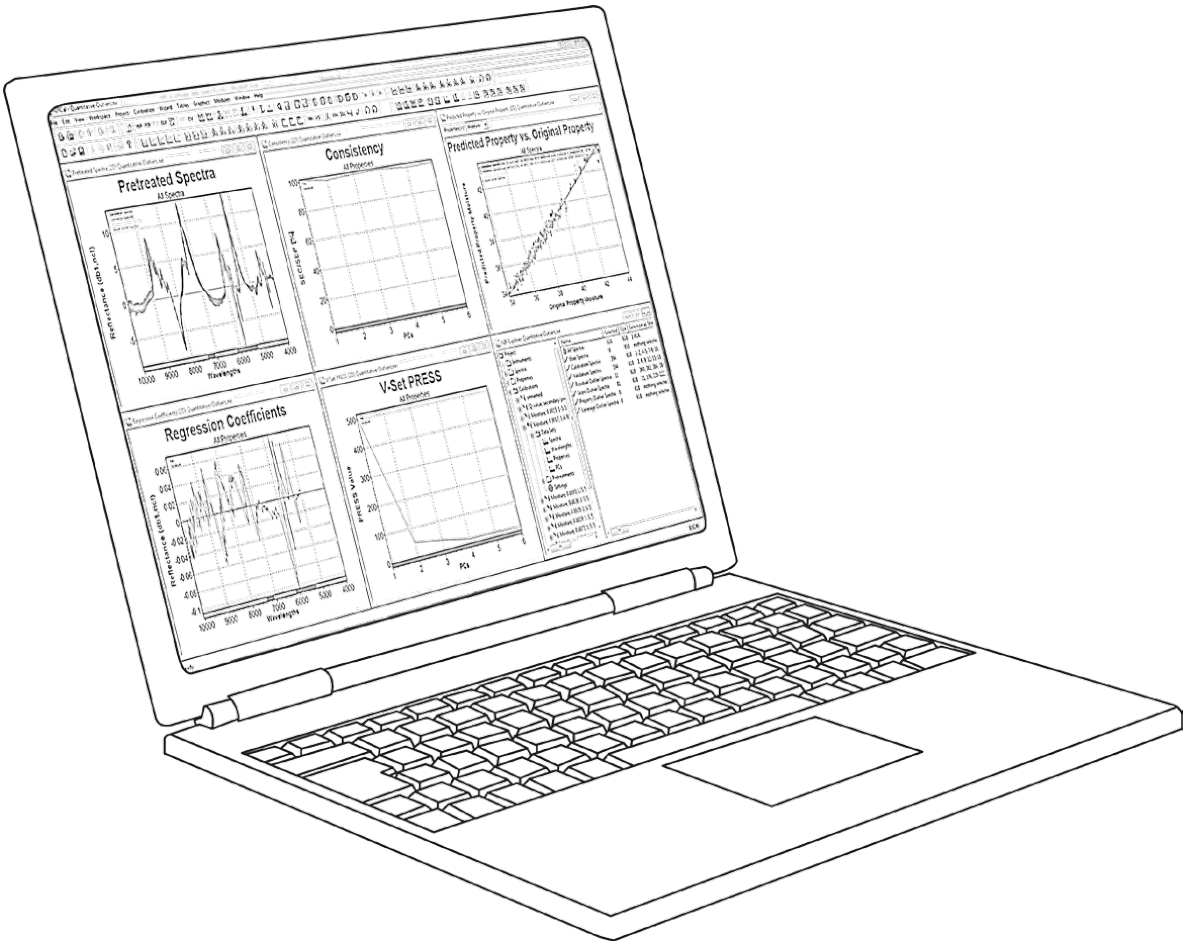
# NIRCal 5.6
## Operation Manual




11593921B en

Imprint

Product Identification:
Operation Manual, NIRCal 5.6
11593921B en

Publication date: 08.2016

BÜCHI Labortechnik AG
Meierseggstrasse 40
Postfach
CH-9230 Flawil 1

E-Mail: quality@buchi.com

# Table of Contents

# Welcome

**Dear NIRCal user,**
NIRCal is recognized as a reliable, comprehensive but also easy to use chemometric software package. It offers a wide variety of tools for method development and optimization.

The NIR spectroscopic technology and the range of applications develops continuously, just as new requirements from users and regulatory authorities. NIRCal is designed to fulfill those requirements today and in the future.

NIR spectroscopy is a powerful technology, which gives insight in product development projects and sound process understanding enabling optimization of processes and quality. NIRCal is designed to support all those tasks. Through high quality plots and informative workspaces, the analyst gets comprehensive overview of the system under study.

Much of the power in multivariate methods like PCA and PLS lies in the information one can get from interpretation of the models. This manual not only describes the functionality of NIRCal but teaches the user how to interpret the models to ensure the maximum return on your investment in NIR technology.

Should you come across any feature which needs improvement or extension, please do not hesitate to contact us. Your feedback helps us to continuously improve our software and is highly appreciated. Please send an e-mail to: info@buchi.com.

Every effort has been made at BUCHI to ensure that the information in this documentation is correct. BÜCHI Labortechnik AG accepts no responsibility for errors or omissions.

Information in this document is subject to change without notice and does not represent a commitment on the part of BÜCHI Labortechnik AG.

The software and/or files may be used or copied only in accordance with the terms of the license agreement.

BÜCHI Labortechnik AG, CH-9230 Flawil, Switzerland, March 2013.

Explanation of the used safety notes:

***NOTE***
Information on technical requirements. Non-compliance can lead to faults, inefficiency and production losses.

This manual is stored under: C:\Program Files\Buchi\NIRSolutions\NIRCal Manual.pdf"

# 1 Introduction

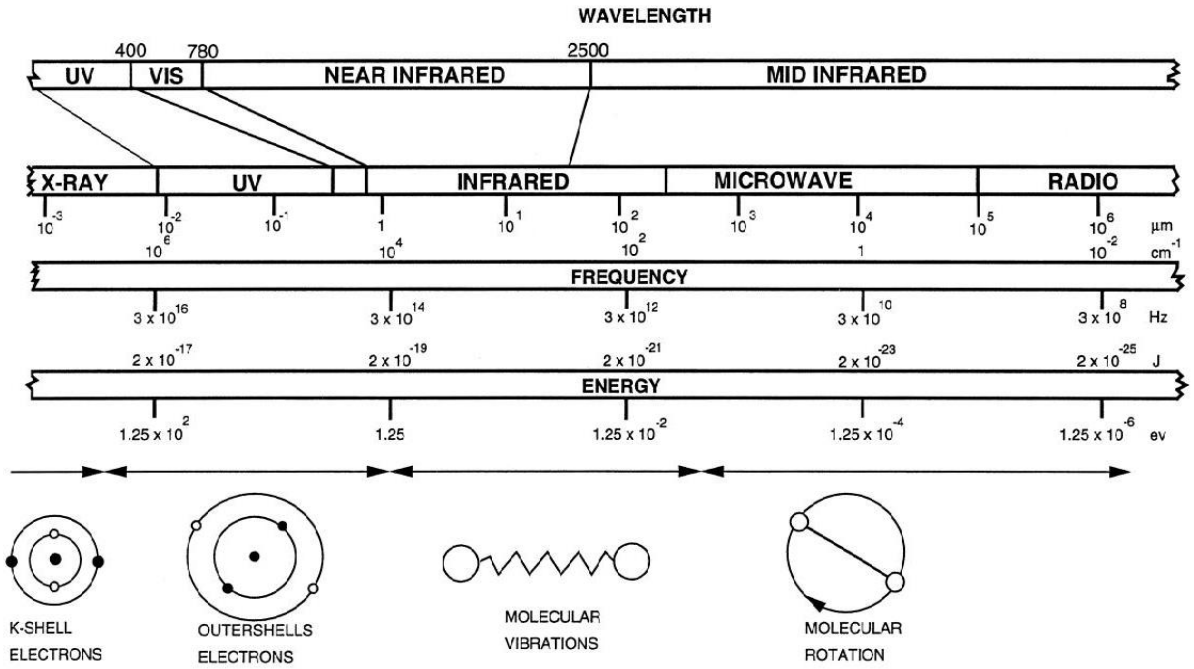
## 1.1 Introduction to NIR

### 1.1.1 Spectroscopy

Light is a fast time dependent sequence of electric and magnetic fields propagating in space.
Light can be characterised with physical properties, like:

| | |
|---|---|
| frequency: | $\nu, w = 2\pi\nu$ |
| wavelength: | $\lambda$ |
| wavenumber: | $k = 1/\lambda$ |
| velocity of propagation: | $c = \lambda\nu$ |
| energy: | $E = h \cdot \nu$ ($h = 6.63 * 10^{-34}$ Js) |

A summary of the electromagnetic spectrum can be seen in the Picture below. NIR lies between the visible and middle infrared range. With this light, the molecular vibrations are activated, similar to the infrared range.




Most of the molecule vibrations take place at a characteristic frequency which lies between $10^{12}$ to $10^{14}$ Hz. The osculating molecule interacts with the electrical field or light when the frequencies are the same.

| | |
|---|---|
| $\upsilon_{char} = \upsilon_{light} = 10^{12} - 10^{14}$ | Hz Infrared Light |

Not only are the basic vibrations absorbed with very high degree of excitation in the infrared range, but also frequencies 2 or 3 times higher. The harmonic frequencies are absorbed with a lower degree of excitation and lie in the near infrared range.
Using an infrared spectrum, the characterisations of molecules is possible. In the NIR spectra region mainly the overtones and combination tones of -CH, -OH and -NH groups are absorbed, so **NIR is suitable mainly for organic substances.**

The absorption bands are very broad and they often overlap, which can cause problems by direct interpretation of the spectra. The most often used evaluation method in spectroscopy, the Lambert-

Beer Law:

$$A = \log 1/T = c \cdot d \cdot \varepsilon$$

has just limited validity in the NIR spectroscopy. The application of chemometrics for the evaluation of NIR spectra is a must.

**Advantages of NIR spectroscopy**
The relative low degree of absorption coefficient of overtones and combination vibrations causes a low degree of absorbance in the NIR region. Solids have a high degree of reflected light and liquids can be measured for path lengths of several mm.

- **no sample preparation is** necessary.

The materials for the optics can be quartz, glass or sapphire, which are

- **cheap and robust**.

## 1.1.2 Chemometrics

NIR spectra are generally characterised by **very broad peaks** and a multitude of oscillation superpositions. Visual evaluation is therefore all but impossible.
Differences in the spectra of similar substances often consist merely of a slight shift or small change in shape of the wide absorption bands. For this reason, NIR spectra are basically evaluated with the aid of mathematical methods, which is why such significance is attached to the chemometric software.
**Chemometric** is the application of **mathematical, statistical** procedures for processing, evaluating and interpreting **large amounts of chemical data** (e.g. NIR spectra). The function of the chemometric software in NIR spectroscopy is to find a statistical correlation between the spectral data and the known (e.g. by laboratory analysis) property values of the samples used for the calibration.

If this **correlation is systematic**, it is possible to **predict** desired **parameters** of unknown samples (e.g. identity, quality, quantity) by recording the spectrum and subsequent evaluation by calculation.

The Büchi NIR spectrometer systems allow **several hundred intensity values** (reflectance / transmittance) of the measured NIR spectral region to be included in the calibration. In order to be able to draw maximum benefit from the measured region, the **Principal Component Analysis** (PCA) is applied.

The Principal Component Analysis for qualitative calibration allows the identification of different substances or similar product qualities. There are 2 ways to use PCA:

- **Cluster** calibrations and the
- **SIMCA** method

For quantitative analyses, three different calculation procedures have been implemented:

- Multiple Linear Regression (MLR),
- Pricipal Component Regression (PCR) and
- Partial Least Squares Regression (PLS).

These methods can be tested with a user made selection of **independent validation samples** or by using each sample for the calibration in **cross validation**.

*Chemometrics is the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods. [International Chemometrics Society (ICS)]*

### 1.1.3 Cluster calibration

Qualitative calibrations are used for identification of different chemical substances and for separation of different qualities of the same substances. The possible applications include identification of:

- substances with very different chemical characteristics;
- chemically similar substances;
- acceptable and rejectable qualities of a given substance.

The possible methods are Cluster or SIMCA. Both methods use PCA with the difference that Cluster method is always used for a group of similar substances, while for each substance a calculation is performed with SIMCA.

**Choosing the Calibration Samples**

Verifying substances in the laboratory often means, ascertaining if a sample can be assigned to a specific category (property); e.g., when checking incoming raw materials in a pharmaceutical company to see if the incoming raw material is the product that was expected or not. The question can be answered efficiently by recording an NIR-spectrum of the raw material and analyzing the spectrum with a qualitative calibration. The raw material can be correctly identified that it belongs to the expected category or falsely identified that it does not belong to the expected category.

To obtain a useful, qualitative calibration, first, calibration samples should be measured that **cover all expected allowable variations** of the quality of the product. For each property, several samples must be collected to cover variations such as different particle size, temperature, moisture or supplier. To obtain a representative set of spectra, we recommend measuring samples from at least **five to fifteen different batches** of each product that have been collected **over a period** of at least 6 months. This will ensure that all variations within a product will be represented in a calibration. Collecting samples can speed up the building of a proper calibration. But only stable and unchanged substances can be used.
Only **samples that have been tested with reference analysis** should be used for the calibration.
The combinations will be chosen randomly or with an adequate experimental design. Two thirds of these samples are composed to be the basic calculation data for the calibration. The remaining third is used for testing the calibration.

***NOTE***
When selecting the spectra in the calibration- and validation-set, it is important to assign all spectra of one sample either to the calibration or to the validation set.

**Calculating the Qualitative Calibration**

The spectra of different substances show the physical and chemical characteristics of each substance. Not all spectral differences are associated with the searched differences. “False“ differences arise, for example through a varying presentation of the sample because of different particle size or other noninteresting but allowable variations of the substance. Such “false“ differences can be reduced or partly eliminated with the help of appropriate **data pretreatments** during the calculation.

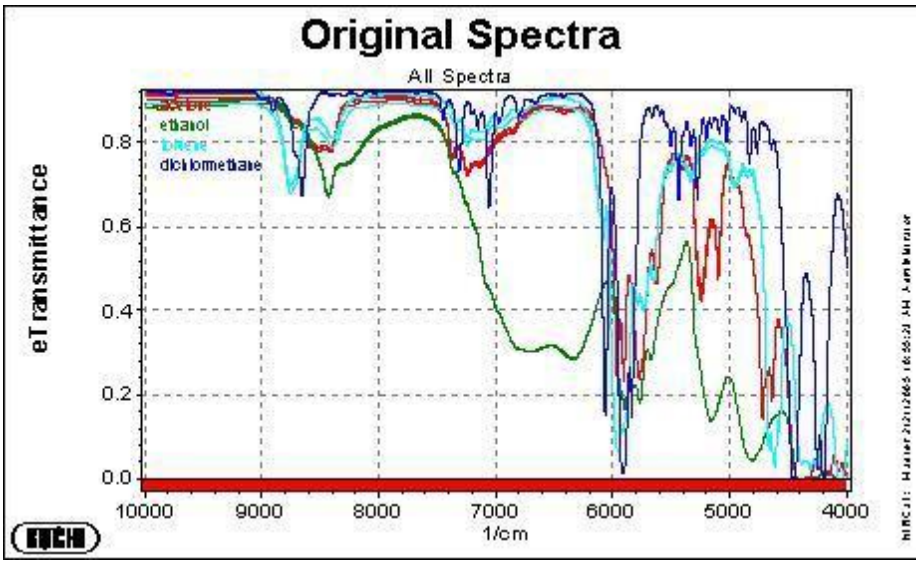



Figure.1: Solvent spectra (acetone, ethanol, toluene and dichlor-methane), measured through glass cuvettes with measurement option liquid. The slight shifts (light scattering) are not of interest and should be reduced with data pretreatments.

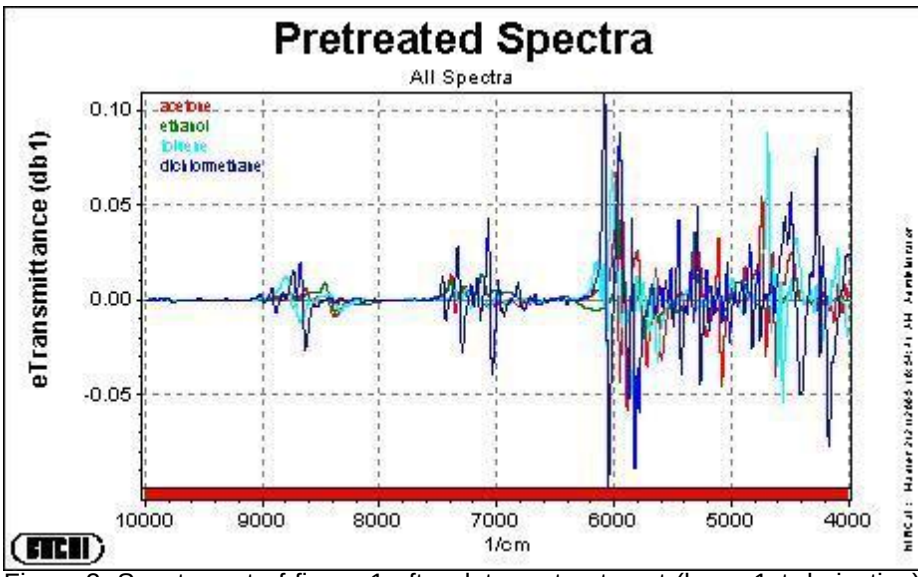



Figure.2: Spectra out of figure 1 after data pretreatment (here: 1st derivation) Systematic differences of the four solvents are clear and reproducible

For the actual calibration, the **Cluster** calibration using **PCA** has been used. This means a calculation of a new illustration of the spectral information with the target to show the main differences in the data set. Spectra appear after the calculation in the new illustration as points in a 2D or 3D Plot.*

The calculation can be performed automatically by the Calibration Wizard. According to the required input by the Wizard concerning measurement method and dedicated calibration type, different calibrations are calculated and sorted by a specific quality attribute (Q-Value). For an optimal result, the wavelength selection, the data pretreatments and the optimal number of PCs must be chosen adequately. This is automatically done by the software.
Good qualitative calibrations can be recognised as the single spectra are found in well separated tolerance regions -Clusters- where each represent only one of the possible categories (properties).

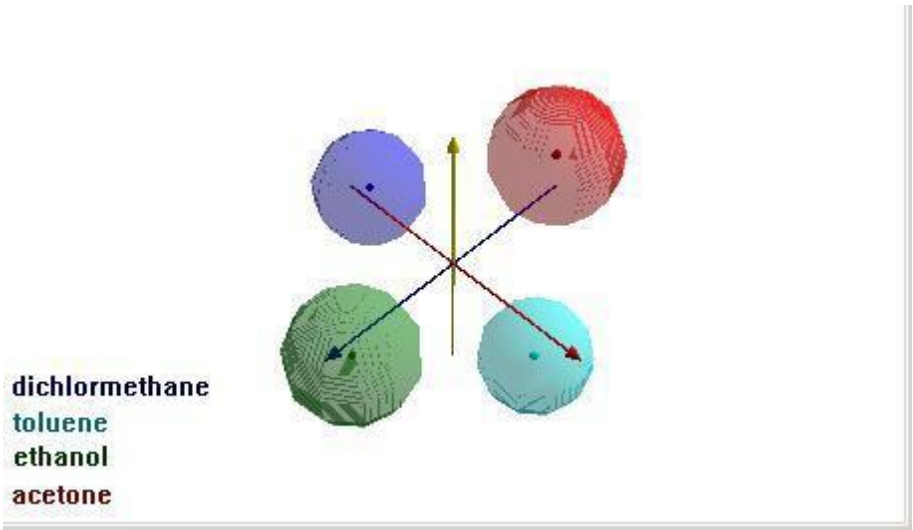



Figure 3: After transformation with the PCA, the spectra appear as points assorted in clusters, which are well separated from each other.

* In case of many different categories (properties) more than three delineation axes are required to show all important differences. It is not necessarily the case, that the first three axes are the most important ones.

**Inspection of the Qualitative Calibration**

The attributes of a good calibration are:

- all tolerance regions of the single categories (Clusters) are cohesive;
- all tolerance regions are convex and engaged consistently;
- all calibration and validation spectra are within a valid range and assigned correctly;
- all other substance spectra can not be predicted false with a certain calibration;
- Q-Value is closest possible to 1.

**Application of a qualitative calibration and interpretation of the results**

During the application of a qualitative calibration, it is determined whether a new measured spectrum can be associated with a calibrated category.
A measured spectrum is identified as OK when these three criteria are fulfilled:

- the residual is smaller than the allowed limit;
- the distance is within the allowed tolerance (the spectrum is in a known cluster),
- the identity of the measured substance (substance-ID entry) matches the identity of a substance used to build up the calibration model.

### 1.1.4 Quantitative calibration

NIR Spectra can be seen as fingerprints, which are characteristic for a certain substance. When investigating a substance mixture, different fingerprints are superposed within a complex spectrum. The concentration rate of the substances is present in the spectra, but cannot be seen. Target of a quantitative calibration is, to calculate applicable filters that establish the coherence of the measured intensities and the concentrations of the single components.

The quantitative calibration serves as the determination of parameters such as **concentration** (e.g. water content, blending ratio, hydroxyl number, etc.) or **physical properties** (e.g. density, viscosity).

**Choosing the calibration samples**

As a basis for the calibration work a set of reference data is used for which the interesting concentrations have been determined with the referring lab method. For each component (property) at least 15 samples must be used.

**Important**:

For the choice of the reference samples, **all components of the mixture** (and eventually varying parameters like temperature) **must be considered**, even components not being calibrated. For complex compositions, like foodstuffs, the requested amount of samples might be quite high.

The choice of calibration samples must be carried out very carefully. Spot test samples taken from the production process might not meet the demands, because the reference values typically show small variations around a nominal value. Suitable calibration samples should whenever possible be chosen with the aid of an experimental design or specially produced. The consistent and independent distribution has to be checked in any case.

Suggestions:

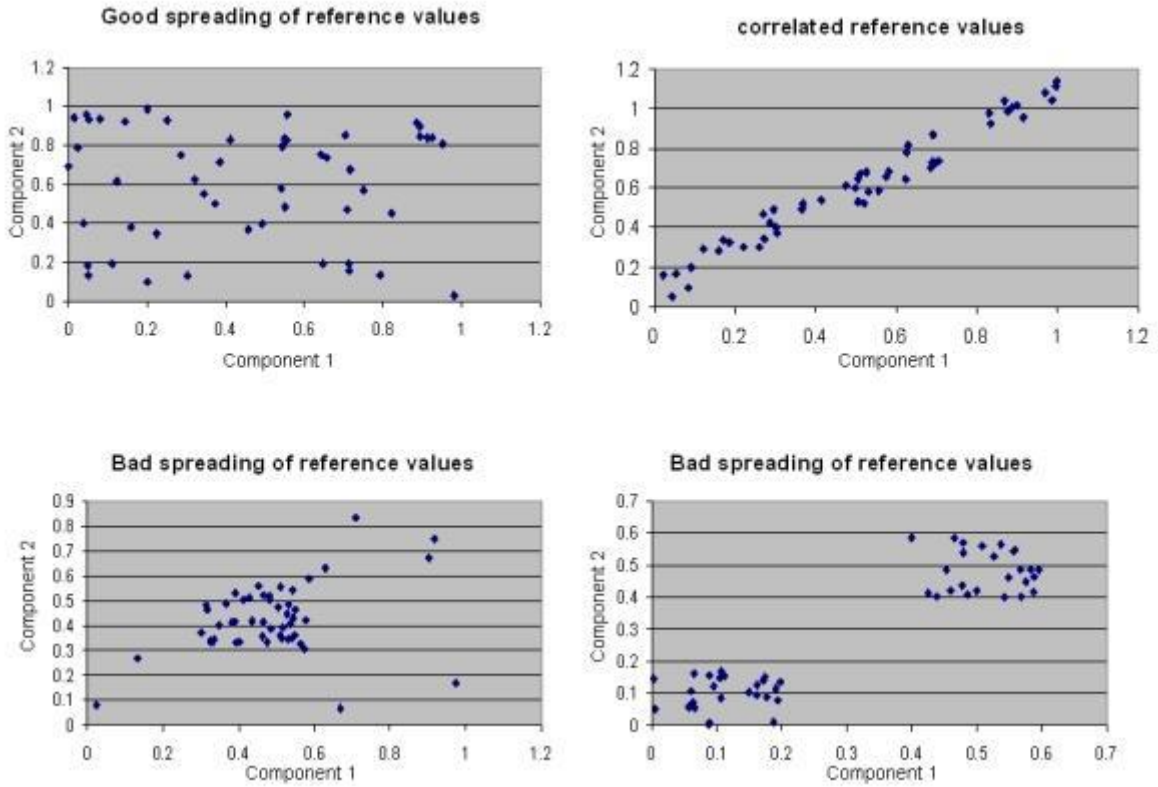


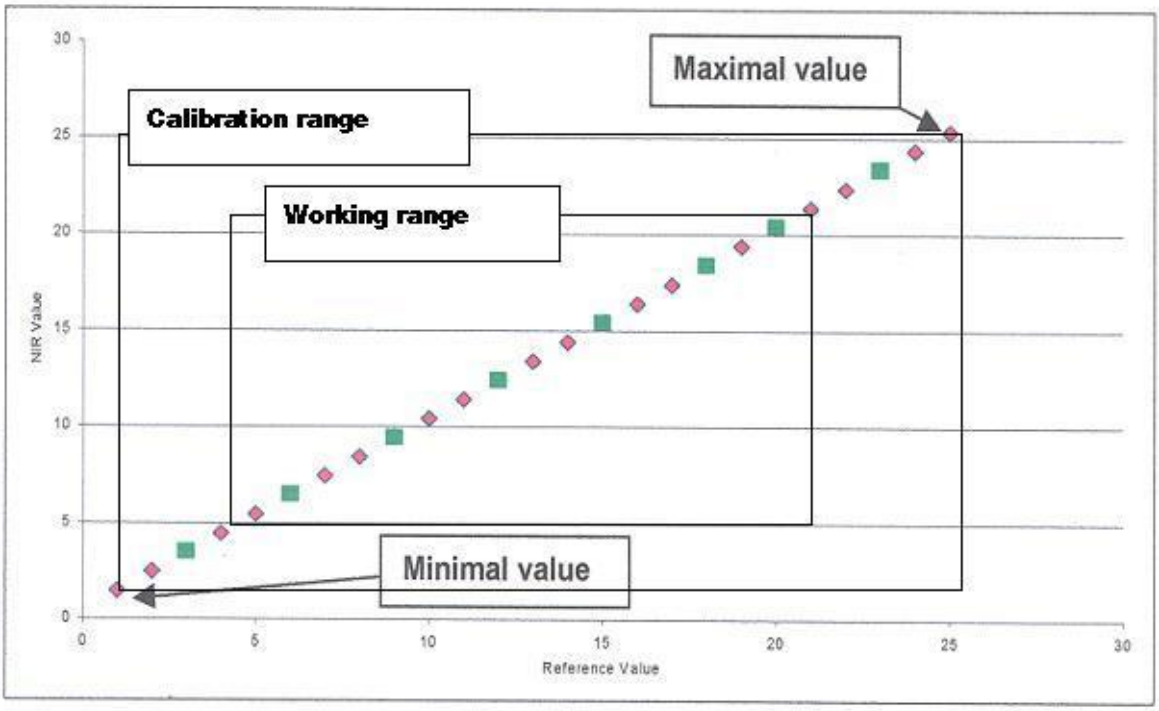

Figure 1: The distribution of the reference values is checked with the help of 2D plots.

The NIR **calibration range** should cover at least 20 times the error range of the reference method. For example for $H_2O$ determination, the range of error of ± 0.2% leads to a minimum NIR-calibration range of 4% around an expected mean value. Otherwise the error of the reference method has a high influence on a calibration. In order to avoid extrapolation for the NIR calibration, select the calibration range slightly wider than the working range.

- **Working range**: concentration range of a parameter that usually is measured in a product;
- **Calibration range**: for the calibration, the concentration range must be set broader than the working range.

As a rule of thumb; **60 samples should ideally cover the calibration range homogeneously**. 2/3 of all samples are used for calibration, 1/3 of all samples are used for validation. Spectra of the same sample must remain together and be put either in the calibration set or the validation set. For the calibration set it is recommended to use at least 10 samples per parameter to be determined.

Spectra with the minimum and maximum concentration values (extreme values) must always be selected as calibration spectra.




An incorrect distribution of the samples with two separated concentration ranges should be avoided. Either an independent calibration should be generated for each range or additional samples, covering the missing range, should be measured.

For reference measurement, the probe and reference material need to be absolutely clean. This is valid for the measurement of calibration spectra as well as for the measurement of application spectra. The measuring option and measurement conditions should be identical for generation of both, the calibration and routine application use. For quantitative measurements, the fibre optics probe should be in a fixed position.

The samples should not be measured with increasing or decreasing concentrations, but randomly distributed concentration.

The accuracy of the laboratory method has a huge influence on the quality of the NIR calibration. It is important to use the most accurate laboratory method (not the weight loss with IR quick drying for moisture determination, but the drying oven method or better the water determination by titration according to Karl Fisher).
The time between the laboratory determination and collecting the NIR spectra should be as short as possible.

**Calculating the Quantitative Calibration**

The suggested calculation methods are called **PCR** (principal component regression) and **PLS** (partial least squares regression). The algorithms mostly lead to similar results.
The choice of the relevant wavelength ranges and data pretreatments can improve the result of the calibration. With the applied data pretreatments unimportant spectral variations can be suppressed or the coherence between intensities and concentrations can be simplified. For example it is recommended to convert reflectance spectral data into absorbance data that depend directly on the concentration values (according the Beer-Lambert law, which is limited valid for liquids in NIR). The difficulties in choosing the data pretreatments are that they always have combined influence and that the effects depend on the applied order. Theoretical considerations take a back seat in the practice. Often the model can be optimised by trying different variations, that are calculated.

Concentration = Bo + B1*Amplitude1 + B2*Amplitude2 + + Bn*Amplituden [1],

These establish the coherence between the measured amplitudes and the searched concentrations. This equation is strongly simplified, a detailed explanation is shown in chapter "Calibration Methods"
Link : Principal component regression (PCR) & Partial least squares regression (PLS)

The choice of parameters and calculation methods can be done automatically by using the Calibration Wizard.

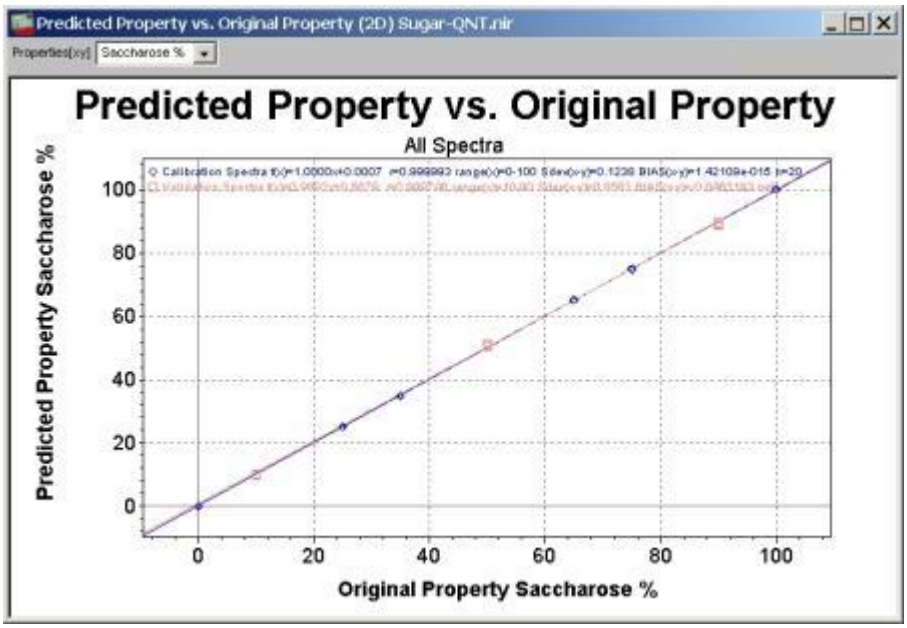



Figure 2: For the review of a quantitative calibration the prediction values according to equation [1] are compared with the reference values. In an ideal case, the corresponding points are lying on the 45° calibration curve through the zero point both for the calibration and validation samples.

**Inspection of the quantitative calibration**

| | |
|---|---|
| Precision | the SEC and SEP provide the magnitude of the standard deviation for the calibration set and the independent validation set. The two values should be **as small as possible**, but they are likely to be comparable with the **standard deviation of the conventional laboratory method**. With an acceptable calibration, the two values are also **roughly equal** (Consistency: »100) |
| Accuracy | the V-Set Bias provides information on the **average deviation** of the predicted values from the true values. This value gives information on a systematic deviation of the calibration and therefore should be as **close to zero** as possible. The C-Set Bias is zero by definition. |
| Regression coefficients, r | show how well the predicted values (NIR values) match the reference values (original property values) on average. |
| Q-Value | for optimisation, different parameters are integrated in this quality factor. The Q-Value lies between 0 ( = inoperative calibration) and 1 ( = ideal calibration) |

**Possible reasons for bad calibration results**

Outliers can be recognised for the differently calculated calibrations as big differences between predicted values and reference values appear.

Remedy: Outliers must be erased from the calibration as well as the validation set. This always must be combined with a careful clarification of the reason for the appearance of the Outlier.

To find out if a reference value or the measured spectra must be regarded as an Outlier, the score plots should be reviewed (Graphics / Scores / 2D-Scatter).

- In case spectra breaking ranks, show clearly deviating Scores and Residuals (Graphics / Spectra / Residuals), these spectra are real outlier and should be eliminated from the calibration / validation.
- In case there are big differences between the reference values and the predicted values, but the Scores of the referring spectra do not have particular deviations with high probability, the Outliers appear because of false reference value. The lab determination should be repeated.
- Group of samples with systematically deviations: this effect can be seen from time to time when samples are evaluated their reference values have been determined in laboratories not using exactly the same reference methods. Here only an alignment of the reference methods can help.
- Significantly different results depending on the chosen classification of the samples into the calibration and validation set: the number of used samples is too small, for instance because of not considered, hidden properties.

Remedy: Selective completion of the master data set that all possible variations flow into the calculation.

**Application of a quantitative calibration and interpretation of the results**

During the application of a quantitative calibration, the referring concentration values of the measured spectrum are calculated according to equation [1] and indicated.
These calculations always lead to a result, even when performing faulty measurements. (e.g. sample of completely different, not calibrated substance class). The calculation of the concentration is therefore enlarged with two further checks:

- the concentration values found must be within the original calibration range, a warning- and action-**limit** can be adjusted as well.
- make sure that the new measured spectrum matches the calibration spectra and that the **Residual** is within the allowed limit. A warning appears, when the residual of the measured spectra is bigger than the maximum allowed residual. Results of the spectra with Residual outlier are not taken for the average calculation of multiple measurements.

## 1.2 Introduction to NIRCal

### 1.2.1 General

The work flow for typical NIR Application with Büchi NIRWare is:

1. **Create an Application** in NIRWare Application Designer **for reference measurements**

2. **Data Acquisition** Spectra collection with NIRWare Operator
   NIR- Spectroscopy

3. **Assign Property** value to spectra with NIRWare Sample Manager
   Spectra + Reference Analysis

4. **Calibration** build mathematical models with **NIRCal 5**
   Correlation known Spectra <- -> Concentrations / Identity

5. Copy and modify the **Application for routine measurements**

6. **Assign** the created **calibration** to the routine application

7. Use the mathematical models/calibrations within an **Application** in the NIRWare Operator for **Predictions**
   Unknown Spectra -> Concentrations / Identity

In case a calibration after the first trial is not delivering the correct results, the steps 3-7 should be repeated, the calibration should be expanded with new samples and tested again.

All data (like spectra, property value, projects, calibrations, etc) are saved in the **NIRWare Database**.

***NOTE***
To be able to take full profit from all the possibilities of the NIRCal chemometric software, a **Buchi training** is highly recommended.

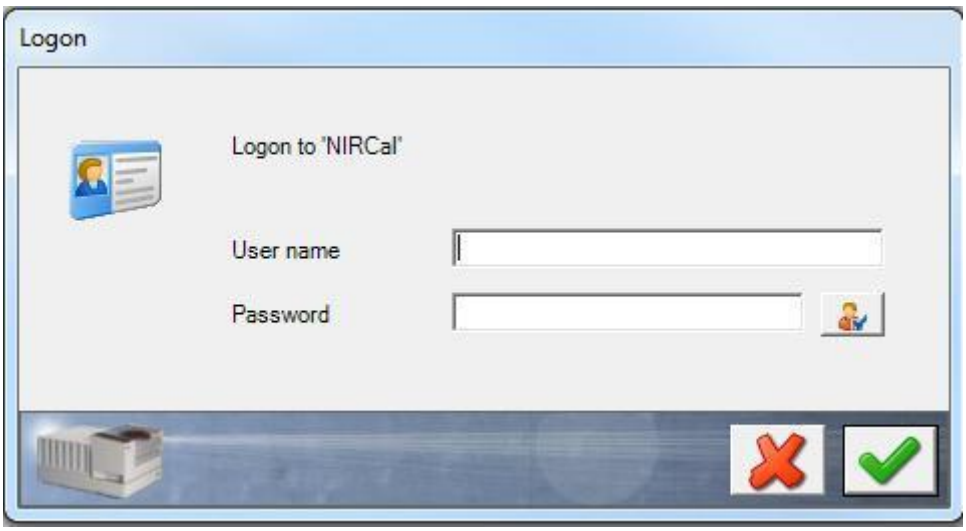
### 1.2.2 Starting NIRCal

Windows 7: Start / All Programs / BUCHI / NIRSolutions /NIRCal

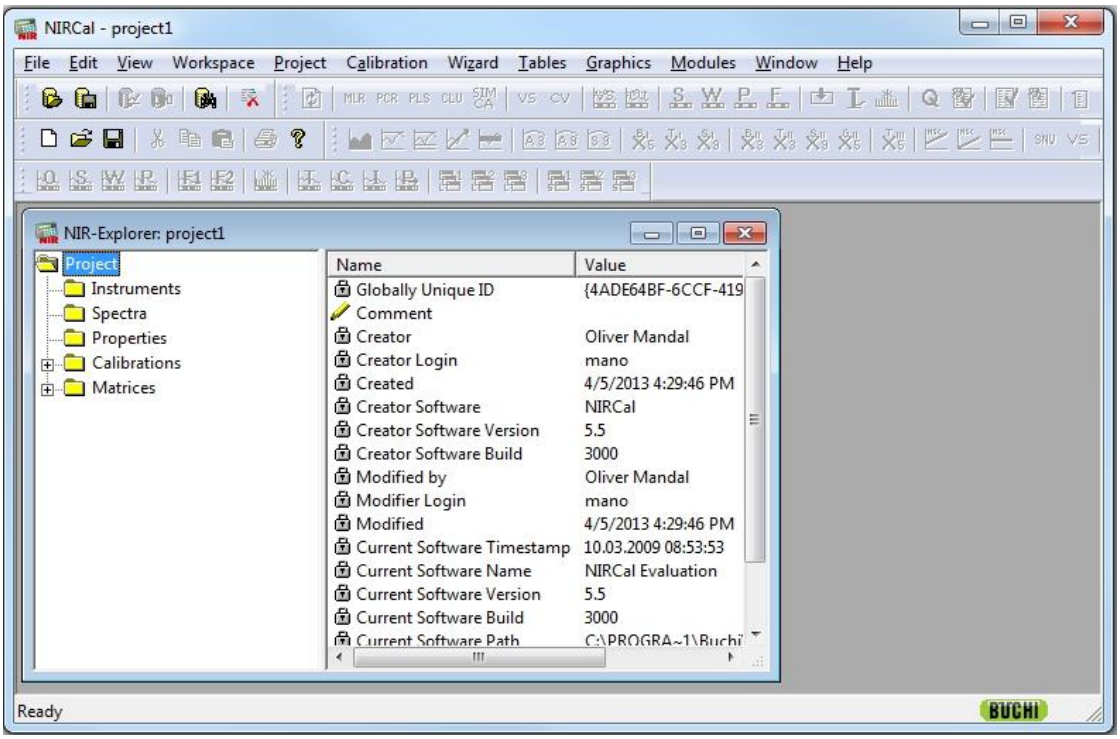
Windows 10: Write “NIRCal” in the main screen and navigate to

programs Icon: 

To use NIRCal Log On with User name and Password.




User rights and setup is managed with the NIRWare Suite "Security Designer". Please refer to the corresponding documentation within the SW-Manual for NIRWare.




NIRCal opens an empty project.

### 1.2.3 NIRCal Project

NIRCal stores the imported spectra and their properties , which belong together in a **project**.

A project contains:

- **spectra:** the imported spectra as they were measured;
- **properties**, the property names with their values, which belongs to the loaded spectra and can not be modified in NIRCal;
- the **calibrations**, which have the information about the used data selections and the calculated results;
- **matrices**, which have all chemometric results of the active calibration.

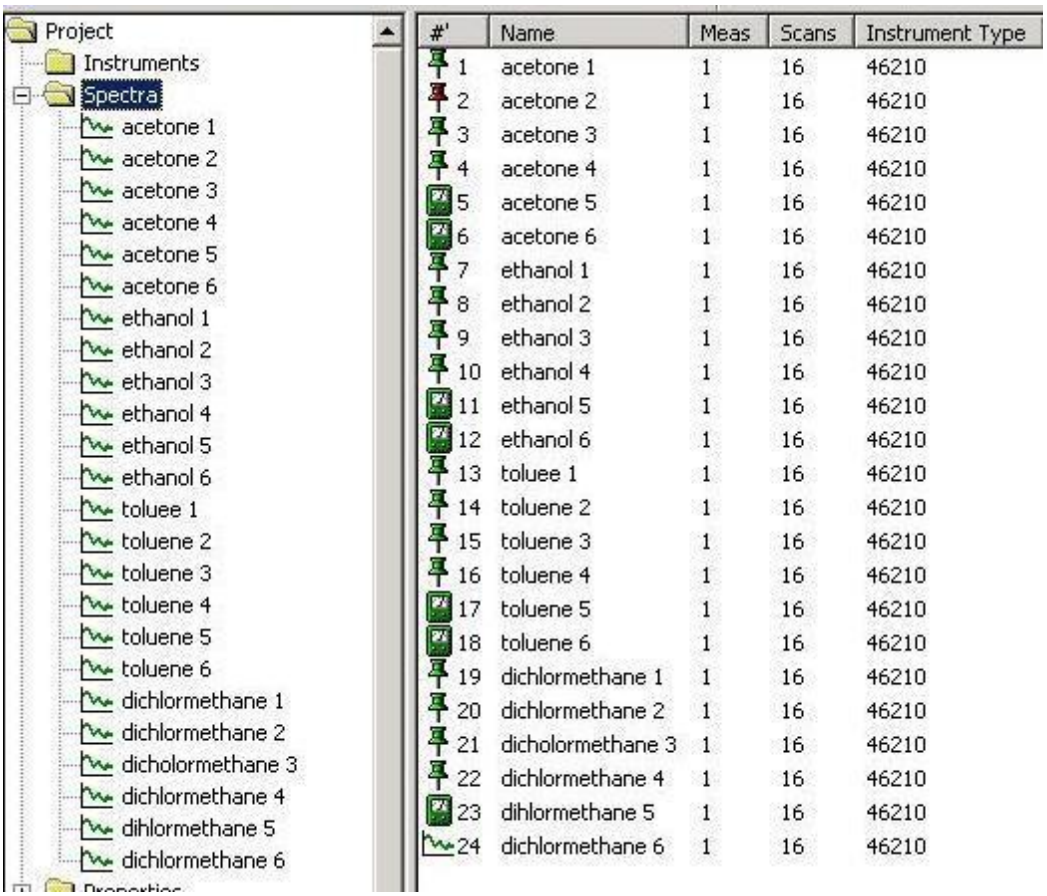

Note: Instruments is not used in NIRCal 5 anymore.

Related Topics for Spectra import:
Search and Import Spectra from NIRWare Database
Import Spectra from File
Convert and Import spectra from other instruments to DB
Use NIRCal for any type of Data

***NOTE***
It is possible to import and export spectra between NIRWare databases within the Software "**NIRWare Management Console - Administrative Tools**"

### 1.2.4 Spectra Overview

All spectra measured or imported into this project can be seen here.
Items marked with a yellow pen can be edited by pressing F2 or a double click.




Explanation of the symbols (right window):

Spectra belonging to the C-Set

Spectra belonging to the V-Set

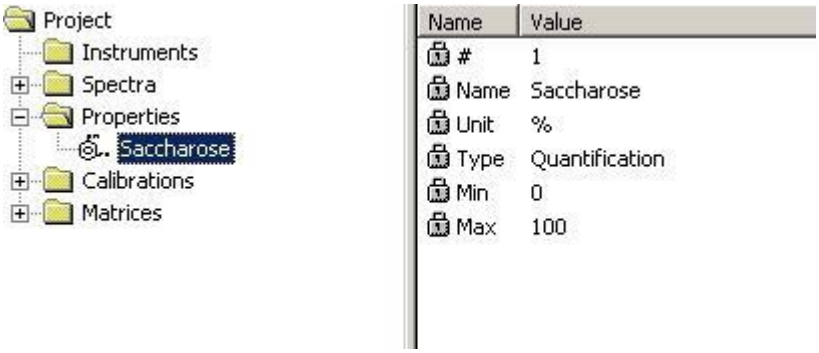
Spectra in unused Set = U-Set (these spectra are not used for calibration nor for the validation of a calibration

Spectra in C- and V-Set at the same time. Overlapping is not allowed.

Red color indicates selection. These spectra for example can be assigned to the user set and be plotted separately.

### 1.2.5 Properties Overview

All properties in this project can be seen here.
Items marked with a yellow pen can be edited by pressing F2 or a double click.



Min/Max represents the calibration range for quantitative properties. For qualitative properties Min/Max is shown as 0/1.

## 1.2.6 Calibration Validation Methods

To be able to judge the performance of a calibration a set of independent validation samples is necessary.

**Validation Set (VS)**

Normally all spectra within a project are divided into 2 sets with a suggested ratio of 2/3 to 1/3. The two sets should be completely independent from each other.

- C-Set (Calibration Set)
- V-Set (Validation Set)

Spectra in the V-Set are not used for the calibration, the V-Set spectra are used like unknown samples to judge the quality of the calibration (internal validation set). Only the C-Set spectra are involved in the loading calculation.
Enough spectra of the sample should be available.

**VS can be used for all calibration methods.**

**Cross Validation (CV)**

Cross validation (CV) uses all samples as the calibration set for **quantitative calibrations** except one sample (or a small group of samples) which is left out.
Validation is accomplished by predicting the left out samples and by systematically varying the selection of left out samples. The procedure is time consuming because for each selection a calibration has to be calculated. The method is especially useful when the total number of samples is small (< 50 samples).
Full cross validation (FCV) means that n-calibrations are calculated so that one spectrum has been left out and all other are in a calibration.

Limitations

- only available for **PCR** and **PLS**;
- needs at least 2 CV groups or at least 4 C-Set spectra for one-leave-out (full cross validation; FCV);
- will delete the V-Set spectra selection, in case it is not empty.

## 1.2.7 Calibration Methods

**Qualitative Calibrations / Identification**

Target is to identify the **membership** of a sample to a property group. The property groups can be chemically completely different or similar to the same substance.
Both implemented method are using PCA:

- Cluster Analysis CLU
- SIMCA

**Quantitative Calibrations**

Target is to determine the **concentration value** such as content in %, OH-value or **physical parameters** like density, viscosity.
In NIRCal implemented algorithms are:

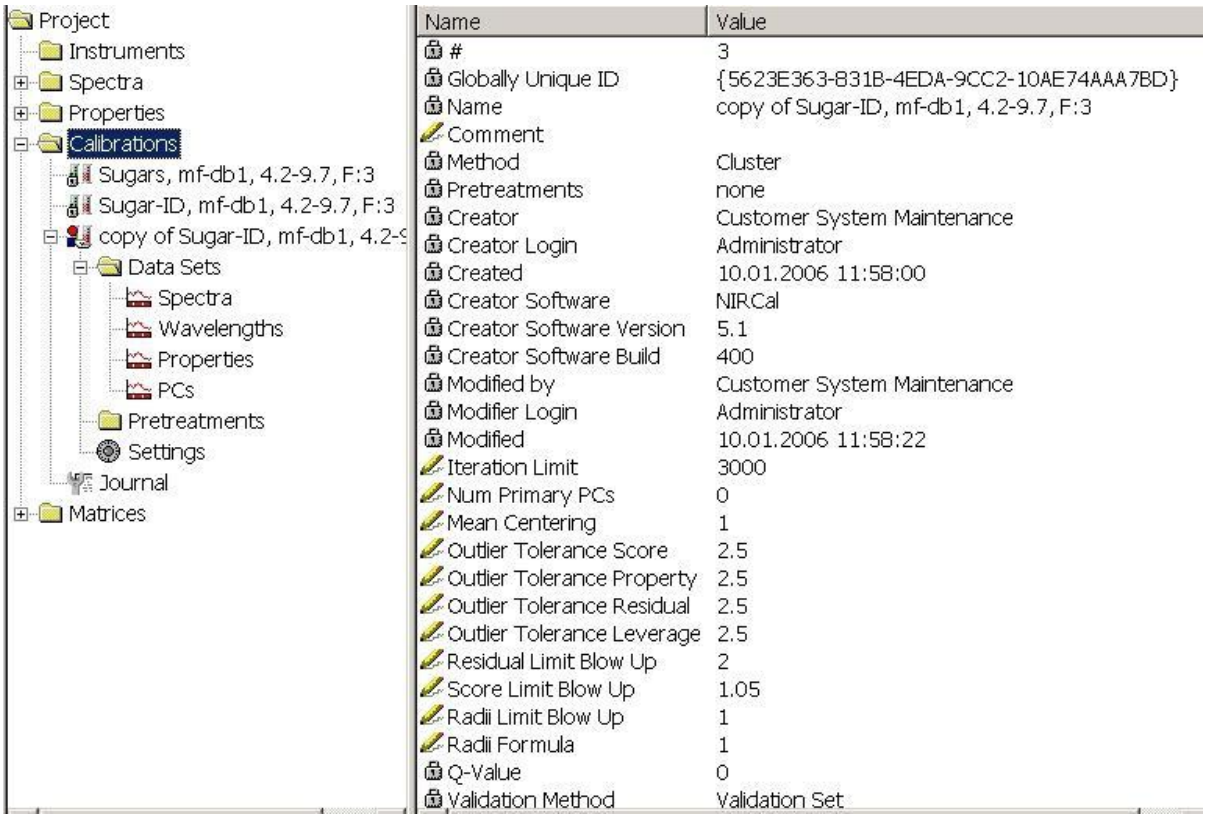
- Principal component regression PCR
- Partial least squares regression PLS
- Multiple linear regression MLR

PCR, PLS, PCA (CLU) and SIMCA are principal components based methods, MLR and also **library search** with spectra comparison are spectra based methods.

- Calibration — Inputs are the measured spectra with their properties and with the reference values together with the calibration data selection.
  Outputs are the calibration for the properties together with the validation for the validation set (V-Set) spectra or the result of Cross Validation (CV).
- Application — Inputs are the calibration and a measured spectrum.
  Outputs are the predicted property value(s), calibration limits or hitlist; spectral residuals.

### 1.2.8 Calibrations Overview

All calibrations calculated in this project can be seen here.
Items marked with a yellow pen can be edited by pressing F2 or a double click.



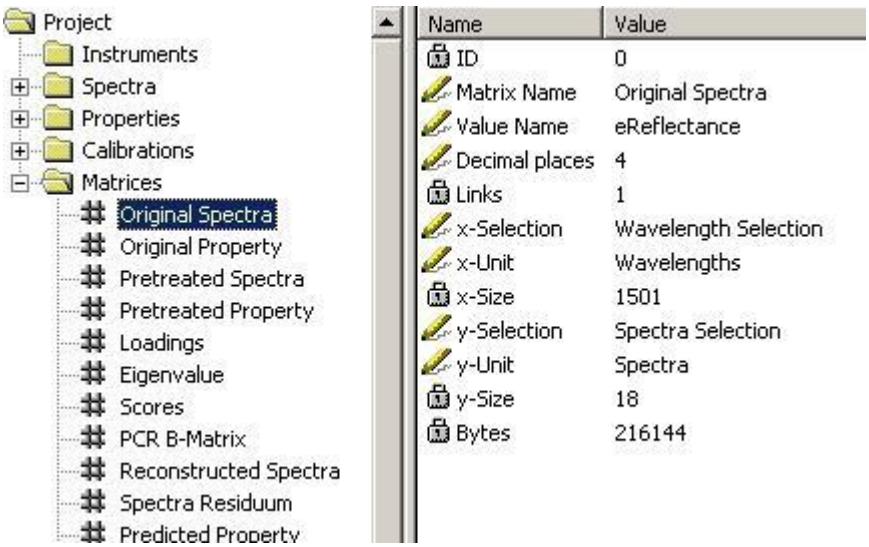

New (active) calibration, not yet stored in the database

Stored calibration not in “Edit” mode (availability for “Edit” mode depends on Lifecycle status)

Stored calibration in “Edit” mode

### 1.2.9 Matrices Overview

The matrices shown here depend on the selected calibration and validation method.
Items marked with a yellow pen can be edited by pressing F2 or a double click.




### 1.2.10 Data

Raw data like **spectra** and **property values** are loaded from the NIRWare Database.
Spectra and property values, loaded from the database, are **not modifiable** in NIRCal anymore.
Pretreatments never change the original spectra.

These data are stored in a NIRCal **Project**. Each project has a first calibration, which has an organiser function and this can not be deleted. Calibration specific data like selections (Spectra, Wavelength, Pretreatments,Principal Components) are stored for each **calibration** separately together with the selected chemometric methods calculated data. Approved calibrations, which are released for the **application** according the Lifecycle procedure, can be used in the NIRWare Operator.
Link: Calibration Handling according to the Lifecycle.

The main and intermediate data are stored in **matrices** (e.g. Original Spectra - and Original Property matrices), which are available in table or graphic form. Intermediate data e.g. Scores and Loadings are calculated on demand by the active calibration settings.
Matrices can be exported to MATLAB as .dat files or to EXCEL as xlm files.

- MATLAB : NIR-Explorer / Matrices / “Loadings” / PopUp Menu / Export.
- EXCEL: NIR-Explorer / Matrices / “Loadings” / PopUp Menu / Table. In the Table PopUp Menu / Export Table…

Link: Matrices

NIRCal 5 can work as **File- and as Database - oriented software**.
With NIRCal it is possible to **import** data from files and to store data to **files**. NIRCal 5.4 downward compatibility of spectra and properties is supported to NIRCal 5.2 as .nir, .ncf and JCAMP-DX file. In NIRCal 5.2 the NIRCal 5.4 files can be imported with: File/ Import/ Spectra.
The loaded spectra can be also stored in the **database**. A spectra conversion from the older NIRCal 4 data or from other software is possible in this way. The following data types can be imported:

- Projects in format ".nir";
- Spectra with properties in format ".nsf, .bmp, .dat, .csv, .spc, .dx, .jdx, .jcm, .nir "
- BCAP-Series in format ".S??" (Spectra file)

It is not possible to import data from NIRCal 4 ".ncf" format, because ncf-files do not include spectral data.
It is important to take care of the **compatibility** of the spectra from different instruments!

Link: Convert and Import..in NIRWare DB.

### 1.2.11 Selections

**Data selections** are needed to group the data before the calculation of a calibration.
For each calibration the selections can be different.

#### Calibration method selection

- Choose the method according to the target of the application.

For quantitative applications there are MLR, PCR or PLS and for qualitative application Cluster or SIMCA are available.

#### Validation method selection

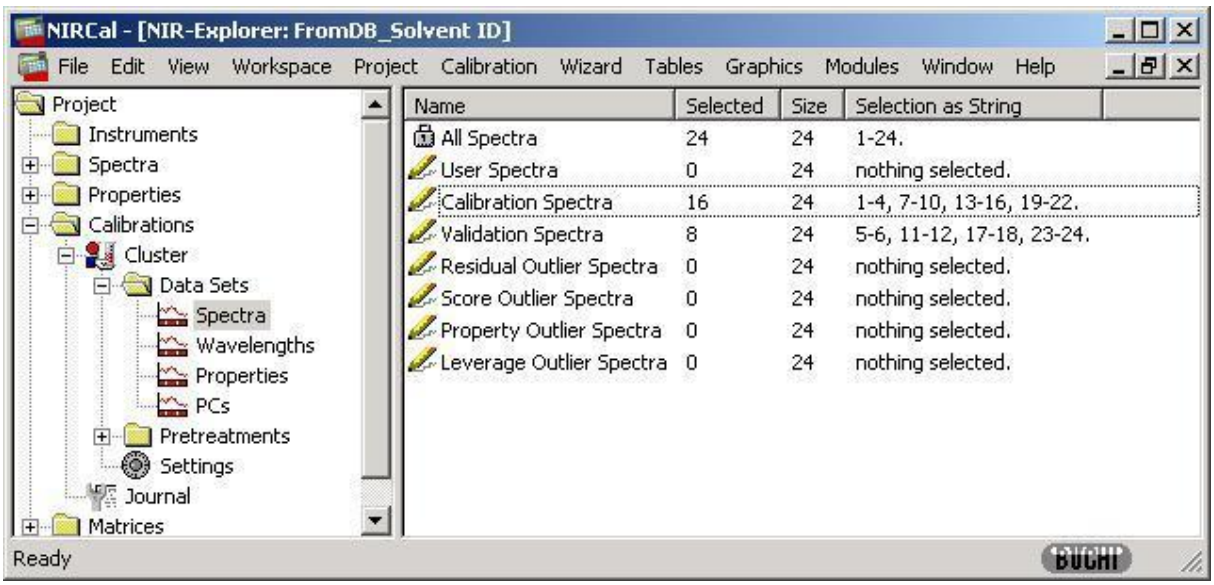
To test the validity of a calibration **internal test** samples are used in the validation mode: **VS**. For quantitative calibration the Cross Validation: **CV** is available.

The default method is VS: Validation Set, where the V-Set spectra should be selected by the user. You can also use the spectra selection wizard.

#### Data Sets Selection

***Spectra Selection***

The selection can be made in NIR-Explorer:

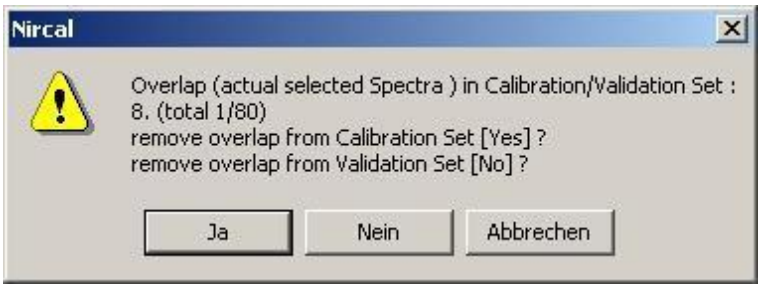



| | |
|---|---|
| All Spectra | not modifiable, determined through the number of imported spectra. |
| User Spectra | mainly used for visual comparison, any selection is acceptable, this selection has no correlation with the C-, V- or U-Set spectra. |
| Calibration Spectra | spectra selected for the calibration: C-Set, about 2/3 of all spectra for VS mode. |
| Validation Spectra | spectra selected for the validation: V-Set (1/3 for VS). |
| Outlier Spectra | result of the "Outlier Detection" wizard for the statistical residual, score, property value and leverage outliers in C- and / or V -Set |

Each of these sets, except of "All Spectra", can be edited by the user.

Select spectra for the calibration (Calibration Spectra = C-Set) and for the validation of the calibration (Validation Spectra = V-Set). The C-Set will be used for the calculation of the calibration, the V-Set is used as an internal test of the calibration.

The same spectra should not be defined as C-Set and V-Set Spectra at the same time. If so then the spectra will be marked with a question symbol in the spectra overview and following message appears, that easily allows to remove the spectra set overlap automatically:




Link: Chemometrics / Selections / Spectra Data Set

**Wavelengths selection:**

- All Wavelengths: the measured wavelength/ wavenumber range and the number of wavelength / wavenumber data points, which is determined by the used instrument and by the resolution;

- Calibration Wavelengths: only the selected range will be taken into account for the calculation and later for the application.

Link: Chemometrics / Selections / Wavelength Data Set

**Properties selection:**

- All Properties: number of properties belonging to the imported spectra;

- Calibration Properties: only the selected properties will be taken into account for the calculation and later for the application. Normally all or **several** will be selected for **Cluster** calibration, only **one** should be selected for **quantitative** and for **SIMCA** calibration method. Link: Chemometrics / Selections / Properties Data Set

**Principal Components Selection:**

- All PCs: the number of **primary principal components** used for the calibration, they are used for the spectra reconstruction and for the Residual calculation;

- Calibration PCs: only the **selected PCs** will be taken into account for the calculation and later for the application, they are used in quantitative calibration for the **property value** calculation or in qualitative calibration for the property separation and for the **allowed tolerance area** calculation.

Link: Chemometrics / Primary Principal Components & Secondary Principal Components

**Pretreatments selection:**

The measured spectra can be modified before the calculation with several pretreatments or different combinations of them.

***NOTE***
The original spectra are still the same and will not be modified.

The list of in NIRCal 5 implemented pretreatments indicates the huge flexibility of the software. The pretreatments used in the calibration will be applied in the application also.

**Blow Up Parameter:**

The allowable limits are calculated with a default blow up limit. These limits can be adjusted by the user:

- for quantitative calibrations the scores and residual blow up limits;
- for qualitative calibrations the scores, residual and radii blow up limits and the radii calculation formula.

Link: Chemometrics / Blow Up Limits

**Outlier detection limit**

To calculate the statistical outliers the limit can be edited by the user or can be calculated according the T-distribution for the actual calculation.
Link: Chemometrics / Outlier Detection

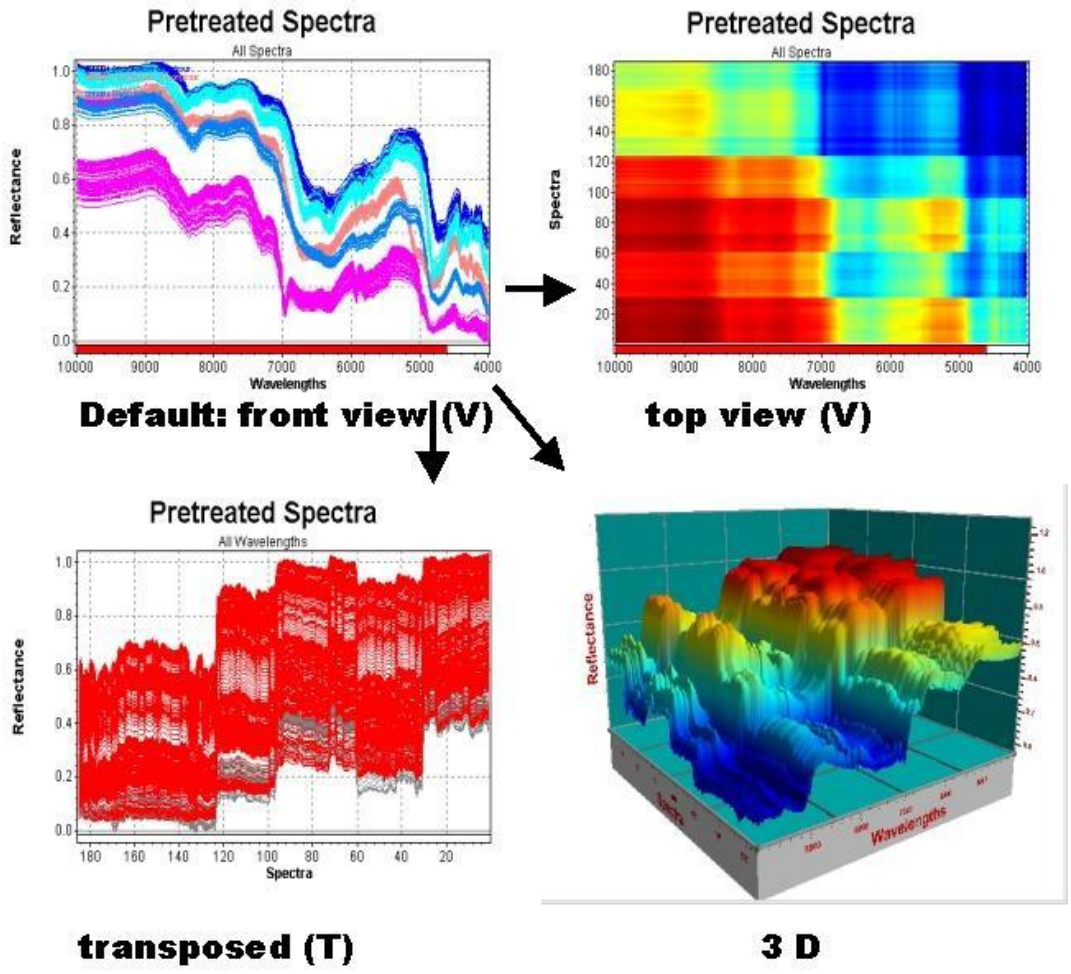
## 1.2.12 Graphics

All data that is available in table format can be plotted graphically as well.

In NIRCal 5 there are 1, 2 and 3 dimensional plots available.

The 1D scatter plot is new in NIRCal 5: it shows only one selected line or column of a matrix. This scatter plot can be used for selections.
The pop-up menu "Options/Show line" or the keyboard shortcut "W " can change the point to a line.

Each 2D plot can be shown as a **front view** (this is normally the default adjustment), **top view** and **transposed**.

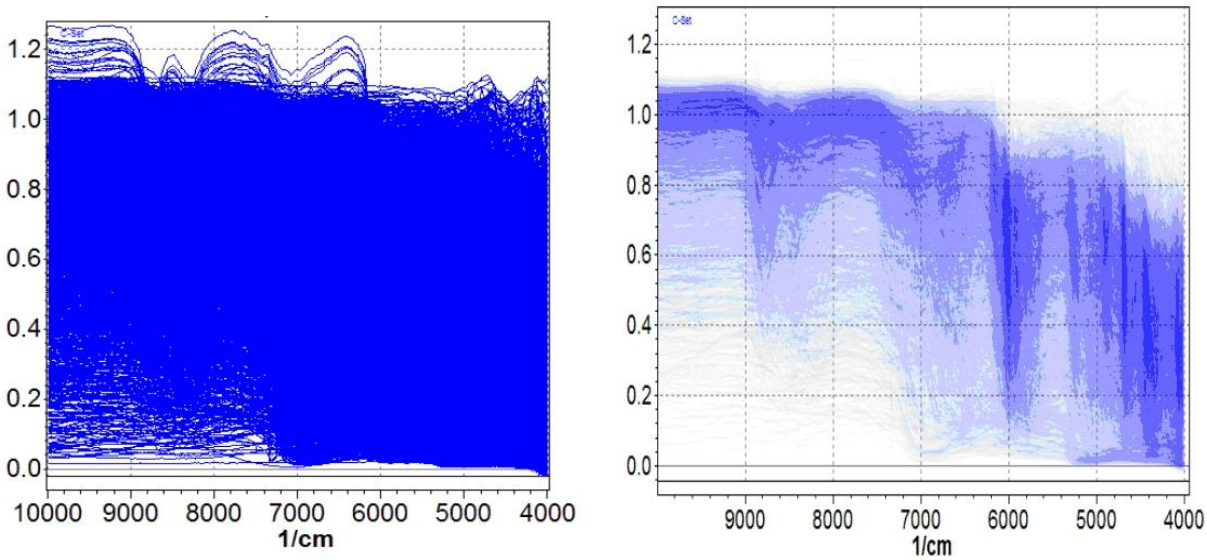


The plots can show **different selections**, while the **selection** plotted is always visible on the **subtitle**.

The use of modern rendering techniques, such as Anti-Aliasing and Alpha-Blending (new hardware accelerated features of newer graphic cards), allows to eliminate jagged edges and stair-stepped lines. NIRCal applies these features to improve the graphical performance.

Anti-Aliasing (hotkey: a) eliminates jagged edges and stair-stepped lines. This can be switched on permanently under: Edit / Options / 2 D Plot. Opening the 2 D plots is quicker with Anti-Aliasing using new graphic cards, but can be slower with an older driver.

Alpha-Blending (hotkey: b) allows a partial transparency with steps of 1, 1/2, 1/4, 1/8. 1/32, 1/64 (by pressing b several times) and helps to look through huge amount of data. The Alpha-Blending level can only be seen from the data color intensity: the more often the point is drawn, the brighter the color. Each NIRCal rendering can be stopped immediately by pressing the ESC key.
Anti-Aliasing and Alpha-Blending are very useful for huge amount of spectra especially for 2-D spectra plots (original, pretreated and residuum); Score plots (vs. PC or vs. Scores) and calibration curve (original vs. Predicted).




On the left side the common drawing style is used, where no details can be seen. Anti-aliasing and alpha-blending (right side) allow to look through a huge amount of data and to detect hidden structures.

**Note: In case a one dimensional matrix - which is a vector- is plotted, the subtitel is "All One". "One" indicates the first selected "dimension" and not the value of the vector!**

The colors are also selection dependent.
The plots can be opened as Table with the keyboard shortcut "G".

On each plot the X- and Y-axis can be flipped using the pop-up menu: Option/ Flip X-Axis (x) and Option/ Flip Y-Axis (y).

**2D combined scatter plots**: beside the standard 2D plots it is possible to combine matrices, even if they have only with one dimension matching. This can be created manually in the NIR-Explorer selecting the two matrices and opening in pop-up menu the "2D-Combined Scatter".
There are several plot combinations available in NIRCal 5: under: View.

These plot combinations, which are fixed, can help to make the data selections easier (spectra, wavelengths, properties, PCs) with useful graphics.
The **dependency plots** can help to find hidden correlation between spectra and different users, or time, or instruments.

The users can create own plot combinations and can save them under: **Workspace**.

### 1.2.13 Protocols

The term "protocol" in NIRCal means **report**.

