IUFOST2006/859
Understanding co-variant patterns in biological data by the use of multiblock methods

A. Kohler[a], M. Hanafi[b], D. Bertrand[b] and E.M. Qannari[b]

[a]Centre for Biospectroscopy and Data Modelling, Matforsk, Norwegian Food Research Institute, 1430 Ås, Norway

[b]ENITIAA, Nantes, La Géraudière BP 82225, 44322 Nantes cedex 03, France

achim.kohler@matforsk.no

Modern science and engineering are increasingly interdisciplinary and in many cases different sets of multivariate data (multiblock data) are collected by very different and highly advanced measurement techniques. This situation has raised the need for data analytical methods that can help to extract information from the large volumes of data, to reveal covariant patterns between the different data blocks and to establish predictive models using the different sources of data. A multiblock situation appears whenever different types of data are collected from corresponding samples. Examples are process technology, where measurements are performed at different stages of processes with the same or different measurement techniques, functional genomics where genetic, molecular biologic, phenotypic methods are used to understand the causal chain from genotype to phenotype (Oust et. al., 2006), and biospectroscopy where highly interpretative spectroscopic methods are applied in order to elucidate biochemical mechanisms (Bertram et al., 2006). In all theses situations a large number of samples (N) is measured with different techniques and data blocks Xb with (b=1,2,oe,B) are obtained for every measurement principle b. The total set of measurements can be described by the supermatrix $X=[X1,X2,...,XB]$ where the single blocks have a row to row correspondence, i.e. the same row is referring to the same sample. In the paper at hand we discuss different multiblock approaches for the treatment of multiblock data in food science. As examples we consider data from process technology, functional genomics and biospectroscopy in general. The spectral data sets contain spectroscopic data as FT-IR, Raman, Fluorescence and low field H-NMR data and other data types like HPLC or gene expression data. The data are obtained from designed experiments. The different multiblock methods are discussed with respect to their capability to establish relationship between the interpretative signals in the different data sets and with respect to their ability to extract information with respect to the design.

References

1. Oust A, Moen B, Martens H, Rudi K, Næs T, Kirschner C, Kohler A 2006 Analysis of co-variation patterns in gene expression data and FT-IR spectra, Journal of Microbiological Methods, in press.

2. Bertram C, Kohler A, Boecker U, Ofstad R, Andersen HJ 2006 Heat-induced changes in myofibrillar protein structures and myowater of two pork qualities. A Combined FT-IR spectroscopic and low-field NMR relaxometry study, Journal of Agricultural and Food Chemistry, in press.