

L01. Chemometric data analysis — a holistic process with many elements

K. Esbensen

Aalborg University Esbjerg, Esbjerg, Denmark

Chemometric data analysis has a natural focus on the statistical/data analytical methods applied, or developed. While this may seem understandable, there is not always a complementary emphasis on the data analysis objective, nor of data quality. A holistic understanding of chemometric data analysis must (at least) include the following element

1. Is data generation (observation) active (controllable) or passive (WYSIWYG)?
2. If active, is use made of an appropriate experimental design?
3. If passive, is appropriate consideration made to representativity, Theory of Sampling (TOS)?
4. Are the variables appropriate for the data analysis objective/method - e.g. what constitute proper criteria for problem-specific variable selection?
5. Does the realized matrix (objects/variables) support the data analysis method?
6. Is the realized data quality commensurate with the specific data analysis method requirements?

It is advantageous to conceptualize of chemometric data analysis as a process starting with a thorough specification of the data analysis objective in the relevant problem context (from which follows an optimal choice of method), followed by experimental design and/or representative sampling before chemical, physical (other) analysis generates the data. Proper consideration of all such pre-data analysis issues is necessary to increase the potential of meaningful/useful data analysis results. Indeed, most often the only guarantee for interpretable results lies already in the pre-data stages. Finally: in holistic data analysis the crucial role of proper validation reigns supreme, without which no scientific merit at all.