Analysis of the dynamic variation in multilevel metabolomics data

Jeroen J. Jansen[1], Huub C.J. Hoefsloot[1], Jan van der Greef[2,3], Age K. Smilde[1,2]

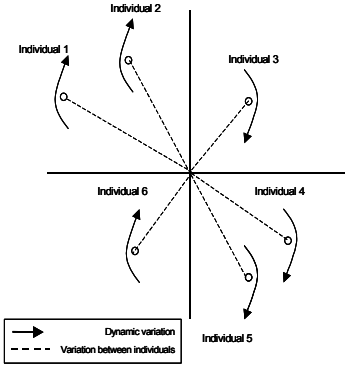
1) Biosystems Data Analysis, Faculty of Science, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands
2) TNO Nutrition and Food Reseach, PO Box 360, 3700 AJ Zeist, The Netherlands
3) Beyond Genomics, 40 Bear Hill Road, Waltham, MA 02451, United States

Website: http://www.science.uva.nl/its/ E-Mail: jjansen@science.uva.nl

Question:

When I perturb the metabolism of an animal, what is its response?
Or rephrased for data analysts:
How do I characterize the dynamic variation in metabolomics data?

A priori knowledge about the variation in time-resolved metabolomics data (represented in two dimensions)



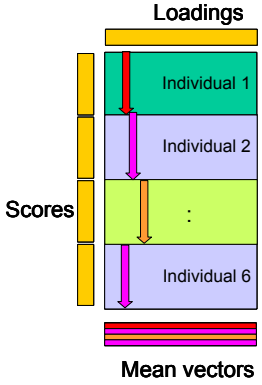
Question:
How do I focus the data analysis on the **dynamic** variation and remove the **static** variation between individuals?
Answer: Use a method that takes the **multilevel** structure of the data into account

- SMART (Keun *et al.*)
- MSCA (Jansen *et al.*)
- **Multilevel** Batch Processing [1]

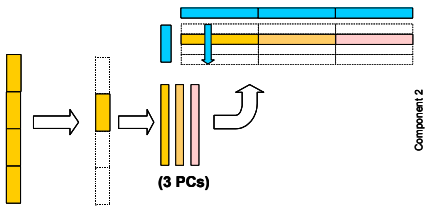
[1] *Following a suggestion of Tim Ebbels presented at SSC7 in Marienhamn, Finland*

Multilevel Batch Processing

Step 1: lower model
PCA with **block centering**

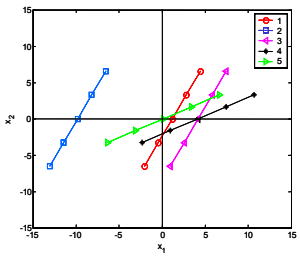


Step 2: upper model
Upper model from Batch Processing



- Block centering removes the static variation between individuals.
- lower model scores then focus on the dynamic variation
- upper model scores give **quantitative comparison** of the dynamic variation

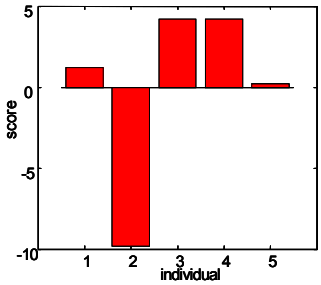
Simulated dataset



Containing:
- 5 individuals
- 2 different trajectories
- 5 measurement occasions

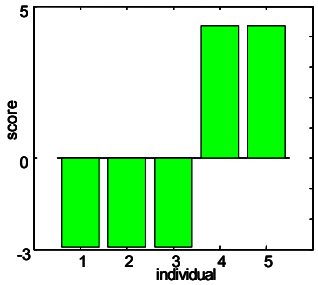
- Data contains static variation between the individuals.
- Has to be removed to analyse dynamic variation.

Column Centering



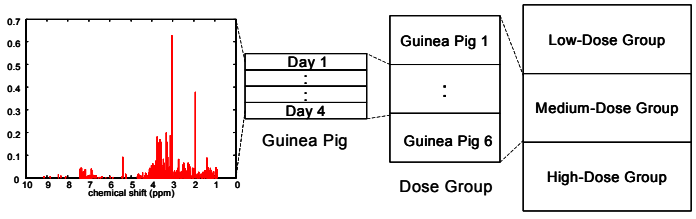
Specific differences in dynamic variation are not captured

Block Centering



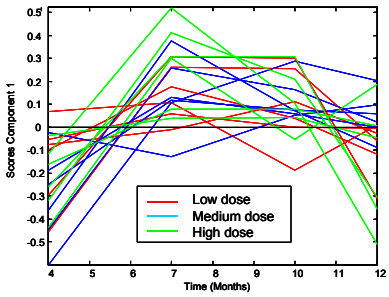
Specific differences in dynamic variation are captured

Time-Resolved Metabolomics Dataset



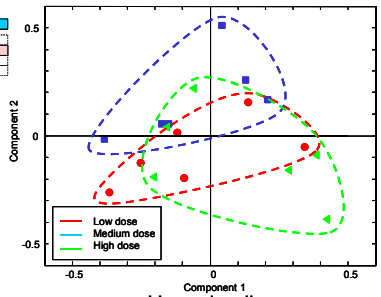
- Guinea Pigs that suffer from Osteoarthritis
- Measured at **4 time-points** during 12 months
- **3 doses** of Vitamin C
- **6 guinea pigs** *per dose group*
- NMR Spectra containing **253 chemical shifts**

Lower model scores



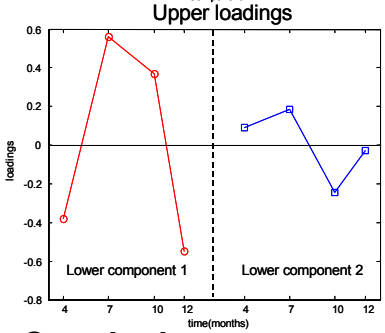
2 components are chosen for the lower model

Upper scores



- No clear differences between scores of different dose groups
- No deviating scores in upper model
- No differences in spread of scores of different dose groups

Upper loadings



- These time-profiles give directions of the largest variation in dynamic behaviour.
- Using these results the differences in dynamic behaviour can be characterized.

Conclusions

- Time-resolved metabolomics data contains multiple types of variation: it has a multilevel structure.
- When data is block centered, the model describes the **dynamic** variation in the data
- This can be used in Multilevel Batch processing to quantitatively compare the dynamic variation between the individuals
- Instead of PCA, also a PLS model can be used.

Jansen,J.J., Hoefsloot,H.C.J., van der Greef,J., Timmerman,M.E., and Smilde,A.K. Multilevel Component Analysis of time-resolved metabolomics data (submitted)
Keun,H.C., Ebbels,T.M., Bollard,M.E., Beckonert,O., Antti,H., Holmes,E., Lindon,J.C., and Nicholson,J.K. (2004) Geometric trajectory analysis of metabolic responses to toxcicity can define treatment specific profiles, Chem.Res.Toxicol., 17, 579-587.