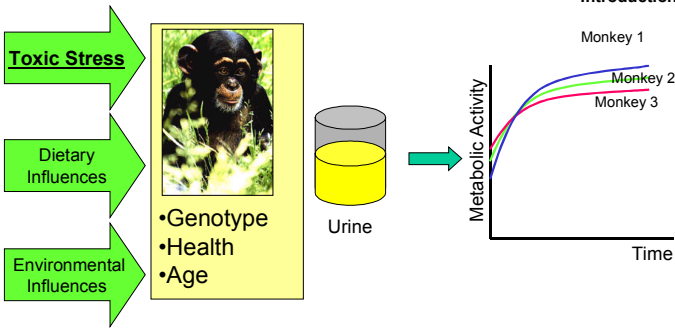




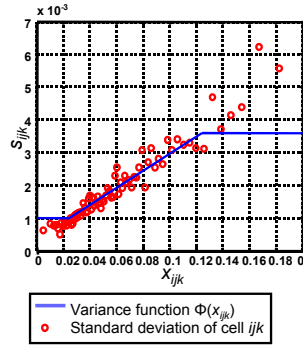
# Exploratory analysis of <sup>1</sup>H-NMR data of Monkey urine

Jeroen J. Jansen<sup>1</sup>, Hans F.M. Boelens<sup>1</sup>, Age K. Smilde<sup>1</sup>, Jan van der Greef<sup>2</sup>

<sup>1</sup> Process Analysis and Chemometrics, Department of Chemical Engineering, Faculty of Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam  
<sup>2</sup> TNO Nutrition and Food Research, PO Box 360, 3700 AJ Zeist  
 Beyond Genomics, 40 Bear Hill Road, Waltham, MA 02451



External influences together with genotypical and fitness-properties of an organism determine the metabolism of an organism and therefore the contents of its urine. The purpose of this research is finding (temporal) differences between male and female monkeys. This part of the research is focused on overall gender-based differences. The composition of a urine can be determined by e.g. NMR Spectroscopy



## B. Heteroscedastic Error

$x_{ijk}$  is sorted in increasing value.  $x_{ijk}$  and  $s_{ijk}$  are averaged for a predefined number of points. The linear relationship in the adjoining figure is obtained. This shows the error is heteroscedastic. It is not possible or very hard to compensate for heteroscedasticity with a common data transformation (log, square root or Box/Cox). The relationship between  $x_{ijk}$  and  $s_{ijk}$  can be described by a variance function (indicated with a blue line). PCA-W can be used to deal with the heteroscedastic error using the variance function.

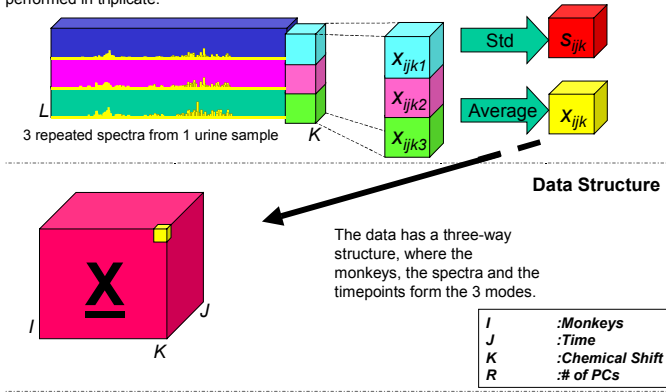
Minimization criterion:  
 PCA :  $\|X - TP^T\|^2$   
 PCA-W :  $\|W*(X - TP^T)\|^2$

Weights  $w_{ijk} : w_{ijk}(x_{ijk}) = \frac{1}{\Phi(x_{ijk})}$

PCA-W can also be used to incorporate missing data: the corresponding weights are set to 0

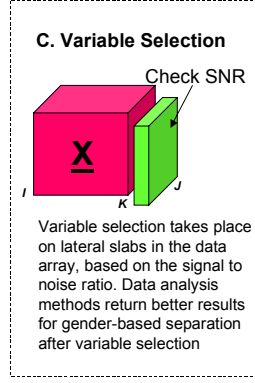
## Measurements

Urine of 10 monkeys is analyzed for 33 days during 3 months using <sup>1</sup>H-NMR. 5 monkeys are male and 5 female. All measurements are performed in triplicate.



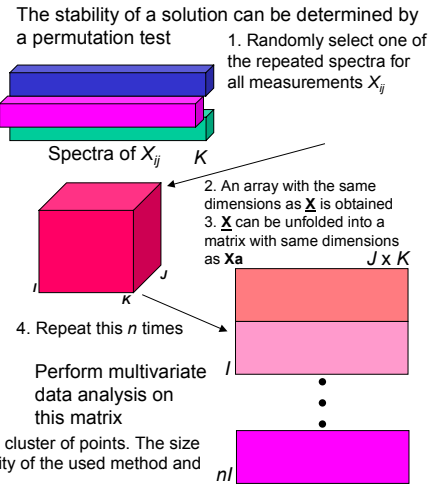
## Research

- A. Alternative methods of unfolding  $\underline{X}$
- B. A method is given that deals with the heteroscedastic error
- C. A method for variable selection based on the signal-to-noise ratio is proposed
- D. A permutation test is given to determine the stability of the solutions of different methods



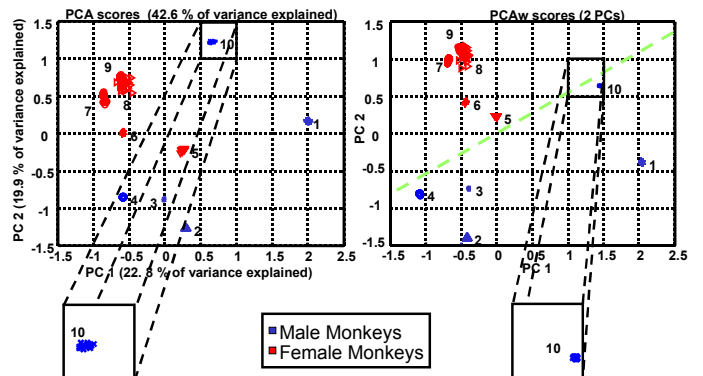
Variable selection takes place on lateral slabs in the data array, based on the signal to noise ratio. Data analysis methods return better results for gender-based separation after variable selection

## D. Permutation Test



The scores of this analysis will be a cluster of points. The size of this cluster will indicate the stability of the used method and of the data preprocessing.

## Results from the Permutation operation for PCA and PCA-W



- In most cases the scores of PCA-W return tighter clusters than PCA
- For both methods a clear clustering of the female monkeys is visible
- There is a better separation between male and female monkeys with PCA-W
- Theoretically PCA-W is preferred: the variance function works efficient.
- Monkey # 8 exhibits deviating behavior: this monkey was probably ill

## A. Unfolding

