Abstract

Some aspects of chemometric modeling for near infrared spectroscopic analysis

Yizeng Liang

Research Centre of Modernization of Traditional Chinese Medicines, Central South University, Changsha 410083, P. R. China
Corresponding author: yizeng_liang@263.net

Introduction

Some important aspects of chemometric modeling for near infrared spectroscopic analysis are discussed in this presentation. As we know, most systems analysed in near infrared spectroscopy are quite complex, quite different from the classic samples analysed by common instruments with standards. From this point of view, the aim of NIR analysis is somewhat like qualitative and quantitative characteristic analysis for samples instead of for pure chemical compounds as classic analysis is. For NIR spectral analysis, most analytical systems are a kind of generalised "grey" system, since they always need some other chemical or instrumental analysis methods as a standard for calibration. Thus, variable selection and robust regression for NIR chemometric modeling are of special importance.

Materials and Methods

Hereby we propose a new strategy to deal with this kind of analysis for NIR spectra. The strategy is named as model population analysis (MPA). It works mainly in three successive steps: (1) obtain a sub-dataset by Monte Carlo sampling; (2) establish a sub-model for each sub-dataset; (3) statistically analyse some interesting outputs of all the sub-models from four different spaces in NIR spectral analysis, that is, variable space, sample space, parameter space and/or model space. Based on such results, some new algorithms for variable selection, robust diagnosis etc can be developed.

Results and Discussion

With the help of MPA, several case studies for variable selection and robust regression will be discussed in some detail to show its outstanding performance for NIR spectral analysis. The results from NIR spectral analysis for determination of fat and protein in liquid milk showed that the ARMSEP (average root of mean square of errors of prediction) of the NIR model could be significantly improved, that is, ARMSEP=0.1936 before variable selection and ARMSEP=0.0931 after variable selection for fat and ARMSEP=0.1000 before variable selection and ARMSEP=0.0550 after variable selection for protein, respectively.

Conclusion

Model population analysis may be a potential methodology for chemometric modeling of NIR spectral analysis.

Reference paper as:
Y. Liang (2012). Some aspects of chemometric modeling for near infrared spectroscopic analysiss (abstract), in:
Proceedings of the 15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley,C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, p. 140.