#### Calibration Protocol

The **calibration protocol** is an important validation document, which contains all information about a calibration, like the user specific data selection and the results of the chemometrical calculations applied for the C- and V-Set spectra in the project. The calibration protocol is stored within the calibration.
Link: **Calibration protocol**

#### Validation Protocol

The **prediction protocol** is an in important validation document, which contains the results of a prediction of a calibration applied for other spectra, mainly not existing in the project. The prediction protocols help to find out possible interfering substances for the qualitative calibrations.
Link: **Prediction protocol**

# 2 Tutorials

## 2.1 Qualitative

### 2.1.1 Qualitative Tutorial

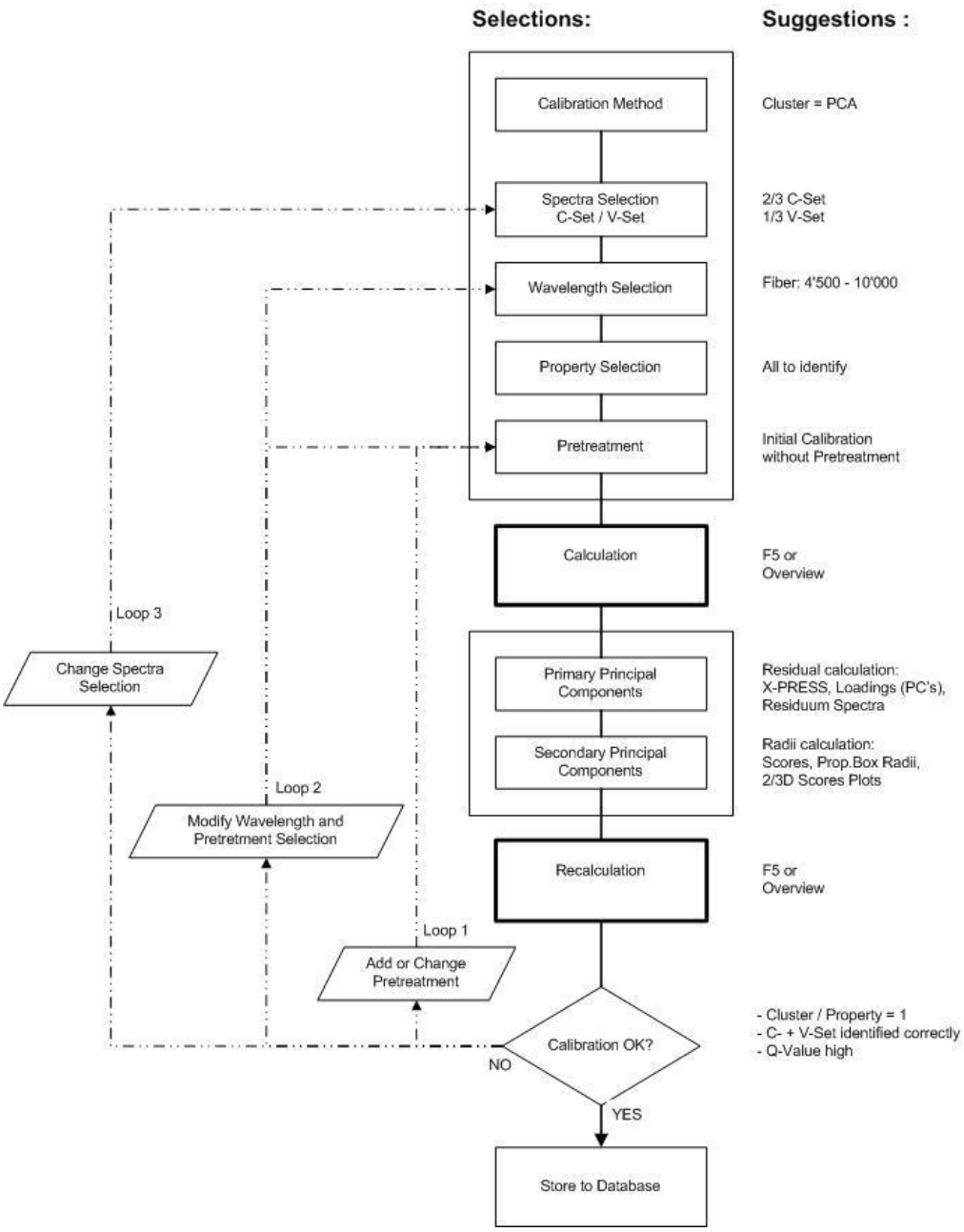
Qualitative calibrations are used for **identification** of different chemical substances and for separation of different qualities of the same substances.
Either **Cluster** or **SIMCA** method can be used for identification, both using Principal Component Analysis PCA.

The Cluster method is explained first.

For detailed explanation to the method see Link : Cluster (CLU) and Link : SIMCA

### 2.1.2 Flow Chart Qualitative

The flow chart shows the way how to build a qualitative calibration using the Cluster method.



***NOTE***
Loop 1 to Loop 3 represent the sequence for optimising a calibration. **After each change in the selection, the calculation and principal component selection should be repeated.**

### 2.1.3 Selecting the Calibration Method

The calculation method can be selected in the Menubar: [Calibration / Method / CLU Cluster] or by clicking the icon 


### 2.1.4 Selecting the Calibration and Validation Spectra

Samples of known characteristics, both chemical and physical are used to generate qualitative calibrations. For each substance classes 5-15 different batches should be used for the calibration. The measured spectra are divided into two sets:

- Calibration Spectra: spectra selected for the calibration: C-Set, about 2/3 of all measured spectra, at least 3 spectra.
  Generally, the calibration spectra should contain all "**acceptable extreme information**" to define the limits of acceptance.

- Validation Spectra: about 1/3 of all measured spectra selected for the validation V-Set at least 2 spectra. Only if a calibration treats the validation spectra equal to the calibration spectra, the settings can be considered as OK.

***NOTE***
These two groups of spectra should be:
- independent from each other;
- no overlapping allowed.

**Suggested selection**: Blockwise: 6-3, if always 3 spectra per batch where measured. In case some extreme samples are in the V-Set, see Loop 3.

It is **possible to leave out** some spectra from the C- and from the V-Set, these spectra are in the **Unused Set** = U-Set:
C-Set + V-Set + U-Set = All Spectra

The U-Set is visible in the calibration results (see Calibration Protocol), but will not influence the calibration and validation results e.g. will not be considered for the Q-Value calculation.

Link: Spectra selection

### 2.1.5 Selecting the Calibration Wavelengths

The exact wavelength / wavenumber range measured is dictated by the instrument type used for the spectrum measurement. The selected wavelength / wavenumber range depends on the application and the measuring option used.

***NOTE***
In general the calibration wavelength / wavenumber range should be as wide as possible.

Suggested wavenumber range for NIRFlex N-500 with measuring options

- Solids and Liquids: 4'000-10'000 $cm^{-1}$;
- Fiber Optics: 4'500-10'000 $cm^{-1}$.
- Solids with Tablet Accessory: 6'000-11'520 $cm^{-1}$.

The calibration wavelengths define the spectrum range, where the mathematical algorithms PCA is applied.

**Suggested selection**: use all measured wavenumber in the first calculation, otherwise see Loop 2.

Link Wavelength selection

### 2.1.6 Selecting the Calibration Properties

The calibration properties are the substances to be identified in the application. The mathematical algorithm of PCA will be applied for the selected properties. Normally **all** qualitative properties that must appear in the calibration model must be **selected**.

**Suggested selection**: select all properties.

Link Calibration property selection

### 2.1.7 Applying Data Pretreatments

Data pretreatments are used to eliminate non important effects or to enlarge minor effects of the measurements.

**Suggested selection**: perform the **first calculation without pretreatment**, using the spectra as they have been measured. The first step **to optimize** the calibration is to **add** and **change** the pretreatments.

Link: Pretreatment selection

### 2.1.8 Performing a calibration (calculation)

After selecting the **spectra** (C-Set and V-Set spectra), the **wavelength** range (calibration wavelength) and the **properties** (calibration properties), the chemometric parameters (principal components, scores, residuals,etc) will be calculated.

For the calculation of the PCA the software use only the spectra selected into the C-Set referring to the calibration wavelength and the choice of the data pretreatments. The spectra of the V-Set will only be used to prove and judge the calibration

An overview of the calculation can be obtained in the Menubar: **View / Overview.**

The result of the calibration will depend on the selected number of **primary and secondary principal components**, the first step is to decide on the number of PCs.

The calibration can be optimised **manually** or using the Calibration Wizard. Here the manual selection criteria are described.

### 2.1.9 Primary Principal Components

Primary principal components are the PCs, which are used for reconstruction of the spectra. They determine the residual value. The more primary PCs used, the smaller the allowed residual of the calibration.

Recommended graphics for primary PCs selection:

- X-PRESS function,
- Loadings graphic;
- Spectra Residuum plots.

**Suggested selection**: avoid overfitting, do not use too many primary PCs. Link
Primary PCs

Adjust the desired number of primary PCs.
Link: Adjusting Primary PCs

## 2.1.10 Secondary Principal Components

The secondary or calibration PCs are used for the **separation of the different substances** and are responsible for the **tolerance radii calculation**. The number of secondary PCs is limited by the number of primary PCs.

Recommended graphics for the selection:

- 2- and 3-D Scores plots,
- Scores against Spectra and Property Box Radii.

These plots are part of the Overview-Plot.

**Suggested selection**: use just as many secondary PCs as necessary to get a selective calibration. Link: Secondary PCs

Adjust the desired number of secondary PCs.
Link: Adjusting Secondary PCs

**Setting the Radii Formula and Radii Blow up Limit**

Calculation of the tolerances can be made with two different formulas and with different blow up limits.

***NOTE***
It may be necessary to **reduce the blow up limit for different chemical substances, when Formula 1 is used** in order to increase the sensitivity of the calibration.

Edit the value for Radii Blow Up Limits or Formula.

***NOTE***
After each new selection a recalculation should be performed again.

## 2.1.11 Judging the Calibration

In a good calibration, the different properties should appear as a **separated and ideally compact cluster**. The proper clusters are visible in the Overview.

For judging the quality of calibration the following criteria are used:

- **Cluster per Property**: should be **one**, so only one cluster for each property:
- **Spectra Residuals too big: zero** (no residual outlier);
- **Property Residuum**: should be **zero** which means, that all spectra are in the right cluster;

The 2nd column of the Overview contains important control windows.

- Q-Value: should be as close to 1 as possible;
- Calibration Protocol: show the important adjustments and result of the calibration
- Prediction Protocol show possible wrong identification for substance spectra not involved in the calibration

After several calibration **optimizations and running the automatic calibration wizard**, there will be several calibrations in the project. Sort the calibrations with a click on the Name or Q-Value.

***NOTE***
Keep only the best calibration and delete all others.

## 2.1.12 Save Calibration - Lifecycle

Save the project and the calibrations to the database. Only approved calibrations according to Lifecycle will be available in the NIRWare Application Designer. See Calibration handling for further details.

### 2.1.13 SIMCA

It is possible to transfer a cluster calibration into several SIMCA calibrations. See Transform Cluster to SIMCA.
This tool creates for all C-Set properties one SIMCA calibration, the name of the calibration is "SIMCA + calibrated property name".

## 2.2 Quantitative

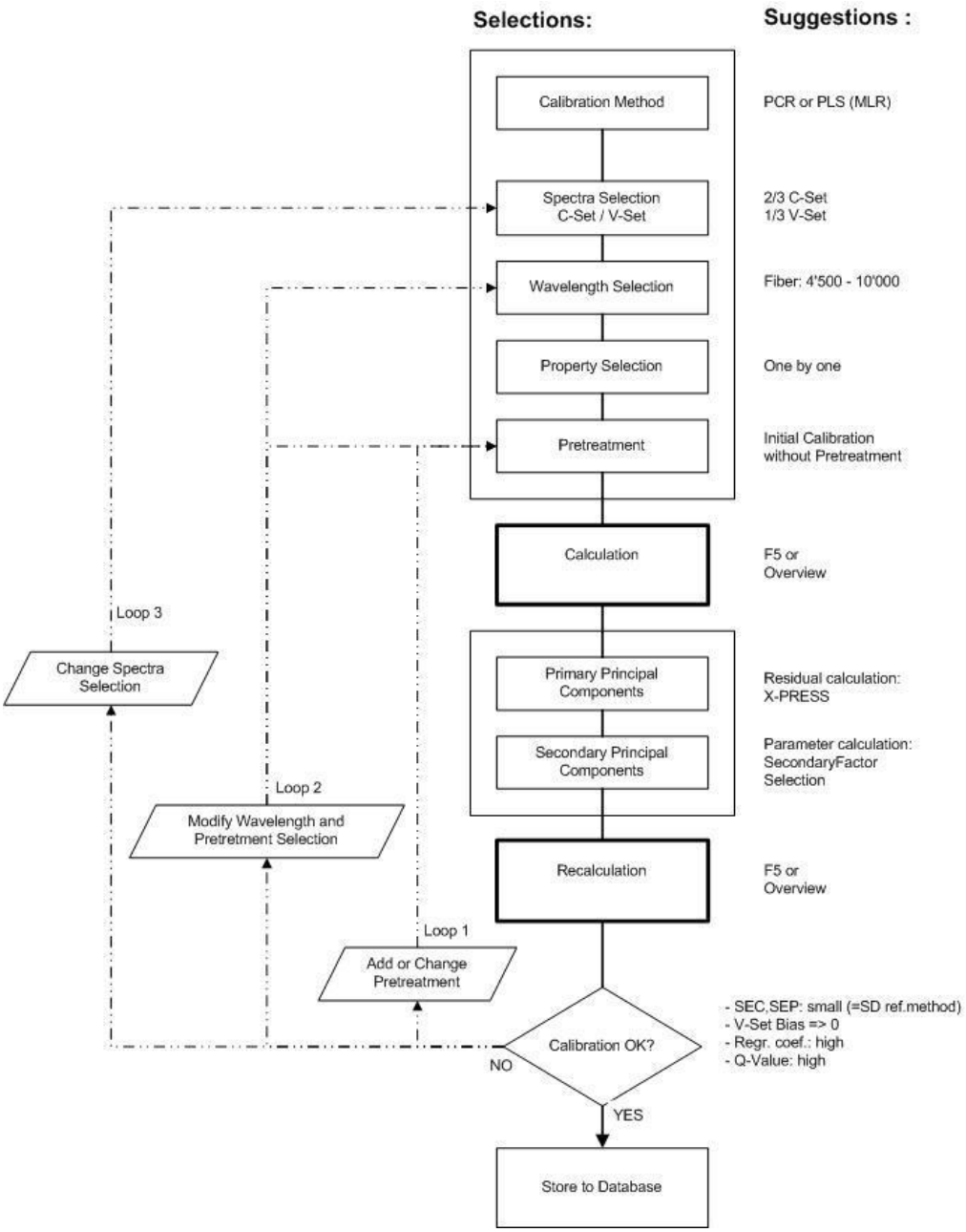
### 2.2.1 Quantitative Tutorial

Quantitative calibrations are used for the **determination** of different concentrations or physical parameters.
For quantitative calibration either the the **PCR** or **PLS** method can be used, less useful is the MLR method.

For detailed explanation see: PCR and PLS

### 2.2.2 Flow Chart

The flow chart shows how to build a quantitative calibration



***NOTE***
Loop 1 to Loop 3 represent the sequence for optimising a calibration. **After each change in the selection, the calculation and principal component selection should be repeated.**

### 2.2.3 Selecting the Calibration Method

For the quantitative calibration, basically two calculation methods are available:

- Principal Component Regression (PCR) consists of a Principal Component Analysis (PCA) with subsequent MLR.
- PLS Partial Least Squares Regression calculates the PCs with iteration in several steps, whereas spectral information and property values are considered.

The calculation method can be selected in the Menubar: Calibration / Method / PCR Principal Component Regression or Calibration / Method / PLS Partial Least Square regression or by clicking the icon PCR or PLS

### 2.2.4 Selecting the Calibration and Validation Spectra

Samples of known characteristics, both chemical and physical are used to generate quantitative calibrations. Ideally 60, but minimum 10 samples for each parameter with different concentration should cover the calibration range homogeneously. The measured spectra are divided into two sets:

- Calibration Spectra: spectra selected for the calibration: C-Set, about 2/3 of all measured spectra.
  Generally, the calibration spectra should contain all “extreme information” to define the limits of acceptance. The spectra with the highest and lowest property values should always belong to the C-Set.

- Validation Spectra: about 1/3 of all measured spectra selected for the validation V-Set. The V-Set should be spreaded over the whole calibration range, possible equally. Only if a calibration treats the validation spectra equal to the calibration spectra, the settings are considered as OK.

***NOTE***
These two groups of spectra should be:

- independent from each other;

no overlapping allowed.

**Suggested selection**: use the selection in the **calibration curve** graphic. In case some extreme samples are in the V-Set, see Loop 3.

It is **possible to leave out** spectra from the C- and from the V-Set (e.g. unknown property values), these spectra are in the Unused Set = U-Set:

C-Set + V-Set + U-Set = All Spectra

The U-Set is visible in the calibration results (see Calibration Protocol), but will not influence the calibration and validation results e.g. will not be considered for the Q-Value calculation.

Link: Spectra selection

### 2.2.5 Selecting the Calibration Wavelengths

The exact wavelength / wavenumber range measured is dictated by the instrument type used for the spectrum measurement. The selected wavelength / wavenumber range depends on the application and the measuring option used.

***NOTE***
In general the calibration wavelength / wavenumber range should be as wide as possible.

Suggested wavenumber range for NIRFlex N-500 with measuring options

- Solids and Liquids: 4'000-10'000 $cm^{-1}$;
- Fiber Optics: 4'500-10'000 $cm^{-1}$.
- Solids with Tablet Accessory: 6'000-11'520 $cm^{-1}$.

The calibration wavelengths define the spectrum range where the mathematical algorithms is applied.

**Suggested selection**: use all measured wavenumber in the first calculation, otherwise see Loop 2. Link

Wavelength selection

### 2.2.6 Selecting the Calibration Properties

The calibration property is the parameter, that is required to determinate in the application. The mathematical algorithm will be applied for the selected property values

**Suggested selection**: create separate calibration for each parameter - single property calibrations: only **one quantitative property is selected** in each calibration.

***NOTE***
Quantitative calibration with several properties can not be stored to the Database, because it can not be used in the application.

Link Calibration property selection

### 2.2.7 Applying Data Pretreatments

Data pretreatments are used to eliminate non important effects or to enlarge minor effects of the measurements.

Suggested selection: perform the first calculation without pretreatment, using the spectra as they have been measured. The first step to optimize the calibration is to add and change the pretreatments.

Link: Pretreatment selection

### 2.2.8 Performing a calibration (calculation)

After selecting the spectra (C-Set and V-Set spectra), the wavelength range (calibration wavelength) and the properties (calibration properties), the chemometric parameters (principal components, scores, residuals, etc) will be calculated.

For the calculation of PCR or PLS, the software uses only the spectra selected into the C-Set referring to the calibration wavelength and the choice of the data pretreatment's. The spectra of the V-Set will only be used to prove and judge the calibration

An overview of the calculation can be obtained in the Menubar: **View / Overview.**

The result of the calibration will depend on the selected number of **primary and secondary principal components.** The first step is to decide on the numbers of PCs.

The calibration can be optimised manually or by using the Calibration Wizard. Here the manual selection criteria are described.

### 2.2.9 Primary Principal Components

Primary principal components are the PCs, which are used for reconstruction of the spectra. They determine the residual value. The more primary PCs used, the smaller the allowed residual of the calibration.

Recommended graphics for primary PCs selection:

- X-PRESS function,
- Loadings graphic;
- Spectra Residuum plots.

**Suggested selection**: avoid overfitting, do not use too many primary PCs. Link
Primary PCs

Adjust the desired number of primary PCs.
Link: Adjusting Primary PCs

### 2.2.10 Secondary Principal Components

The secondary or calibration PCs are used for the **parameter** calculation. The number of secondary PCs is limited to the number of primary PCs.

The target is the best matching between the original reference values and the predicted NIR values. This can be seen in the "Predicted Property vs. Original Property" plot and can be judged with statistical values like:

- standard errors: SEC / SEP;
- V-Set Bias;
- regression coefficients;
- PRESS-values;

These plots are part of the Overview-Plot.

**Suggested selection**: use the optimal number of secondary PCs, which gives the best result for all spectra in the C-Set and V-Set.
Link: Secondary PCs

Adjust the desired number of secondary PCs.
Link: Adjusting Secondary PCs

***NOTE***
After each new selection a recalculation should be performed again.

### 2.2.11 Judging the Calibration

The target is the best matching between the original reference values and the predicted NIR values.

This can be seen in the calibration curve and can be judged with statistical values like:

| | | |
|---|---|---|
| Precision | SEC/SEP | as small as possible (around the standard deviation of the reference method) |
| Accuracy | V-Set Bias | around 0 |
| regression coefficients | r | close to 1 |
| Q-Value | | close to 1 |
| Consistency | | around 100 (80-110) |

These values are documented in the Calibration Protocol.

After several calibration optimizations and running the automatic calibration wizard, there will be several calibrations in the project. Sort the calibrations with a click on the Name, Q-Value or e.g. SEP.

***NOTE***
Keep only the best calibration and delete all others.

### 2.2.12 Save Calibration - Lifecycle

Save the project and the calibrations to the database. Only approved calibrations according to Lifecycle will be available in the NIRWare Application Designer. See Calibration handling for further details.

### 2.2.13 Create a Quantitative Calibration with Cross Validation

The above tutorial is valid for a user selected validation set (VS method).
It is possible to create a quantitative calibration using the Cross Validation (CV method), in that case **all spectra should be in the C-Set**, the **V-Set should be empty**.

# 3 Chemometrics

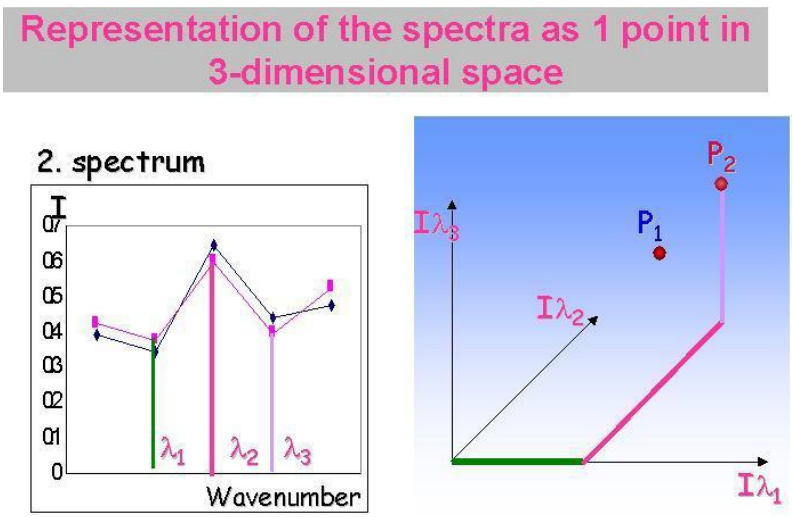
## 3.1 Calibration Methods

### 3.1.1 Principal Component Analysis: PCA

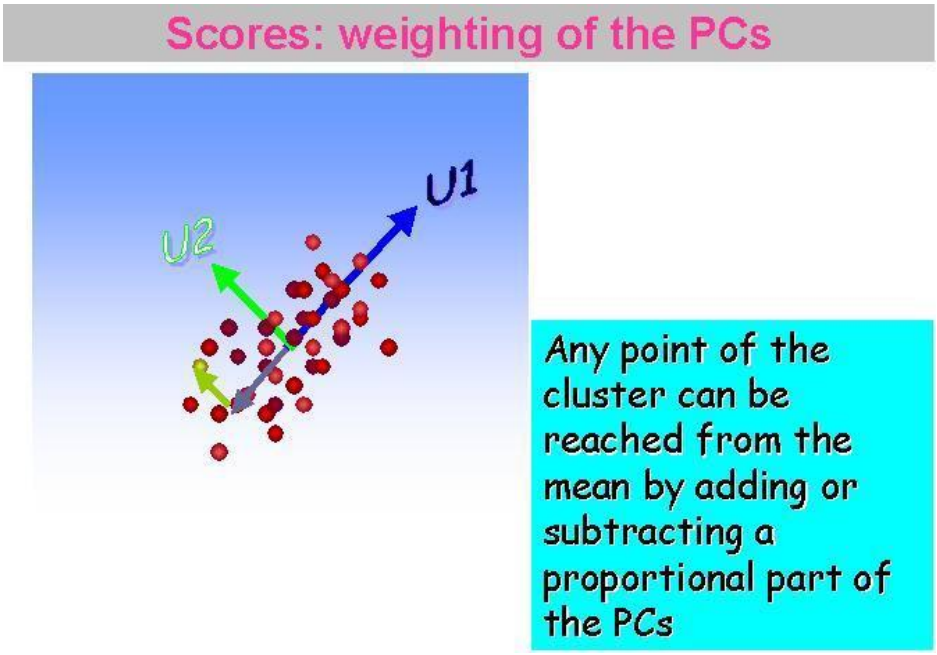
Principal Component Analysis is a **mathematical**, **statistical** evaluation of a **large amount of chemical data**. In this case the chemical data are the measured NIR spectra.
PCA is made for two reasons:
- to reduces the data amount without loosing necessary information. Noise is truncated by the number of primary PCs;
- to evaluate the measured spectrum automatically after creating a calibration.

With today's powerful computers, the prime object is no longer the reduction of the data volume. Today, the main goal of PCA is to **find and automatically evaluate** characteristics of identity, quality and quantity in the spectra.

Each spectrum measured with **NIRFlex N-500** consists of **1.501 data**, which correspond to the intensity values of the 1.501 support points on the wavenumber scale.
In order to obtain a good calibration, a large number of spectra is needed. For 100 substance spectra, this already gives us 150.100 data points, which places an enormous computing workload on computers.
To achieve acceptable computing times, the spectral data are therefore efficiently **compressed** with the aid of PCA without **loosing any important information**. For this purpose, PCA utilises the redundancy occurring in the spectra. With PCA, so-called **principal components** are extracted which are statistically independent from each other and which are therefore orthogonal relative to one another, yet are still capable of adequately **reconstructing the original spectra**.
The PCA will always be performed with the calibration spectra set in the selected wavenumber with the selected pretreatment.
A geometric explanation will serve to visualize the PCA: it is not possible to imagine a space of 1.501 dimensions (selected wavelengths), with each wavelength or wavenumber corresponding to a dimension. But in this space, a spectrum can be represented as a point. For three dimensions, this can be shown graphically:




In mathematical terms this point is equivalent to a vector with 1.501 components (I1....I1501). Several calibration spectra produce a "cloud" of points - a cluster- in space. For a set of spectra or points in the 1.501-dimensional space, a coordinate's transformation is now performed in a way that the new origin comes to lie in the mean centre of all the spectra - mean centering - and the new space directions - principal components - lie along the greatest variance in the spectra.
The new space directions are calculated in such a way that the features with the widest variances - differences- of all spectra are included in the first space directions and the higher space directions gradually evolve into noise. Space directions, which contain no any other information than noise, are no longer taken into account.




A **reduction** of the dimensions is performed when the number of PC's (i) is not higher than 1.501. Through this type of data reduction it is **impossible to lose information**.
With today's powerful computers, the prime object is no longer the reduction of the data volume. Today, the main goal of PCA is **to find and automatically evaluate** characteristics of identity, quality and quantity in the spectra.
As a result of the PCA, the following is obtained:

**Mean spectrum: <I (k)>**
Calculated by averaging the intensity values at each wavelength. The centre of the new coordinate system is shifted to that point: **mean centering**.

Formula: $\langle I(k) \rangle = [\, I_1(k) + I_2(k) + \ldots\ldots I_N(k)] \,/\, N$

**Principal Components: $U_i(k)$**
New space directions in the point cluster, which are also called principal components. PCs are artificial differential spectra.
Each calibration can have i PCs (default is: 15).

**The mean spectrum and PCs are always fixed for the calibration.**

**Scores: v in**

Weightings of each PC after the pretreated spectrum has been transformed into the cluster. A score is the portion of a PC used for the reconstruction of the original spectrum. Each spectrum has different and up to max. i scores.

**Residuum: R n (k)**

The difference between the pretreated spectrum and the reconstructed spectrum is the **residuum spectrum**. When the residuum is summed across the wavelength, a number is obtained, the **Residual**.

**The scores and the residual are variables for each spectrum.**

**Leverage**

The direct distance of the spectrum from the centre of the coordinate system in the score place.

**Reconstruction of a spectrum**

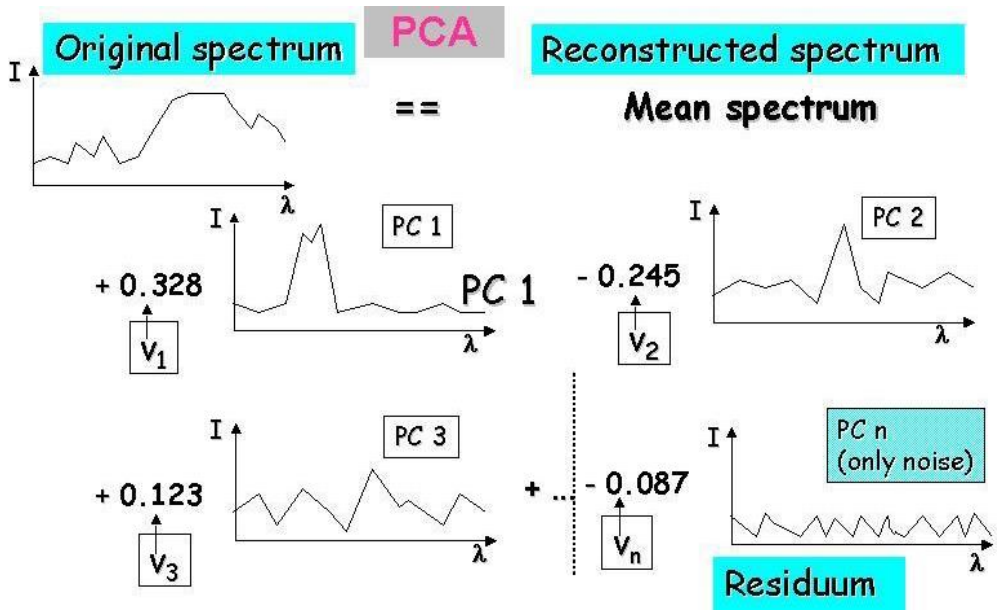
Now, **each spectrum can be reconstructed** on the basis of a sum through multiplying of scores and PCs.
Any desired spectrum I n (k) of the calibration set is developed:

$$I_n(k) = \langle I(k) \rangle + \sum_{i=1}^{\max i} v_{in} \bullet U_i(k) + R_n(k)$$

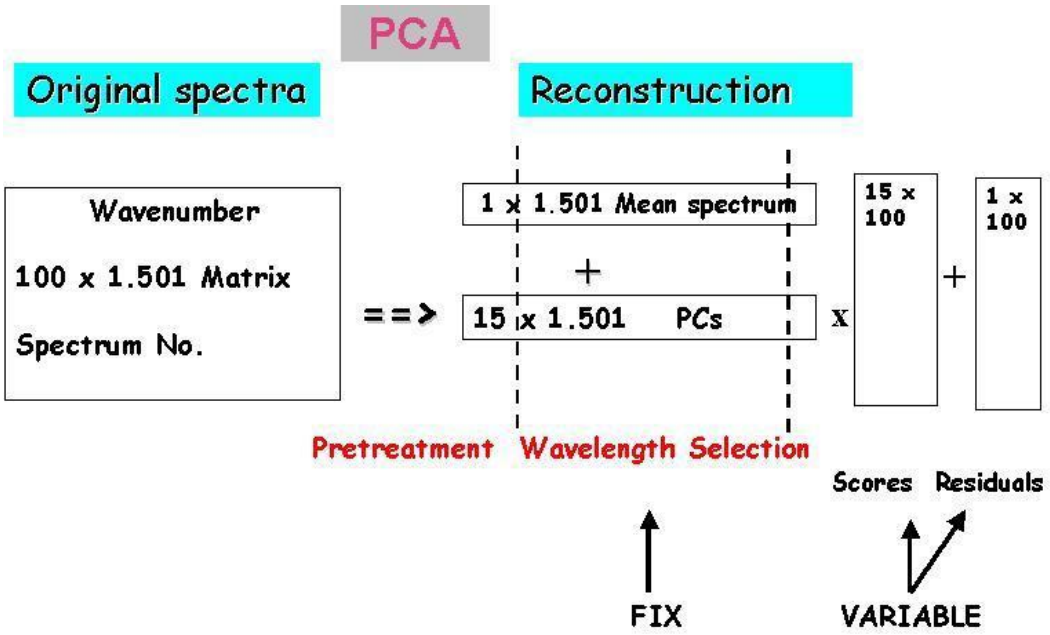
Formula:

For the reconstruction it is now only necessary to **save the scores v in and the residual R n of each individual spectrum**, since the **mean spectrum and PCs are constant** for the entire set of spectra in a calibration.
The spectrum can also be described as the linear combination of the PCs -U i (k)- and their scores (mean-centred data matrix).

The following figure shows the reconstruction of a spectrum:




Mathematically speaking, the Principal Component Analysis is then a breakdown of the spectrum matrix into 2 smaller matrices. This matrix operation can be represented graphically for, e.g., 15 PC's in the following way:

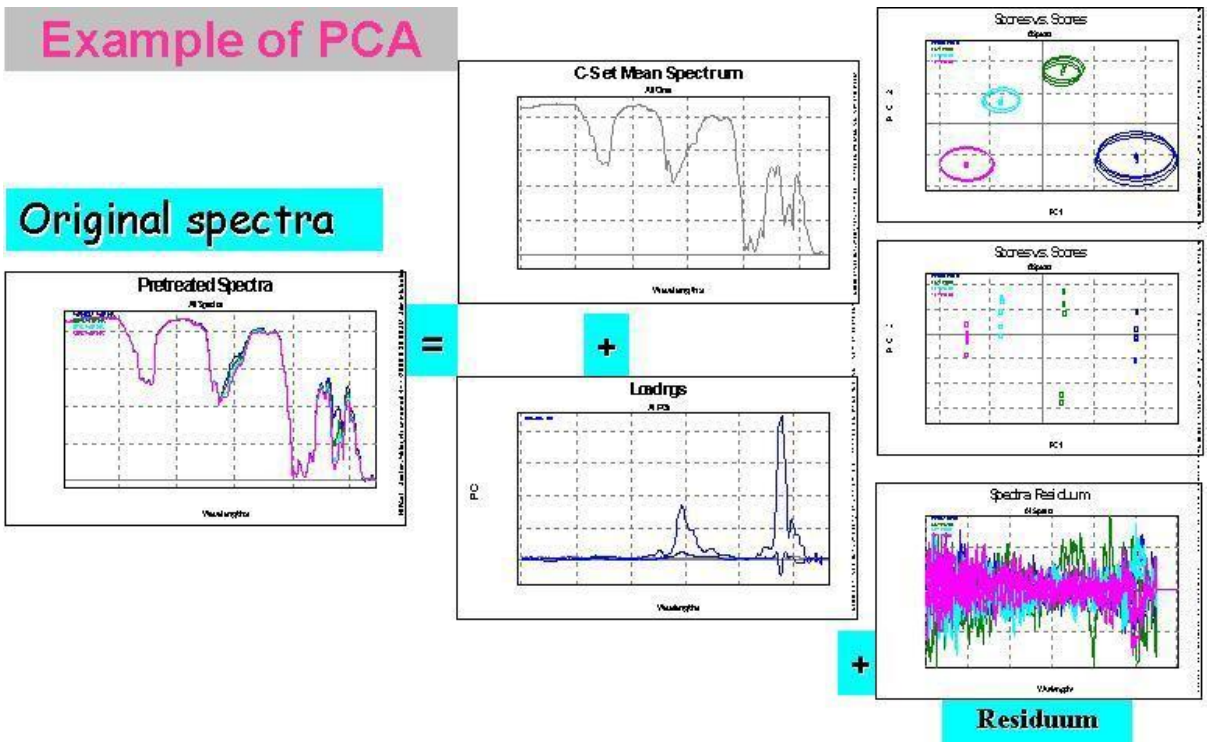



For 100 spectra with1501 data points the following data reduction is obtained with 15 PCs:

100 x 1.501 ==> 1.501 + [15x1.501] + [100x 15] + 100
150.100 ==> 1.501 + [22.515] + [1.500] + 100

**Example for a Principal Component Analysis:**

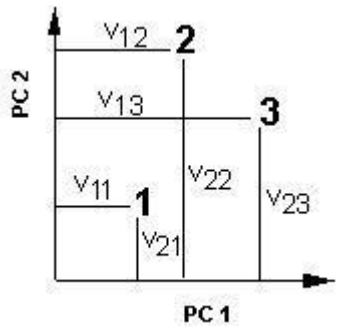
4 different acetone qualities: without and with 0.3 %; 0.7 % and 1.0 % added water. There are only 3 PCs necessary to reconstruct the spectra. The scores according to PC 1 and PC 2 are repeatable. The residuum spectra have only noise character.




As a result of the PCA, we obtain those PCs which themselves represent spectra and scores for each spectrum.

The scores can be represented in two- or three-dimensional PC plots. Each number represents a spectrum, v in its score.

Here, the scores of the spectra 1, 2 and 3 of the PCs 1 and 2 are graphically represented. Each spectrum with "i" PCs will also have "i" scores. The closer together the points in the plot, the "more similar" the spectra. It is now possible to break down an unknown spectrum with regard to these two PCs, i.e. to determine the scores. If this spectrum is located, e.g. in the region of 1, it will be identified as 1.




User-allocated properties of the spectra (e.g. quantity, good/poor quality, identity) do not have any effect on the Principal Component Analysis.

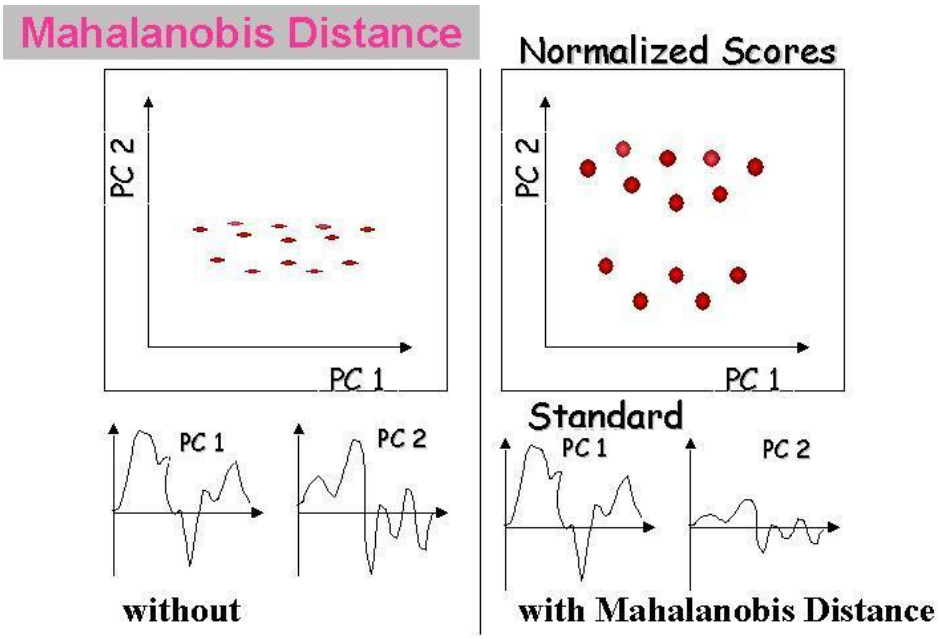
**The Mahalanobis distance**

The introduction of Mahalanobis distances means an artificial scaling of the scores with the square scores sum being normalised. At the same time this leads to a stretching or compression of the PCs, since the product obtained from the score and PC is not changed.
A new normalisation is performed so that the scores v in of the spectra n will retain roughly the same magnitude as the PC index increases. Scores are variables without unit that must only be considered relative to one another.
The purpose behind all this is to make physical or chemical properties which have only slight effects on the spectral data and which therefore only manifest themselves in higher PCs as visible as those clearly shown in the spectra.

The scores are normalised as follows:

Formula: $$\sum_{i=1}^{\max.i} \left(v_{in}\right)^2 = 1$$

For this reason, the points in the 2-D plots are evenly distributed, i.e. the scores of all PCs have the same average magnitude. On the other hand, individual PCs are appropriately reduced or increased. For further evaluation, only the normalised scores are used.




The Büchi NIRCal software package only works with the normalised distances, the user can not see the result without it.

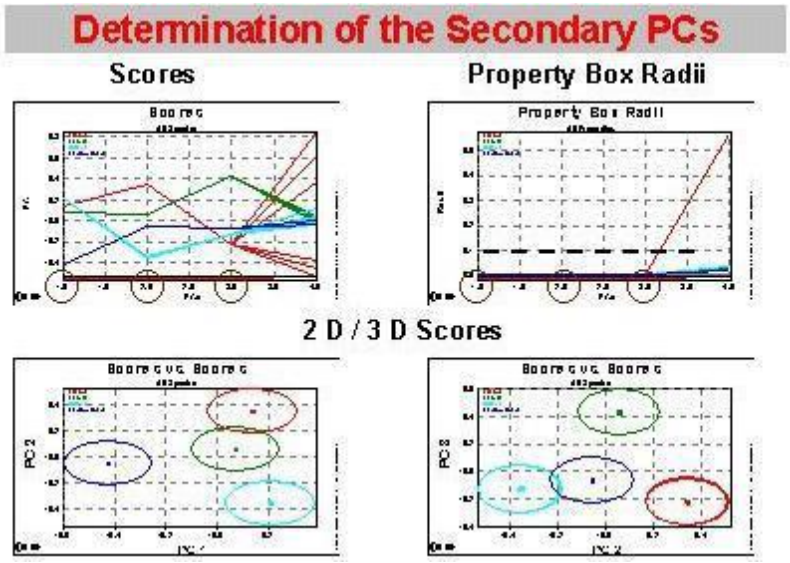
### 3.1.2 Cluster: CLU

Goal: to identify different chemical substances using the **PCA** and as a result to get a **well separated cluster** area in the scores plot **for each substance**.

The clusters are created according the selected **secondary PCs**. Secondary PCs are that PCs among the primary PCs, which shows a **clear separation** of the substances and the scores are **good repeatable**. These secondary PCs are responsible and used for the **tolerance radii calculation**.

Tools for the selection of secondary PC's:

- the scores of the spectra will be shown dependent to the PCs: **Scores against Spectra**. The scores of all spectra of each substance should be close together, but separated from the scores of all other substances;
- the **Property Box Radii** plot shows the repeatability [(max-min) scores / 2] of the scores for each property. PCs showing small Property Box Radii values (normally below 0.1) are important for the calibration;
- 2 and 3 dimensional **score plots**: PCs having repeatable scores should be selected.

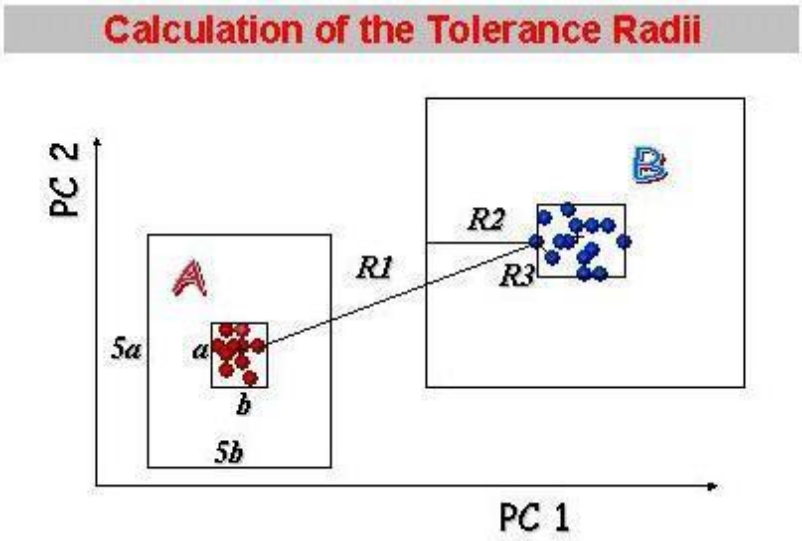


Additional tool:

- The Score Disorder values show how effectively a particular PC separates different properties (substances) from each other. By scanning the score values in one direction of a PC and counting the changes between the membership of A or B the disorder value is achieved. If a PC completely separates all calibration properties, the **disorder values** is the [number of **calibration properties - 1]**. PCs with small disorder values are possible calibration PCs.

The selected number of secondary PCs should be adjusted and the calibration should be recalculated.
Link: Secondary PCs

The **tolerance ring radii** are determined according to the **secondary PCs**.
For this calculation the **smallest possible rectangle** (rectangle in 2D, a cuboid in 3D, an n-dimensional cuboid in n-dimensions) constructed around each C-Set. The sides of the rectangles (a and b) are parallel to the axis of the PCs. The distance from the center to the side of the smallest possible rectangle is a measure of the **extension** of the C-Set, this is the so called "**Property Box Radii**". A '**virtual rectangle**' is created **5 times greater** than the rectangle around the C-Set.




The following distances will be calculated:

- R1 smallest distance between a spectrum (substance B) to the closest spectrum of a different property (substance A).
- R2 smallest distance between a spectrum (substance B) to the side of its virtual rectangle.
- R3 smallest distance between a spectrum (substance B) to the closest spectrum of its own property (substance B).
- R4 mean value out of all R3 distances of the same property.

With these distances it is possible to calculate tolerance circles with a **radius** r for every calibration spectra by using Formula 1, 2 or 3.

**Formula 1:**

Min. of r = R1/2 * f
or r = (R1+R2)/4 * f

The smaller of the two possible values for the radius "r" is used. The default setting for the **Radii Blow Up factor "f"** is 1.
Depending on the extension of the cluster (R2) and the distance between the two closest clusters (R1), the circles are closer or further away from each other. With a **Radii Blow Up** of 1 two circles **can just touch each other** but do not overlap.
For **chemically different substances** the Radii Blow up (f) may be **reduced** (0< f <1) in order to increase the sensitivity of the calibration.

NOTE: It is not suggested to have a Radii Blow up (f) higher than 1: it can cause overlapping rings!

**Formula 2**

If generally small tolerances are required, Formula 2 can be used:

min. of r = R1*0.499 (0.499: to avoid overlapping)
or r = R3 * Pre Blow Up * f
or r = R4*0.5*Pre Blow Up * f

The smallest of the three possible values for the radius “r” is used.

The pre Blow Up factor is 5, this is an **empirically** evaluated value. The default Radii Blow Up f = 1, this value can be adapted by the user. In general it should be **increased** to get connected spheres.

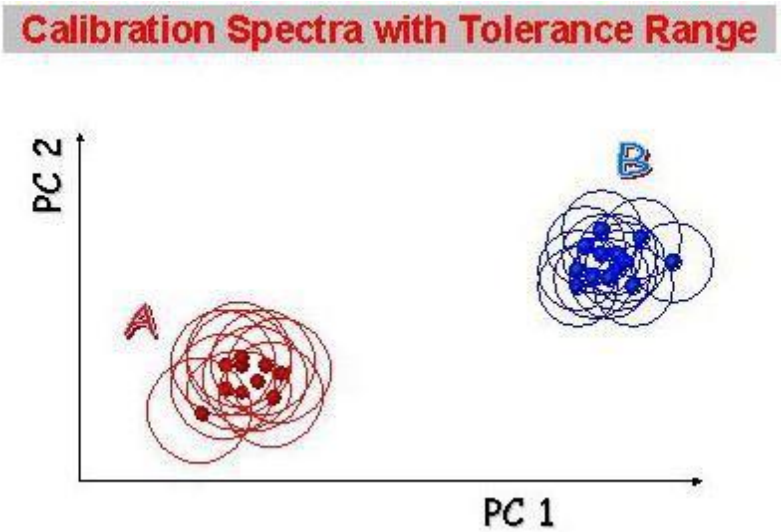
**Formula 3**

It is used **only for SIMCA** calibration, where only one substance is calculated (there is no R1).

Min. of r = R2 * 0.5
or r = R3 * Pre Blow Up * f
or r = R4 * 0.5 * Pre Blow Up * f

The smallest of the three possible values for the radius “r” is used.

The tolerance circles can now be plotted for all calibration spectra:




These tolerance ranges will show if the sum of the radii of one property generates only **one cluster**: cluster / property should be one. In this case the number of the secondary PC selection is OK. If more than one cluster is generated, the number of the secondary PCs is not optimal or other calculation (e.g. pretreatment) should be tried.

**Assessing the calibration**

***Cluster per Property***

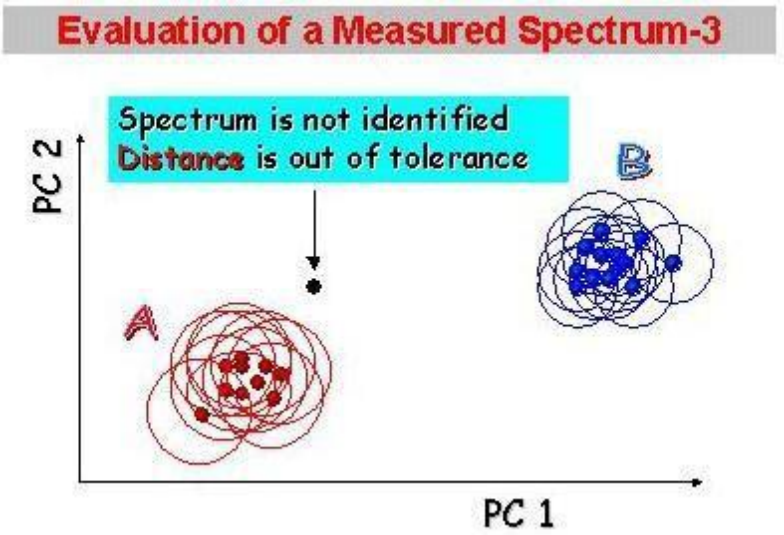
The Cluster per Property plot shows if all tolerance circles build one connected cluster for each property.
Here only a value of **1 is acceptable**.

**All spectra in the C- and V-Set should be identified correctly**

The identification is made according to the distance in the scores plot and residual:

***Distance***

The distance to the next calibration spectra should be **smaller**, than the **tolerance ring radius** of the neighboring calibration spectrum. In this case the distance criterion is OK, the **spectrum is in a cluster.**




**Property Residuum zero** means, all spectra are in the correct cluster.
Here only a value of **0 is acceptable**.
Property Residuum **+1** means, that a spectrum is **outside** the cluster: it is **not identified**.
Property Residuum **-1** means, that a spectrum is **in a wrong cluster**: it is **false identified**.

***Residual***

The [(maximum residual of a C-Set) * 2] is the max. **allowed residual** for the calibration and later for the application.
The default residual Blow up factor is 2, it may be changed by the user (it is not suggested to use smaller values as 1).
**Spectra Residual too** big should be for all spectra **zero**.

These three criteria are showed in the Overview plot in the 2nd column.

### 3.1.3 SIMCA

SIMCA is a calibration method used for identification of substances. Using SIMCA a Principal Component Analysis (PCA) is made for each substance/property in the project, but **each calibration is made for only one substance**.

Cluster calibrations can be transformed to several SIMCA calibrations. See Transform Cluster to SIMCA.
This tool creates for all C-Set properties one SIMCA calibration, the name of the calibration is "SIMCA + calibrated property name".

All SIMCA calibrations take over the following default parameters:

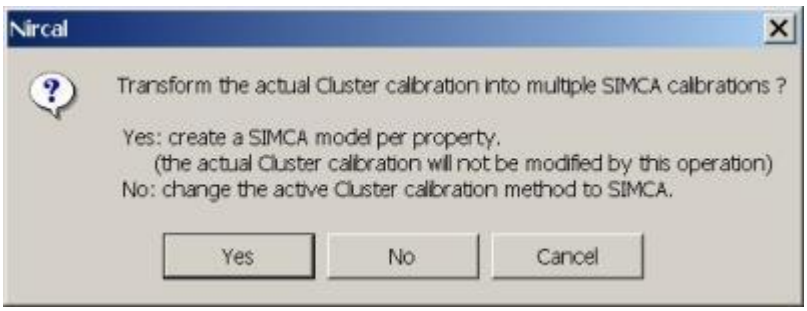
- the spectra C-Set and V-Set selection of the last active cluster calibration;
- the wavelength selection from the selection of the last active cluster calibration;
- pretreatments of the last active cluster calibration;
- the "Mean centering after Pretreatments" is still switched on;
- represent the mean value spectrum;

***NOTE***
Switching the **"Mean centering after Pretreatment" OFF**, the first principal component represent almost the mean value spectrum. This is suggested for only one property in the calibration, which is always the case for SIMCA. SIMCA can also be created without a transformation from a cluster calibration.

- the number of primary principal components is selected with the "Factor Selection Wizard";
- the number of secondary principal components is selected also with the "Factor Selection Wizard" according the Q-Value.

### 3.1.4 Transform Cluster to SIMCA

An active Cluster Calibration can be transformed to SIMCA by changing the calibration method to SIMCA. Before the wizard starts with the calculation the following pop-up window appears:

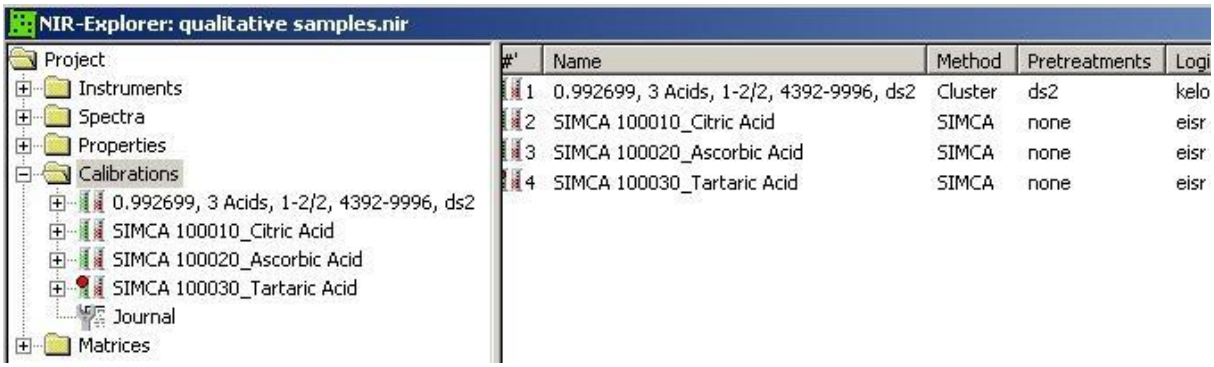



Yes Transform the actual Cluster Calibration into multiple SIMCA Calibration using default SIMCA Parameters. The existing Cluster Calibration will not be modified.

No Change only to SIMCA method without any calculation.

Cancel Cancel the transformation and keep the current method.

For each C-Set Property of the Cluster Calibration a separate SIMCA Calibration is created.

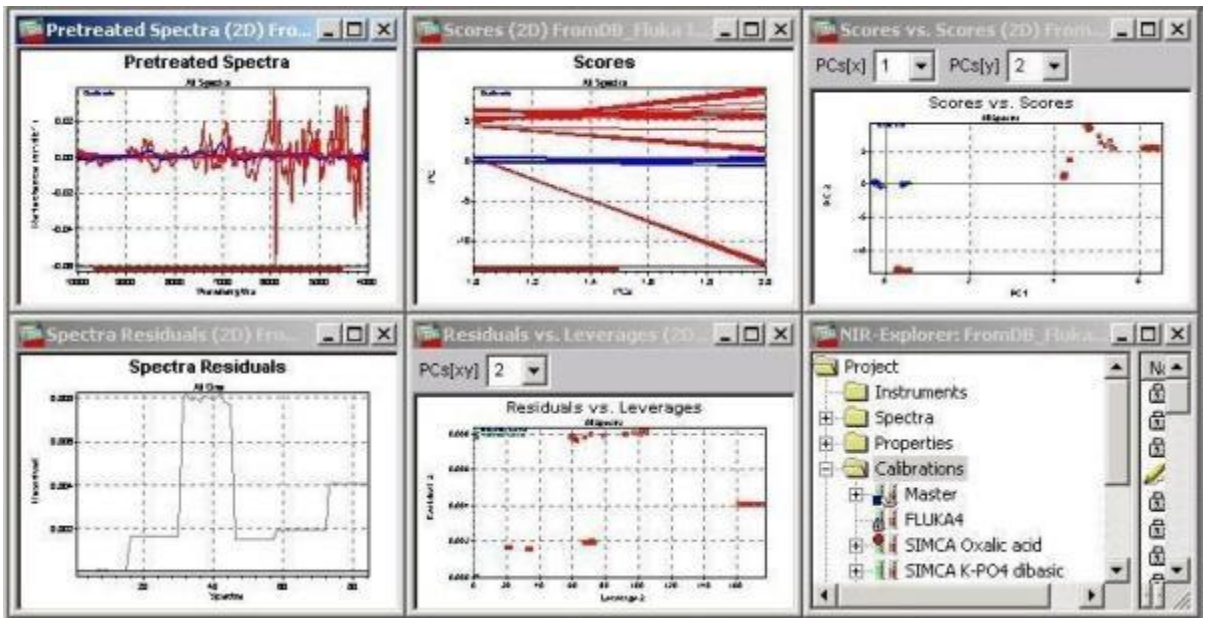



Hereby the SIMCA default parameters are used and calculated. An automatic estimation of primary and secondary PCs is made.

**SIMCA Overview Plot**

The Overview plot is automatically opened after the SIMCA calculation is finished. Each calibration has the name:

- **SIMCA** + **substance name** (property name) selected for the calibration.



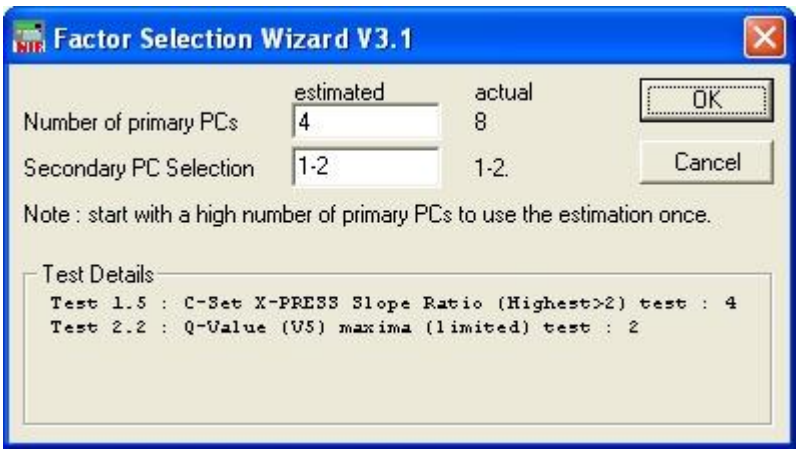

The Overview plot contains the following plots:

| | Plot name | Description |
|---|---|---|
| 1 | Pretreated Spectra | using the same pretreatment, as the cluster calibration |
| 2 | scores against PCs | scores against the PCs for each spectra in the project |
| 3 | scores vs. scores | in this score plot the calibrated substance is around the center (Mean Centering distance is calculated for only the C-set spectra), each other spectra, which are not in the calibration, have normally huge score values |
| 4 | spectra residual | the residuals of the calibrated substance spectra are normally smaller, as each other substance residuals |
| 5 | residuals vs. leverages | the so-called **Coomans plot** shows the residuals against the leverages (leverage is the direct distance of a spectrum in the score plot from the centrum) |
| 6 | NIR-Explorer | |

For a SIMCA calibration the limits are also called:

- Residual : outer model distance;
- Leverage : inner model distance.

**PC (Factor) Selection Wizard**




**Primary PC selection:**
Test 1.5 : C-Set X-PRESS Slope Ratio Highest test

w(i+1) = ( y(i) - y(i+1) ) / ( y(i+1) - y(i+2) )
Limit = 2
PC i for highest i where w(i) > Limit Precise

**Secondary PC selection:**
Test 2.2 : Q-Value maxima (limited) test

PC i for Max ( QValue(i) ) I < NumPrimaryPCs

NIRCal calculates the **allowed residual** using the **primary principal components** for reconstruction.
Default parameter for residual
Residual Blow Up = 2.5

Allowed residual for calibration is 2.5 x max. C-Set Residual.
For each C-Set spectrum NIRCal creates a **tolerance sphere** using the **Formula 3** for radii calculation according to the Mahalanobis distances with the **secondary PCs**. This calibration sphere "inner" space defines the area for a substance.
Default parameters for scores and radii:
Scores Blow Up = 1.05

- Radii Formula = 3

Radii Blow Up = 2.5

**SIMCA Q-Value**

The Q-Value for SIMCA calibrations can not take into account the "Property Interference" value, because there is only one cluster type in each calibration. This value is always Zero. This causes a slightly higher Q-Value against the Cluster calibrations, in case no outliers are in the SIMCA calibration.

**SIMCA Method Validation**

SIMCA allows that the principal component spaces **cover each other** or can **partially overlap** another spaces. To check possible overlapping the "Prediction Protocols" can be used, as it is also suggested for the Cluster calibrations.

The number "**Total not identified, Cluster BAD (&)**" and „**Total not identified, Cluster OK (%)**" can indicate the correct adjustment of the Blow up (Radii and Residual).
SIMCA tolerance spheres normally lie close to the PC center, other spectra can lie also here, which causes losts of "**Total not identified, Cluster BAD (&)**" cases. This can be reduced by reducing the

Radii Blow Up limit.
The number of "**Total not identified, Cluster OK (%)**" can be reduced by reducing the number of primary PCs or by increasing the Residual Blow Up limit.

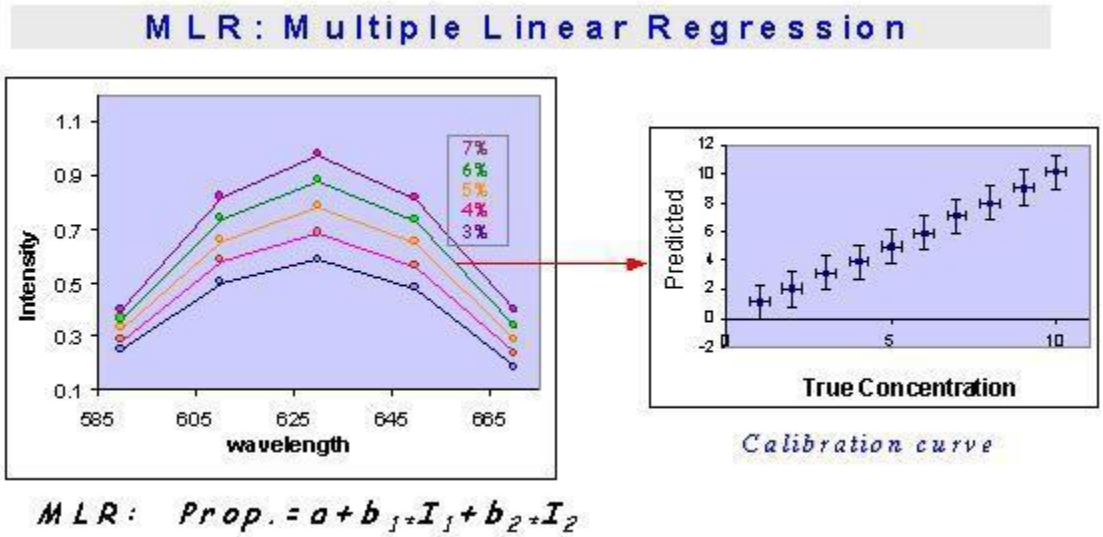
**Using SIMCA in application**

For the identification of an unknown substance the residual should be below the allowed limit and the unknown spectrum distance should be smaller as the allowed tolerance sphere to the nearest known calibration spectrum (inside the "inner space").
In the application mode there are 2 answers possible:

| Result | Residual | Distance |
|---|---|---|
| Identified | OK | OK |
| Not identified | not OK | not OK |
| Not identified | not OK | OK |
| Not identified | OK | not OK |

## 3.1.5 Multiple Linear Regression: MLR

Multiple Linear Regression is an extension of the linear regression to several dimensions.
The analysis is based on a few selected wavelengths and does not require any PCA calculation. In this procedure, the properties are calculated through intensity values and correlation coefficients, e.g. it is valid for two selected wavelengths ($I_1$ and $I_2$).




Where:

| | |
|---|---|
| Prop | property of the „n“th Spectrum |
| a | intercept |
| $b_1$ | correlation coefficient of the first wavelength |
| $I_1$ | intensity at the selected (first) wavelength |

**Note:** $I_1$ and $I_2$ must describe independent characteristics. Select at least 3 wavelengths.

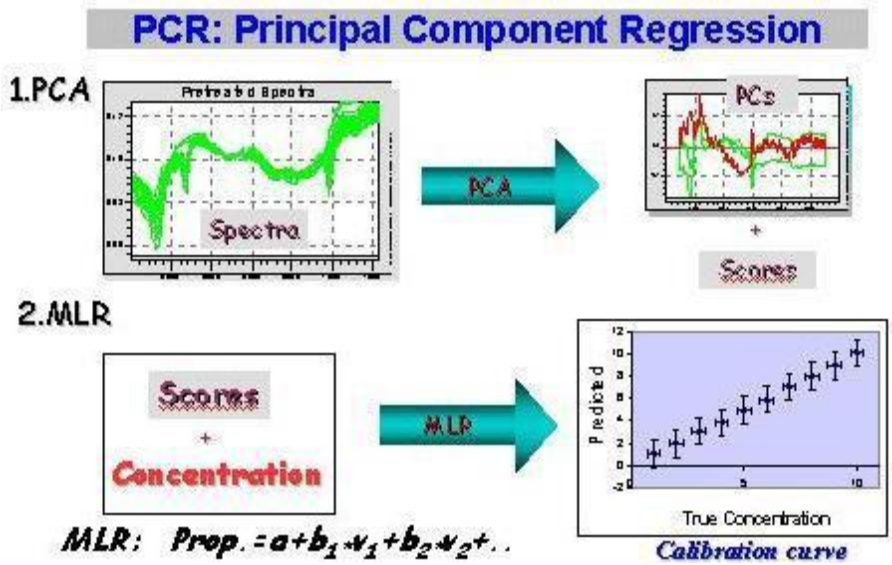
Because with MLR only few wavenumber (min. 3) are used and the rest of the measured 1501 (in case of NIRFlex N-500) are automatically discarded, this simple method is not suggested to use.

The residual cannot be used for outlier detection during the application because of the extreme wavelength reduction.

This method is only suggested for **filter instruments**. For Interferometers (full wavelength range) it is suggested to use PCR or PLS.

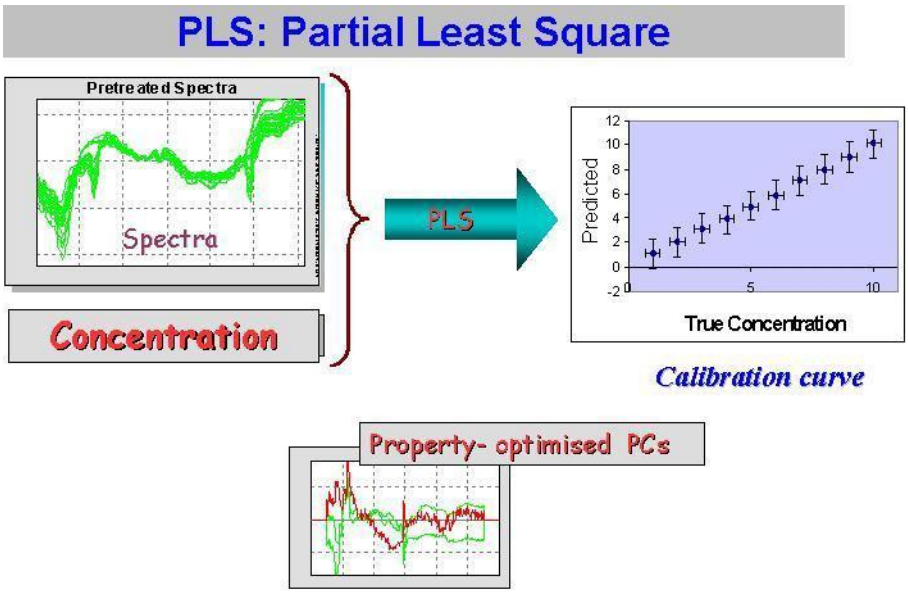
### 3.1.6 Principal Component Regression: PCR

Principal Component Analysis (PCA) with subsequent MLR is called Principal Component Regression (PCR). As a first step, the principal components and scores are calculated with **PCA**. The second step is a multiple linear regression **MLR** using the scores and property values (concentrations).
Since the calculation of the principal components is performed with the spectral data - independently of the subsequent regression calculation for the correlation of the quantitative values - any number of parameters can be simultaneously included in a PCR calibration. This also means that the relevant PCs for the determination of the property are not necessarily the ones describing the biggest spectral variations.




### 3.1.7 Partial Least Squares Regression: PLS

Partial Least Squares Regression (PLS) calculates the PC's with iteration in several steps, with spectral information and property values being taken into account simultaneously.
This calculation method is more up to date than the PCR. Based on the principle of recursion, PC's and scores are also calculated as with PCR, but the quantitative reference values are included in the calculation from the beginning.




Each of the calculated PC's in the PLS procedure contains information about the original property values (true concentration) of the samples, with the first PCs (unlike PCR) always showing the highest correlation.

If two parameters are not systematically correlated, the mathematical approximation of the spectra via PLS can never be performed for both parameters when each parameter is calibrated using its own PLS. For this reason, it is recommended to calculate properties which are not systematically correlated (e.g. ethanol and acetone contents in any given solvent mixture) in separate calibrations.
Therefore, whereas PCR reduces spectral data to the most dominate dimensions, the PLS aims at the most relevant dimensions (relevant here means: best match between predicted and original values). With PLS, the PC's are calculated exactly in relation to the highest correlation in the first PC.

## 3.2 Calibration Validation Methods

### 3.2.1 Validation Set (VS)

**C-Set and V-Set**

C-Set (Calibration Set).
From all spectra within a project only the spectra which are in the C-Set are used for the calculation of the calibration.

V-Set (Validation Set).
From all spectra within a project only the spectra which are in the V-Set are used for an internal validation of the calibration.

Normally for VS mode 2/3 of the spectra are selected as C-Set and 1/3 are in the V-Set. This can be done with Toolbox “Set Creation”.

### 3.2.2 Cross Validation (CV)

**Cross Validation Method**

Instead of dividing the samples into two groups, a calibration set and a validation set, all samples are used for calibration in CV mode. Several calibration runs are performed with all samples except a small group with which the actual calibration is tested. This group is changed for all trial runs. The validation results of all the runs are stored and lead to a standard error of the cross validation (SECV) which compares well with the standard error of predictions (SEP) obtained in VS-mode.

In case each CV group consists of only one sample (one leave out) the method is called **full cross validation**. This is the **default setting** for CV-grouping in NIRCal.

Cross validation is recommended for calibrations based on a small amount of samples. If this number is larger than 50, NIRCal suggests to use a validation set instead (this comment can be suppressed; see NIRCal Configuration Edit / Options / Calibration Defaults).

