

## Application of Ion-Mobility and Advanced Statistical Methods for Small Molecule Profiling

John Shockcor

Director of Metabolic Profiling Business Development Visiting Fellow, Dept of Biochemistry, University of Cambridge

# Five Simple Questions

I have 2 sample groups A and B

- Are the groups different? If so:
  - What has higher in A
  - What has lower in A
  - What is absent in A
  - What is unique to A

POSSTRI F."

# I have 2 sample groups A and B

Five Simple Questions

- Are the groups different? If so:
  - What has higher in A
  - What has lower in A
  - What is absent in A
  - What is unique to A

These are MARKERS of the difference between A and B



## How is Do We Find These Markers?



©2007 Waters Corporation



- It is possible to mine the data using multivariate statistics.
- We first analyze the groups using GC or LC/MS
- We then tabulate all the observed masses and their chromatographic retention times with advanced computational methods.
  - These mass/retention time pairs become the variables used for statistical analysis.



 Multivariate statistics then allows us to reduce thousands variables (mass/retention time pairs) down to a simple 2 or three dimensional maps.

These maps shows the difference, if any, between the groups and provides us with a list of the variables which contribute to the difference.

## How Spectral Information is displayed



 Spectrum (observation) becomes a point in PCA
Scores plot

/aters

POSSIBLE."

THE SCIENCE OF

 Variables (m/z\_RT) shown in PCA
Loadings Plot

 Using plots together allows trends in the sample spectra to be interpreted in terms of m/z

#### Why Choose to use Multivariate Statistics? Waters THE SCIENCE OF WHAT'S POSSIBLE"

## Why not :

- Hierarchical Clustering
- Heat make
- ANOVA
- T tests

With MarkerLynx it is possible to export your data to any statistical program you like.

# Woters

#### Short and wide data sets

- Few observations (N)
- Many variables (K)
- -Noisy data
- Missing data/excluded regions
- Multiple objectives

#### Implications

- High degree of correlation (Many variables are related)
- Difficult to analyse with conventional methods
- Require methods for simplification and visualisation





# Multivariate Statistics The Basics



## Principle Component Analysis PCA

A multivariate statistical approach that facilitates the identification of differences or similarities between groups



Data table  $\rightarrow$  variable space

	var. 1	var. 2	var. 3				
1							
2							
3					ĺ		
4				$\rightarrow$			
5							
6							
Ν							

The whole table yields a swarm of points in variable space





•Centering – move centre of point swarm to the variable origin







The first principal component  $(PC_1)$ is set to describe the largest variation in the data, which is the same as the direction in which the points spread most in the variable space

<u>The Score value</u>  $(t_{i1})$  for the point i is the distance from the projection of the point on the 1:st component to the origin.

 $PC_1$  hence is the first latent variable in a new coordinate system that describes the variation in the data.





The second principal component  $(PC_2)$ is set to describe the largest variation in the data, Perpendicular (orthogonal) to the 1:st component



•This is the scores plot similarities or differences between samples can now be seen.

•A corresponding loading plot describes the variables relationships

•allows interpretation of the scores plot by showing which variables are responsible for similarities and differences between samples.

•The perpendicular distance from the object to the projection on the plane is the residual of the two PCs.

•Two PCs make a plane (window) in the Kdimensional variable space. The points are projected down onto the plane which is lifted out and viewed as a two dimensional plot.









The loading (p) is described as the cosine of the angle between the original variable and the PC.

With  $p_{x,1} = \cos(\theta_{x,1})$  and  $p_{x,2} = \cos(\theta_{x,2})$ and  $\theta_{x,1}$ : angle between axe (rt<sub>x</sub>, *m*/*z*<sub>x</sub>) and PC1 and  $\theta_{x,2}$ : angle between axe (rt<sub>x</sub>, *m*/*z*<sub>x</sub>) and PC2



Mice Loadings Plot



SIMCA P+ 10.5 - 28/07/2004 10:3\*

# Interpretation of PCA Waters

#### Scores

- Observations (spectra)
- Trends, patterns, groups



#### Loadings

- Variables (m/z)
- Correlation, influence





### **Advanced Methods**

**OPLS and the S-Plot** 



 Often the effect we are looking for is masked by other unwanted variation

- OPLS is able to rotate the projection so that the model focuses on the effect of interest
- Here we want to focus on control vs. treated but gender is the bigger influence on X
- OPLS causes a rotation so that the first OPLS component shows the between class difference



