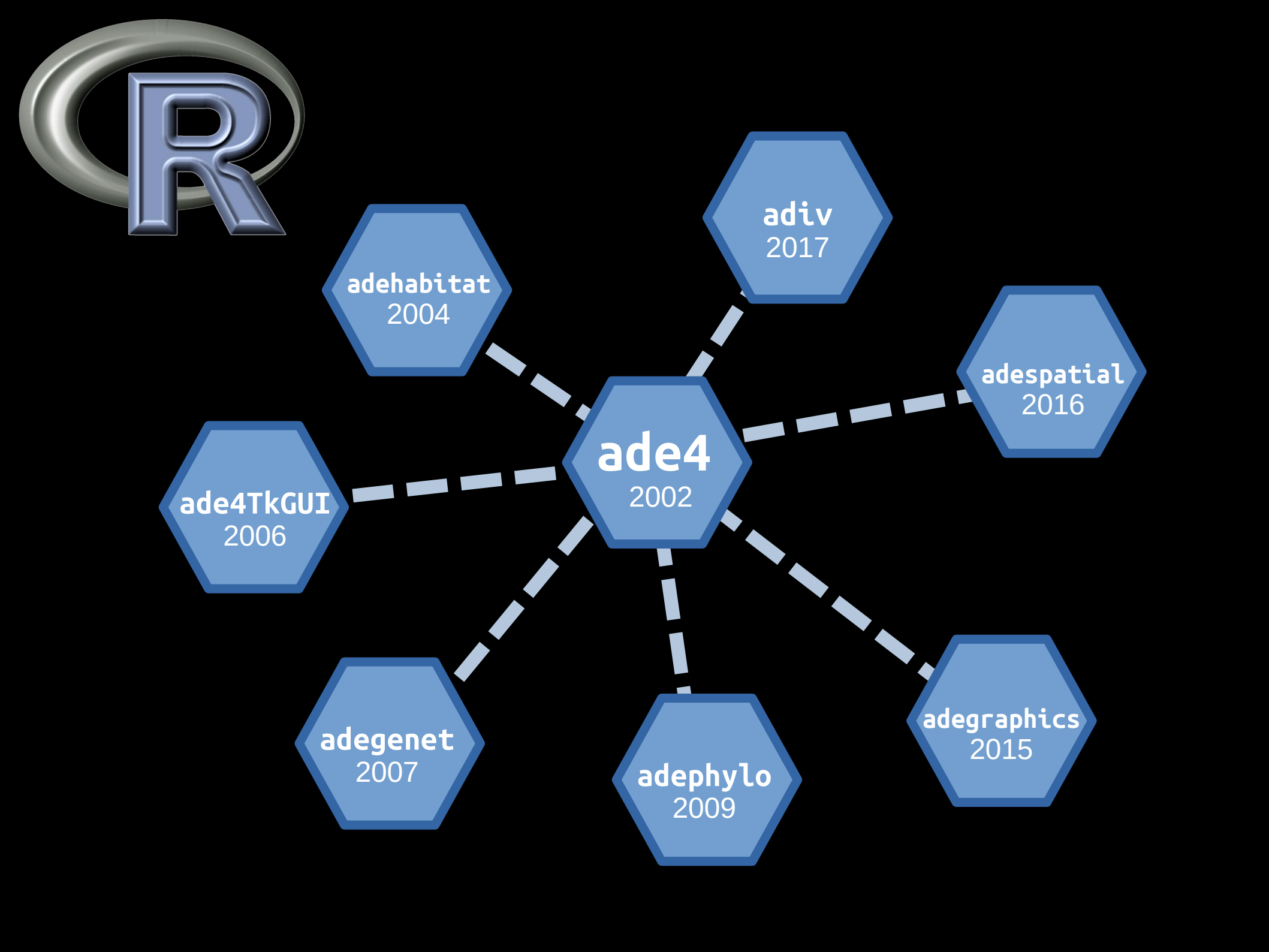
Interactive Multivariate Data Analysis in R with the ade4 and ade4TkGUI Packages

ade4
2002

Journal of Statistical Software
September 2007, Volume 22, Issue 4. *http://www.jstatsoft.org/*

The ade4 Package: Implementing the Duality Diagram for Ecologists




adehabitat
2004
adiv
2017
ade4
2002
adespatial
2016
ade4TkGUI
2006
adegenet
2007
adephylo
2009
adegraphics
2015

Ecologie des communautés et analyses multivariées

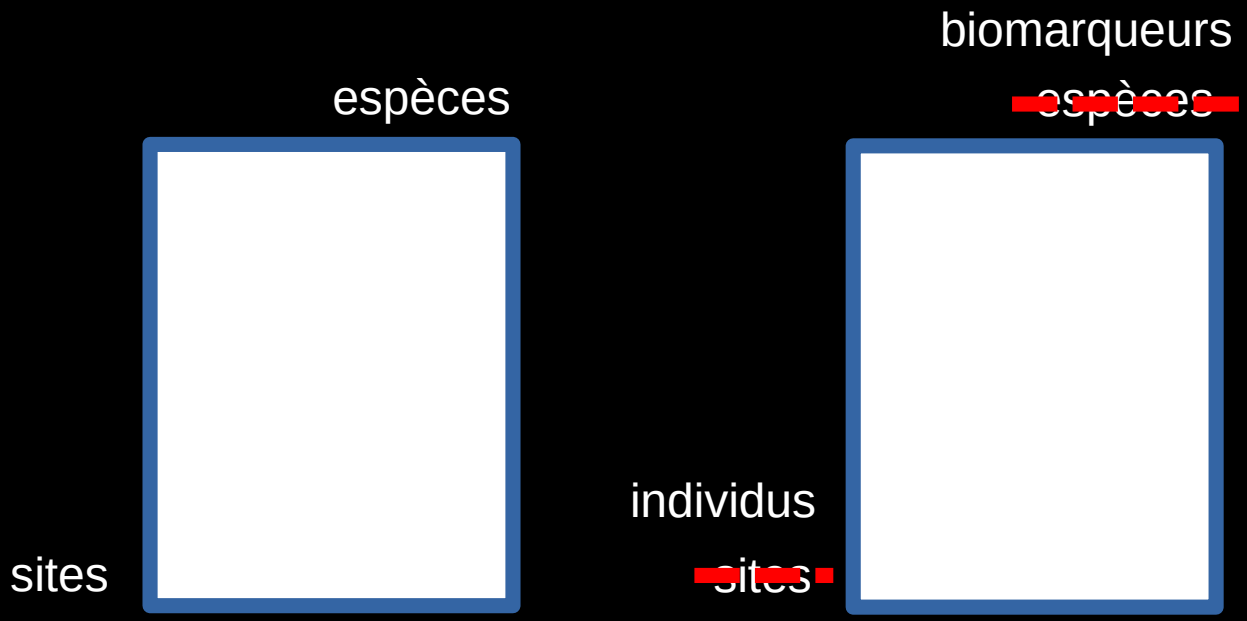
- quelles questions ?
- quelles données ?
- quelles méthodes ?

De l’écologie des communautés à la génomique ...

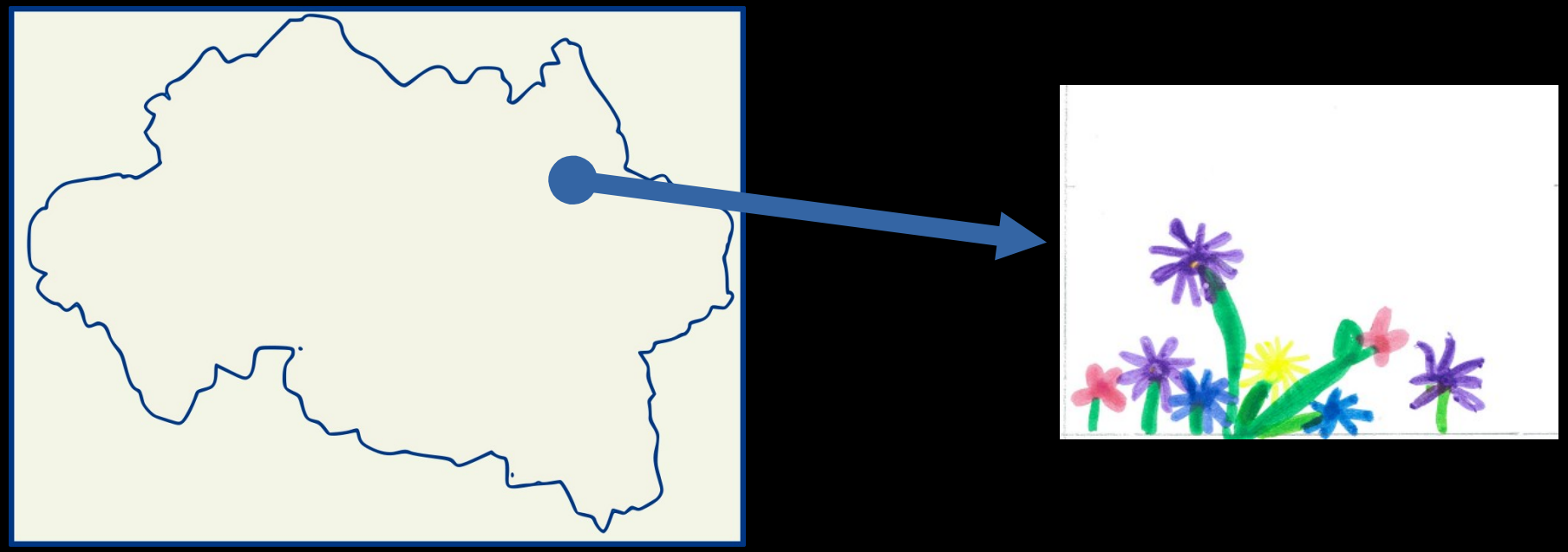
Ecologie des communautés et analyses multivariées

- quelles questions ?
- quelles données ?
- quelles méthodes ?

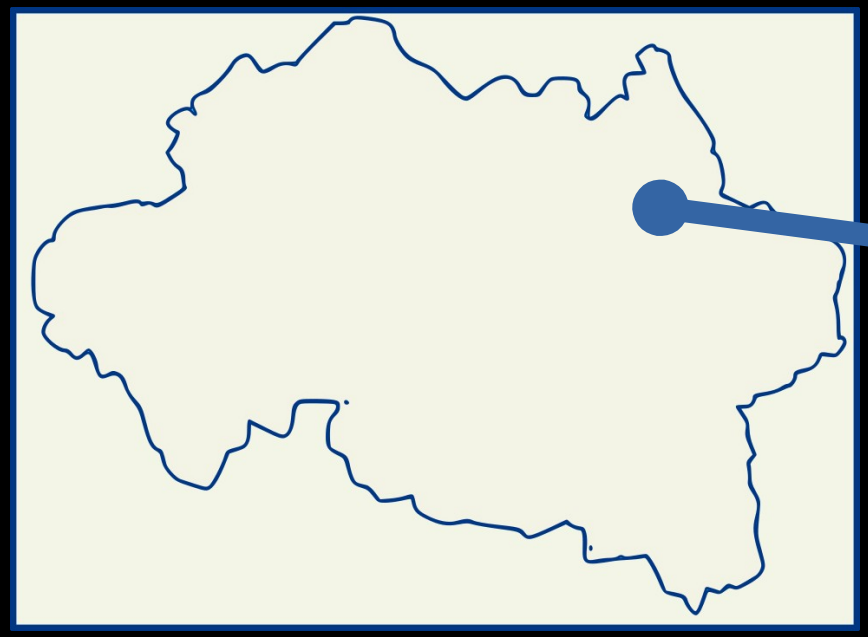
De l’écologie des communautés à la génomique ...

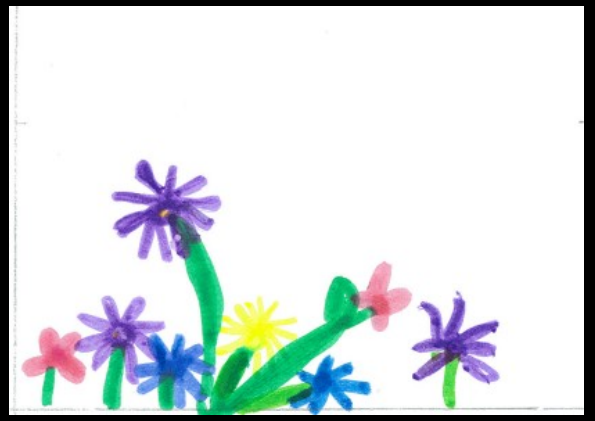


Une communauté

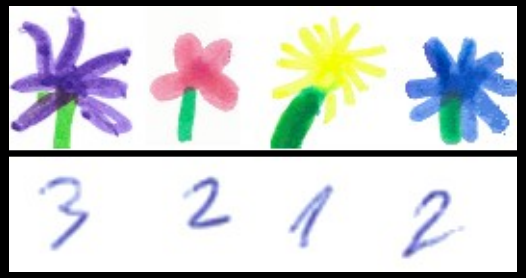


Une communauté



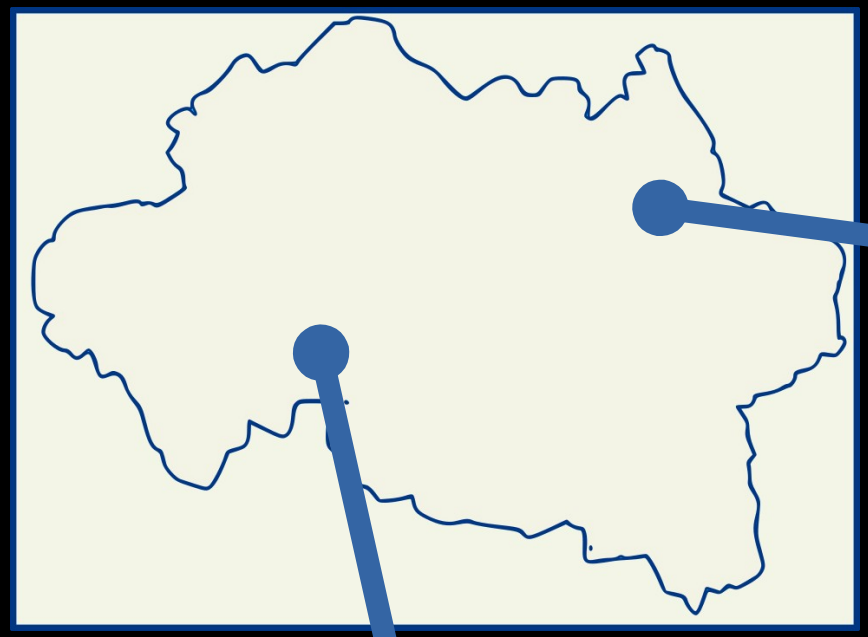


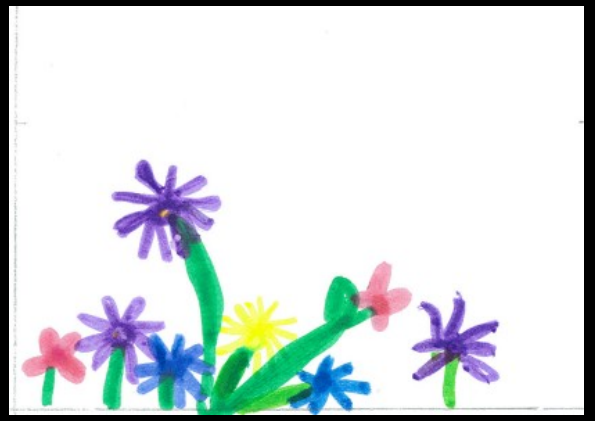
Un vecteur :



