

Analysis of Longitudinal Metabolomics data using MSCA

Jeroen J. Jansen¹, Huub C.J. Hoefsloot¹, Jan van der Greef^{2,3}, Marieke E. Timmerman⁴, Age K. Smilde^{1,2}

- 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

Introduction

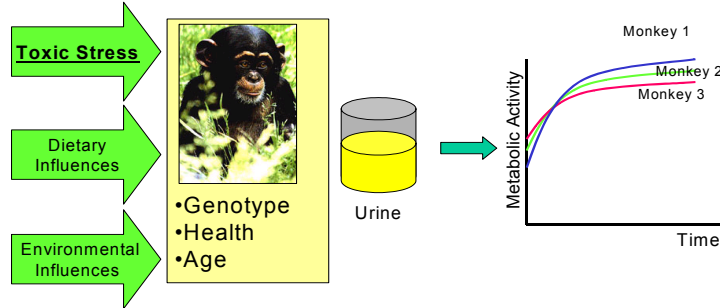
Metabolomics:

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

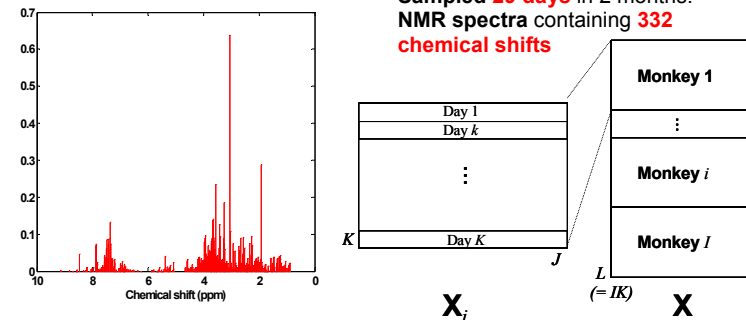
How: Analysis of chemical composition of Body Fluids (**urine**)

Using: ¹H-NMR spectroscopy and multivariate data analysis methods

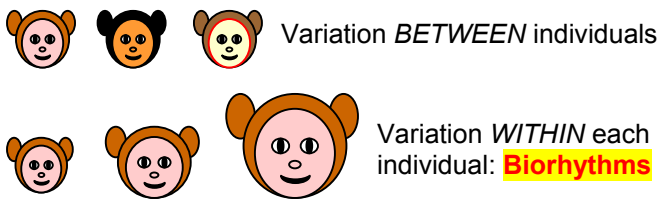
This Research: Analysing **normal** metabolism of Rhesus Monkeys



Dataset



Variation in the data



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

Methods

Principal Component Analysis

Minimize:

$$f_{PCA} = \|\mathbf{X} - \mathbf{1}_L \mathbf{m}^T - \mathbf{T} \mathbf{P}^T\|^2$$

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

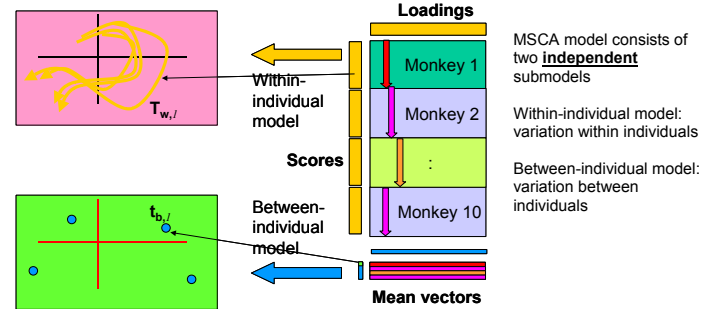
Multilevel Simultaneous Component Analysis:

Minimize:

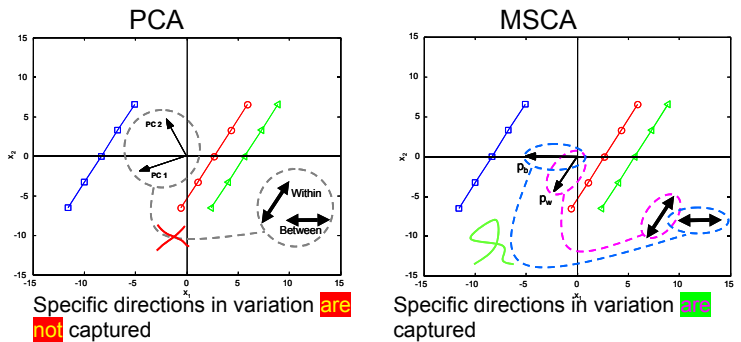
$$f_{MSCA} = \sum_{i=1}^I \left(\|\mathbf{X}_i - \mathbf{1}_K \mathbf{m}^T - \mathbf{1}_K \mathbf{t}_{b,i} \mathbf{P}_b^T - \mathbf{T}_{w,i} \mathbf{P}_w^T\|^2 \right)$$

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

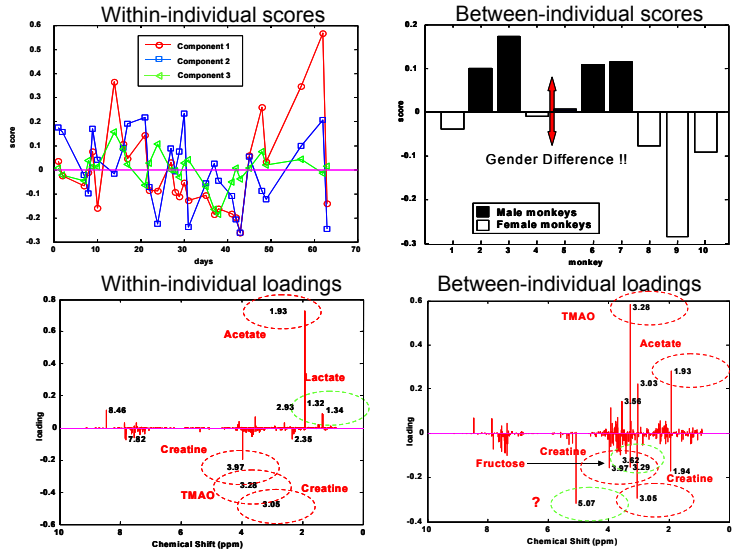
MSCA Model



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