Chemometrics and Validation in Vibrational Spectroscopy

J. Schmitt[1,2] and Th. Udelhoven[1]

[1]University of Trier, FB VI, Spectroscopy Group, [2] Synthon, analytics and pattern recognition, Martinstr.8, 54317 Gusterath, email: synthon_analytics@t-online.de

Vibrational spectroscopy as a new tool in microbiology and biomedical applications has witnessed an enormous progress in the past. NIR, FT-IR and FT-Raman spectroscopy is used in manyfold techniques to identify bacteria, yeasts and fungi, characterize mammalian cells, tissues, body fluids and for spectral imaging. Along with the development of applications and the detailed spectral information available, very high dimensional patterns, consisting of hundreds of spectral elements occur. If all components of patterns are taken into equal account, the process of pattern classification becomes extremely complicated and even impossible to achieve reasonable practical results. Therefore, it is of great importance to properly select a small part from a large number of spectral components for the classification or quantification process. Many new techniques are proposed for this task called feature selection and feature extraction; PCA, PLS, Wavelet-Transforms and statistical parameter in combination with genetic algorithms and several optimization routines.

Based on a large dataset of an animal study with almost 300 individuals and repeated measurements, the problems arising with the different methods of feature selection will be discussed. In the following step, the preselected features are used as input variables for the classification model. Many powerful classification techniques, require a data splitting in training and validation datasets or using validation strategies like leave-k-out procedures. How representative are these datasets e.g from biological samples and where are the pitfalls in validating our models for practical purposes ? Several examples on the same dataset will demonstrate the different results after classification and validation. When new techniques are to gain confidence in chemometrics and practical applications of vibrational spectroscopy, a well described and thoroughly tested validation and evaluation procedure is indispensable.