***NOTE***
All spectra of one sample must be assigned to the same CV-group.

***NOTE***
For CV mode all spectra should be assigned to the C-Set.

Limitations:

- Cross validation is available for PCR and PLS only
- Cross validation requires at least 4 spectra in the C-Set
- For cross validation at least 2 CV groups have to be assigned

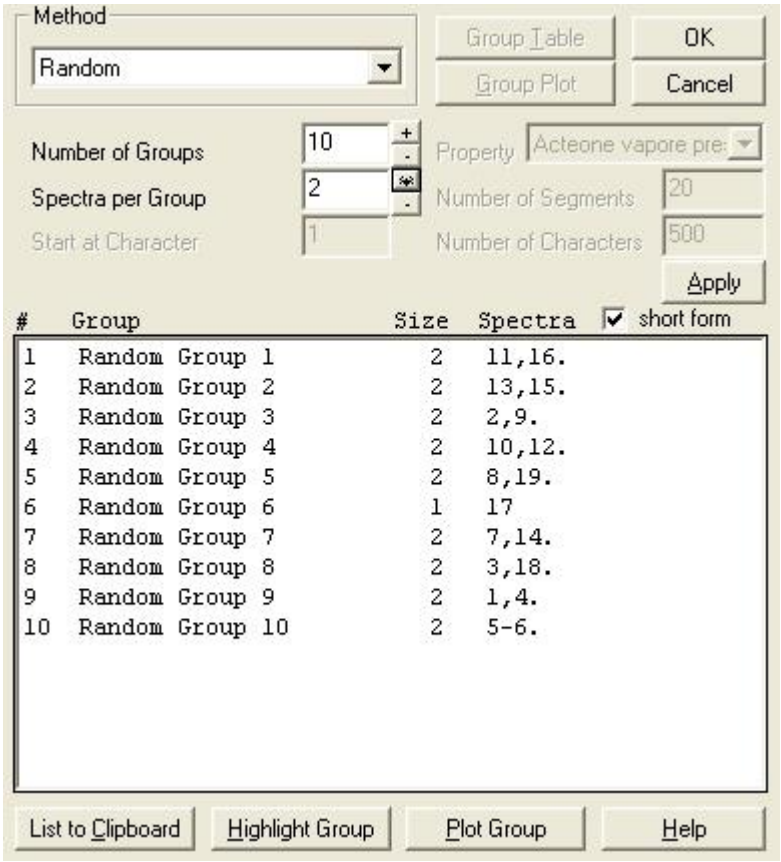
Cross validation will delete the V-Set assignment

**Cross Validation Grouping**

Menu: Calibration \ Change Data Sets \ Edit CV Groups...

Icon: 

In order to define the cross validation groups, the CV Group selector is opened.




The various possibilities for selections are summarized in the following table.

Methods

| | |
|---|---|
| One leave out | Each spectrum represents a group. Full cross validation |
| Alternate | The spectra are grouped to different groups one after the other.<br>**Number of Groups** and **Spectra per Groups** can be varied, they have an interdependency. |
| Sequence | Consecutive spectra are grouped to the same group.<br>**Number of Groups** and **Spectra per Groups** can be varied, they have an interdependency. |
| Random | Spectra are grouped randomly. The groups are filled to the maximal amount of **Spectra per Group.** |
| Property Segments | For the selected Property the range (min - max value) grouped into segments. An empty segment will not build a group. |
| Property Equal | All spectra with the identical property value are grouped. The property values are compared over all properties. |
| Spectra Name | All spectra with the same spectra name are grouped. The characters to compare can be defined with **Start at Character** and **Number of Characters**. |
| Spectra Name (autom.) | **Number of Characters** is appraised incrementally until the spectra can be grouped to more than one group.<br>**Start at Character** is always 1. |

| | |
|---|---|
| Custom assign Group to Spectra | This is the default method when the CV Group Selector is started . In this mode it is possible to display a plot (**Group Plot**) and/or a table (**Group Table**) with the CV Group Index |

Buttons / Selections

| | |
|---|---|
| Method | Select a method from the drop-down list. |
| Group Table | Display table: CV Group Index.<br>Enabled only for the method: Custom assign Group to Spectra |
| Group Plot | Display plot: CV Group Index.<br>Enabled only for the method: Custom assign Group to Spectra |
| OK | OK |
| Cancel | Cancel |
| Number of Groups | |
| Spectra per Group | |
| Start at Character | Used for Spectra Name<br>Default is 1, the character comparison starts at the first character. (e.g. Start at Character=10 then all Spectra names that are shorter than 10 characters like "Name xy","Batch01","1" are grouped into 1 group. |
| Property | |
| Number of Segment | |
| Number of Characters | Used for Spectra Name<br>Default is 500 characters, up to 500 characters are compared. |
| Short form | Enabled : e.g. 1-5<br>Disabled: e.g. 1,2,3,4,5 |
| List to Clipboard | The group list is copied to the clipboard. |
| Highlight Group | All spectra of a group are highlighted red in NIRCal-plots. |
| Plot Group | Select a group and plot |

After creating or changing the groups, the CV calculation should be performed again.

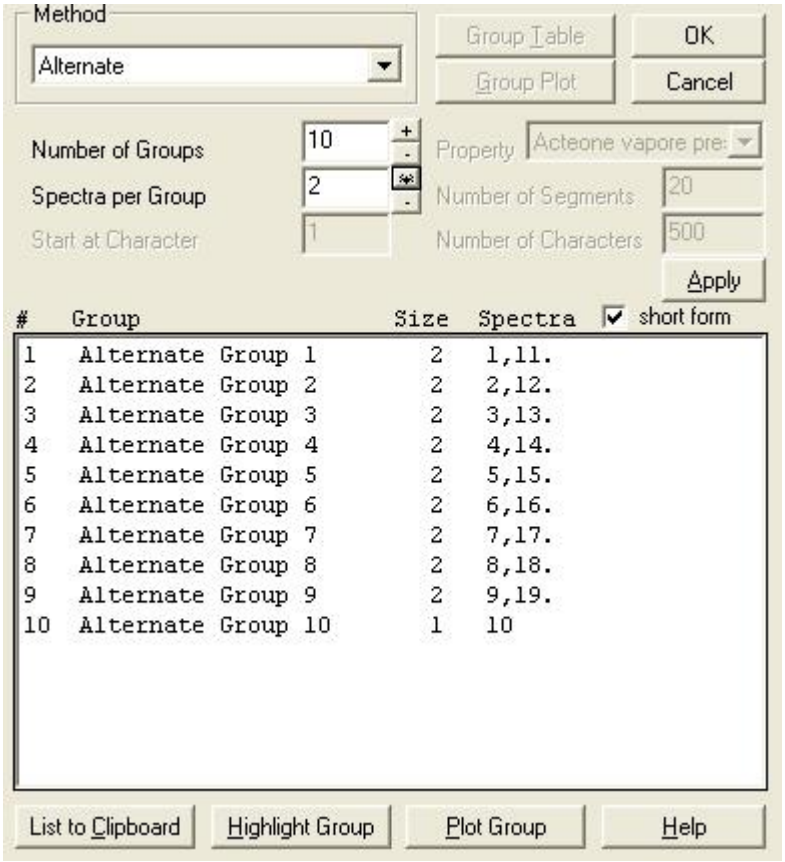
**CV Methods**

***One leave out (FCV)***

Each spectrum represents a group: Full Cross Validation.

***Alternate***

The spectra are grouped successively by increasing number (1st spectrum to the 1st group, 2nd spectrum to the 2nd group, etc.).

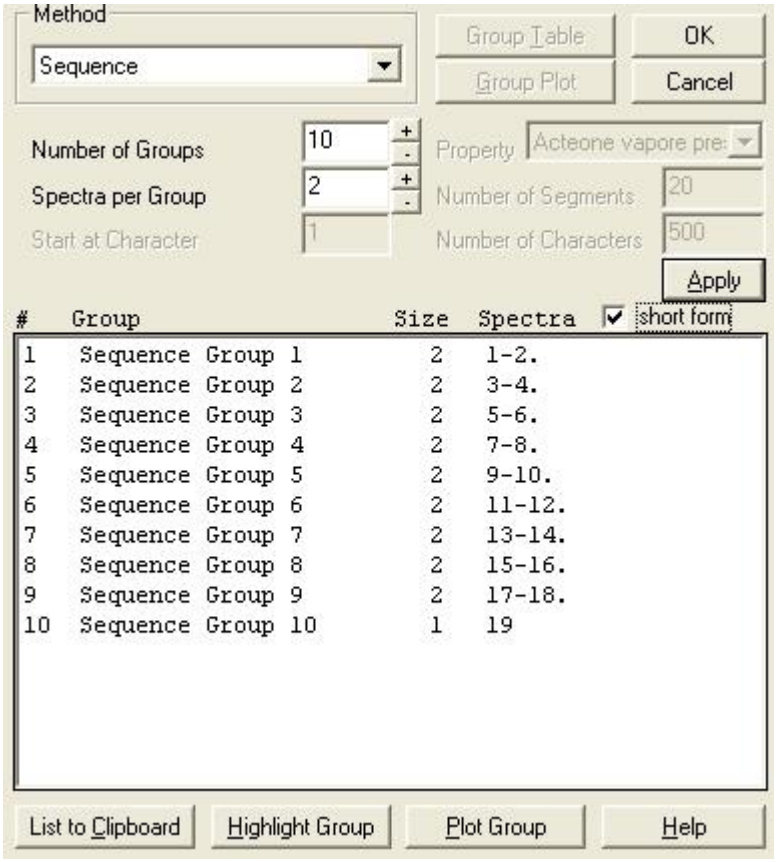



The number of Groups or the number of Spectra per Groups can be changed by clicking on the + or - symbols.


The number of Groups and Spectra per Groups are depending on each other.

***Sequence***

The spectra are grouped with consecutive number (1st to 3rd spectra to the 1st group, 4th to 6th spectra to the 2nd group, etc.).

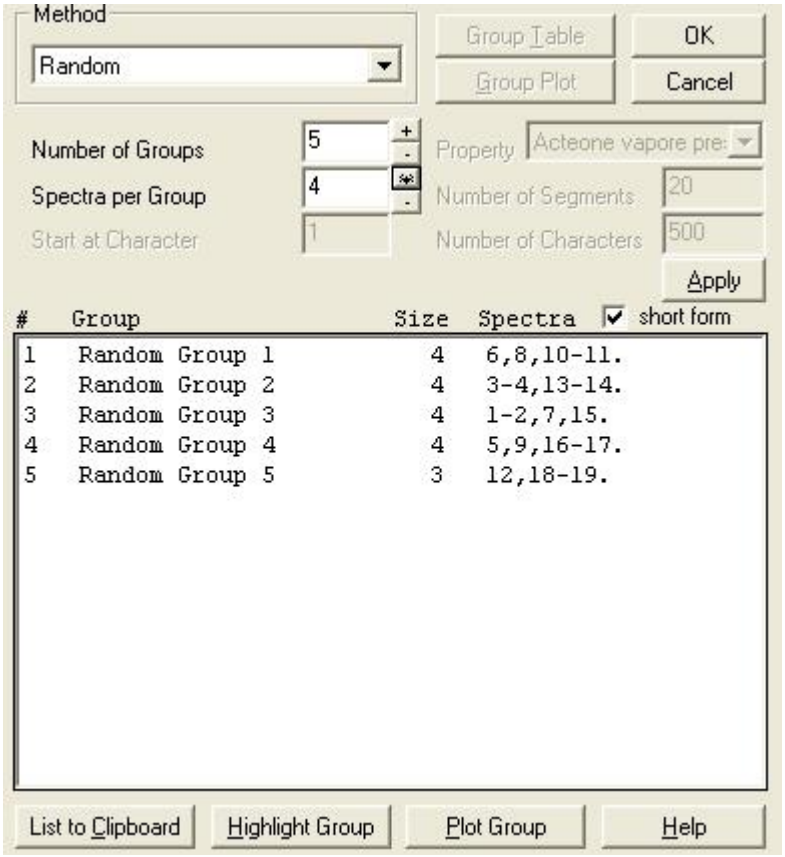



The number of Groups or the number of Spectra per Groups can be changed clicking on the + or - symbols.


The number of Groups and Spectra per Groups are depending on each other.

***Random***

The spectra are grouped randomly.

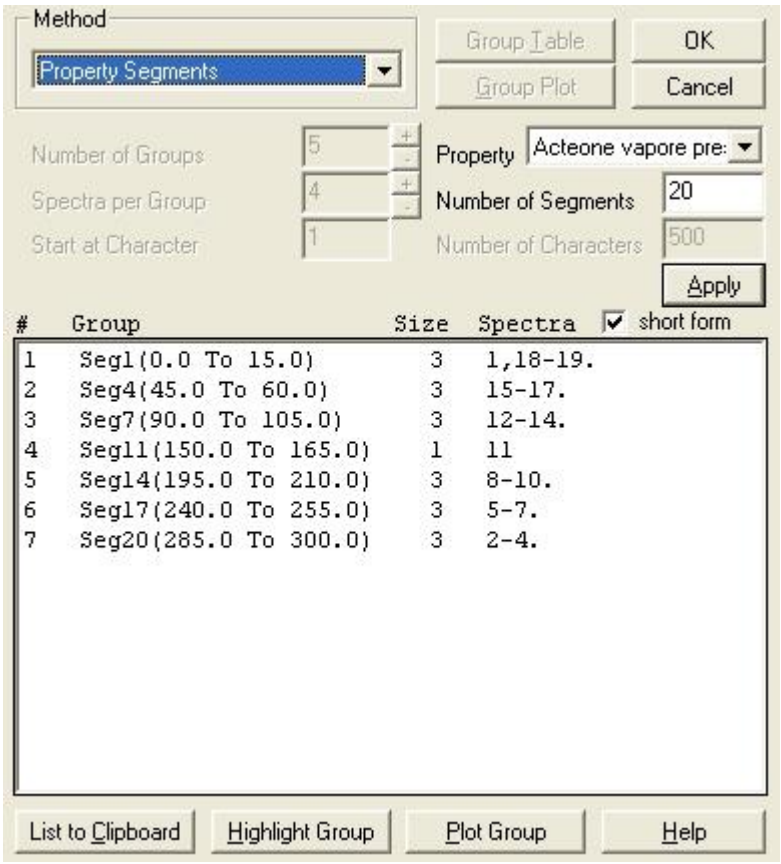



The number of Groups or the number of Spectra per Groups can be changed clicking on the + or - symbols.


The number of Groups and Spectra per Groups are depending on each other.

The smallest number of Group is 2.

***Property Segments***




For the selected property the min.-max. value range will be divided into Number of Segments (default: 20) Each spectrum, which has the value belonging to a segment, builds a group.

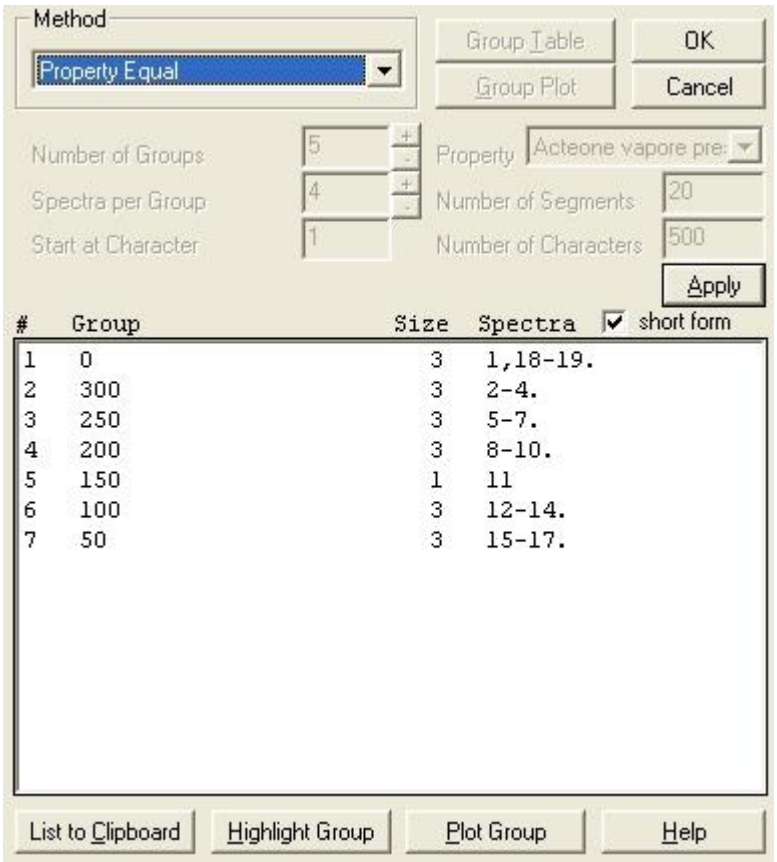
Segments without property value are empty (here e.g. concentration range from 5 till 10 %).

The cross validation calculates the groups for the selected property. Only one property must be selected.

***Property Equal***

Spectra with the same property value (concentration) selected for a group.

Each property is taken into account.

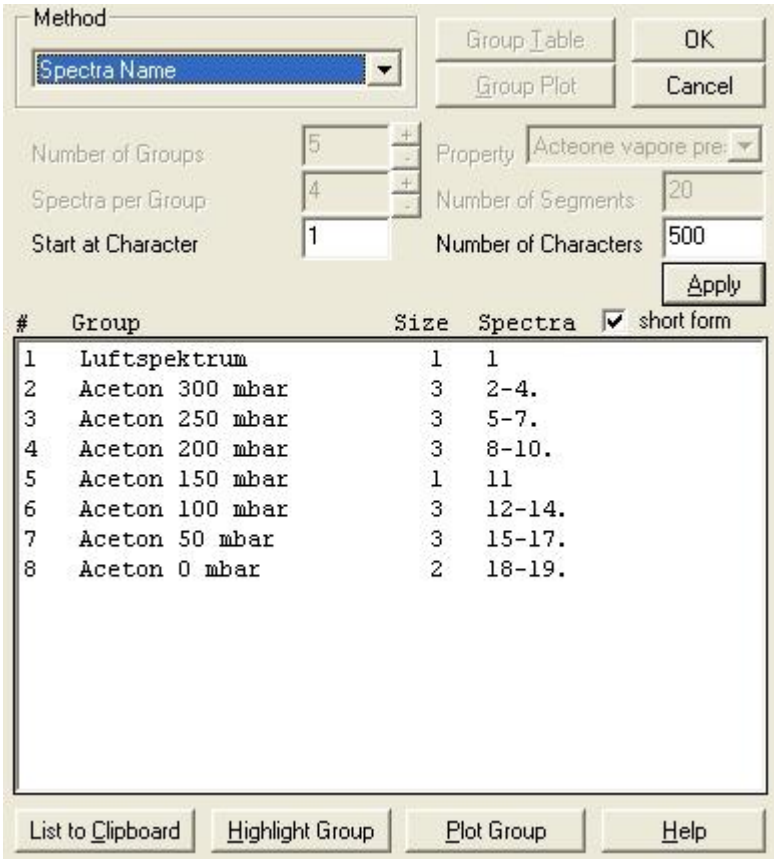


***Spectra Name***

Spectra with the same name are selected for a group.

The comparison will start from the 1st character (default), but it can be changed by the user.

The length of spectra name can be limited by: Number of Characters. Default: 500.

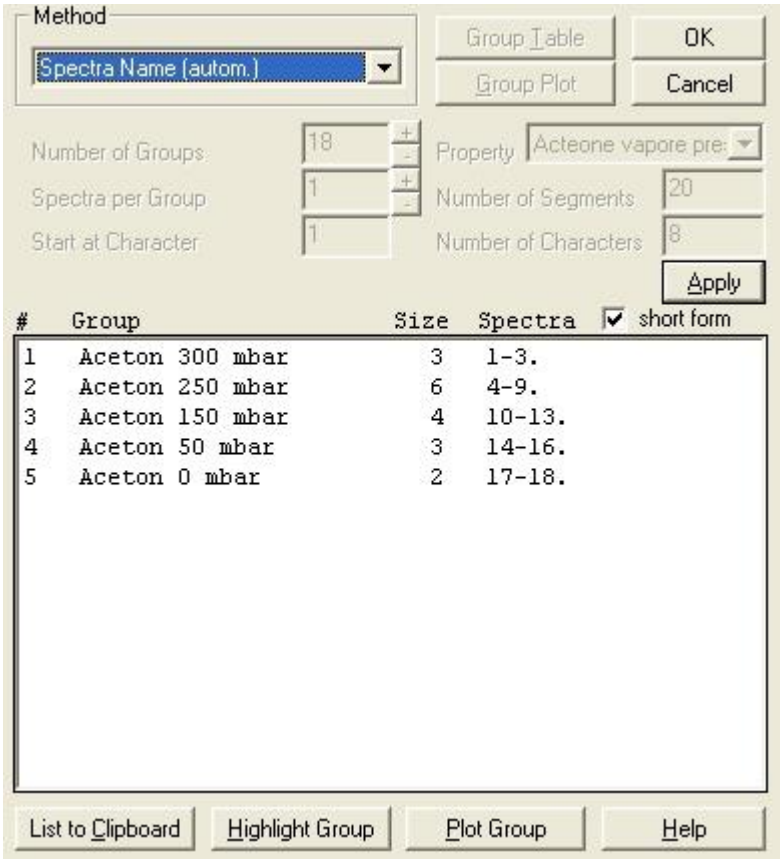


***Spectra Name (autom)***

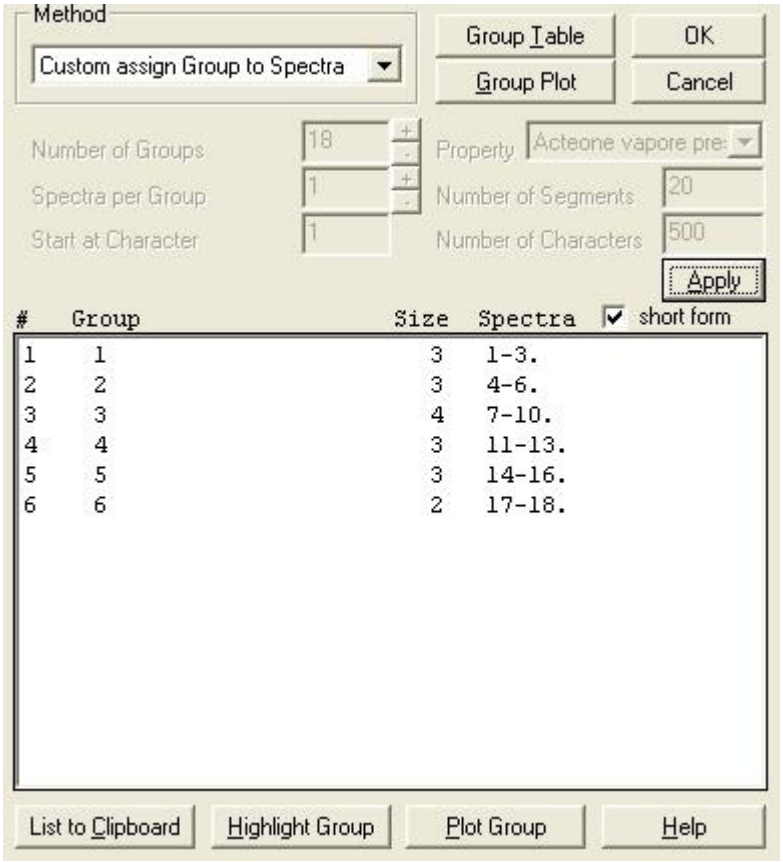
Spectra with the same name are selected for a group.

The comparison will start always from the 1st character.

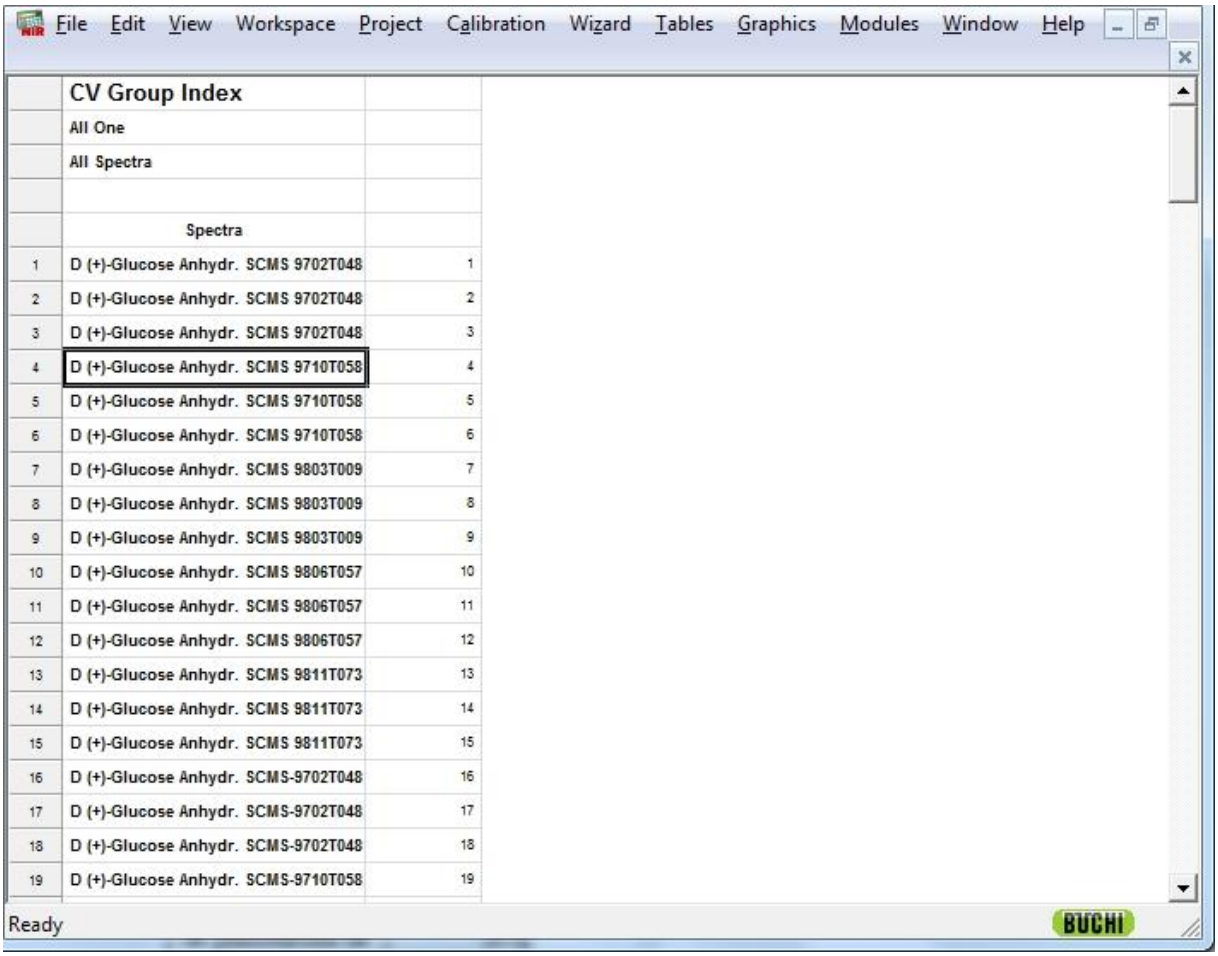
The length of spectra name will automatically be determined.



***Custom assign Group to Spectra***




For custom-defined grouping, open the Matrix CV Group Index with the button "Group Table".




Group Index numbers define to which group a spectrum is assigned.

There are 4 different types of selection possible:

| Group name | In C-Set | In CV-Group (CV Group Index > 0) |
|---|---|---|
| Unused Spectra | No | No |
| CV-unused Group | No | Yes |
| CV-permanent C-Set | Yes | No |
| CV-mutable V-Set | Yes | Yes |

Spectra with Group Index zero and not selected in the C-Set: **unused spectra.**

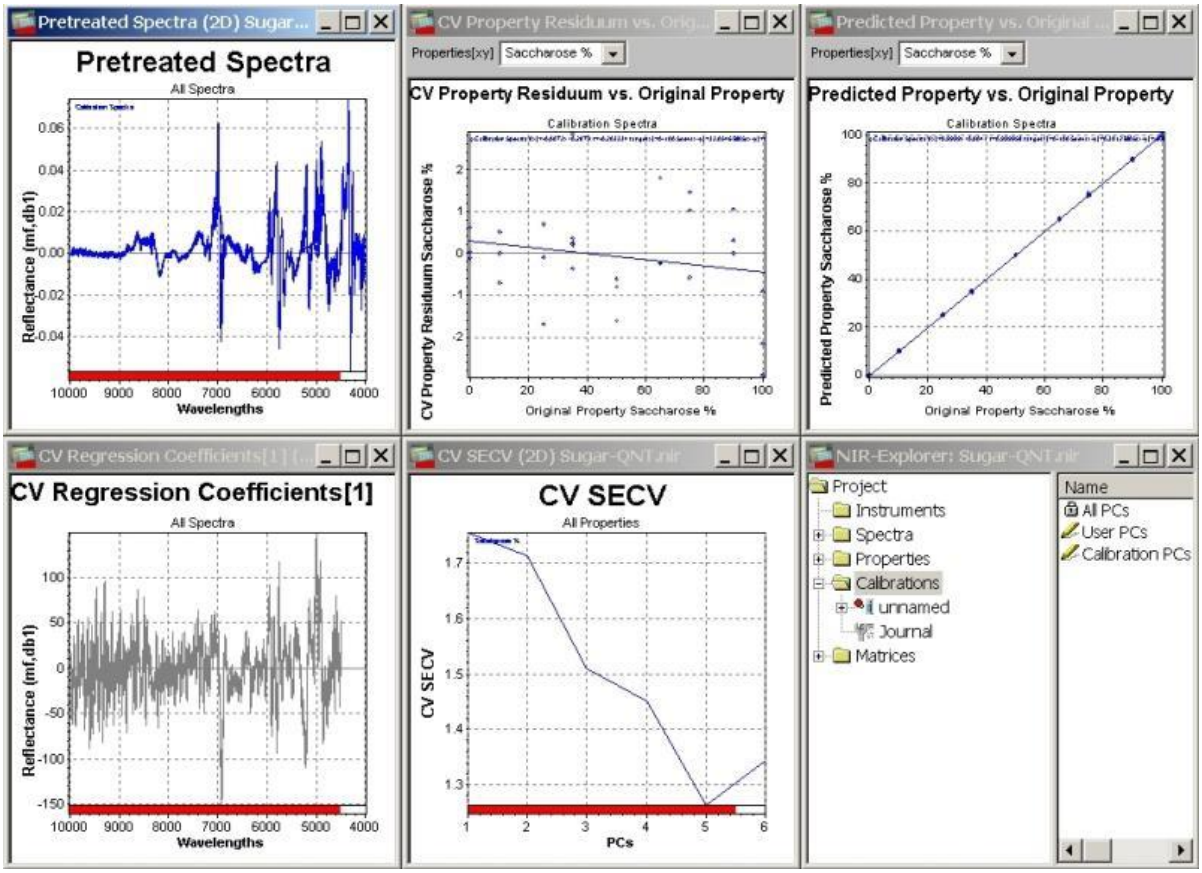
Spectra with Group Index higher than 0, but the spectra are not selected in the C-Set build the **CVunused Group.**
Advantage of this group: outliers, which belong to a spectra group, can be removed from the C-Set without making a new spectra group. In case the spectra should be removed from this group, the Group Index should be edited to zero.

Spectra with Group Index zero and selected in the C-Set: **CV-permanent C-Set**. These spectra will **never be left out** from the calibration, for each calculation they will be used within the C-Set.
Advantage of this group: spectra with extreme values (min., max.) can be permanently kept in the C-Set, which is highly recommended.
Spectra with Group Index higher than 0 and selected in the C-Set: **CV-mutable V-Set**. These spectra are removed once out of the C-Set and used as V-Set during the CV cycles. These spectra are listed in the calibration protocol with their Group number/Group Index and the spectra belonging to each group.

**CV Plots**

It is possible to show the result of each calibration step (e.g. CV Predicted Property) and also the result of the final calibration (e.g. Predicted Property).

The Overview contains the following plots:




[1] The 1st C-Set property is chosen automatically for the plots.

- Pretreated Spectra:
  it is suggested to start the first calibration without pretreatment, but try later some pretreatments and combinations.

- CV Property Residuum vs. Original Property:
  property residuum = original property -predicted property. This plot shows for each sample group the property residuum for the calibration, which had this group in the V-Set during the CV. A small CV Property Residuum and a regression coefficient between original property and predicted property residuum close to 0 shows a stable calibration. Spectra with big deviations are possible outliers and removing them from the C-Set can improve the calibration.

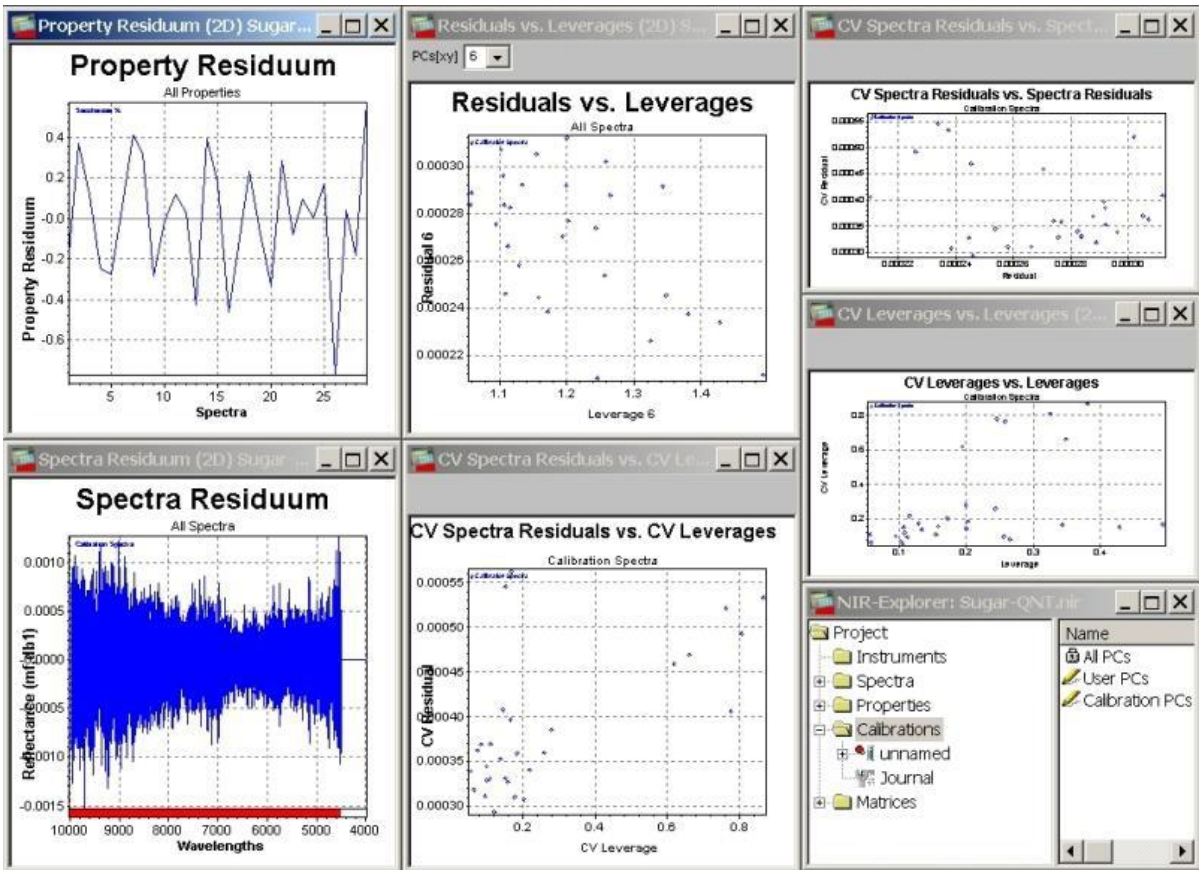
- Predicted Property vs. Original Property:
  This plot shows the result of the final calibration for the spectra which are in the calibration set (unused spectra are not visible per default: it can be changed by the user under Visibility).

- CV Regression Coefficients [1] (called the property spectra in NIRCal 4.21)
  Shows the coefficients of the linear relationship between the NIR amplitudes (of the pretreated spectra) and the selected C-property.
  [1] refers to the 1st C-property. In general for each application only one C-Set property is allowed, so for each property a separate calibration is necessary (NIRCal could handle more, but NIRWare is designed for only one property / calibration for quantitative measurements).

- CV SECV; see further details under Matrix CV SECV.
  In general the first local minimum of CV SECV for the secondary PC selection will be taken, in

case several minimum SECV exist. Use the Factor Selection Wizard. The selected number of secondary PCs will be red.

Beside these plots there are several other plots to choose outlier spectra or to make a better view of the results. The plots under  has the following plots:




- Property Residuum:
  The property residuum of each spectra showed with the final calibration.

- Spectra Residuals vs. Leverages:
  The final calibration result with the used primary PCs are showed for the calibration spectra.

- CV Spectra Residuals vs. Spectra Residuals:
  The spectra residual with the final calibration against the residuals with the CV calculation results are compared. Big difference shows, which spectra groups can be outliers.

- Spectra Residuum:
  = pretreated spectrum-reconstructed spectrum. The residuum of the final calibration is shown.

- CV Leverages vs. Leverages:
  The leverages for the final calibration against the CV leverages are ploted.

- CV Spectra Residuals vs. CV Leverages:
  Large residuals together with large leverage (points in the upper right square) are typical for possible outliers.

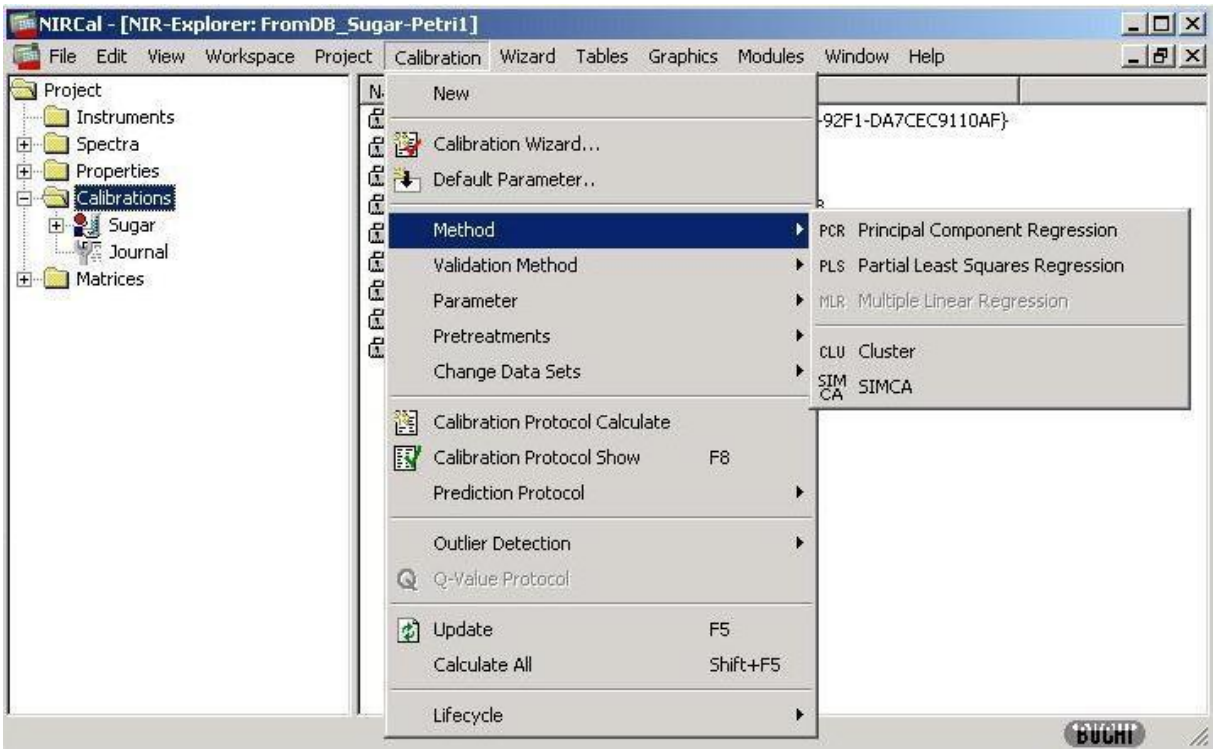
## 3.3 Selections

### 3.3.1 Calibration Method Selection

Menu: Calibration / Method / ...

Icon: PCR for Principal Component Regression / PLS for Partial Least Squares / MLR for Multiple Linear Regression

CLU for Cluster / SIM CA for SIMCA





Example: For the calibration of Saccharose the method PCR is selected and for the calibration of Lactose PLS is chosen.

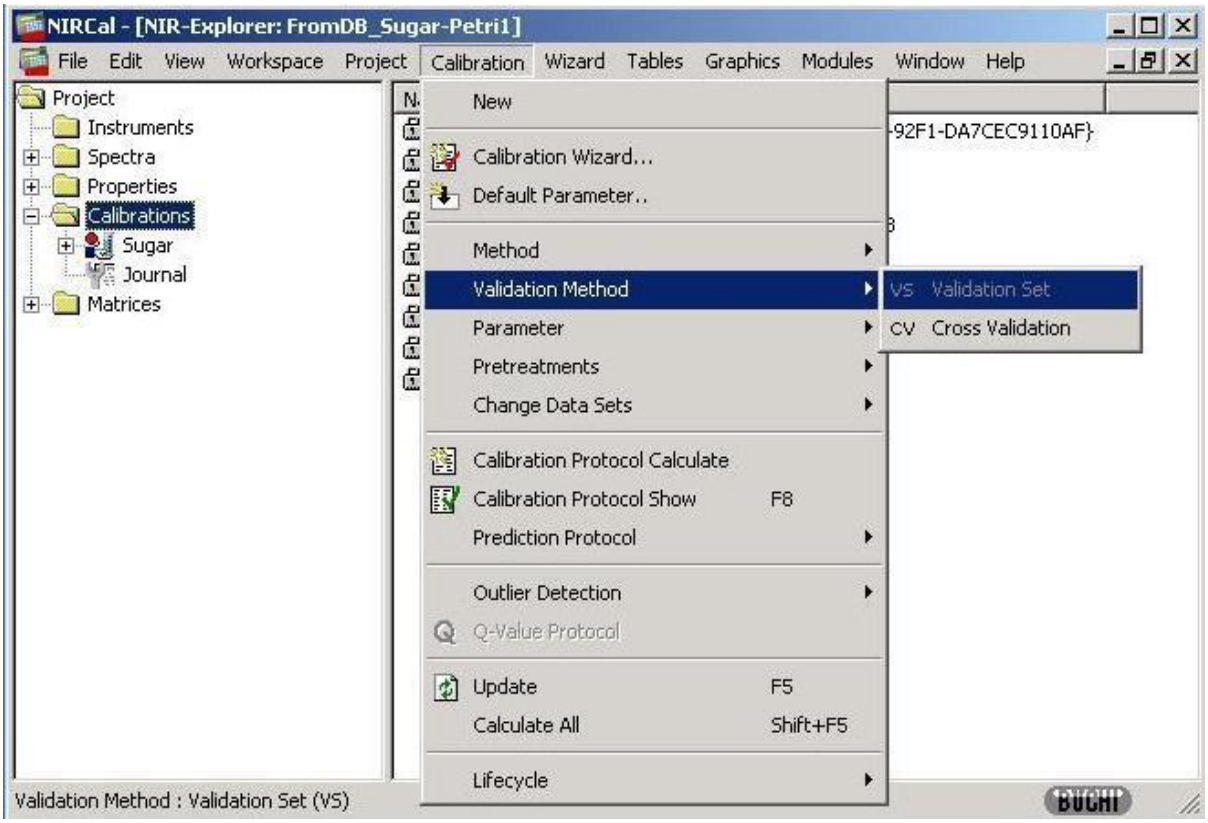
### 3.3.2 Validation Method Selection

Menu: Calibration / Validation Method / ...

Icon: VS for Validation Set

CV for Cross Validation

The default method is VS (Validation Set). For quantitative calibrations CV (Cross Validation) is available.




### 3.3.3 Data Sets

Data sets are permanent selections that are stored and loaded with the project.

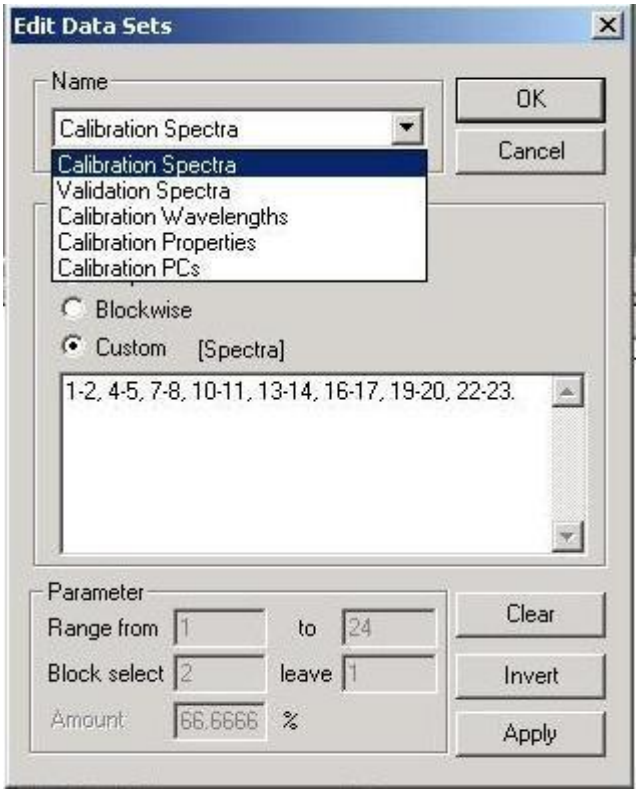
- Spectra Data Set
- Wavelength Data Set
- Properties Data Set
- PC Data Set

In case one of the above data set is left empty by the user the software automatically will create a selection with all data within the project (e.g. wavelengths property).

### 3.3.4 Edit Data Sets Dialog

Menu: Calibration / Change Data Set / Edit Data Sets...

Icon:




Under the drop-down list **Name** you can choose to edit the selections of the sets:

- Calibration Spectra
- Validation spectra
- and more

***NOTE***
Click **Apply** after a selection, click **OK** to close the Dialog.

**Invert**: a simple tool to invert the selections previously made. It is very useful when dealing with large numbers of spectra.

Method

| | |
|---|---|
| Monte Carlo Random | this selection is suggested where there is only one spectrum for each substance was measured but is not recommended (measurement mistake is not clear). 3 measured spectra of the same sample could be separated. |
| Sequence | e.g. 70% of the spectra are selected automatically. The measured spectra will be separated time dependent.<br>3 measured spectra of the same sample can be separated. |
| Blockwise | to distribute 2/3 of all measured spectra into the calibration set, 6 spectra are selected to the C-Set and 3 left out for V-Set in the range from 1 to the last measured spectra. This is the most common selection method. |
| Custom (Spectra) | user selected by spectrum number. To separate two spectra blocks, a comma and space are used. |

Parameter

| | |
|---|---|
| Range from ... to ... | The first and the last spectra index. |
| Block select ... leave ... | Number of selected and left out spectra per block. |
| Amount ... % | Amount of selected spectra in %. |

### 3.3.5 Spectra Data Set

Samples of known characteristics, both chemical and physical are used to generate calibrations. Measurement conditions should remain constant for all samples using the full spectrum range. Several sample from different batches should be collected for a robust calibration and each samples should be analysed in the laboratory with the classical method. Only acceptable samples can be used for the calibration. The measured spectra are normally divided into two sets.

- **Calibration Spectra (**spectra selected for the calibration) C-Set, about 2/3 of all measured spectra (min. 3). Only the C-Set spectra are used for the calculation of the principal components and the calibration limits.
  Generally, the calibration spectra should contain all “extreme information” to define the limits of acceptance. For quantitative calibration the spectra with the highest and lowest property values always belong to the calibration spectra.

- **Validation Spectra**, about 1/3 of all measured spectra selected for the internal validation V-Set.
  Only if a calibration treats the validation spectra equally to the calibration spectra, the settings are considered as OK

***NOTE***
These two groups of spectra should be:

- independent from each other: all spectra of one sample should belong to the same set;
- no overlapping allowed.

It is **possible to leave out** some spectra from the C- and from the V-Set, these spectra are in the **Unused Set** = U-Set:

C-Set + V-Set + U-Set = All Spectra

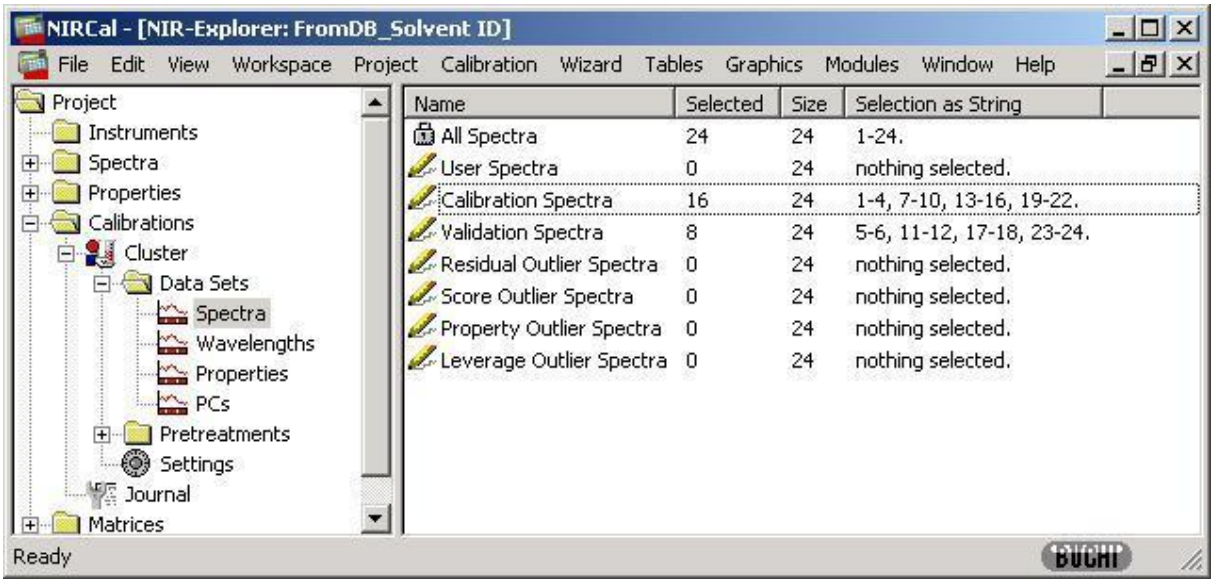
The U-Set is visible in the calibration results, but will not influence the calibration and validation results e.g. will not be taken for the Q-Value calculation. The calibration protocol is stored within the calibration.

C- and V-Set can be selected by the user in different ways:

- in the NIR-Explorer;
- in the Property table;
- in the graphic.

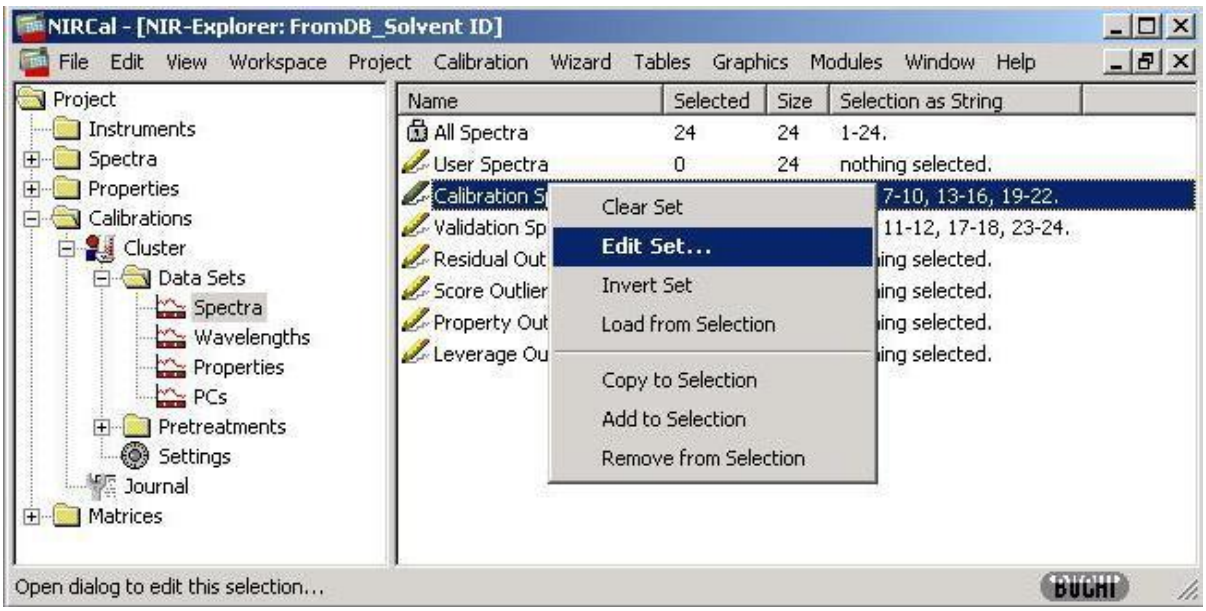
**Spectra Selection in the NIR-Explorer**

Open NIR-Explorer

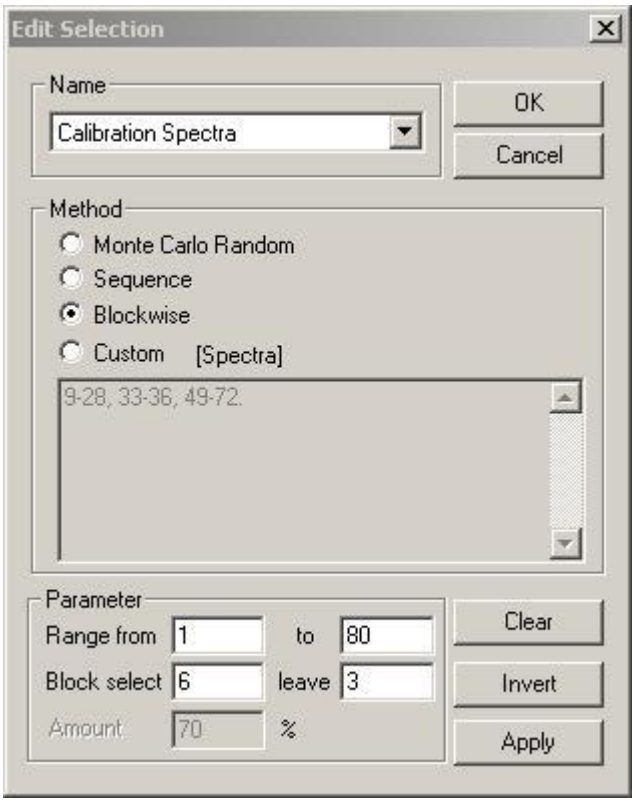



Steps sequence:

1. Open the folder "**Calibrations**" by clicking on the box + in front of the folder or double clicking the folder
2. Open the active calibration, indicated by the red dot.
3. Open the folder "**Data Sets**".
4. Select "**Spectra**".




5. Select "**Calibration Spectra**" in the right part of the window. To open the Edit window press the right mouse button and click on **Edit Set** or double click on the selected line. Choose **Blockwise** with Block select 6 and leave 3, press **OK** for applying this selection.
For the " **Validation Spectra**" the **Blockwise** method is selected, press first **Apply**, than **Invert** and press **OK** to get the rest.




**Selection methods:**

**Monte Carlo Random**: this selection is suggested, when only one spectrum for each substance was measured, which is not recommended (measurement mistake is not clear). For 3 spectra / samples this method is not suggested, while the 3 spectra can be randomly in C- or V-Set.

**Sequence**: e.g. 70 % of the spectra are selected automatically into the C-Set. 3 spectra / sample can be separated!

**Blockwise**: to distribute 2/3 of all measured spectra into the C-Set, 6 spectra are selected to the C-Set and 3 left out for the V-Set of the range from 1 to 80. This is the most common selection method.

**Custom**: spectra selected by spectrum number in the project. To separate two spectra blocks, a comma and space are used.

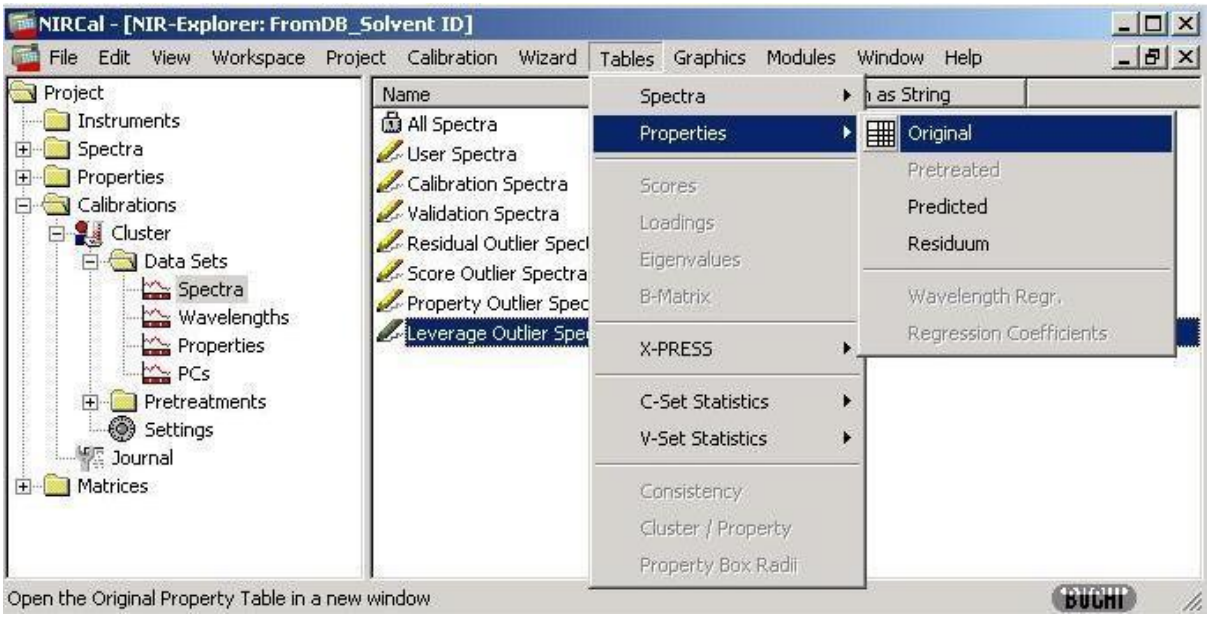
***NOTE***

When each sample has been measured three times, all three spectra should be designated to either the calibration or validation set.
Therefor it is ideal to use **Blockwise** selection e.g. if each sample was measured with 3 spectra block select 6 and leave 3 can be used. Blockwise selection is only recommended for qualitative calibrations.

Choose Blockwise with Block select 6 and leave 3, press OK for applying this selection for the C-Set. For the " **Validation Spectra**" the Blockwise method is selected, press first **Apply**, than **Invert** and press **OK** to get the rest.

**Spectra Selection in the Property Table**

Open the table Original Property in Menubar: Tables / Properties / Original:




Original Property (Table) Sugar-QL.nir

**Original Property**

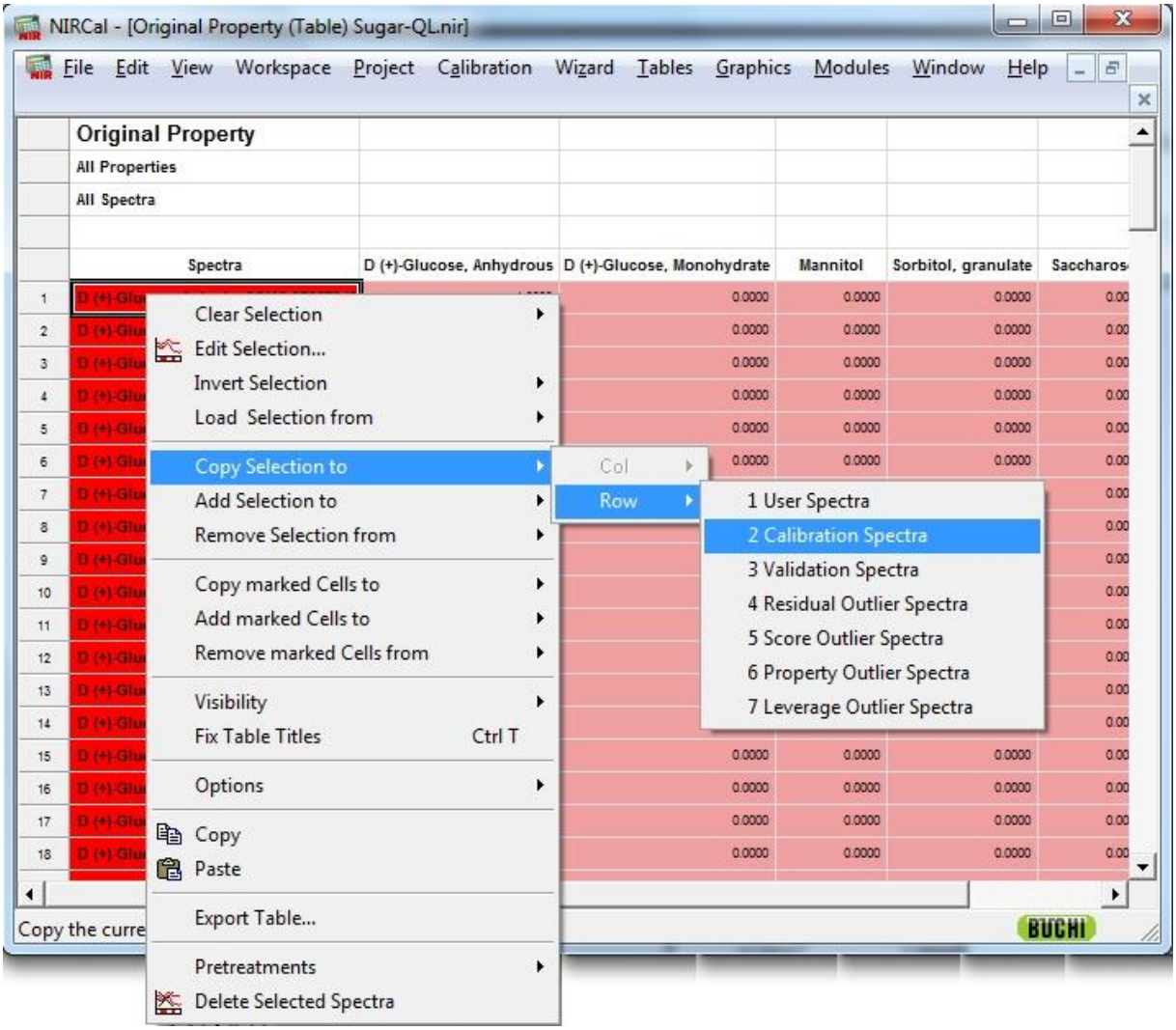
All Properties

All Spectra

| | Spectra | D (+)-Glucose, Anhydrous | D (+)-Glucose, Monohydrate | Mannitol | Sorbitol, granulate | Saccharose |
|---|---|---|---|---|---|---|
| 1 | D (+)-Glucose Anhydr. SCMS 9702T048 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2 | D (+)-Glucose Anhydr. SCMS 9702T048 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 3 | D (+)-Glucose Anhydr. SCMS 9702T048 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 4 | D (+)-Glucose Anhydr. SCMS 9710T058 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 5 | D (+)-Glucose Anhydr. SCMS 9710T058 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 6 | D (+)-Glucose Anhydr. SCMS 9710T058 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 7 | D (+)-Glucose Anhydr. SCMS 9803T009 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 8 | D (+)-Glucose Anhydr. SCMS 9803T009 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 9 | D (+)-Glucose Anhydr. SCMS 9803T009 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 10 | D (+)-Glucose Anhydr. SCMS 9806T057 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 11 | D (+)-Glucose Anhydr. SCMS 9806T057 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 12 | D (+)-Glucose Anhydr. SCMS 9806T057 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 13 | D (+)-Glucose Anhydr. SCMS 9811T073 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 14 | D (+)-Glucose Anhydr. SCMS 9811T073 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 15 | D (+)-Glucose Anhydr. SCMS 9811T073 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 16 | D (+)-Glucose Anhydr. SCMS-9702T048 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 17 | D (+)-Glucose Anhydr. SCMS-9702T048 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 18 | D (+)-Glucose Anhydr. SCMS-9702T048 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 19 | D (+)-Glucose Anhydr. SCMS-9710T058 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 20 | D (+)-Glucose Anhydr. SCMS-9710T058 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 21 | D (+)-Glucose Anhydr. SCMS-9710T058 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 22 | D (+)-Glucose Anhydr. SCMS-9803T009 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 23 | D (+)-Glucose Anhydr. SCMS-9803T009 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 24 | D (+)-Glucose Anhydr. SCMS-9803T009 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

The table consists of the spectrum number, the name of spectrum with batch number and the property membership (1 indicates: it belongs to a property, 0 indicates: it does not belong to a property).

Creating the **selection**: mark the first selected row with the mouse, **press the left button only once**. Press the "**Shift-key**" and **double click on the last marked row**. All selected spectra are highligthed in red colour.




To **remove** spectra from the selection: mark the selected row with the mouse, press the **left button** only once. Press the "**Ctrl-key**" and double click on the row.
When the selection is created, the spectra need to be copied to the calibration spectra and/or the validation spectra:

1. press the right mouse button in the graphic and **Copy Selection to - Row - Calibration Spectra**;
2. for setting the Validation Spectra select in the Popup-Menus the command **Invert Selection**. The rest of the spectra will be selected;
3. copy this selection in the Popup-Menus with the command **Copy Selection to - Row - Validation Spectra.**

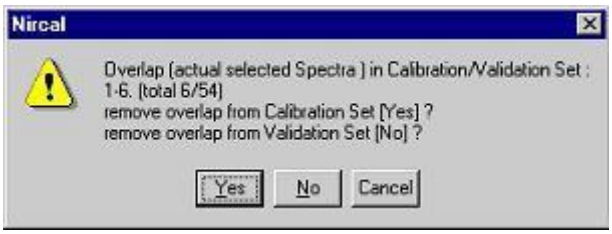
***NOTE***
The menu "Copy Selection to" **overwrites** the existing selection in the data sets.

***Add and Remove Selection from Datasets***

Mark the spectra rows of the position to be changed in the selection. Press the right mouse button and add to or remove this selection from the desired data set. In this way, the selection made before will be enlarged / reduced with only the new selected data.

Spectra should not be designated to both, Calibration and Validation-Set.

Should this happen the system forces the user to make a clear decision.
Clicking "Yes" will remove the overlapping spectra from the **Calibration Set** and keep them in the Validation Set.




**Spectra Selection in Graphic**

A very practical way to make the spectra selection is in the calibration plot using the mouse. This is especially useful for the **quantitative** calibration because in the plot "Predicted Property vs. Original Property" the concentration distribution of the C- and V-Set is visible.

Click on the  Overview-Button in the NIRCal-Toolbar. If the C-Set and the V-Set are empty, a message will appear. Confirm by clicking on the Yes button to get the Overview.

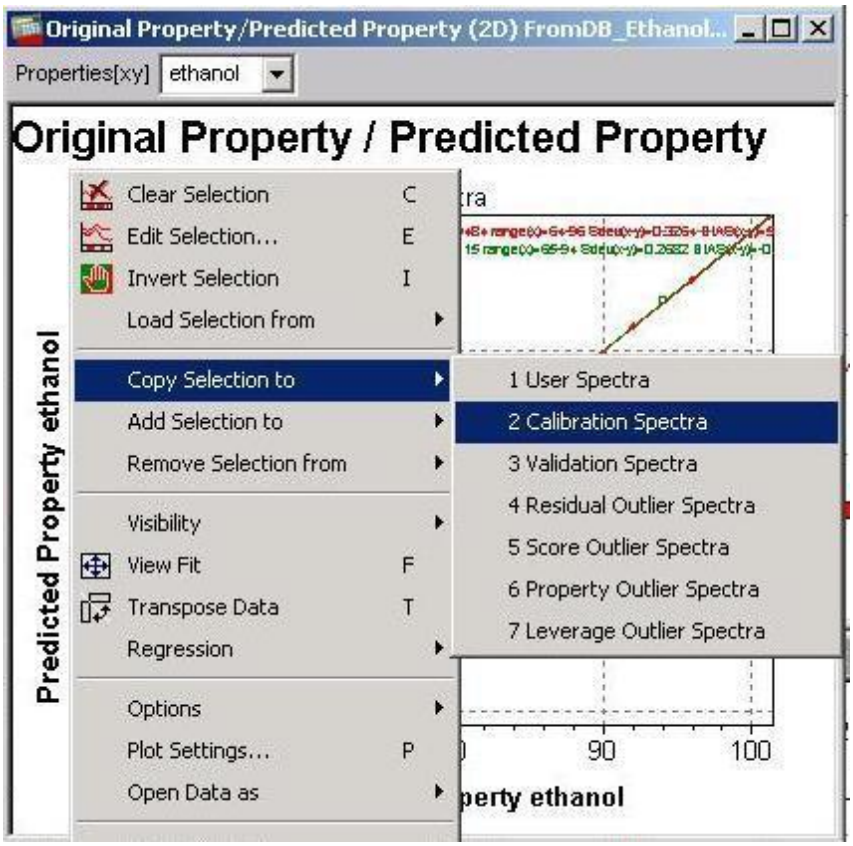
Enlarge the window ”Predicted Property vs. Original Property" in the 3rd column.
The ”Predicted Property vs. Original Property” can be opened in the Menubar: Graphics / Properties / Original vs. Predicted as well:

Press the “Minus”-button on the keyboard to expand the X-and Y-ranges of the graphical display.
Clear the selection made before by pressing the right mouse button and click on **Clear Selection**.
To create selections in a graph, the function of the mouse must be changed from the Zoom function to the Window Select function.

- Press the right mouse button and choose "Mouse Select" under "Options". The symbol of the mouse will change from  to  .
- Draw a box around the spectra with the mouse keeping the **left mouse** button **pressed**.
- Draw boxes around all the other spectra you want to select by **holding the “Shift-key**”. All selected spectra will be highlighted in red as soon as the left mouse button is released.

***NOTE***
While adding spectra DO NOT release the “Shift-key”.

To remove spectra from the selection: Draw a box around the spectra you wish to deselect with the left mouse button pressed while holding the “Ctrl-key”. Release the left mouse button still holding the “Ctrlkey”. All remaining selected spectra will stay highlighted in red.




- When the selection is created, the spectra can be copied to the calibration spectra, the validation spectra or the user spectra.

1. press the right mouse button in the graphic and **Copy Selection to - Row - Calibration Spectra**;
2. for setting the Validation Spectra select in the Popup-Menus the command **Invert Selection**. The rest of the spectra will be selected;
3. copy this selection in the Popup-Menus with the command **Copy Selection to - Row - Validation Spectra.**

***NOTE***
The function “Copy Selection to” overwrites the existing selection in the data sets.

**Add and Remove Selection from Dataset**

Mark the spectra, which should be removed from a selection.

Press the right mouse button and use "Add / Remove Selection to". This will enlarge or reduce the previously selection with the new spectra.

***NOTE***
By copying, adding and removing spectra to (from) the sets, **a homogeneous distribution** between calibration and validation spectra over the whole concentration range can be achieved.

