CSC - COPENHAGEN SCHOOL OF CHEMOMETRICS

NSCPHD1191 - LPhD191

June 1st – July 3rd, 2015

Main responsible: José Manuel Amigo
Description of the course:

Spectroscopy and Chemometrics (SPECC) group is more than happy to announce the new edition of the PhD course:

Copenhagen School of Chemometrics - CSC

CSC is a five-week school designed to be an introduction to different key aspects of advanced data analysis in science. CSC addresses to PhD students/post-docs, associate professors, etc. who want to acquire or refresh basic knowledge on multivariate data analysis from different disciplines (Chemistry, Physics, Biology, Geology, Environmental Sciences, etc.). CSC also addresses to researchers working in industry or research laboratories who want to implement multivariate data analysis in their daily research environment.

CSC is thought to be an intensive school. Therefore, CSC will be held in five weeks structured in ten different seminars/workshops where the students will be more than welcome to work with their own data together with high-qualified teachers.

Targets of CSC:

CSC aims at being a platform for:

- **Learning basic and advanced data analysis methods:** CSC is specifically designed for researchers who want to start using data analysis in their routine work.

- **Sharing knowledge and interchange ideas between students covering different scientific backgrounds:** One of the key points of CSC is the interaction between the students to discuss issues and troubleshooting always within the framework of scientific data analysis and performance. The reports days and the workshop will offer the opportunity to the students to discuss, share and improve their main issues and initiatives in a professional environment.
- Meeting world-wide recognized experts of Multivariate Data Analysis in an open discussion forum environment: CSC will count on teachers that are well-recognized experts on chemometrics and multivariate data analysis in their respective fields. This, at the same time, will offer the possibility of opening new collaborative frameworks between students and teachers.

- Flexibility in the seminars and ECTS credits: The students can chose to attend the seminars which they consider more relevant for their research. There is no a minimum of seminars that the student must attend. Also, they will have the opportunity to deepen into any multivariate method in the WORKSHOP days. In total, attending the whole school, one student can obtain 12 ECTS credits.

Timetable, topics and lecturers:

The timetable and topics for CSC - 2015 are:

<table>
<thead>
<tr>
<th>Week</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LinAl</td>
<td>LinAl</td>
<td>LinAl</td>
<td>DoE</td>
<td>DoE</td>
</tr>
<tr>
<td>2</td>
<td>EXPLORE</td>
<td>EXPLORE</td>
<td>EXPLORE</td>
<td>REGRESS</td>
<td>REGRESS</td>
</tr>
<tr>
<td>3</td>
<td>CLASS</td>
<td>CLASS</td>
<td>VarSel</td>
<td>VarSel</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>MCR</td>
<td>MCR</td>
<td>MCR</td>
<td>3-Way</td>
<td>3-Way</td>
</tr>
<tr>
<td>5</td>
<td>MSCP</td>
<td>MSCP</td>
<td>MetaboMetrics</td>
<td>MetaboMetrics</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Acronym</th>
<th>ECTS</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinAl</td>
<td>1,5</td>
<td>Linear Algebra for Multivariate Data Analysis</td>
</tr>
<tr>
<td>DoE</td>
<td>1</td>
<td>Introduction to Design of Experiments</td>
</tr>
<tr>
<td>EXPLORE</td>
<td>1,5</td>
<td>Multivariate data exploration and pre-processing. PCA</td>
</tr>
<tr>
<td>REGRESS</td>
<td>1</td>
<td>Multivariate calibration methods. Partial Least Squares</td>
</tr>
<tr>
<td>CLASS</td>
<td>1</td>
<td>Linear Classification in multivariate data analysis. KNN, LDA, SIMCA, PLS-DA</td>
</tr>
<tr>
<td>VarSel</td>
<td>1</td>
<td>Variable selection methods in multivariate regression and classification</td>
</tr>
<tr>
<td>MCR</td>
<td>1,5</td>
<td>Multivariate Curve resolution</td>
</tr>
<tr>
<td>3-Way</td>
<td>1</td>
<td>Introduction to 3-Way methods. PARAFAC, PARAFAC2 and Tucker3</td>
</tr>
<tr>
<td>MSCP</td>
<td>1</td>
<td>Workshop on Multivariate Statistics Process Control</td>
</tr>
<tr>
<td>MetaboMetrics</td>
<td>1,5</td>
<td>Workshop on Metabolomics Data Analysis</td>
</tr>
</tbody>
</table>
All the seminars run from 9 a.m. to 5 p.m. with a lunch break. Organization of the breaks and the final timetable for each seminar will be under the responsibility of the corresponding teacher.

Lecturers:

1.- LinAl: Dr. Morten A. Rasmusen.
2.- DoE: Dr. Frans van den Berg.
3.- EXPLORE: Prof. Rasmus Bro.
4.- REGRESS: Dr. José Manuel Amigo.
5.- CLASS: Dr. Davide Ballabio.
6.- VarSel: Dr. Åsmund Rinnan.
7.- MCR: Dr. José Manuel Amigo.
8.- 3-WAY: Dr. Rasmus Bro.
9.- MSPC: Dr. Frans van den Berg
10.- MetaboMetrics: Prof. Age Smilde and Dr. Frans van der Kloet.