Diversité α

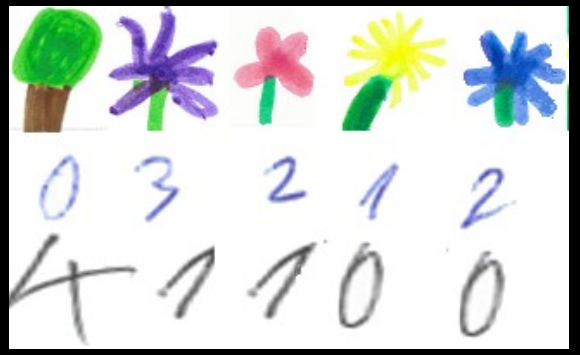
Deux communautés



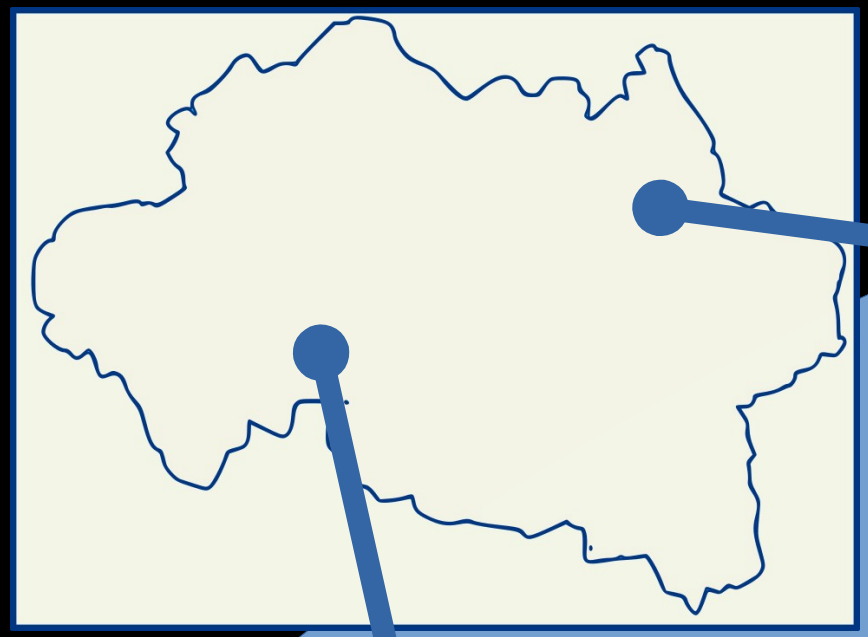


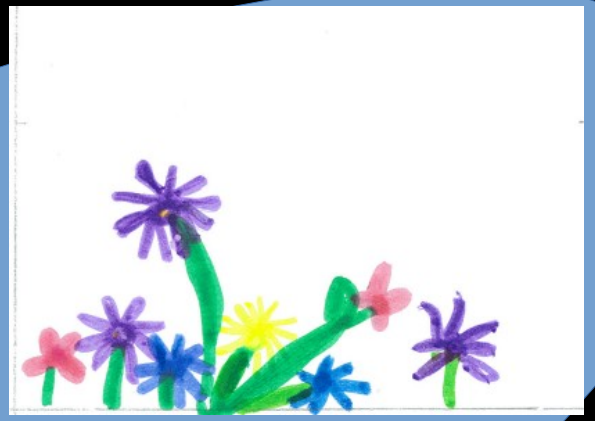
Diversité α_1

Diversité α_2



Deux communautés



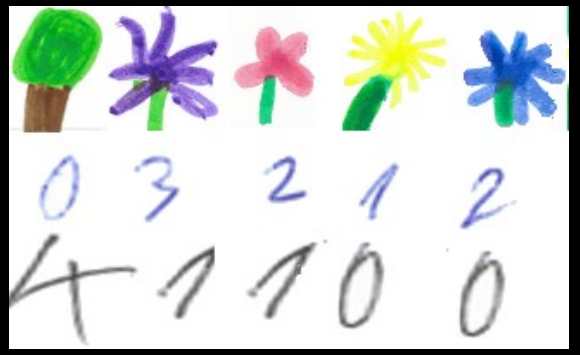


Diversité α_1

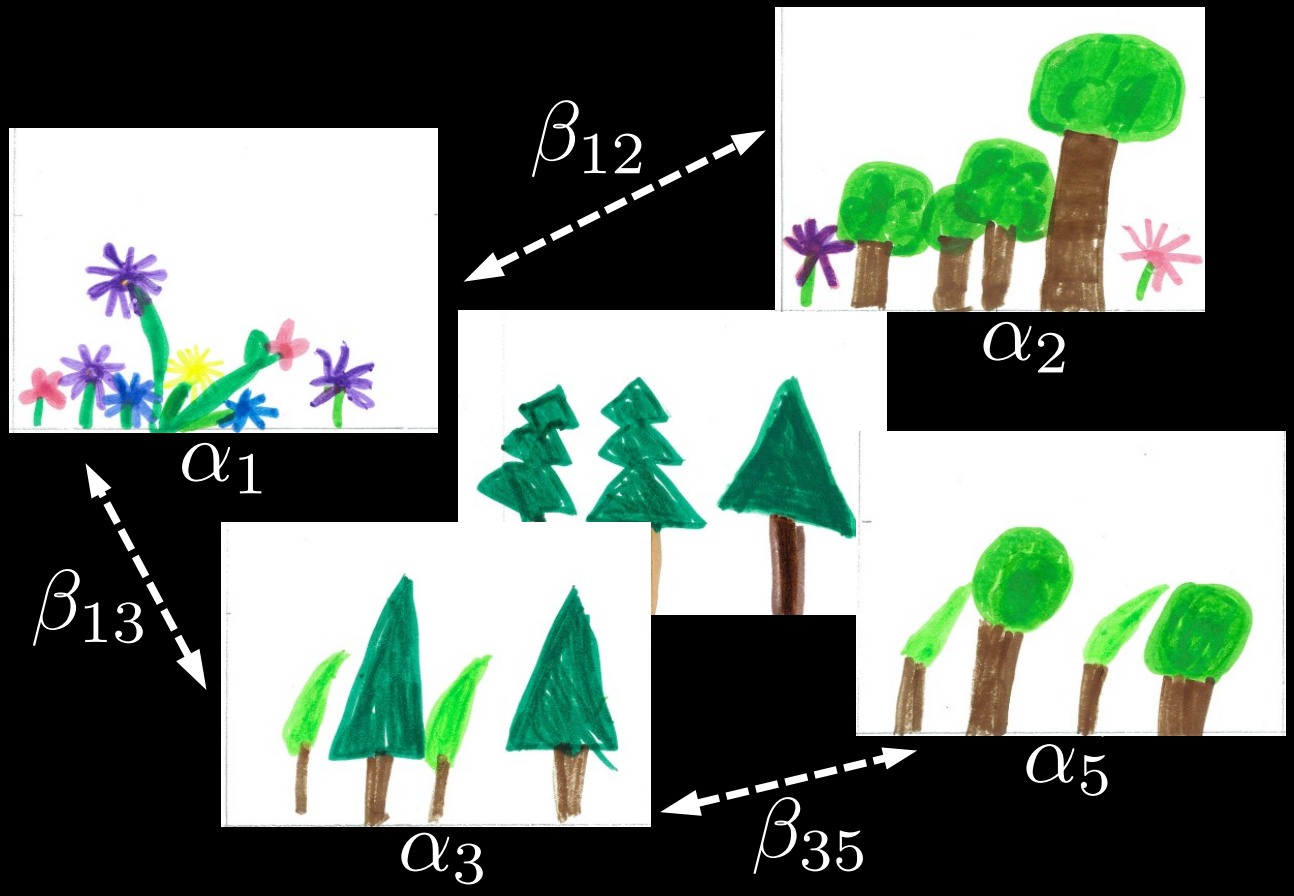
Diversité β_{12}

Diversité α_2

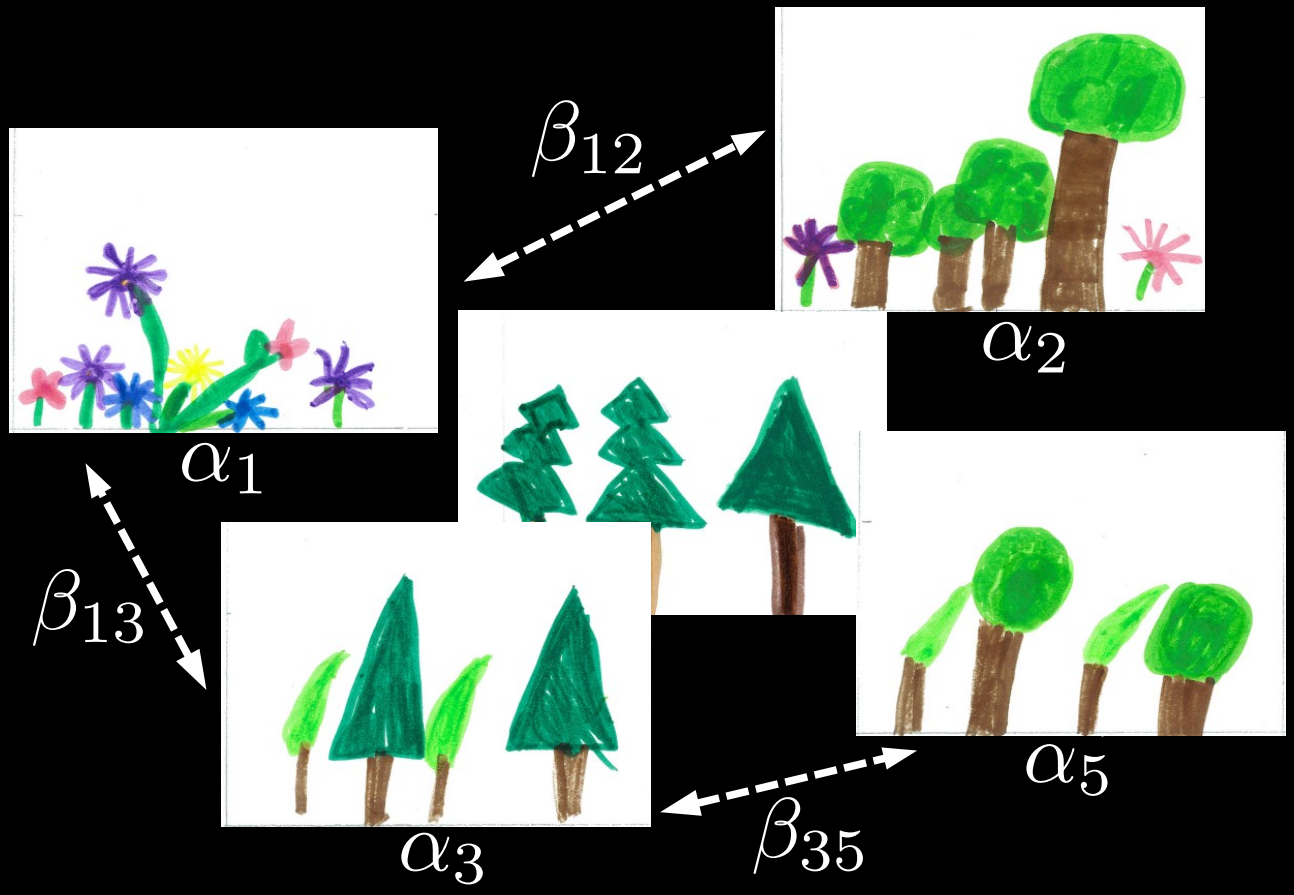
Diversité γ

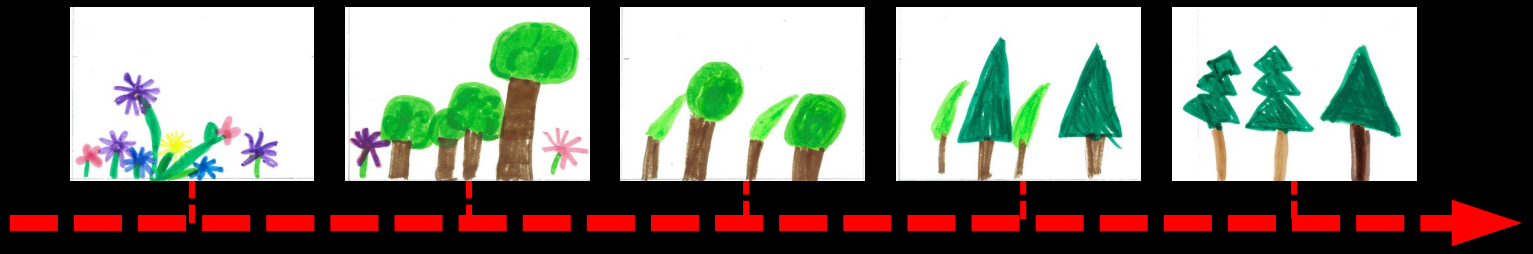


Ordination des communautés

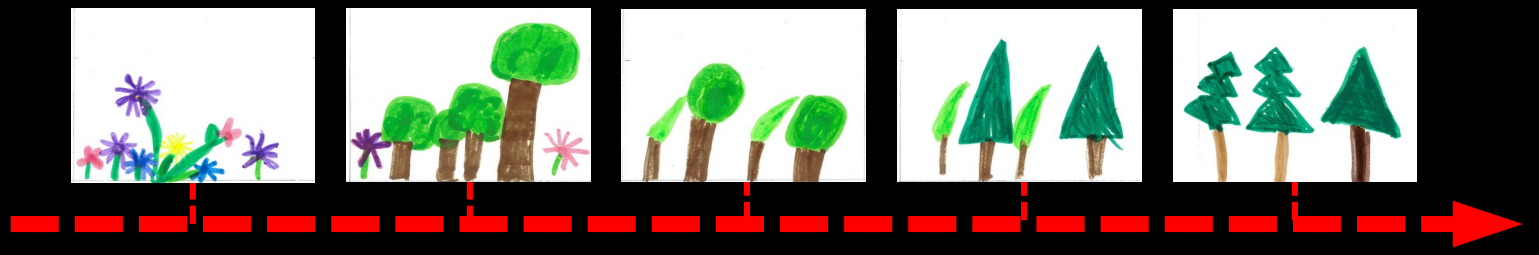


Ordination des communautés

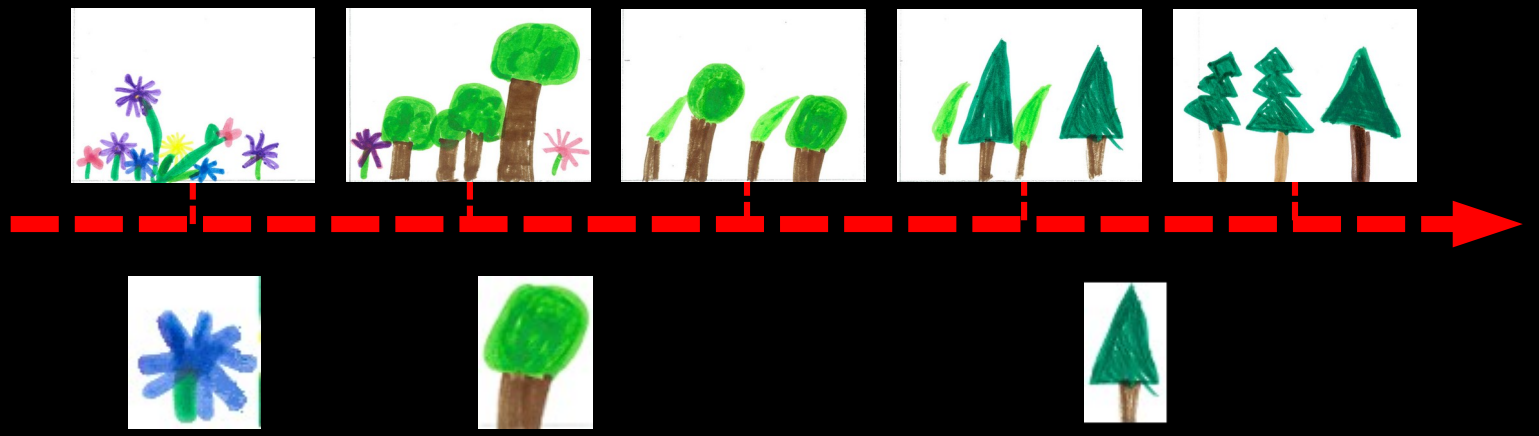




Ordination simultanée

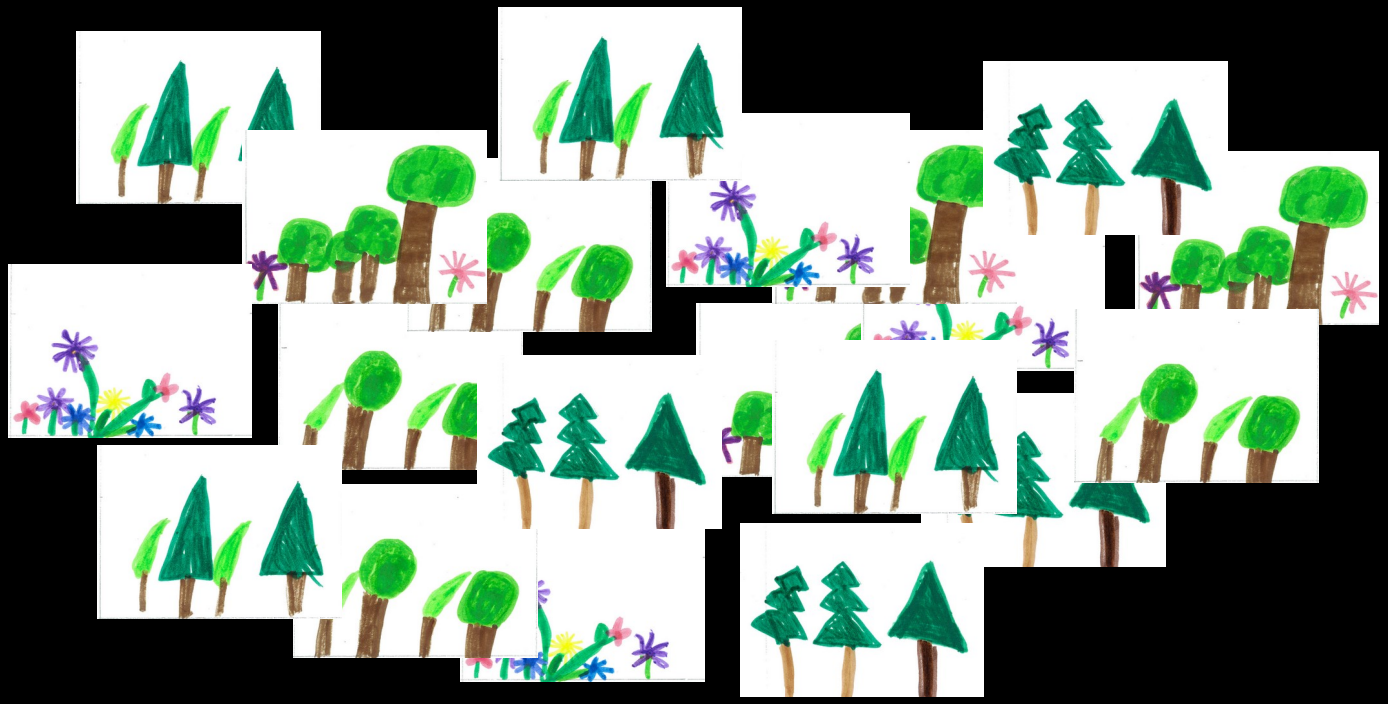


Ordination simultanée

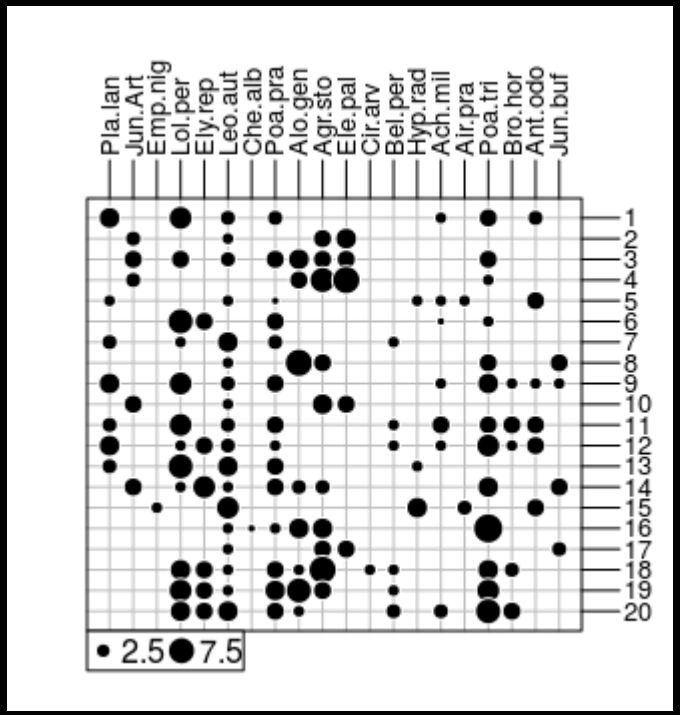


Moyenne pondérée
(*weighted averaging*)

