Training in ade4 in R - Module II: Advanced methods

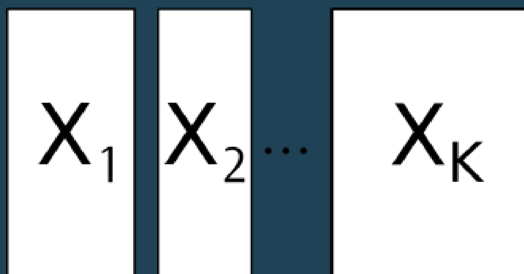
K-table methods

Stéphane Dray

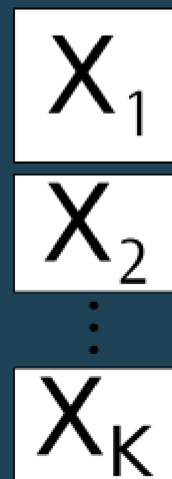
2025-10-19

Introduction

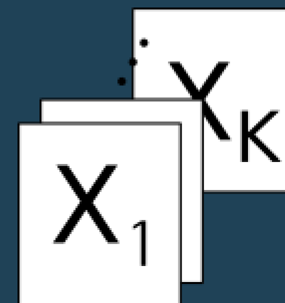
We consider situations involving the analysis of multiple tables. This data structure encompasses the following cases:



K tables with different variables measured on the same *n* individuals



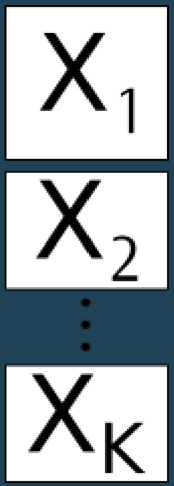
K tables with the same *p* variables measured on different individuals



K tables with the same *p* variables measured on the same *n* individuals

Example: `jv73`

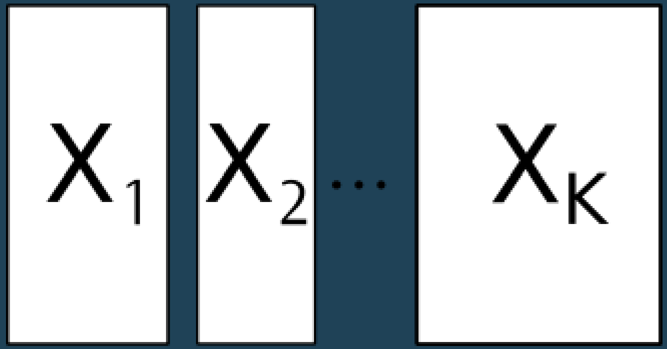
- Measurements of 12 physico-chemical variables for 92 sites
- The sites belong to 12 rivers



Is the typology of variables (physico-chemical gradients) similar in all rivers?

Example: `friday87`

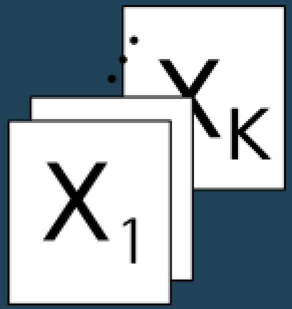
- Abundance of 91 macro-invertebrate species sampled 16 ponds
- The species are grouped in 10 taxonomic groups (Hemiptera, Trichoptera, ...)



Is the typology of sites (similarities in community composition) identical for all taxonomic groups?

Example: `meaudret`

- Measurements of 10 physico-chemical variables for 5 sites
- Each table corresponds to a date (season)

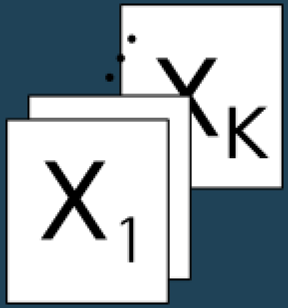


Is the typology of sites (similarities in physico-chemical aspects) identical for all dates?

Are the temporal patterns (similarities in physico-chemical aspects) identical for all sites?

Example: `bf88`

- Abundance of 70 bird species in 4 regions
- Each table corresponds to a vegetation stage along a gradient of 6 (open → closed habitat)



Is the typology of vegetation stages (community composition) identical for all regions?

Are the regional differences in community composition identical for all stages?

K-table methods

Simultaneous analysis of several tables preserving the multiblock design of the data. It allows to:

- Identify a common structure to all tables (blocks)
- Identify similarities/differences between tables

  $\longrightarrow$ based on the common dimension (either variables or individuals)
- Identify which variables are involved in the structures
- Identify which individuals are involved in the structures

General principles

- (Step 0: Define cross-product matrices to obtain matrices with same dimensions)
- Step 1: Define an "average table" (common table for MFA)
- Step 2: Analysis of this table to describe the common structure
- Step 3: Project each table and its rows and columns on the common structure

Different alternatives

- **Partial Triadic Analysis** is restrictred to data cubes where all tables have the same individuals and variables
- **Multiple Factor Analysis** is restricted to K-tables with at least the same individuals
- **Multiple Co-Inertia Analysis** is restricted to K-tables with at least the same individuals
- **STATIS** allows to deal with K-tables with at least the same individuals (STATIS on WD) or at least the same variables (STATIS on VQ)

