



**HAL**  
open science

# Variations intra-individuelles de données omiques L'Analyse Triadique Partielle Journée "Statistique et Génomique"

Denis Laloë, Tatiana Zerjal

## ► To cite this version:

Denis Laloë, Tatiana Zerjal. Variations intra-individuelles de données omiques L'Analyse Triadique Partielle Journée "Statistique et Génomique". École thématique. Journées Statistiques et Génomique, Paris, France. 2018. ⟨hal-04447190⟩

**HAL Id: hal-04447190**

**<https://hal.inrae.fr/hal-04447190v1>**

Submitted on 8 Feb 2024

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.



HAL Authorization

# Variations intra-individuelles de données omiques

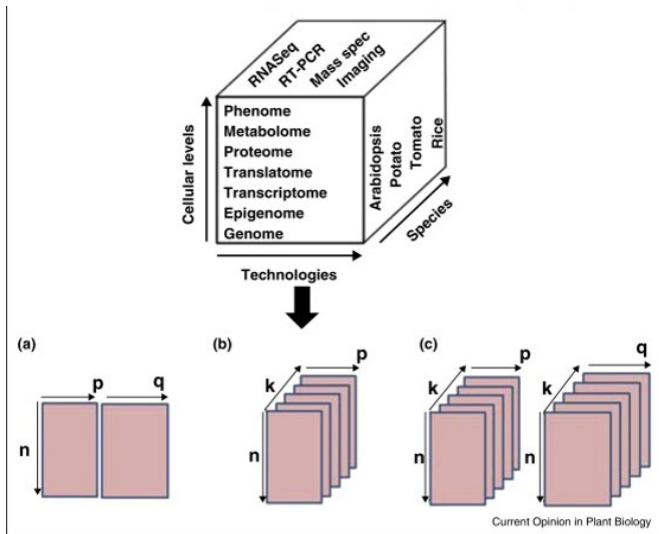
## L'Analyse Triadique Partielle

### Journée "Statistique et Génomique"

Denis Laloë & Tatiana Zerjal  
GABI - PSGen

12 octobre 2018

# Intégromique



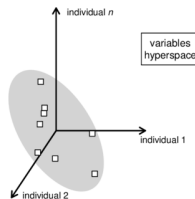
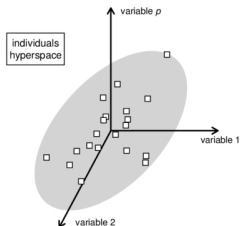
# Analyse géométrique des données

## Analyses multivariées

## Analyses factorielles

- Approche descriptive / exploratoire
- Approche géométrique : représentation de données sous forme de nuages de points
- Graphiques

# Le schéma de dualité Observations / Variables

cloud of  $n$  rows (individuals)cloud of  $p$  columns (variables)

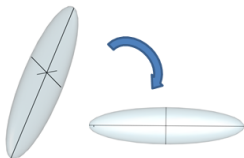
# Le schéma de dualité

## Observations / Variables

### The same transformation mechanism

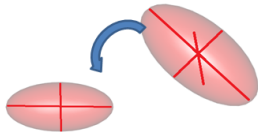
- Rotation
- Dimension reduction

Hyperspace of observations  
 $n$  points in  $\mathbb{R}^p$



A measure of distance between  
observations

Hyperspace of variables  
 $p$  points in  $\mathbb{R}^n$



A measure of distance between  
variables

**A mathematical formalisation**  
**The duality diagram theory**

# Le schéma de dualité

## Observations / Variables

Maximisation of the correlation between variables and components

$$V = X'X/n$$

$$VA = \Lambda$$

$$A'A = I$$

Principal axes

$$\text{Variable scores}$$

$$C = X'B$$

Best approximation (rank  $l$ )  
Eckart and Young

$$\begin{array}{ccc} & Q = I_p & \\ \langle p \rangle & \xrightarrow{\quad} & \langle p \rangle \\ X^t \uparrow & & \downarrow X \\ \langle n \rangle & \xleftarrow{\quad} & \langle n \rangle \\ & M = \frac{I_n}{n} & \end{array}$$

