A/Prof. Olivier Thas
Ghent University, Belgium

Title:
Principal Bi-correlation Analysis: A new multivariate data analysis method for the integration of three data sources, with applications in early drug development

Abstract:
Classical multivariate data analysis methods were developed for the analysis of one or at most two data sources. In many areas of application a need has grown for statistical methods that can deal with three data sources simultaneously. In this talk the focus is on a new problem in early drug development. Data is available on (1) the chemical structure of molecules, (2) measurements on the bio-activity of these chemical compounds and (3) gene-expressions induced by these compounds in selected cell lines. We have developed new multivariate statistical methods that analyse these three data sets simultaneously. In particular, the method aims to find a sparse linear combination of genes so as to maximise its joint correlation with a single chemical fingerprint feature and a single bio-assay outcome. In a second stage the results are aggregated over the full chemical structure data set. The chemical fingerprints selected by this procedure are expected to act on the bio-assay outcome through a genomic pathway as identified by the sparse linear combination of the gene expression levels. The method is applied to real data sets from the pharmaceutical industry.

Authors: Olivier Thas (1,2) and Federico Mattiello (1)

Affiliations:
(1) Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Belgium
(2) National Institute for Applied Statistics Research Australia (NIASRA), School of Mathematics and Applied Statistics, University of Wollongong, Australia