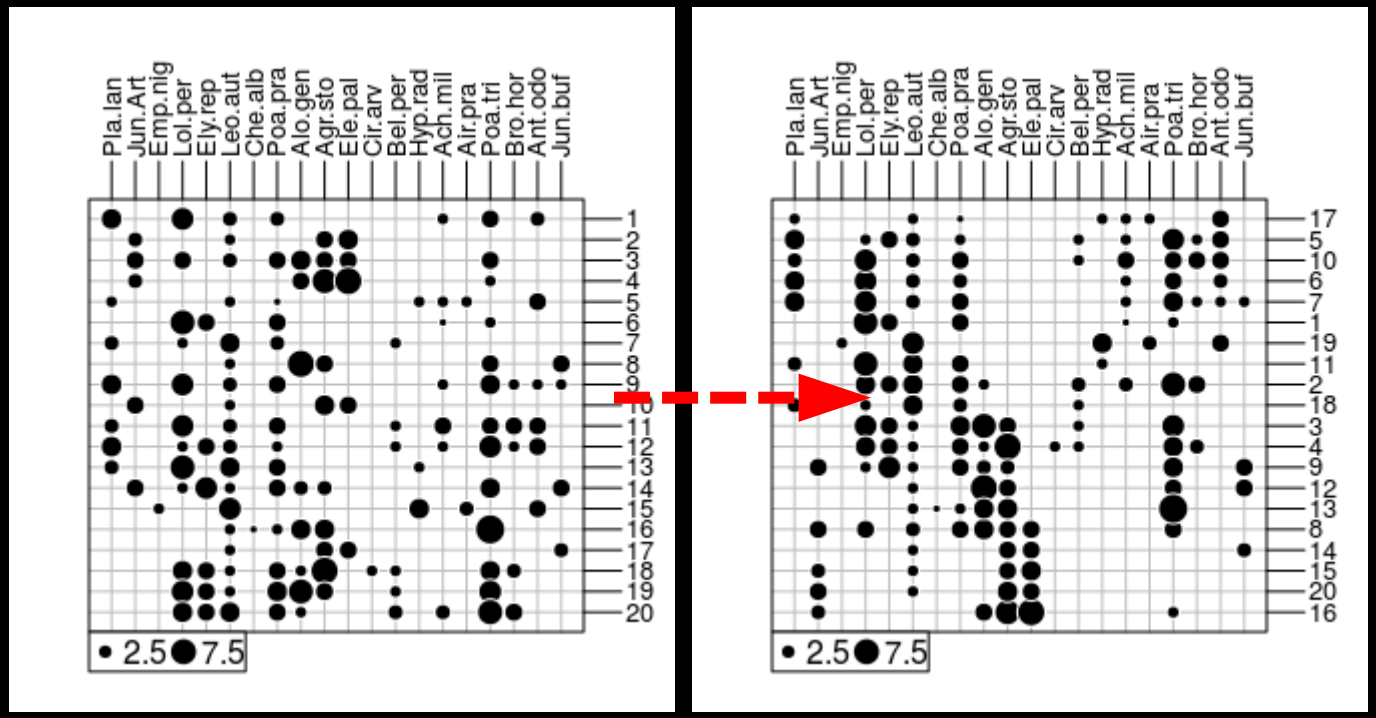
Un casse-tête



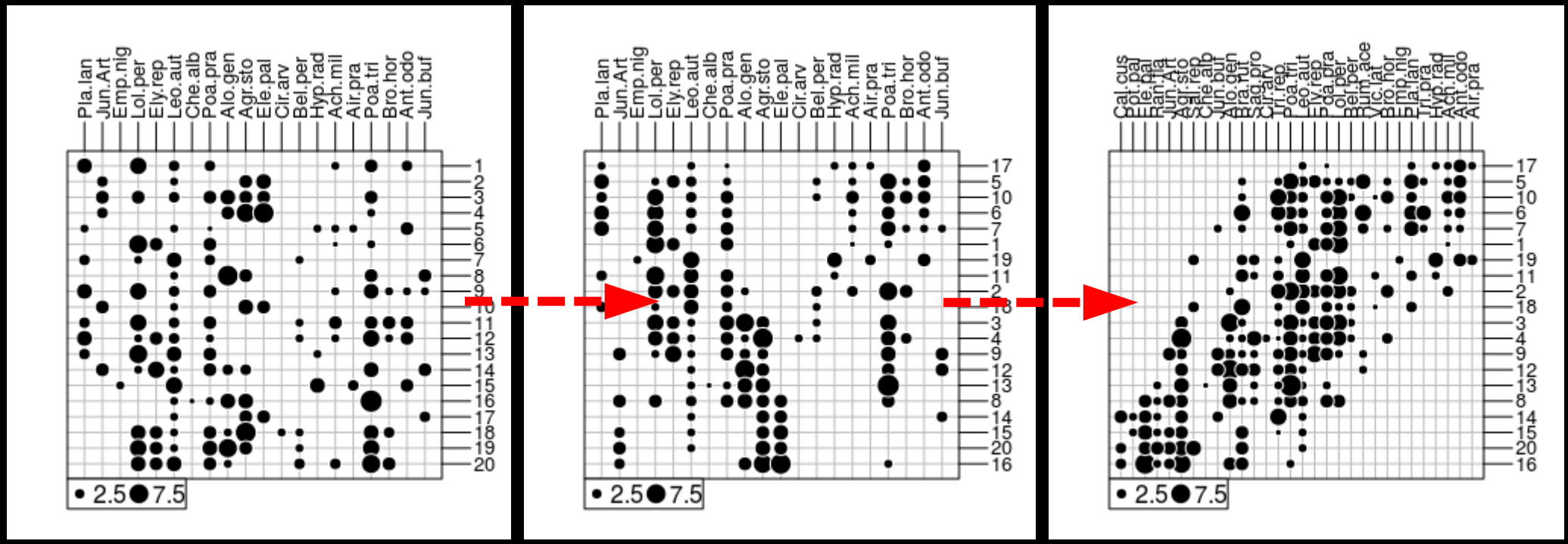
Ordonner un tableau



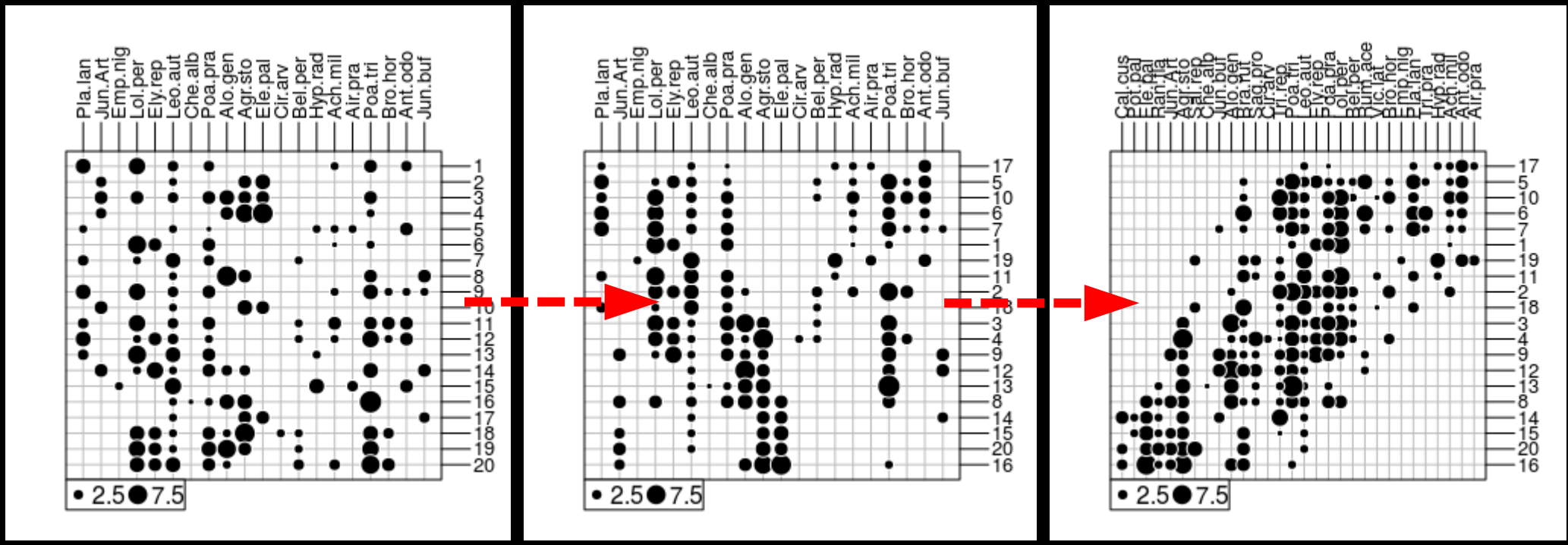
Ordonner un tableau



Ordonner un tableau



Ordonner un tableau



Besoin de méthodes objectives permettant d'ordonner les sites et les espèces et d'identifier les principaux axes de variation de diversité.

D'une question écologique à une problématique méthodologique

OBJECTIVE METHODS FOR THE CLASSIFICATION O
VEGETATION

III. AN ESSAY IN THE USE OF FACTOR ANALYSIS*

By D. W. GOODALL†

(*Manuscript received April 5, 1954*)

GRADIENT ANALYSIS OF VEGETATION*

By

R. H. WHITTAKER

AN ORDINATION STUDY OF
A CHALK GRASSLAND COMMUNITY

By M. P. AUSTIN*

AN ORDINATION OF THE UPLAND FOREST COMMUNITIES OF SOUTHERN WISCONSIN*

J. ROGER BRAY† AND J. T. CURTIS
Department of Botany, University of Minnesota, Minneapolis, Minnesota
Department of Botany, University of Wisconsin, Madison, Wisconsin

GEOMETRIC MODELS IN ECOLOGY

I. THE THEORY AND APPLICATION OF SOME ORDINATION METHODS

By L. ORLOCI

PRINCIPAL COMPONENT ORDINATION AND SIMULATED VEGETATIONAL DATA[1]

I. NOY-MEIR
Department of Biogeography and Geomorphology, Research School of Pacific Studies
Australian National University, Canberra, Australia
AND
M. P. AUSTIN
Division of Land Research, CSIRO, Canberra, Australia

THE APPLICATION OF QUANTITATIVE METHODS TO VEGETATION SURVEY

II. SOME METHODOLOGICAL PROBLEMS OF DATA FROM RAIN FOREST

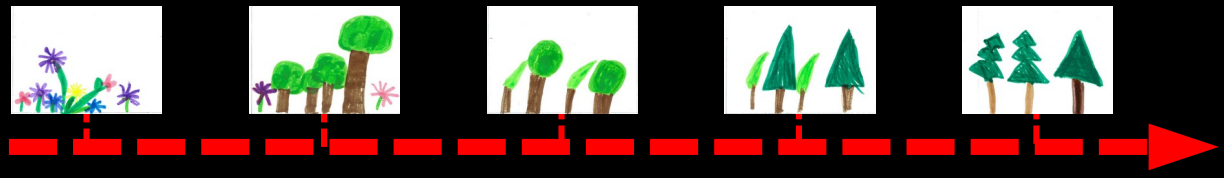
By M. P. AUSTIN* AND P. GREIG-SMITH

GEOMETRIC MODELS IN ECOLOGY

II. AN EVALUATION OF SOME ORDINATION TECHNIQUES

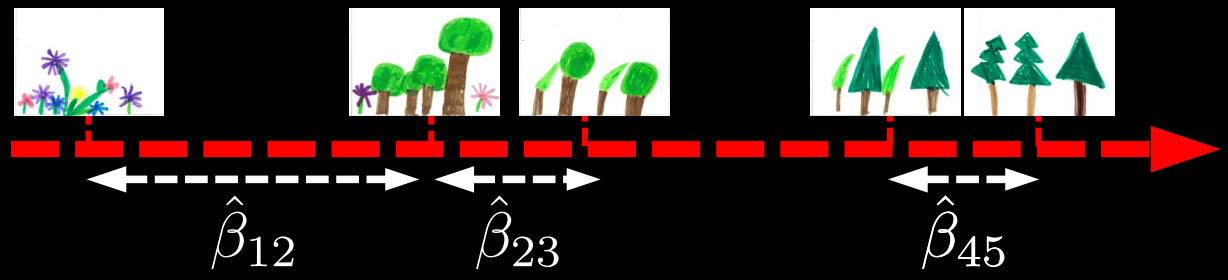
By M. P. AUSTIN AND L. ORLOCI*

L'ordination des tableaux écologiques



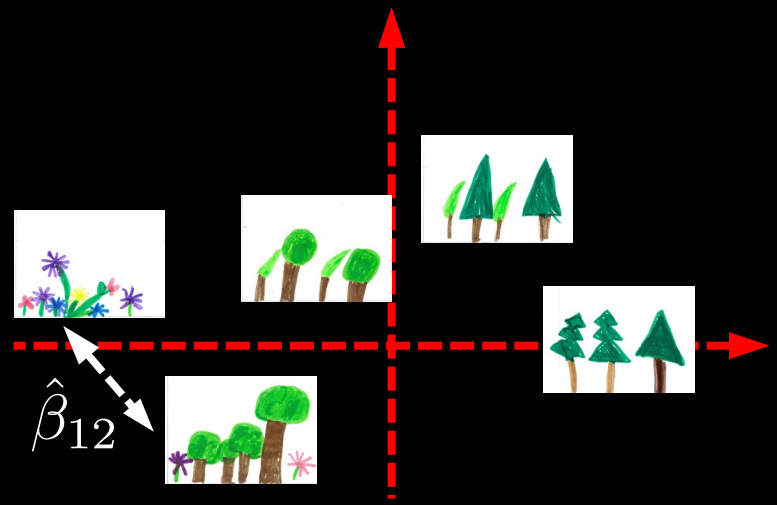
L'ordination des tableaux écologiques

- Scoring



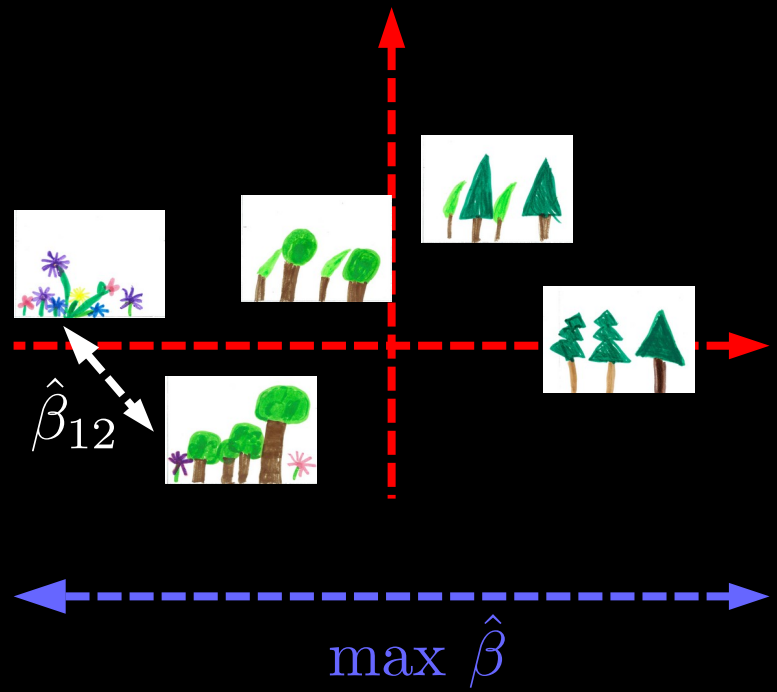
L'ordination des tableaux écologiques

- Scoring
- Multi-dimensionnel



L'ordination des tableaux écologiques

- Scoring
- Multi-dimensionnel
- Maximise la diversité



Analyses multivariées

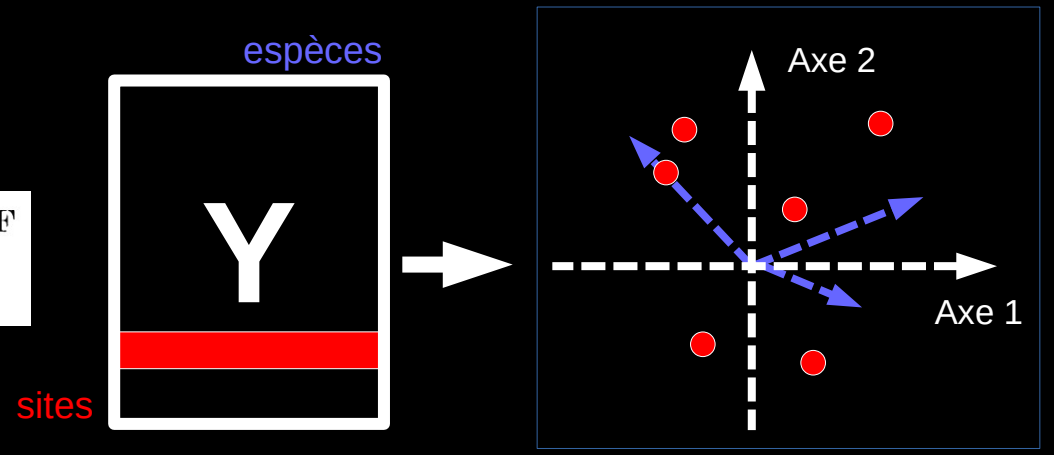
OBJECTIVE METHODS FOR THE CLASSIFICATION OF VEGETATION

III. AN ESSAY IN THE USE OF FACTOR ANALYSIS*

By D. W. GOODALL†

(*Manuscript received April 5, 1954*)

ACP (1954)



AN ORDINATION OF THE UPLAND FOREST COMMUNITIES OF SOUTHERN WISCONSIN*

J. ROGER BRAY† AND J. T. CURTIS

PCoA (1957)

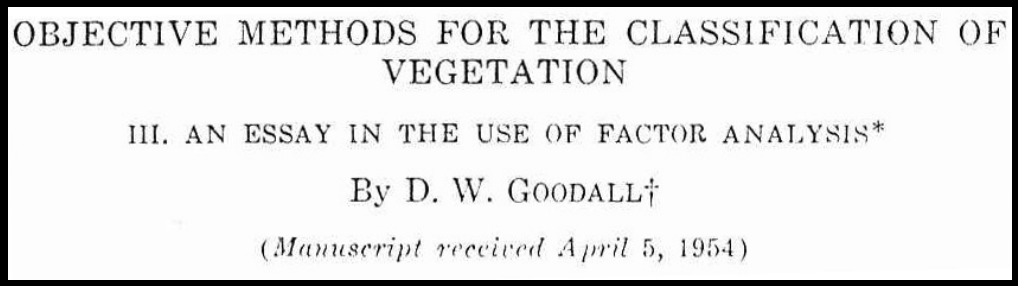
RECIPROCAL AVERAGING: AN EIGENVECTOR METHOD OF ORDINATION

BY M. O. HILL*

AFC (1973)

Analyses multivariées...

