

## **L07. Multi-block and Path Methods – The next generation of chemometric methods**

*A. Höskuldsson*

*Centre for Advanced Data Analysis, Kgs Lyngby, Denmark*

When data are collected for modelling tasks there is usually available information that should be included in the modelling task. E.g., data from a process plant is basically of multi-block type. Each sub-process generates separate data. Similarly, organisational diagrams may suggest sub-divisions of data. Here are presented methods where data are organised in a directional network of data blocks. The data blocks are arranged in directional paths. Each data block can lead to one or more data blocks. It is assumed that there are given a collection of input data blocks. Each of them is supposed to describe one or more intermediate data blocks. The output data blocks are those that are at the end of the paths and have no succeeding data blocks. The optimisation procedure finds weights for the input data blocks such that the size of the total loadings for the output data blocks is maximised. When the optimal weight vectors have been determined, the score and loading vectors for the data blocks in the network are determined. Appropriate adjustment of the data blocks is carried out at each step. Regression coefficients are computed for each data block that show how the data block is estimated by data blocks that lead to it. Methods of standard regression analysis are extended to present methods. Three types of 'strengths' of relationship are computed for each set of two connected data blocks. First is the strength in the path, second the strength where only the data blocks leading to the last one is use and third if only the two are considered. Cross-validation and other standard methods of linear and non-linear regression are carried out in a similar manner.

In industry processes are organised in different ways. It can be useful to model the processes in the way they are carried out. By proper alignment of sub-processes, overall model can be specified. There can be several useful path models during the process, where the data blocks in a path are the ones that are actual or important at given stages of the process. An important aspect of these methods is that we get score vectors for each data block. Thus we can study the role of each data block in the network.