Diagonalisation

$$X'X \quad \quad \quad XX'$$

mêmes valeurs propres non nulles

$$\lambda_1 > \lambda_2 > \dots > \lambda_r > 0$$

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$$

Transition formulae

$$X\Lambda^{-0,5} = B$$

$$X'B\Lambda^{-0,5} = A$$

Singular value decomposition

$$X = B\Lambda^{0,5}A'$$

$$\hat{X}_l = \sum_{i=1, l} \sqrt{\lambda_i} b_i a_i'$$

Maximisation of the dispersion of individuals

Observations

$$W = XX' / n$$

$$WB = \Lambda$$

$$B'B = I$$

Principal components

Observation scores

$$L = XA$$

## Déclinaison du schéma de dualité

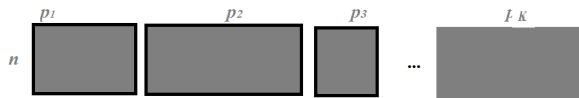
Analyse	Nature des données	Tableau de données	Pondération lignes	Pondérations colonnes
ACP	Quantitatif	$\frac{x_{ij} - x_{.j}}{s_j}$	$\frac{1}{n}$	I
AFC NS	Tableau de contingence	$\frac{f_{ij} - f_{i.} f_{.j}}{f_{i.} f_{.j}}$	$diag(f_{1.}, \dots, f_{i.})$	I
AFC	Tableau de contingence	$\frac{f_{ij} - f_{i.} f_{.j}}{f_{i.} f_{.j}}$	$f_{i.}$	$f_{.j}$
ACM	Qualitatif (tableau disjonctif)	$\frac{\delta_{ikj} - f_{kj}}{f_{kj}}$	$\frac{1}{n}$	$diag(f_i)/n$ facteurs

# Analyses multitables

Ensemble d'individus décrits par plusieurs groupes de variables

- Relations entre tableaux (interstructure)
- Recherche d'un compromis
- Intrastructure, trajectoires (comment une analyse partielle reproduit l'analyse compromis)

# Analyses multitable

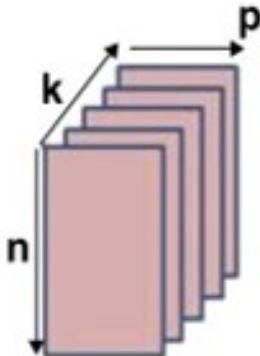


## Plusieurs tableaux

- Corrélations canoniques généralisées
- Co-inertie multiple
- STATIS
- Analyse Factorielle Multiple

# Analyse Triadique Partielle

*Thioulouse et Chessel, 1987*



Analyse d'un cube de données, vu comme une séquence de tables à deux dimensions

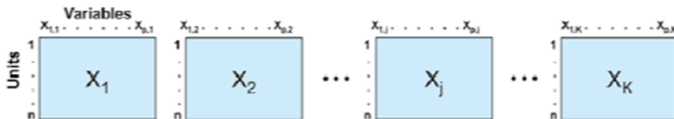
# Analyse Triadique Partielle

Bertrand & Maumy, 2010

<http://www.bentley.edu/csbigs/Vol4-1/bertrand.pdf>

Using Partial Triadic Analysis for Depicting the Temporal Evolution of Spatial Structures: Assessing Phytoplankton Structure and Succession in a Water Reservoir

## ■ K tables



## ■ Common typology: the « compromise »



Interstructure: similarities between the different tables

Compromise: Common structure derived from each table

2

# Analyse Triadique Partielle

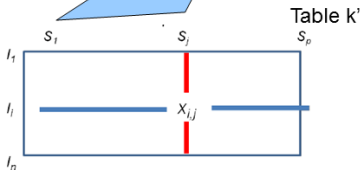
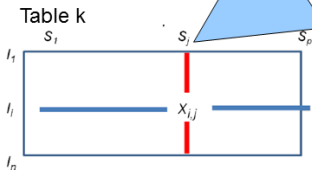
## Interstructure

Relation entre deux tableaux (mêmes lignes, mêmes colonnes) :  $R_v$  coefficient

$$R_v(X_k, X_{k'}) = \frac{\text{Cov}_v(X_k, X_{k'})}{\sqrt{\text{Var}_v(X_k)}\sqrt{\text{Var}_v(X_{k'})}}$$

$$= \frac{\sum_{i=1}^p \text{cor}(X_j^i, X_k^i)}{p}$$

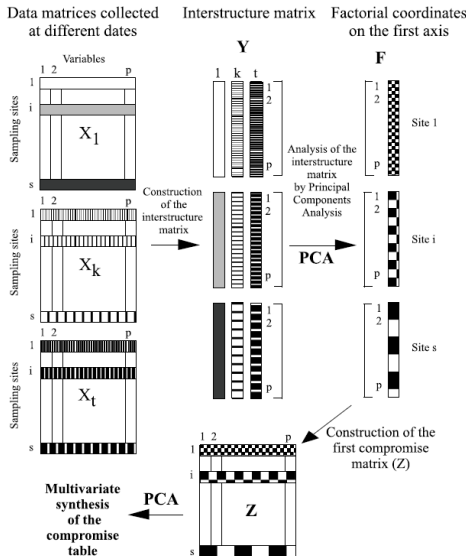
- Calcul des  $p$  corrélations gène à gène
- $R_v$  : moyenne des  $p$  corrélations