… et diversité



OBJECTIVE METHODS FOR THE CLASSIFICATION OF VEGETATION

III. AN ESSAY IN THE USE OF FACTOR ANALYSIS*

By D. W. GOODALL†

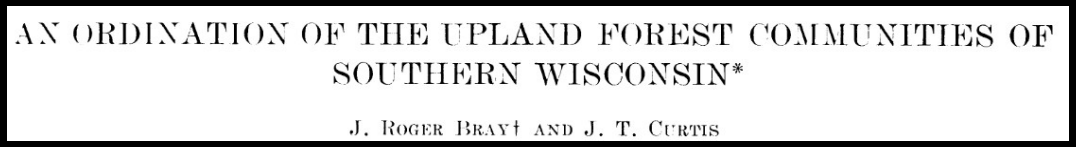
(*Manuscript received April 5, 1954*)

ACP (1954)

PRINCIPAL COMPONENTS BIPLOTS AND ALPHA AND BETA DIVERSITY[1]

CAJO J. F. TER BRAAK

ACP et α-Simpson (1983)



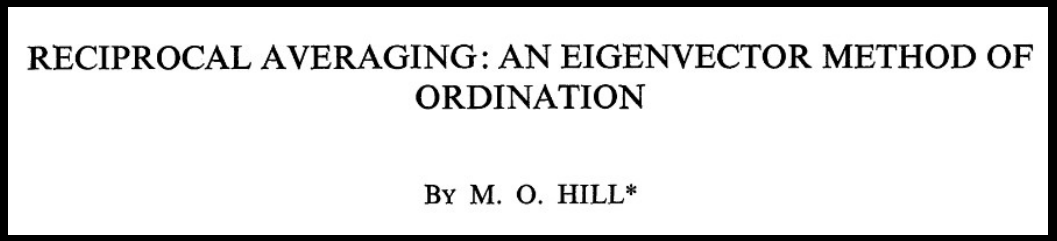
AN ORDINATION OF THE UPLAND FOREST COMMUNITIES OF SOUTHERN WISCONSIN*

J. ROGER BRAY† AND J. T. CURTIS

PCoA (1957)

CONSISTENCY BETWEEN ORDINATION TECHNIQUES AND DIVERSITY MEASUREMENTS: TWO STRATEGIES FOR SPECIES OCCURRENCE DATA

RAPHAËL PÉLISSIER,[1,4] PIERRE COUTERON,[2] STÉPHANE DRAY,[3] AND DANIEL SABATIER[1]



RECIPROCAL AVERAGING: AN EIGENVECTOR METHOD OF ORDINATION

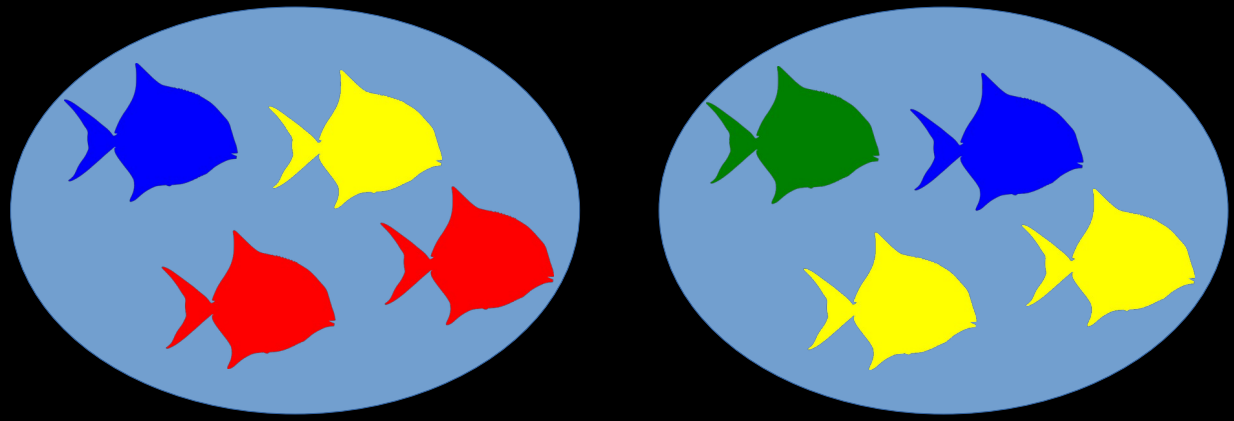
BY M. O. HILL*

AFC (1973)

AFC / ANSC et α,β,ɣ-Richesse / Simpson (2003)

Diversités

Diversité spécifique



espèces

Y

sites

Indice de Simpson

$$Q = 1 - \sum p_i^2$$

Diversités

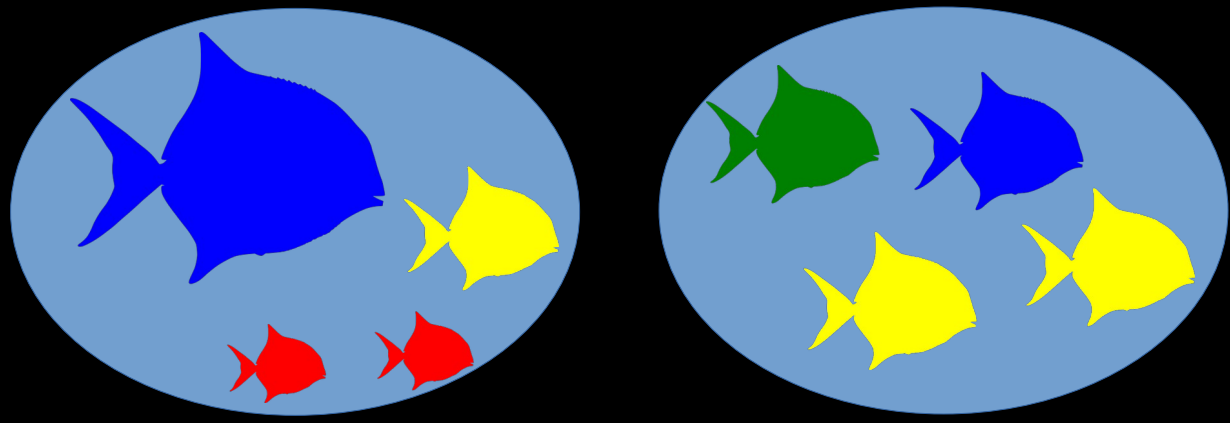
Diversité fonctionelle

D

espèces

Y

sites

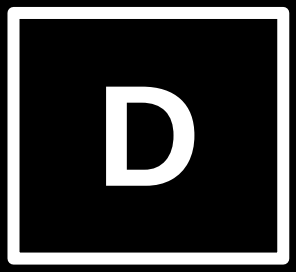


Indice de Rao

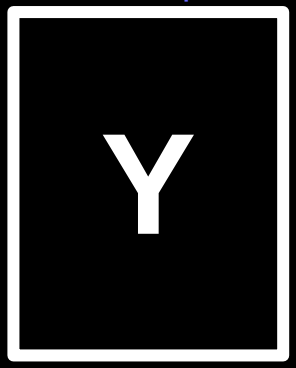
$$Q = \sum \sum p_i p_j d_{ij}^2$$

Analyse multivariées et diversité

DPCoA



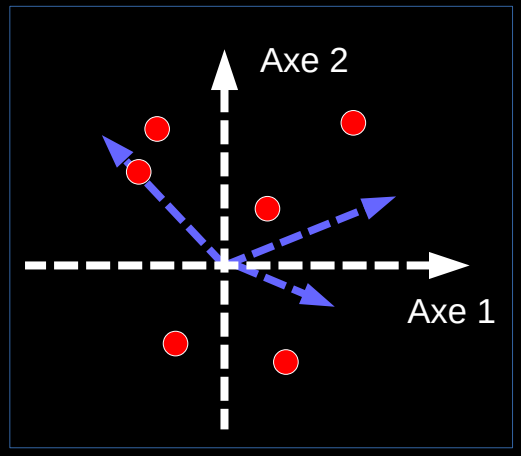
espèces



sites

From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis

Sandrine Pavoine*,*, Anne-Béatrice Dufour*, Daniel Chessel*



Inertie totale = indice de Rao

Réconcilier statistiques et écologie

« *A major purpose is* ***interpretation*** *[…]*

and not simply the representation of numerical relationships among samples or species in a hyperspace with a limited number of axes. »

Patrons ↔ Facteurs / Processus

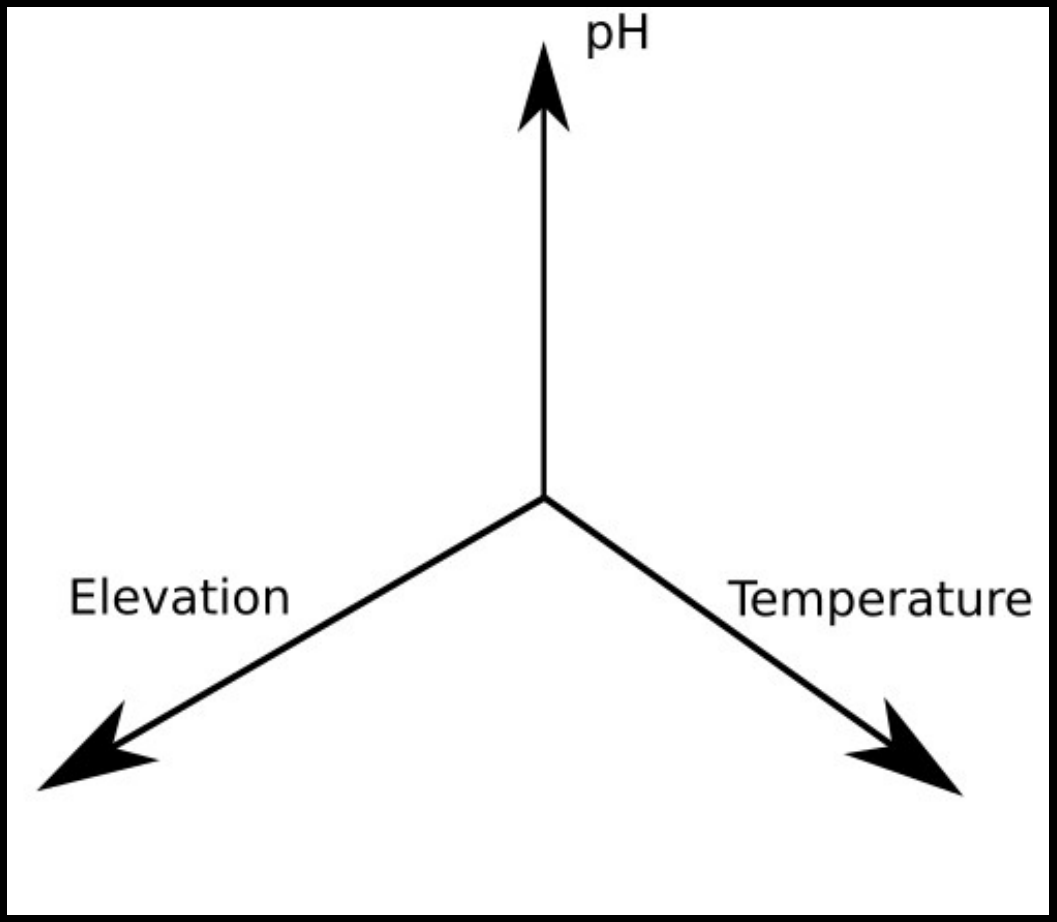
Gauch & Wentworth (1976) Vegetatio

Réconcilier statistiques et écologie

« *A major purpose is* ***interpretation of community relationships to environment,*** *and not simply the representation of numerical relationships among samples or species in a hyperspace with a limited number of axes.* »

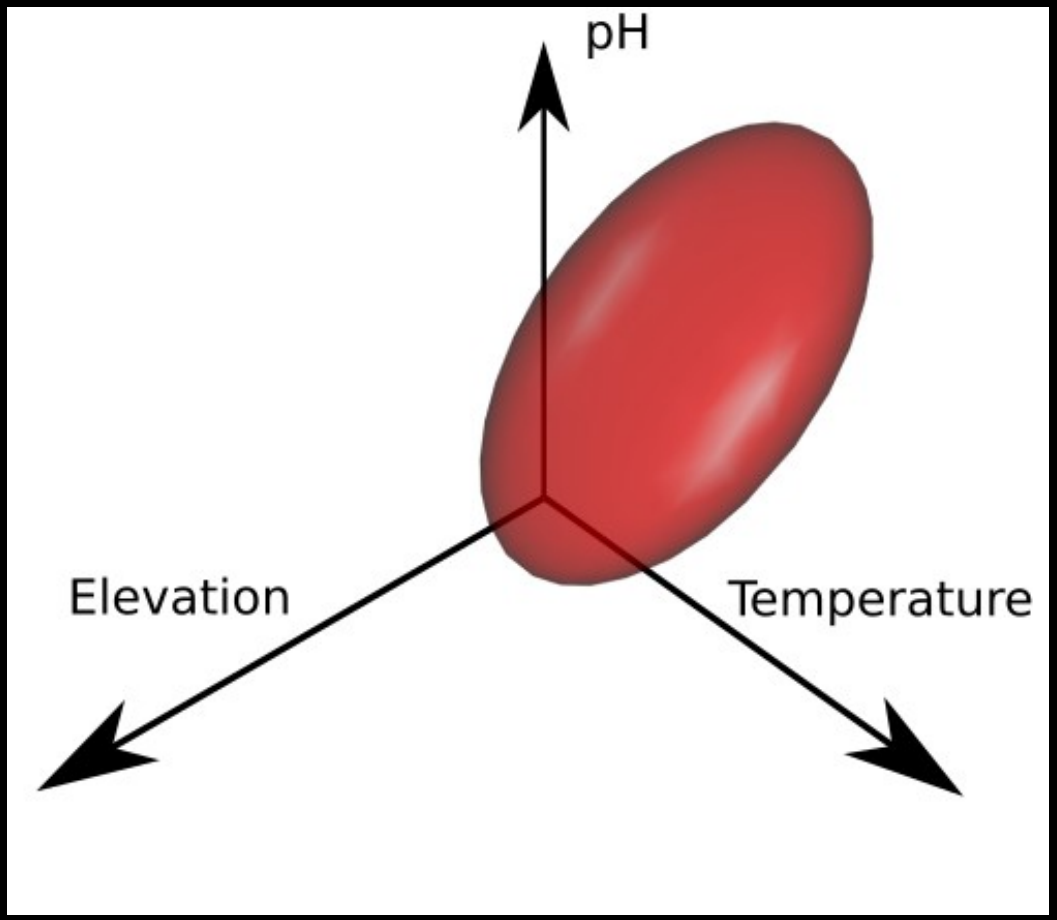
Gauch & Wentworth (1976) Vegetatio

The niche theory



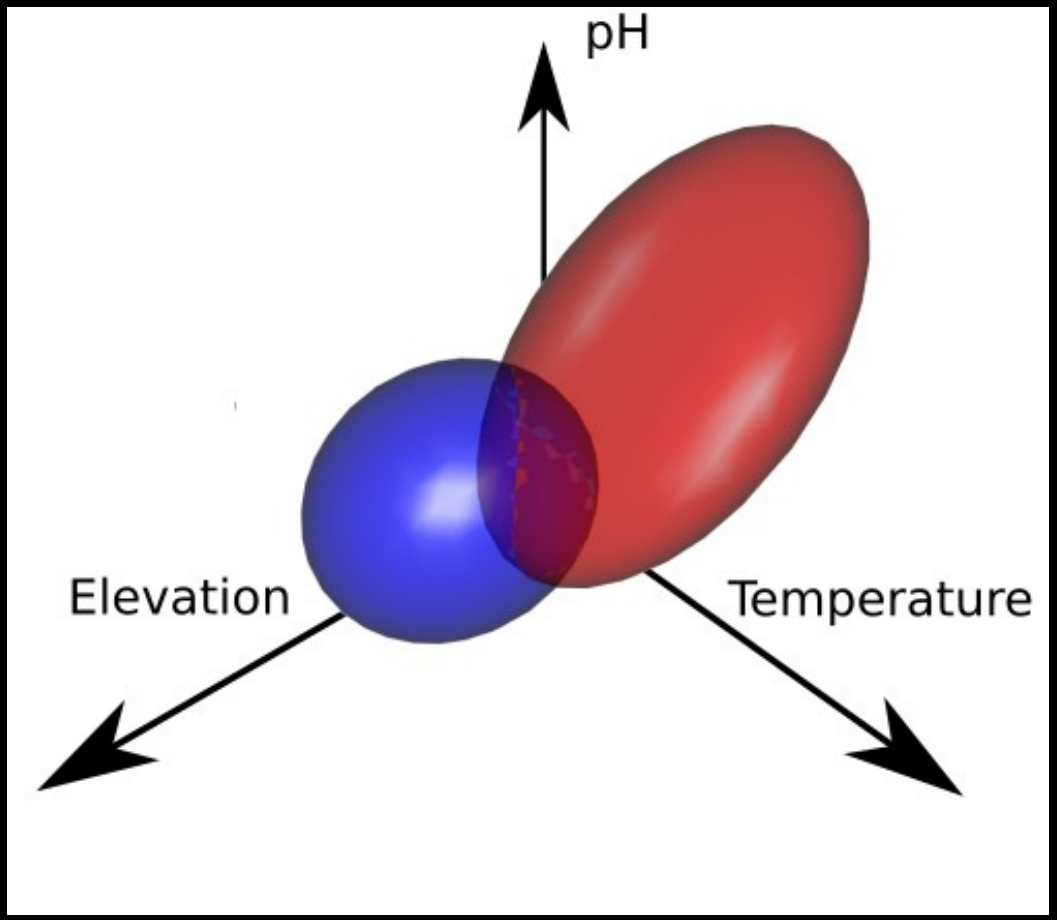
Hutchinson (1957) Population studies:

The niche theory



Hutchinson (1957) Population studies: animal ecology and demography

The niche theory



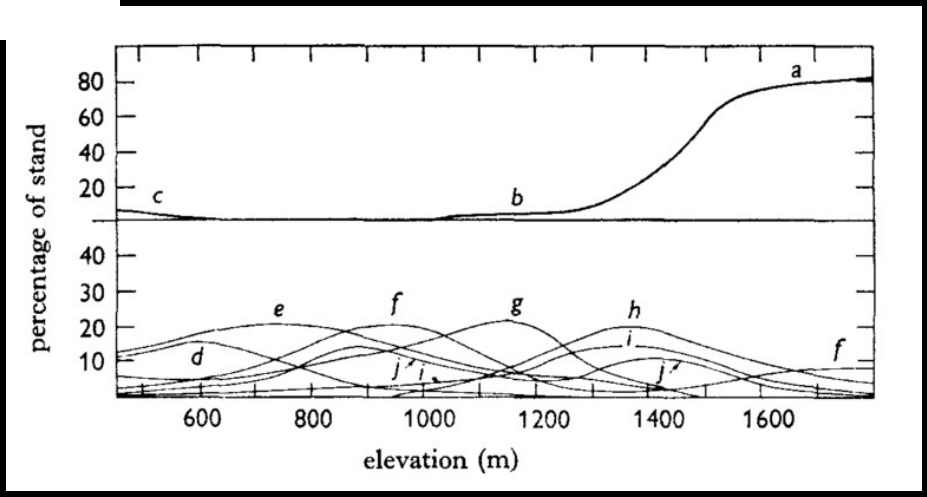
Hutchinson (1957) Population studies: animal ecology and demography

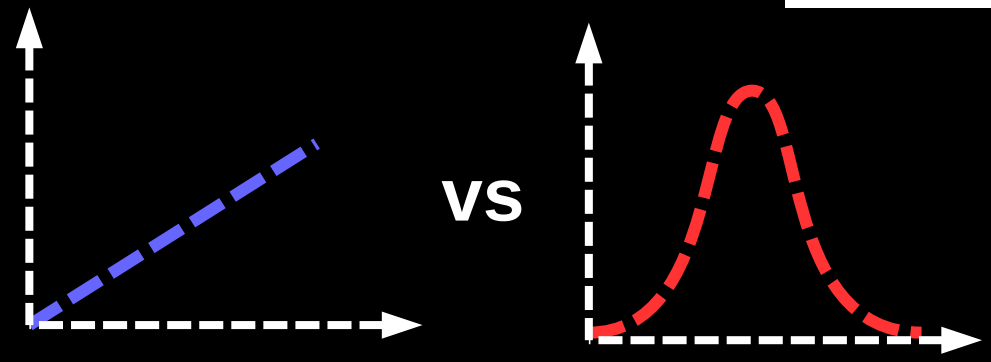
Gradient sous-jacent

GRADIENT ANALYSIS OF VEGETATION*

By

R. H. WHITTAKER





Whittaker (1967) Biological Reviews

Vers une ordination optimale

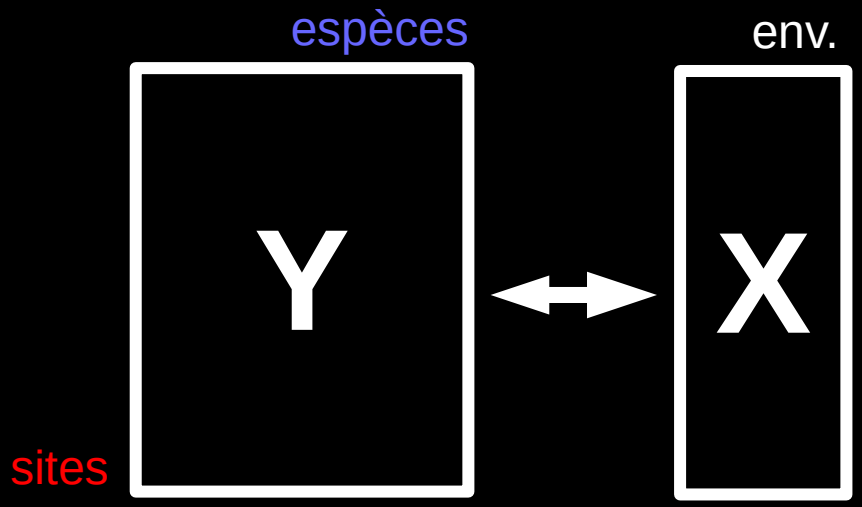
« *Ecologists would welcome a technique that could (a) extract* ***main directions of community-and-environmental variation*** *on the basis of* ***simultaneous consideration*** *of vegetational and environmental data and (b) yield, in the axes extracted, an effective and* ***interpretable*** *ordination without the vulnerability to distortion that affects some other techniques.* »

Gauch & Wentworth (1976) Vegetatio

Ordination directe

CANONICAL CORRESPONDENCE ANALYSIS:
A NEW EIGENVECTOR TECHNIQUE FOR MULTIVARIATE
DIRECT GRADIENT ANALYSIS[1]

CAJO J. F. TER BRAAK



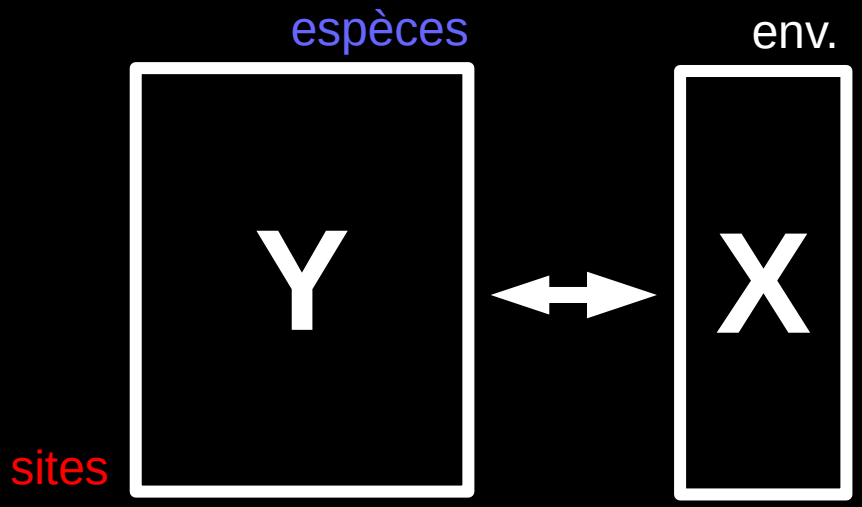
Ordination directe

CANONICAL CORRESPONDENCE ANALYSIS:
A NEW EIGENVECTOR TECHNIQUE FOR MULTIVARIATE
DIRECT GRADIENT ANALYSIS[1]

CAJO J. F. TER BRAAK

Contrainte environnementale

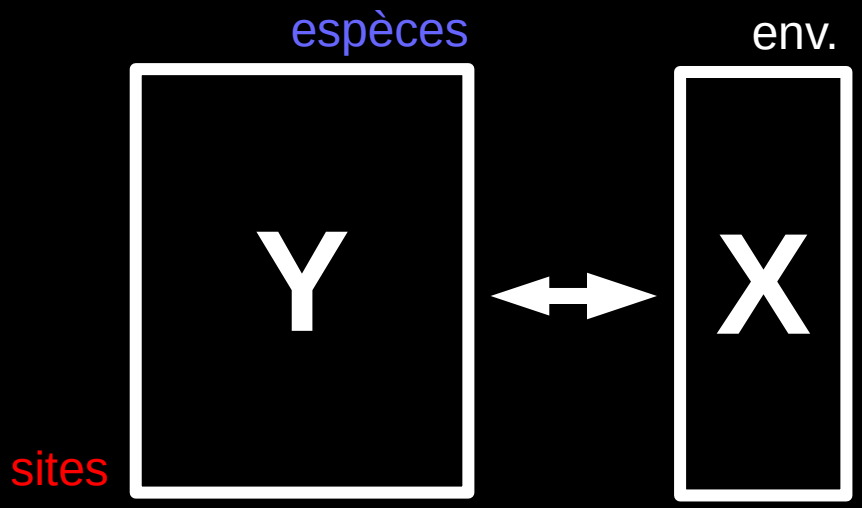
$\mathbf{z} = f(\mathbf{X}) = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \ldots + a_P\mathbf{x}_p$

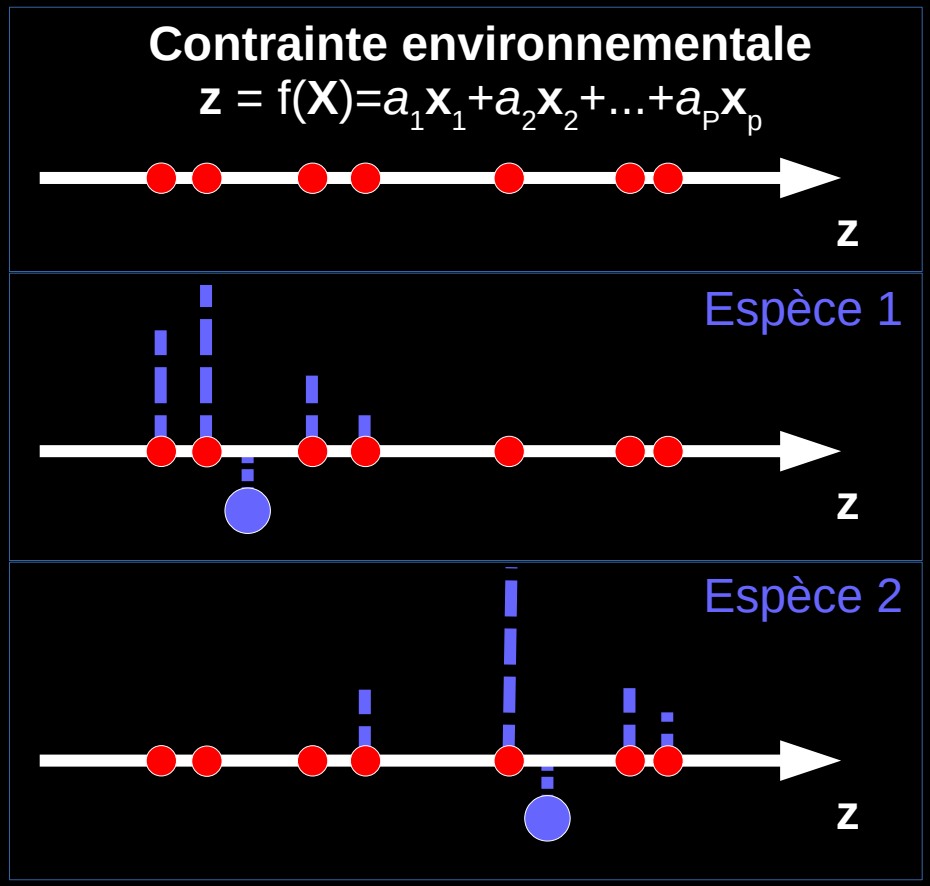


Ordination directe

CANONICAL CORRESPONDENCE ANALYSIS:
A NEW EIGENVECTOR TECHNIQUE FOR MULTIVARIATE
DIRECT GRADIENT ANALYSIS[1]

CAJO J. F. TER BRAAK

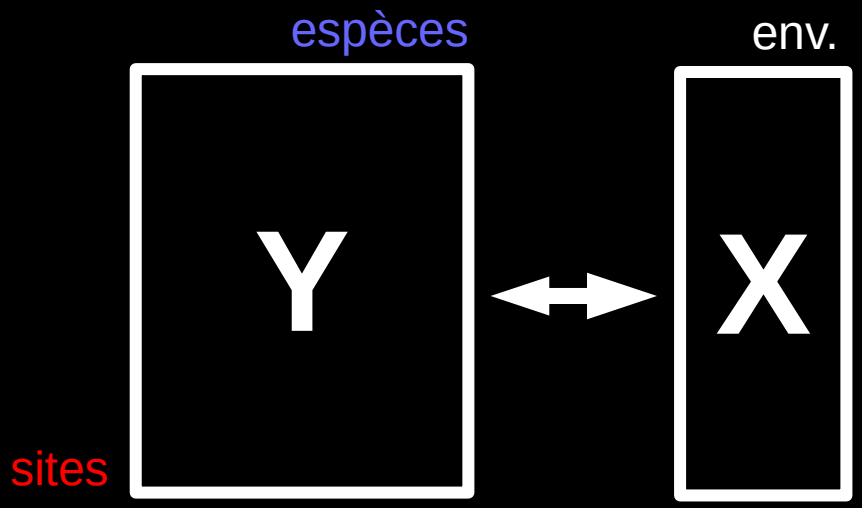


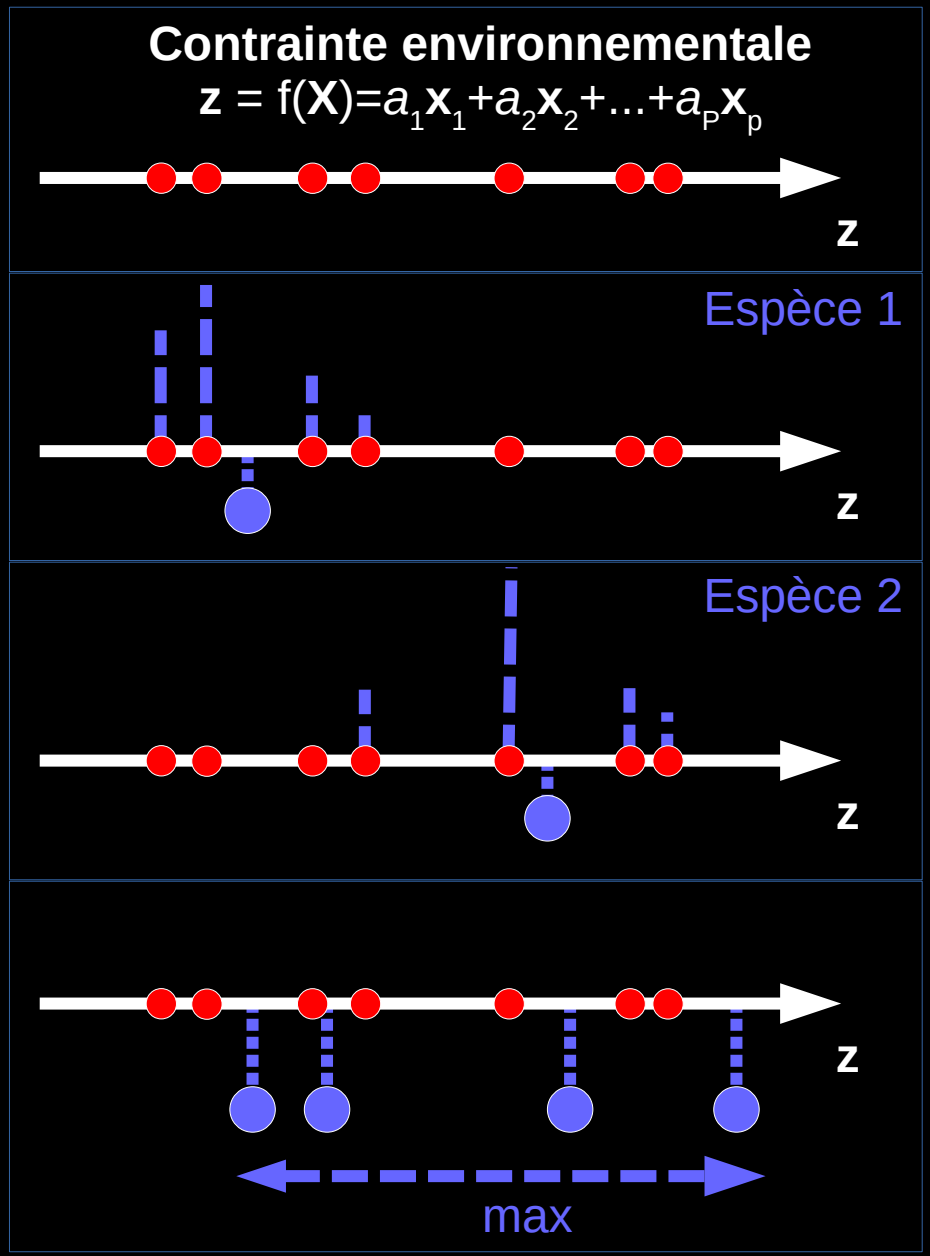


Ordination directe

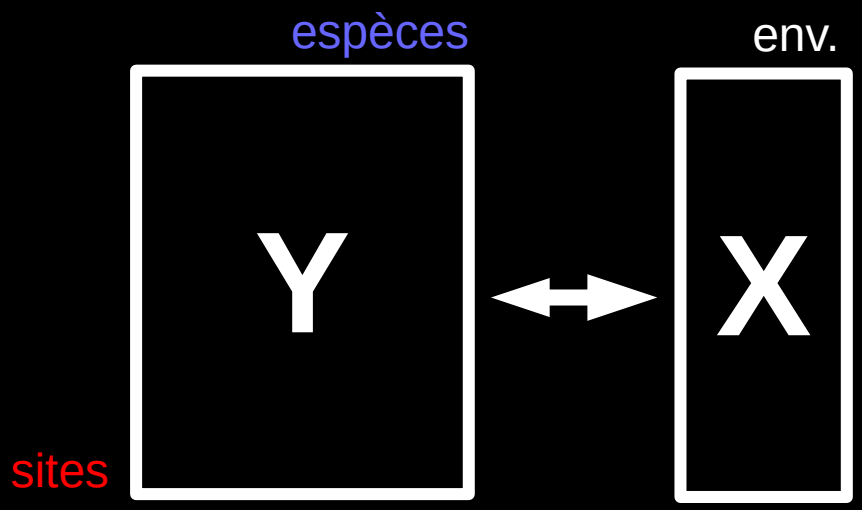
CANONICAL CORRESPONDENCE ANALYSIS:
A NEW EIGENVECTOR TECHNIQUE FOR MULTIVARIATE
DIRECT GRADIENT ANALYSIS[1]

