



# Analysis of Longitudinal Metabolomics data using MSCA

Jeroen J. Jansen<sup>1</sup>, Huub C.J. Hoefsloot<sup>1</sup>, Jan van der Greef<sup>2,3</sup>, Marieke E. Timmerman<sup>4</sup>, Age K. Smilde<sup>1,2</sup>

- 1) Biosystems Data Analysis, Faculty of Science, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands
- 2) TNO Nutrition and Food Research, PO Box 360, 3700 AJ Zeist, The Netherlands
- 3) Beyond Genomics, 40 Bear Hill Road, Waltham, MA 02451, United States
- 4) Heymans Institute of Psychology, DPMG, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

Website: <http://www.science.uva.nl/its/> E-Mail: [jjansen@science.uva.nl](mailto:jjansen@science.uva.nl)

## Introduction

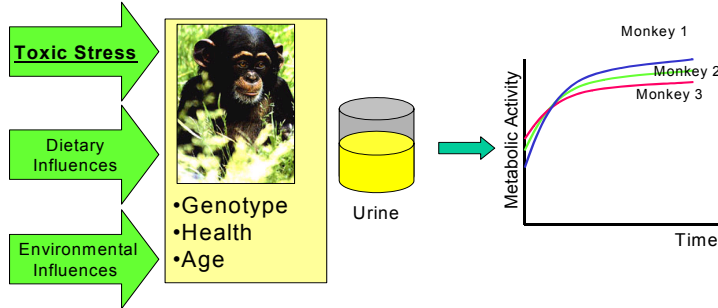
### Metabolomics:

Investigating the metabolism of an organism, as influenced by **Internal** and **External** Factors.

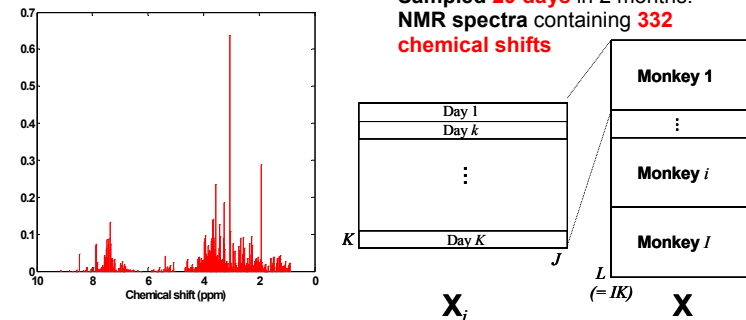
**How:** Analysis of chemical composition of Body Fluids (**urine**)

**Using:** <sup>1</sup>H-NMR spectroscopy and multivariate data analysis methods

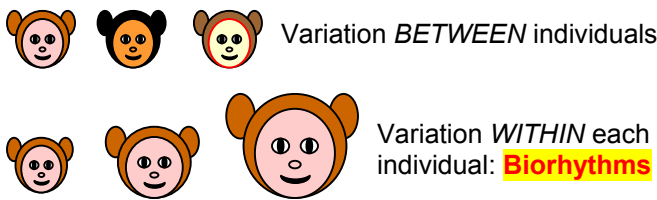
**This Research:** Analysing **normal** metabolism of Rhesus Monkeys



## Dataset



## Variation in the data



Different sources of variation: This is **Multilevel data**

## Methods

### Principal Component Analysis

Minimize:

$$f_{PCA} = \left\| X - 1_L m^T - TP^T \right\|^2$$

- Does not differentiate between different sources of variation in the data

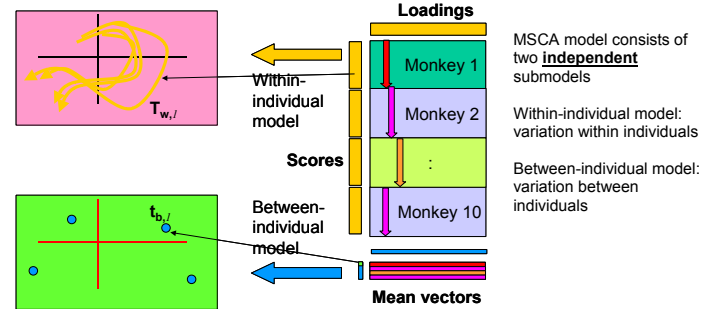
### Multilevel Simultaneous Component Analysis:

Minimize:

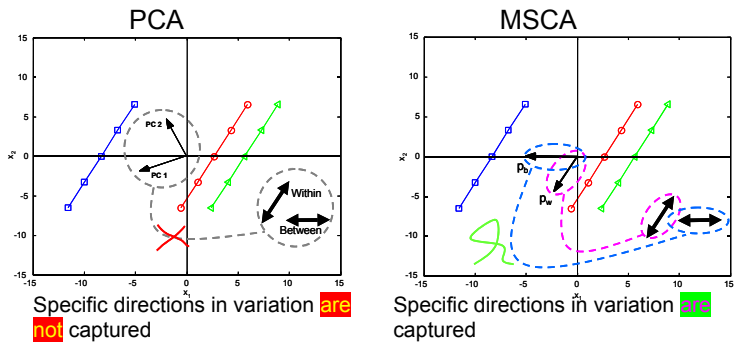
$$f_{MSCA} = \sum_{i=1}^I \left( \left\| X_i - 1_K m^T - 1_K t_{b,i} P_b^T - T_{w,i} P_w^T \right\|^2 \right)$$

- Differentiates between different sources of variation in the data and thereby improves interpretation
- Is a restricted version of PCA

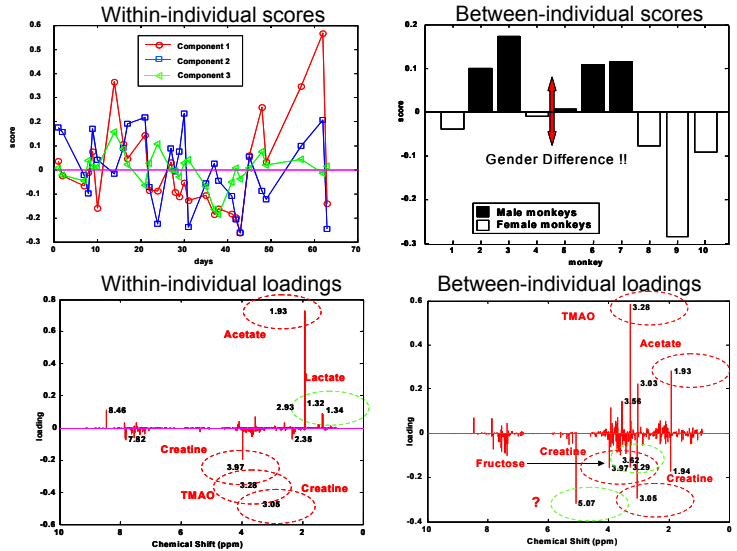
## MSCA Model



## Testset



## Results



Similarities and differences in loadings show similarities and differences in variation on both levels

## Conclusions

- PCA does not give a good interpretation of the variation in multilevel data
- MSCA can handle different sources of variation in the data by making different submodels for each source
- MSCA gives a better view on each of the different sources of variation than PCA: the gender difference in the between-individual scores is not clearly visible in PCA and the biorhythms in the data are characterized better by MSCA