

Automated optimisation of NIR calibrations

R. Bossart, H. Kellerhals, H. Keller and J. Oelichmann

Büchi Labortechnik AG, Meierseggstr. 40, 9230 Flawil, Switzerland

Introduction

Near-infrared spectroscopy has become firmly established and widely acknowledged as a mature and non-destructive quality assurance method. It is well known that the successful application of NIR spectroscopy is highly dependant on the proper selection of calibration and validation samples as well as the robustness, sensitivity, selectivity and portability of the chemometric calibrations.

To establish and optimise a NIR calibration, a large number of different parameters are available to the user. A choice between the various algorithms like Cluster Analysis, Principal Components Regression (PCR) or Partial Least Squares Regression (PLSR) has to be made and wavelengths and “factors” have to be selected. Additionally data pre-treatments need to be used, which can be combined in any required sequence and in any number. The quality of the calibration is evaluated by means of different statistical terms like regression coefficient, PRESS, SEE, SEP, BIAS and others. The assessment and optimisation of a calibration as the basis for its successful use in the routine application greatly depends on the experience of the user.

The required amount of calibration and validation samples and especially the “mysterious” chemometric calibrations are in many cases a big hurdle for wider acceptance of NIR spectroscopy.

Automated calibration development and optimisation

The principle workflow of the automated calibration development and optimisation is outlined in Figure 1. The heart of this “Calibration Wizard” is the knowledge base, which makes use of the extensive experience during the development of numerous NIR applications. It selects various wavelengths ranges and data pre-treatment, which are reasonable for the application to be solved. An important step during the calibration development is the correct selection of Primary and Secondary Factors, for which various statistical measures are used. The quality of the calibrations is judged by the Q-value. This quality index evaluates the most important statistical measures and qualifies the calibrations using a number between 0 and 1. During the iterative process a large number of calibrations will be calculated.

The individual steps are described in more detail in the following sections.

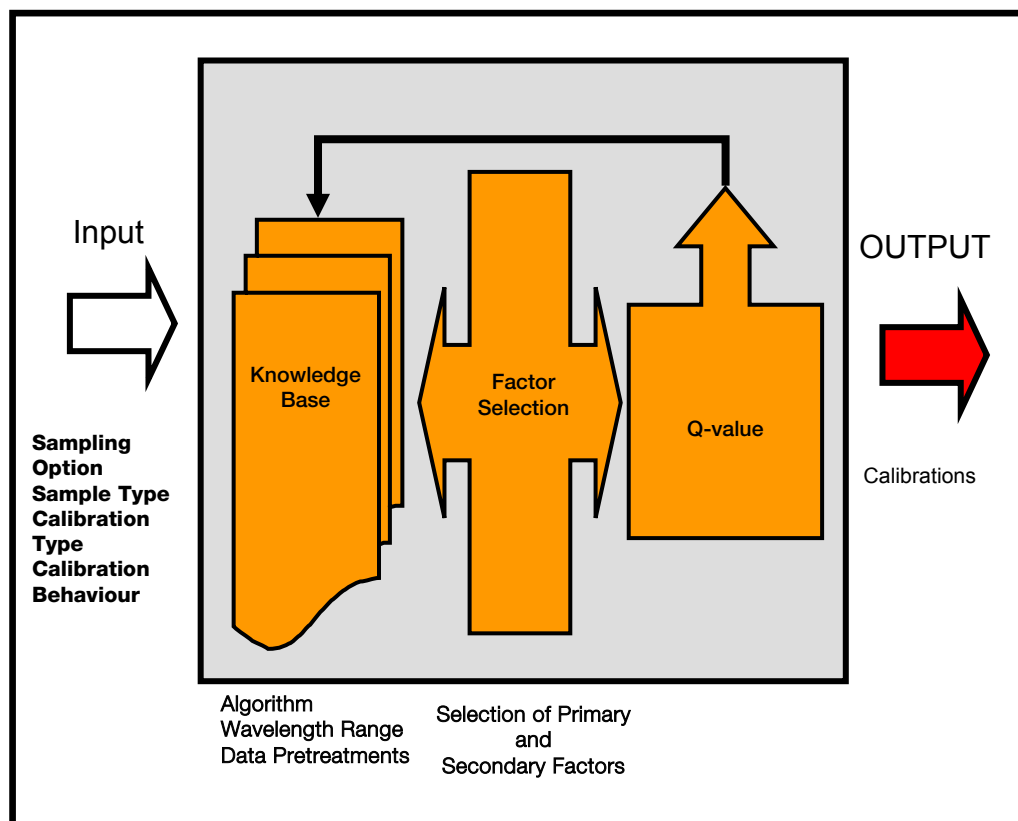


Figure 1. Principle workflow during the automated calibration development and optimisation.