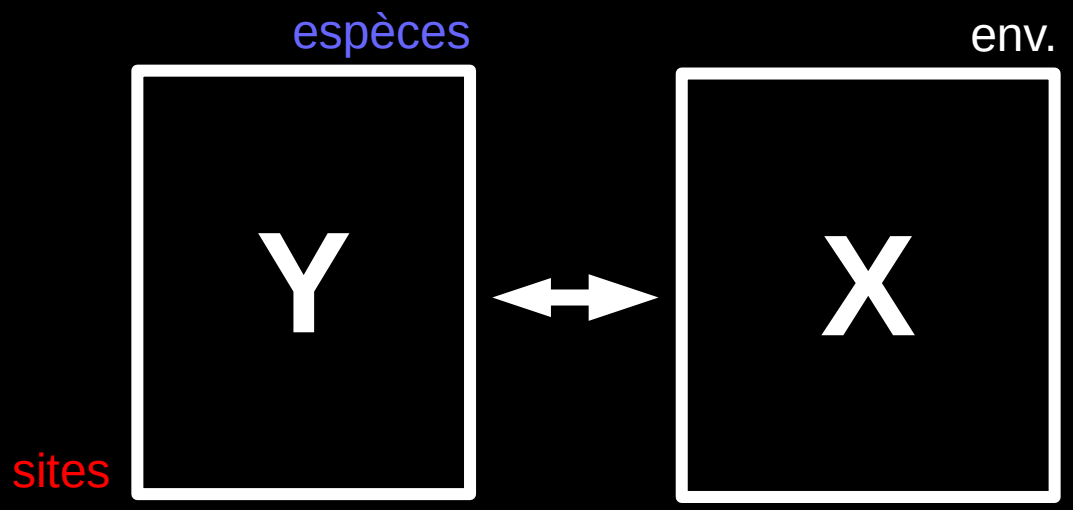
CAJO J. F. TER BRAAK





Régression ou co-variation





ACPVI
- analyse des redondances
- analyse canonique des correspondances

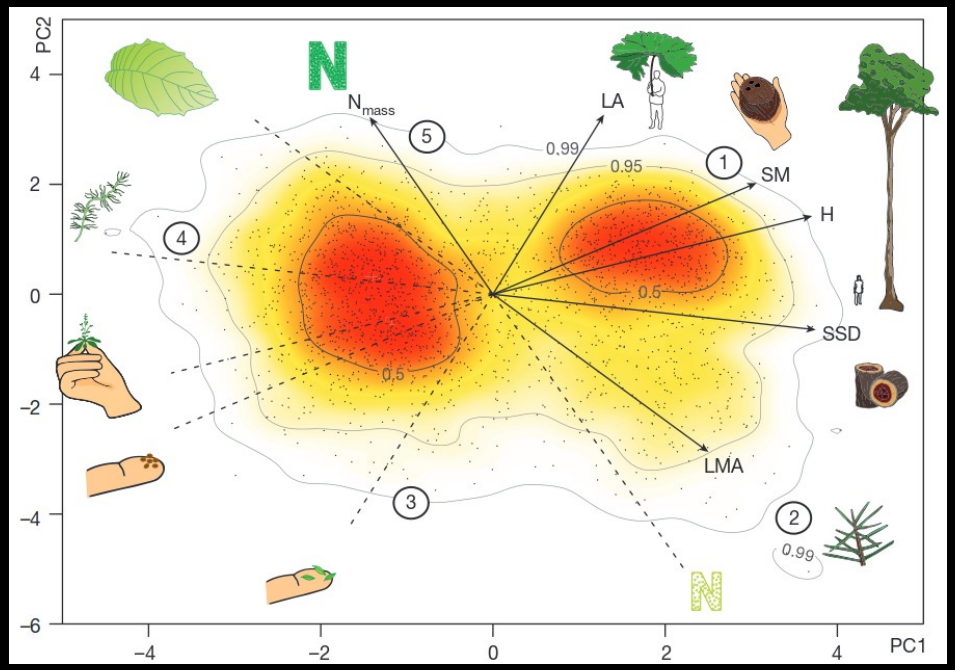
Co-inertia analysis: an alternative method for studying species–environment relationships

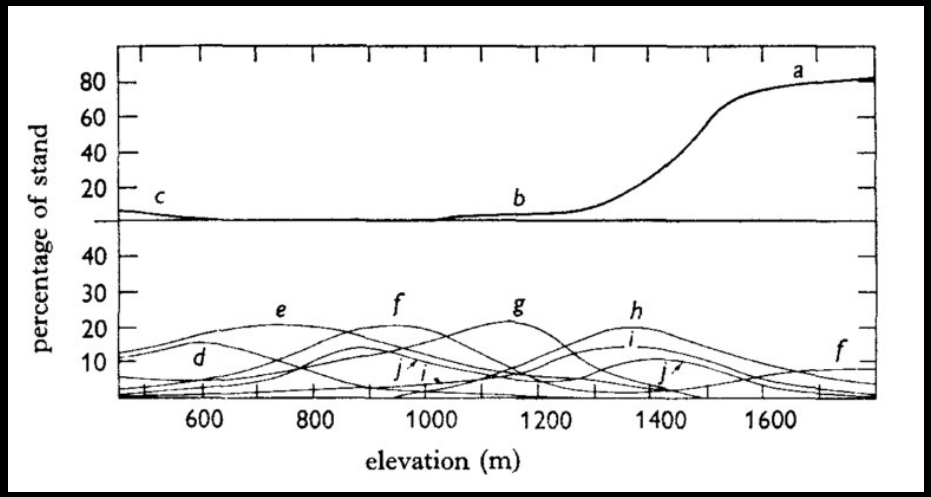
SYLVAIN DOLÉDEC* AND DANIEL CHESSEL

max $R^2_{Y/X}$

max cov^2 (Ya, Xb)

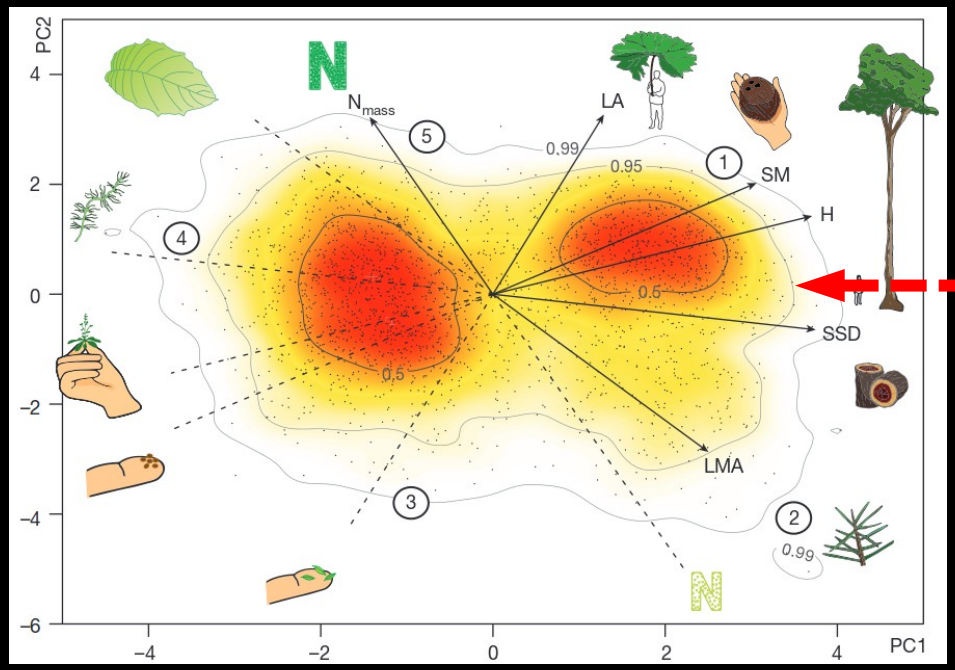
Lien traits-environnement

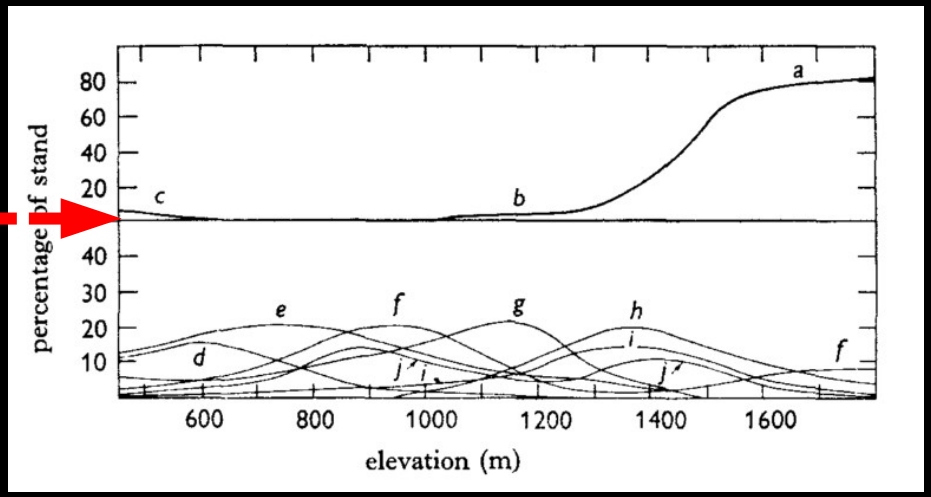




Diaz et al (2016) Nature

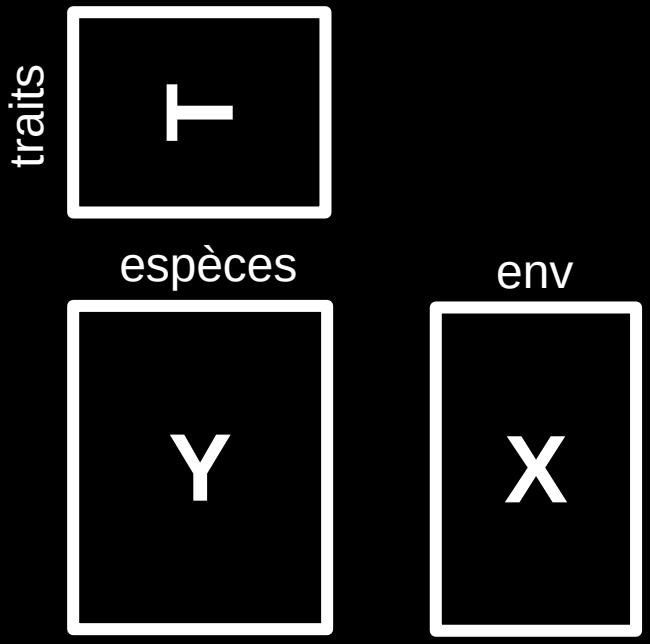
Lien traits-environnement



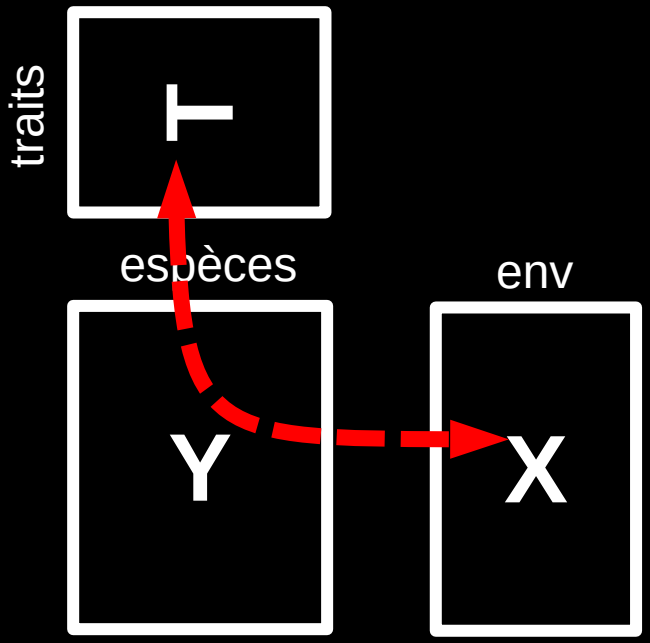


Diaz et al (2016) Nature

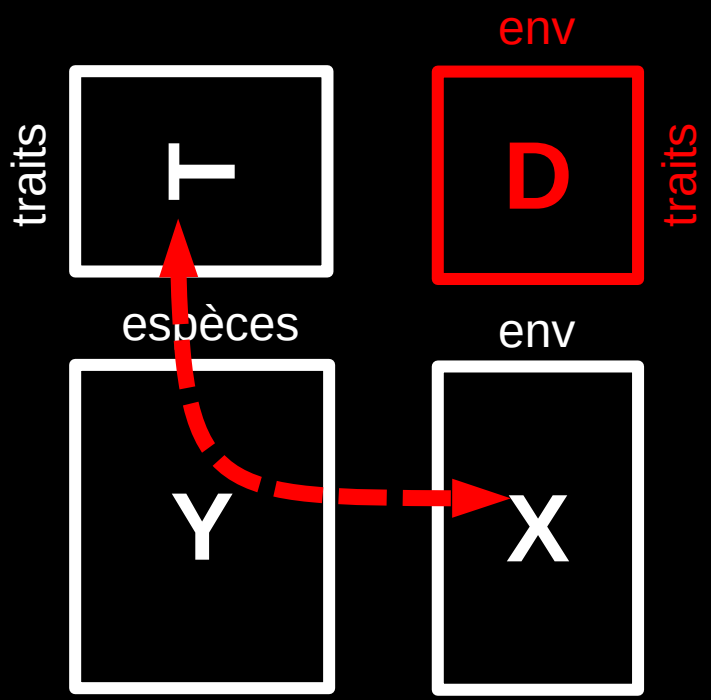
Le quatrième coin



Le quatrième coin



Le quatrième coin



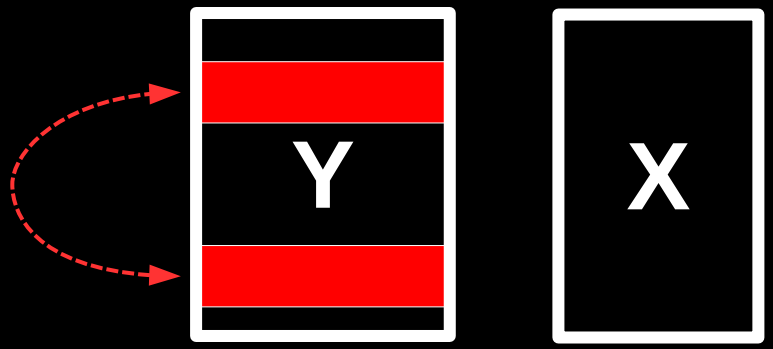
Le quatrième coin :

$$\mathbf{D} = \mathbf{T}^\top \mathbf{Y}^\top \mathbf{X}$$

- Ordination : analyse RLQ
- Test d’hypothèses par permutation

Test d'hypothèses

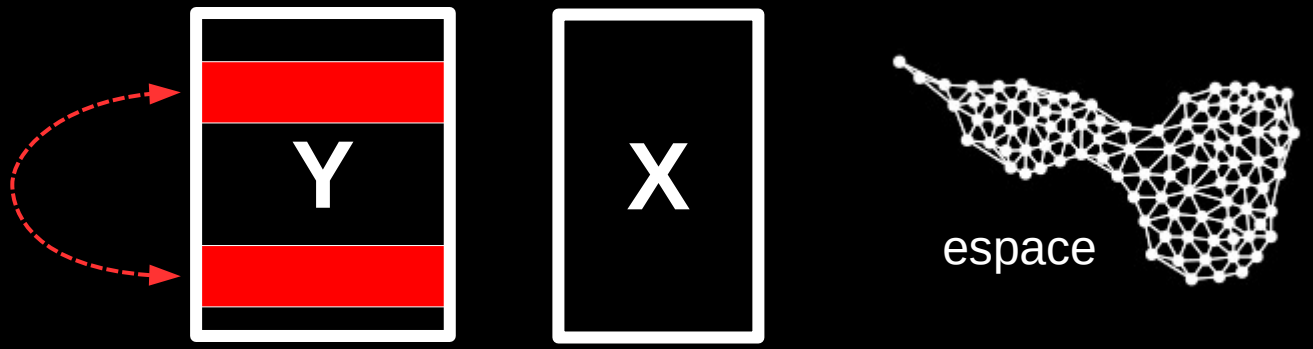
H_o : **Y** ⇹ **X**
H_1 : **Y** ↔ **X**



- Compute a statistic S_{obs} on observed data (e.g., total inertia)
- Randomize data (e.g., permutations)
- Compute S_{sim} on randomized data
- Repeat 2-3 a number of times (e.g., 999)
- Compare S_{obs} to the distribution of S_{sim} and take the appropriate decision

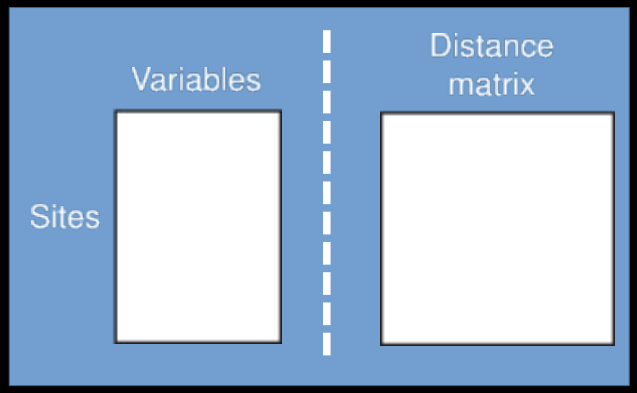
Test d'hypothèses

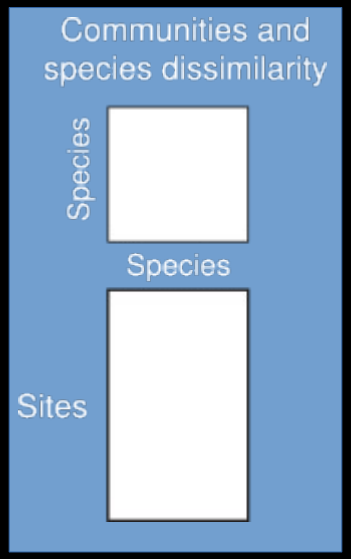
H_o : **Y ⇹ X**
H_1 : **Y ↔ X**

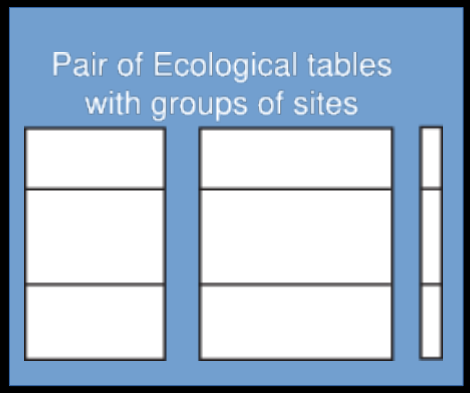


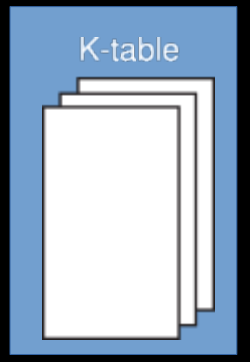
- Compute a statistic S_{obs} on observed data (e.g., total inertia)
- Randomize data (e.g., constrained permutations)
- Compute S_{sim} on randomized data
- Repeat 2-3 a number of times (e.g., 999)
- Compare S_{obs} to the distribution of S_{sim} and take the appropriate decision