# Analyse Triadique Partielle

**Fig. 1.** General scheme of the Partial Triadic Analysis: construction of the interstructure matrix  $Y$  and extraction of the compromise table  $Z$ . The interstructure between matrices  $X_1 \dots X_t$  corresponding to the sampling dates is analysed by a simple PCA of  $Y$ . The compromise tables are derived from the coordinates of the variables at the different sites on the principal components of the PCA of  $Y$ . The figure illustrates the construction of the first compromise table from the first principal component of the simple PCA of  $Y$ . Compromise tables are then analysed by a simple PCA

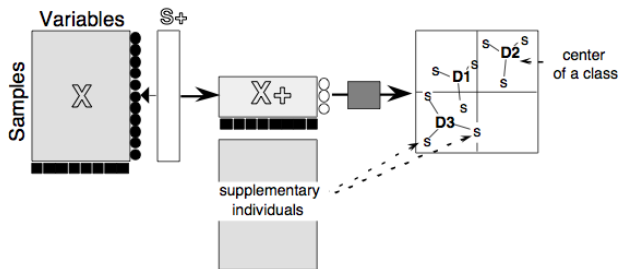
Rossi, (*Pedobiologia*, 2003)



# Projet Chickstress

- 2 lignées de poules R+, R-
- Adaptation à un régime suboptimal
- 4 groupes Lignées \* Traitement
- Pour chaque individu, transcriptome sur 4 tissus:
  - Foie,
  - Sang,
  - Tissu adipeux,
  - Hypothalamus

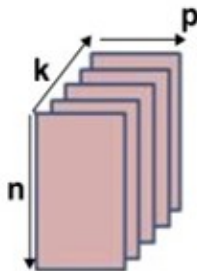
## Analyses entre groupes



Dufour, 2010

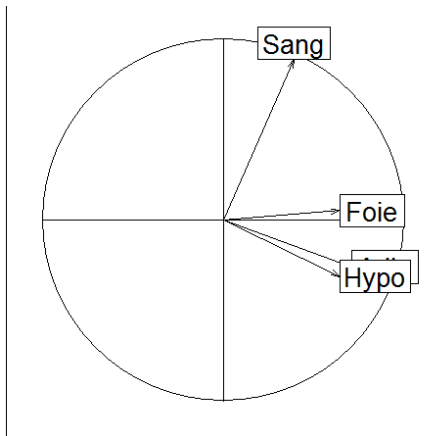
- Alternative robuste à l'analyse discriminante;
- Modélisation (PCAIV / Analyse de redondance);
- Tests de permutation

## Analyse Triadique Partielle / ACP entre groupes



- $n = 4$  (112 individus, dans 4 groupes Lignées \* Traitement),
- $k = 4$  (4 tissus)
- $p = 4286$  (Gènes exprimés dans les 4 tissus)