### 3.3.6 Wavelength Data Set

The exact wavelength / wavenumber range depends on the instrument type used for the spectrum measurement and their settings (e.g. resolution).
The selected wavelength / wavenumber range depends on the application and the measuring option used. The calibration wavelengths define the spectrum range used by the mathematical algorithms, PCA, PCR or PLS.

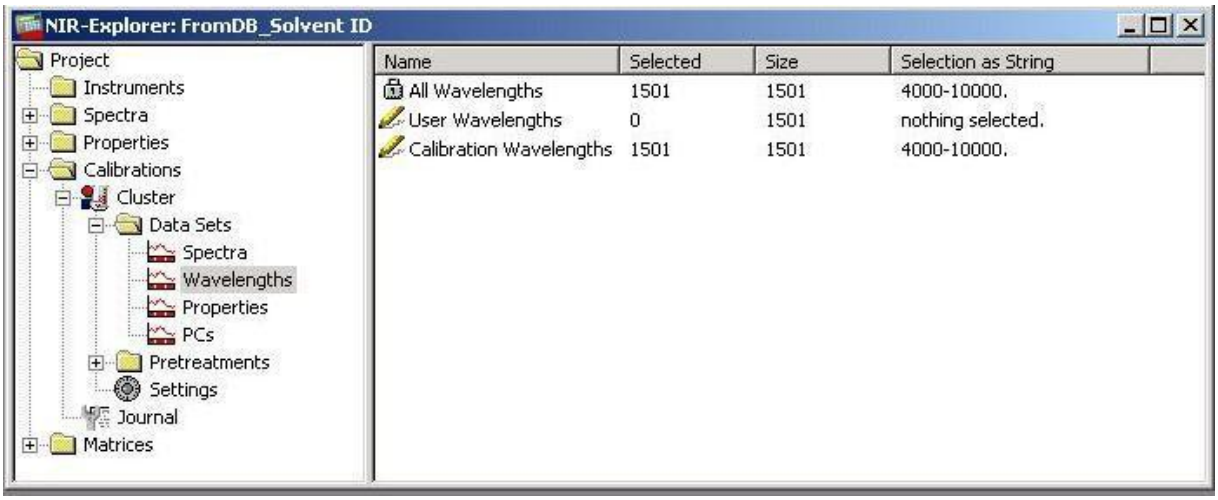
Calibration Wavelengths are the wavelengths used to create the calibration model.
Removing certain wavelengths from a calibration can lead to some improvement.

***NOTE***
In general, the calibration wavelength / wavenumber range should be as wide as possible.

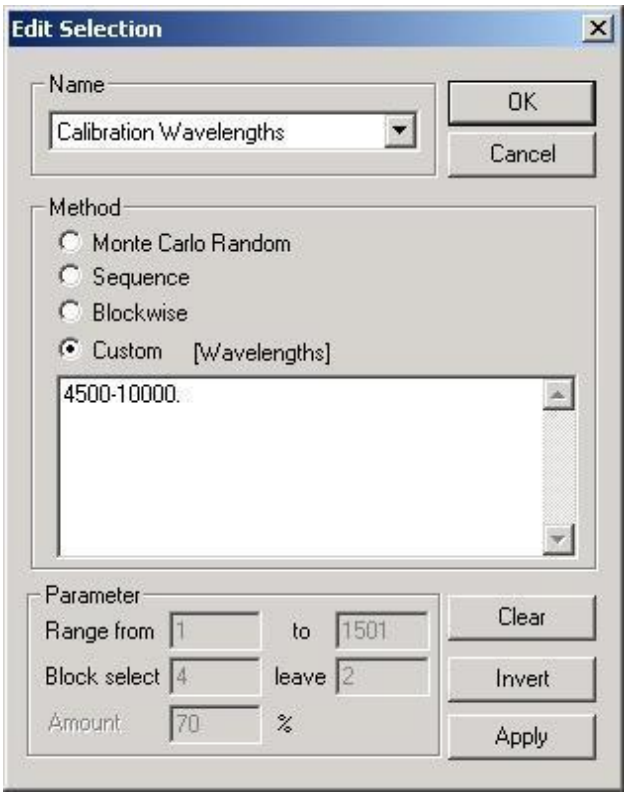
Suggested wavenumber range for NIRFlex N-500 for measuring options with

- Solids and Liquids: 4'000-10'000 $cm^{-1}$;
- Fiber Optics: 4'500-10'000 $cm^{-1}$.
- Solids with Tablet Accessory: 6'000-11'520 $cm^{-1}$.

**Selecting the Calibration Wavelengths in the NIR-Explorer**



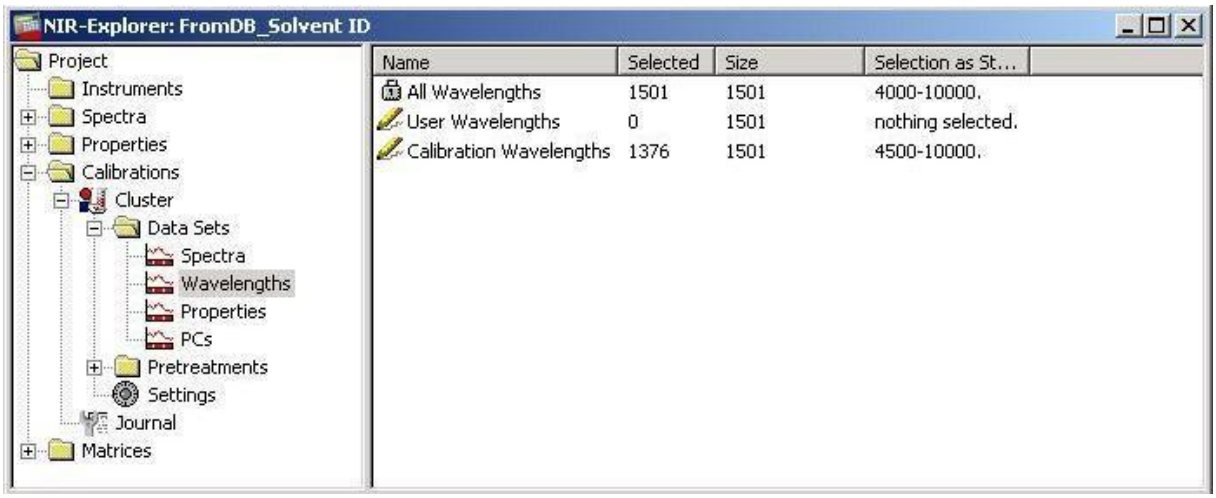

To make the selection use the NIR-Explorer: double click "**Calibration Wavelengths"** and the Edit selection dialog appears




Steps sequence:

1. Select Custom;
2. Type in the selection;
3. Click on OK.

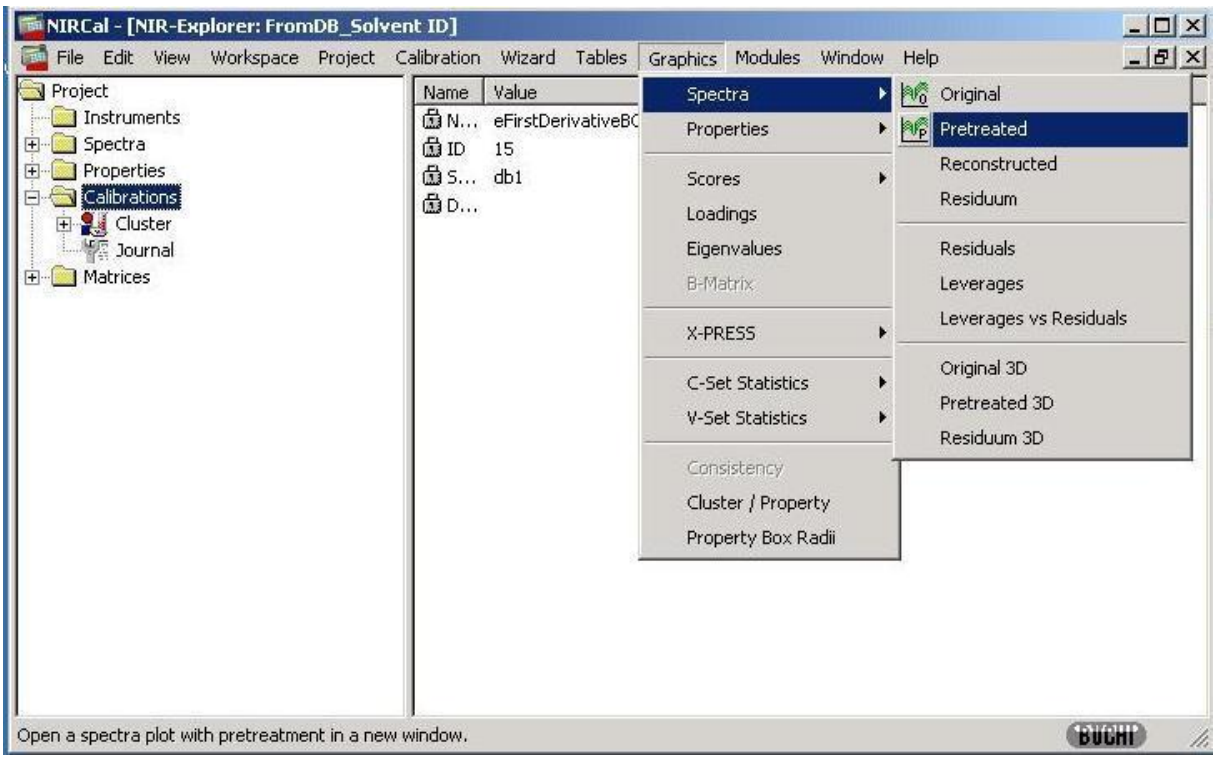
***NOTE***
It is recommended to use only **custom** selection.




In the project, the selected wavenumber as data points and range are shown (here: 1376 data points in the range of 4'500-10'000 $cm^{-1}$ ).

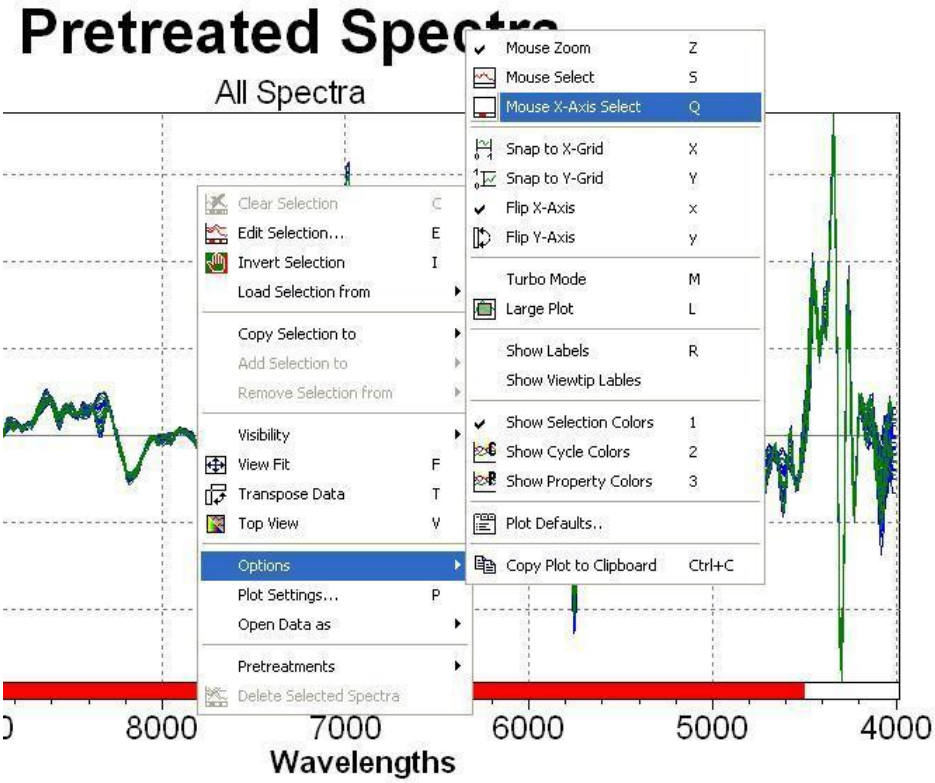
**Selecting Calibration Wavelengths using Graphics**

To review and select the suitable calibration wavelength range, several graphics can be used.
An example is shown with the **pretreated spectra**, but it can also be done in the original spectra, loading or property wavelength regression graphic in the same way.

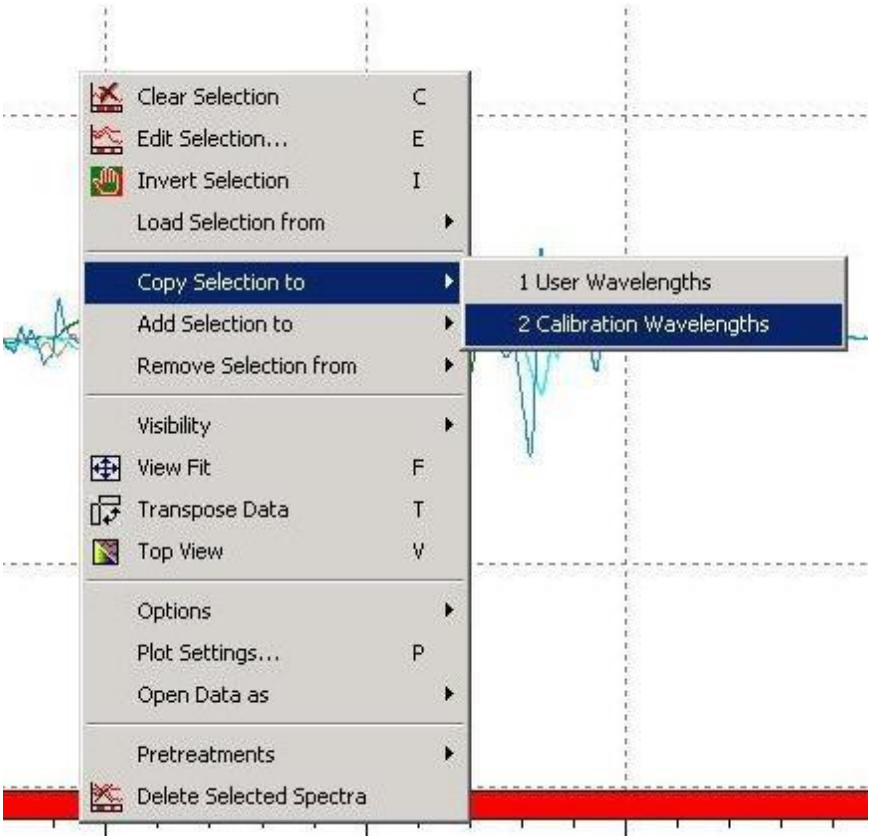



Open the pretreated spectra in Menubar: **Graphics / Spectra / Pretreated**.

When the graphic is opened, press the right mouse button and choose:**Options / Mouse X-Axis Select**.




Now it is possible to select the wavelengths with the mouse. The cursor position as wavenumber can be read in the status bar. Keep the left button pressed for selection. The selected range is marked with red color.
Copy the selected range into the calibration wavelengths with the right mouse button Popup-Menu: **Copy Selection to / Calibration Wavelengths**.

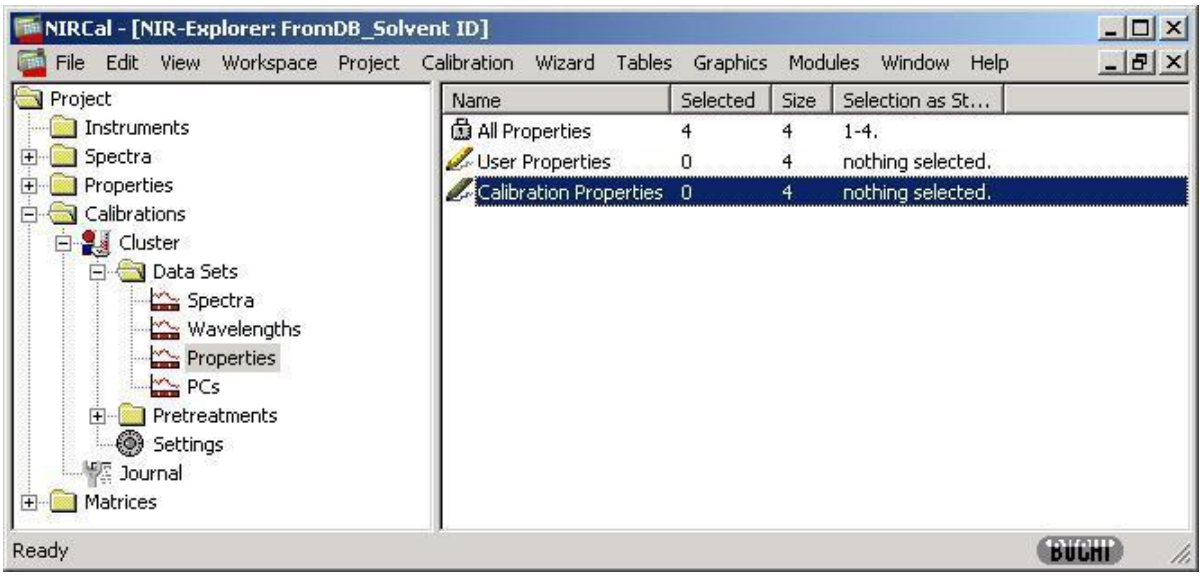


### 3.3.7 Properties Data Set

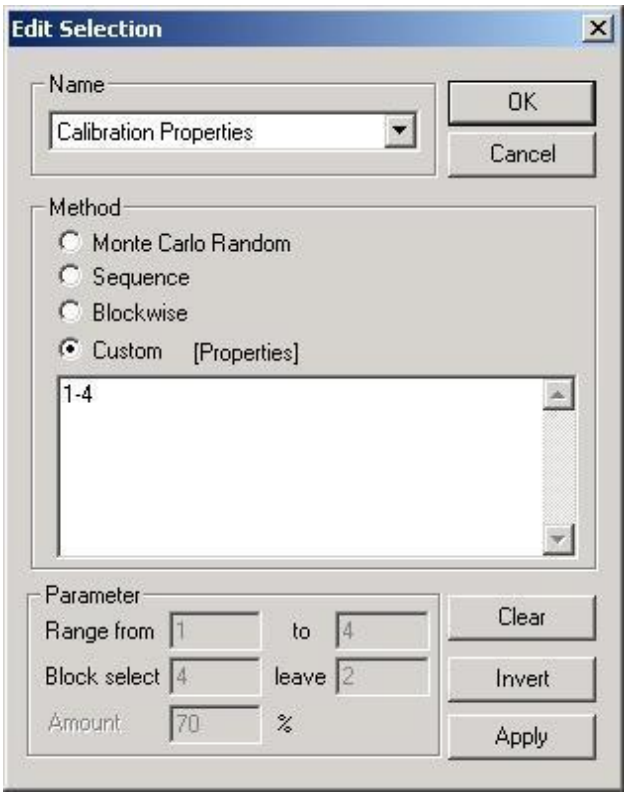
In qualitative methods the calibration properties are the substances that are required to be identified in the application.
In cluster calibration the mathematical algorithm of PCA will be applied for the selected properties. Normally all qualitative properties are used in the calibration.

In quantitative methods the calibration property (single property calibration) is the selected parameter which should be predicted.

**Selecting the Calibration Properties in the NIR-Explorer**

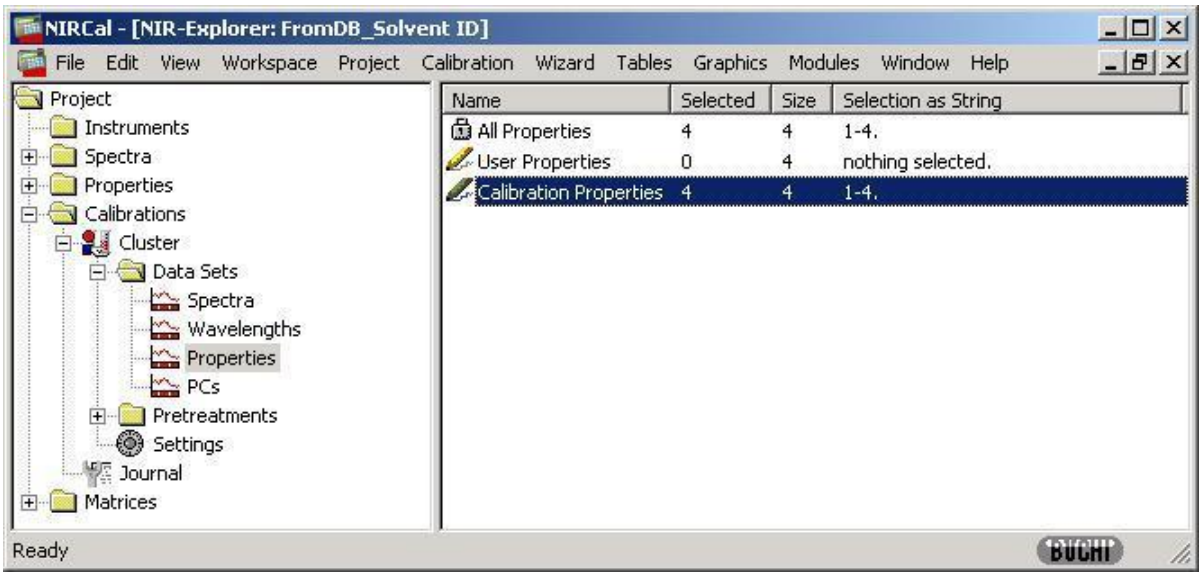



To make the selection use the NIR-Explorer: double click "**Calibration Properties**" and the Edit selection dialog appears.



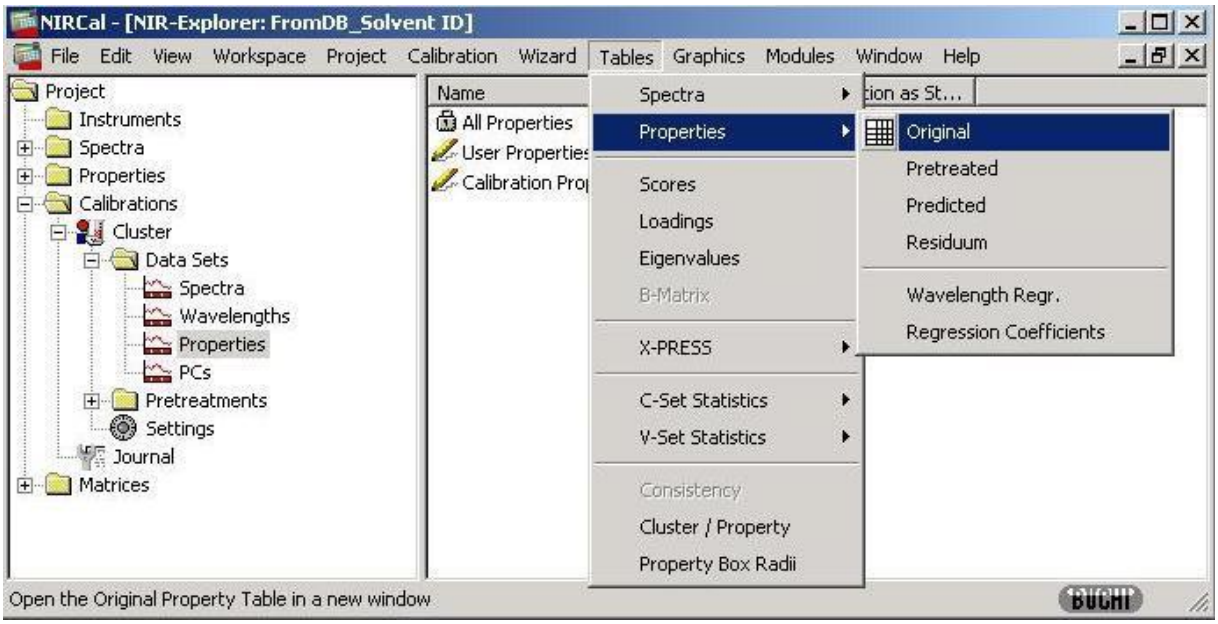

Steps sequence:

1. Select Custom;
2. Type in the selection;
3. Click on OK.

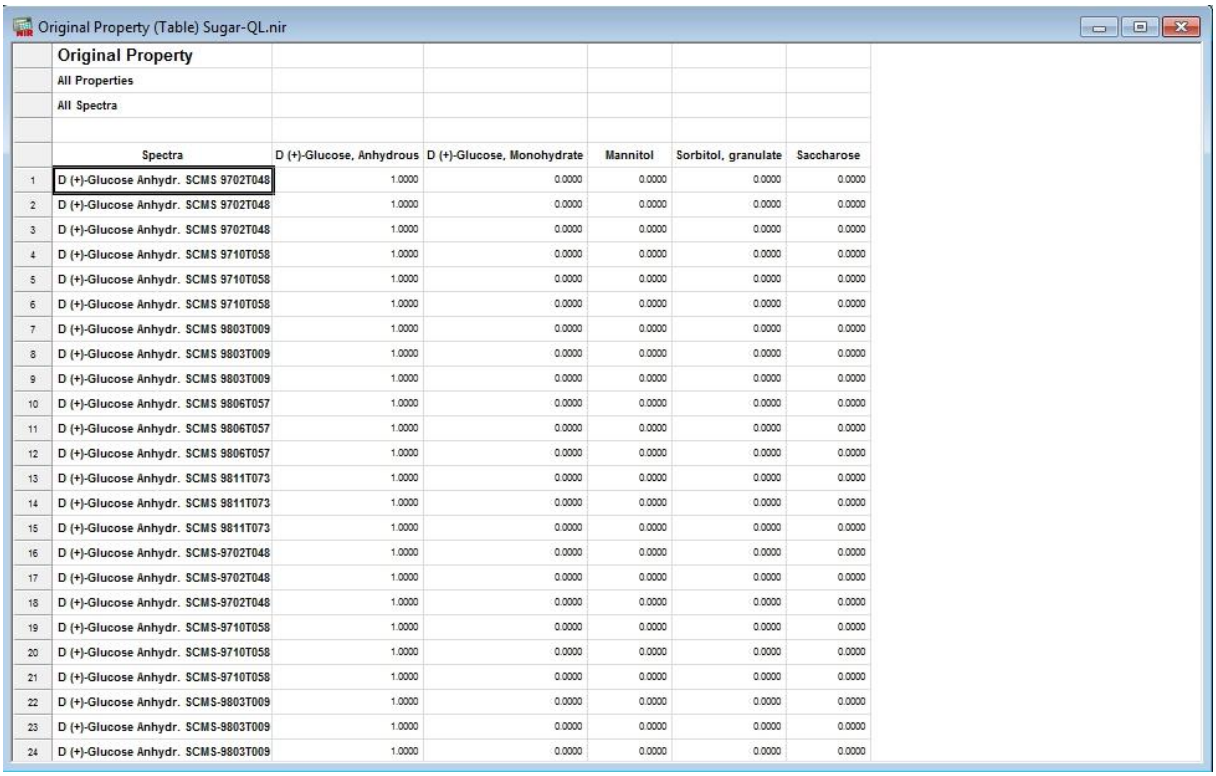



In the project the selection is shown

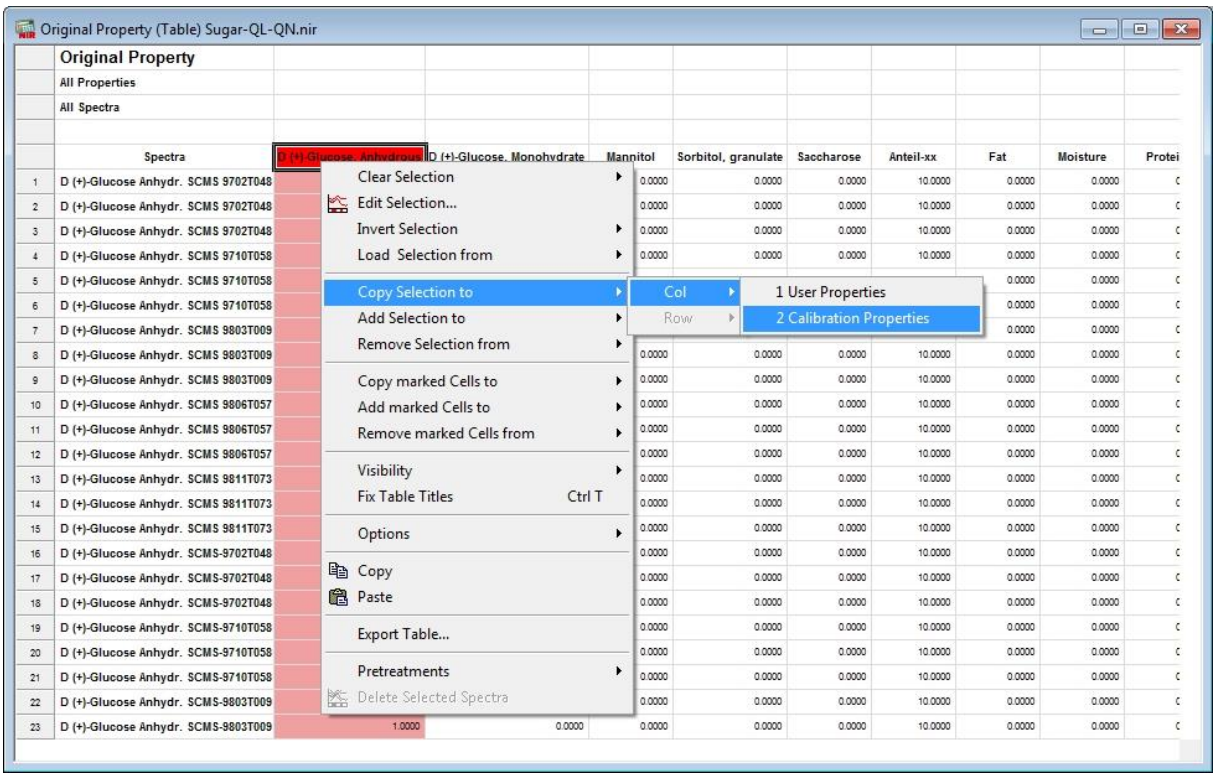
**Selecting the Calibration Properties using the Property Table**



Open the property table.




Select the desired column.




Copy the selection into calibration properties.

### 3.3.8 PC Data Set

Principal Components or Loadings (old name Factors) are artificial difference spectra.

All PCs: Number of selected primary PCs.

User PCs: use this set as a selection buffer (e.g. for visibility).

Calibration PCs: these secondary PCs are used to create the calibration model. Calibration PCs or secondary PCs form a subset of the primary PCs of the calibration.

## 3.4 Calibration Wizard

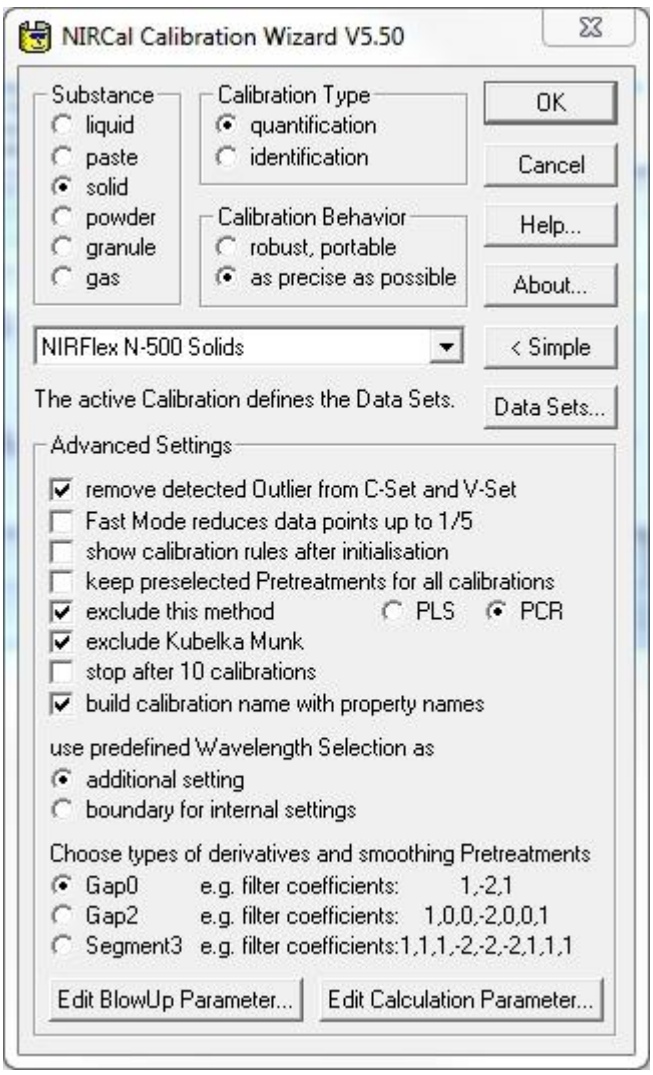
### 3.4.1 Calibration Wizard

Menu: Wizard / Calibration Wizard

Icon: 

The quality (robustness, sensitivity, selectivity, portability...) of a calibration for quantification and/or identification mainly depends on the data selections: choice of calibration and validation spectra, wavelength range, data pretreatments and PCs used.
The chemometric software NIRCal offers the calibration wizard for an easy and fast generation of calibrations without a profound understanding of chemometrics. The **automatic Calibration Wizard** guides the user. While the user only has to answer a few questions about the used samples, the installed sampling options or the expected quality of the calibration, the automatic calibration uses its knowledge base for selecting spectra, wavelength ranges or the number of PCs. The pretreatment selector for permutation and calculating several pretreatment combinations is an integral part of the calibration wizard. After all calibration is calculated the summary is documented and the five best calibrations are stored, all other deleted automatically. The calibration wizard can be started without any data selection before, but a precalculation can be useful: the wizard never selects properties! To do the data selection press the **Data Sets** button.

The Calibration Wizard with the **Advanced** option:




1. Select type of substance.
2. Select type of calibration.
   Important for PCR/PLS or CLU method selection!
3. Select behavior of calibration.
4. Select used measuring option.
5. Start the wizard with OK.

**Advanced Settings:**

| | |
|---|---|
| remove detected Outlier from C-Set and V-Set | Based upon outlier detection module with c=2.5, removed spectra are listed in the calibration protocol.<br>Not in all cases it is advantageous to use this option. |
| Fast mode reduces data points up to 1/5 | Every 5. wavelength value will be used to reduce the calculation time.<br>Not suggested for the end calculation! Use "Turbo Mode" (M) in plots e.g. Loading plot. |
| show calibration rules after initialisation | The used pretreatment, wavelength and method selection is showed; "**Use all rules**" with up to 2000 combinations are available (the time of calculation will increase dramatically).<br>  |
| keep preselected Pretreatments for all calibrations | The pretreatment or pretreatment sequence, used in the active calibration, is kept and the pretreatments from the wizard are added afterwards. |
| exclude this method | For quantitative calibration the PLS or PCR method can be eliminated to reduce the calculation time. |
| exclude Kubelka Munk | Kubelka Munk will not be used if this option is checked. |
| stop after 10 calibrations | Calculating 10 calibrations reduce the time for the calculation (mainly for demo purposes). |
| built calibration name with property names | Adequate for single property calibrations; to be switched off for a larger number of properties. |

The **wavelength selection** can be influenced in one of these ways:

| | |
|---|---|
| additional setting | The wavelength selection used in the active calibration, will be used additional to the predefined selection. |

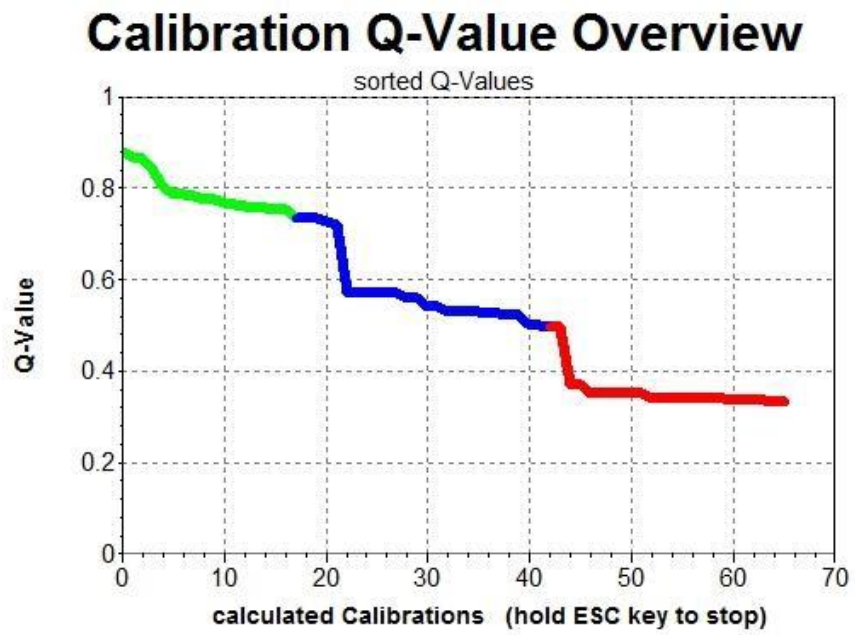
or

| | |
|---|---|
| boundary for internal settings | The wavelength selection is never bigger, as the selection in the active calibration. The boundary selection can also include gaps. |

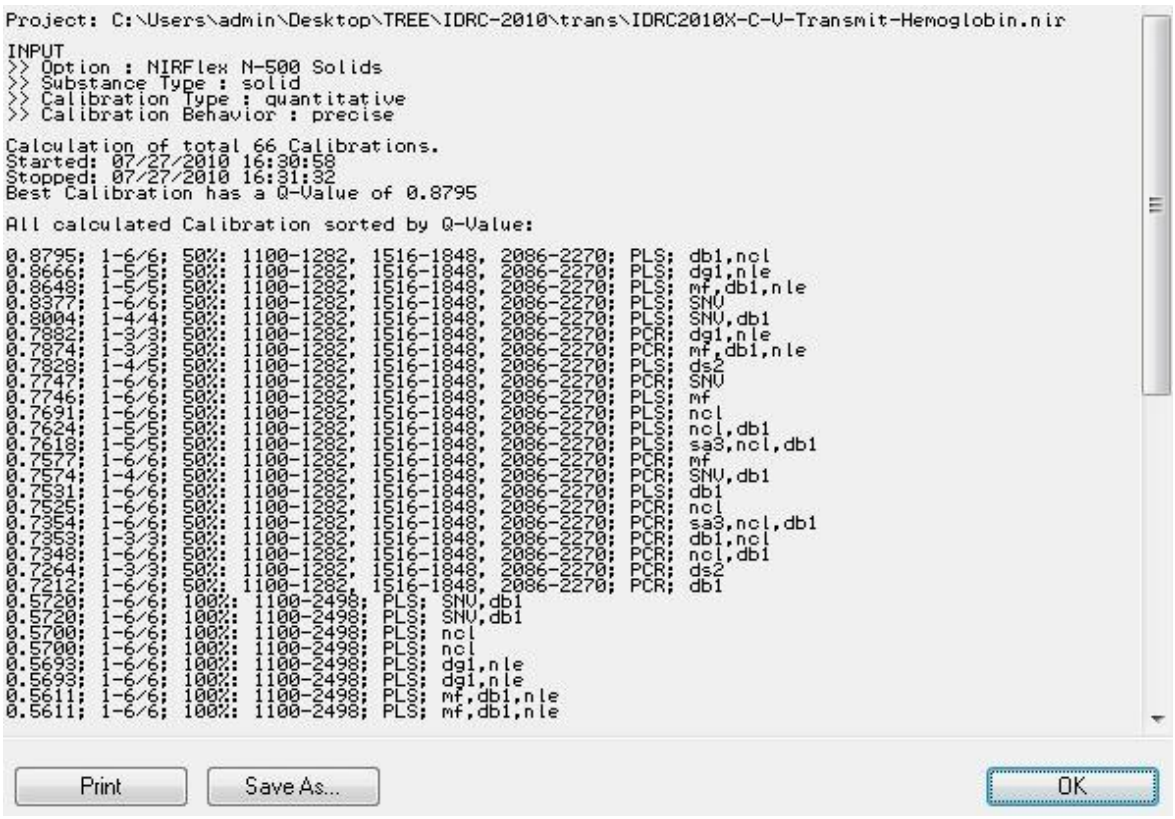
For the pretreatments of derivatives and smoothings there are 3 possibilities:
- Gap0: all data points are in the calculation;
- Gap2: each 3rd data point is in the calculations;
- Segment3: there is an additional 3 point smoothing (as3) for the Gap2 pretreatment.

With these selections predefined wavelength and pretreatment combinations and the calculation algorithm will be activated.

The wizard can be stopped with the keeping the **Esc** button pressed.




The Q-Value is a measure of the quality of the calibrations. It ranges from 0 to 1. The higher the Q-Value, the better the calibration.
Q-Value 1 is theoretical, in practice not achievable, if the value is higher as 0.75 (green), the calibration is acceptable, between 0.5-0.75 (blue) the calibration is useable, but not very accurate. Calibrations with Q-Value below 0.5 (red) should be inspected very carefully before routine use. The general observation is that qualitative calibrations yield considerably better Q-Value than quantitative calibrations. The Q-Value is a very good tool for the judgement of calibrations especially when comparing different calibrations.
After all calibrations have been calculated the results are summarized, the calibrations are sorted by the Q-Value. In addition to the calculated property the used secondary (calibration) / primary factors = PCs, the selected wavelengths regions, the used algorithm and pretreatments are listed as well.



It is possible to save or print this list by clicking on the appropriate buttons.

The ten best calibrations are kept in the project. The calibration with the highest Q-Value is set active. For this calibration the results are summarized using the Overview. It is possible to inspect the active calibration manually.

***NOTE***
It is the responsibility of the user to judge the calibrations and to release them for routine use. The Q-Value is one among other helpful tools for the judgement of the quality of the calibrations.
It is important to test the calibration with independent well characterized samples

# 3.5 Pretreatments

## 3.5.1 Pretreatments

NIR spectra are influenced by various parameters. Variations of chemical and physical properties of samples as well as the measurement process and changes at the spectrometer will have an influence on the spectrum. These effects will mainly appear as problems with:

- overlapping absorption bands;
- non-linearity;
- light scattering;
- random noise;

One possibility to overcome these problems is to improve the signal by mathematical transformations of the spectra using pretreatments. They are used to improve the quality of the spectra and to minimize unwanted effects.

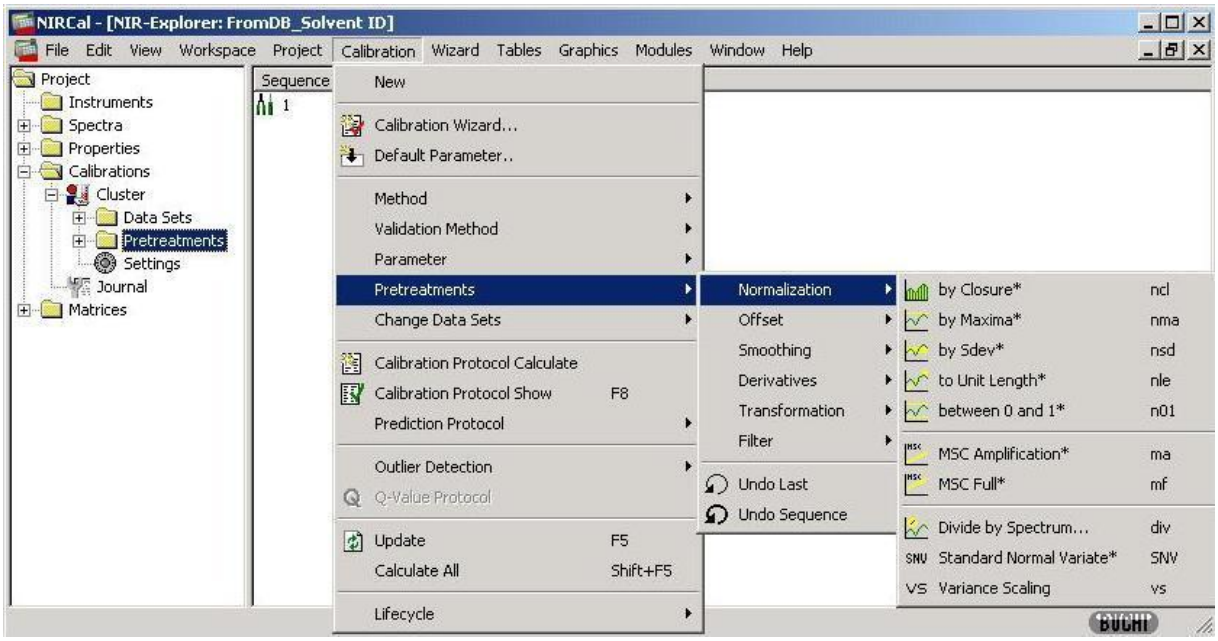
***NOTE***
Pretreatments do not change or affect the original spectra.

NIRCal provides a variety of data pretreatments, there are 34 pretreatment possibilities available in 6 groups. Each pretreatment can be combined with another and the **order of combination** is also important. Some pretreatments, which have a star "*" behind the name, are wavelength dependent. In this way there are a hugh number of combinations available. The size of a pretreatment sequence is only limited by the memory.

***NOTE***
According to our experience it is not suggested to use more than 3 pretreatments. Be aware of trashing your data to nonsense by misusing pretreatments.

**Applying Pretreatments**

The pretreatment selection is available in the Menubar: Calibration / Pretreatments.




The pretreatments can be selected in the toolbar via Icons as well; or select Pretreatments in the NIR Explorer and use the right mouse button; or select Pretreatments in the NIR Explorer, or in the Pretreated Spectra Plot and use the right mouse button.

**Removing Pretreatment**

 with Undo Last it is possible to cancel the last pretreatment;

 with Undo Sequence the whole sequence of pretreatments will be canceled.

### 3.5.2 Available Pretreatments

| Pretreatment | Type | Short |
|---|---|---|
| Normalization | by Closure* | ncl |
| | by Maxima* | nma |
| | by Sdev* | nsd |
| | to Unit Length* | nle |
| | between 0 and 1* | n01 |
| | MSC Amplification * ** | ma |
| | MSC Full* ** | mf |
| | Divide by Spectrum | div |
| | Standard Normal Variate* | SNV |
| | Variance Scaling** | vs |
| Offset | Subtract DC* | dc |
| | Shift Negative to Zero* | n2z |
| | MSC Offset* ** | mo |
| | Add constant | +c |
| | Mean Centering ** | mc |
| | Subtract Spectrum | sub |
| Smoothing | Average 3 points | sa3, sa3g2 |
| | Average 9 points | sa9, sa9g2 |
| | Savitzky-Golay 9 points | sg9, sg9g2 |
| Derivatives | 1st BCAP 5 points | db1, db1g2 |
| | 1st Taylor 3 points | dt1, dt1g2 |
| | 1st Savitzky-Golay 9 points | dg1, dg1g2 |
| | 2nd BCAP 3 points | db2, db2g2 |
| | 2nd Taylor 3 points | dt2, dt2g2 |
| | 2nd Savitzky-Golay 9 points | dg2, dg2g2 |
| | 2nd Taylor 3 points, Segment5, Gap5 (Linear Filter - with fixed coefficients) | ds2, ds2g2 |
| | 3rd Taylor 5 points | dt3, dt3g2 |
| Transformation | Absorbance $Log_{10}(1/x)$ | log |
| | Absorbance inverse $1/(10^{-x})$ | ilg |
| | 2nd Derivative/Logarithm | SDL |
| | Kubelka Munk | kmu |
| | Square $x^2$ | sqr |
| | Reciprocal 1/x | 1/s |
| Filter | Linear Filter - with editable coefficients | flt |

* These pretreatments are **wavelength dependent**. The used wavelength is the selected calibration wavelength, or can be edited in the NIR-Explorer under Pretreatments.
*** g2 stands for Gap2-filtering. These pretreatments have been adapted for improved performance for N-500 spectra.

** **MSC** and **Mean Centering** are also depending on the **C-Set spectra selection**. NIRCal handles this dependency automatically. The necessary data is stored in the pretreatments itself so they can also operate automatically in the predictor of the application.

Legend of formula on the following pages:
Capital Letters: Vectors
Small Letters: Scalars
T: Transmittance or Reflectance A
Absorbance
S: Spectrum
h: delta X, distance of base point on the x-axis

## 3.5.3 Normalization

**Normalization**

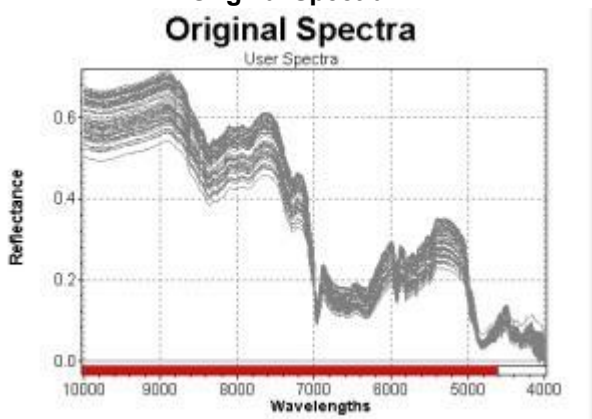
The aim of normalization is to reduce baseline variations.

MSC is used to reduce or to increase baseline effects caused by scattering.
Tip

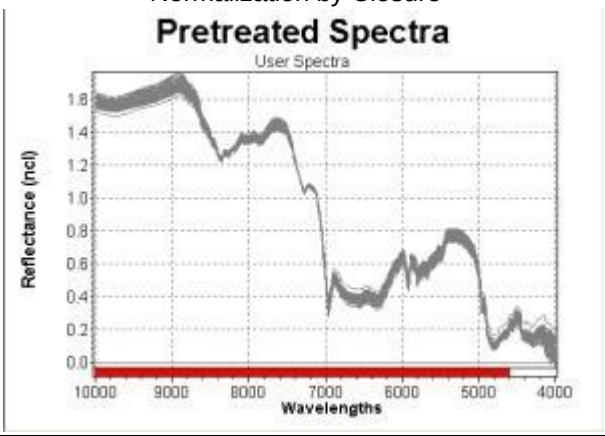
- for solids
- reduce particle size effects
- reduce pressure difference

Overview of all Normalization pretreatments:

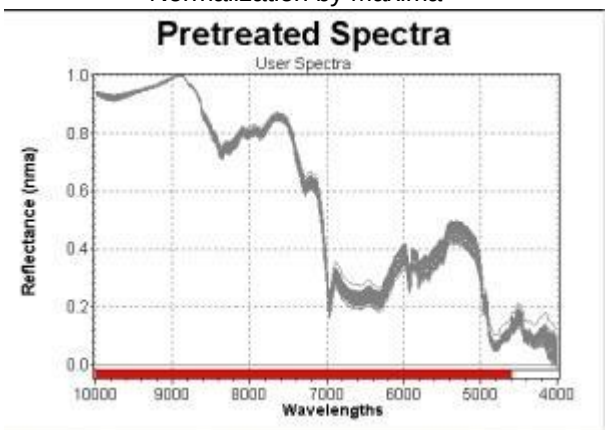
**Original Spectra**



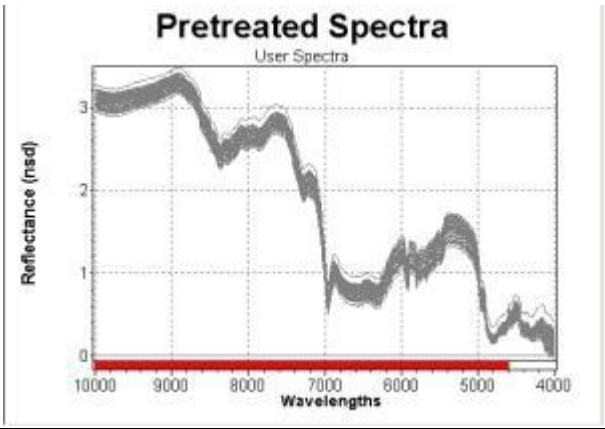

Normalization by Closure*



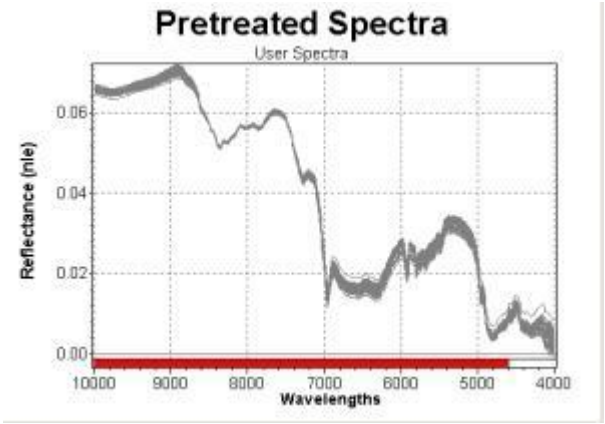

Normalization by Maxima*



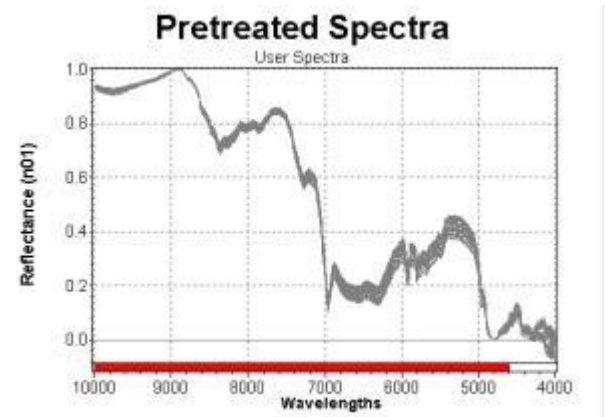
Normalization by Sdev*



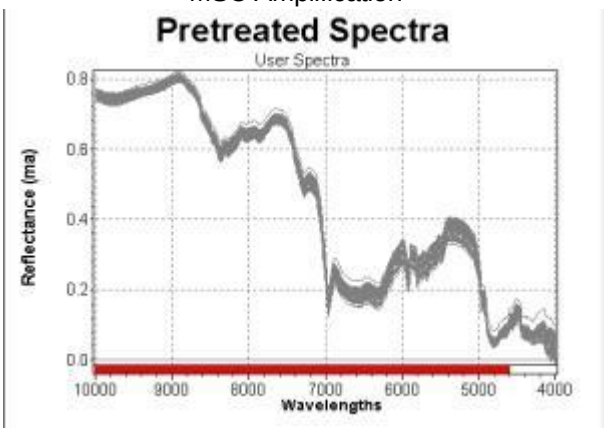

Normalization to Unit Length*



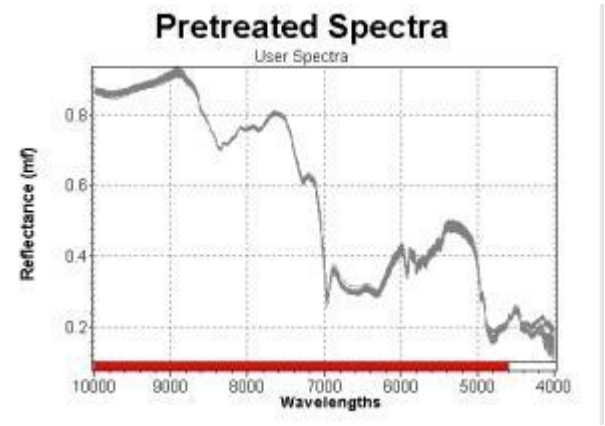

Normalization between 0 to 1*



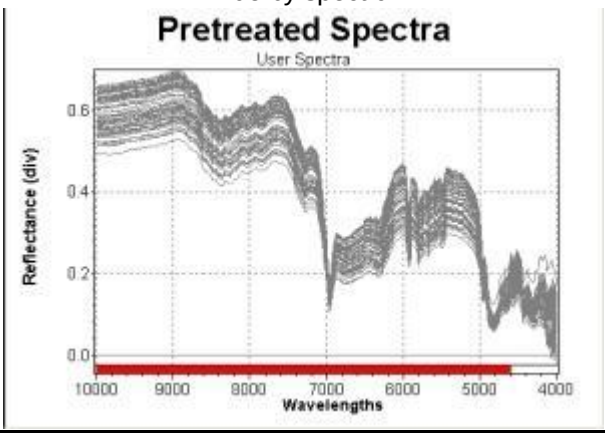

MSC Amplification*



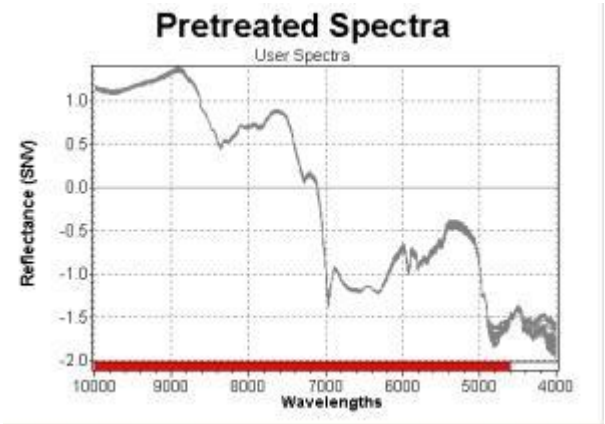

MSC Full*



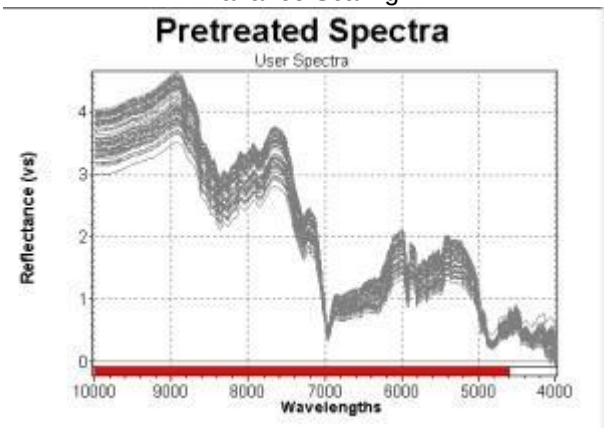

Divide by spectrum




Standard Normal Variate*

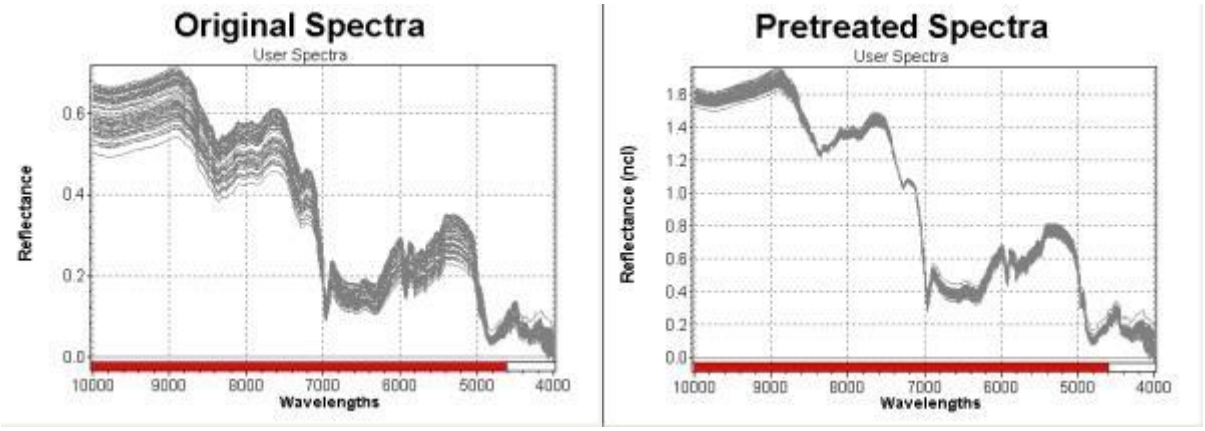



Variance Scaling



* These pretreatments are **wavelength dependent**.

**Normalization by Closure**




Formula:

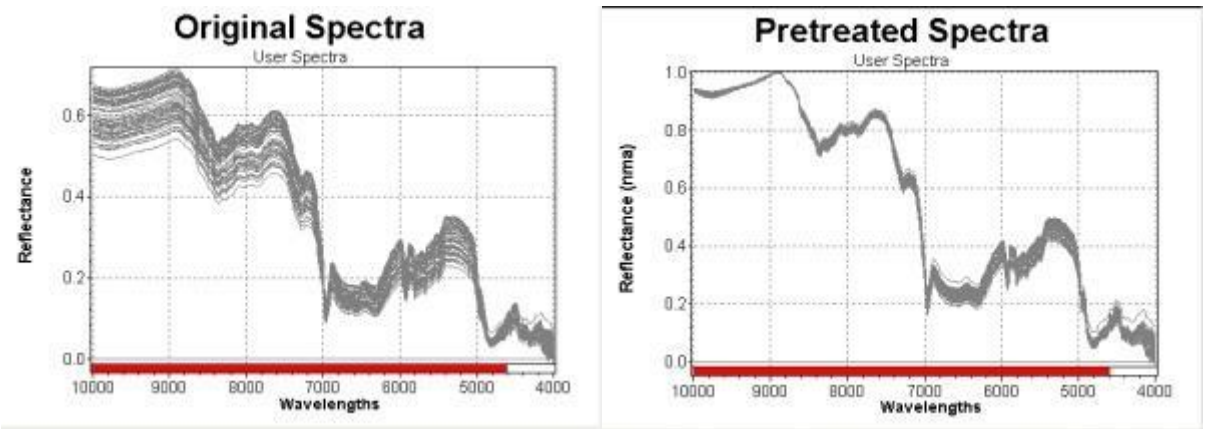
$$NC = \frac{N \cdot T}{\sum_{i=1}^{N} T_i}$$

Martens, Naes 1989, p. 337

Use
Reduction of baseline variations.

Type
The Calibration wavelengths are used when the pretreatment is added; the wavelength range can be changed afterwards in the pretreatment that can be different from the calibration wavelength.

**Normalization by Maxima**



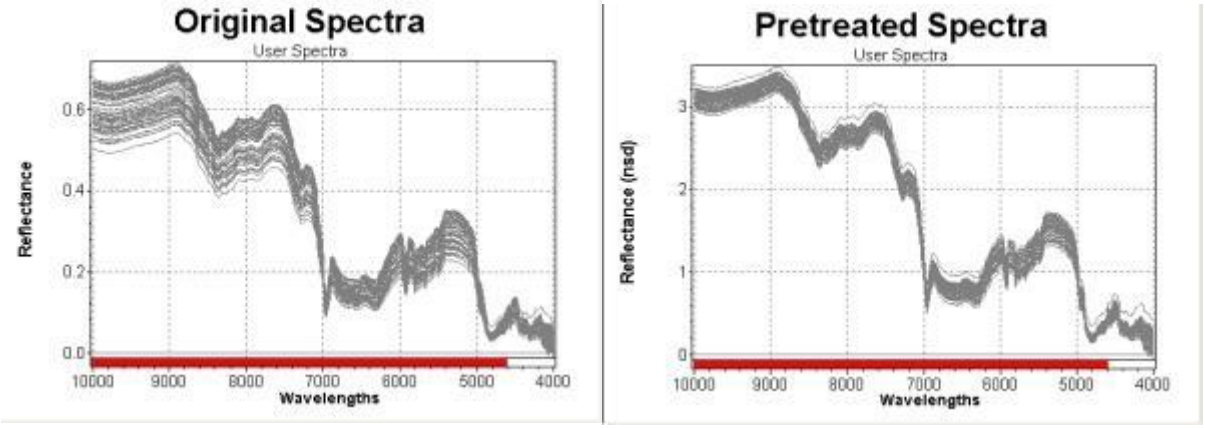

$$N = \frac{T}{\max(T_i)}$$

Use
Reduction of baseline variations.

Type:
The Calibration wavelengths are used when the pretreatment is added; the wavelength range can be changed afterwards in the pretreatment that can be different from the calibration wavelength.

**Normalization by Sdev**

Division of each spectrum through the Standard Deviation of its Intensity Value within the Wavelength selection.



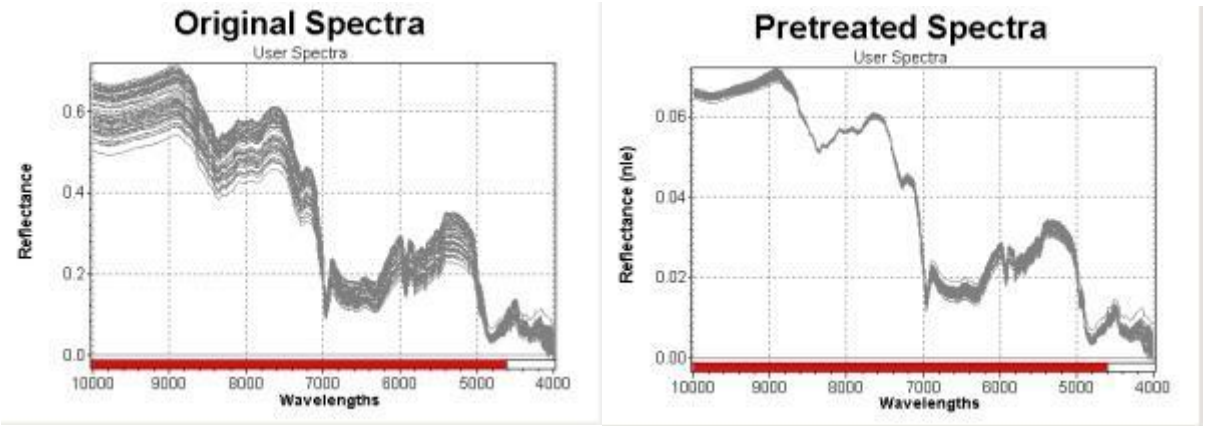

$$X_{nsd} = \frac{x}{Sdev(x)}$$

Use
Reduction of baseline variations.

Type:
The Calibration wavelengths are used when the pretreatment is added; the wavelength range can be changed afterwards in the pretreatment that can be different from the calibration wavelength.

**Normalization to Unit Length**

Vector Normalization to Unit Length.

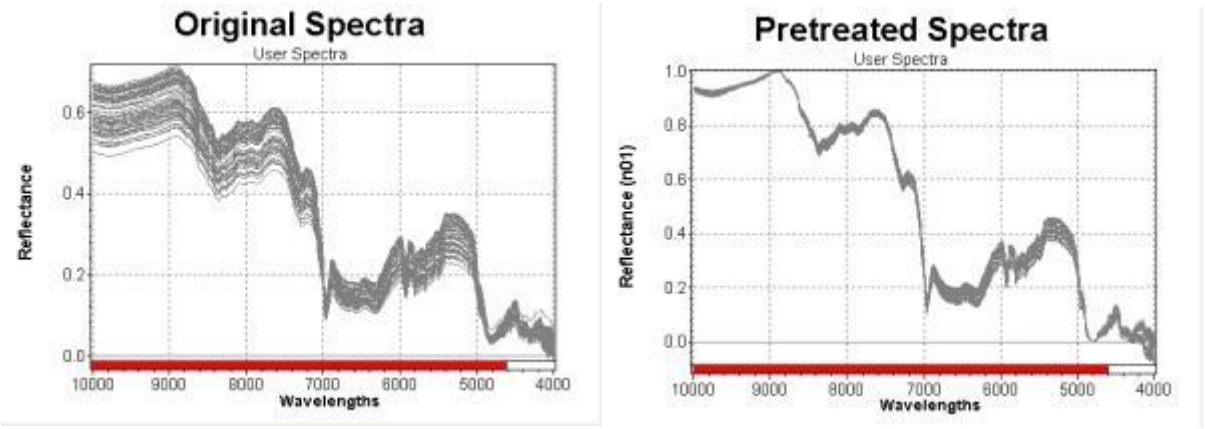



$$VN = \frac{T}{\|T\|} = \frac{T}{\sqrt{T_i^2 + T_{i+1}^2 + \ldots + T_n^2}}$$

Use
Reduction of baseline variations.

Type
The Calibration wavelengths are used when the pretreatment is added; the wavelength range can be changed afterwards in the pretreatment that can be different from the calibration wavelength.

**Normalization between 0 to 1**

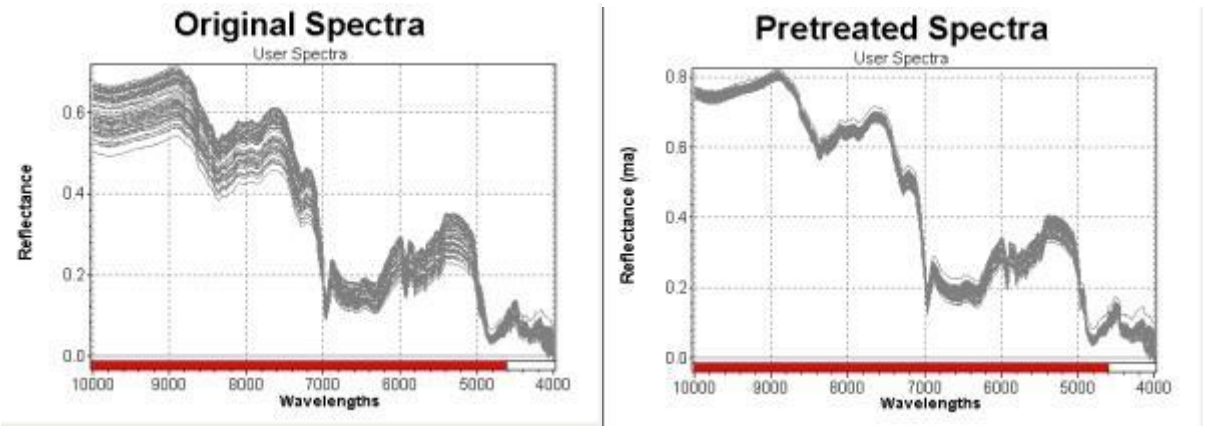



$$S = \frac{S - \min(S)\vec{1}}{\max\left|S - \min(S)\vec{1}\right|}$$

Use
Reduction of baseline variations.

Type
The Calibration wavelengths are used when the pretreatment is added; the wavelength range can be changed afterwards in the pretreatment that can be different from the calibration wavelength.