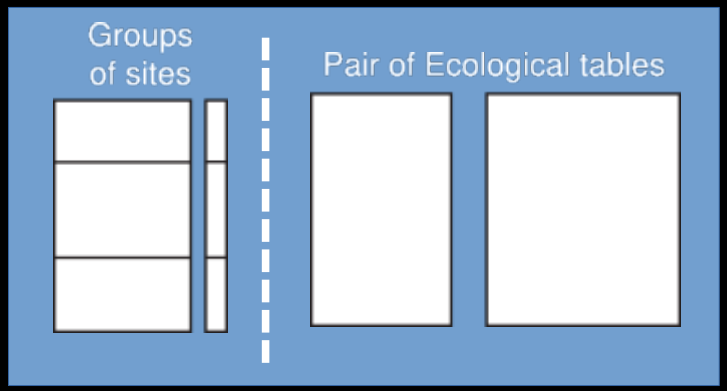
Plus de structures de données variées, plus de méthodes

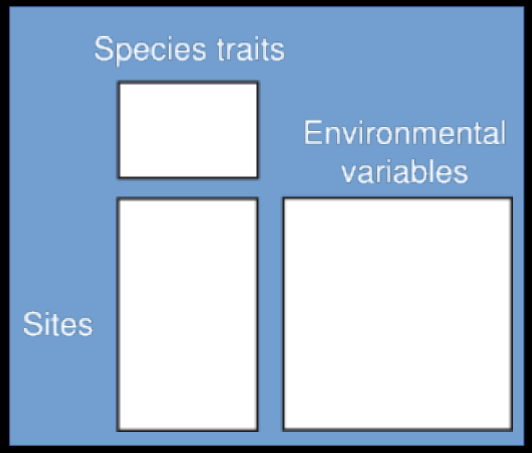


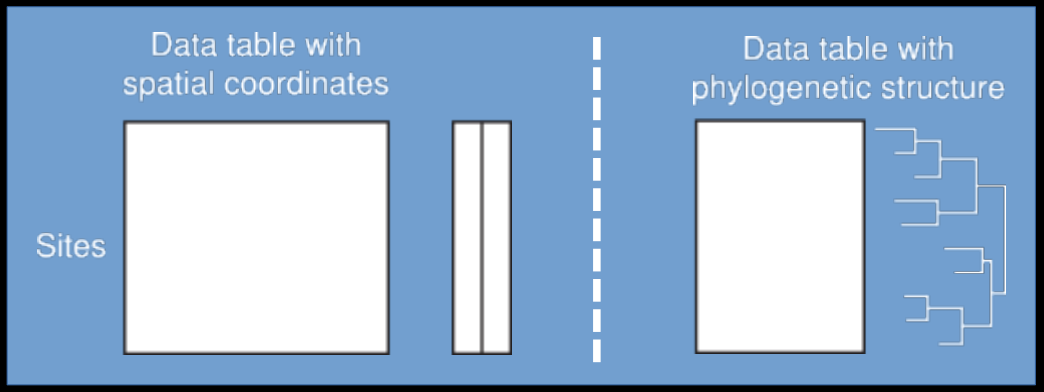


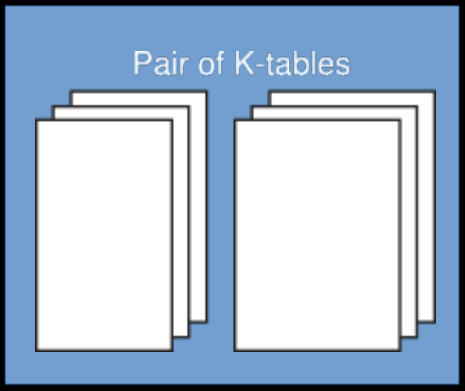










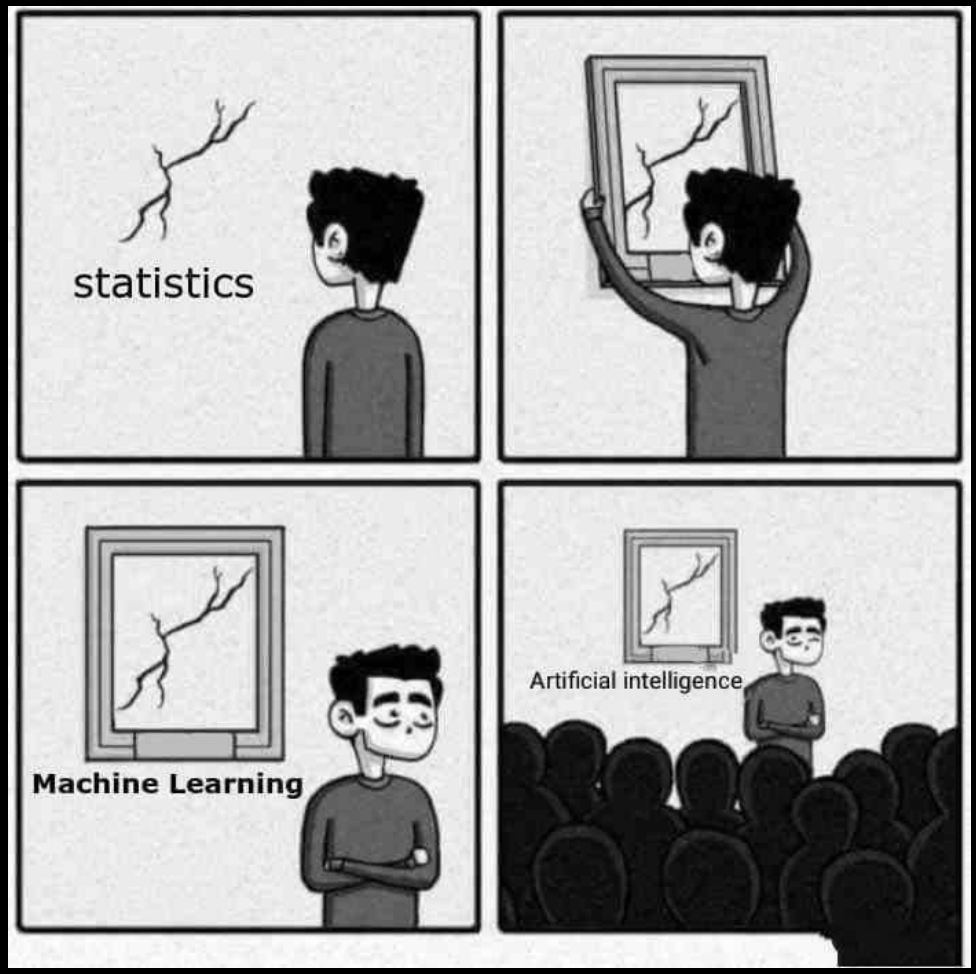


Plus de questions, plus de données

Données massives, hétérogènes, spatio-temporelle, etc.

En résumé

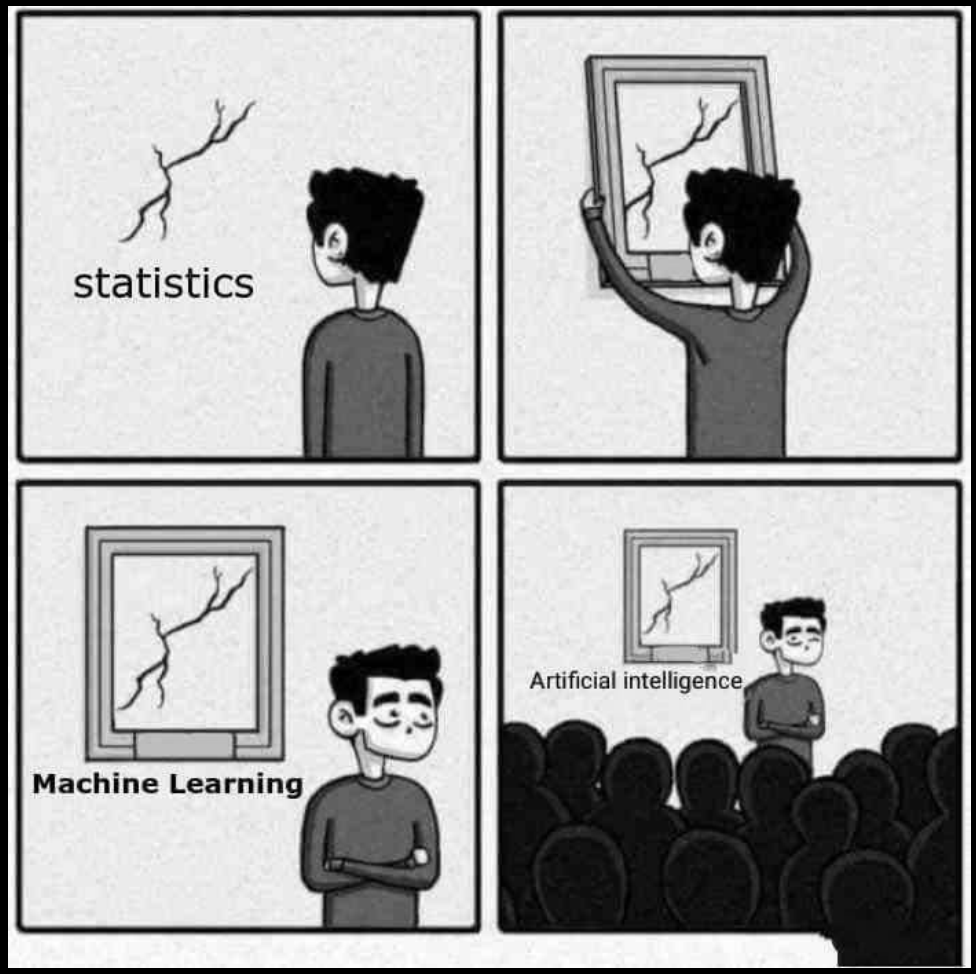
Des approches anciennes



En résumé

Des approches anciennes **mais**

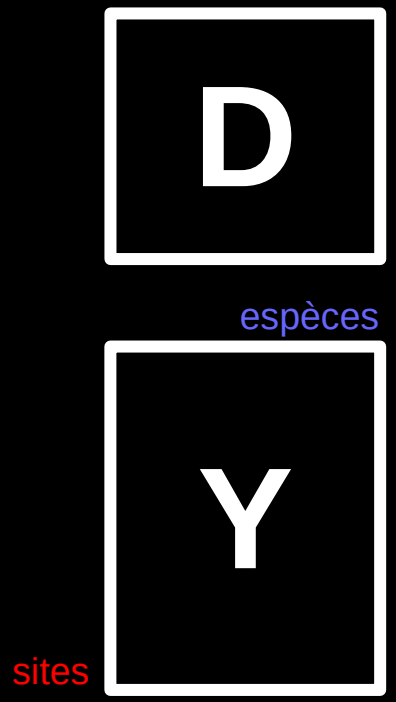
- lien fort avec les théories en écologie
- simplicité de mise en œuvre
- grande diversité d’outils disponibles
- encore des développements



De l'écologie à la génomique

Considering external information to improve the phylogenetic comparison of microbial communities: a new approach based on constrained Double Principal Coordinates Analysis (cDPCoA)

S. DRAY,*† S. PAVOINE‡§ and D. AGUIRRE DE CÁRCER¶



DPCoA

Application en métagénomique

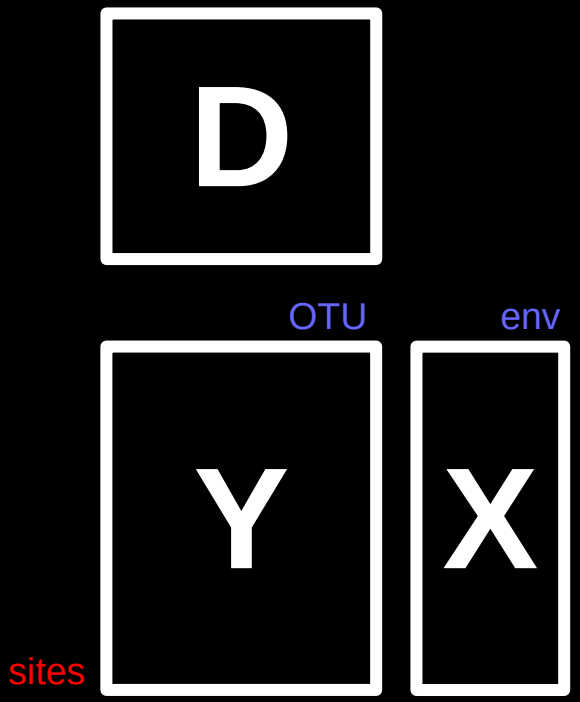
Y : échantillons de sol x OTUs

D : matrice de distances entre OTUs

De l'écologie à la génomique

Considering external information to improve the phylogenetic comparison of microbial communities: a new approach based on constrained Double Principal Coordinates Analysis (cDPCoA)

S. DRAY,*† S. PAVOINE‡§ and D. AGUIRRE DE CÁRCER¶



DPCoA

Application en métagénomique

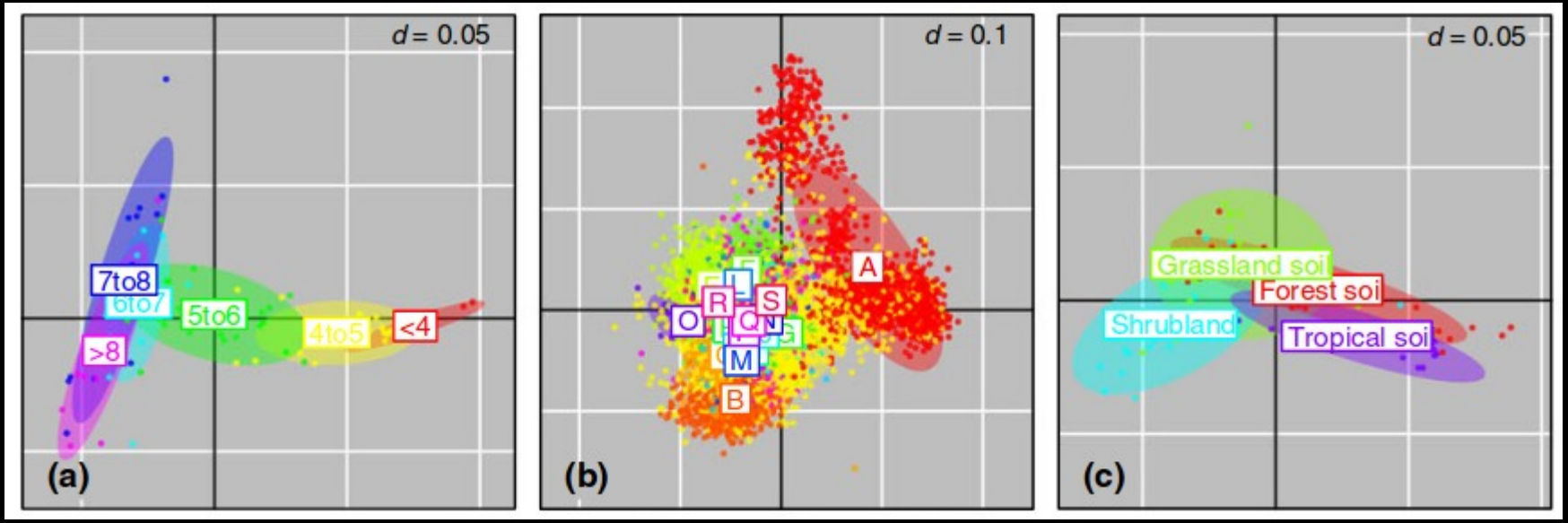
Y : échantillons de sol x OTUs

D : matrice de distances entre OTUs

X : pH et type de végétation

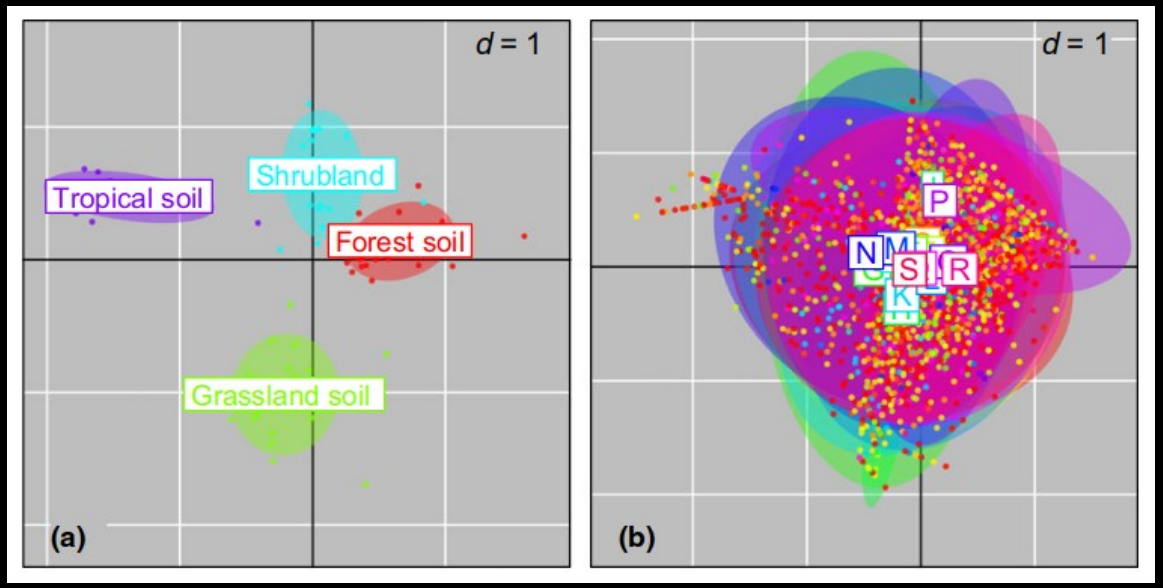
Diversité phylogénétique des communautés microbiennes du sol