Control vs. Treated

JATIC POSSTRIE"



# The S-Plot for OPLS-DA Waters The Science of WHAT'S POSSIBLE."

Developed by A team of scientists from:

•The University of Umea

•Umetrics

•Waters

Wiklund S, Johansson E, Sjöström L, Mellerowicz EJ, Edlund U, Shockcor JP, Gottfries J, Moritz T, Trygg J. Visualization of GC/TOF-MS-Based Metabolomics Data for Identification of Biochemically Interesting Compounds Using OPLS Class Models. Anal Chem. 2008 Jan 1;80(1):115-122. Variable Confidence (correlation)



#### Variable Contribution (covariance)

# S-Plot vs Loadings High in Wild Type

#### Complex Mixture.M6 (OPLS), WT vs. nor p[Comp. 1]/p(corr)[Comp. 1]



#### Complex Mixture.M7 (OPLS), WT vs. rin p[Comp. 1]/p[Comp. 2] Colored according to model terms



# S-Plot vs Loadings High in Wild Type

# Complex Mixture.M7 (OPLS), WT vs. rin p[Comp. 1]/p(corr)[Comp. 1]



#### Complex Mixture.M7 (OPLS), WT vs. rin p[Comp. 1]/p[Comp. 2] Colored according to model terms





## MarkerLynx XS Software Demo



### Ion-Mobility /oa-TOF Mass Spectrometry And its Application to the Study of Lipids

John Shockcor Director of Metabolic Profiling Business Development, Waters Corp. Visiting Fellow, Dept of Biochemistry, University of Cambridge

# Lon mobility Waters The science of what's possible."

- Separation of ions as they drift through a gas under the influence of an electric field
- Rate of drift is dependent on the ion's mobility through the gas
- Mobility is dependent on factors such as,
  - Size
  - Shape
  - Charge







Thomson, J. J.; Rutherford, G. P., "Conduction of Electricity Through Gases", Dover, NY, 1928.

Thomson, J. J.; "Rays of Positive Electricity", Green and Co., London, England, 1933.







## Ion Mobility Spectrometry Perspective:- The Benefits

#### Rapid orthogonal separation capability



- Improve peak capacity
- Provide novel functionality
- Enhance performance of the mass spectrometer
- IMS is an orthogonal separation
- How can T-Wave technology be implemented ?


















# Data clean-up step using the drift time function

THE SCIENCE OF WHAT'S POSSIBLE."

ers











# Trap fragmentation (CID-IMS)







# 3D representation of TAP for Verapamil





#### **Lipidomics using Ion Mobility**









#### ids and 10 $\mathbf{D}$

THE SCIENCE OF WHAT'S POSSIBLE."



### MS<sup>E</sup> analysis for Lipids and TAG's showing polar head group for PC's

#### Waters THE SCIENCE OF WHAT'S POSSIBLE."











### PC 18:0/18:2 vs. PC 18:1/18: Fragmentation Differences







#### Drift times for different TAG acy carbon numbers









# Acknowledgements

#### The Metabolic Profiling Group

- Jose Castro-Perez
- Kate Yu
- Emma Marsden-Edwards
- Henry Shion

#### **Our Colleagues**

- Andy Baker
- Richard Lock
- Steve McDonald
- Tim Riley
- Richard Gilpin
- Jeff Goshawk
- Iggy Kass

#### **A Few of Our Collaborators**

- Johan Trygg U. of Umea
- > Erik Johansson Umetrics
- Jules Griffin U. Of Cambridge
- > Kathie Knights Flinders U.
- > John Miners Flinders U.
- Andrew Tyler Novartis
- United States Department of Agriculture
- Jeremy Nicholson Imperial College
- Elaine Holmes Imperial College
- Liz Want Imperial College
- The MetaPhor Project (EU)
- Gerhard Wagner Harvard Medical Ctr.
- Colorado State U.