STATIS

This method is very flexible and allows to deal with the three different types of K-tables. It consists in 3 main steps:

- **The interstructure** is the analysis of the relations between the individual data sets
- **The compromise** consists in deriving an optimal set of weights from the interstructure to compute the best common representation of the data sets. This consists in performing the PCA of a consensus table
- **The intrastructure** consists in studying the variation of the different data sets relative to the compromise structure

RV coefficient

It measures the link between two tables. It varies between 0 and 1. If the tables have the same rows, we have:

$$RV(\mathbf{X}_k, \mathbf{X}_l) = \frac{COVV(\mathbf{X}_k, \mathbf{X}_l)}{\sqrt{VAV(\mathbf{X}_k)}\sqrt{VAV(\mathbf{X}_l)}}$$

where the vectorial covariance is

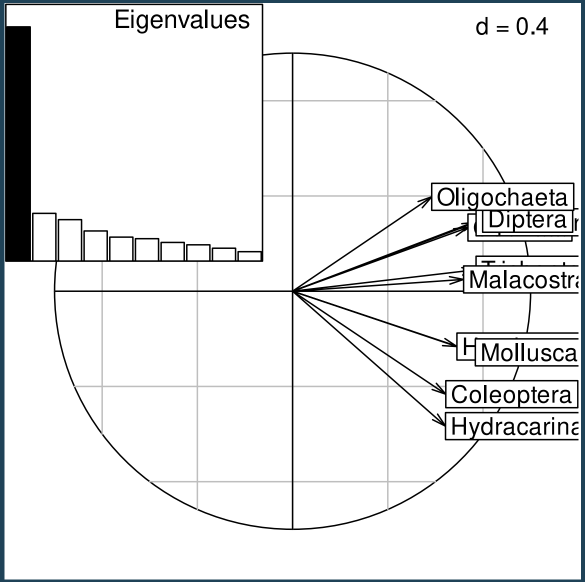
$$COVV(\mathbf{X}_k, \mathbf{X}_l) = Trace(\mathbf{X}_k \mathbf{Q}_k {\mathbf{X}_k}^\top \mathbf{D} \mathbf{X}_l \mathbf{Q}_l {\mathbf{X}_l}^\top \mathbf{D}) = Trace(\mathbf{W}_k \mathbf{D} \mathbf{W}_l \mathbf{D})$$

and the vectorial variance is

$$VAV(\mathbf{X}_k) = Trace(\mathbf{X}_k \mathbf{Q}_k {\mathbf{X}_k}^\top \mathbf{D} \mathbf{X}_k \mathbf{Q}_k {\mathbf{X}_k}^\top \mathbf{D}) = Trace(\mathbf{W}_k \mathbf{D} \mathbf{W}_k \mathbf{D})$$

The interstructure

It consists in the diagonalization of the $K \times K$ matrix of RV coefficients

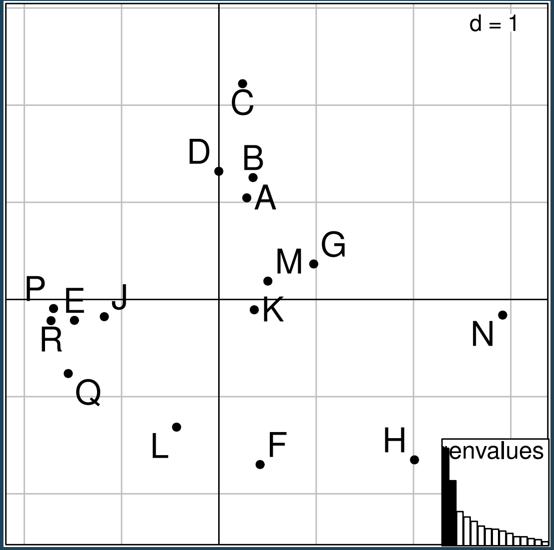


The compromise

Let $\alpha^\top = (\alpha_1 \ \ldots \ \alpha_k \ \ldots \ \alpha_K)$ be the first eigenvector of the Interstructure. The Compromise is defined as

$$\sum_{k=1}^{K} \alpha_k \frac{\mathbf{W}_k \mathbf{D}}{\sqrt{VAV(\mathbf{X}_k)}}$$

Let $\mathbf{\Lambda}$ and $\mathbf{U}$ be the eigenvalues and the eigenvectors of $\mathbf{WD}$ ($\mathbf{U}^\top \mathbf{DU} = \mathbf{I}$). The coordinates of individuals of the compromise are given by $\mathbf{L} = \mathbf{WDU\Lambda}^{\frac{1}{2}}$.



The intrastructure

- Variables of each table $\mathbf{X}_k$ are represented by the scores $\mathbf{C}_k = \mathbf{X}_k{}^\top \mathbf{D}\mathbf{U}$.
- Individuals of each table are represented by $\mathbf{L}_k = \mathbf{W}_k \mathbf{D}\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}$.