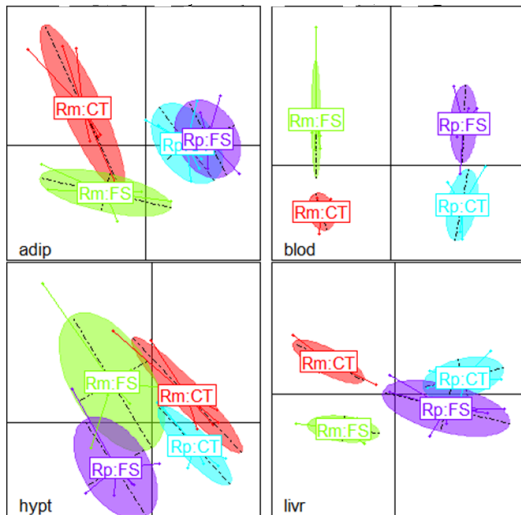
# Interstructure



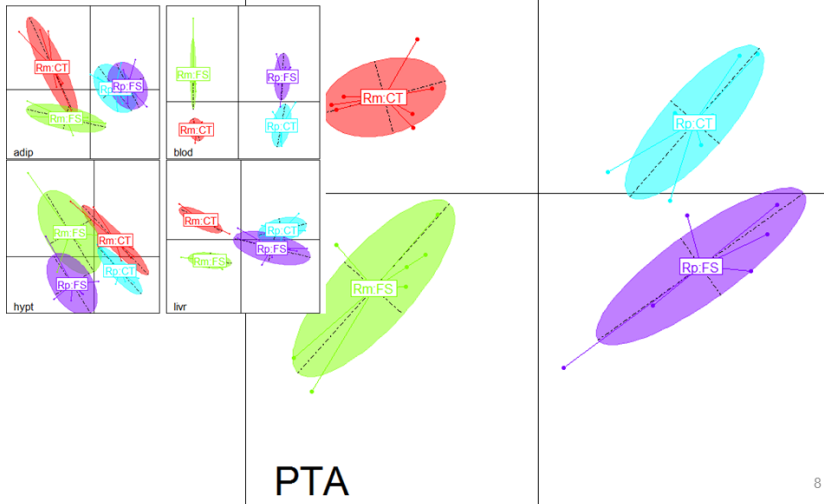
	Adip	Sang	Hypo	Foie
Adip	1.00	0.09	0.25	0.23
Sang	0.09	1.00	0.08	0.12
Hypo	0.25	0.08	1.00	0.16
Foie	0.23	0.12	0.16	1.00

# ACP entre groupes

## Analyses séparées par tissu



# Compromis



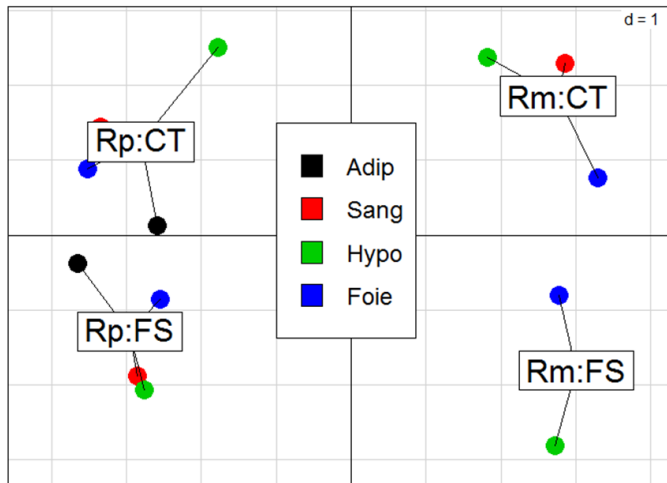
# Compromis et Intrastructure

$k$  tables  $\longrightarrow k + 1$  points

- $k$  points par analyse séparée
- 1 point compromis, au centre des  $k$  points