Definition of calibration and validation sets

For a chemometric calibration independent sets of calibration and validation spectra are needed. This classification can be done either automatically on a random basis or manually defining rules (Figure 2).

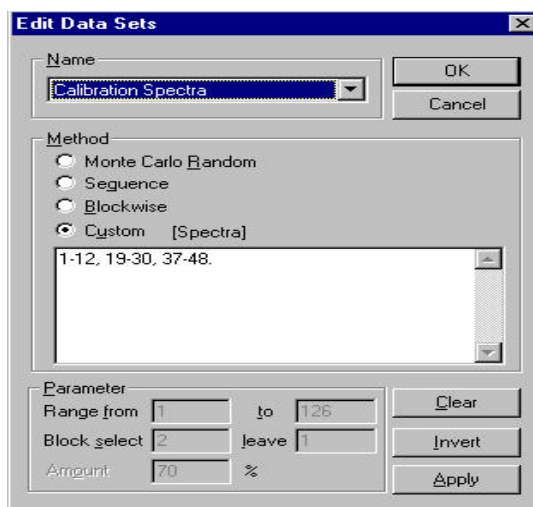


Figure 2. Possibilities for the definition of calibration and validation sets.

Required information

Having been fed with information about the nature of the samples, the sampling option used and the type and required characteristic of the calibration, the automatic calibration uses a comprehensive knowledge base for the development of an optimised calibration (Figure 3).

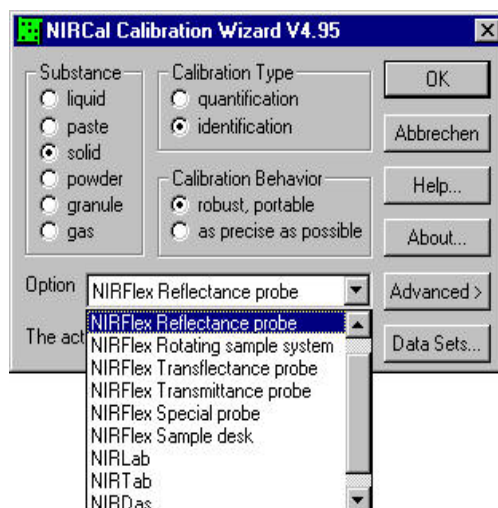


Figure 3: Required information before performing an automated calibration development.

The parameters to be used for calibration development and optimisation are highly dependent on the required information, namely the sample, the sampling options and the calibration type and characteristic. The dependencies are demonstrated in Figure 4.

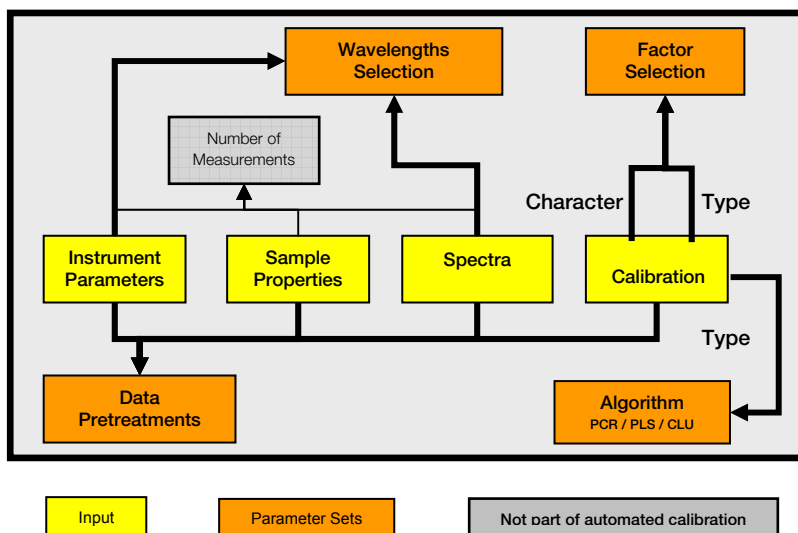


Figure 4. Elements of the knowledge base showing the dependencies between the hardware used, sample properties, required calibration type and characteristic and the calibration parameters.

Selecting the correct number of primary and secondary factors

The Primary Factors are used for the reconstruction of the spectra. They are also used for the determination of the residual values. The correct selection of the Secondary or Calibration Factors is important for the description of the properties to be determined and the quality of the calibrations.

For the selection of the primary factors the X-PRESS (Predictive Residual Sum of Squares of the spectra, Figure 5) is the most important measure.