DPCoA



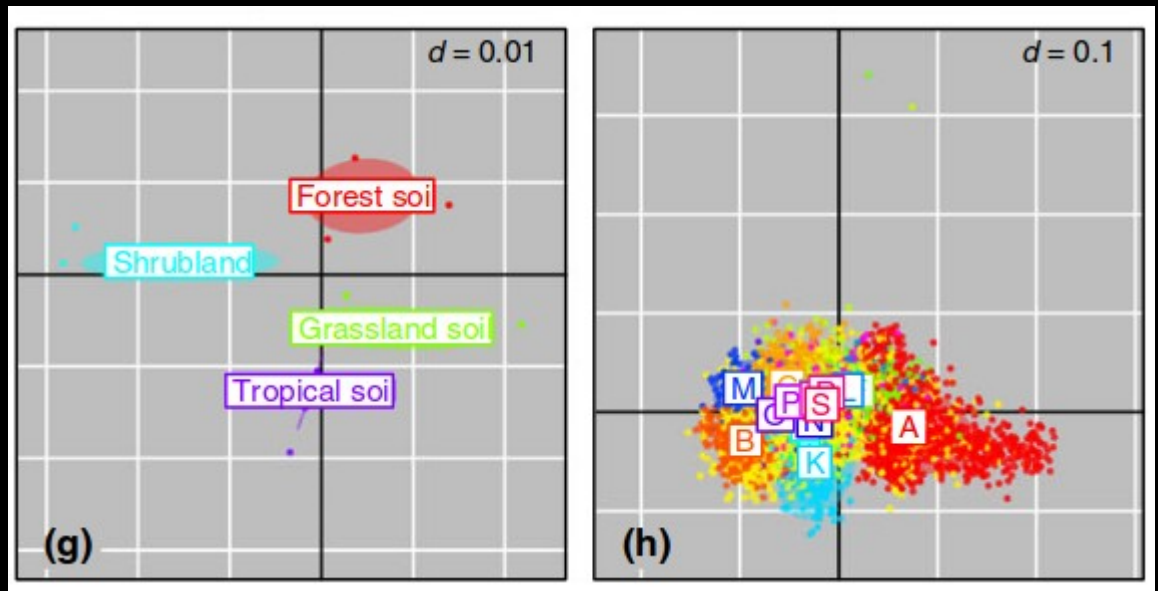
Les différences de composition des communautés entre types (21.5 %) sont essentiellement dues aux différences de pH

Diversité phylogénétique des communautés microbiennes du sol



AFC
intra pH
inter types

DPCoA
intra pH
inter types

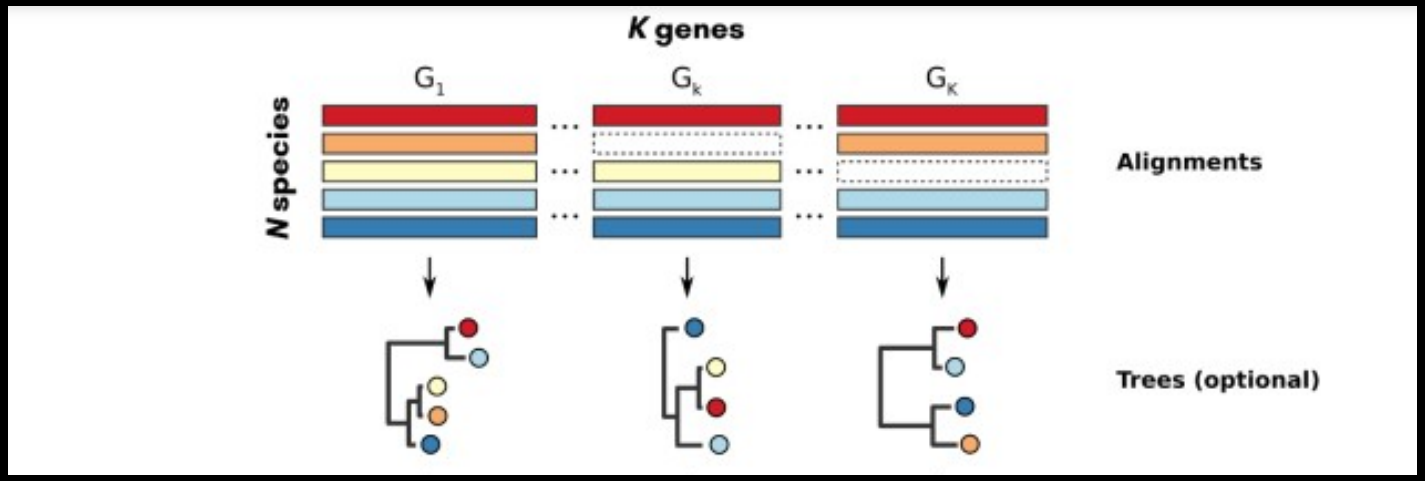


54.5 % non expliquée par le pH
5.45 % non expliquée par le pH mais expliquée par le type

Congruence d’arbres de gènes

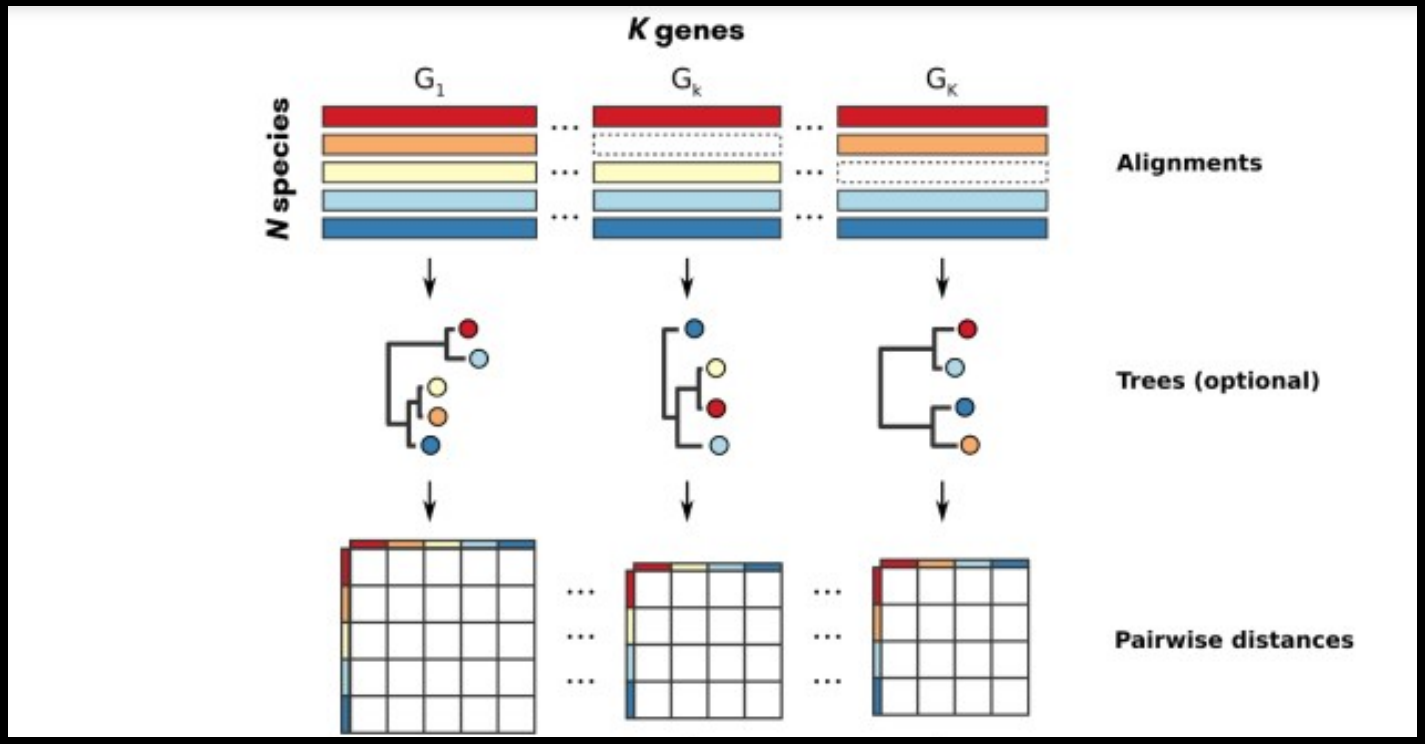
PhylteR: efficient identification of outlier sequences in phylogenomic datasets

Aurore Comte[1,2,†], Théo Tricou[3,†], Eric Tannier[3,4], Julien Joseph[3], Aurélie Siberchicot[3], Simon Penel[3], Rémi Allio[5], Frédéric Delsuc[6], Stéphane Dray[3], Damien M. de Vienne[3,*]

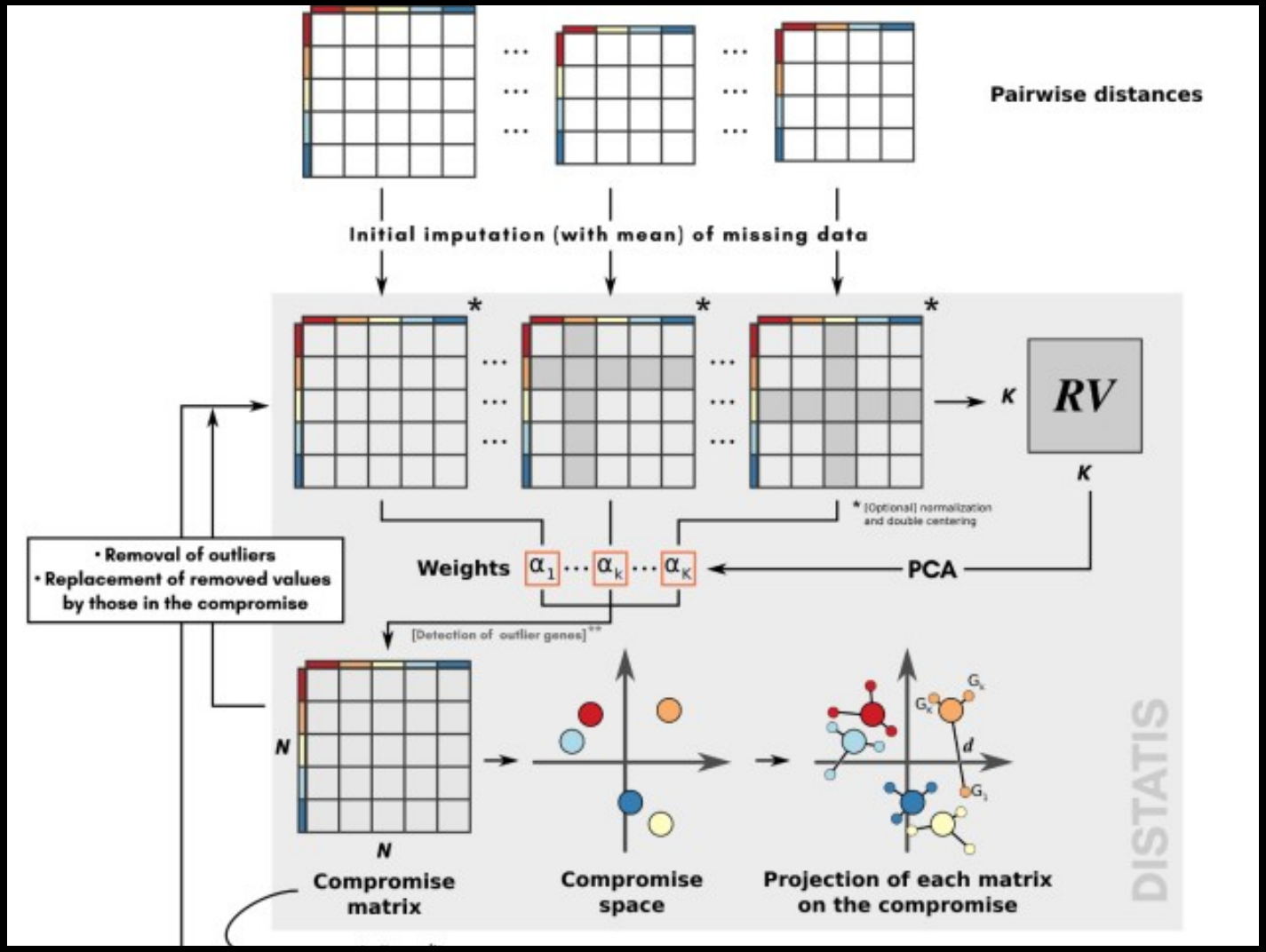


- Est-ce que tous les gènes racontent la même histoire évolutive ?
- Quels gènes et/ou espèces représentant une « anomalie » ?

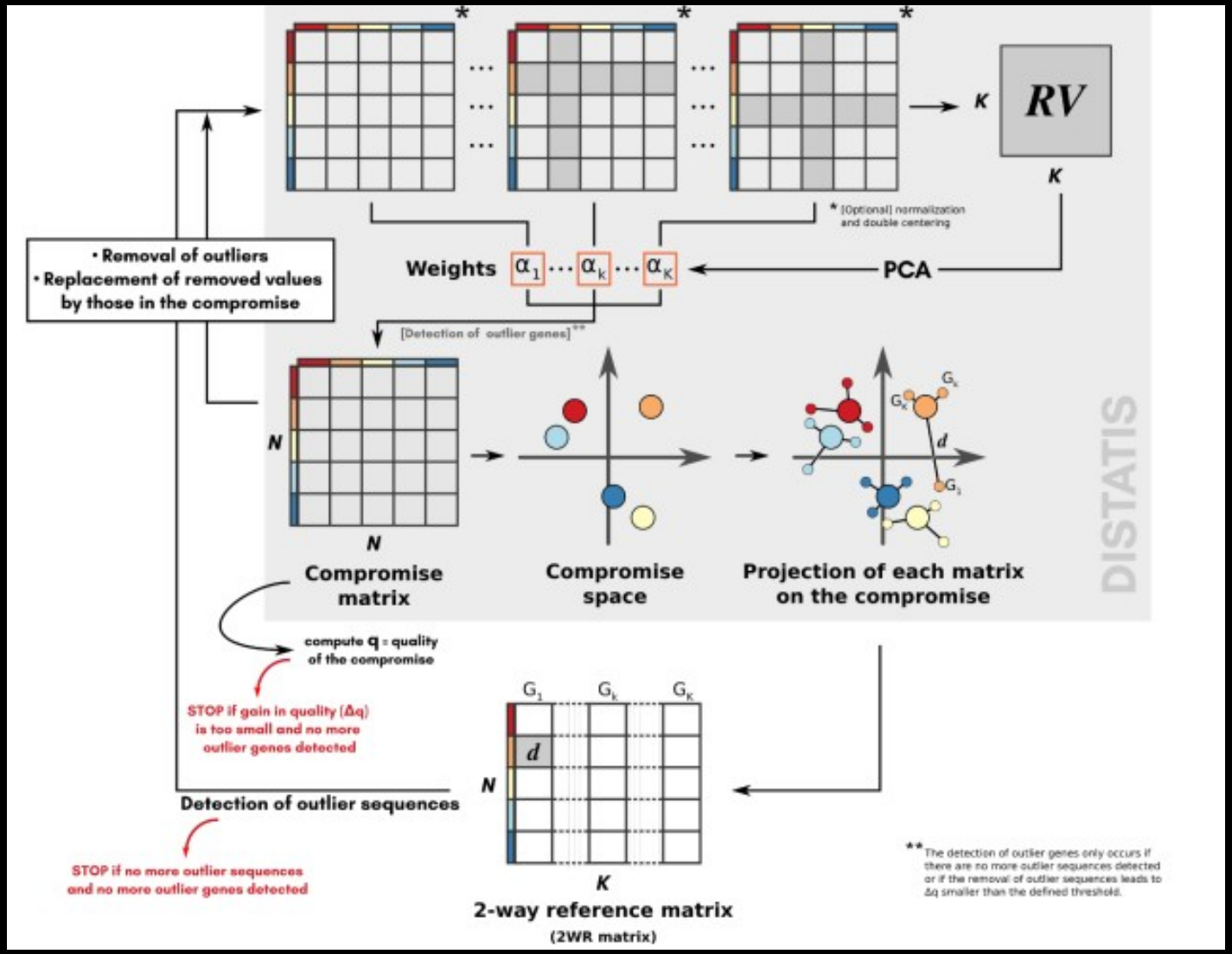
Congruence d’arbres de gènes



Congruence d’arbres de gènes



Congruence d’arbres de gènes



Analyses multivariées et génomique

The community ecology perspective of omics data

Stephanie D. Jurburg[1,2,3*], François Buscot[2,4], Antonis Chatzinotas[1,2,3], Narendrakumar M. Chaudhari[2,5], Adam T. Clark[6], Magda Garbowski[2,7], Matthias Grenié[2,3], Erik F. Y. Hom[2,8], Canan Karakoç[1,2,9], Susanne Marr[2,10,11], Steffen Neumann[2,11], Mika Tarkka[2,4], Nicole M. van ... Anna Heintz-Buschart[1,3]

A multivariate approach to the integration of multi-omics datasets

Chen Meng[1], Bernhard Kuster[1,2], Aedín C Culhane[3,4*] and Amin Moghaddas Gholami[1*]

phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data

Paul J. McMurdie, Susan Holmes*

Dimension reduction techniques for the integrative analysis of multi-omics data

Chen Meng*, Oana A. Zeleznik*, Gerhard G. Thallinger, Bernhard Kuster, Amin M. Gholami and Aedín C. Culhane

Statistical analysis of metagenomics data

M. Luz Calle*

Integration of omics data to unravel root microbiome recruitment☆

Anouk Zancarini[1,2], Johan A Westerhuis[2], Age K Smilde[2] and Harro J Bouwmeester[1]

MiBiOmics: an interactive web application for multi-omics data exploration and integration

Johanna Zoppi[1], Jean-François Guillaume[2], Michel Neunlist[1] and Samuel Chaffron[3,4*]

The community ecology perspective of omics data

Stephanie D. Jurburg[1,2,3*], François Buscot[2,4], Antonis Chatzinotas[1,2,3], Narendrakumar M. Chaudhari[2,5], Adam T. Clark[6], Magda Garbowski[2,7], Matthias Grenié[2,3], Erik F. Y. Hom[2,8], Canan Karakoç[1,2,9], Susanne Marr[2,10,11], Steffen Neumann[2,11], Mika Tarkka[2,4], Nicole M. van [illegible], Anna Heintz-Buschart[1,3]

A multivariate approach to the integration of multi-omics datasets

Chen Meng[1], Bernhard Kuster[1,2], Aedín C Culhane[3,4*] and Amin Moghaddas Gholami[1*]

phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data

Paul J. McMurdie, Susan Holmes*

Dimension reduction techniques for the integrative analysis of multi-omics data

Chen Meng, Oana A. Zeleznik, Gerhard G. Thallinger, Bernhard Kuster, Amin M. Gholami and Aedín C. Culhane

Statistical analysis of metagenomics data

M. Luz Calle*

Integration of omics data to unravel root microbiome recruitment☆

Anouk Zancarini[1,2], Johan A Westerhuis[2], Age K Smilde[2] and Harro J Bouwmeester[1]

MiBiOmics: an interactive web application for multi-omics data exploration and integration

Johanna Zoppi[1], Jean-François Guillaume[2], Michel Neunlist[1] and Samuel Chaffron[3,4*]

Merci

Et bonne retraite Denis