For PhD students:

Each seminar accounts for 1 or 1.5 ECTS (See previous table). All PhD students who want to obtain the corresponding ECTS credits from the seminars will be obliged:

1) To attend the corresponding seminars.

2) To deliver the corresponding reports (see below). The length, content and the final delivering date of the reports will be specified by the corresponding lecturer.
Detailed description for each seminar:

1.- LinAI. Linear Algebra for Multivariate Data Analysis

Linear Algebra is the basis of the Multivariate Data Analysis. Therefore, it is important to give it a central position in any school about multivariate data analysis. This seminar will be a clear introduction to the linear algebra behind the multivariate models that will be presented in the school.

**Previous knowledge needed:** None.

**Software:** Matlab.

2.- DOE. Introduction to Design of Experiments

The basic theory and practice of Design of Experiments and statistical inference is revisited. The aim is to give a practitioner idea or reminder about the main features and uses of DoE. The seminar will be based on teaching hours and some guided exercises.

**Previous knowledge needed:** Basic statistics.

**Software:** JMP. A full demo of JMP can be downloaded for free at:

http://www.jmp.com/applications/doe/

3.- EXPLORE

Principal Component Analysis has become the most powerful and versatile tool for analyzing data tables in Analytical Sciences. Here we present a course to show the main benefits and drawbacks of PCA when it is used for different kind of analytical data: Spectroscopy, environmental assessment, experiments performance, chromatography, etc. Moreover, preprocessing of different type of data will be also addressed in the seminar.

**Previous knowledge needed:** None.
Software: Matlab and/or PLS-Toolbox. A full demo version of PLS-Toolbox can be downloaded at:

http://www.eigenvector.com/

4.- REGRESS

If PCA is the keystone of pattern recognition methods, PLS is the keystone of multivariate calibration methods. This seminar will give a general overview of different multivariate calibration strategies and will focus in Partial Least Squares regression.

Previous knowledge needed: Basic knowledge of PCA and linear algebra.

Software: Matlab and PLS-Toolbox. A full demo version of PLS-Toolbox can be downloaded at:

http://www.eigenvector.com/

5.- CLASS.

The seminar is focused on the theory and practice of two linear classification tools (SIMCA and PLS-DA). The seminar will be based on teaching hours with guided exercises and practical sessions with real cases.

Previous knowledge needed: Basic knowledge of PCA.

Software: Matlab.

6.- VarSel.

This new seminar aims at revisiting the most important variable selection methods for regression and classification purposes with the aim at improving the performance of the models.

Previous knowledge needed: Basic knowledge of multivariate regression and classification methods.
Software: Matlab.

7.- MCR

In the last years, curve resolution techniques are gaining importance in modeling of different analytical data types. Especially, Multivariate Curve Resolution has widely demonstrated its usefulness in kinetic modeling, solving problems in chromatographic data (peak resolution/deconvolution) and hyperspectral images. This seminar will offer a general overview of curve resolution methodologies and will focus in multivariate curve resolution applied to different analytical problems.

Previous knowledge needed: Basic knowledge of PCA.

Software: Matlab and MCR-ALS Toolbox. MCR-ALS Toolbox can be downloaded for free at: http://www.mcrals.info/

8.- 3-Way

Three-way analysis, covering PARAFAC, PARAFAC2 and Tucker3, is the curve resolution method for three-dimensional and multi-dimensional structures.

Previous knowledge needed: Basic knowledge of PCA.

Software: Matlab and PLS-Toolbox. A full demo version of PLS-Toolbox can be downloaded at:

http://www.eigenvector.com/

9.- MSPC

Nowadays, there are more and more companies that want to implement fast and reliable multivariate statistical methods for real-time control of the performance of different processes. In this seminar, the main Multivariate Statistics Process Control (MSPC) tools will be revisited and different real examples of how MSPC has been implemented in companies will be shown.
Previous knowledge needed: Basic knowledge of PCA, regression methods and linear algebra.

Software: Matlab.

10.- MetaboMetrics

Undoubtedly, metabolomics has become an important tool to address metabolic pathways in different biological systems. This seminar/workshop aims at offering a general introduction to the main metabolomics data analysis methods, including preprocessing of metabolomics data. Moreover, Analysis of variance-Simultaneous Component Analysis (ASCA) and Metabolic Network Inference will be explained and exemplified.

Previous knowledge needed: Basic knowledge of PCA, regression methods and linear algebra.

Software: Matlab.

Location, subscription and fees:

Location: CSC will be held at the department of food sciences of the University Of Copenhagen, Denmark. The exact location will be specified later.

Subscription: For joining us, just send an e-mail to Dr. José Manuel Amigo (imar@food.ku.dk) or Ms. Jeanette Venla Hansen (jvh@food.ku.dk). The e-mail must contain: Full name, position, university/company, country, which seminars to attend, need of ECTS credits?

Maximum number of participants: The maximum number of participants will be 15 for each seminar.

Fees: The seminars are free of charge for everybody. NEVERTHELESS, there will be a fee of 2000 Danish Krone for not coming when joining.