**MSC Amplification**



$$T = a\vec{1} + b\overline{T} + \vec{E}$$

$$MSC_{full} = \frac{T - a\vec{1}}{b}$$

$$MSC_{offset} = T - a\vec{1}$$

$$MSC_{amplification} = \frac{T}{b}$$

$$\overline{T} = Mittelwert$$

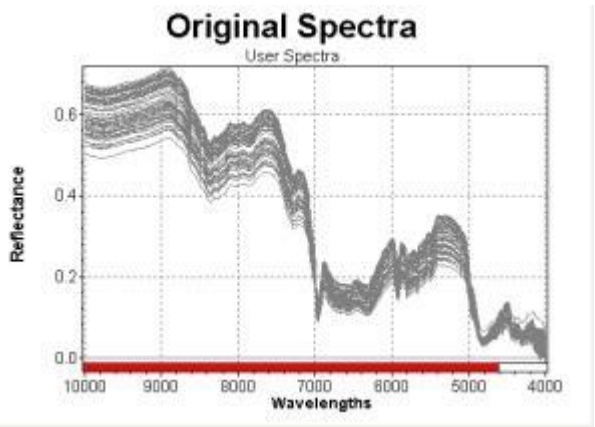
$$\vec{E} = Error$$

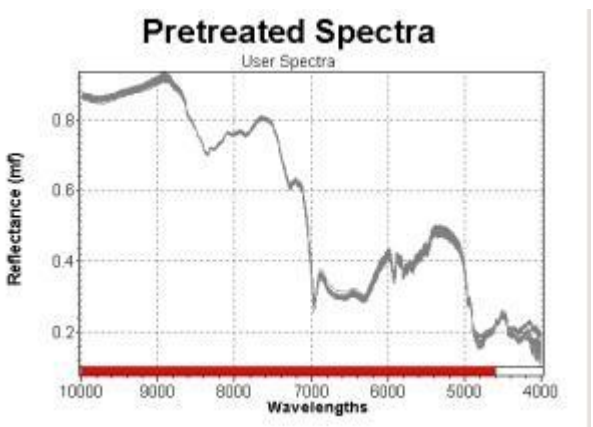
Use
Can increase baseline effect, can be good for particle size separation.

Type
This pretreatment is not depending on the calibration wavelengths.

**MSC Full**

MSC Multiplicative Scatter Correction (full)




$$T = a\vec{1} + b\overline{T} + \vec{E}$$

$$MSC_{full} = \frac{T - a\vec{1}}{b}$$

$$MSC_{offset} = T - a\vec{1}$$

$$MSC_{amplification} = \frac{T}{b}$$

$$\overline{T} = Mittelwert$$

$$\vec{E} = Error$$

Use
Eliminates scattering effects. Reduction of baseline variations.

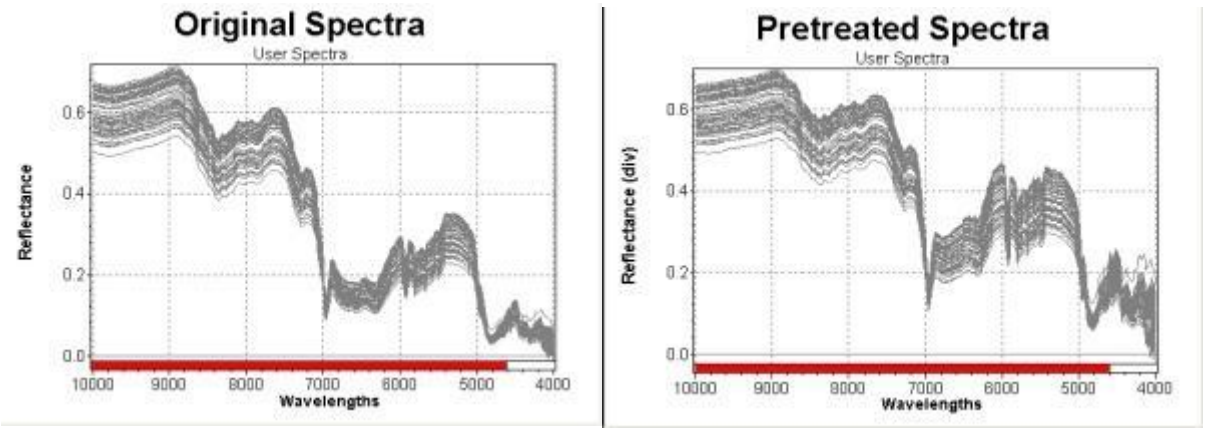
Type
The Calibration wavelengths are used when the pretreatment is added; the wavelength range can be changed afterwards in the pretreatment that can be different from the calibration wavelength.

**Divide by Spectrum**

When selecting this pretreatment the number of the spectrum, which should be used for the division, has to be entered.

In this way it is possible to enhance the differences in the data set.



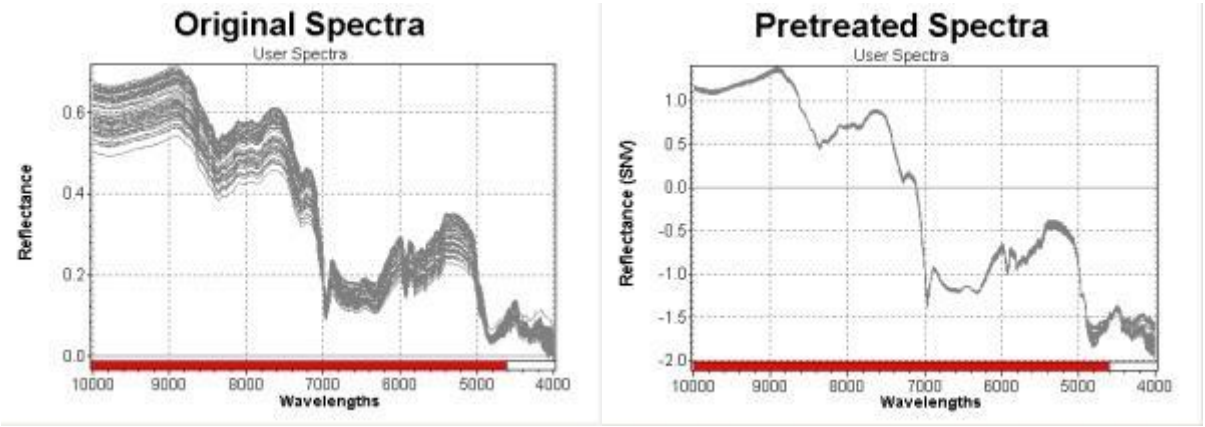



***NOTE***
Having applied Divide by Spectrum the selected spectrum (spectrum No. 10 in this example) must not be put into the C-Set; its ordinate values will contain 1.0 only. The selected spectrum is copied into the pretreatment once. The selected spectra contains all previous pretreatments.

Type
This pretreatment is not depending on the calibration wavelengths.

**Standard Normal Variate**

The SNV transformation centers each spectrum and then scales it by its own standard deviation (mean zero and variance equal to one).
It corrects shifts on the ordinate.




Y_SNV = (y - mean(y)) / Sdev (Y)

Use
Reduction of baseline variations.

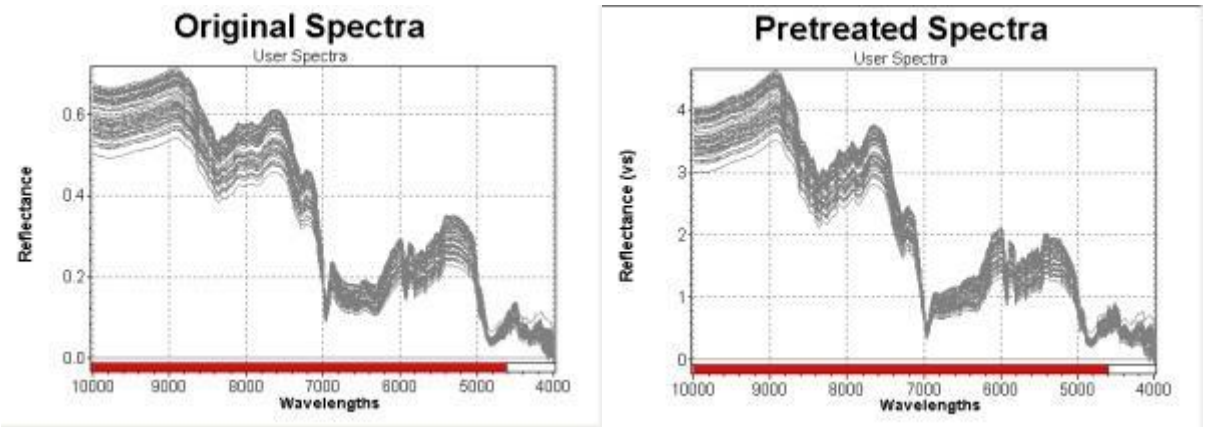
Type
The Calibration wavelengths are used when the pretreatment is added; the wavelength range can be changed afterwards in the pretreatment that can be different from the calibration wavelength.

**Variance Scaling**

The spectra are divided by the standard deviation vector of the C-Set spectra. It is dependent on the C-Set selection.

***NOTE***

NIRCal handles the changing of the C-Set spectra automatically (by a refresh F5 or recalculation) and changes the the standard deviation and the result of the pretreatment. The standard deviation vector is stored in the pretreatment, so it can be used in the predictor and application.




$$VS = \frac{T}{s}$$

Use
Can increase baseline variations, can be good for particle size separation.

Type
This pretreatment is not depending on the calibration wavelengths.

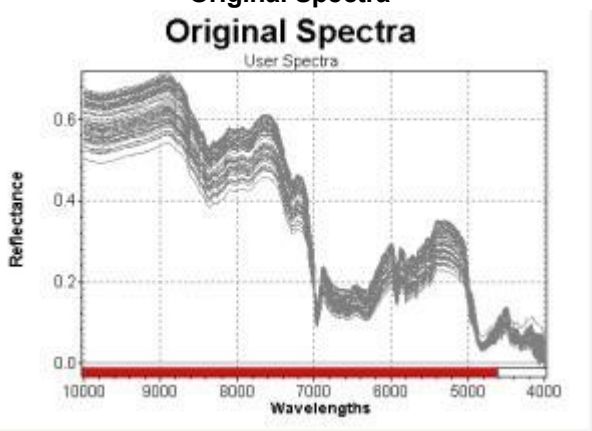
### 3.5.4 Offset

**Offset**

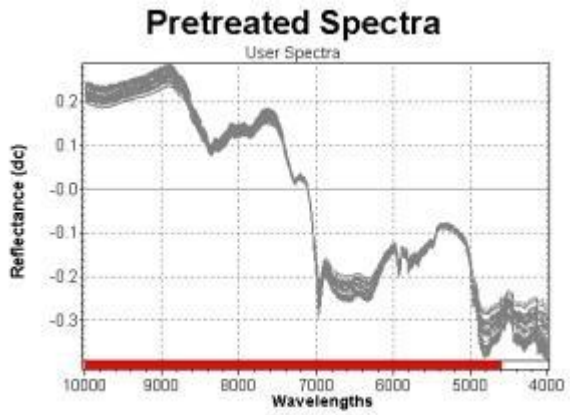
The aim of Offset is to make baseline correction which caused by scattering.

Overview of all Offset pretreatments:

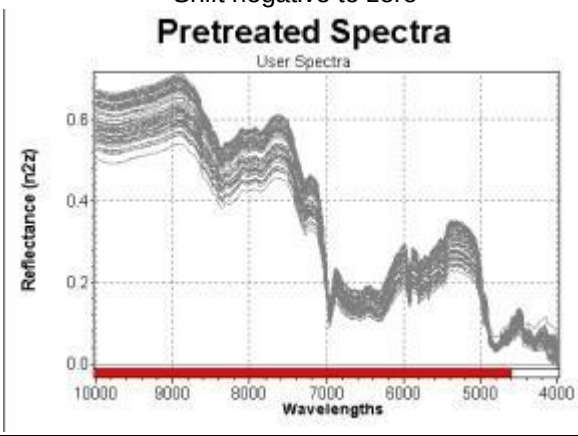
**Original Spectra**



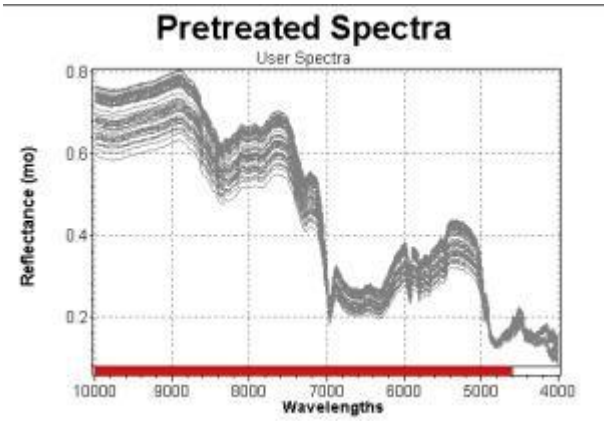

Subtract DC*



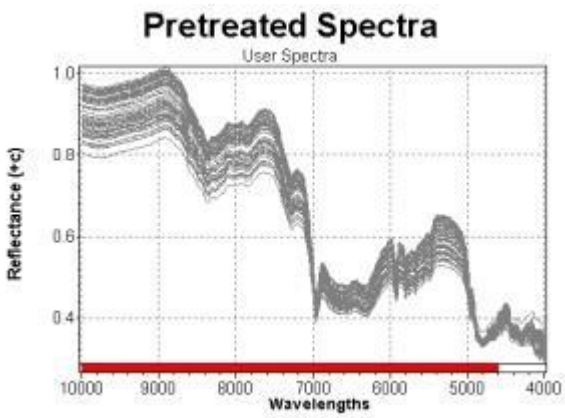

Shift negative to zero*



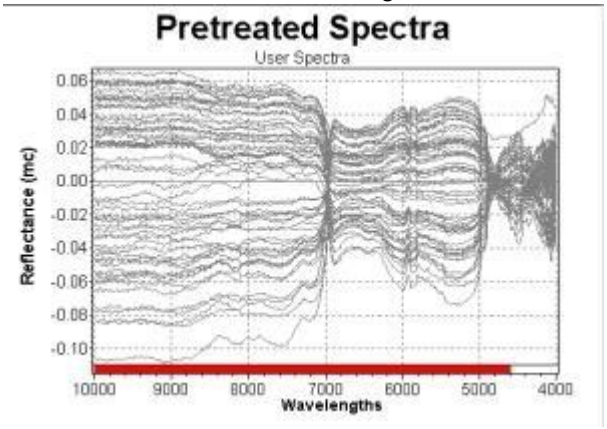

MSC Offset*



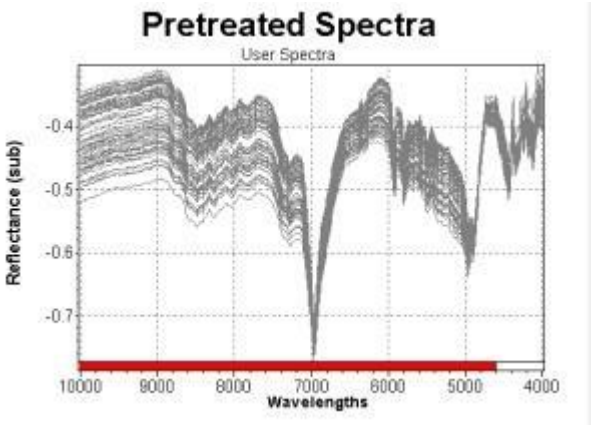

Add Constant



Mean Centering



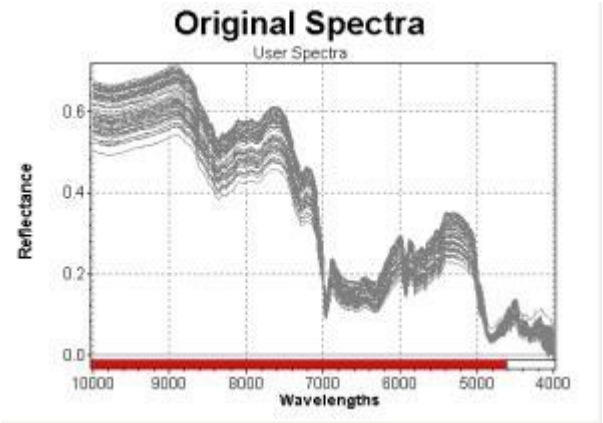

Subtract spectrum

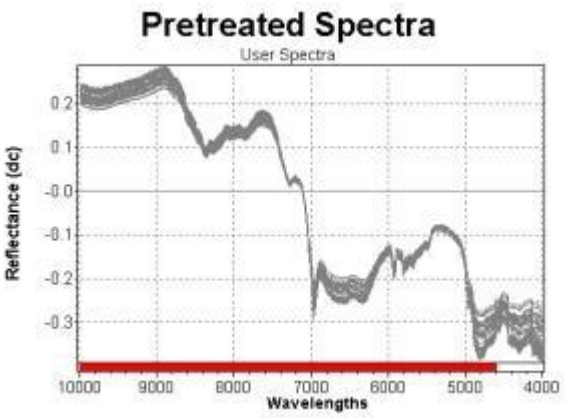



* These pretreatments are wavelength dependent.

**Subtract DC**

The integral of the spectrum is subtracted as a scalar from the spectrum.




Formula:

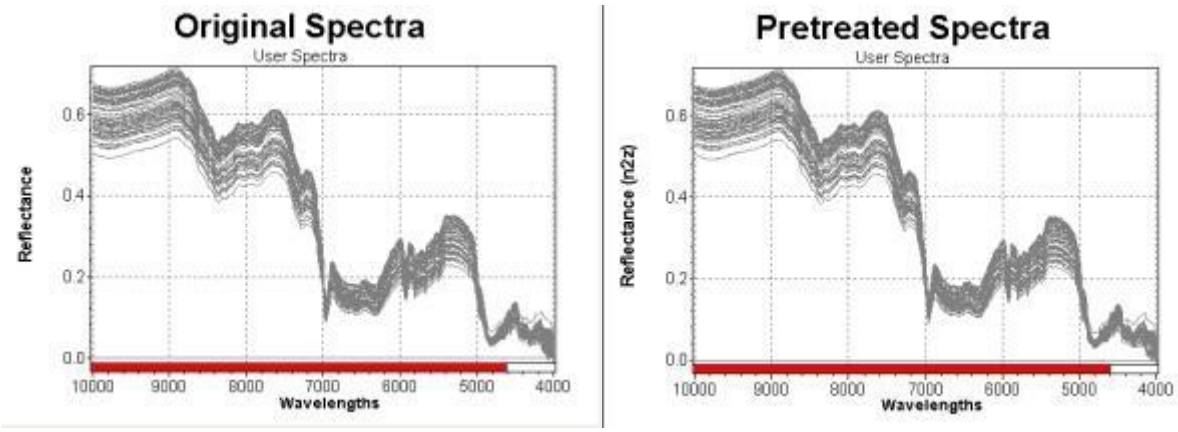
$$BC = T - 1\frac{\sum_{i=1}^{n} T_i}{N}$$

Use
Baseline correction of spectra.

Type:
The Calibration wavelengths are used when the pretreatment is added; the wavelength range can be changed afterwards in the pretreatment that can be different from the calibration wavelength.

**Shift Negative to 0**

In this example there is no effect because the spectra do not have any negative values. If there are negative values in a spectrum, a constant of the same absolute value as the largest negative value will be subtracted from the spectra. There will be no more negative values in the spectrum after the pretreatment has been performed; the minimum ordinate value will be zero.



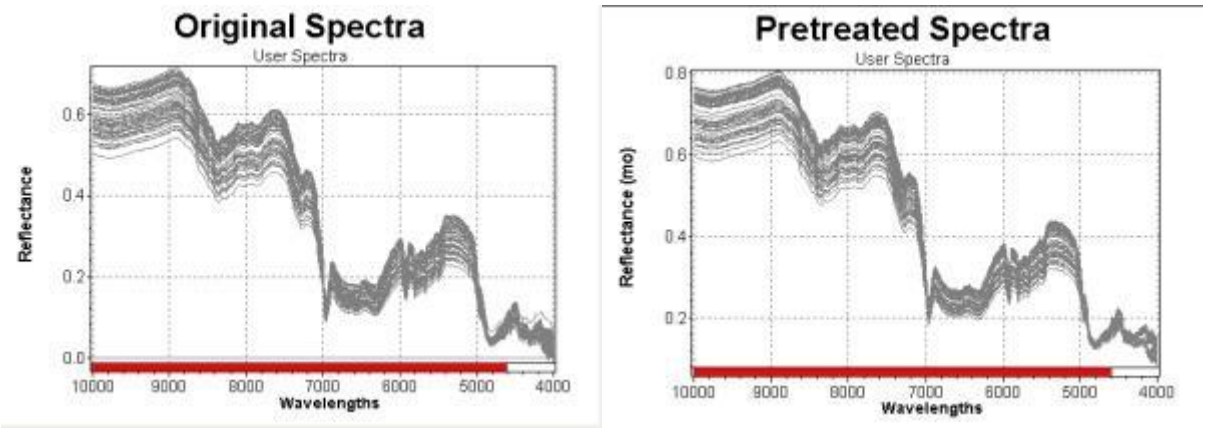

Formula:

$$S = S - \min(S)\vec{1} \quad \Lambda \quad \min(S) < 0$$

Use
For each spectrum the minimum -if it is smaller than zero- is substracted from the spectrum.

Type:
The Calibration wavelengths are used when the pretreatment is added; the wavelength range can be changed afterwards in the pretreatment that can be different from the calibration wavelength.

**MSC Offset**



Formula:

$$T = a\vec{1} + b\overline{T} + \vec{E}$$

$$MSC_{full} = \frac{T - a\vec{1}}{b}$$

$$MSC_{offset} = T - a\vec{1}$$

$$MSC_{amplification} = \frac{T}{b}$$

$$\overline{T} = Mittelwert$$

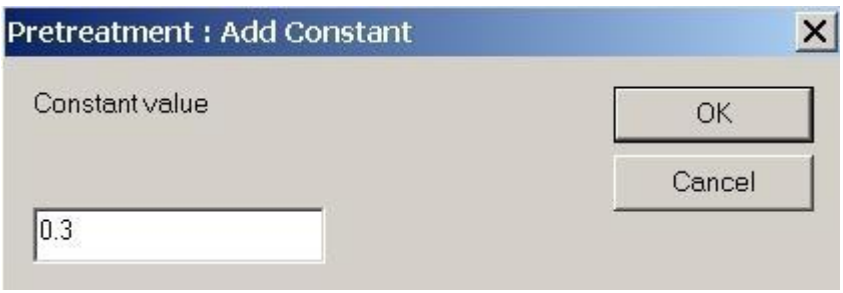
$$\vec{E} = Error$$

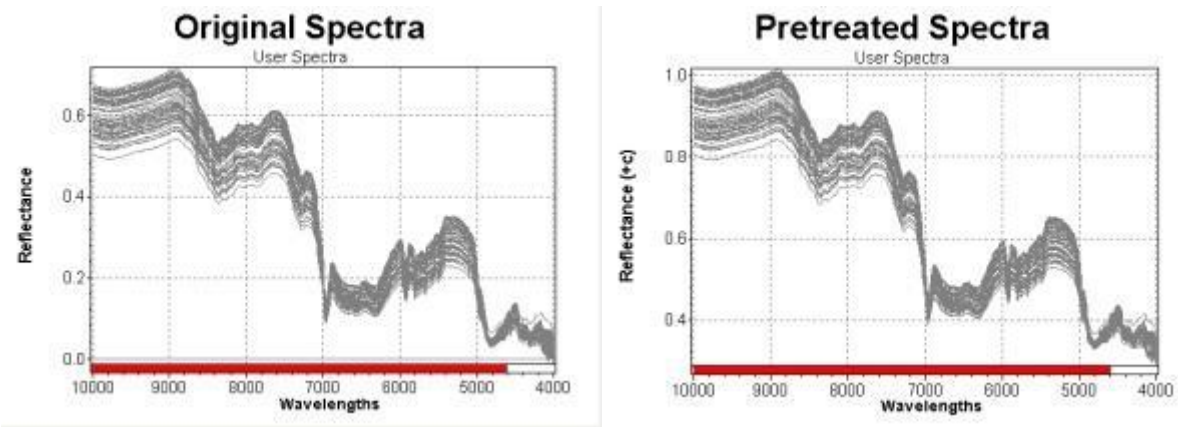
Use
Can increase the baseline shift, can be used for particle size separation.

Type:
The Calibration wavelengths are used when the pretreatment is added; the wavelength range can be changed afterwards in the pretreatment that can be different from the calibration wavelength.

**Add constant**

When selecting this pretreatment the constant to be added has to be entered first.




e.g. Constant = 0.3

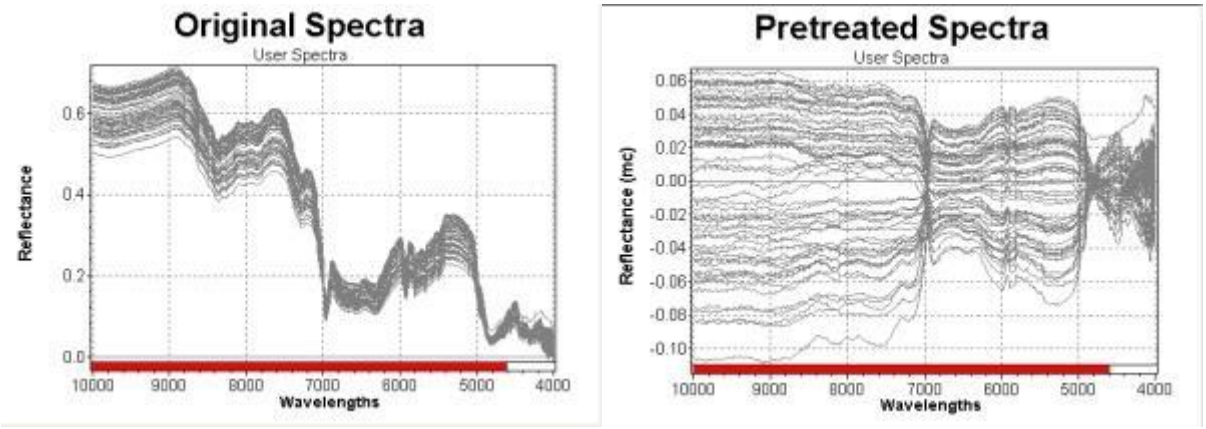
This pretreatment can be used for the correction of systematic errors.

***NOTE***
To subtract a constant simply enter a negative constant.

$$T = T + a\vec{1}$$

**Mean Centering**

The mean spectrum of the C-Set spectra will be subtracted from each spectra, only the deviation is remaining.




***NOTE***

Do not use Mean Centering as the last step in the pretreatment sequence when the Mean Centering has already been activated in the Calculation Parameters dialog box.

The Mean Centering defined in the Calculation Parameters box is calculated directly after the pretreatment sequence.

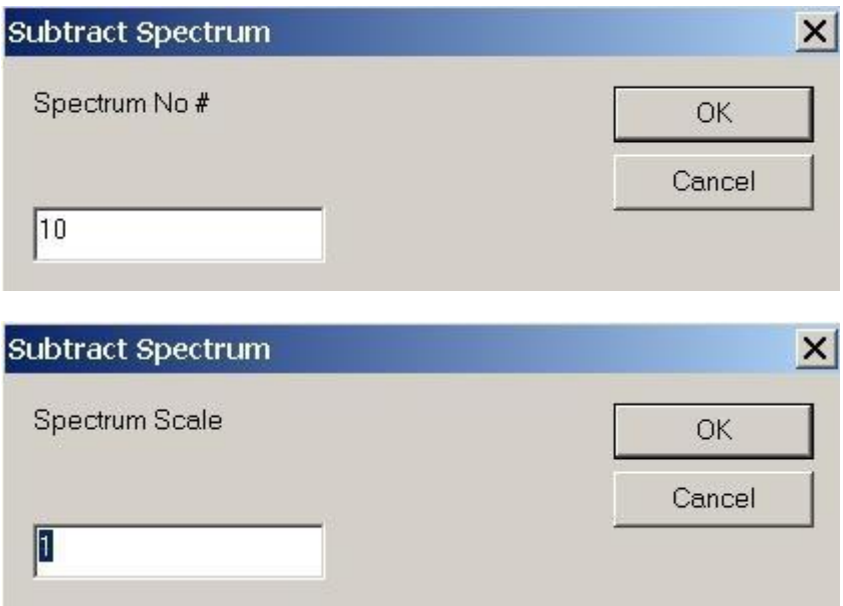
Type

This pretreatment is not depending on the calibration wavelengths.

**Subtract spectrum**

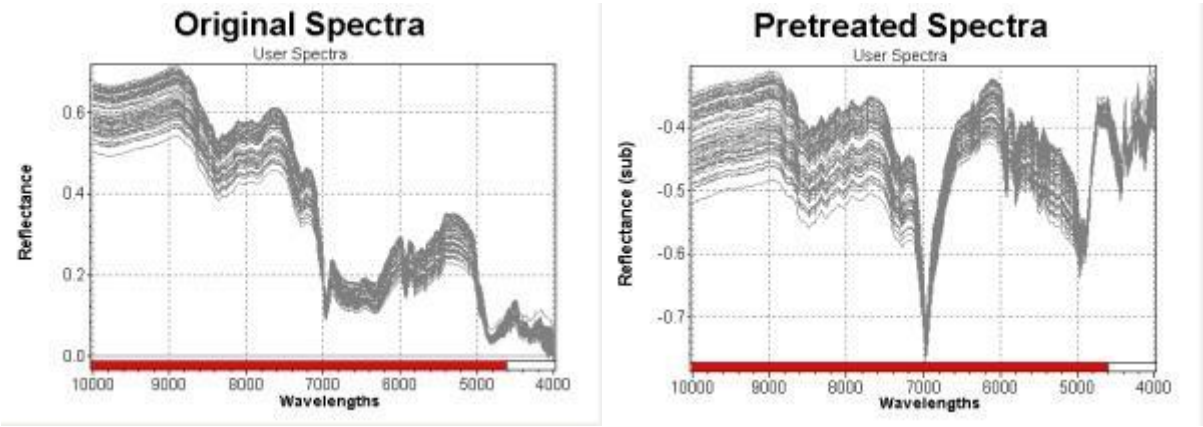
When selecting this pretreatment the number of the spectrum, which should be subtracted from the other spectra, has to be entered.

In this way it is possible to enhance the differences in the data set.




***NOTE***

The scale factor can also be negative to add a spectrum.




Type
This pretreatment is not depending on the calibration wavelengths.

## 3.5.5 Smoothing

**Smoothing**

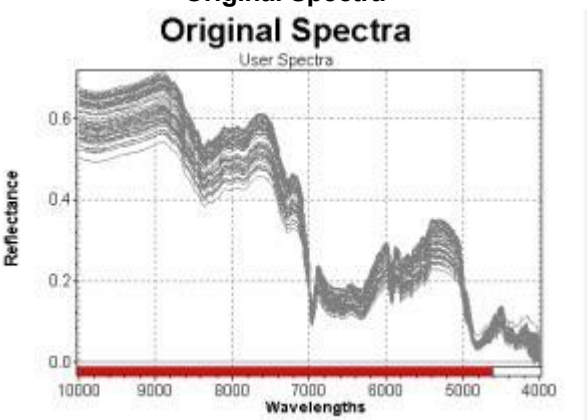
Smoothing is used to reduce the noise level in spectra. However when using smoothing the spectral resolution will be affected. The smoothing function should be chosen according to the original data.

***NOTE***

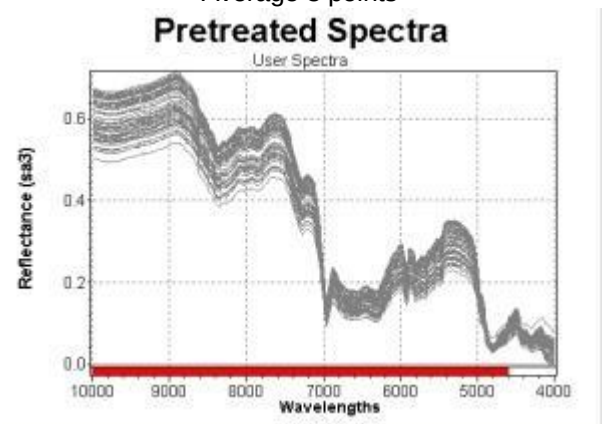
Please choose the smoothing function carefully. No loss of spectral information should appear.

Overview of all Smoothing pretreatments:

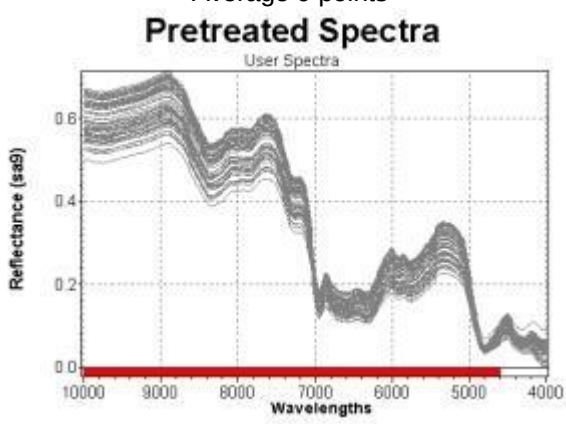
**Original Spectra**



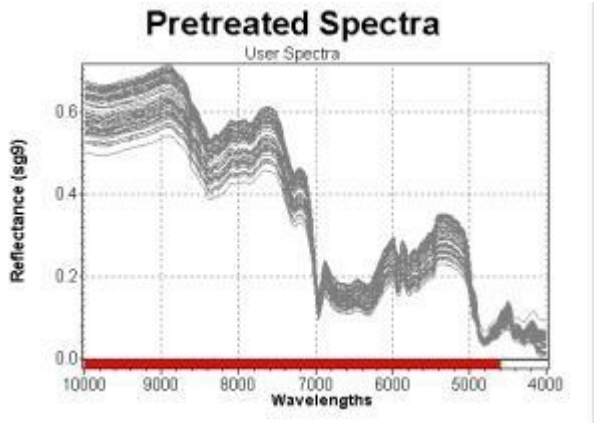

Average 3 points



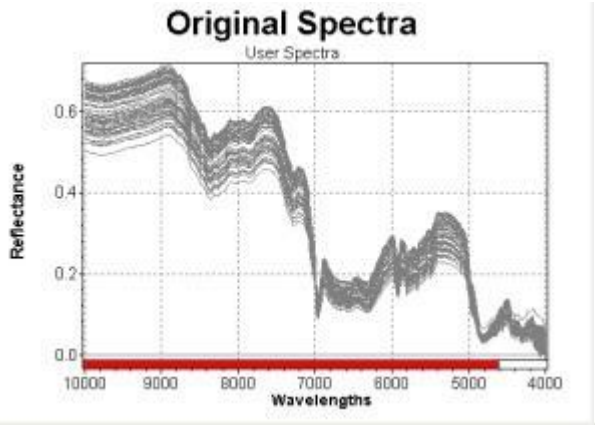

Average 9 points

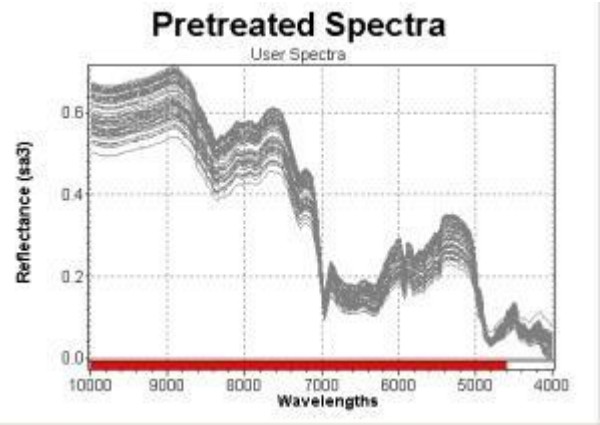


Average Savitzky-Golay 9 points




**Average 3 points**



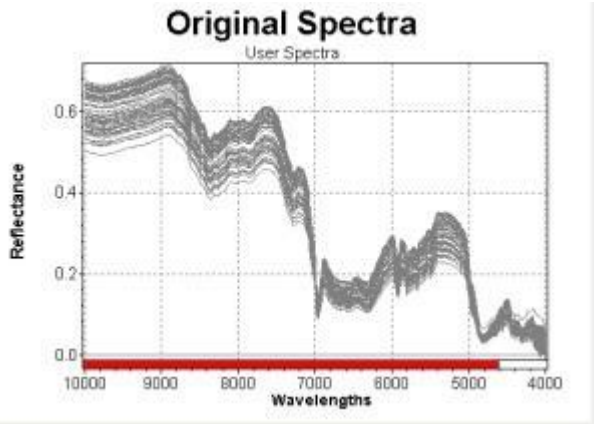



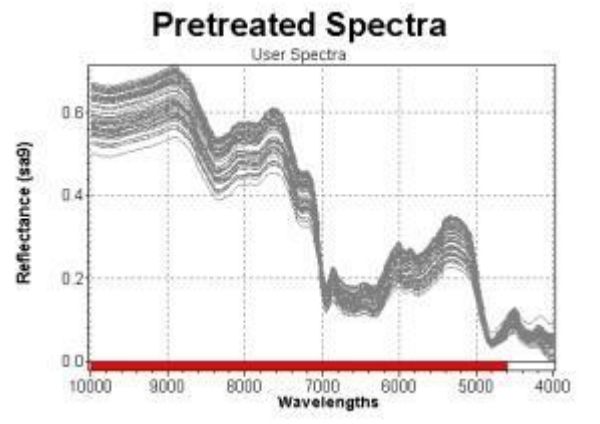

Formula:

$$gm_i = \frac{T_{i-1} + T_i + T_{i+1}}{3}$$

Type
This pretreatment is not depending on the calibration wavelengths.

**Average 9 points**



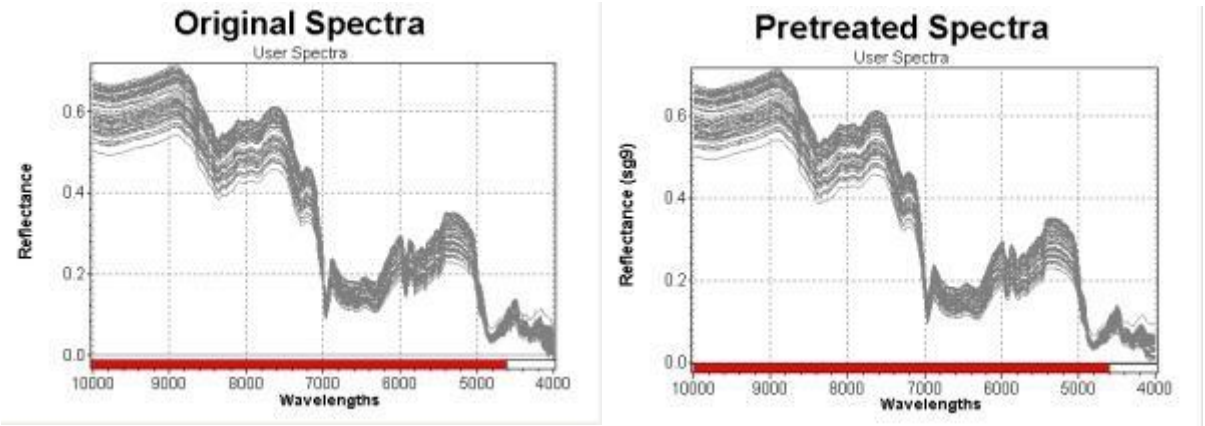
Formula:

$$gm_i = \frac{T_{i-4} + T_{i-3} + T_{i-2} + T_{i-1} + T_i + T_{i+1} + T_{i+2} + T_{i+3} + T_{i+4}}{9}$$

Type
This pretreatment is not depending on the calibration wavelengths.

***NOTE***
The information loss can be very high.

**Average Savitzky-Golay 9 points**




Savitzky-Golay Smoothing (=zero-order derivative), 9 points, cubic

$$f_{smooth}(x_i) = \frac{-21f(x_{i+4}) + 14f(x_{i+3}) + 39f(x_{i+2}) + 54f(x_{i+1}) + 59f(x_i) + 54f(x_{i-1}) + 39f(x_{i-2}) + 14f(x_{i-3}) - 21f(x_{i-4})}{231}$$

Type
This pretreatment is not depending on the calibration wavelengths.

## 3.5.6 Derivatives

**Derivatives**

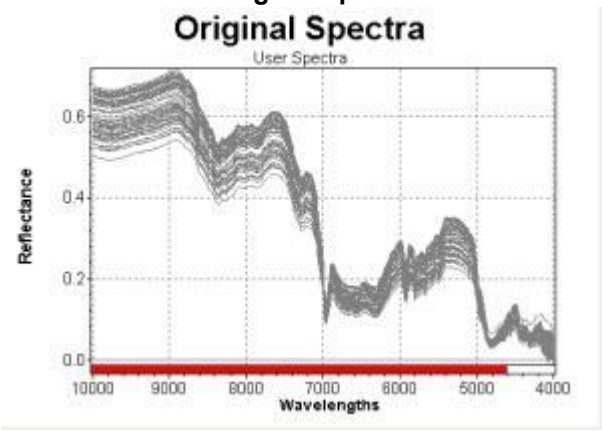
A derivative is used to reduce baseline effects and to increase smaller absorption peaks (shoulder effect). The first derivative will eliminate a linear ordinate offset. The second derivative will eliminate a sloping baseline.

***NOTE***
When using a derivative the signal-to-noise ratio will decrease. Therefore often it will be necessary to combine it with smoothing.

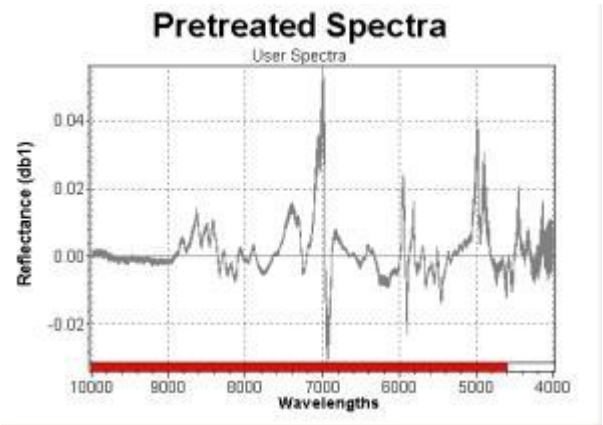
Many derivative calculations are based on polynom fitting, therefore because of the mathematics involved, some smoothing will be applied anyway.

Overview of all Derivative pretreatments:

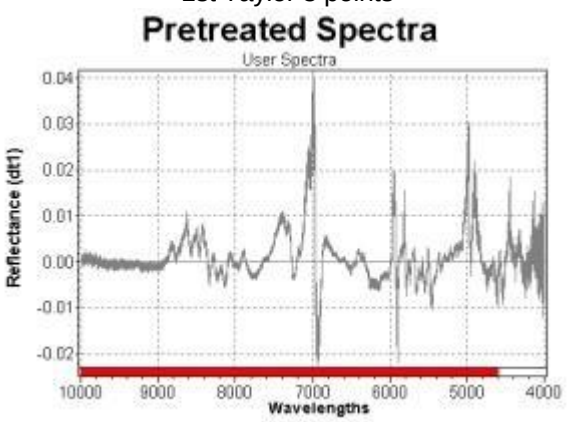
**Original Spectra**



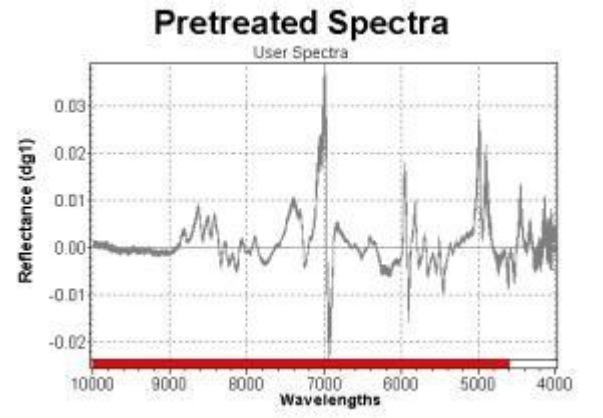

1st BCAP



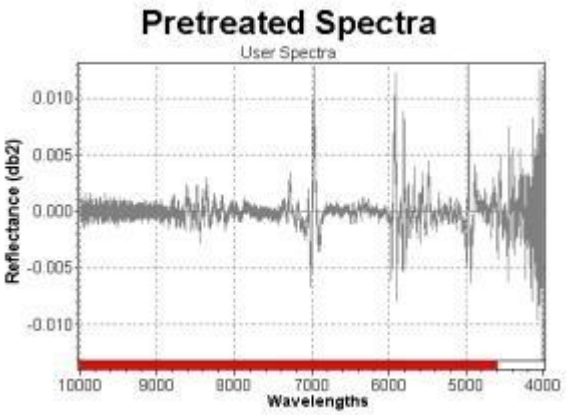

1st Taylor 3 points



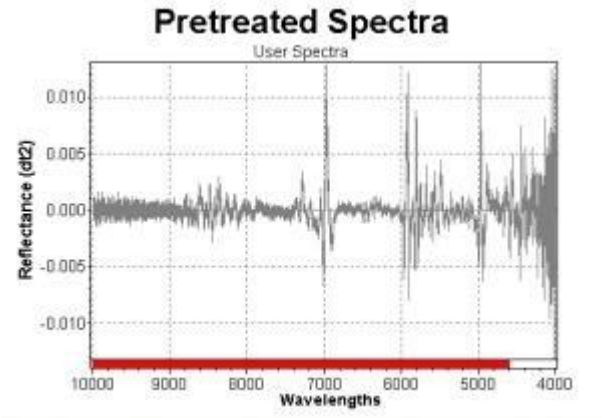

1st Savitzky-Golay 9 points



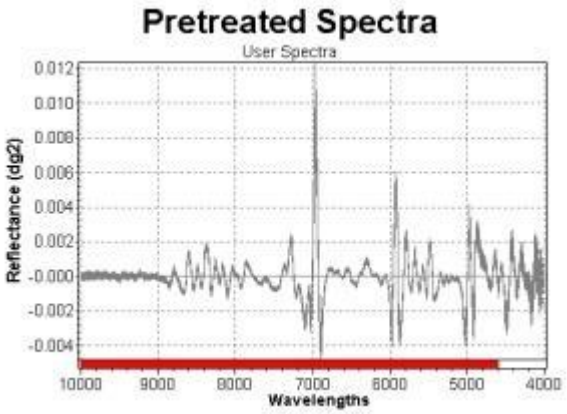

2nd BCAP



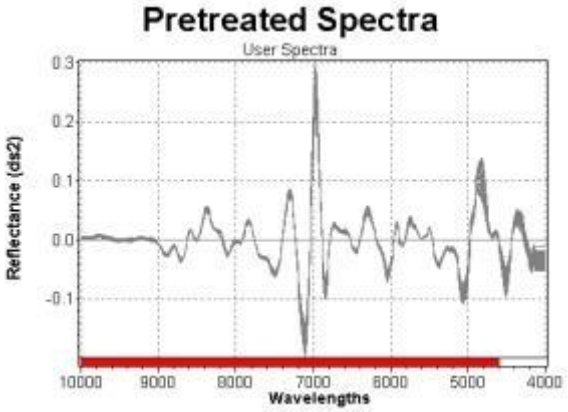

2nd Taylor 3 points



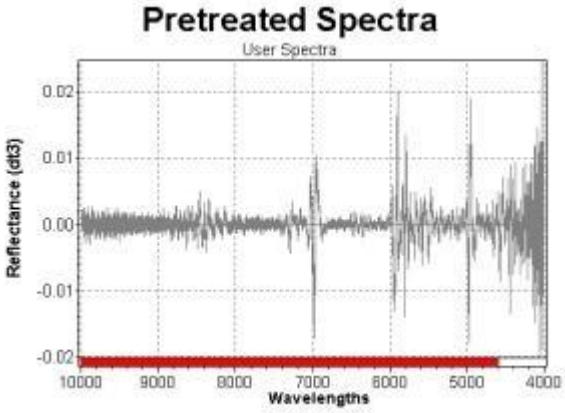

2nd Savitzky-Golay 9 points



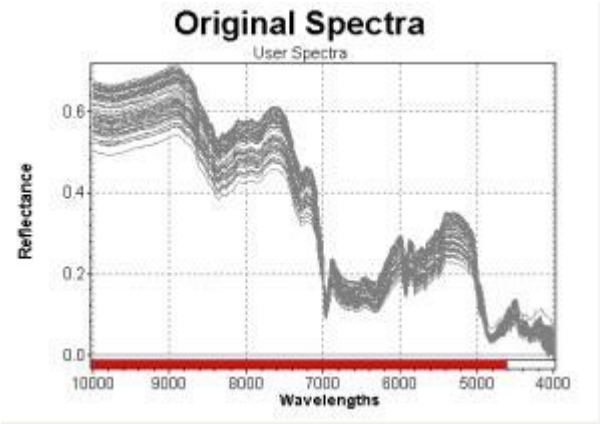
2nd Taylor 3 Points Segment5 Gap5

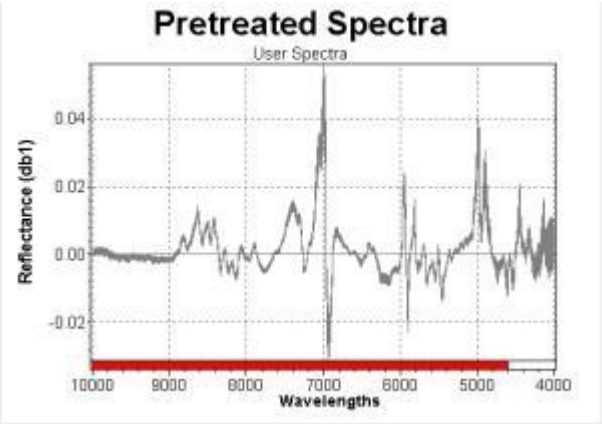



3rd Taylor 5 points




**1st BCAP**



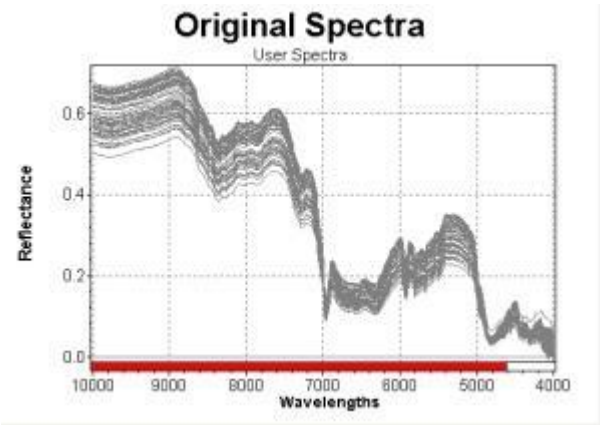



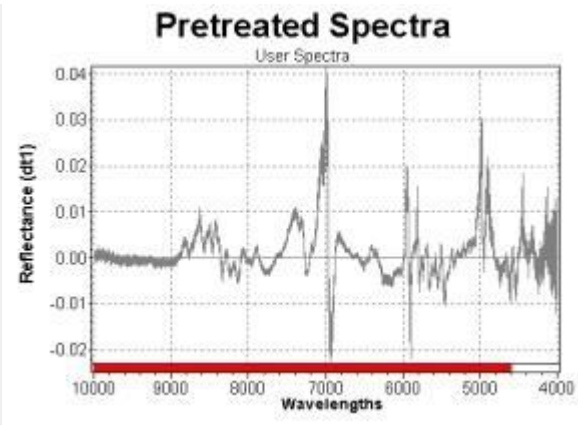

Formula:

$$f'(x_i) = \frac{f(x_{i+2}) + f(x_{i+1}) - f(x_{i-1}) - f(x_{i-2})}{4}$$

Type
This pretreatment is not depending on the calibration wavelengths.

**1st Taylor 3 points**



Formula:

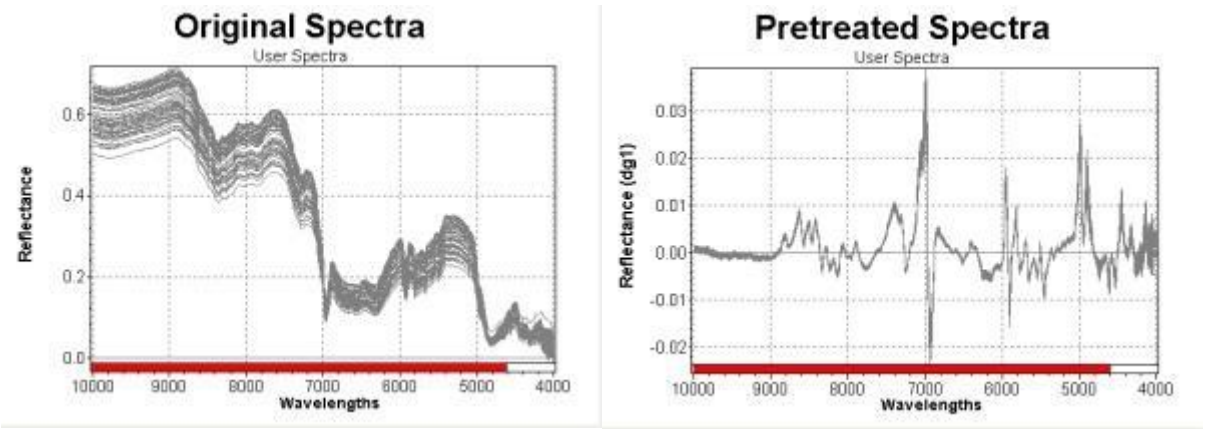
$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{h^2}; \text{centered}$$

Type
This pretreatment is not depending on the calibration wavelengths.

***NOTE***
More sensitive to noise than Savitzky-Golay 1st derivative, can be combined with smoothing (before or after).

**1st Savitzky-Golay 9 points**

More useful for spectra with very sharp absorption bands with high noise spectra, for detecting very small wavelength shifts.




Formula:
Savitzky-Golay derivative works with orthogonal polynoms (here cubic)

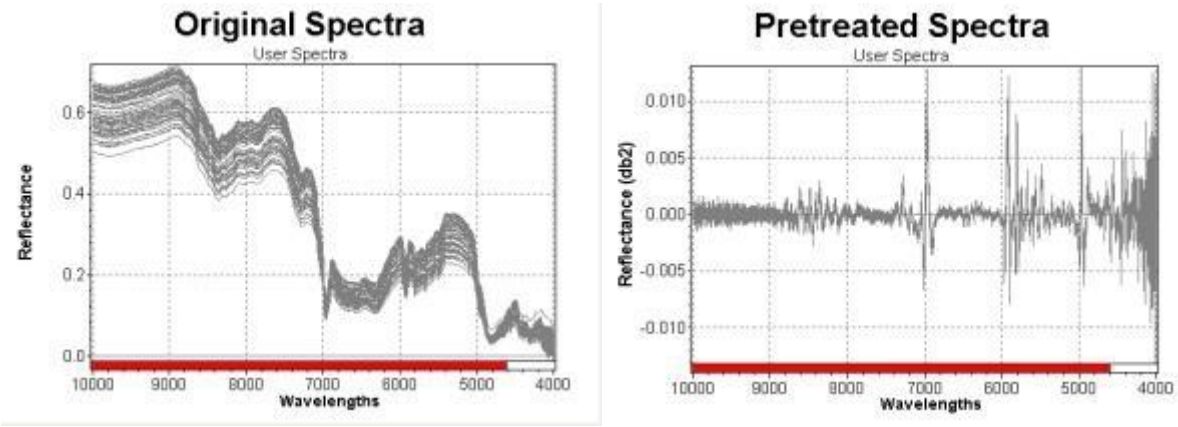
$$f'(x_i) = \frac{-86f(x_{i+4}) + 142f(x_{i+3}) + 193f(x_{i+2}) + 126f(x_{i+1}) - 126f(x_{i-1}) - 193f(x_{i-2}) - 142f(x_{i-3}) + 86f(x_{i-4})}{1188}$$

Type
This pretreatment is not depending on the calibration wavelengths.

Literature:
Savitzky, Golay Analytical Chemistry Vol. 36, No. 8, July 1964, p.1627-1639, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures"

***NOTE***
This cited paper contains some errors in some coefficients.

**2nd BCAP**



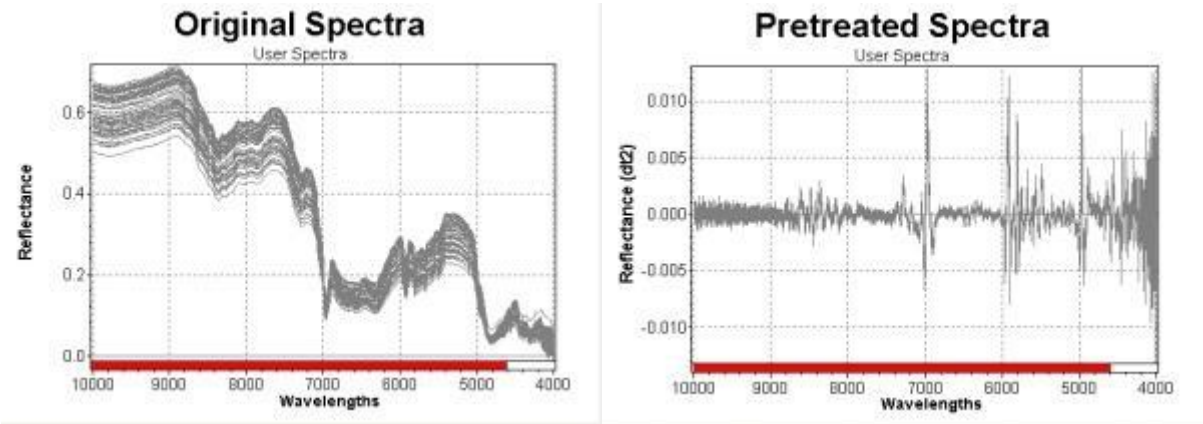

Formula:

$$f''(x_i) = \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1})}{2}$$

Type
This pretreatment is not depending on the calibration wavelengths.

***NOTE***
The result is mainly noisy spectra.

**2nd Taylor 3 points**



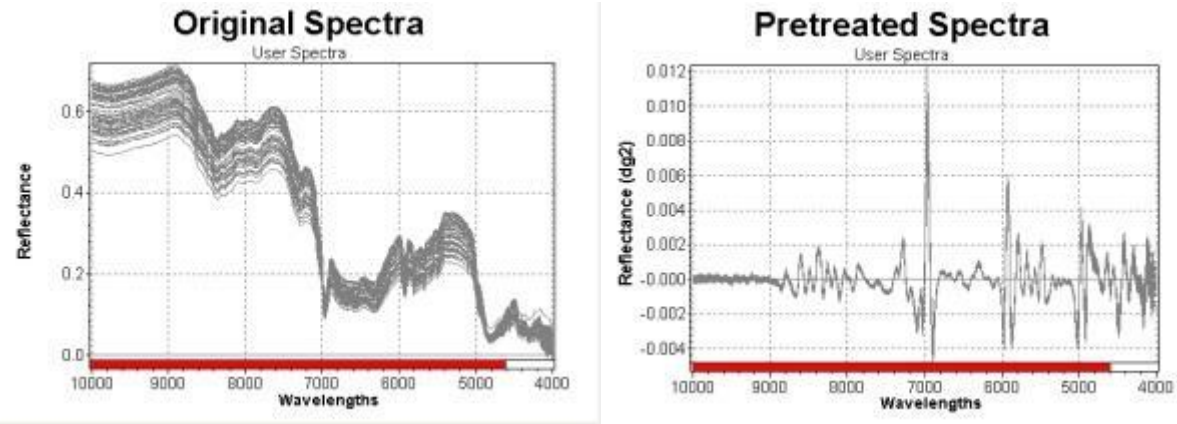

Formula:

$$SD = T''$$

$$f''(x_i) = \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1})}{h^2}; \text{centered, "boxcar 2nd derivative"}$$

Type
This pretreatment is not depending on the calibration wavelengths.
***NOTE***
The result is mainly noisy spectra.

**2nd Savitzky-Golay 9 points**




Formula:

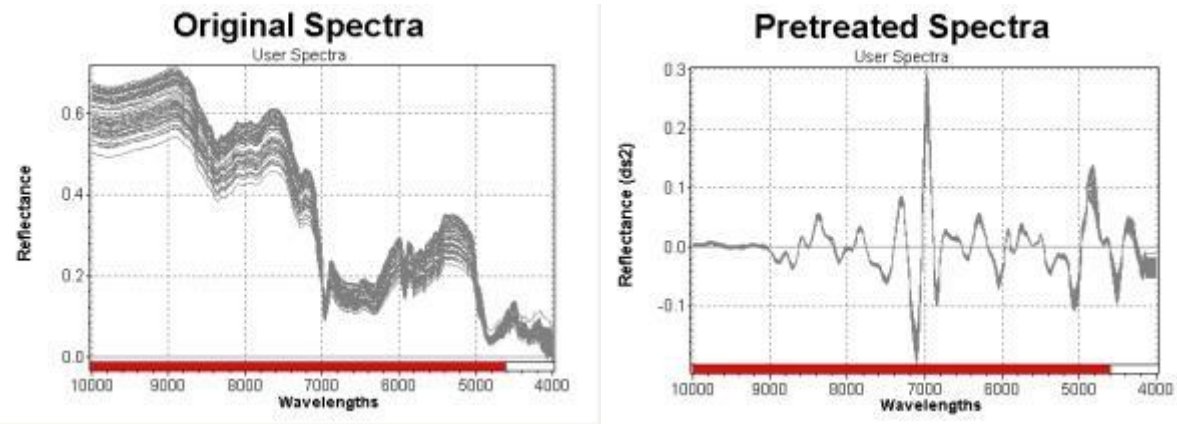
$$f''(x_i) = \frac{28f(x_{i+4}) + 7f(x_{i+3}) - 8f(x_{i+2}) - 17f(x_{i+1}) - 20f(x_i) - 17f(x_{i-1}) - 8f(x_{i-2}) + 7f(x_{i-3}) + 28f(x_{i-4})}{462}$$

Type
This pretreatment is not depending on the calibration wavelengths.

Literature:
Savitzky, Golay Analytical Chemistry Vol. 36, No. 8, July 1964, p.1627-1639, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures"

***NOTE***
This cited paper contains some errors in some coefficients.

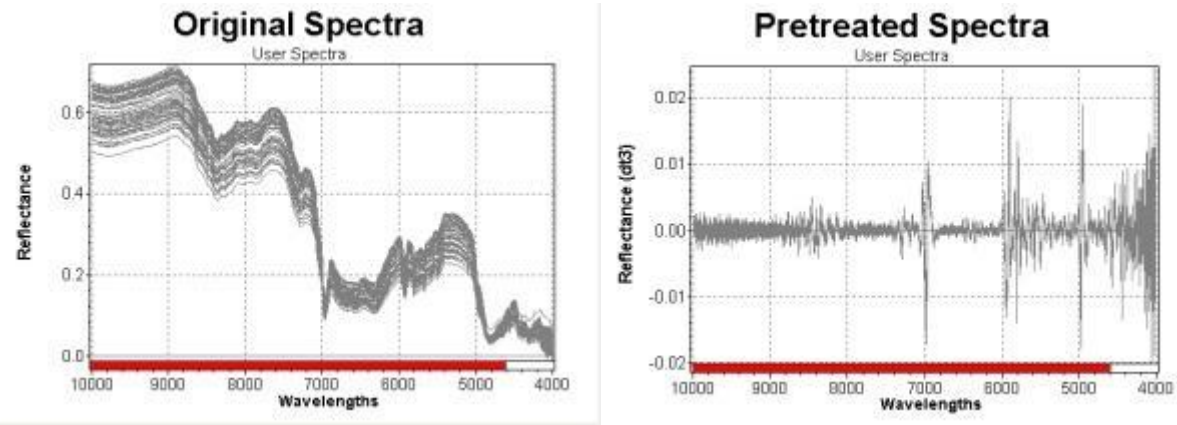
**2nd Taylor 3 Points Segment5 Gap5**




Formula:
See Linear Filter

Type
This pretreatment is not depending on the calibration wavelengths.

**3rd Taylor 5 points**




Formula:

$$f'''(x_i) = \frac{f(x_{i+2}) - 2f(x_{i+1}) + 2f(x_{i-1}) - f(x_{i-2})}{2h^3}; \text{centered}$$

Type
This pretreatment is not depending on the calibration wavelengths.

***NOTE***
The result is mainly very noisy spectra.

## 3.5.7 Transformation

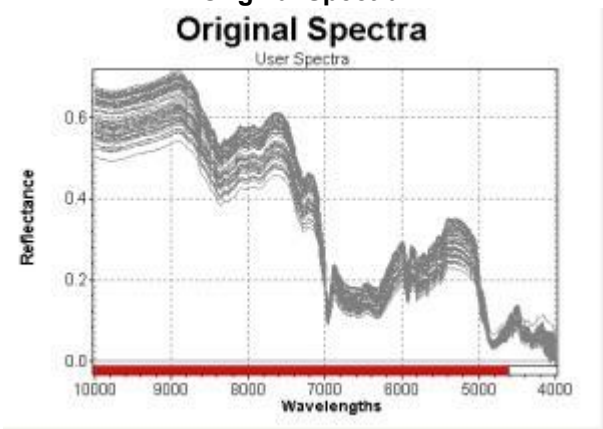
**Transformation**

Transformation can be applied to modify the absorption peaks.

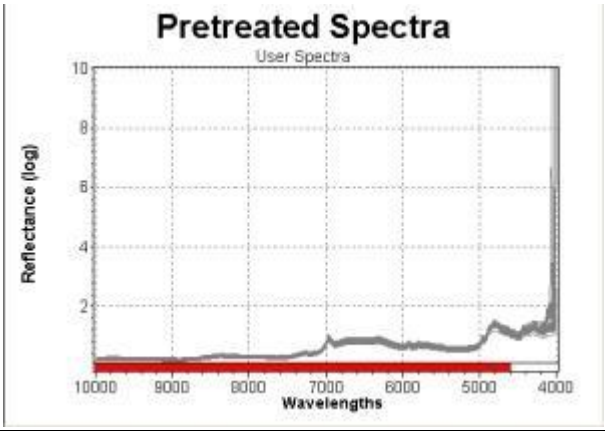
Tip: for liquids, in case the absorption peaks are not too high (thickness is too high). Not suggested for solids (scattering).

Overview of all Transformation pretreatments:

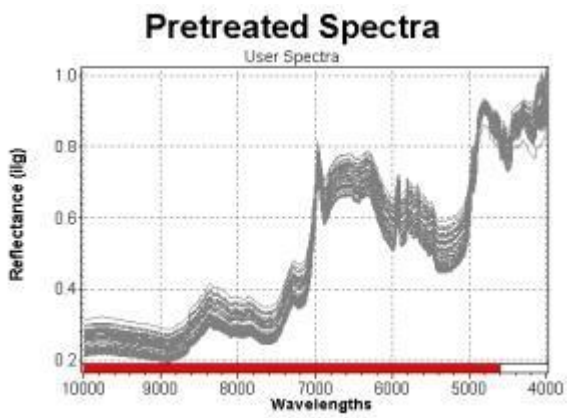
**Original Spectra**



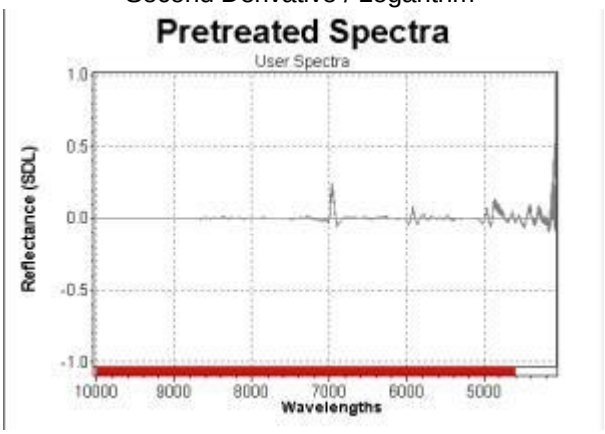
Absorbance



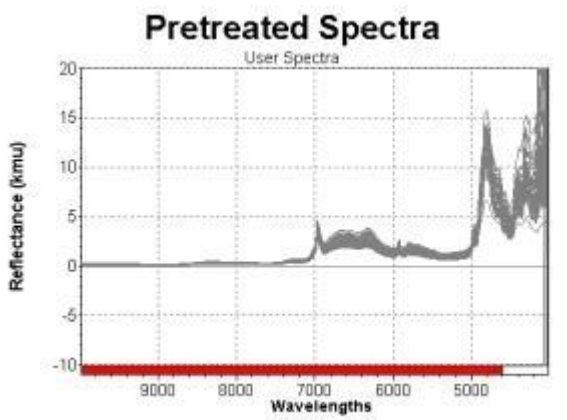

Absorbance invers



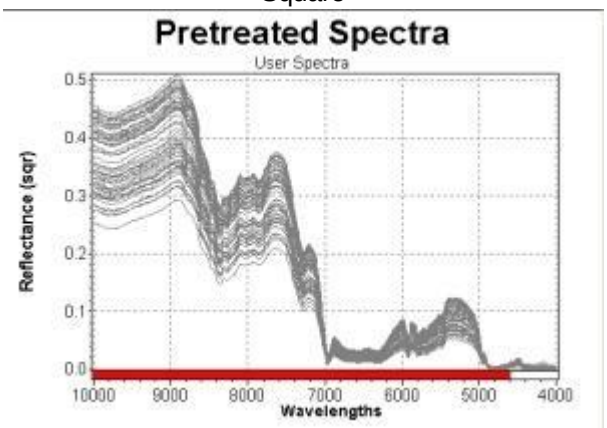

Second Derivative / Logarithm



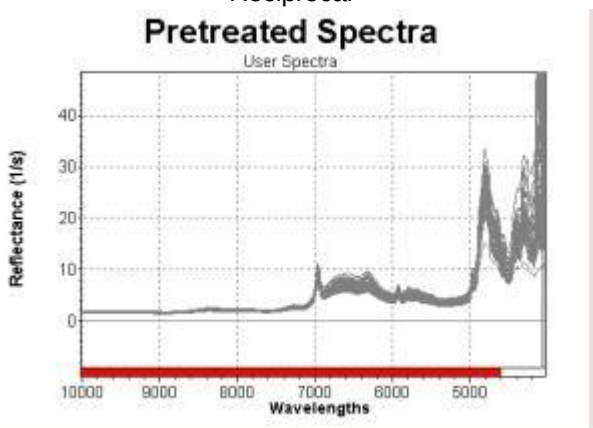

Kubelka Munk




Square

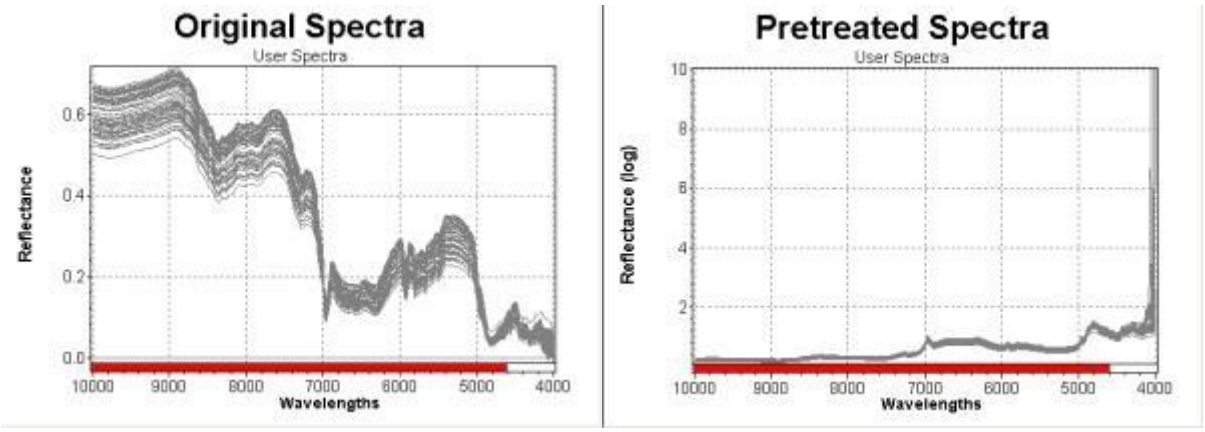



Reciprocal



**Absorbance**

It is useful in case the Lambert-Beer law is valid (not too high absorption).