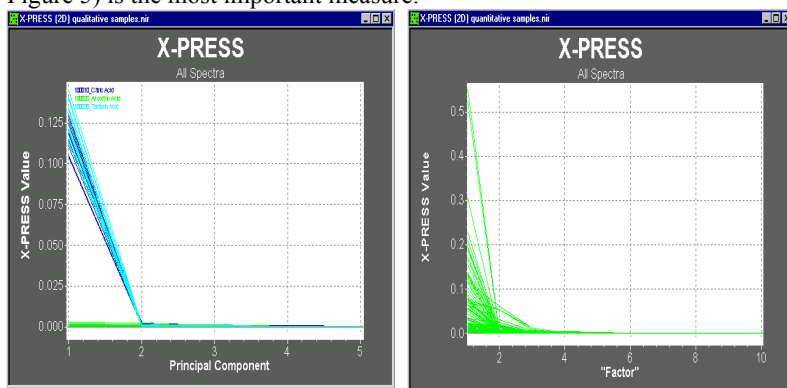


Figure 5. Examples for X-PRESS (qualitative left, quantitative right).

The selection of the secondary factors is based on different measures for qualitative and quantitative calibrations (see Table 1).

Judging the quality of the calibrations

During an iterative optimisation process various combinations of wavelengths ranges and data pre-treatments are applied in order to yield the best possible calibration. The difficulty during the

automatic process is to judge the quality of intermediate results. During the automatic optimisation a Q-value is calculated for every intermediate calibration.

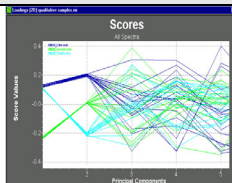
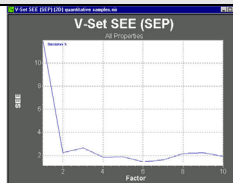
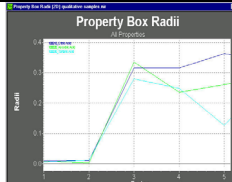
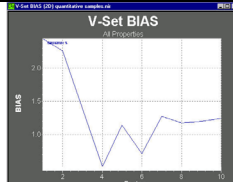
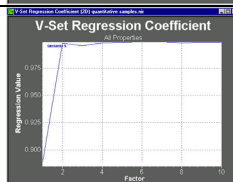
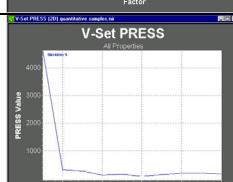
The Q-value takes all relevant statistical measures into account. It is calculated for qualitative and quantitative calibrations in different ways.

For **qualitative calibrations** information on the correct identification of the spectra in the calibration and the validation set, the number of clusters and the distribution of the clusters in the factor space are considered.

For **quantitative calibrations** *SEE*, *SEP* and *BIAS* as well as residuals and regression for the calibration and the validation set are taken into account.

The Q-value (Figure 6) qualifies all calibrations with a number between 0 (useless) and 1 (ideal).

Table 1. Important measures for the selection of the secondary factors.

Qualitative Calibrations		Quantitative Calibrations	
Scores		SEE and SEP, Highest importance	
Property Box Radii		v-set BIAS Highest importance	
		c-set, v-set regression Important	
		Consistency Important	
		c-set, v-set PRESS Minor importance	

The automatically generated calibrations are accessible to the user for evaluation. Therefore he is able to select that calibration for routine use, which is most robust and delivers the most reliable results.

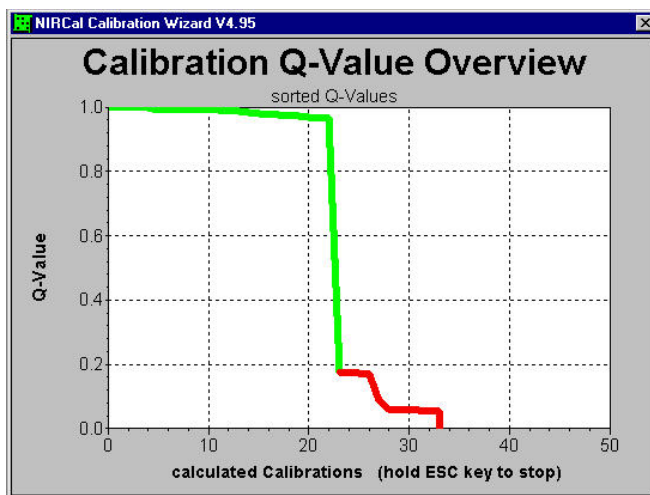


Figure 6. Display of the Q-value for the calculated calibrations.

Conclusion

The automated calibration procedure delivers robust calibrations which are ready for routine use. One advantage offered by this approach is a more objectified method of calibration development, which is much less affected by the users chemometric knowledge than manual calibration. From the point of view of industrial users of NIR spectroscopy, this calibration tool offers a particularly large potential for reducing the calibration costs, which will enhance the acceptance of the method.

Reference

1. R. Bossart, J. Grabinski, *Automatic Calibration Method* United States Patent 6,480,795 B1