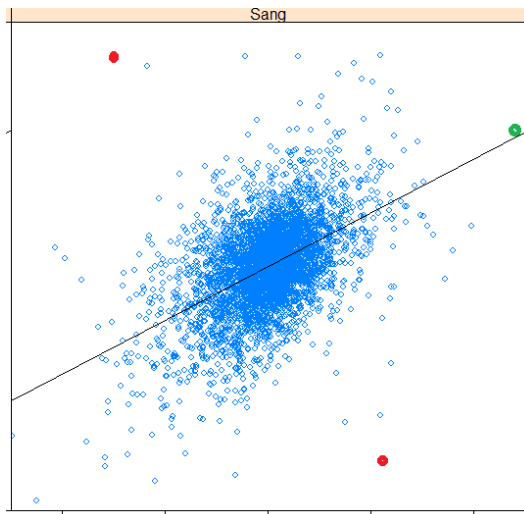
# Compromis et Intrastructure

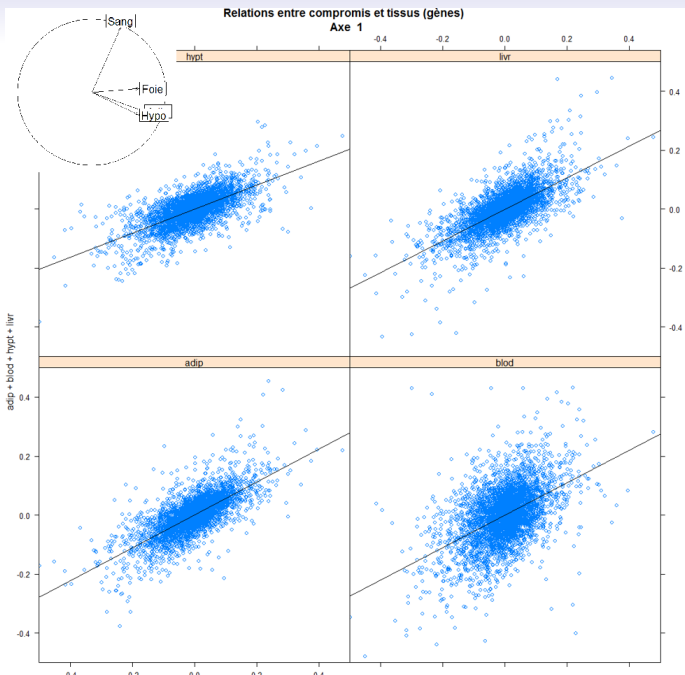
## Les groupes d'animaux



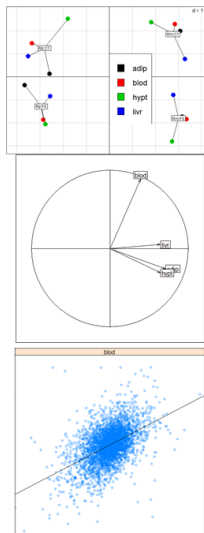
# Compromis et Infrastructure

## Les gènes





## Conclusion

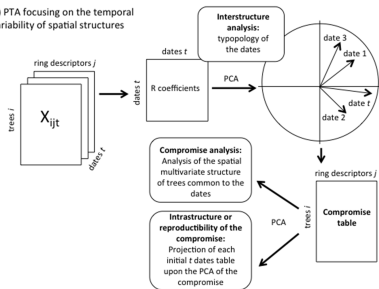


- Identification d'une structure compromis entre individus ou groupes
- Relations entre les transcriptomes des tissus produisant cette structure compromis
- Relations entre l'expression des gènes selon les tissus

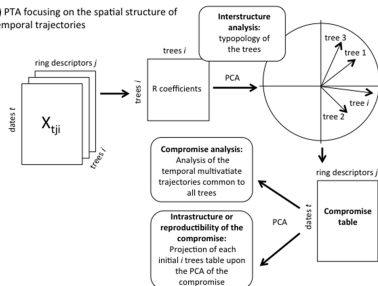
# Conclusion

## Pourquoi "Partiel"

A) PTA focusing on the temporal variability of spatial structures



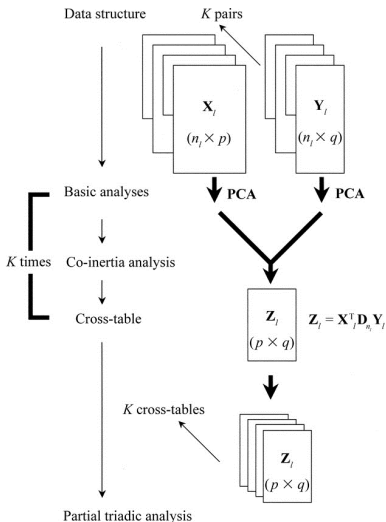
B) PTA focusing on the spatial structure of temporal trajectories



Rossi et al, 2014

# Conclusion - Extensions

## STATICO



# Références

- Bertrand, F., & Maumy, M. (2010). Using partial triadic analysis for depicting the temporal evolution of spatial structures: assessing phytoplankton structure and succession in a water reservoir. *Case Studies Business, Industry Govern Stat*, 4, 23-43.
- Dufour, A.B., 2010. Les analyses en composantes principales inter et intra-classes. <https://pbil.univ-lyon1.fr/R/pdf/tdr621.pdf>
- Rajasundaram, D., & Selbig, J. (2016). More effort?more results: recent advances in integrative ?omics? data analysis. *Current opinion in plant biology*, 30, 57-61.
- Rossi, J. P., Nardin, M., Godefroid, M., Ruiz-Diaz, M., Sergent, A. S., Martinez-Meier, A., ... & Rozenberg, P. (2014). Dissecting the space-time structure of tree-ring datasets using the partial triadic analysis. *PLoS one*, 9(9), e108332
- Thioulouse, J., & Chessel, D. (1987). Multi-table analysis of ecological factors. I. From description of ecosystem status to that of ecosystem function using triadic analysis. *ACTA OECOL., OECOL. GEN.*, 8(4), 463-480.
- Thioulouse, J., Simier, M., & Chessel, D. (2004). Simultaneous analysis of a sequence of paired ecological tables. *Ecology*, 85(1), 272-283.
- ade4. <http://pbil.univ-lyon1.fr/ade4/>. Dray, S., & Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software*, 22(4), 1-20.
- ExPosition. cf Beaton, D., Fatt, C. R. C., et Abdi, H. (2014). An ExPosition of multivariate analysis with the singular value decomposition in R. *Computational Statistics & Data Analysis*, 72, 176-189.