Formula:

$$A = \log\ (1\ /\ T)$$

Implemented as:

$$A = -\log_{10}(T)$$
$$= -\frac{\log_e(T)}{\log_e(10)}$$
$$= -0.4342944819021\log_e(T)$$
$$T \in R^+$$

Type

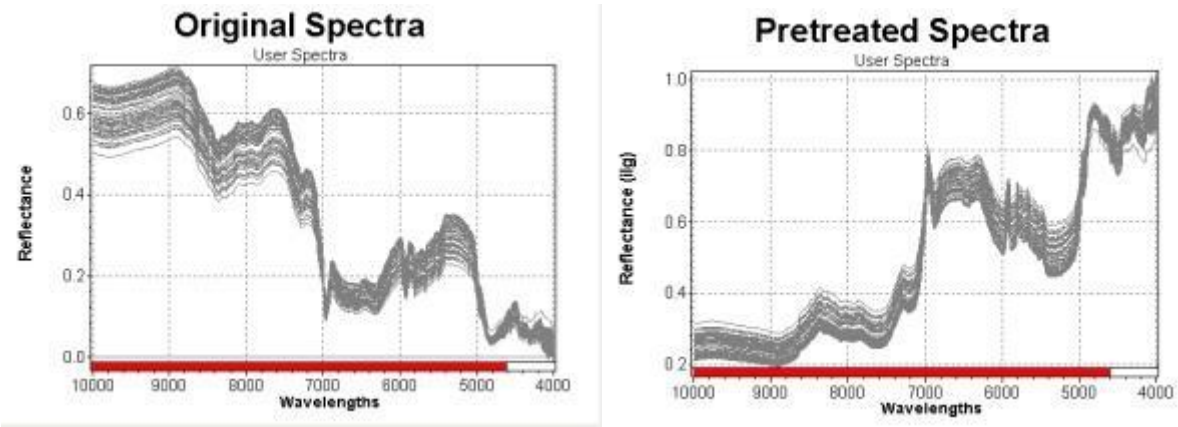
This pretreatment is not depending on the calibration wavelengths.

***NOTE***

NIRCal 5 calculates maximum 10 absorption units (means: R = 0.0000000001).

**Absorbance inverse**

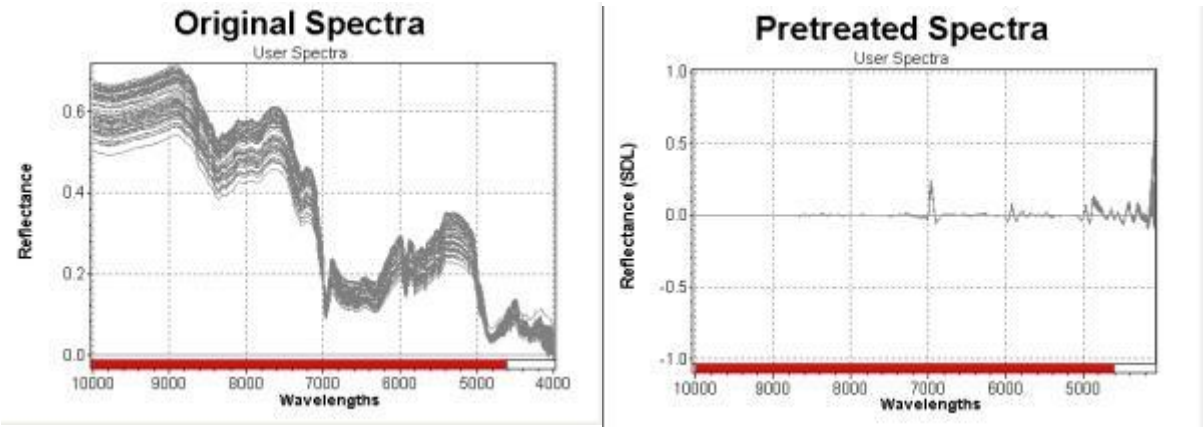
This is simply the inverse function of the ordinate transformation into absorbance; i.e. an absorbance spectrum will be transformed into transmission/reflectance.




Formula:
$ilg = 10^{-x}$

Type
This pretreatment is not depending on the calibration wavelengths.

**Second Derivative / Logarithm**

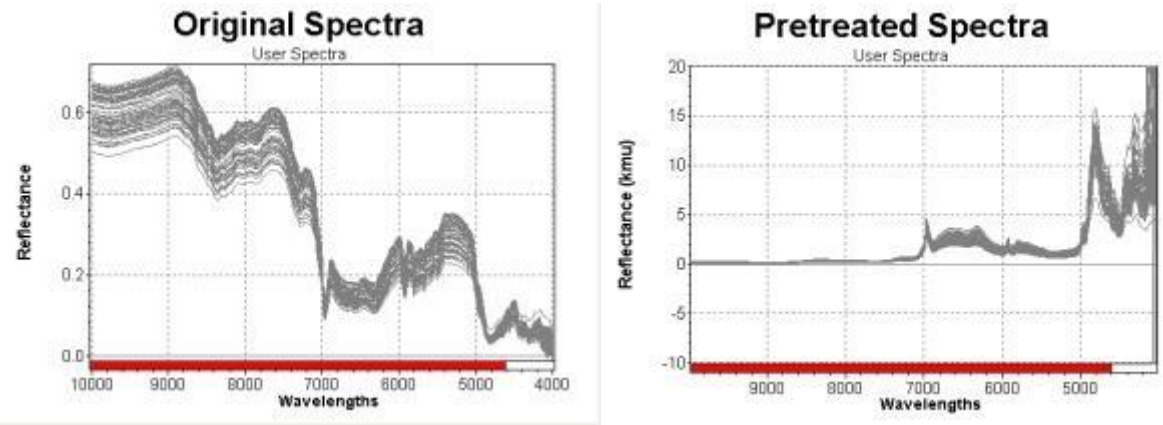



Formula:

$$SDL = \frac{A''}{\log(1+A)}$$

Type
This pretreatment is not depending on the calibration wavelengths.

**Kubelka Munk**



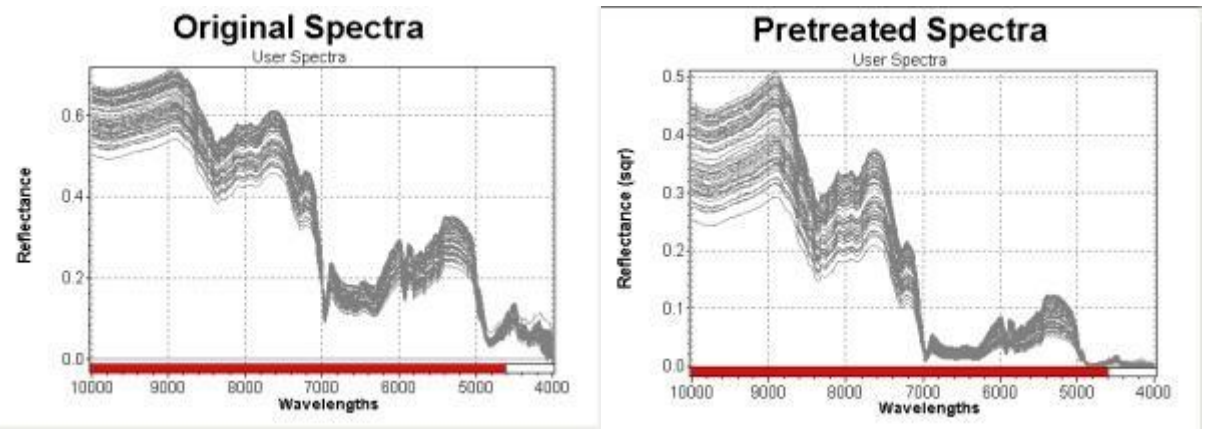

Formula:

$$KM = \frac{(1-T)^2}{2T}$$

Type
This pretreatment is not depending on the calibration wavelengths.

**Square**

Square is a tool for contrast amplification in spectra.



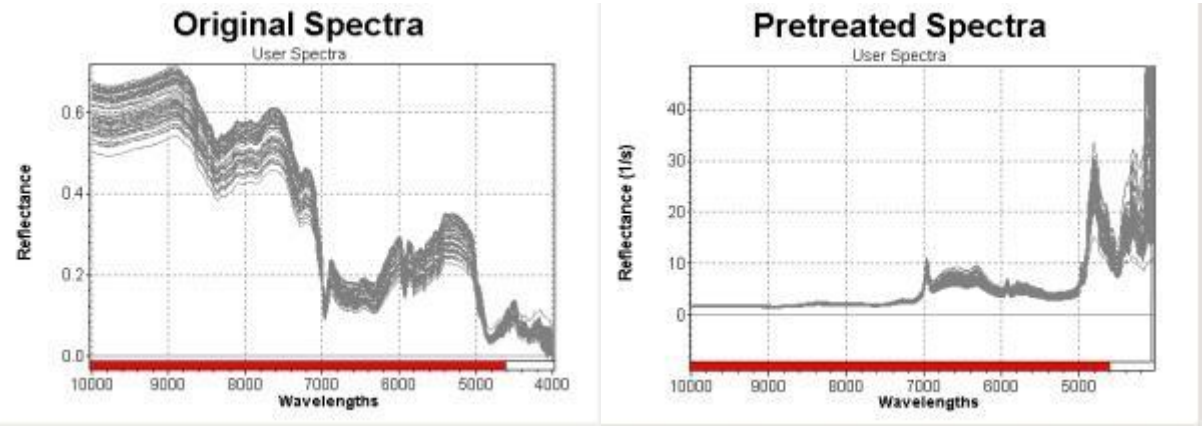

Formula:

$$SQ = S^2$$

Use
Contrast enhancement.

Type
This pretreatment is not depending on the calibration wavelengths.

**Reciprocal**




Formula:

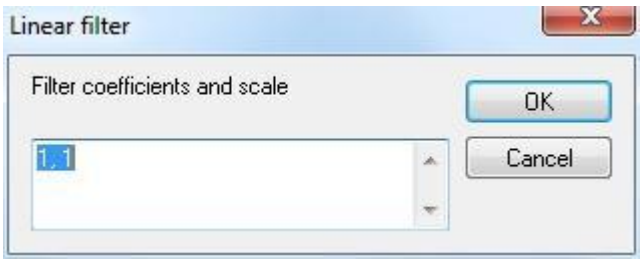
$$S = \frac{1}{S}$$

Type
This pretreatment is not depending on the calibration wavelengths.

## 3.5.8 Filter

**Linear Filter**

Linear filters allow the application of user defined pretreatments. For instance it is possible to apply average smoothing, derivative, low and high band pass filtering. The default linear filter in NIRCal contains a neutral setting of 1, 1.




It is best to show examples for the explanation of linear filters.
The coefficients for the lienar filter will be explained for a Savitzky-Golay first derivative. In the original paper (Anal. Chem. 36, 1627-1639 (1964) the following formula was given using orthogonal cubic polynomes (nine points):

$$f'(x_i) = \frac{-86f(x_{i+4}) + 142f(x_{i+3}) + 193f(x_{i+2}) + 126f(x_{i+1}) - 126f(x_{i-1}) - 193f(x_{i-2}) - 142f(x_{i-3}) + 86f(x_{i-4})}{1188}$$

This formula will be transferred into the following coefficients for linear filters:
-86, 142, 193, 126, 0, -126, -193, -142, 86, 1188

The coefficient for the highest x-value x(i+4) is -86, for the lowest x-value x(i-4) the coefficient is +86. The coefficient is 0 for that term, where there is no contribution; in this example it is in the middle, i.e. x(i+0). The last value of the coefficients is the Scale Factor. The sum will be divided by the scale factor.
As next example a three point average smoothing is used. The formula is: x'(i
= ( x(i+1) + x(i) + x(i-1) ) / 3

The corresponding coefficients for this example are: 1, 1, 1, 3

Linear filters are very useful when spectra recorded with a different resolution have to be treated similarly to original spectra with e.g. only one third of data points (step from NIRFlex N-400 to NIRFlex N-500). For similar prediction results of calibrations smoothing and deriving functions should be adapted. In the following tables the linear filters for the basic smoothing- and derivative functions are summarized.

| Derivatives | Linear Filter Coefficients |
|---|---|
| First Derivative BCAP (db1) | 1,1,0,-1,-1,4 |
| First Derivative Taylor 3 points (dt1) | 1,0,-1,2 {= First Derivative Finite Differences} |
| First Derivative Savitzky-Golay 9 points (dg1) | -86,142,193,126,0,-126,-193,-142,86,1188 |
| Second Derivative BCAP (db2) | 1,-2,1,2 |
| Second Derivative Taylor 3 points (dt2) | 1,-2,1,2 )* |
| Second Derivative Savitzky-Golay 9 points (dg2) | 28,7,-8,-17,-20,-17,-8,7,28,462 |
| Second Taylor 3 Points Segment5 Gap5 (ds2) | 1,1,1,1,1,0,0,0,0,0,-2,-2,-2,-2,-2,0,0,0,0,0,1,1,1,1,1,5 |
| {Second Derivative Finite Differences *} | {1,0,-2,0,1,1} |
| Third Derivative Taylor 5 Points (dt3) | 1,-2,0,2,-1,2 |
| {Third Derivative Finite Differences} | {1,0,-3,0,3,-1,1} |

*) due to upward compatibility reasons, identical to BCAP
{…} not included in NIRCal 5; for “Finite Differences” see: Norris and Williams in “NIR-Technology in Agricultural and Food Industries” p. 48

| Smoothing | Linear Filter Coefficients |
|---|---|
| Average 3 Points (sa3) | 1,1,1,3 |
| Average 9 Points (sa9) | 1,1,1,1,1,1,1,1,9 |
| Savitzky-Golay 9 Points (sg9) | -21,14,39,54,59,54,39,14,-21,231 |

There are various ways for adapting Filters for the step between data of e.g. the NIRFlex N-400 and NIRFlex N-500. As a simple recipe we suggest to repeat each filter coefficient 3 times and adjust the scale or insert zero coefficients between the coefficients and also here adjust the scale. The recipe is illustrated for the second derivative BCAP (db2) below.

| Number of data points: 500 | Linear Filter Coefficients |
|---|---|
| 500 | 1,-2,1,2 |
| | Various alternatives: |
| 1501 | 1,1,1,-2,-2,-2,1,1,1,3 |
| 1501 | 1,0,0,-2,0,0,1,1.2 |
| 1501 | 1,0,-2,0,1,0.7 |

For new NIRFlex N-500 datasets gap2 pretreatements have been introduced with NIRCal 5.2. The following tables show the linear filter coefficients for gap2 pretreatements.

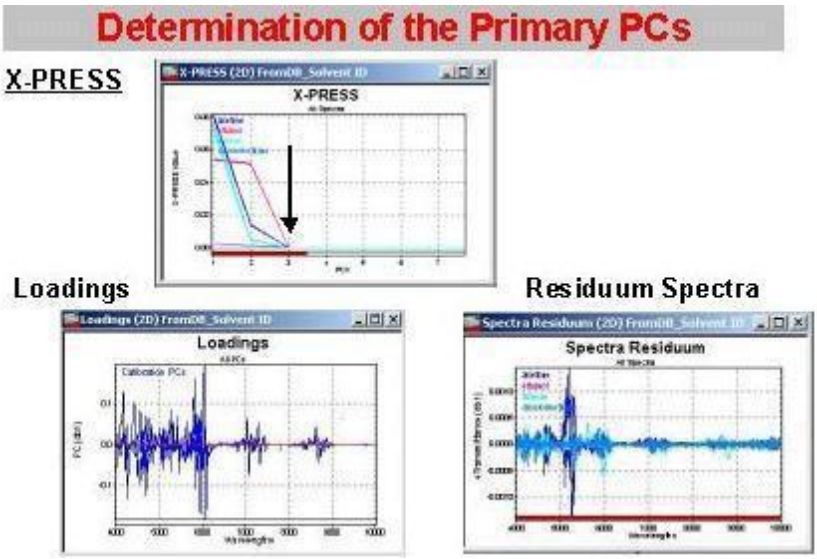
| Gap2 Derivatives | Linear Filter Coefficients |
|---|---|
| First Derivative BCAP Gap2 (db1g2) | 1,0,0,1,0,0,0,0,0,-1,0,0,-1, 4 |
| First Derivative Taylor 3 Points Gap2 (dt1g2) | 1,0,0,0,0,0,-1, 2 |
| First Derivative Savitzky-Golay 9 Points Gap2 (dg1g2) | -86,0,0,142,0,0,193,0,0,126,0,0,0,0,0,-126,0,0,-193,0,0,-142,0,0,86, 1188 |
| Second Derivative BCAP Gap2 (db2g2) | 1,0,0,-2,0,0,1, 2 |
| Second Derivative Taylor 3 Points Gap2 (dt2g2) | 1,0,0,-2,0,0,1, 2 *) |
| Second Derivative Savitzky-Golay 9 Points Gap2 (dg2g2) | 28,0,0,7,0,0,-8,0,0,-17,0,0,-20,0,0,-17,0,0,-8,0,0,7,0,0,28,462 |
| Second Taylor 3 Points (Segment5 Gap5) Gap2 (ds2g2) | 1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-2,0,0,-2,0,0,-2,0,0,-2,0,0,-2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1, 5 |
| Third Derivative Taylor 5 Points Gap2 (dt3g2) | 1,0,0,-2,0,0,0,0,0,2,0,0,-1, 2 |

| Gap2 Smoothing | Linear Filter Coefficients |
|---|---|
| Average 3 Points Gap2 (sa3g2) | 1,0,0,1,0,0,1,3 |
| Average 9 Points Gap2 (sa9g2) | 1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,9 |
| Savitzky-Golay 9 Points Gap2 (sg9g2) | -21,0,0,14,0,0,39,0,0,54,0,0,59,0,0,54,0,0,39,0,0,14,0,0,-21, 231 |

## 3.6 Primary PCs

Primary PCs are used for the reconstruction of measured and pretreated spectra. The optimum number of primary PCs will be achieved if the spectra are described down to the limit of noise of the spectrometer. Primary PCs influence the spectra reconstruction and therefore are responsible for the residuum spectra and the residual.
The more primary PCs are used, the smaller the allowed residual of the calibration. **If too many primary PCs are selected, a calibration is overfitted**. In this situation only the calibration spectra will be identified correctly but not necessarily spectra of additional validation samples represented in the project.
**If too few primary PCs are selected, the calibration might not be selective enough for the products (underfitted)**.
Tools for the selection of the optimal number of primary PCs:

The X-PRESS function shows from which PCs on, an additional PC does not improve the reconstruction of the spectra. The smallest number of PCs that still shows changes should be selected.
The Loadings / PCs themselves can be judged. PCs that appear noisy should not be used.
The Residuum spectra can be checked. The amplitude of these spectra with the selected number of primary PCs should be about the same.




In this example, 3 primary PCs are necessary, each loading contains spectral information. PCs 4 to 10 are very small and have only noise with no useful information.
Check the number of primary PCs with the help of the graphic X-PRESS (in the Overview, or open the graphic in the Menubar : **Graphics / X-PRESS**) and Loading.

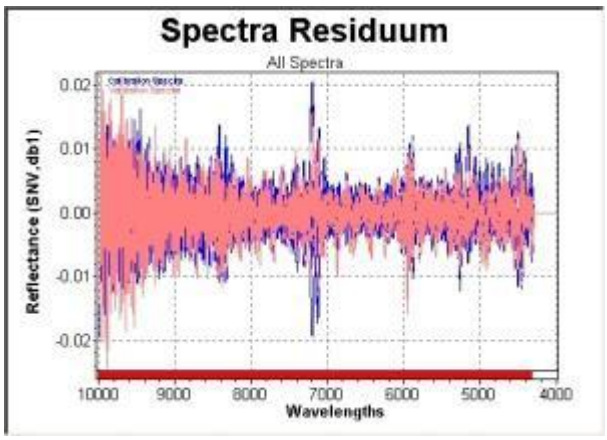
Additional tool:

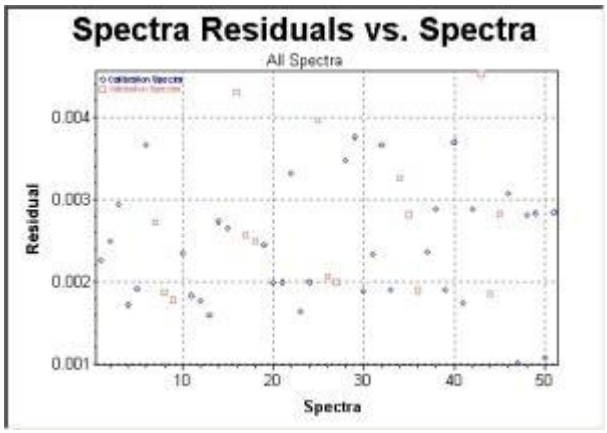
The Eigenvalue of the PCs may provide information of the information content of the PCs. The reliability of the Eigenvalue is dependent on the applied data pretreatment.

The selected number of primary PCs should be adjusted and the calibration recalculated.

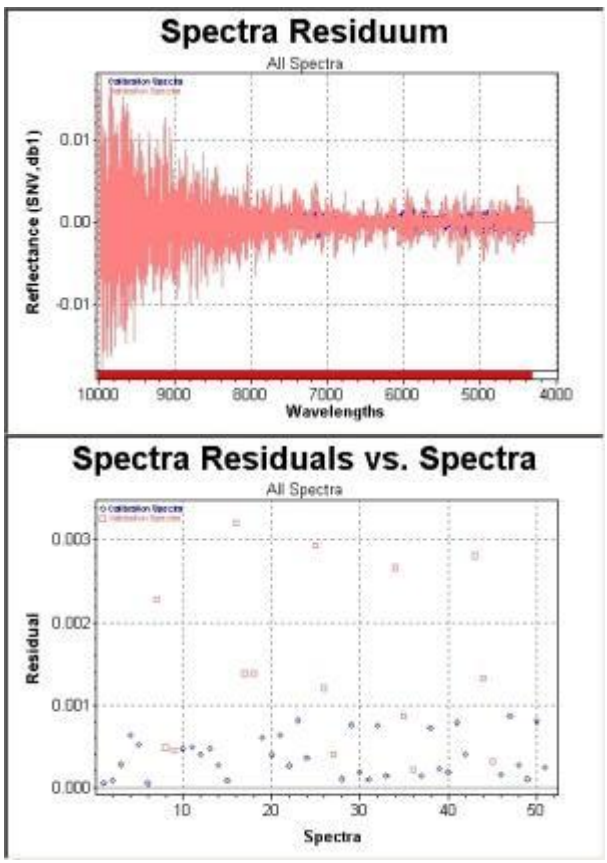
The **more primary PCs** are used, the **smaller** the allowed **residual** of the calibration.

- If the primary PCs are selected correctly, the residuum spectra are similar for the C- and V-Set.
  Example of correct primary PC selection:



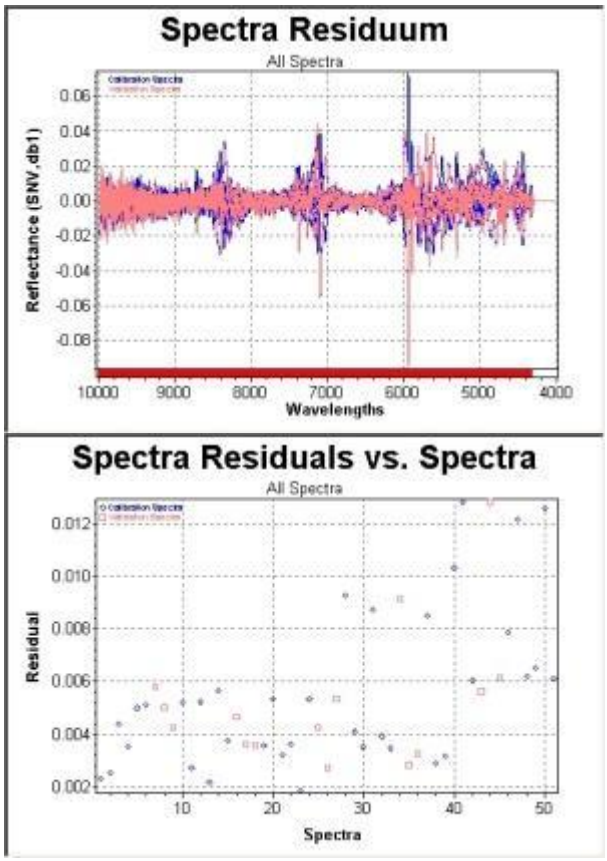



- If **too many primary PCs** are selected, a calibration is **overfitted**. In this case only the calibration spectra will be identified correctly but not necessarily the validation samples in the project.
  Example of overfitting:




The residuals of the C-Set are much lower, as for the V-Set.

- If **too few primary PCs** are selected, the spectra are **underfitted**, the calibration might not be selective enough for the products.
  Example of underfitting:



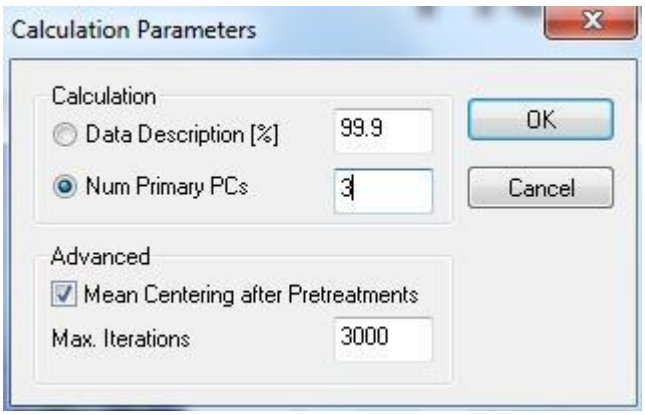

There are spectral characteristics, which are still not fitted, the residuum spectra are bigger, as the noise of the instrument.

***NOTE***
Avoid underfitting and overfitting! The PCs / Factor Selection Wizard can help with an estimate.

## 3.7 Adjusting Primary PCs (Calculation Parameters)

Menu: Calibration \ Parameter \ Calculation

Icon:




The primary PCs are edited in the Dialog called Calculation Parameters. To check the result of the Setting of the Primary Principal components open the graphic Spectra Residuum: Menu: **Graphics / Spectra /** Residuum:

***NOTE***
Avoid overfitting: do not use too many principal components with little information.

Calculation

| | |
|---|---|
| Data Description [%] | To determine the number of primary PCs with a certain Data description (in %).<br>Available for **Cluster** and **PCR** Method only. |
| Num Primary PCs | Change the value at Num Primary PCs to the number defined by the graphic "X-PRESS". The limit of maximum number PCs is number of C-Set spectra. |

Advanced

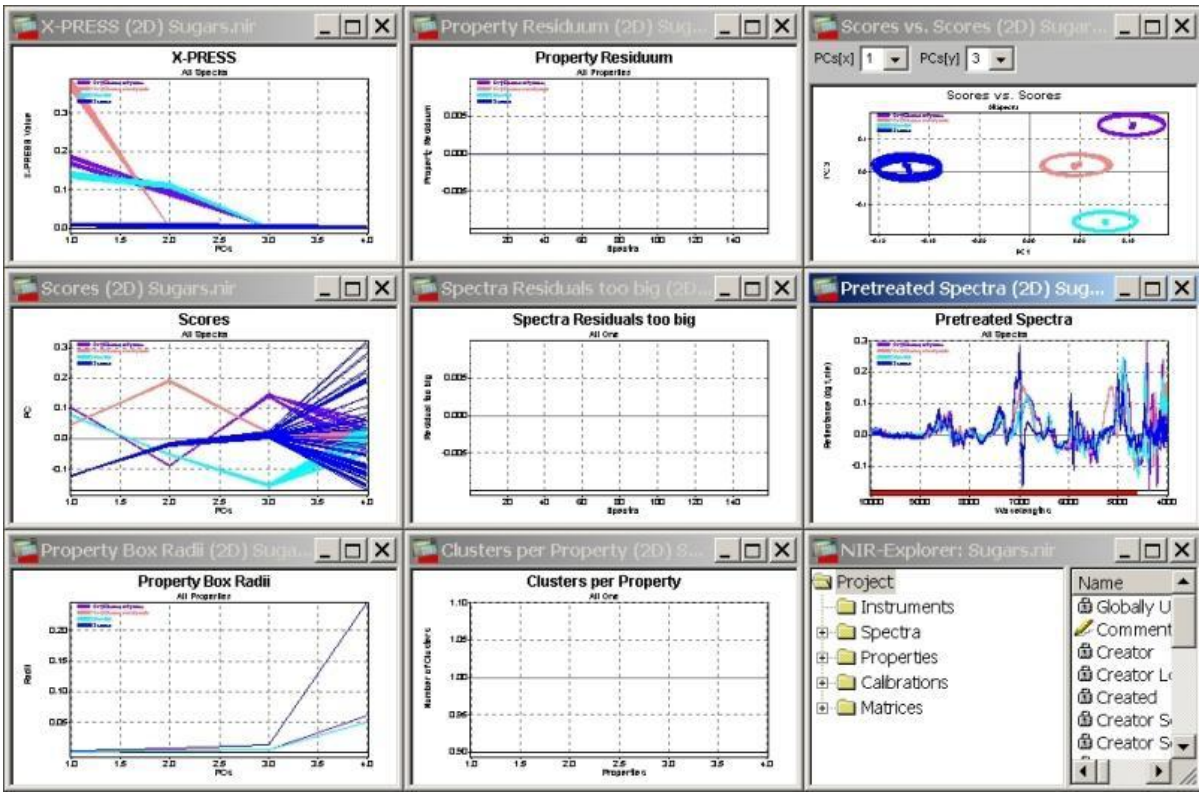
| | |
|---|---|
| Mean centering after Pretreatments | Sets the Mean Centering after finishing Pretreatments ON or OFF.<br>**All implemented algorithms (PCA, PCR, PLS) prefer a mean centering of the data.**<br>In very special situation it is possible to try the calibration without mean centering (e.g. SIMCA). |
| Max. Iterations | The maximum number of iterations for the internal PCA, PCR and PLS algorithm.<br>When the maximum iterations are reached the algorithm will stop the calculation at the actual PC. All the following calculations will be done with this actual number of PCs instead the desired number of primary PCs.<br>Tip: for very difficult data set, set the number of iterations up to 3'000.<br>NOTE: Max. iteration limit is the number of C-Set spectra. |

## 3.8 Secondary PCs

### 3.8.1 Secondary PCs for qualitative Calibration

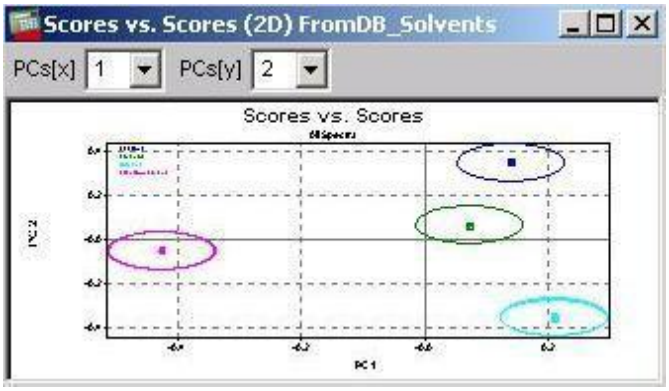
The secondary / calibration principal components are responsible for the **separation of the different substances** and are used for the **tolerance radii calculation**. Several graphics or tables can help decide which principal components will lead to a successful separation of the properties.

The most comprehensive results are in the Menubar: **View / Overview**.



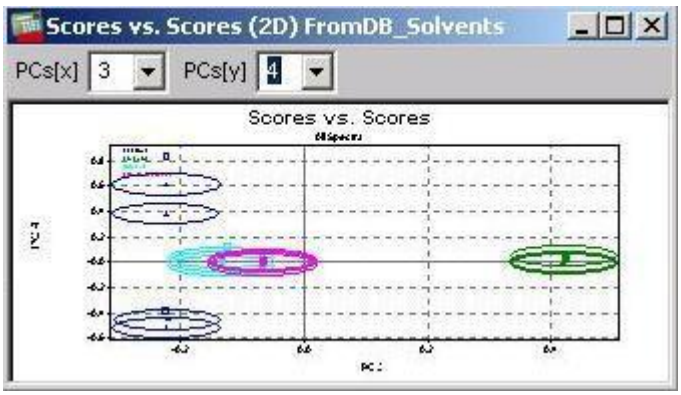
**View Scores Graphic**

Enlarge the **Scores / Scores** window in the Overview (3th column, 1st window).




There is a clear separation of the 4 different properties spectra according to the PC 1 and 2.

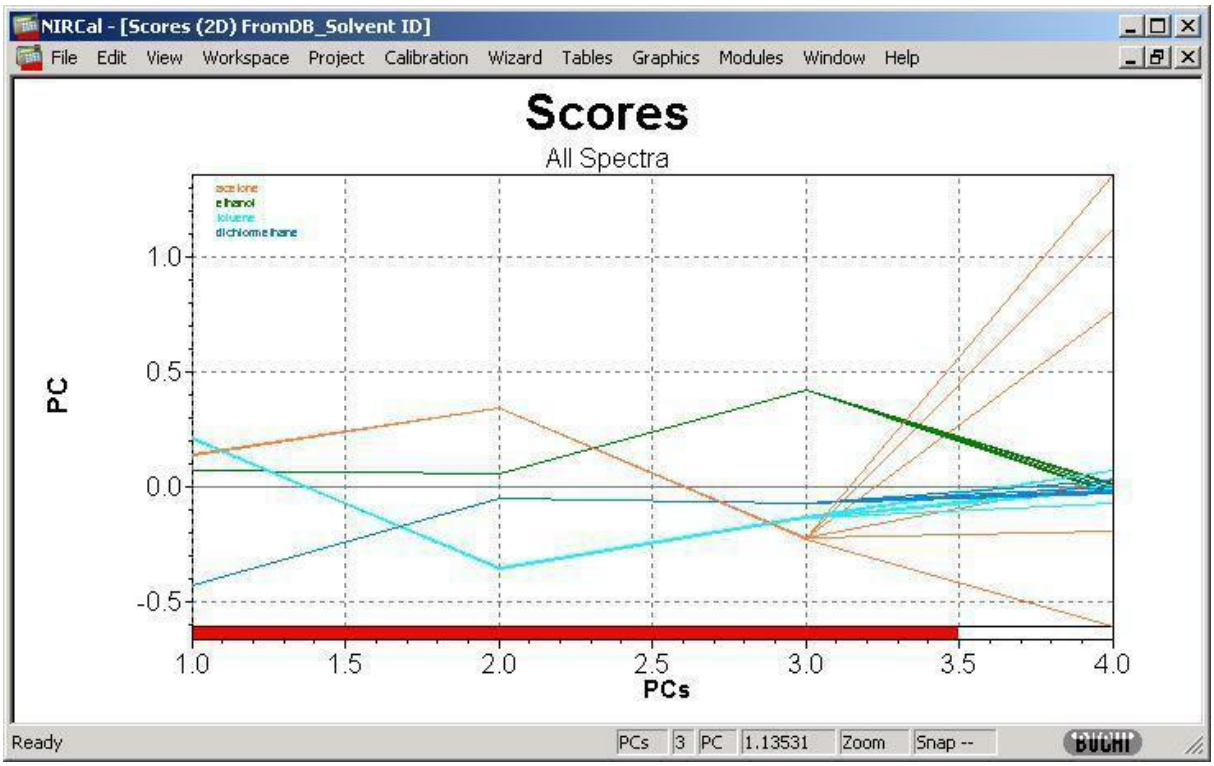
There is no separation with the scores of PC 4.




PC 1, 2 and 3 are secondary PCs, the PC 4 is not chosen for a secondary PC.

**Scores against Spectra:**

Enlarge the **Scores** window in the Cluster Method Overview (1st column, 2nd window). Good reproducible scores belong to PCs 1, 2 and 3, the PC 4 scores are not repeatable => PC 4 can not be used as secondary / calibration PC.

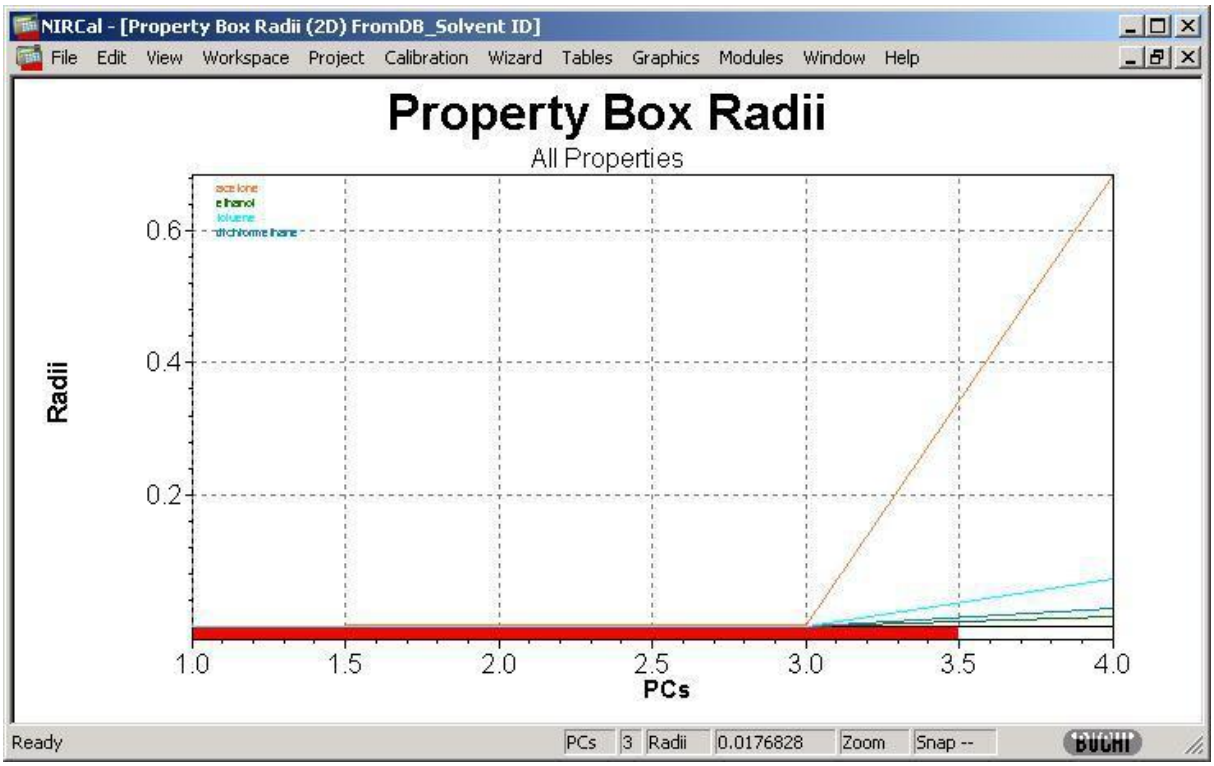



**Property Box Radii:**

Enlarge the **Property Box Radii** window in the Overview (1st column, 3rd window). Property Box Radii represents the scattering of the scores

- small values (normally below 0.1) mean repeatable
- scores, big values mean bad repeatability.

The scores of PC 4 are not repeatable, PC 4 can not be used as calibration PC.

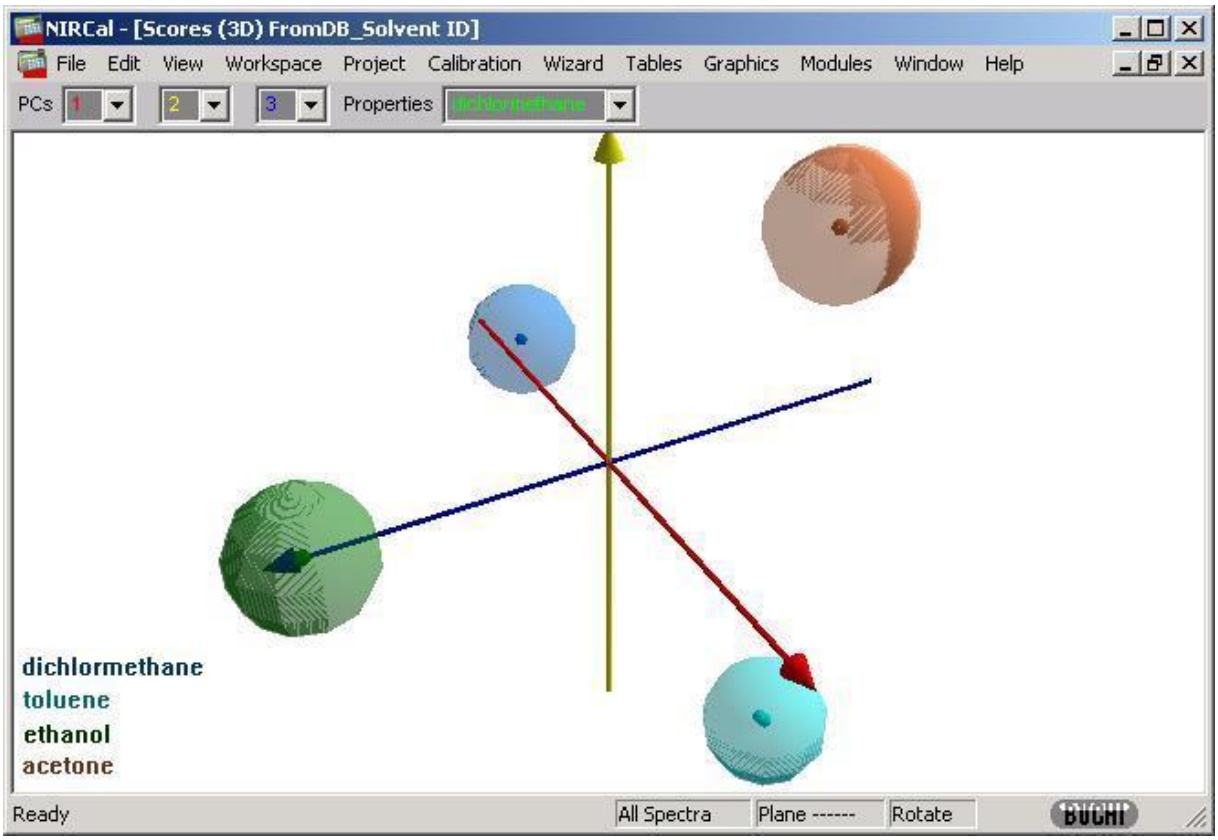



Additional tool:

The **Property Score Disorder** values show how effective a particular PC separates different properties (substances) from each other. By scanning the score values in one direction of a PC and counting the changes between the membership of A or B, the disorder value is achieved. If a PC completely separates all calibration properties, **the smallest disorder value is: [the total number of calibration properties-1]**.
PCs with small disorder values are possible calibration PCs.

**Summarising all the information available from the graphics for this example, the first 3 principal components (1-3) should be used as secondary PCs.**

The separation with 3 selected calibration PCs is visible in a 3 D graphics.




The selected number of secondary PCs should be adjusted and the calibration recalculated.

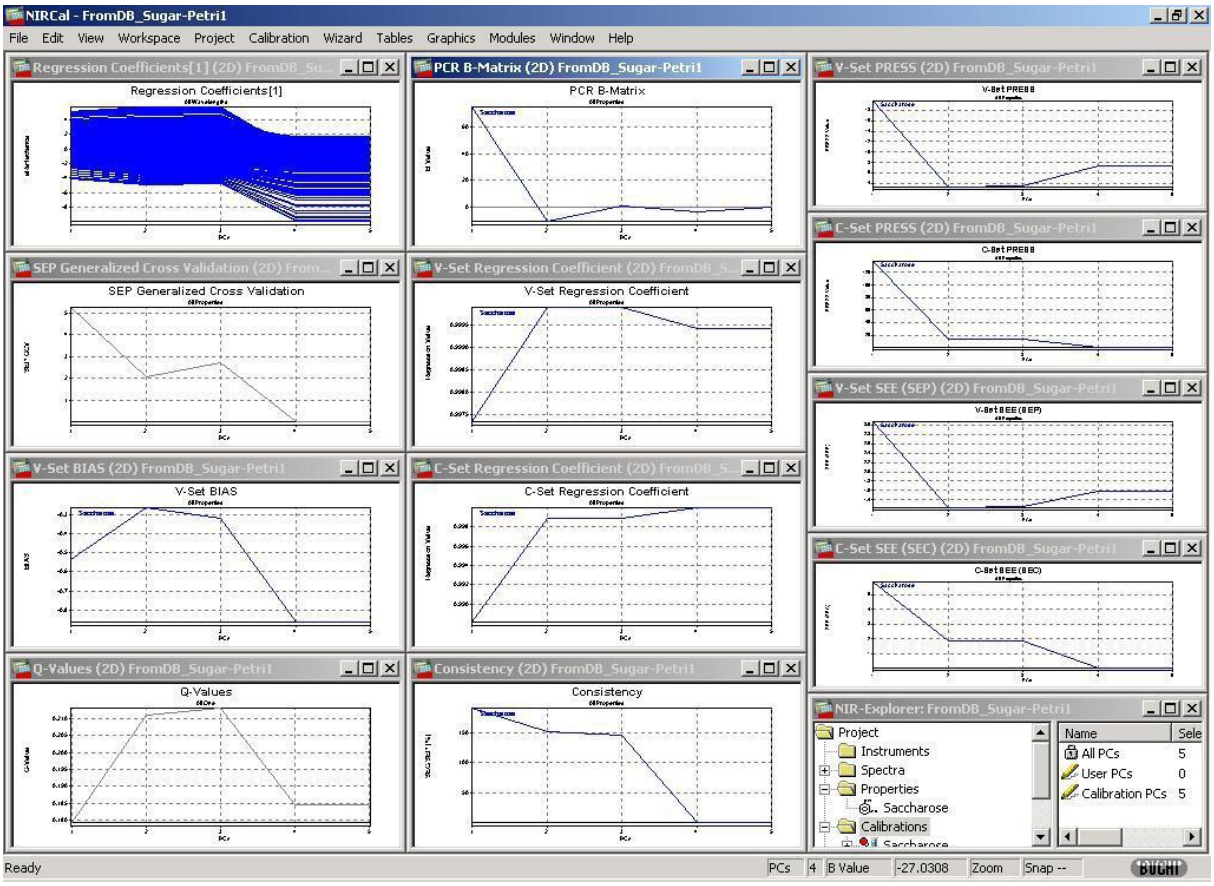
***NOTE***
In case more than 3 secondary PCs are selected, there is no possibility to see the clusters with the real radii in a 3D-Plot (more than 3D is not existent).
Try adjusting the xyz axis using the drop down list.

***NOTE***
To see the scattering in multi dimensions use Grafics / Scores / Multi 2D-Scatter.

## 3.8.2 Secondary PCs for quantitative Calibration

The secondary / calibration principal components are responsible for the **parameter calculation** and for the quality of the calibration. Several statistical values can help this decision.

The most comprehensive results are in the Menubar: **View / Secondary PCs Selection**.




Select the number of PCs, where:

- the in the plot regression coefficients[1] the PCs, which have similar constant value, are good for the calibration, big deviation indicates over fitting;
- the SEP Generalized Cross Validation is small (about the value of the standard deviation of lab method);
- the V-Set Bias is around zero;
- the Q-Value is high;
- the absolute value of the PCR B-Matrix is high (not available by PLS);
- the V- and C-Set regression coefficients are as close to one as possible;
- the consistency is around 100;
- the V- and C-Set PRESS are as small as possible;
- the V- and C-Set SEP and SEE (SEC) are as small as possible and are similar (consistency).

Summarising all the information available from the graphics for this example, the first 3 principal components (1-3) should be used as secondary PCs.

It is possible that different numbers of PCs are ideal for different selection criteria. In this situation the different secondary PCs should be adjusted, the calibrations recalculated and the results compared.

In general for the C-Set a higher number of PCs always improve the result. For the V-Set, after a certain number of PCs the result can be even worse. The optimum should be selected.

The selected number of secondary PCs should be adjusted and the calibration recalculated.

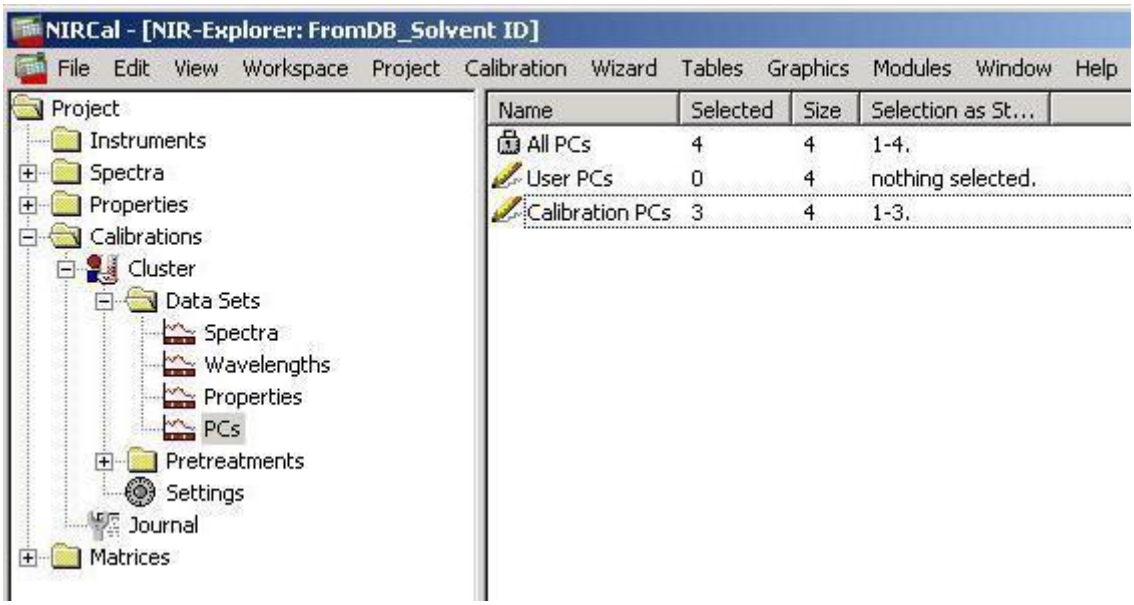
**Note**: The **PLS** algorithm calculates the PC‘s with the highest correlation to the property values, that means the **secondary PC selection can not have a gap**. The real selection is always "**1 to the last selected**". For 1-2, 4-7 secondary PC‘s selection the internal used secondary PC`s are 1-7 due to the PLS algorithm. This allows easy switching between the methods form PCR to PLS and back without losing the PC selection.

## 3.9 Secondary PC Selection

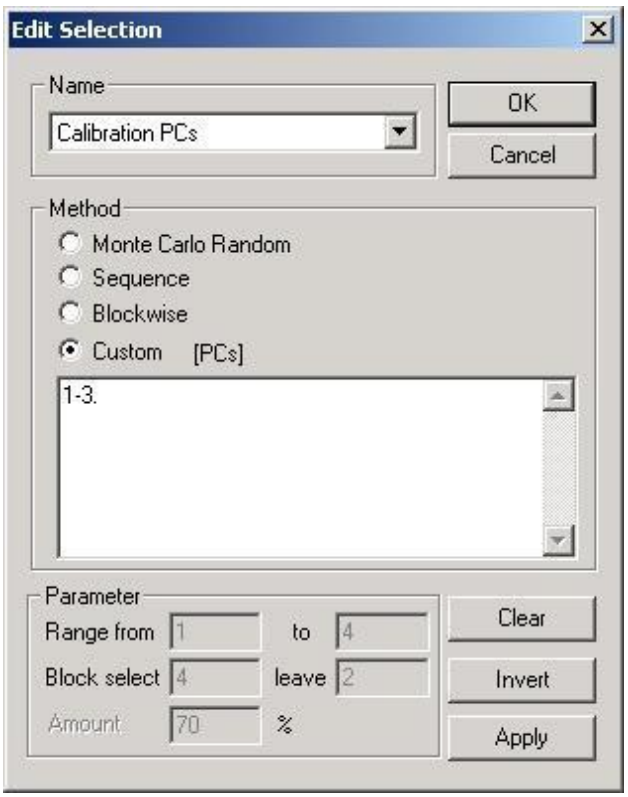
The number of secondary PCs is limited by the number of primary PCs. From this number of PC any number can be selected using the NIR-Explorer or the Edit dta Set dialog.

***NOTE***
In PLS a secondary PC selection like in the PCR or CLU is not possible, NIRCal uses all PCs from PC 1 to the last selected PC as secondary PCs.

### 3.9.1 Secondary PC Selection using NIR-Explorer

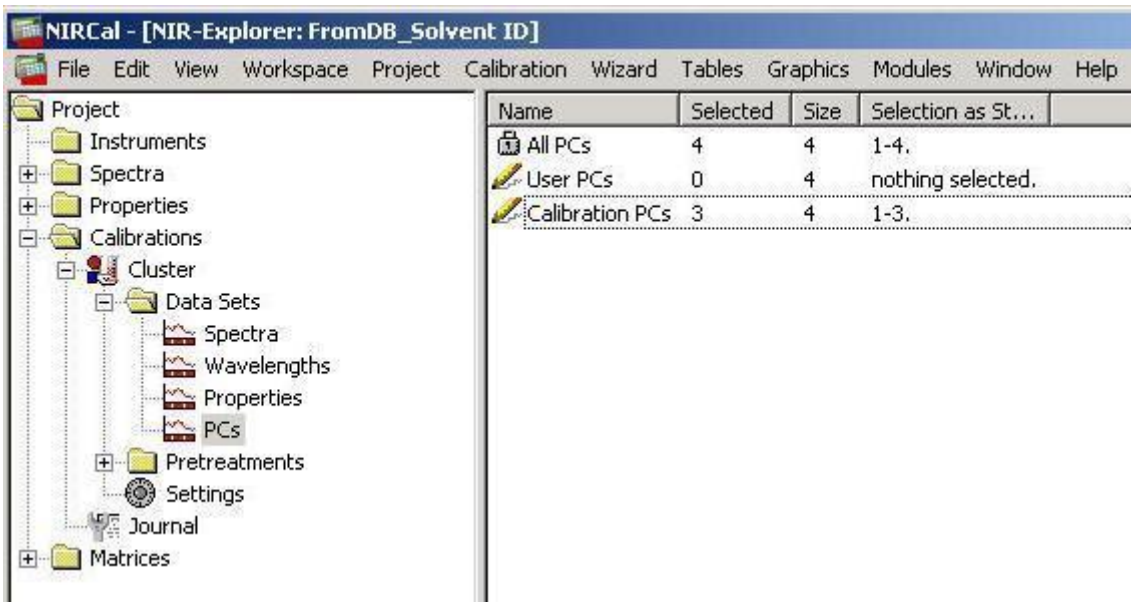



By double clicking on **Calibration PCs**, the selection dialog appears




Steps sequence:

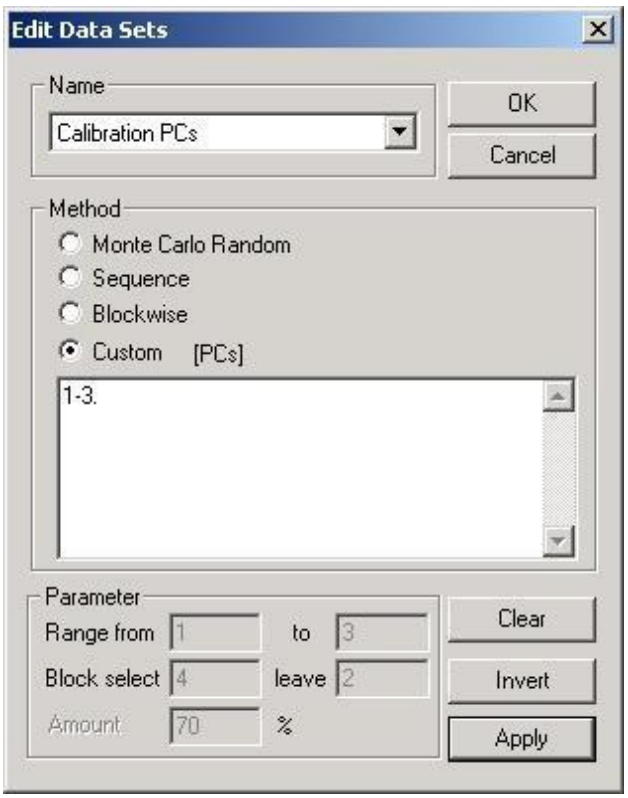
1. Select **Custom**;
2. **Type** in the selection;
3. Click on **OK**.




In the calibration the selection is shown.

### 3.9.2 Secondary PC Selection using Edit Data Sets Dialog

Click on the icon:  or open the Edit Data Sets dialog under the Menubar: **Calibration / Change data Sets / Edit Data Sets**.



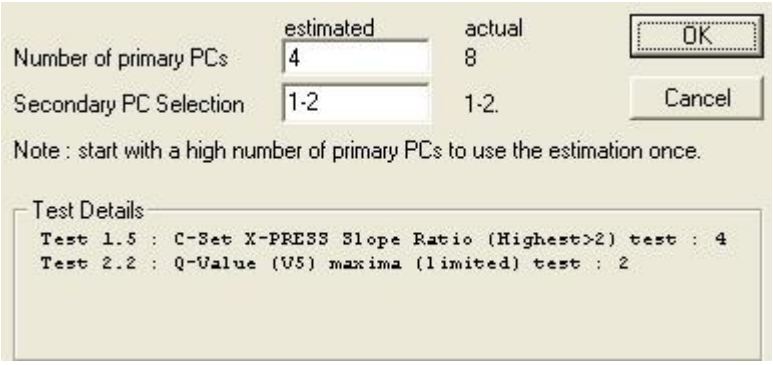

Select **Calibration PCs** from the drop down menu. Select **Custom**, **edit** the selection and press **Apply**.

***NOTE***
After setting the secondary principal components, the calibration should be recalculated.

## 3.10 PCs (Factor Selection Wizard)

Menu: Calibration / Change Data Sets / PC

Icon: 




Under **actual** the current settings for primary and secondary PCs (Factors) are displayed.

Under **estimated** the wizard suggest an optimized selection. The selections are changed manually.
See: Adjusting Primary PCs and Adjusting Secondary PCs

***NOTE***
The estimated selections can only be smaller then the actual selections in the current project. Therefore it is recommended to change the selections for the PCs to a **relative high value**.

Related Topic: SIMCA, Transform Cluster to SIMCA

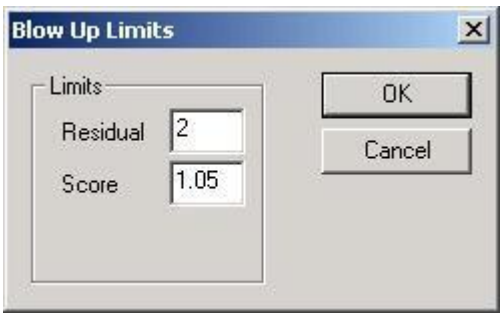
## 3.11 Blow Up Limits

Menu: Calibration / Parameter / Blow Up Limits

Icon: 

The dialog is different depending on whether a quantitative or qualitative method is selected. They can be used for the fine tuning of the calibrations.

### 3.11.1 Blow Up limits quantitative calibration

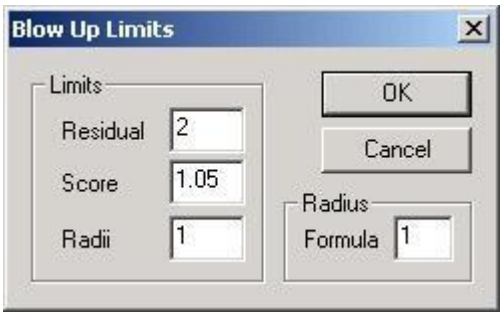


Limits

Residual The maximum residual, allowed for the application is determined by two (2) times the maximum residual of the calibration set. The default value is 2 and can be edited by the user.

Scores The calibration spectra "maximum and minimum Scores x Blow Up limit" are used for possible score outlier detection in the prediction protocol. Possible outliers are searched with the secondary PCs. The default value is 1.05 and can be edited by the user.

### 3.11.2 Blow Up limits qualitative calibration




For a qualitative calibration (Cluster, SIMCA) the dialog box additionally has the possibility to edit the Radii Blow Up limit and the Radius Formula.

Limits

Radii For all calibration spectra a radius is calculated and will be used for the application. These radius can be enlarge or decreased by editing the Radii Blow Up limit. The default value is 1 and can be edited by the user.
Reducing the radii will decrease the robustness, but increase the sensitivity of the calibration.

***NOTE***
Formula 1: it is suggested to **reduce** the Radii Blow Up limit (0<f<1) **for chemically different substances** to increase the sensitivity of the calibration.

**Do not use higher blow up limit as 1 with Formula 1: it can cause overlapping.**

Radius Formula

1 Formula 1: Radius = min(R1,(R1+R2)/2)/2 * Radii Blow Up

2 Formula 2: Radius = min(min(0.499*R1, R3*5*Radii Blow Up), R4*2.5*Radii Blow Up )

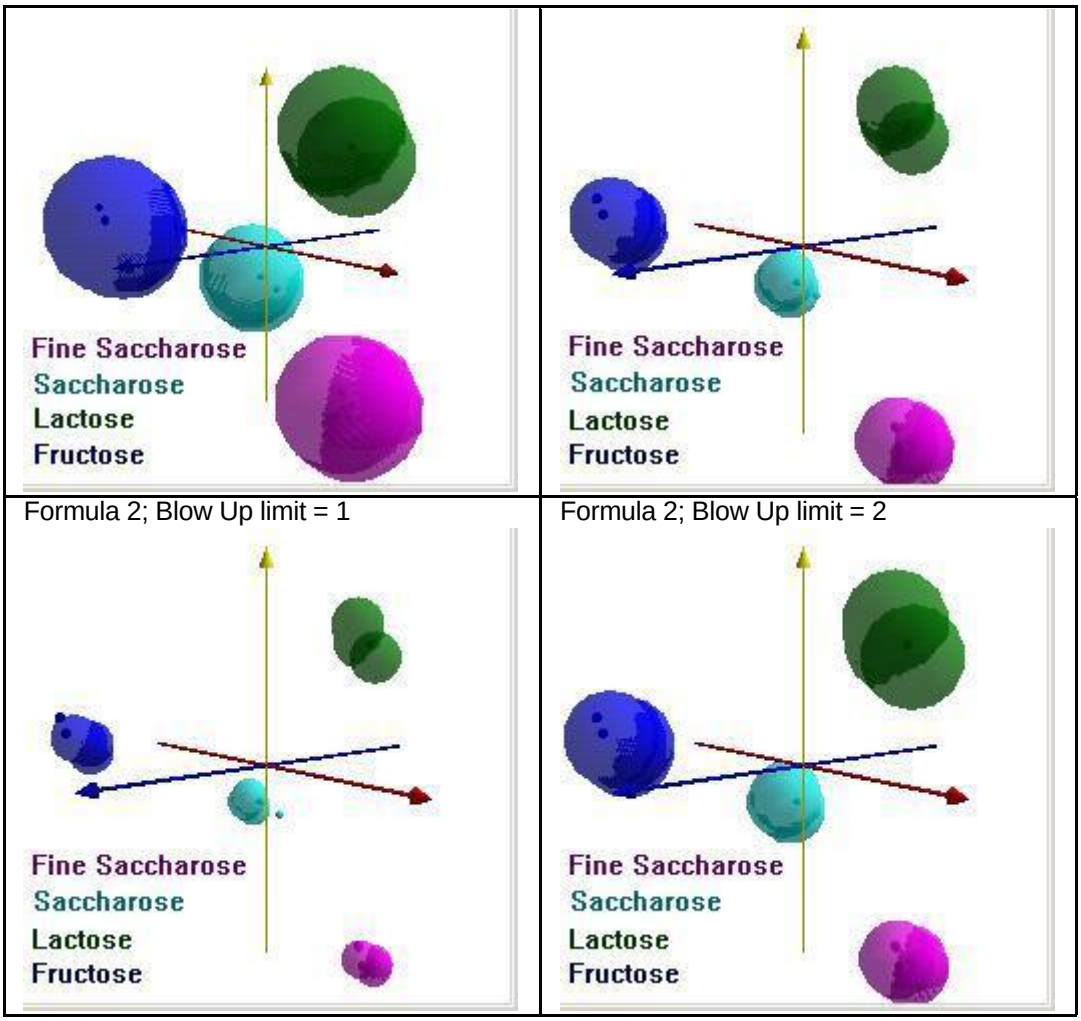
3 Formula 3: Radius = min(min( 0.5*R2, R3*5*Radii Blow Up), R4*2.5*Radii Blow Up)

See details in Cluster.

***NOTE***
Formula 1 or 2 are recommended for Cluster calibration.
Formula 3 is recommended for SIMCA.

Examples for different Formula and Radii Blow Up limits:

| Formula 1; Blow Up limit = 1 | Formula 1; Blow Up limit = 0.5 |
|---|---|




Reducing the radii results more selective calibration but increases the danger to get separated cluster for the same substances (Cluster / Property> 1).

## 3.12 Outlier Detection

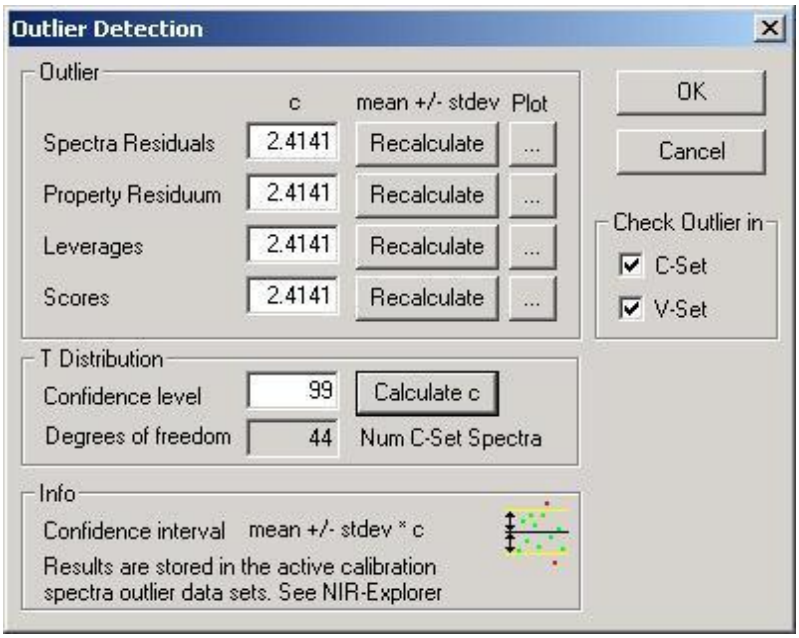
Menu: Calibration / Outlier Detection / Advanced

Icon: 

Outliers are spectra, that differentiate strongly from spectra of the same class. They can be detected

**visually**, check the spectra in the menu Graphics, in the calibration curve or in e.g. score plots;
**statistically** with the "Outlier Detection" module.

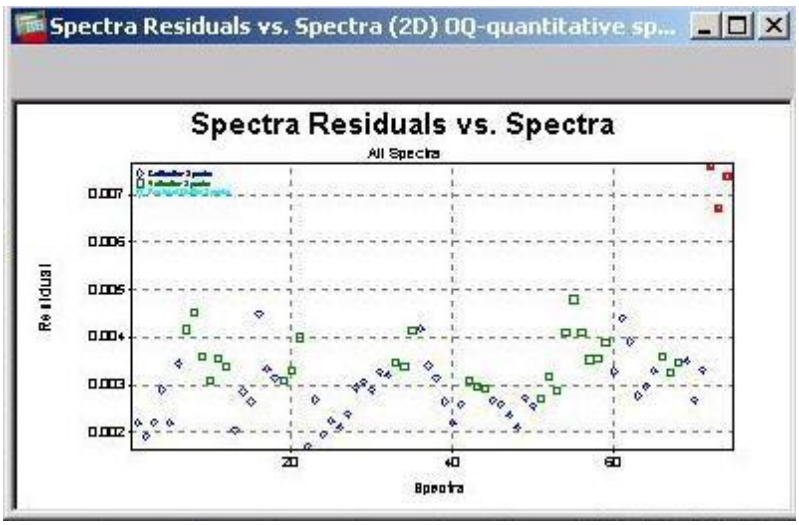
The principle of the calculation with the **advanced** "**Outlier Detection**" is explained for the residual outliers.
The average (mean) and the standard deviation (stdev) of the C-Set residuals are calculated. Each spectrum outside the confidence interval will be sorted as outliers, they are stored in the NIR-Explorer.




This calculations can be made for the **C-Set** and/or **V-Set** as selected by the user.

The confidence interval e.g. for 99 % level (mean value +/- c * stdev) can be determined according the number of C-set spectra: **Calculate c**.

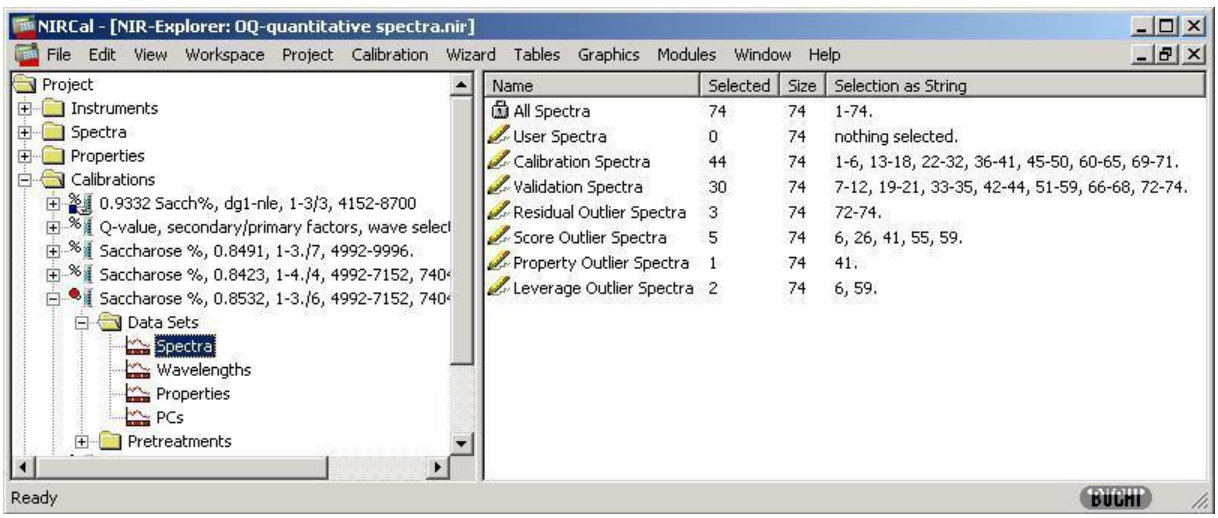
Press **Recalculate** beside the 4 different types of outliers to start the calculations. The found outlier type spectra are marked red after pressing Recalculate."




By pressing **Plot**, the graphic is opened with scatter plot for easier spectra selection.

Outliers are marked with a sign and colored as shown in the plot legend.

Outlier spectra are listed in the NIR-Explorer under the spectra selection of the active calibration:




Spectrum No. 72-74 are identified as residual outliers with this module (residuals: 0.0067-0.0075).

Spectrum No. 16 in the calibration set has the max. residual value (0.00449) and defines the limit for the calibration and application:

**Max. allowed residual = Max. C-Set residual * 2**

2: default Residual Blow up, can be adjusted by the user.
In this example, these spectra are not real residual outliers for NIRCal. In case these spectra would be real outliers, they can be selected as C-set spectra to entlarge the limits or eliminated from the calibration (U-Set = unused set).

There are two different types of outliers:

1. False measurement:
   If for example air bubbles in liquid are measured or particles pollute the sample just in front of the optics of a probe, an outlier spectrum is measured. As three spectra are collected from every sample, it is easy to recognize these spectra. The spectra of one and the same sample should be similar. Nevertheless variations due to the production process will also manifest in the NIR spectra. Spectra originating from false measurements should be deleted or eliminated from the selection (unused).

2. Samples "out of specification":
   If all spectra of one sample are different to spectra from other batches of the same material, the sample itself may be considered as an outlier. The difference can be caused by several parameters, for example changes in the production process, the type and/or the composition of single substances in a product were changed and/or varied, or the sample is polluted. The user has to verify what caused the difference and if the particular sample can still be used for its purpose.

Before outliers are deleted, a careful clarification of the reason should be made for the appearance of the outlier.

For **quantitative calibration** to find out if the reference value or the measured spectra must be regarded as an outlier, the score plots should be reviewed (Graphics / Scores / 2D-Scatter). Spectra breaking ranks, show clearly deviating scores and residuals (Graphics / Spectra / Residuals). Is that not the case, the reference value can be considered as false.

If there are big differences between the reference values and the predicted values, but the scores do not have particular deviations, with high probability, the outliers appear because of false reference values.

**Groups of samples with systematically deviations**

This effect can be seen from time to time when samples are evaluated their reference values have been determined in laboratories not using exactly the same reference methods. Here only an alignment of the reference methods can help.

**Significantly different results depending on the chosen classification of the samples into the C- and V-Set**

The number of used samples is too small, for instance because of not considered, hidden properties. Remedy: selective completion of the master data set that all possible variations flow into the calculation.

## 3.13 Q-Value

Menu: Calibration / Q-Value Protocol

Icon: 

In NIRCal calibrations can be rated via the Q-Value. The Q-Value sums up all important criteria.

The Q-Value is part of the **calibration protocol**, there is no separate calculation necessary

After several calibration optimization and running the automatic calibration module (Step 2: Calibrate), there are several calibrations in the project. Sort the calibrations with a click on the Name or Q-Value.

The calibration are sorted with increasing or decreasing Q-Value.

***NOTE***
Keep only the best 2-4 calibrations and delete all others.

Further details can be found in the Calibration Protocol or in the description of the matrix Q-Value.

## 3.14 Change X-Unit Wavelength / Wavenumber

Menu: Project / Spectra / Change X-Unit to

Icon:  Changes the x-axis units for the display of spectra to nm.

 Changes the x-axis units for the display of spectra to 1/cm.

## 3.15 Convert and Import spectra from other instruments to DB

It is possible to **import** spectra measured with:

NIRVis and NIRFlex N-400
NIRLab N-200
other instruments (in case the wavelength range is similar)

into the NIRWare Database. In order to use those spectra with the **NIRFlex N-500** the s**pectral data need to be converted** prior to saving them into the Database to reach the **compatibility**. The spectra should be **stored into the database**, the project should be closed, opened again and the **calibration** should be created.

These steps are explained in detail below.
Typically following conversions can be made with Büchi instruments:

| | **NIRVIS / NIRFlex N-400** | **NIRLab N-200** | **Convert** | **NIRFlex N-500** |
|---|---|---|---|---|
| | Reflectance | Absorbance | --> | Reflectance |
| [1/cm] | 4'008 - 9996 | 3'999,67 - 10'001,1 | --> | 4'000 - 10'000 |
| Resolution | 25 | 8 | | 8* |
| Step | 12 | 3,85696 | | 4 |
| Datapoints | 500 | 1'557 | --> | 1'501 |
| Remarks | Values at 4000, 4004 are repeated from 4008 and 10'000 is repeated from 9996 | Transformed by ilg (inverse logarithm) Pretreatment. | | |

* can be 16, but this is not suggested.

### 3.15.1 Procedure of Spectra Import

NIRCal 5 can work as **File- and as Database - oriented software**.

In case the **File - oriented** mode is used, the projects can be opened and stored as **".nir" Files** (as NIRCal V 4.21), all function of editing and deleting of names, properties and values is allowed, also the use of "Find and Replace" module (like in NIRCal 4.21).

***NOTE***
Calibrations stored in file format cannot be used together with NIRFlex N-500.

This mode should be used to **copy** spectra together, which are stored in different projects before conversion.
Use "File/ Import/ Project" or "File/ Import/ Spectra".

***NOTE***
**NIRLab spectra** can cause problem by importing with the error: "Spectra are not compatible with Project Spectra because: not the same resolution in the Data Vector".

In this case **export** the spectra as **JCAM.DX** file, use "**Fixed Format**".

Import the spectra in an existing project as JCAM.DX file.

**BCAP Series** can be also imported using "File/ Import/ Project" and using "All Files(*.*)" as file type. The Spectra files **" .Sxx" should be selected** (not the Head files ".Hxx").

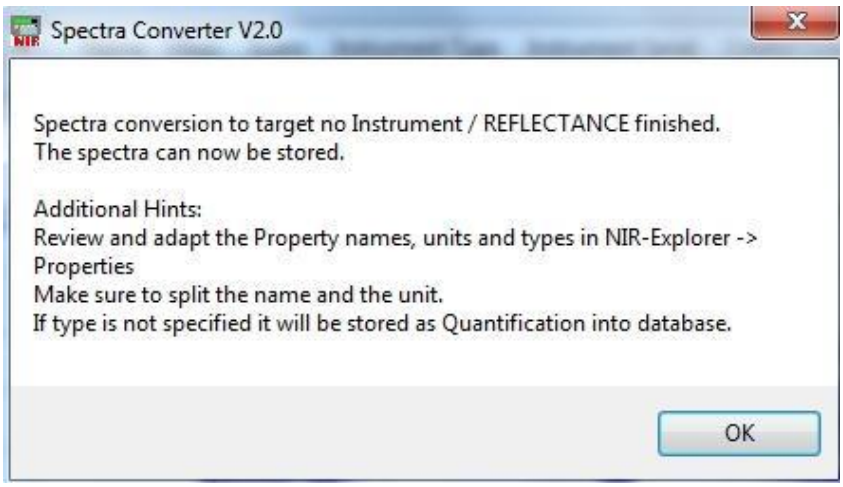
In **Database - oriented** mode the projects are **stored in the Database**, the function of editing in NIRCal 5 is strictly restricted. Deleting is allowed according the adjusted **Lifecycle** template.

### 3.15.2 Procedure of Spectra Conversion:

1 Open a new empty NIRCal project.
2 Import one NIRFlex N-500 template spectrum into this empty project.

**NOTE**
There are template spectra stored on the NIRSolution CD: Operation Manual / NIRFlex N-500 Template Spectra.
Choose a spectrum according the **the y-Unit** and **wavenumber range**, intended to use later with NIRFlex N-500.

3 Start the Module / **Spectra Converter**
4 Define the folder, where the source spectra are stored. Spectra which have *.nir, *.nsf, *.jdx, *.dx, *.jcm, *.spc and*.s?? file format, can be imported and transformed.
5 The spectra will be imported and converted into the project and the first template spectrum is automatically deleted from the project. At the end the following information is shown:



After conversion the spectra are available.

***NOTE***
E.g. each spectrum will have 1501 data points between 10.000 and 4.000 cm-1. In case some data are missing in the original spectra, the last measured value will be repeated, eg. by spectra measured between 10.000 and 4.500 cm-1, the value at 4.500 cm-1 will be repeated till 4.000 cm-1. To keep the spectrum as a constant graph (better for derivatives).

**Do not use the "Spectra Converter" module twice in a project!**

6 Check the name of the property in the project and make sure, that it has always the same spelling, as it is already used in the database. Consider capitalisation!

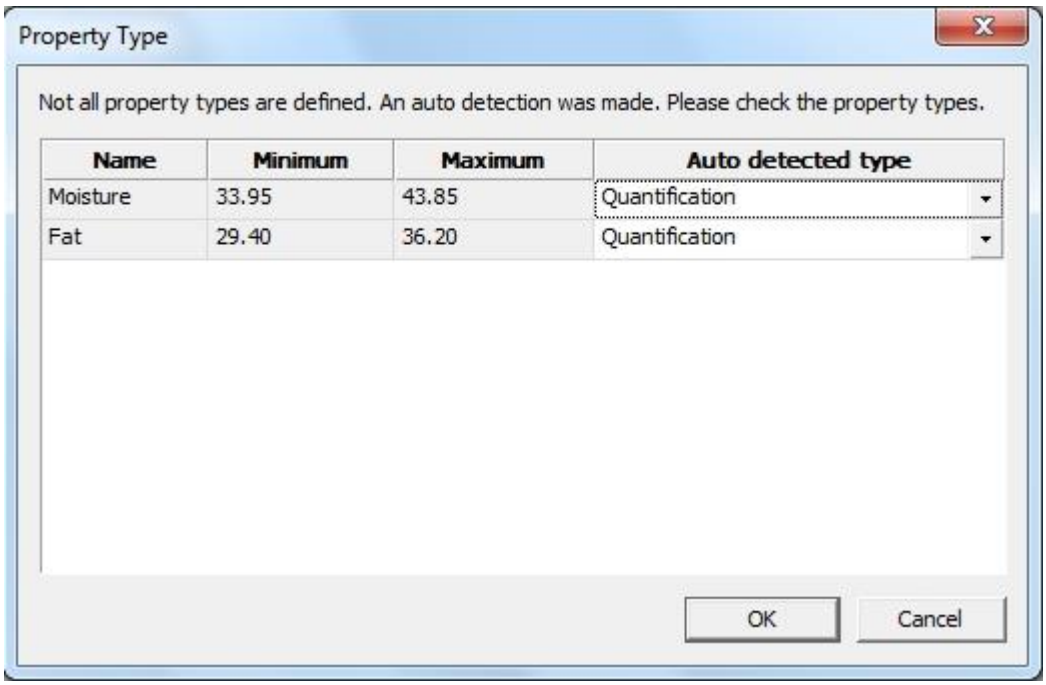
Adjust the property **unit**.
**Unit**: it is necessary to separate the unit from the property name (quantitative).

Deleting not necessary property and spectra from the project is still possible. Check the property values (Table Properties Original).

7 Save the spectra/project into the database with "File / Database / Save Project As..".

Attention: in case the project has huge number of spectra, this step can be time consuming.

(7a) A dialog appears with the property types (Quantification/Identification) selected automatically. Review the types before saving and make changes via the drop-down menu if necessary.




**The project will be closed automatically.**

(7b) Edit the property name, - value and - unit in NIRWare **Sample Manager** in case it was not made in NIRCal 5 before saving the project into the DB.

8 **Reopen the NIRCal-project**.

9 Create the **calibration**.

In case a NIRCal 4.21 calibration is available, open the project with NIRCal 4.21 and copy the data selection (C-V-Set spectra, calibration wavelength) into the new NIRCal 5 calibration.

***NOTE***
**Calibration wavelength**: deselect wavelength range, which were not measured with the original spectra. By several instrument spectra in the same project use the smallest measured range.

**NIRLab** spectra have not integer wavenumber; the wave selection is after conversion rounded to integer.

In case the calibration was calculated with a special wave selection, it is suggested to make several trials to find the best wavelength selection.

Apply the **pretreatments** as used in the existing calibration.

***NOTE***
**Pretreatment**: use "Absorbance" (log) as additional **first** data pretreatment by NIRLab N-200 spectra, in case not "Absorbance inverse" (ilg) was used as first data pretreatment. This can be used for each calculation with the calibration wizard activating "Keep preselected Pretreatments for all calibrations".

**NIRFlex N-400** spectra have **3 times more data points** after the conversion. Using the same data pretreatment give other result. In case e.g. "1st Derivative BCAP" was used in NIRCal 4, after the spectra conversion "1st Derivative BCAP Gap2" will nearly give the same result in NIRCal 5. The same with other derivatives and smoothing.

10 Perform the first calculation (Overview) and adjust the number of primary and secondary PCs according the existing calibration.

***NOTE***
Compare the results of the calibrations. The Q-Value should be different for quantitative calibration, while the calculation equation is improved that leads to slightly different Q-Values.

11 Increase the robustness of the calibration with spectra recently measured with NIRFlex N-500.

12 Optimize the calibration and save the best according the **Lifecycle**.

13 Close and save the project. The created calibration is now available in the **NIRWare Application Designer**.

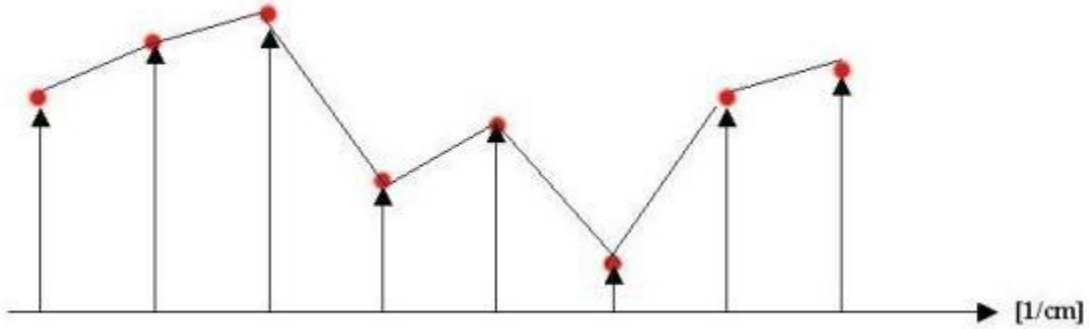
## 3.15.3 Explanation of the Spectra Converter

Cubic spline function is used for the spectra conversion. The cubic function will be fitted on the data stepwise with 3rd degree polynom:

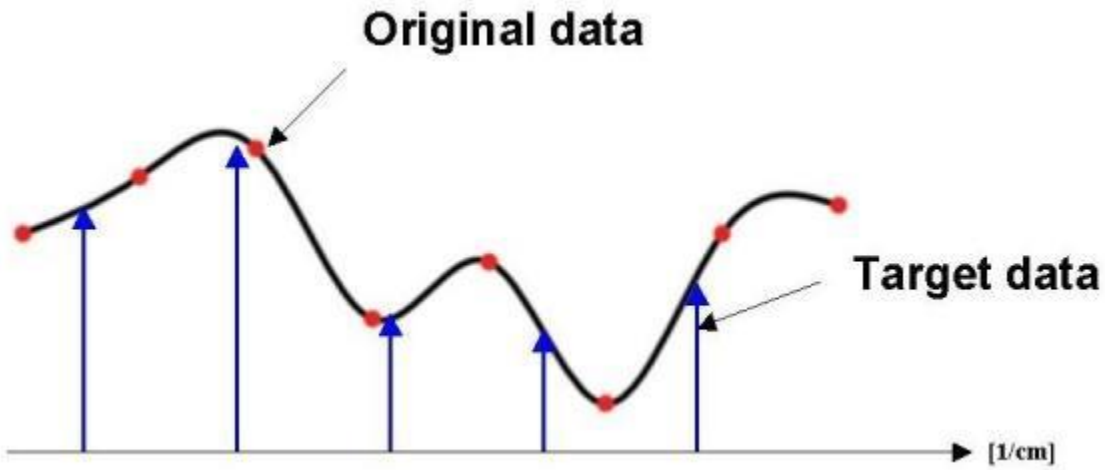
$y(x) = a * x^3 + b * x^2 + c * x + d$

The Y-value for the desired X-values (wavenumber) are extracted from the resulting function curve.

The result is: not a linear connection of the points




but a spline function.




## 3.16 Calibration Handling

In a NIRCal project several calibrations can be handled. Their status can be different according the project is stored as **file** or into the **database**. The calibrations, stored in **file format** are always **editable**.
The **calibrations which are stored in the database**, underlying the **Lifecycle** and can have "**Created**" with "editing" or "**Approved**" status with "**read-only**" mode.
The first calibration in the NIRCal project is always the master calibration.

***NOTE***
The typical features of the **master calibration**:

- This calibration **organises the spectra** in the project: **all spectra should be imported activating this calibration; otherwise they will not be stored in the project!**
- Can have only **"Created"** LC mode; so it is always editable. It is not possible to put to "Approved" LC state and it can not be deleted with "Lifecycle / Delete";
- Can only be deleted if the project is deleted with "File/ Database/ Delete Project" (new feature in NIRCal 5.4).
- It is not suggested to use this calibration for an application: for tracability of the results use only approved calibrations for applications.

Approved calibrations, which are already used for application, are different from the approved calibrations, which are never implemented in any application. Approved calibrations, already used in application for measurement can not be deleted.

### 3.16.1 Calibrations without Lifecycle

Each new project in NIRCal 5 has an "**unnamed" or master calibration**, which is created automatically by importing spectra into the empty project.
This calibration and every calibration created with the menu "**Calibration / New**" has an "**all**" **selection** for the C-Set spectra, wavelength and property selection automatically and has "editable" status.
Running the calibration wizard, several new calibrations will be generated, the calibration with the highest Q-Value is active. The calibrations created by the wizard are **outside the normal Lifecycle, unless they are saved in the database!**

Calibrations, which are created but **not yet stored in the database** are editable and can be deleted by selecting them in the NIR-Explorer on the right side and pressing "Delete".

**Rename the Calibration**

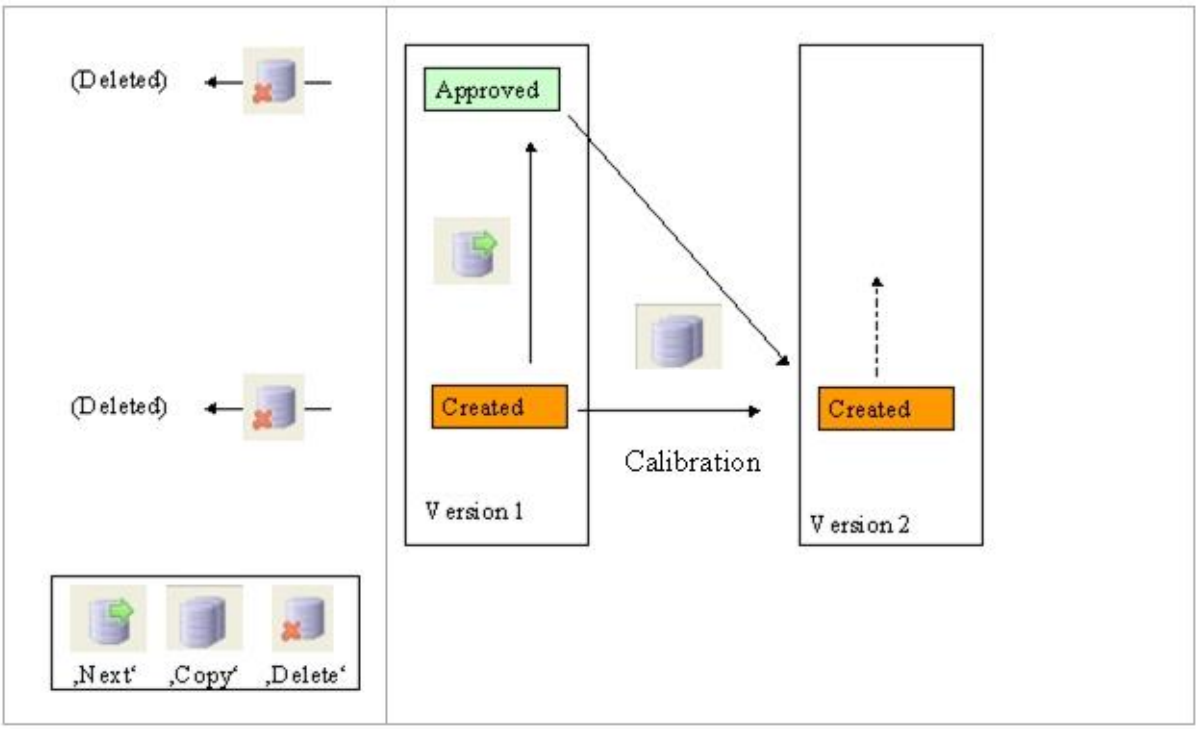
The name of an **active, editable calibration** can be changed by clicking the right mouse button and selecting: "**Rename**". **It is suggested to give a short, but clear name for each calibration; data as extention is useful for approved calibrations.**

***NOTE***
The calibration name is part of the Calibration Protocol which is created during calculation. If the calibration name is changed afterwards, make sure **to do a full calculation to get an updated calibration protocol**.

***NOTE***
Give an appropriate name to the calibration, this allows a correct connection for the application.
Rename is not available for Approved calibrations.

### 3.16.2 Calibrations with Lifecycle

After saving the project into the **database**, the calibrations are also stored and from that moment underlying the **Lifecycle**.
Calibration status in all Lifecycle Template "Unregulated, ER, ERES":

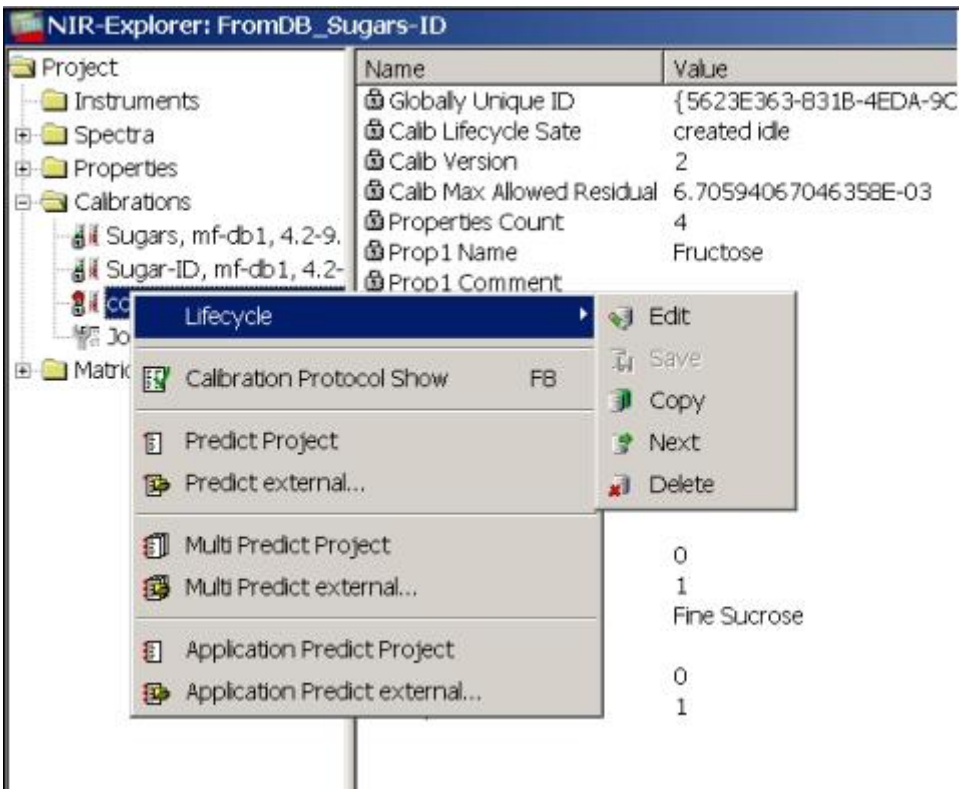



Allowed Lifecycle functions are visible, the grey fields are not available.

Note: Audit trail will log all important transitions.

**Lifecycle: Edit**

A calibration stored in the DB has the "**created**" status and " **read-only"** mode. The active calibration can be edited with the **Lifecycle / Edit**.



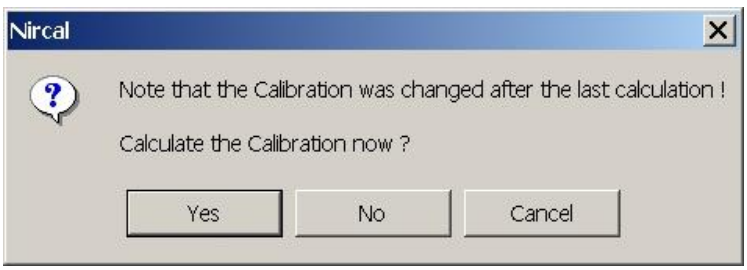

**There is no possibility to edit an approved calibration.**

***NOTE***

If the **first calibration in the project is approved**, this calibration **can never be extended** with new spectra. Work around: use the spectra GUIDs to create the same project, the calibration protocol can be a valuable help.

**Lifecycle: Save**

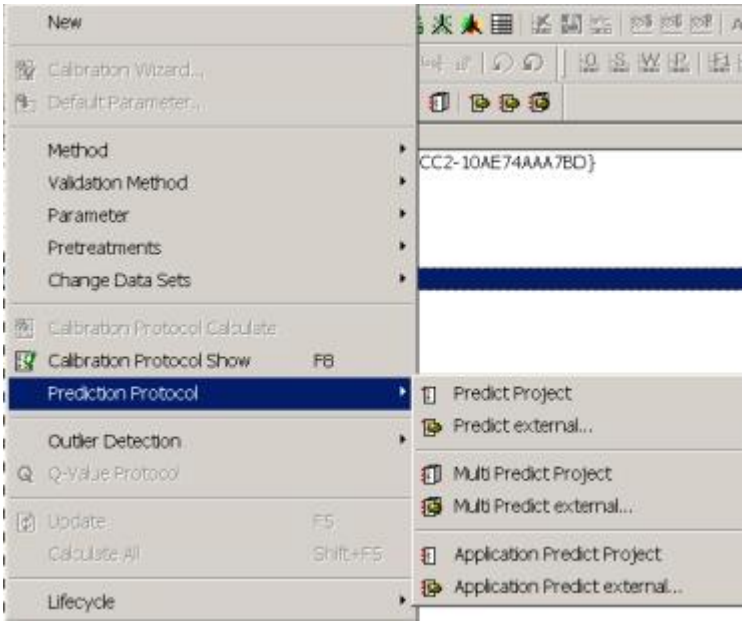
Calibrations after optimization can be stored into the database with this LC function. It is not available, in case the project is not yet stored in the database.
NIRCal tests each **quantitative calibration**, if only **one property** has been selected (single property calibration). In case there are several properties, the calibration is not stored, the user gets a warning. There is a question in case after a modification the calculation is not yet made:




Yes: the calculation is necessary for "Approved" status.
No: gives the possibility to save the calibration into the database without recalculation (quick save at the end of the day).

***NOTE***

Prediction Protocols are only available for calibrations, which are calculated and not modified after the calculation.




**Lifecycle: Copy**

A copy of the active calibration can be created with this LC function. There is a request for calculation in case after a modification the calculation is not yet made.

The copy is automatically **saved** in the DB and set to **active** with the status: **Created.**
The copy has the **same GUID**, as the original calibration, but a **higher version number.**
For the copy of a calculated calibration the Prediction Protocols are available.
Each calibration can be copied.
The copied calibration can be edited, modified and used later -after the necessary steps of saving- for the application.

**Lifecycle: Next**

To put a created calibration to the **approved** status, this LC function is used. If the calibration is not yet calculated, there is a request for calculation.

***NOTE***
It is not possible to put the **first master calibration** to the next status.
After the necessary calculation the calibration is stored and it`s quality is tested. The Q-Value default limit is:

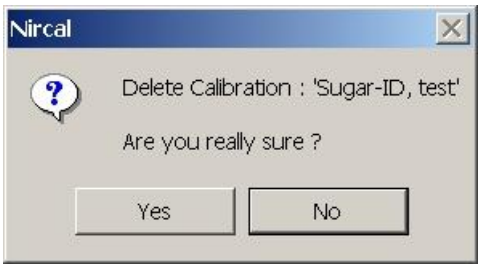
- 0.8 for qualitative and
- 0.6 for quantitative calibrations.

It is possible to put a calibration to a higher status with lower Q-value.

***NOTE***
Do not use a calibration with outliers!

**Lifecycle: Delete**

Calibrations, which are created but **not yet stored in the DB**, can be deleted by selecting them in the NIR-Explorer and pressing "Delete".
Calibrations, which are **already stored in the DB but not used for application**, can be deleted only with: **Lifecycle / Delete.**
The question about deleting should be answered before:




***NOTE***
It is not possible to delete the first master calibration.
Calibrations, which are **embedded in an application**, can never be deleted.


***NOTE***
Calibrations, which are **already used for sample measurement** in the application, can be never deleted.

**Optimize the Calibration**

In order to optimize a calibration manually, the first calibration can be edited or copied: **Lifecycle / Edit or Copy**.
Running the calibration wizard, several new calibrations are available, the calibration with the highest Q-Value is activated.
A new empty calibration can be created: menu **Calibration / New**.

**Save the Calibration for Application**

The **adjustment of the best calibration** -mainly wavelength, pretreatment and PCs selection- **is suggested to be copied into the first calibration** in order to keep these adjustments for further changes. Unnecessary calibrations should be deleted before saving the project into the Database.

The best, activated and validated calibration should be stored for the application, for this the **Lifecycle / Next** should be used.
The calibration will get the **Approved** state using **Lifecycle / Next**.

***NOTE***
Only calibrations with Lifecycle state "**Approved**" are available for use in the routine **applications for the operator user group**.
The Designer and Administrator user groups can use created or checked calibrations for test.

**Validate the Calibration**

In the calibration process released samples with known properties should be measured together with the calibration samples. These samples should be placed in the V-Set of the calibration: **internal validation**.
When a calibration is created and released for the application, the application should be also validated: **external validation**. This is made with several **new, known samples**, which are not yet in the calibration: **operation qualification - OQ**. These samples should be measured also with the conventional laboratory method. In case the results of NIR and lab are the same (identification, qualification) or they are within the allowed tolerances (quantitative determination), the calibration can be used **for the routine measurement** till the next **performance qualification - PQ**.

**Extend the Calibration**

In case during the external validation there are **unacceptable differences -property value or residual-**, the **calibration should be extended** with the spectra of the external validation samples. These spectra are stored in the DB. The necessary steps are:

- **editing the property values** in the **Manager Console / Sample Manager** (see NIRWare manual);
- **import the spectra** into the original NIR-project;
- **extend the calibration**.

***NOTE***
**Spectra import** is suggested to make in the **activated first calibration.** Approved calibrations, which were already used in the application, can not be edited and modified anymore.

For creating an **extended calibration**, the new **imported spectra** should be selected as **C-Set** spectra, in this way the missing information is fully integrated in the new model. A copy of the optimized master calibration should be created with: **Lifecycle / Copy**. This copy should be stored in the database **Lifecycle / Next** and used for the **application**.

### 3.16.3 Project Handling

A project with existing calibrations can be stored under an other name in the Database.

***NOTE***
**These calibrations are in this case only linked, but not copied!**
Changes in a calibration causes changes in the calibration of both "projects" (after deleting a calibration in a project it will be lost in both projects).

In case with the same spectra e.g. for **different quantitative properties, several calibrations have to be created**, the spectra should be loaded into new projects. In each project calibrations for just one property are evaluated and optimized.

## 3.17 Protocols

### 3.17.1 Calibration Protocol Qualitative

Menu: Calibration / Calibration Protocol Calculate or Show

Icon:  Short Key: F8

The calibration protocol is an important validation report giving all information about the data selection and result of the chemometrics calculation. The calibration protocols are stored in the DB (or .nir project file).
Explanation of the content of the calibration protocols:

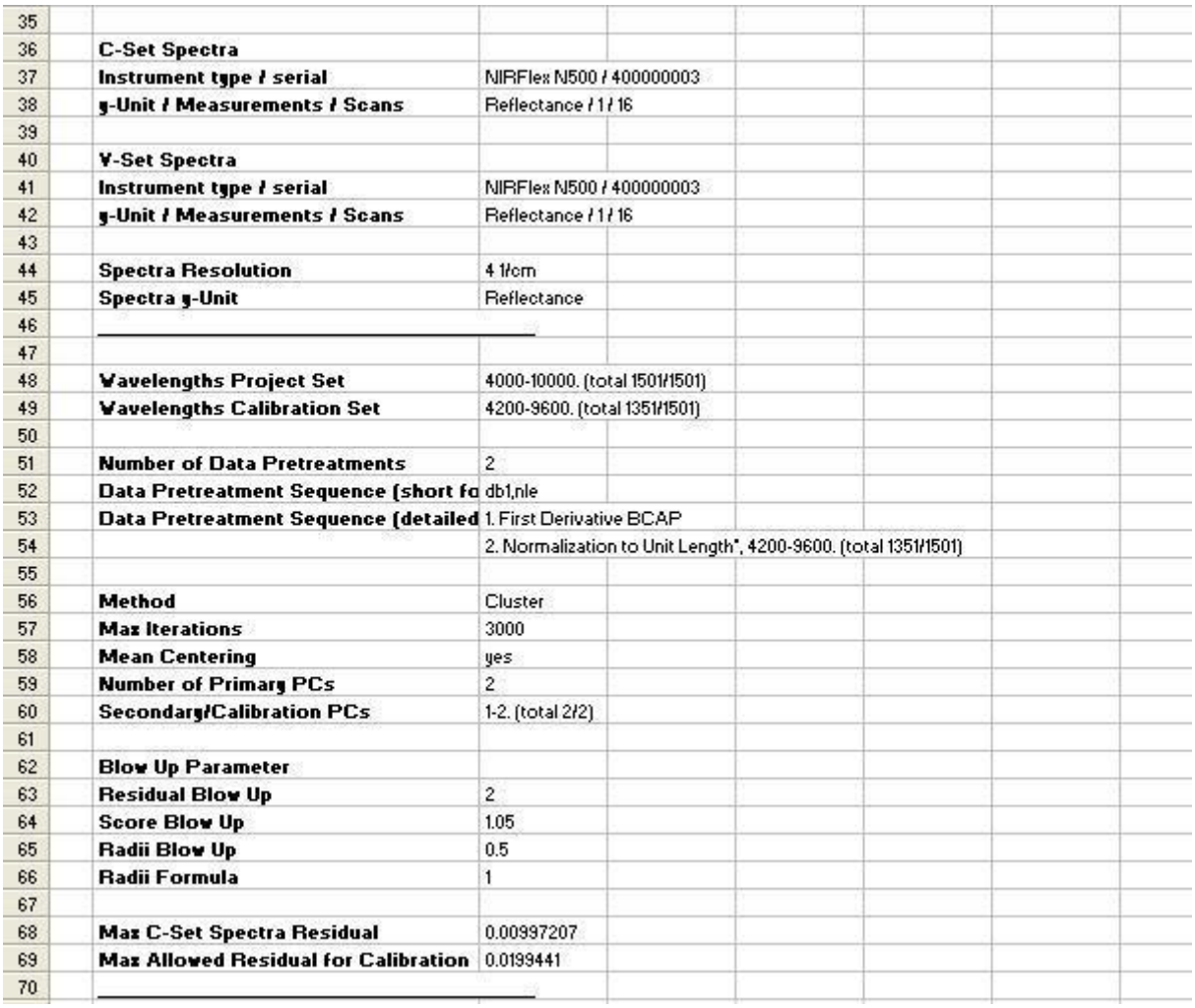
**Cluster Calibration**

The first section has the information about the user specific data selection (row 1-64).

| 1 | Calibration Protocol | |
|---|---|---|
| 4 | | |
| 5 | | |
| 6 | **User** | Customer System Maintenance |
| 7 | **Date/Time** | 27.01.2006 14:17:13 |
| 8 | **Software** | NIRCal V5.1 (Build 600) |
| 9 | **Project File Name** | |
| 10 | **Project Comment** | |
| 11 | **Project GUID** | {DBADF1B8-3FA0-4970-BA00-2B49E741CB70} |
| 12 | **Calibration Name** | ID-Amino-V3 |
| 13 | **Calibration Comment** | |
| 14 | **Calibration GUID** | {482178F5-F73E-4AA7-A9D1-A86BA0F97C30} |
| 15 | **Calibration Version** | 3 |
| 16 | **Calibration Lifecycle Sate** | created editing |
| 17 | | |
| 18 | | |
| 19 | **Properties in Project** | Ornitin-aspartat, L-Alanin, Propafenon-HCl, Cellulose, PEG, Stearic acid, |
| 20 | | Myristyl alkohol, Kollidon, Ketone Resin, Mg-Stearate, Verapamil-HCl. (total 11/11) |
| 21 | **Properties in Calibration Set** | Ornitin-aspartat, L-Alanin, Propafenon-HCl. (total 3/11) |
| 22 | | |
| 23 | | |
| 24 | **Spectra in Project** | 25 |
| 25 | **Spectra in Calibration Set** | 18 |
| 26 | **Spectra in Validation Set** | 7 |
| 27 | | |
| 28 | **Spectra in Calibration Set** | 1-2, 4-5, 7-8, 10-11, 13-14, 16-19, 21-23, 25. (total 18/25) |
| 29 | | |
| 30 | **Spectra in Validation Set** | 3, 6, 9, 12, 15, 20, 24. (total 7/25) |
| 31 | | |
| 32 | **Spectra unused (U-Set)** | nothing selected. (total 0/25) |
| 33 | | |
| 34 | **Validation Method** | Validation Set |

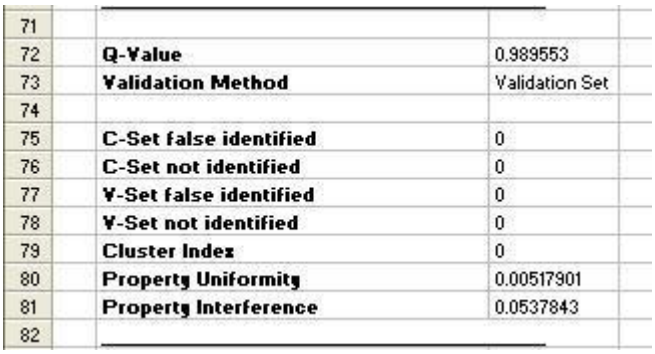
| Row | Name | Description |
|---|---|---|
| 1 | Calibration Protocol | Header |
| 6 | User | |
| 7 | Date/Time | Date and time of the calculation |
| 8 | Software | Current software version |
| 9 | Project File Name | Empty for database, name and place of the project for file |
| 10 | Project Comment | Normally empty. The comment section of the project. |
| 11 | Project GUID | **G**lobal **U**niq **ID**entity of the project |
| 12 | Calibration Name | Name of the calibration |
| 13 | Calibration Comment | Normally empty |
| 14 | Calibration GUID | **G**lobal **U**nique **ID**entity of the calibration |
| 15 | Calibration Version | Indicates the modification of the calibration. The Calibration Lifecycle Version is incremented by Lifecycle copy (or by XML import to avoid duplicates). |
| 16 | Calibration Lifecycle State | not available anymore |
| 19 | Properties in Project | Name and number of all properties in the project |
| 21 | Properties in Calibration Set | Name and number of calibrated properties |
| 24 | Spectra in Project | Total number of spectra in the project |
| 25 | Spectra in Calibration Set | Total number of spectra selected in the C-Set |
| 26 | Spectra in Validation Set | Total number of spectra selected in the V-Set |
| 28 | Spectra in Calibration Set | Spectra selected in the C-Set |
| 30 | Spectra in Validation Set | Spectra selected in the V-Se |

| Row | Name | Description |
|---|---|---|
| 32 | Spectra unused (U-Set) | Spectra not used for calibration |
| 34 | Validation Method | Selected Validation method: Validation Set or Cross Validation (quantitative) |



| | | |
|---|---|---|
| 36 | C-Set Spectra | Instrument statistic |
| 37 | Instrument type / serial | Used instrument type and it`s serial number |
| 38 | y-Unit / Measurements / Scans | Measuring principle / Number of repeated measurements and scans of a spectrum |
| 40 | V-Set Spectra | Instrument statistic |
| 41 | Instrument type / serial | Used instrument type and it`s serial number |
| 42 | y-Unit / Measurements / Scans | Measuring principle / Number of repeated measurements and scans of a spectrum |
| 44 | Spectra Resolution | Used data resolution |
| 45 | Spectra y-Unit | Measuring principle |
| 48 | Wavelengths Project Set | Measured wavelength range |
| 49 | Wavelengths Calibration Set | Selected wavelength range (selected datapoints / all data points) |
| 51 | Number of Data Pretreatments | ~~Number of applied data pretreatment~~ |
| 52 | Data Pretreatment Sequence (short) | ~~Short name of the applied data pretreatment~~ |
| 53 | Data Pretreatment Sequence (detailed) | ~~Full name of the applied data pretreatment~~ |

| Row | Name | Description |
|---|---|---|
| 56 | Method | Selected calculation method / algorithm |
| 57 | Max Iterations | Calculation steps before default stop (max. 3000) |
| 58 | Mean Centering | Mean centering after pretreatment is activated: yes / no |
| 59 | Number of Primary PCs | Number of selected primary principal components |
| 60 | Secondary/Calibration PCs | Numbers of selected secondary principal components |
| 62 | Blow Up Parameter | Title |
| 63 | Residual Blow Up | Factor for residual limit calculation (default: 2) |
| 64 | Score Blow Up | Factor for score limit calculation (default: 1.05) |
| 65 | Radii Blow Up | Factor for radii limit calculation (default: 1) |
| 66 | Radii Formula | Formula for radii calculation: 1 / 2 (SIMCA: 3) |
| 68 | Max C-Set Spectra Residual | Highest value of the C-Set residuals |
| 69 | Max Allowed Residual for Calibration | Residual limit of the calibration for the application (max.* 2) |



| | | |
|---|---|---|
| 71 | | |
| 72 | **Q-Value** | 0.989553 |
| 73 | **Validation Method** | Validation Set |
| 74 | | |
| 75 | **C-Set false identified** | 0 |
| 76 | **C-Set not identified** | 0 |
| 77 | **V-Set false identified** | 0 |
| 78 | **V-Set not identified** | 0 |
| 79 | **Cluster Index** | 0 |
| 80 | **Property Uniformity** | 0.00517901 |
| 81 | **Property Interference** | 0.0537843 |
| 82 | | |

| | | |
|---|---|---|
| 72 | Q-Value | Q-Value of the calibration |
| 73 | Validation Method | Selected Validation method: only Validation Set |
| 75 | C-Set false identified | Should be zero (indicates wrong user settings) |
| 76 | C-Set not identified | Should be zero (indicates wrong user settings) |
| 77 | V-Set false identified | Should be zero (indicates wrong sample or calibration) |
| 78 | V-Set not identified | Should be zero (residual or distance problem) |
| 79 | Cluster Index | (No. Cluster- No. Property): should be zero |
| 80 | Property Uniformity | Should be small: the spectra spreading in the clusters is uniform |
| 81 | Property Interference | Should be small: the clusters are independent from each other |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 83 | | | | | | | |
| 84 | **Property Overview** | Num Cluster | C num Spec | V num Spec | U num Spec | | |
| 85 | **Total Sum** | 3 | 18 | 7 | 0 | | |
| 86 | **Ornitin-aspartat** | 1 | 6 | 3 | 0 | | |
| 87 | **L-Alanin** | 1 | 6 | 2 | 0 | | |
| 88 | **Propafenon-HCl** | 1 | 6 | 2 | 0 | | |
| 89 | | | | | | | |
| 90 | **Property Separation** | Q- | Q+ | Interference | Extension | Distance | Nearest Prop |
| 91 | **Total Sum** | | | 0.161353 | 0.0873482 | 1.62702 | |
| 92 | **Ornitin-aspartat** | Dmin | | 0.0547229 | 0.0295032 | 0.539138 | L-Alanin |
| 93 | **L-Alanin** | Imax Emax Dmin | | 0.0694985 | 0.0374692 | 0.539138 | Ornitin-aspartat |
| 94 | **Propafenon-HCl** | | Imin Emin Dma | 0.0371314 | 0.0203758 | 0.548747 | Ornitin-aspartat |
| 95 | | | | | | | |
| 96 | **Property Outlier** | C out Clu | V out Clu | C false Clu | V false Clu | C Resid big | V Resid big |
| 97 | **Total Sum** | 0 | 0 | 0 | 0 | 0 | 0 |
| 98 | **Ornitin-aspartat** | 0 | 0 | 0 | 0 | 0 | 0 |
| 99 | **L-Alanin** | 0 | 0 | 0 | 0 | 0 | 0 |
| 100 | **Propafenon-HCl** | 0 | 0 | 0 | 0 | 0 | 0 |
| 101 | | | | | | | |
| 102 | **Property Outlier U-Set** | | U out Clu | | U false Clu | | U Resid big |
| 103 | **Total Sum** | | 0 | | 0 | | 0 |
| 104 | **Ornitin-aspartat** | | 0 | | 0 | | 0 |
| 105 | **L-Alanin** | | 0 | | 0 | | 0 |
| 106 | **Propafenon-HCl** | | 0 | | 0 | | 0 |
| 107 | | | | | | | |

| Row | Name | Description |
|---|---|---|
| 84 | Property Overview | Number of clusters and selected spectra |
| 85 | Total Sum | Num Cluster = Total number of clusters, should be equal to the number of calibration properties |
| 86 | 1st Property | **Num Cluster = should be 1, the spreading of C-and V-Set spectra about 2/3-1/3** |
| 87 | 2nd Property | **Num Cluster = should be 1, the spreading of C-and V-Set spectra about 2/3-1/3** |
| 88 | 3rd Property | **Num Cluster = should be 1, the spreading of C-and V-Set spectra about 2/3-1/3** |
| 90 | Property Separation | Information about the influence of the properties on the separation * |

* Description of the property separation:

Q- = Imax, Emax and Dmin indicate bad influence on the substance separation Q+ = Imin, Emin and Dmax indicate good influence on the substance separation
Nearest Prop = Name of the property with the smallest distance

Judging criteria:

Distance = Mahalanobis distance to the closest property (spectrum to spectrum of the other cluster; see Property Adjacency)
Extension = The max. Mahalanobis distance of each cluster in the factor space (secondary principal components)
Interference = Extension divided by distance
Imin / Imax = Extreme values regarding the parameter Interference
Emin / Emax = Extreme values regarding the parameter Extension
Dmin / Dmax = Extreme values regarding the parameter Distance

| | | |
|---|---|---|
| 96 | Property Outlier | Number of outlier spectra outside a cluster, in false cluster, residual too big in C- and V-Set |
| 97 | Total Sum | **All should be 0** |
| 98 | 1st Property | **All should be 0** |
| 99 | 2nd Property | **All should be 0** |
| 100 | 3rd Property | **All should be 0** |
| 102 | Property Outlier U-Set | Number of outlier spectra outside a cluster, in false cluster, residual too big in U-Set |
| 103 | Total Sum | Normally the U-Set is empty, the sum is also zero |

| Row | Name | Description |
|---|---|---|
| 104 | 1st Property | Normally all are zero |
| 105 | 2nd Property | Normally all are zero |
| 106 | 3rd Property | Normally all are zero |

***NOTE***
Number of cluster of each property must always be 1.
Number of any outliers in the C-and V-Set must always be 0.

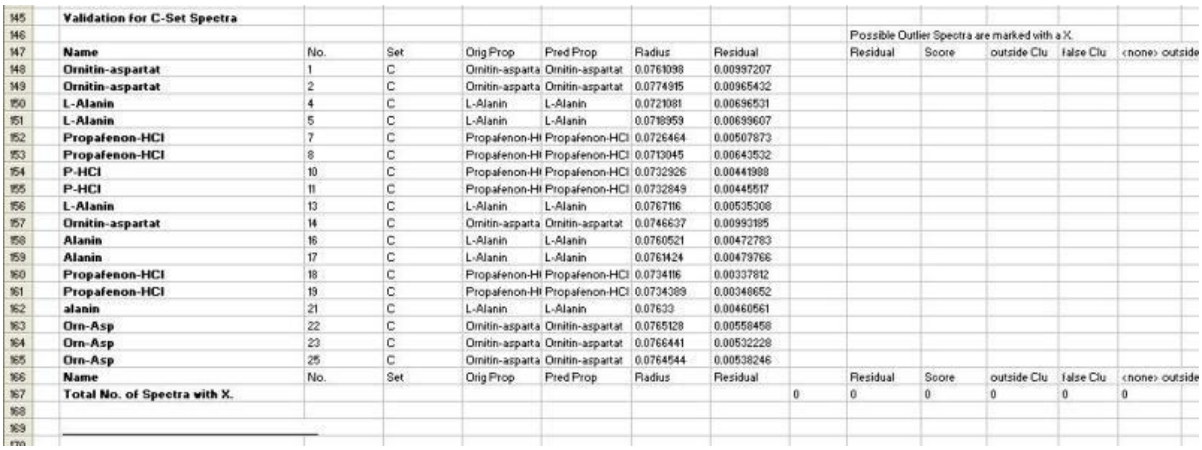
| | | |
|---|---|---|
| 109 | **Legend** | |
| 110 | **Number of Clusters** | Num Cluster |
| 111 | **Num C-Set Spectra** | C num Spec |
| 112 | **Num V-Set Spectra** | V num Spec |
| 113 | **Num U-Set Spectra** | U num Spec |
| 114 | **Distance to nearest foreign C-Set Spectra** | Distance |
| 115 | **Name of nearest Property** | Nearest Prop |
| 116 | **Num C-Set Spectra outside of Cluster** | C out Clu |
| 117 | **Num V-Set Spectra outside of Cluster** | V out Clu |
| 118 | **Num C-Set Spectra in false Cluster** | C false Clu |
| 119 | **Num V-Set Spectra in false Cluster** | V false Clu |
| 120 | **Num C-Set Spectra with Residual too big** | C Resid big |
| 121 | **Num V-Set Spectra with Residual too big** | V Resid big |
| 122 | **Num U-Set Spectra outside of Cluster** | U out Clu |
| 123 | **Num U-Set Spectra in false Cluster** | U false Clu |
| 124 | **Num U-Set Spectra with Residual too big** | U Resid big |
| 125 | | |
| 126 | | |

| | | |
|---|---|---|
| 109-124 | Explanation of the legend | Short form |

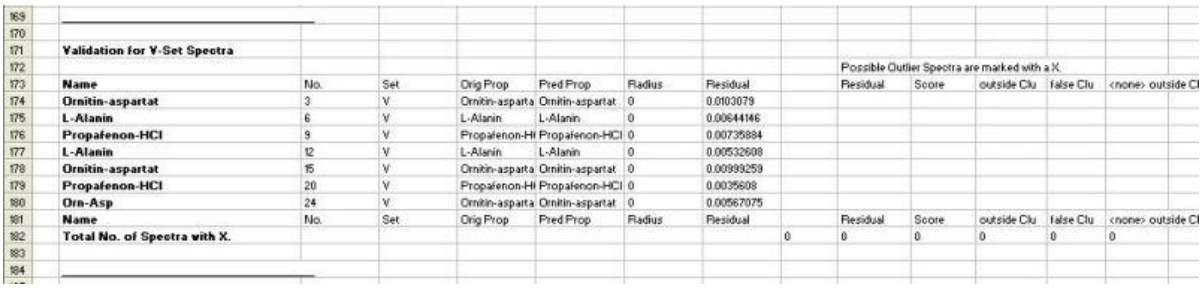
Some statistics:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 126 | | | | | | | |
| 127 | | | | | | | |
| 128 | **C-Set Spectra Residual** | Overview by C-Set Properties | | | | | |
| 129 | | | | | | | |
| 130 | **C-Set Property** | Min Residual | Min at Spec | Max Residual | Max at Spec | Mean | Sdev |
| 131 | **Ornitin-aspartat** | 0.00532228 | 23 | 0.00997207 | 1 | 0.00764126 | 0.00242659 |
| 132 | **L-Alanin** | 0.00460561 | 21 | 0.00699607 | 5 | 0.00557426 | 0.00111922 |
| 133 | **Propafenon-HCl** | 0.00337812 | 18 | 0.00643532 | 8 | 0.00454229 | 0.00112817 |
| 134 | | | | | | | |
| 135 | | | | | | | |
| 136 | **C-Set Spectra Radii** | Overview by C-Set Properties | | | | | |
| 137 | | | | | | | |
| 138 | **C-Set Property** | Min Radius | Min at Spec | Max Radius | Max at Spec | Mean | Sdev |
| 139 | **Ornitin-aspartat** | 0.0746637 | 14 | 0.0774915 | 2 | 0.0763127 | 0.000929506 |
| 140 | **L-Alanin** | 0.0718959 | 5 | 0.0767116 | 13 | 0.0748734 | 0.00223665 |
| 141 | **Propafenon-HCl** | 0.0713045 | 8 | 0.0734389 | 19 | 0.0728965 | 0.000832398 |
| 142 | | | | | | | |
| 143 | | | | | | | |

| | | |
|---|---|---|
| 128-133 | C-Set Spectra Residual | Statistic of residual: min.-max. residual with spectrum number, mean value and standard deviation |
| 136-141 | C-Set Spectra Radii | Statistic of radii: min.-max radii with spectrum number, mean value and standard deviation |



| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 145 | **Validation for C-Set Spectra** | | | | | | | | | | | | |
| 146 | | | | | | | | Possible Outlier Spectra are marked with a X. | | | | | |
| 147 | **Name** | No. | Set | Orig Prop | Pred Prop | Radius | Residual | Residual | Score | outside Clu | false Clu | <none> outside | |
| 148 | **Ornitin-aspartat** | 1 | C | Ornitin-aspartat | Ornitin-aspartat | 0.0761098 | 0.00997207 | | | | | | |
| 149 | **Ornitin-aspartat** | 2 | C | Ornitin-aspartat | Ornitin-aspartat | 0.0774915 | 0.00965432 | | | | | | |
| 150 | **L-Alanin** | 4 | C | L-Alanin | L-Alanin | 0.0721081 | 0.00696531 | | | | | | |
| 151 | **L-Alanin** | 5 | C | L-Alanin | L-Alanin | 0.0718959 | 0.00699607 | | | | | | |
| 152 | **Propafenon-HCl** | 7 | C | Propafenon-Hl | Propafenon-HCl | 0.0726464 | 0.00507873 | | | | | | |
| 153 | **Propafenon-HCl** | 8 | C | Propafenon-Hl | Propafenon-HCl | 0.0713045 | 0.00643532 | | | | | | |
| 154 | **P-HCl** | 10 | C | Propafenon-Hl | Propafenon-HCl | 0.0732926 | 0.00441988 | | | | | | |
| 155 | **P-HCl** | 11 | C | Propafenon-Hl | Propafenon-HCl | 0.0732849 | 0.00445517 | | | | | | |
| 156 | **L-Alanin** | 13 | C | L-Alanin | L-Alanin | 0.0767116 | 0.00535308 | | | | | | |
| 157 | **Ornitin-aspartat** | 14 | C | Ornitin-aspartat | Ornitin-aspartat | 0.0746637 | 0.00993185 | | | | | | |
| 158 | **Alanin** | 16 | C | L-Alanin | L-Alanin | 0.0760521 | 0.00472783 | | | | | | |
| 159 | **Alanin** | 17 | C | L-Alanin | L-Alanin | 0.0761424 | 0.00479766 | | | | | | |
| 160 | **Propafenon-HCl** | 18 | C | Propafenon-Hl | Propafenon-HCl | 0.0734116 | 0.00337812 | | | | | | |
| 161 | **Propafenon-HCl** | 19 | C | Propafenon-Hl | Propafenon-HCl | 0.0734389 | 0.00348652 | | | | | | |
| 162 | **alanin** | 21 | C | L-Alanin | L-Alanin | 0.07633 | 0.00460561 | | | | | | |
| 163 | **Orn-Asp** | 22 | C | Ornitin-aspartat | Ornitin-aspartat | 0.0765128 | 0.00558456 | | | | | | |
| 164 | **Orn-Asp** | 23 | C | Ornitin-aspartat | Ornitin-aspartat | 0.0766441 | 0.00532228 | | | | | | |
| 165 | **Orn-Asp** | 25 | C | Ornitin-aspartat | Ornitin-aspartat | 0.0764544 | 0.00538246 | | | | | | |
| 166 | **Name** | No. | Set | Orig Prop | Pred Prop | Radius | Residual | Residual | Score | outside Clu | false Clu | <none> outside | |
| 167 | **Total No. of Spectra with X.** | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 |
| 168 | | | | | | | | | | | | | |
| 169 | | | | | | | | | | | | | |

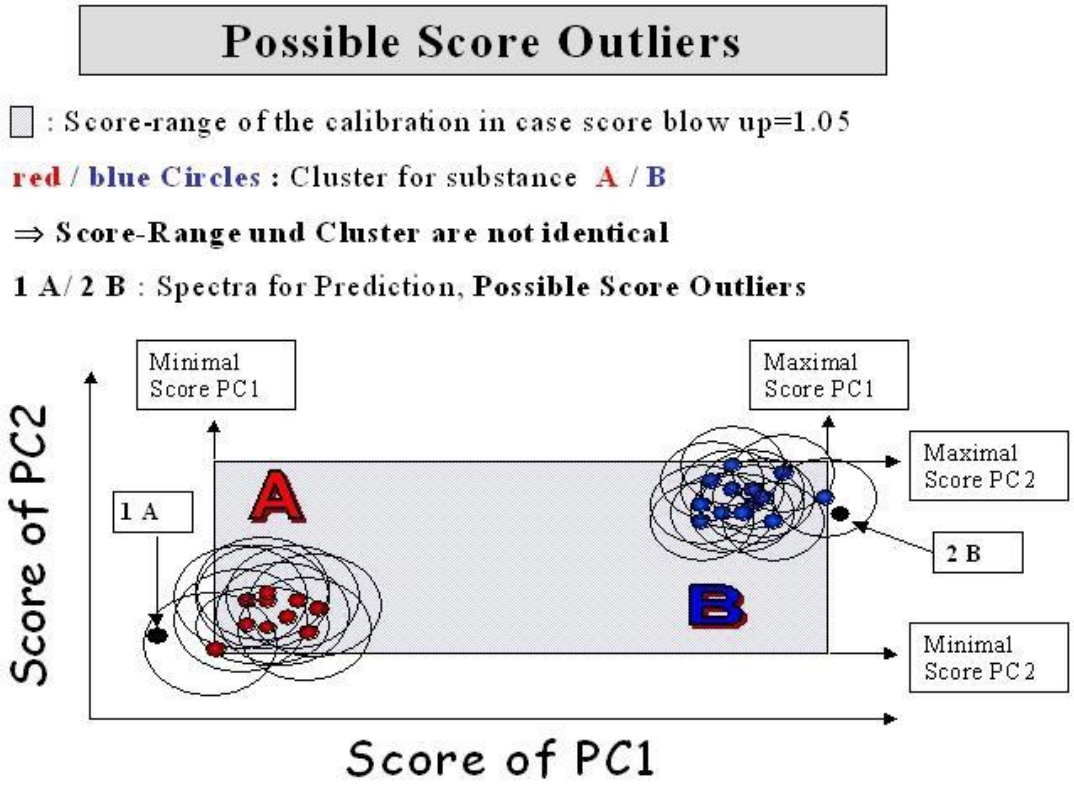


| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 169 | | | | | | | | | | | | | |
| 170 | | | | | | | | | | | | | |
| 171 | **Validation for V-Set Spectra** | | | | | | | | | | | | |
| 172 | | | | | | | | Possible Outlier Spectra are marked with a X. | | | | | |
| 173 | **Name** | No. | Set | Orig Prop | Pred Prop | Radius | Residual | Residual | Score | outside Clu | false Clu | <none> outside Cl | |
| 174 | **Ornitin-aspartat** | 3 | V | Ornitin-aspartat | Ornitin-aspartat | 0 | 0.0103079 | | | | | | |
| 175 | **L-Alanin** | 6 | V | L-Alanin | L-Alanin | 0 | 0.00644146 | | | | | | |
| 176 | **Propafenon-HCl** | 9 | V | Propafenon-Hl | Propafenon-HCl | 0 | 0.00735884 | | | | | | |
| 177 | **L-Alanin** | 12 | V | L-Alanin | L-Alanin | 0 | 0.00532608 | | | | | | |
| 178 | **Ornitin-aspartat** | 15 | V | Ornitin-aspartat | Ornitin-aspartat | 0 | 0.00999259 | | | | | | |
| 179 | **Propafenon-HCl** | 20 | V | Propafenon-Hl | Propafenon-HCl | 0 | 0.0035608 | | | | | | |
| 180 | **Orn-Asp** | 24 | V | Ornitin-aspartat | Ornitin-aspartat | 0 | 0.00567075 | | | | | | |
| 181 | **Name** | No. | Set | Orig Prop | Pred Prop | Radius | Residual | Residual | Score | outside Clu | false Clu | <none> outside Cl | |
| 182 | **Total No. of Spectra with X.** | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 |
| 183 | | | | | | | | | | | | | |
| 184 | | | | | | | | | | | | | |

Information of the C-Set and V-Set spectra:

| Row | Name | Description |
|---|---|---|
| 145-166 | Validation for C-Set Spectra | List of C-Set spectra with predicted and original property, radii, residual. Number of possible outliers should be 0.<br>.The last column shows the GUID of each spectra. |
| 167 | Total No.of Spectra with X | It should be all type zero. |
| 171-181 | Validation for V-Set Spectra | List of V-Set spectra with predicted and original property, radii (0), residual. Number of possible outliers should be 0 except Scores outliers**.<br>The last column shows the GUID of each spectra. |
| 182 | Total No.of Spectra with X | It is mostly all type zero, scores outliers should be carefully controlled. |

** Explanation of **possible score outliers**:
These spectra are identified correctly, but they are possible outliers according to the score. As score tolerance the **min. and max. score** of the C-Set spectra are taken. These extreme values are multiplied by the "**Score Blow Up**", which is equal to 1.05 by default. But the **tolerance rings give a higher limit**.




The spectrum 2B is not a real outlier, but spectrum 1A is dangerous: it is better to select into the C-Set.

### 3.17.2 Calibration Protocol Quantitative

Menu: Calibration / Calibration Protocol Calculate or Show

Icon:  Short Key: F8

The calibration protocol is an important validation report giving all information about the data selection and result of the chemometrics calculation. The calibration protocols are stored in the DB (or .nir project file).
Explanation of the content of the calibration protocols:

#### Quantitative Calibration

The first section has the information about the user specific data selection and similar, as for the Cluster calibration (row 1-58).

| | | |
|---|---|---|
| 59 | **Blow Up Parameter** | |
| 60 | **Residual Blow Up** | 2 |
| 61 | **Score Blow Up** | 1 |
| 62 | | |
| 63 | **Max C-Set Spectra Residual** | 0.00028157 |
| 64 | **Max Allowed Residual for Calibration** | 0.00056314 |
| 65 | _____ | |
| 66 | | |
| 67 | **Q-Value V5** | 0.871706 |
| 68 | **Validation Method** | Validation Set |
| 69 | | |
| 70 | **C-Set Residual too big** | 0 |
| 71 | **V-Set Residual too big** | 0 |
| 72 | | |
| 73 | **Num Properties** | 1 |
| 74 | **Rel. Consistency** | 0.0381583 |
| 75 | **Weighted BIAS** | 0.00185398 |
| 76 | **Validity** | 0.0004727 |
| 77 | **Comparability** | 0.00027496 |
| 78 | **Precision** | 0.0106114 |
| 79 | **Weighted Accuracy** | 0.0557919 |
| 80 | _____ | |

**Result of the calculation**

| Row | Name | Description |
|---|---|---|
| 59 | Blow Up Parameter | Title |
| 60 | Residual Blow Up | Factor for residual limit calculation (default: 2) |
| 61 | Score Blow Up | Factor for score limit calculation (default: 1.05) |
| 63 | Max C-Set Spectra Residual | Highest value of the C-Set residuals |
| 64 | Max Allowed Residual for Calibration | Residual limit of the calibration for the application (max.* 2) |
| | | **Q-Value protocol** |
| 67 | Q-Value V5 | Q-Value of the calibration according version 5. The higher the Q-Value the best the calibration. |
| 68 | Validation Method | Selected Validation method: Validation Set or Cross Validation |
| 70 | C-Set Residual too big* | Number of C-Set spectra with too high residual = rejection of known (theoretically: 0 but always tested!) |
| 71 | V-Set Residual too big* | Number of V-Set spectra with too high residual = rejection of unknown (should be zero) |
| 73 | Num Properties | Number of properties in the calibration (should be always 1 for the application) |
| 74 | Rel. Consistency* | The C- and V-Set should have similar low standard error of prediction (BIAS corrected):<br>Abs(SEC-SEP)/(Abs(SEP)+1). Should be close to 0 |
| 75 | Weighted BIAS* | The absolute V-Set BIAS should be as close to zero as possible:<br>Abs(Vset-BIAS)/Abs(Range) |
| 76 | Validity* | The V-Set regression coefficient should be near to 1:<br>1-VsetRegr. Validity should be close to 0 |
| 77 | Comparability* | C- and V-Set should have similar high regression coefficient of prediction:<br>Abs(CsetRegr-VsetRegr). Comparability should be close to 0 |
| 78 | Precision* | V-Set standard error of prediction (BIAS corrected) should be low:<br>SEP/Abs (Range) |
| 79 | Weighted Accuracy* | V-Set residual sum of squares=RSS should be low (not BIAS corrected):<br>Abs(Sum((Orig-Predicted)^2)/Abs (Range) |

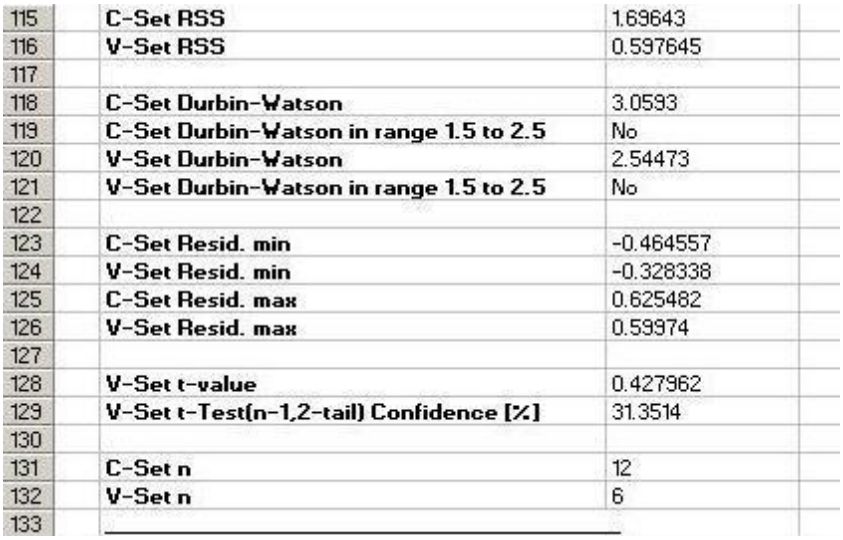
* used with different weighting values.

| | | | | |
|---|---|---|---|---|
| 82 | **Property Statistics** | ethanol | | |
| 83 | | | | |
| 84 | **C-Set BIAS** | 0 | | |
| 85 | **V-Set BIAS** | -0.0593272 | | |
| 86 | **C-Set SEE (SEC)** | 0.39271 | | |
| 87 | **V-Set SEE (SEP)** | 0.339566 | | |
| 88 | **Consistency** | 115.65 | | |
| 89 | | | | |
| 90 | **C-Set Regression Coefficient** | 0.999252 | | |
| 91 | **V-Set Regression Coefficient** | 0.999527 | | |
| 92 | **C-Set Regression Intercept** | 0.117782 | | |
| 93 | **V-Set Regression Intercept** | 0.212944 | | |
| 94 | **C-Set Regression Slope** | 0.99852 | | |
| 95 | **V-Set Regression Slope** | 0.998064 | | |
| 96 | | | | |
| 97 | **C-Set Orig. min** | 64 | | |
| 98 | **V-Set Orig. min** | 65 | | |
| 99 | **C-Set Orig. max** | 96 | | |
| 100 | **V-Set Orig. max** | 94 | | |
| 101 | **C-Set Orig. mean** | 79.5833 | | |
| 102 | **V-Set Orig. mean** | 79.3333 | | |
| 103 | **C-Set Orig. sdev** | 10.1575 | | |
| 104 | **V-Set Orig. sdev** | 11.0393 | | |
| 105 | | | | |
| 106 | **C-Set Pred. min** | 64.1316 | | |
| 107 | **V-Set Pred. min** | 65.2528 | | |
| 108 | **C-Set Pred. max** | 95.8965 | | |
| 109 | **V-Set Pred. max** | 94.2257 | | |
| 110 | **C-Set Pred. mean** | 79.5833 | | |
| 111 | **V-Set Pred. mean** | 79.3927 | | |
| 112 | **C-Set Pred. sdev** | 10.15 | | |
| 113 | **V-Set Pred. sdev** | 11.0232 | | |
| 114 | | | | |

**Property statistics**

| Row | Name | Description |
|---|---|---|
| 84 | C-Set BIAS | Always zero by definition. |
| 85 | V-Set BIAS | Average deviation between the predicted = NIR and original = lab method values. Should be as close to zero as possible (no systematic deviation). |
| 86 | C-Set SEE (SEC) | Standard error of estimation of the C-Set. Should be as small as possible, but comparable to the standard deviation of the laboratory method. |
| 87 | V-Set SEE (SEP) | Standard error of prediction of the V-Set. Should be as small as possible, but comparable to the standard deviation of the labor method. |
| 88 | Consistency | Relation between the standard error of estimation of C- and V-Set: SEC/SEP * 100.<br>Should be around 100 (80-110). |
| 90 | C-Set Regression Coefficient | "r" shows, how well the predicted values of C-Set match the original values on average. Should be as close to 1 as possible (r > 0.9). ***NOTE***:r is not r^2 (r-square), r can also be nagative. |
| 91 | V-Set Regression Coefficient | "r" shows, how well the predicted values of V-Set match the original values on average. Should be as close to 1 as possible (r > 0.9) |
| 92 | C-Set Regression Intercept | r= a + bx, where "a"= intercept of the C-Set. Should be as close to 0 as possible |
| 93 | V-Set Regression Intercept | r= a + bx, where "a"= intercept of the V-Set. Should be as close to 0 as possible |
| 94 | C-Set Regression Slope | r= a + bx, where "b"= slope of the C-Set. Should be as close to 1 as possible |

| Row | Name | Description |
|---|---|---|
| 95 | V-Set Regression Slope | r= a + bx, where "b"= slope of the V-Set. Should be as close to 1 as possible |
| | | Statistics of the original property values (labor method, X-axis) |
| 97 | C-Set Orig. min. | The smallest property value in the C-Set |
| 98 | V-Set Orig. min. | The smallest property value in the V-Set |
| 99 | C-Set Orig. max. | The highest property value in the C-Set |
| 100 | V-Set Orig. max. | The highest property value in the V-Set |
| 101 | C-Set Orig. mean. | The mean property value of the C-Set |
| 102 | V-Set Orig. mean. | The mean property value of the V-Set |
| 103 | C-Set Orig. sdev | The standard deviation of the property values in the C-Set |
| 104 | V-Set Orig. sdev | The standard deviation of the property values in the V-Set |
| 106-113 | | Statistics of the predicted property values (NIR, Y-axis). These values should be comparable to the similar values of the original property values. |



| | | |
|---|---|---|
| 115 | C-Set RSS | 1.69643 |
| 116 | V-Set RSS | 0.597645 |
| 117 | | |
| 118 | C-Set Durbin-Watson | 3.0593 |
| 119 | C-Set Durbin-Watson in range 1.5 to 2.5 | No |
| 120 | V-Set Durbin-Watson | 2.54473 |
| 121 | V-Set Durbin-Watson in range 1.5 to 2.5 | No |
| 122 | | |
| 123 | C-Set Resid. min | -0.464557 |
| 124 | V-Set Resid. min | -0.328338 |
| 125 | C-Set Resid. max | 0.625482 |
| 126 | V-Set Resid. max | 0.59974 |
| 127 | | |
| 128 | V-Set t-value | 0.427962 |
| 129 | V-Set t-Test(n-1,2-tail) Confidence [%] | 31.3514 |
| 130 | | |
| 131 | C-Set n | 12 |
| 132 | V-Set n | 6 |
| 133 | | |

| | | |
|---|---|---|
| 115 | C-Set RSS | C-Set residual sum of squares = RSS should be low: Sum((Orig-Predicted property)^2) = Sum (e^2) |
| 116 | V-Set RSS | V-Set residual sum of squares = RSS should be low: Sum((Orig-Predicted property)^2) = Sum (e^2) |
| 118 | C-Set Durbin-Watson | dw=Durbin-Watson = Sum of consecutive/succesive Residual Difference Square / Sum of Residual Square LINK: Durbin-Watson Factor Statistics |
| 119 | C-Set Durbin-Watson in range 1.5 to 2.5 | Is the dw value between 1.5 and 2.5? Answer: Yes or No |
| 120 | V-Set Durbin-Watson | dw=Durbin-Watson = Sum of consecutive/succesive Residual Difference Square / Sum of Residual Square |
| 121 | V-Set Durbin-Watson in range 1.5 to 2.5 | Is the dw value between 1.5 and 2.5? Answer: Yes or No |
| 123 | C-Set Resid. min | Biggest negative deviation in the C-Set property values |
| 124 | V-Set Resid. min | Biggest negative deviation in the V-Set property values |
| 125 | C-Set Resid. max | Biggest deviation in the C-Set property values |
| 126 | V-Set Resid. max | Biggest deviation in the V-Set property values |
| 128 | V-Set t-value | Original property and the Predicted property are compared via a paired t-test to show if the results are statistically equivalent |
| 129 | V-Set t-value (n-1, 2-tail) Confidence (%) | Significance level, where is no evidence for a difference between the Original and Predicted property results |
| 131 | C-Set n | Number of C-Set spectra |
| 132 | V-Set n | Number of V-Set spectra |

Explanation of V-Set t-test:
Original property and the Predicted property are compared via a paired t-test to show if the results are statistically equivalent.

$$t = \frac{|\bar{e}|}{s_d / \sqrt{n}}$$

where:

- e is the mean residual (original - predicted property);
- d is the standard deviation of the residuals.

Example:
The two-sided critical value of students'"t" e.g. at 5% significance level for n-1 degrees of freedom is protocoled as V-Set t (5%,n-1) and the calculated t as V-Set t-value. If V-Set t (5%, n-1) is greater than V-Set t-value, there is no evidence for a difference between the Original and Predicted results.
NIRCal estimates and protocols the significance level, where is no evidence for a difference between the Original and Predicted results.

V-Set t-Test (n-1,2-tail) Confidence [%]: the higher the value the best the confidence of the V-Set results.

See reference: [Anthony C. Moffat, Andrew D. Trafford, Roger D. Jee and Paul Graham. "Meeting the International Conference on Harmonisation's Guidelines on Validation of Analytical Procedures: Quantification as exemplified by a near-infrared reflectance assay of paracetamol in intact tablets.", The Royal Society of Chemistry, Analyst, 2000, 125,1341-1351].

***NOTE***

- the V-Set spectra selection has a huge influence on the t-value;
- the t-test reacts also negative on calibrations with very small V-Set BIAS. The t-test for the C-Set has no sence, because the C-Set BIAS is per default 0. t = Abs( BIAS ) / ( SEP / Sqrt(n) )

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 135 | | | | | | | | |
| 136 | **Validation for** | C-Set Spectra | | | | | | |
| 137 | **Property** | ethanol | | | | | | |
| 138 | | | | | | | | |
| 139 | **Spectrum Name** | No. | Orig | Pred | Orig - Pred | Extrapolation | Residual | Residual too big |
| 140 | **EtOH 96** | 1 | 96 | 95.8965 | 0.103519 | | 6.13798e-005 | |
| 141 | **EtOH 64** | 2 | 64 | 64.1316 | -0.131623 | | 9.16348e-005 | |
| 142 | **EtOH 67** | 4 | 67 | 66.7091 | 0.290922 | | 0.000167636 | |
| 143 | **EtOH 69** | 5 | 69 | 69.152 | -0.151998 | | 7.68871e-005 | |
| 144 | **EtOH 75** | 7 | 75 | 75.4646 | -0.464557 | | 8.5873e-005 | |
| 145 | **EtOH 76** | 8 | 76 | 75.3891 | 0.610904 | | 0.00028157 | |
| 146 | **EtOH 78** | 10 | 78 | 78.4499 | -0.449885 | | 0.000225863 | |
| 147 | **EtOH 79** | 11 | 79 | 79.1449 | -0.144927 | | 0.000197002 | |
| 148 | **EtOH 83** | 13 | 83 | 82.3745 | 0.625482 | | 9.48781e-005 | |
| 149 | **EtOH 86** | 14 | 86 | 86.2338 | -0.233807 | | 0.000112266 | |
| 150 | **EtOH 90** | 16 | 90 | 89.6391 | 0.360851 | | 0.000163933 | |
| 151 | **EtOH 92** | 17 | 92 | 92.4149 | -0.414881 | | 0.000243941 | |
| 152 | **Spectrum Name** | No. | Orig | Pred | Orig - Pred | Extrapolation | Residual | Residual too big |
| 153 | | | | | | | | |
| 154 | **Validation for** | V-Set Spectra | | | | | | |
| 155 | **Property** | ethanol | | | | | | |
| 156 | | | | | | | | |
| 157 | **Spectrum Name** | No. | Orig | Pred | Orig - Pred | Extrapolation | Residual | Residual too big |
| 158 | **EtOH 65** | 3 | 65 | 65.2528 | -0.25276 | | 0.000114213 | |
| 159 | **EtOH 70** | 6 | 70 | 70.0285 | -0.0284836 | | 0.000205743 | |
| 160 | **EtOH 77** | 9 | 77 | 77.3283 | -0.328338 | | 0.000234052 | |
| 161 | **EtOH 81** | 12 | 81 | 80.4003 | 0.59974 | | 0.000144012 | |
| 162 | **EtOH 89** | 15 | 89 | 89.1204 | -0.120379 | | 0.000193084 | |
| 163 | **EtOH 94** | 18 | 94 | 94.2257 | -0.225743 | | 0.000142493 | |
| 164 | **Spectrum Name** | No. | Orig | Pred | Orig - Pred | Extrapolation | Residual | Residual too big |
| 165 | | | | | | | | |
| 166 | | | | | | | | |

Statistics of the C-Set and V-Set:

| Row | Name | Description |
|---|---|---|
| 136-151 | Validation for C-Set Spectra | Name, number of spectrum, original, predicted property and the difference (original minus predicted); an X for "extrapolation", in case the value is outside the calibration range; residual value and an X for "residual too big", in case the value is outside the allowed range, spectra GUIDs |
| 154-163 | Validation for V-Set Spectra | |

### 3.17.3 Calibration Protocol Quantitative CV

Menu: Calibration / Calibration Protocol Calculate or Show

Icon:  Short Key: F8

The calibration protocol is an important validation report giving all information about the data selection and result of the chemometrics calculation. The calibration protocols are stored in the DB (or .nir project file).
Explanation of the content of the calibration protocols.

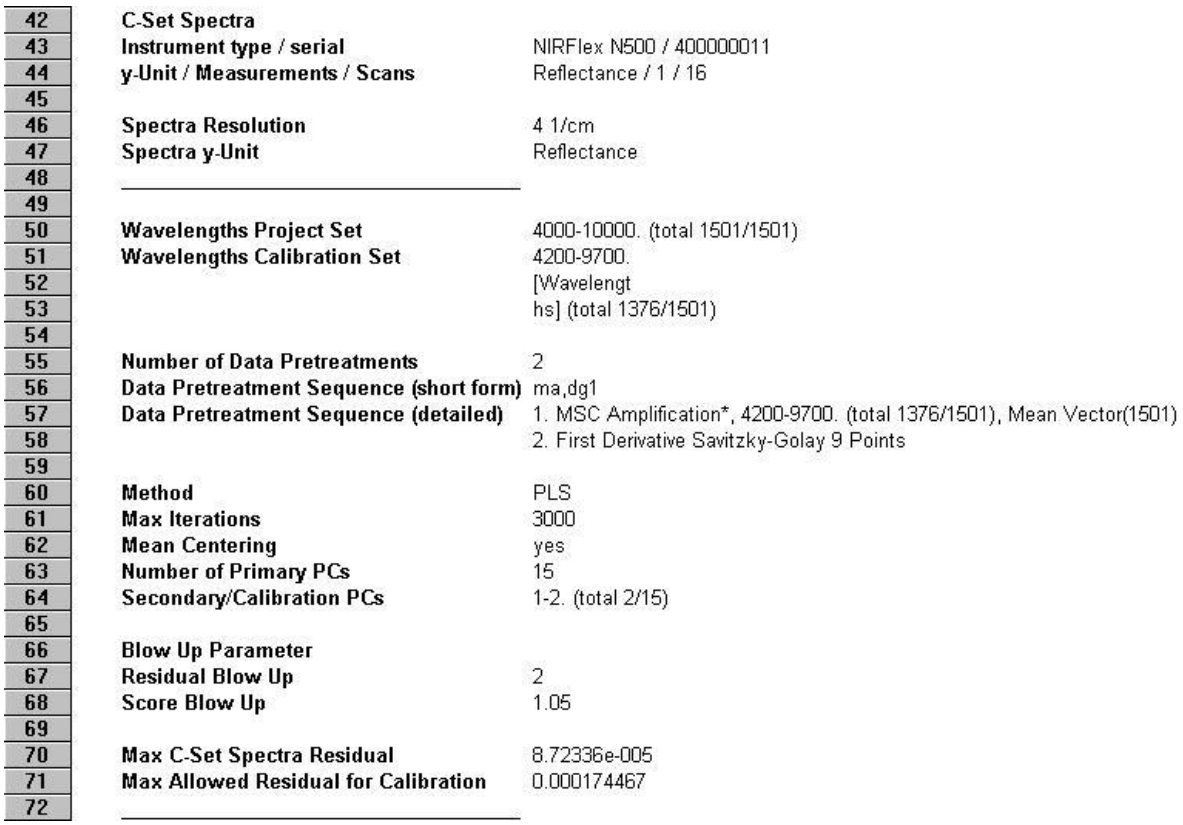
**Quantitative Calibration: Cross Validation**

The first section has the information about the user specific data selection (row 1-58).

| | | |
|---|---|---|
| 1 | **Calibration Protocol** | |
| 4 | | |
| 5 | | |
| 6 | **User** | Customer System Maintenance |
| 7 | **Date/Time** | 11.01.2006 13:42:42 |
| 8 | **Software** | NIRCal V5.1 (Build 400) |
| 9 | **Project File Name** | C:\NIRCal-Data\NIRCAL-5 Projekte\FromDB_Lactose quantitation.nir |
| 10 | **Project Comment** | |
| 11 | **Project GUID** | {98C449B2-729B-4385-8920-12874F9DB400} |
| 12 | **Calibration Name** | Sugar-Quant. |
| 13 | **Calibration Comment** | |
| 14 | **Calibration GUID** | {4923922A-2D1F-43EC-8BCB-F40E5B6AFC7C} |
| 15 | **Calibration Version** | 2 |
| 16 | **Calibration Lifecycle Sate** | created editing |
| 17 | | |
| 18 | | |
| 19 | **Properties in Project** | Lactose, Fine Sucrose. (total 2/2) |
| 20 | **Properties in Calibration Set** | Lactose. (total 1/2) |
| 21 | | |
| 22 | | |
| 23 | **Spectra in Project** | 69 |
| 24 | **Spectra in Calibration Set** | 66 |
| 25 | | |
| 26 | **Spectra in Calibration Set** | 1-55, 57-61, 63-67, 69. (total 66/69) |
| 27 | | |
| 28 | **Spectra unused (U-Set)** | 56, 62, 68. (total 3/69) |
| 29 | | |
| 30 | **Validation Method** | Cross Validation |
| 31 | **CV-mutable V-Set** | 1-42, 52-55, 57-61, 63-67, 69. (total 57/69) |
| 32 | **CV-permanent C-Set** | 43-51. (total 9/69) |
| 33 | **CV-unused Group** | 56, 62, 68. (total 3/69) |
| 34 | **CV Group # : Group Index** | (CV-mutable V-Set) |
| 35 | **1 : 1** | 1, 7, 13, 19, 25, 31, 37, 52, 58, 64. (total 10/69) |
| 36 | **2 : 2** | 2, 8, 14, 20, 26, 32, 38, 53, 59, 65. (total 10/69) |
| 37 | **3 : 3** | 3, 9, 15, 21, 27, 33, 39, 54, 60, 66. (total 10/69) |
| 38 | **4 : 4** | 4, 10, 16, 22, 28, 34, 40, 55, 61, 67. (total 10/69) |
| 39 | **5 : 5** | 5, 11, 17, 23, 29, 35, 41. (total 7/69) |
| 40 | **6 : 6** | 6, 12, 18, 24, 30, 36, 42, 57, 63, 69. (total 10/69) |

| Row | Name | Description |
|---|---|---|
| 1 | Calibration Protocol | Header |
| 6 | User | |
| 7 | Date/Time | Date and time of the calculation |
| 8 | Software | Current software version |
| 9 | Project File Name | Empty for DB, name of the project for file |
| 10 | Project Comment | Normally empty. The comment section of the project. |
| 11 | Project GUID | **G**lobal **U**nique **ID**entity of the project |
| 12 | Calibration Name | Name of the calibration |
| 13 | Calibration Comment | Normally empty |

| Row | Name | Description |
|---|---|---|
| 14 | Calibration GUID | **G**lobal **U**nique **ID**entity of the calibration |
| 15 | Calibration Version | Indicates the modification of the calibration. The Calibration Lifecycle Version is incremented by LifeCycle copy (or by XML import to avoid duplicates). |
| 16 | Calibration Lifecycle State | not available anymore |
| 19 | Properties in Project | Name and number of all properties in the project |
| 20 | Properties in Calibration Set | Name and number of calibrated properties |
| 23 | Spectra in Project | Total number of spectra in the project |
| 24 | Spectra in Calibration Set | Total number of spectra selected in the C-Set |
| 26 | Spectra in Calibration Set | Spectra selected in the C-Set |
| 28 | Spectra unused (U-Set) | Spectra not used for calibration (not in the C-Set) |
| 30 | Validation Method | Selected Validation method: Validation Set or Cross Validation (quantitative) |
| 31 | CV-mutable V-Set | Spectra used as V-Set (C-Set minus CV-permanent C-Set) |
| 32 | CV-permanent C-Set | Spectra used always in C-Set |
| 33 | CV-unused Group | Spectra belonging to a group, but not used for the CV |
| 34-40 | CV-Group #: Group Index | List of the CV-mutable V-Set spectra, belonging to different groups |




| | | |
|---|---|---|
| 42 | C-Set Spectra | Instrument statistic |
| 43 | Instrument type / serial | Used instrument type and it`s serial number |
| 44 | y-Unit / Measurements / Scans | Measuring principle / Number of repeated measurements and scans of a spectrum |

| Row | Name | Description |
|---|---|---|
| 47 | Spectra y-Unit | Measuring principle |
| 50 | Wavelengths Project Set | Measured wavelength range |
| 51 | Wavelengths Calibration Set | Selected wavelength range (selected datapoints / all data points) |
| 55 | Number of Data Pretreatments | Number of applied data pretreatment |
| 56 | Data Pretreatment Sequence (short) | Short name of the applied data pretreatment |
| 57 | Data Pretreatment Sequence (detailed) | Full name of the applied data pretreatment |
| 60 | Method | Selected calculation method / algorithm |
| 61 | Max Iterations | Calculation steps before default stop (max. 3000) Mean |
| 62 | Mean Centering | centering after pretreatment is activated: yes / no Number |
| 63 | Number of Primary PCs | of selected primary principal components |
| 64 | Secondary/Calibration PCs | Numbers of selected secondary principal components |
| 66 | Blow Up Parameter | Title |
| 67 | Residual Blow Up | Factor for residual limit calculation (default: 2) |
| 68 | Score Blow Up | Factor for score limit calculation (default: 1.05) |
| 70 | Max C-Set Spectra Residual | Highest value of the C-Set residuals |
| 71 | Max Allowed Residual for Calibration | Residual limit of the calibration for the application (max.* 2) |

| | | |
|---|---|---|
| 73 | | |
| 74 | **Q-Value V5** | 0.555521 |
| 75 | **Validation Method** | Cross Validation |
| 76 | | |
| 77 | **C-Set Residual too big** | 0 |
| 78 | | |
| 79 | **Num Properties** | 1 |
| 80 | **Rel. Consistency** | 0.381513 |
| 81 | **Weighted BIAS** | 0.000974035 |
| 82 | **Validity** | 0.00334391 |
| 83 | **Comparability** | 0.00200931 |
| 84 | **Precision** | 0.0297861 |
| 85 | **Weighted Accuracy** | 0 |
| 86 | | |

**Result of the calculation**

| Row | Name | Description |
|---|---|---|
| | | **Q-Value protocol** |
| 74 | Q-Value V5 | Q-Value of the calibration according version 5. The higher the Q-Value the best the calibration. |
| 75 | Validation Method | Selected Validation method: Validation Set or Cross Validation |
| 77 | C-Set Residual too big* | Number of C-Set spectra with too high residual = rejection of known (theoretical: 0) |
| 79 | Num Properties | Number of properties in the calibration (should be always 1 for the application) |
| 80 | Rel. Consistency* | The C- and V-Set should have similar low standard error of prediction (BIAS corrected): Abs(SEC-SEP)/(Abs(SEP)+1). Should be close to 0 |
| 81 | Weighted BIAS* | The absolute V-Set BIAS should be as close to zero as possible: Abs(Vset-BIAS)/Abs(Range) |

| Row | Name | Description |
|---|---|---|
| 82 | Validity* | The V-Set regression coefficient should be near to 1: 1-VsetRegr. Validity should be close to 0 |
| 83 | Comparability* | C- and V-Set should have similar high regression coefficient of prediction: Abs(CsetRegr-VsetRegr). Comparability should be close to 0 |
| 84 | Precision* | V-Set standard error of prediction (BIAS corrected) should be low: SEP/Abs (Range) |
| 85 | Weighted Accuracy* | V-Set residual sum of squares=RSS should be low (not BIAS corrected): Abs(Sum((Orig-Predicted)^2)/Abs (Range) |

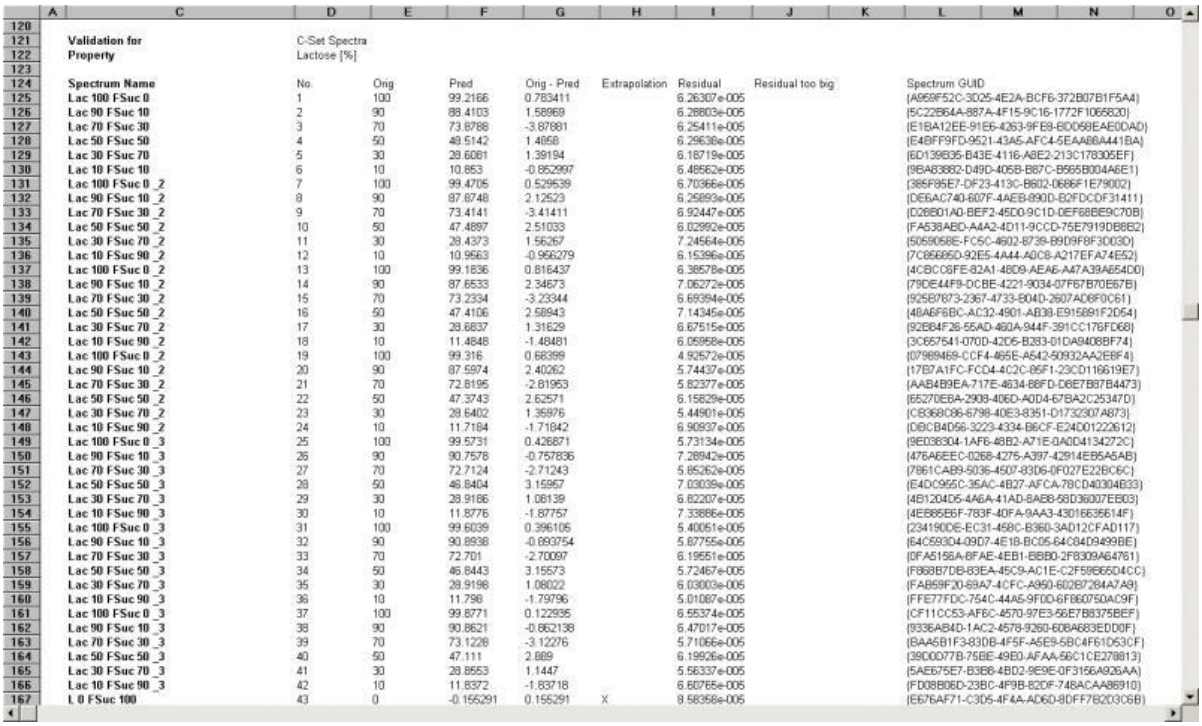
* used with different weighting values.

| | | |
|---|---|---|
| 87 | | |
| 88 | **Property Statistics** | Lactose [%] |
| 89 | | |
| 90 | **C-Set BIAS** | 0 |
| 91 | **C-Set SEE (SEC)** | 1.87989 |
| 92 | **CV SECV** | 3.00342 |
| 93 | | |
| 94 | **C-Set Regression Coefficient** | 0.998665 |
| 95 | **C-Set Regression Intercept** | 0.137817 |
| 96 | **C-Set Regression Slope** | 0.997333 |
| 97 | | |
| 98 | **C-Set Orig. min** | 0 |
| 99 | **C-Set Orig. max** | 100 |
| 100 | **C-Set Orig. mean** | 51.6667 |
| 101 | **C-Set Orig. sdev** | 36.3988 |
| 102 | | |
| 103 | **C-Set Pred. min** | -0.415629 |
| 104 | **C-Set Pred. max** | 99.8771 |
| 105 | **C-Set Pred. mean** | 51.6667 |
| 106 | **C-Set Pred. sdev** | 36.3502 |
| 107 | | |
| 108 | **C-Set RSS** | 229.71 |
| 109 | | |
| 110 | **C-Set Durbin-Watson** | 1.43165 |
| 111 | **C-Set Durbin-Watson in range 1.5 to 2.5** | No |
| 112 | | |
| 113 | **C-Set Resid. min** | -3.87881 |
| 114 | **C-Set Resid. max** | 3.15957 |
| 115 | | |
| 116 | | |
| 117 | **C-Set n** | 66 |
| 118 | | |

Statistics of the C-Set

| | | |
|---|---|---|
| 88 | Property statistics | Selected property |
| 90 | C-Set BIAS | Always zero by definition. |
| 91 | C-Set SEE (SEC) | Standard error of estimation of the C-Set. Should be as small as possible, but comparable to the standard deviation of the laboratory method. |
| 92 | CV SECV | Standard error of estimation of the C-Set. Should be as small as possible, but comparable to the standard deviation of the labor method. |
| 94 | C-Set Regression Coefficient | "r" shows, how well the predicted values of C-Set match the original values on average. Should be as close to 1 as possible (r > 0.9) |
| 95 | C-Set Regression Intercept | r= a + bx, where "a"= intercept of the C-Set. Should be as close to 0 as possible |
| 96 | C-Set Regression Slope | r= a + bx, where "b"= slope of the C-Set. Should be as close to 1 as possible |

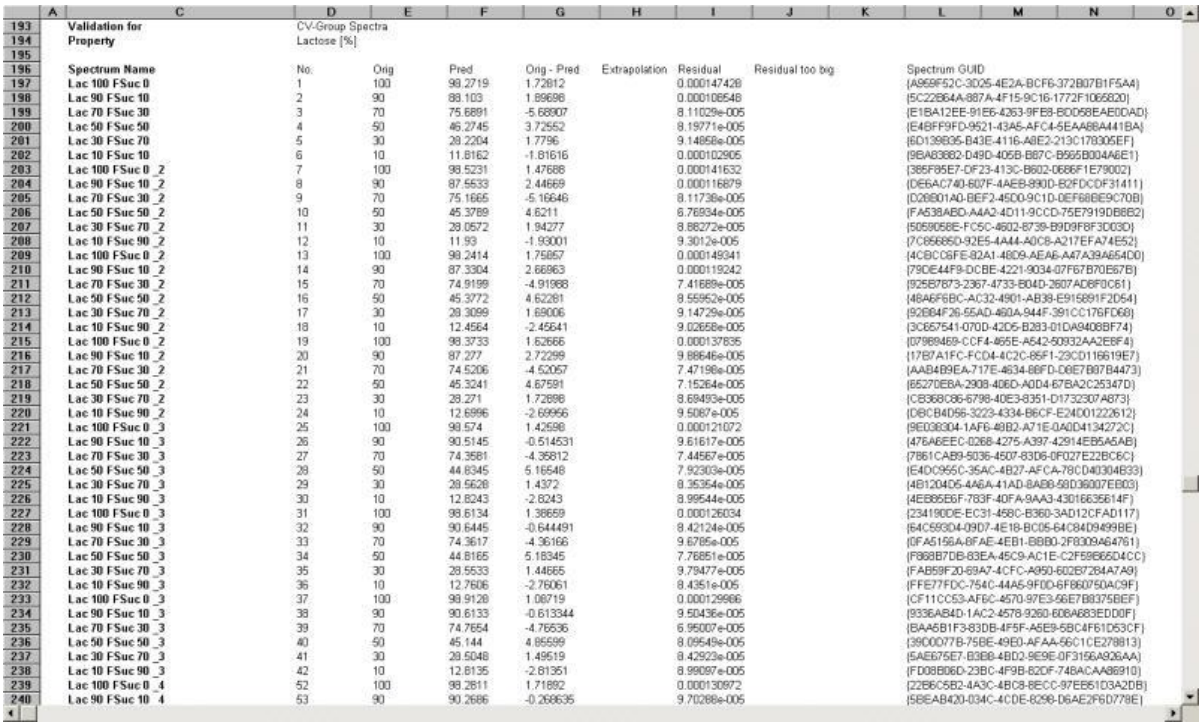
| Row | Name | Description |
|---|---|---|
| | | Statistics of the original property values (labor method, X-axis) |
| 98 | C-Set Orig. min. | The smallest property value in the C-Set |
| 99 | C-Set Orig. max. | The highest property value in the C-Set |
| 100 | C-Set Orig. mean. | The mean property value of the C-Set |
| 101 | C-Set Orig. sdev | The standard deviation of the property values in the C-Set |
| 103-106 | C-Set Pred. | Statistics of the predicted property values (NIR, Y-axis). These values should be comparable to the similar values of the original property values. |
| 108 | C-Set RSS | C-Set residual sum of squares = RSS should be low: Sum((Orig-Predicted property)^2) = Sum (e^2) |
| 110 | C-Set Durbin-Watson | dw=Durbin-Watson = Sum of consecutive/succesive Residual Difference Square / Sum of Residual Square LINK: Durbin-Watson Factor Statistics |
| 111 | C-Set Durbin-Watson in range 1.5 to 2.5 | Is the dw value between 1.5 and 2.5? Answer: Yes or No |
| 113 | C-Set Resid. min | Biggest negative deviation in the C-Set property values |
| 114 | C-Set Resid. max | Biggest positive deviation in the C-Set property values |
| 117 | C-Set n | Number of C-Set spectra |



| | A | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 120 | | | | | | | | | | | | | | |
| 121 | | Validation for | C-Set Spectra | | | | | | | | | | | |
| 122 | | Property | Lactose [%] | | | | | | | | | | | |
| 123 | | | | | | | | | | | | | | |
| 124 | | Spectrum Name | No. | Orig | Pred | Orig - Pred | Extrapolation | Residual | Residual too big | | Spectrum GUID | | | |
| 125 | | Lac 100 FSuc 0 | 1 | 100 | 99.2166 | 0.783411 | | 6.26307e-005 | | | {A959F52C-3D25-4E2A-BCF6-372B07B1F5A4} | | | |
| 126 | | Lac 90 FSuc 10 | 2 | 90 | 88.4103 | 1.58969 | | 6.28803e-005 | | | {5C22B64A-887A-4F15-9C16-1772F1066820} | | | |
| 127 | | Lac 70 FSuc 30 | 3 | 70 | 73.8788 | -3.87881 | | 6.25411e-005 | | | {E1BA12EE-91E6-4263-9FE8-BDD59EAE0DAD} | | | |
| 128 | | Lac 50 FSuc 50 | 4 | 50 | 48.5142 | 1.4858 | | 6.29636e-005 | | | {E4BFF9FD-9521-43A5-AFC4-5EAAB6A441BA} | | | |
| 129 | | Lac 30 FSuc 70 | 5 | 30 | 28.6081 | 1.39194 | | 6.18719e-005 | | | {6D139B35-B43E-4116-ABE2-213C178305EF} | | | |
| 130 | | Lac 10 FSuc 90 | 6 | 10 | 10.853 | -0.852997 | | 6.48562e-005 | | | {9BA83882-D49D-405B-B87C-B565B004A6E1} | | | |
| 131 | | Lac 100 FSuc 0_2 | 7 | 100 | 99.4705 | 0.529539 | | 6.70366e-005 | | | {385F85E7-DF23-413C-B602-0686F1E79002} | | | |
| 132 | | Lac 90 FSuc 10_2 | 8 | 90 | 87.8748 | 2.12523 | | 6.25893e-005 | | | {DE6AC740-607F-4AEB-890D-B2FDCDF31411} | | | |
| 133 | | Lac 70 FSuc 30_2 | 9 | 70 | 73.4141 | -3.41411 | | 6.92447e-005 | | | {D28B01A0-BEF2-45D0-9C1D-DEF68BE9C70B} | | | |
| 134 | | Lac 50 FSuc 50_2 | 10 | 50 | 47.4897 | 2.51033 | | 6.0298e-005 | | | {FA538ABD-A4A2-4D11-9CCD-75E7919DB8B2} | | | |
| 135 | | Lac 30 FSuc 70_2 | 11 | 30 | 28.4373 | 1.56267 | | 7.24564e-005 | | | {50590588-FC5C-4602-8739-B9D9F8F3D03D} | | | |
| 136 | | Lac 10 FSuc 90_2 | 12 | 10 | 10.9563 | -0.956279 | | 6.15396e-005 | | | {7CB56685D-92E5-4A44-A0C8-A217EFA74E52} | | | |
| 137 | | Lac 100 FSuc 0_2 | 13 | 100 | 99.1836 | 0.816437 | | 6.38578e-005 | | | {4CBCC6FE-82A1-46D9-AEA6-A47A39A654DD} | | | |
| 138 | | Lac 90 FSuc 10_2 | 14 | 90 | 87.6533 | 2.34673 | | 7.06272e-005 | | | {79DE44F9-DCBE-4221-9034-07F67B70E67B} | | | |
| 139 | | Lac 70 FSuc 30_2 | 15 | 70 | 73.2334 | -3.23344 | | 6.69364e-005 | | | {92587873-2367-4733-B04D-2607A06F0C61} | | | |
| 140 | | Lac 50 FSuc 50_2 | 16 | 50 | 47.4106 | 2.58943 | | 7.14345e-005 | | | {48A6F6BC-AC32-4901-AB38-E915891F2D54} | | | |
| 141 | | Lac 30 FSuc 70_2 | 17 | 30 | 28.6837 | 1.31629 | | 6.67515e-005 | | | {92B64F26-55AD-460A-944F-391CC176FD6B} | | | |
| 142 | | Lac 10 FSuc 90_2 | 18 | 10 | 11.4848 | -1.48481 | | 6.05958e-005 | | | {3C657541-070D-42D5-B283-01DA9408BF74} | | | |
| 143 | | Lac 100 FSuc 0_2 | 19 | 100 | 99.316 | 0.68399 | | 4.92572e-005 | | | {07989469-CCF4-465E-A542-50932AA2EBF4} | | | |
| 144 | | Lac 90 FSuc 10_2 | 20 | 90 | 87.5974 | 2.40262 | | 5.74437e-005 | | | {17B7A1FC-FCD4-4C2C-95F1-23CD116619E7} | | | |
| 145 | | Lac 70 FSuc 30_2 | 21 | 70 | 72.8195 | -2.81953 | | 5.82377e-005 | | | {AAB4B9EA-717E-4634-86FD-D8E7B87B4473} | | | |
| 146 | | Lac 50 FSuc 50_2 | 22 | 50 | 47.3743 | 2.62571 | | 6.15829e-005 | | | {65270EBA-2908-406D-A0D4-67BA2C25347D} | | | |
| 147 | | Lac 30 FSuc 70_2 | 23 | 30 | 28.6402 | 1.35976 | | 5.44901e-005 | | | {CB368C86-6798-40E3-8351-D1732307A873} | | | |
| 148 | | Lac 10 FSuc 90_2 | 24 | 10 | 11.7184 | -1.71842 | | 6.90937e-005 | | | {D8CB4D56-3223-4334-B6CF-E24D01222612} | | | |
| 149 | | Lac 100 FSuc 0_3 | 25 | 100 | 99.5731 | 0.426871 | | 5.73134e-005 | | | {9E038304-1AF6-48B2-A71E-0A0D4134272C} | | | |
| 150 | | Lac 90 FSuc 10_3 | 26 | 90 | 90.7578 | -0.757836 | | 7.28942e-005 | | | {476A6EEC-0268-4275-A397-42914EB5A5AB} | | | |
| 151 | | Lac 70 FSuc 30_3 | 27 | 70 | 72.7124 | -2.71243 | | 5.85262e-005 | | | {7861CAB9-5036-4507-83D6-0F027E22BC6C} | | | |
| 152 | | Lac 50 FSuc 50_3 | 28 | 50 | 46.8404 | 3.15967 | | 7.03039e-005 | | | {E4DC955C-35AC-4B27-AFCA-78CD40304B33} | | | |
| 153 | | Lac 30 FSuc 70_3 | 29 | 30 | 28.9186 | 1.08139 | | 6.82207e-005 | | | {4B1204D5-4A6A-41AD-8AB8-58D36007EB03} | | | |
| 154 | | Lac 10 FSuc 90_3 | 30 | 10 | 11.8776 | -1.87757 | | 7.33886e-005 | | | {4EB85E6F-783F-40FA-9AA3-43D16635614F} | | | |
| 155 | | Lac 100 FSuc 0_3 | 31 | 100 | 99.6039 | 0.396105 | | 5.40051e-005 | | | {234190DE-EC31-458C-B360-3AD12CFAD117} | | | |
| 156 | | Lac 90 FSuc 10_3 | 32 | 90 | 90.8938 | -0.893754 | | 5.87755e-005 | | | {64C593D4-09D7-4E18-BC05-64C84D9499BE} | | | |
| 157 | | Lac 70 FSuc 30_3 | 33 | 70 | 72.701 | -2.70097 | | 6.19551e-005 | | | {0FA5156A-8FAE-4EB1-BBB0-2F8309A64761} | | | |
| 158 | | Lac 50 FSuc 50_3 | 34 | 50 | 46.8443 | 3.15573 | | 5.72467e-005 | | | {F868B7DB-83EA-45C9-AC1E-C2F59B65D4CC} | | | |
| 159 | | Lac 30 FSuc 70_3 | 35 | 30 | 28.9198 | 1.08022 | | 6.03003e-005 | | | {FAB59F20-69A7-4CFC-A950-602B7284A7A9} | | | |
| 160 | | Lac 10 FSuc 90_3 | 36 | 10 | 11.798 | -1.79796 | | 5.01087e-005 | | | {FFE77FDC-754C-44A5-9FDD-6FB60750AC9F} | | | |
| 161 | | Lac 100 FSuc 0_3 | 37 | 100 | 99.8771 | 0.122935 | | 6.55374e-005 | | | {CF11CC53-AF8C-4570-97E3-56E7B8375BEF} | | | |
| 162 | | Lac 90 FSuc 10_3 | 38 | 90 | 90.8621 | -0.862138 | | 6.47017e-005 | | | {9336AB4D-1AC2-4578-9260-60BA683EDD0F} | | | |
| 163 | | Lac 70 FSuc 30_3 | 39 | 70 | 73.1228 | -3.12276 | | 5.71066e-005 | | | {BAA5B1F3-83DB-4F5F-A5E9-5BC4F61D53CF} | | | |
| 164 | | Lac 50 FSuc 50_3 | 40 | 50 | 47.111 | 2.889 | | 6.19926e-005 | | | {39DDD77B-75BE-49E0-AFAA-56C1CE278813} | | | |
| 165 | | Lac 30 FSuc 70_3 | 41 | 30 | 28.8553 | 1.1447 | | 5.56337e-005 | | | {5AE675E7-B3B8-4BD2-9E9E-0F3156A926AA} | | | |
| 166 | | Lac 10 FSuc 90_3 | 42 | 10 | 11.8372 | -1.83718 | | 6.60765e-005 | | | {FD08B06D-23BC-4F9B-82DF-748ACAA86910} | | | |
| 167 | | L 0 FSuc 100 | 43 | 0 | -0.155291 | 0.155291 | X | 8.58358e-005 | | | {E676AF71-C3D5-4F4A-AD6D-8DFF7B2D306B} | | | |

**Result of the final calibration**

| | | |
|---|---|---|
| 121-190 | Validation for C-Set Spectra | Name, number of spectrum, original, predicted property and the difference (Orig-Pred); an X for "extrapolation", in case the value is outside the calibration range; residual value and an X for "redidual too big", in case the value is outside the allowed range; spectrum GUID |



| | A | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 193 | | Validation for | CV-Group Spectra | | | | | | | | | | | |
| 194 | | Property | Lactose [%] | | | | | | | | | | | |
| 195 | | | | | | | | | | | | | | |
| 196 | | Spectrum Name | No. | | Orig | Pred | Orig - Pred | Extrapolation | Residual | Residual too big | Spectrum GUID | | | |
| 197 | | Lac 100 FSuc 0 | 1 | | 100 | 98.2719 | 1.72812 | | 0.000147428 | | {A969F52C-3D25-4E2A-BCF6-372B07B1F5A4} | | | |
| 198 | | Lac 90 FSuc 10 | 2 | | 90 | 88.103 | 1.89698 | | 0.000108548 | | {5C22B64A-887A-4F15-9C16-1772F1065820} | | | |
| 199 | | Lac 70 FSuc 30 | 3 | | 70 | 75.6891 | -5.68907 | | 8.11029e-005 | | {E1BA12EE-91E6-4263-9FB8-5DD58EAE0DAD} | | | |
| 200 | | Lac 50 FSuc 50 | 4 | | 50 | 46.2745 | 3.72552 | | 8.19771e-005 | | {E4BF9FD-9521-43A5-AFC4-5EAA86A441BA} | | | |
| 201 | | Lac 30 FSuc 70 | 5 | | 30 | 28.2204 | 1.7796 | | 9.14958e-005 | | {6D139B35-B43E-4116-A8E2-213C178305EF} | | | |
| 202 | | Lac 10 FSuc 90 | 6 | | 10 | 11.8162 | -1.81616 | | 0.000102905 | | {9BA83882-D49D-405B-B87C-B565B004A6E1} | | | |
| 203 | | Lac 100 FSuc 0 _2 | 7 | | 100 | 98.5231 | 1.47688 | | 0.000141632 | | {385F85E7-0F23-413C-B602-0686F1E79002} | | | |
| 204 | | Lac 90 FSuc 10 _2 | 8 | | 90 | 87.5533 | 2.44669 | | 0.000116879 | | {DE6AC740-607F-4AEB-890D-B2FDCDF31411} | | | |
| 205 | | Lac 70 FSuc 30 _2 | 9 | | 70 | 75.1665 | -5.16648 | | 8.11738e-005 | | {C2B601A0-BEF2-45D0-9C1D-0EF68BE9C70B} | | | |
| 206 | | Lac 50 FSuc 50 _2 | 10 | | 50 | 45.3789 | 4.6211 | | 6.76934e-005 | | {FA538ABD-A4A2-4D11-9CCD-75E7919DBBB2} | | | |
| 207 | | Lac 30 FSuc 70 _2 | 11 | | 30 | 28.0572 | 1.94277 | | 8.88272e-005 | | {5058058E-FC5C-4602-8739-B9D9F8F3D03D} | | | |
| 208 | | Lac 10 FSuc 90 _2 | 12 | | 10 | 11.93 | -1.93001 | | 9.3012e-005 | | {7C85685D-92E5-4A44-A0C8-A217EFA74E52} | | | |
| 209 | | Lac 100 FSuc 0 _2 | 13 | | 100 | 98.2414 | 1.75857 | | 0.000149341 | | {4CBCC6FE-82A1-48D9-AEA6-A47A39A654DD} | | | |
| 210 | | Lac 90 FSuc 10 _2 | 14 | | 90 | 87.3304 | 2.66963 | | 0.000119242 | | {79DE44F9-DCBE-4221-9034-07F67B70E67B} | | | |
| 211 | | Lac 70 FSuc 30 _2 | 15 | | 70 | 74.9199 | -4.91988 | | 7.41689e-005 | | {925B7873-2367-4733-B04D-2607AD6F0C61} | | | |
| 212 | | Lac 50 FSuc 50 _2 | 16 | | 50 | 45.3772 | 4.62281 | | 8.55952e-005 | | {48A6F6BC-AC32-4901-AB38-E915891F2D54} | | | |
| 213 | | Lac 30 FSuc 70 _2 | 17 | | 30 | 28.3099 | 1.69006 | | 9.14729e-005 | | {92B84F26-55AD-460A-944F-391CC176FD68} | | | |
| 214 | | Lac 10 FSuc 90 _2 | 18 | | 10 | 12.4564 | -2.45641 | | 9.02658e-005 | | {3C657541-070D-42D5-B283-01DA9408BF74} | | | |
| 215 | | Lac 100 FSuc 0 _2 | 19 | | 100 | 98.3733 | 1.62666 | | 0.000137835 | | {07989469-CCF4-465E-A542-50932AA2E8F4} | | | |
| 216 | | Lac 90 FSuc 10 _2 | 20 | | 90 | 87.277 | 2.72299 | | 9.88646e-005 | | {17B7A1FC-FCD4-4C2C-85F1-23CD116619E7} | | | |
| 217 | | Lac 70 FSuc 30 _2 | 21 | | 70 | 74.5206 | -4.52057 | | 7.47198e-005 | | {AAB4B9EA-717E-4634-88FD-D9E7B07B4473} | | | |
| 218 | | Lac 50 FSuc 50 _2 | 22 | | 50 | 45.3241 | 4.67591 | | 7.15264e-005 | | {B5270EBA-2908-406D-A0D4-67BA2C25347D} | | | |
| 219 | | Lac 30 FSuc 70 _2 | 23 | | 30 | 28.271 | 1.72898 | | 8.68493e-005 | | {CB368C86-6798-40E3-8351-D1732307A873} | | | |
| 220 | | Lac 10 FSuc 90 _2 | 24 | | 10 | 12.6996 | -2.69956 | | 9.5087e-005 | | {DBCB4D56-3223-4334-B6CF-E24D01222612} | | | |
| 221 | | Lac 100 FSuc 0 _3 | 25 | | 100 | 98.574 | 1.42598 | | 0.000121072 | | {9E038304-1AF6-49B2-A71E-0A0D4134272C} | | | |
| 222 | | Lac 90 FSuc 10 _3 | 26 | | 90 | 90.5145 | -0.514531 | | 9.61617e-005 | | {476A8EEC-0268-4275-A397-42914EB5A5AB} | | | |
| 223 | | Lac 70 FSuc 30 _3 | 27 | | 70 | 74.3581 | -4.35812 | | 7.44567e-005 | | {7861CAB9-5036-4507-83D6-0F027E22BC6C} | | | |
| 224 | | Lac 50 FSuc 50 _3 | 28 | | 50 | 44.8345 | 5.16548 | | 7.92303e-005 | | {E4DC955C-35AC-4B27-AFCA-78C040304B33} | | | |
| 225 | | Lac 30 FSuc 70 _3 | 29 | | 30 | 28.5628 | 1.4372 | | 8.35354e-005 | | {4B1204D5-4A6A-41AD-8AB8-59D36007EB03} | | | |
| 226 | | Lac 10 FSuc 90 _3 | 30 | | 10 | 12.8243 | -2.8243 | | 8.99544e-005 | | {4EB85E6F-783F-40FA-9AA3-43D16635614F} | | | |
| 227 | | Lac 100 FSuc 0 _3 | 31 | | 100 | 98.6134 | 1.38659 | | 0.000126034 | | {234190DE-EC31-458C-B360-3AD12CFAD117} | | | |
| 228 | | Lac 90 FSuc 10 _3 | 32 | | 90 | 90.6445 | -0.644491 | | 8.42124e-005 | | {64C593D4-09D7-4E18-BC05-64C84D9499BE} | | | |
| 229 | | Lac 70 FSuc 30 _3 | 33 | | 70 | 74.3617 | -4.36166 | | 9.6785e-005 | | {0FA5156A-8FAE-4EB1-BBB0-2F8309A64761} | | | |
| 230 | | Lac 50 FSuc 50 _3 | 34 | | 50 | 44.8165 | 5.18345 | | 7.76851e-005 | | {F868B7DB-83EA-45C9-AC1E-C2F59B65D4CC} | | | |
| 231 | | Lac 30 FSuc 70 _3 | 35 | | 30 | 28.5533 | 1.44665 | | 9.79477e-005 | | {FAB59F20-69A7-4CFC-A950-602B7264A7A9} | | | |
| 232 | | Lac 10 FSuc 90 _3 | 36 | | 10 | 12.7606 | -2.76061 | | 8.4351e-005 | | {FFE77FDC-754C-44A5-9F0D-6F860750AC9F} | | | |
| 233 | | Lac 100 FSuc 0 _3 | 37 | | 100 | 98.9128 | 1.08719 | | 0.000129986 | | {CF11CC53-AF6C-4570-97E3-56E7B8375BEF} | | | |
| 234 | | Lac 90 FSuc 10 _3 | 38 | | 90 | 90.6133 | -0.613344 | | 9.50436e-005 | | {9336AB4D-1AC2-4578-9260-60BA883EDD0F} | | | |
| 235 | | Lac 70 FSuc 30 _3 | 39 | | 70 | 74.7654 | -4.76536 | | 6.95007e-005 | | {BAA5B1F3-83DB-4F5F-A5E9-5BC4F61D53CF} | | | |
| 236 | | Lac 50 FSuc 50 _3 | 40 | | 50 | 45.144 | 4.85599 | | 8.09549e-005 | | {39C0D77B-75BE-49E0-AFAA-56C1CE278813} | | | |
| 237 | | Lac 30 FSuc 70 _3 | 41 | | 30 | 28.5048 | 1.49519 | | 8.42923e-005 | | {5AE675E7-B3B8-4BD2-9E9E-0F3156A926AA} | | | |
| 238 | | Lac 10 FSuc 90 _3 | 42 | | 10 | 12.8135 | -2.81351 | | 8.99097e-005 | | {FD08B06D-23BC-4F9B-82DF-74BACAA88910} | | | |
| 239 | | Lac 100 FSuc 0 _4 | 52 | | 100 | 98.2811 | 1.71892 | | 0.000130872 | | {22B6C5B2-4A3C-4BC8-BECC-97EB51D3A2DB} | | | |
| 240 | | Lac 90 FSuc 10 _4 | 53 | | 90 | 90.2686 | -0.268635 | | 9.70288e-005 | | {5BEAB420-034C-4CDE-8298-D6AE2F6D778E} | | | |

**Result of the CV-steps**

| Row | Name | Description |
|---|---|---|
| 194-253 | Validation for CV-Group Spectra | Name, number of spectrum, original, predicted property and the difference (Orig-Pred); an X for "extrapolation", in case the value is outside the calibration range; residual value and an X for "redidual too big",in case the value is outside the allowed range; spectrum GUID |

## 3.17.4 Prediction Protocol

Menu: Calibration / Prediction Protocol / Predict Project

Icon: 

**Prediction Protocol of quantitative Calibration**

The quantitative calibration will give the list of selected spectra with their name, the spectra number, the residual value. The outlier according residual means, the matrix is different, in this case the predicted property value should be in question. The scores outliers are less important: scores are intermediate results. The spectra with property outliers have extrapolated values: they are outside the calibrated range. The predicted and original property values give the information about the calibration quality.

Introduction

In NIRCal new statistical measures are included in the prediction protocol. The statistics are calculated in compliance with the ISO 12099 standard: Animal feeding stuff, cereals and milled cereal products Guidelines for the application of near infrared spectrometry.

The statistics are included for each property in the prediction protocol. The statistics are calculated using the spectra which are not flagged as outliers. If there are too many outliers, the limits for the calibration should be adjusted in the Calibration -> Outlier Detection -> Advanced menu.

| Calibration Properties | Min | Max | Offset | Slope | RMSEP | SEP | RSD | Bias |
|---|---|---|---|---|---|---|---|---|
| Moisture [%] | 33.95 | 43.85 | -0.1136001 | 1.0029908 | 0.2391134 | 0.2390958 | 0.2394709 | -0.0029012 |
| Fat [%] | 29.4 | 36.2 | 0.0568048 | 0.9985213 | 0.4415973 | 0.4415367 | 0.4423662 | 0.0073186 |

**Important**:
The statistics only gives relevant estimates of calibration performance if the prediction protocol is made using **an independent test set**.

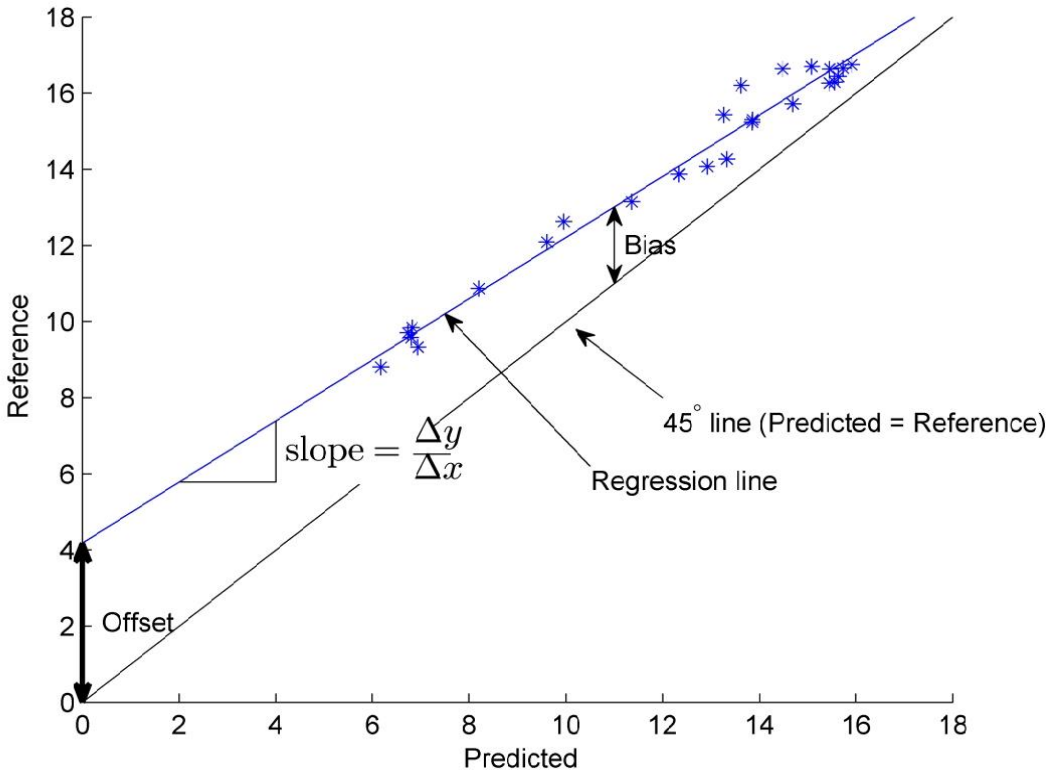
The statistics reported here can deviate from what is reported in the Original vs Predicted Property plots of two reasons: 1) outliers are not included, and 2) the linear regression to find slope and offset is made using the predicted values as the independent variable.

Definitions

The prediction protocol includes the following statistics:

- Offset
- Slope
- RMSEP
- SEP
- RSD
- Bias

The following figure explains some of the statistics.




The vector of difference between the predicted and the reference values for sample $i$ is defined as:

$$e_i = y_i - \hat{y}_i$$

Where $y_i$ is the reference value for sample $i$, $\hat{y}_i$ is the predicted value for sample $i$ and $e_i$ is the residual for sample $i$.

**Bias**: The bias is calculated as the mean of the differences.

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i$$

**RMSEP (Root Mean Square Error of Prediction)**: The RMSEP is an estimate of the random and systematic errors in the predictions.

$$RMSEP = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$$

**SEP (Standard Error of Prediction)**: The SEP is an estimate of the random error.

$$SEP = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(e_i - \bar{e})^2}$$

**Slope and offset**: The slope and offset of the regression line are both indicated in the figure above. They are found as a linear model of the form $y = ax+b$, where y is the reference value, $a$ is the slope, $x$ is the predicted value and $b$ is the offset:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$b = \text{offset} = \hat{\boldsymbol{\beta}}(1)$$

$$a = \text{slope} = \hat{\boldsymbol{\beta}}(2)$$

where $\mathbf{X}$ is the a $n$ x 2 matrix with 1 in the first column and $\hat{\boldsymbol{y}}$ in the second column; $\boldsymbol{y}$ is a column vector of reference values and ˆ is a two element column vector with the offset and slope.

**RSD (Residual Standard Deviation)**: This estimates the random error corrected with the slope.

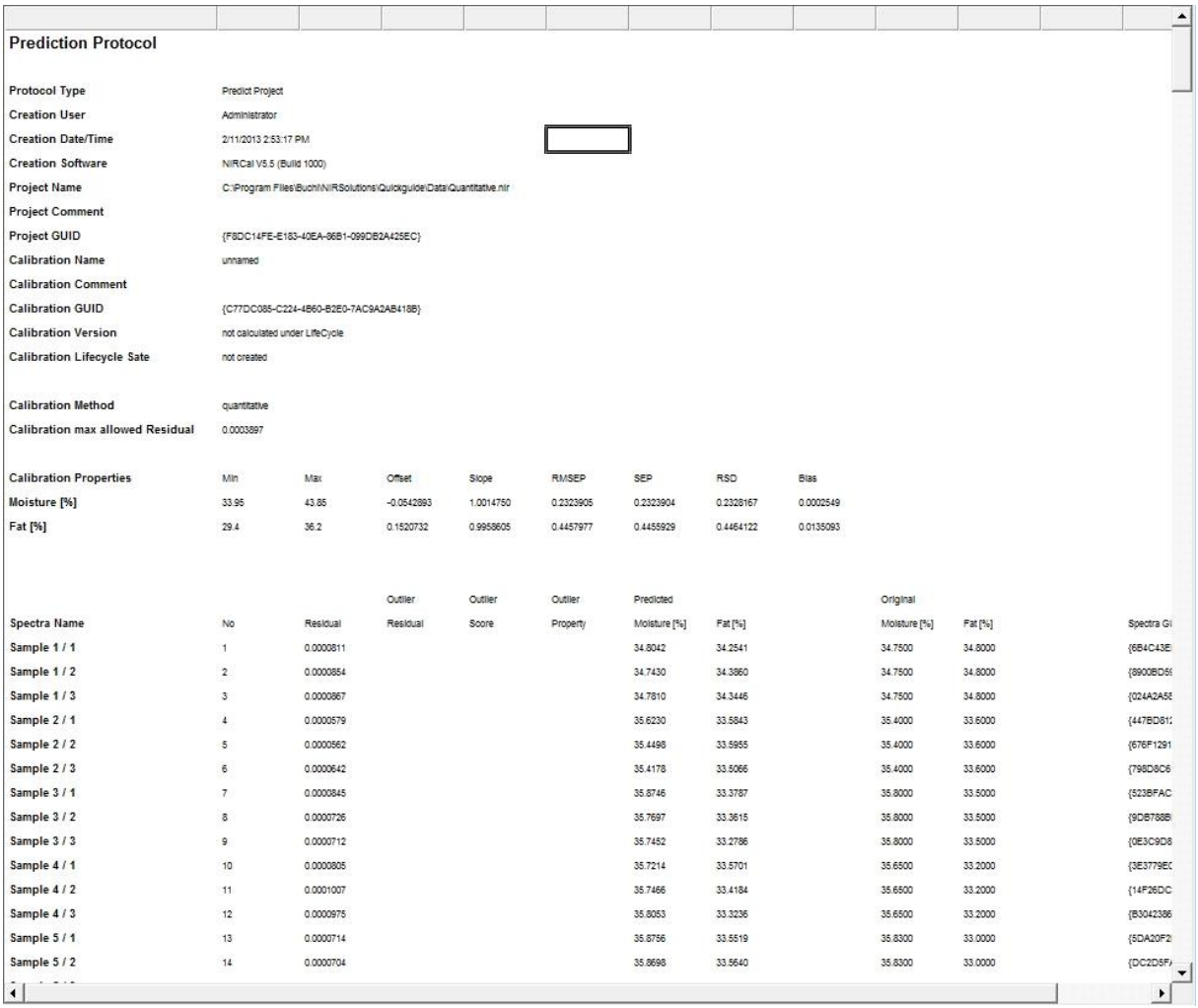
$$RSD = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - a + b\hat{y}_i)^2}$$

**RMSEP vs SEP vs RSD**: The three estimates of random errors each include

- **RMSEP**: Random error + systematic slope + bias
- **SEP**: Random error + systematic slope
- **RSD**: Random error

**Slope and bias correction**

If the prediction protocol is based on an **independent test set**, then the slope and offset calculated in the prediction protocol can be used directly in the management console to make slope and bias correction. **Notice that the offset value from the prediction protocol must be entred in the bias field.**



**Prediction Protocol**

| | |
|---|---|
| Protocol Type | Predict Project |
| Creation User | Administrator |
| Creation Date/Time | 2/11/2013 2:53:17 PM |
| Creation Software | NIRCal V5.5 (Build 1000) |
| Project Name | C:\Program Files\Buchi\NIRSolutions\Quickguide\Data\Quantitative.nir |
| Project Comment | |
| Project GUID | {F8DC14FE-E183-40EA-86B1-099DB2A425EC} |
| Calibration Name | unnamed |
| Calibration Comment | |
| Calibration GUID | {C77DC085-C224-4B60-B2E0-7AC9A2AB418B} |
| Calibration Version | not calculated under LifeCycle |
| Calibration Lifecycle Sate | not created |
| Calibration Method | quantitative |
| Calibration max allowed Residual | 0.0003897 |

| Calibration Properties | Min | Max | Offset | Slope | RMSEP | SEP | RSD | Bias |
|---|---|---|---|---|---|---|---|---|
| Moisture [%] | 33.95 | 43.85 | -0.0542893 | 1.0014750 | 0.2323905 | 0.2323904 | 0.2328167 | 0.0002549 |
| Fat [%] | 29.4 | 36.2 | 0.1520732 | 0.9958605 | 0.4457977 | 0.4455929 | 0.4464122 | 0.0135093 |

| | | | Outlier | Outlier | Outlier | Predicted | | Original | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Spectra Name | No | Residual | Residual | Score | Property | Moisture [%] | Fat [%] | Moisture [%] | Fat [%] | Spectra GI |
| Sample 1 / 1 | 1 | 0.0000811 | | | | 34.8042 | 34.2541 | 34.7500 | 34.8000 | {6B4C43E |
| Sample 1 / 2 | 2 | 0.0000854 | | | | 34.7430 | 34.3860 | 34.7500 | 34.8000 | {8900BD5 |
| Sample 1 / 3 | 3 | 0.0000867 | | | | 34.7810 | 34.3446 | 34.7500 | 34.8000 | {024A2A5 |
| Sample 2 / 1 | 4 | 0.0000579 | | | | 35.6230 | 33.5843 | 35.4000 | 33.6000 | {447BD61 |
| Sample 2 / 2 | 5 | 0.0000562 | | | | 35.4498 | 33.5955 | 35.4000 | 33.6000 | {676F1291 |
| Sample 2 / 3 | 6 | 0.0000642 | | | | 35.4178 | 33.5066 | 35.4000 | 33.6000 | {798D8C6 |
| Sample 3 / 1 | 7 | 0.0000845 | | | | 35.8746 | 33.3787 | 35.8000 | 33.5000 | {523BFAC |
| Sample 3 / 2 | 8 | 0.0000726 | | | | 35.7697 | 33.3615 | 35.8000 | 33.5000 | {9DB768B |
| Sample 3 / 3 | 9 | 0.0000712 | | | | 35.7452 | 33.2786 | 35.8000 | 33.5000 | {0E3C9D8 |
| Sample 4 / 1 | 10 | 0.0000805 | | | | 35.7214 | 33.5701 | 35.6500 | 33.2000 | {3E3779E |
| Sample 4 / 2 | 11 | 0.0001007 | | | | 35.7466 | 33.4184 | 35.6500 | 33.2000 | {14F26DC |
| Sample 4 / 3 | 12 | 0.0000975 | | | | 35.8053 | 33.3236 | 35.6500 | 33.2000 | {B3042386 |
| Sample 5 / 1 | 13 | 0.0000714 | | | | 35.8756 | 33.5519 | 35.8300 | 33.0000 | {5DA20F2 |
| Sample 5 / 2 | 14 | 0.0000704 | | | | 35.8698 | 33.5640 | 35.8300 | 33.0000 | {DC2D5F |

**Prediction Protocol of qualitative Calibration**

The prediction protocol (SIMCA and Cluster) can be configurated: LINK: Cluster Prediction Protocol.

The documented 7 cases are:

| Case | Group | Symbol | Residual | Distance | Name/Property |
|---|---|---|---|---|---|
| 1 | not in a cluster | ! | OK | not OK | not OK |
| 2 | false identified | ? | OK | OK | not OK |
| 3 | identified | * | OK | OK | OK |
| 4 | not identified CLU OK | % | not OK | OK | OK |
| 5 | not identified CLU BAD | & | not OK | OK | not OK |
| 6 | not identified known | - | not OK | not OK | OK |
| 7 | not identified unknown | = | not OK | not OK | not OK |

Example:

| | | | |
|---|---|---|---|
| 1 | **Prediction Protocol** | | |
| 4 | | | |
| 5 | | | |
| 6 | **Protocol Type** | Predict Project | |
| 7 | **Creation User** | Customer System Maintenance | |
| 8 | **Creation Date/Time** | 4/29/2005 10:46:14 AM | |
| 9 | **Creation Software** | NIRCal V5 (Build 3000) | |
| 10 | **Project Name** | Fluka Ident | |
| 11 | **Project Comment** | | |
| 12 | **Project GUID** | {BA28704E-DC90-4B17-A5A9-79C342DF42E2} | |
| 13 | **Calibration Name** | SIMCA Oxalic acid | |
| 14 | **Calibration Comment** | | |
| 15 | **Calibration GUID** | {F6B39CD1-90A5-4C6B-AEB2-4A5295E32695} | |
| 16 | **Calibration Version** | 0 | |
| 17 | **Calibration Lifecycle Sate** | created editing | |
| 18 | | | |
| 19 | **Calibration Method** | qualitative | |
| 20 | **Calibration max allowed Residual** | | 0.0001972 |
| 21 | | | |
| 22 | **Calibration Properties** | Min | Max |
| 23 | **Oxalic acid** | 0 | 1 |
| 24 | | | |
| 25 | **Protocol filter settings** | | |
| 26 | **not in a cluster (!)** | on | |
| 27 | **false identified (?)** | on | |
| 28 | **identified (*)** | on | |
| 29 | **not identified known (-)** | on | |
| 30 | **not identified unknown (=)** | on | |
| 31 | **not identified Clu OK (%)** | on | |
| 32 | **not identified Clu BAD (&)** | on | |

On the top of the protocol about some general information about the calibration, the calibrated properties are listed and the filter setting are visible. The **calibration max. allowed residual** is important to note.

| | | | | Outlier | Outlier | Outlier | | | |
|---|---|---|---|---|---|---|---|---|---|
| 35 | **Spectra Name** | No | Residual | Residual | Score | Property | Predicted | | Original |
| 36 | **75702 Oxalsäure -007** | 1* | 0.0000655 | | | | Oxalic acid | | Oxalic acid |
| 37 | **75702 Oxalsäure -007** | 2* | 0.0000729 | | | | Oxalic acid | | Oxalic acid |
| 38 | **75702 Oxalsäure -007** | 3* | 0.0000677 | | | | Oxalic acid | | Oxalic acid |
| 39 | **75702 Oxalsäure -011** | 4* | 0.0000591 | | | | Oxalic acid | | Oxalic acid |
| 40 | **75702 Oxalsäure -011** | 5* | 0.0000550 | | | | Oxalic acid | | Oxalic acid |
| 41 | **75702 Oxalsäure -011** | 6* | 0.0000516 | | | | Oxalic acid | | Oxalic acid |
| 42 | **75702 Oxalsäure -011** | 7* | 0.0000892 | | | | Oxalic acid | | Oxalic acid |
| 43 | **75702 Oxalsäure -011** | 8* | 0.0000652 | | | | Oxalic acid | | Oxalic acid |
| 44 | **75702 Oxalsäure -011** | 9* | 0.0000886 | | | | Oxalic acid | | Oxalic acid |
| 45 | **75702 Oxalsäure -009** | 10* | 0.0000629 | | | | Oxalic acid | | Oxalic acid |
| 46 | **75702 Oxalsäure -009** | 11* | 0.0000705 | | | | Oxalic acid | | Oxalic acid |
| 47 | **75702 Oxalsäure -009** | 12* | 0.0000734 | | | | Oxalic acid | | Oxalic acid |
| 48 | **75702 Oxalsäure -010** | 13* | 0.0000696 | | | | Oxalic acid | | Oxalic acid |
| 49 | **75702 Oxalsäure -010** | 14* | 0.0000789 | | | | Oxalic acid | | Oxalic acid |
| 50 | **75702 Oxalsäure -010** | 15* | 0.0000691 | | | | Oxalic acid | | Oxalic acid |
| 51 | **0001_448972/1-01-005** | 16= | 0.0016402 | X | X | X | | | K-PO4 dibasic |
| 52 | **0001_448972/1-01-005** | 17= | 0.0016495 | X | X | X | | | K-PO4 dibasic |
| 53 | **0001_448972/1-01-005** | 18= | 0.0016432 | X | X | X | | | K-PO4 dibasic |
| 54 | **0006_60355 K-PO4 dibas_449026/1-01-007** | 19= | 0.0016514 | X | X | X | | | K-PO4 dibasic |
| 55 | **0006_60355 K-PO4 dibas_449026/1-01-007** | 20= | 0.0016510 | X | X | X | | | K-PO4 dibasic |
| 56 | **0006_60355 K-PO4 dibas_449026/1-01-007** | 21= | 0.0016465 | X | X | X | | | K-PO4 dibasic |
| 57 | **0007_60355 K-PO4 dibas_449026/1-01-007** | 22= | 0.0016427 | X | X | X | | | K-PO4 dibasic |
| 58 | **0007_60355 K-PO4 dibas_449026/1-01-007** | 23= | 0.0016447 | X | X | X | | | K-PO4 dibasic |
| 59 | **0007_60355 K-PO4 dibas_449026/1-01-007** | 24= | 0.0016479 | X | X | X | | | K-PO4 dibasic |
| 60 | **0008_60355 K-PO4 dibas_449026/1-01-007** | 25= | 0.0016442 | X | X | X | | | K-PO4 dibasic |
| 61 | **0008_60355 K-PO4 dibas_449026/1-01-007** | 26= | 0.0016471 | X | X | X | | | K-PO4 dibasic |
| 62 | **0008_60355 K-PO4 dibas_449026/1-01-007** | 27= | 0.0016416 | X | X | X | | | K-PO4 dibasic |
| 63 | **0009_60355 K-PO4 dibas_449026/1-01-007** | 28= | 0.0016333 | X | X | X | | | K-PO4 dibasic |
| 64 | **0009_60355 K-PO4 dibas_449026/1-01-007** | 29= | 0.0016388 | X | X | X | | | K-PO4 dibasic |
| 65 | **0009_60355 K-PO4 dibas_449026/1-01-007** | 30= | 0.0016358 | X | X | X | | | K-PO4 dibasic |

Each spectra name is listed with the number. Behind the number there is the symbol, which identification case it is. The actual residual will be compared with the allowed max. residual and in case it is higher, the spectrum is a residual outlier and will be not identified: residual criteria is not fulfilled. The possible scores outliers are less important: LINK: Cluster Cal.Protocol. Under predicted property the filed is empty, if the spectrum is not a cluster. There is the name of the cluster, in case the spectrum is in a certain cluster.
In case the predicted and original property names are matching, it is correct prediction: spectra 1-15.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 93 | 0001_09900 NH4-Oxalate Monohydr_449073/1 | 58 = | 0.0020416 | X | X | X | | | NH4-Oxalate monohydr |
| 94 | 0001_09900 NH4-Oxalate Monohydr_449073/1 | 59 = | 0.0020215 | X | X | X | | | NH4-Oxalate monohydr |
| 95 | 0001_09900 NH4-Oxalate Monohydr_449073/1 | 60 = | 0.0020328 | X | X | X | | | NH4-Oxalate monohydr |
| 96 | 0002_09900 NH4-Oxalate Monohydr_449073/ | 61 = | 0.0020259 | X | X | X | | | NH4-Oxalate monohydr |
| 97 | 0002_09900 NH4-Oxalate Monohydr_449073/ | 62 = | 0.0020141 | X | X | X | | | NH4-Oxalate monohydr |
| 98 | 0002_09900 NH4-Oxalate Monohydr_449073/ | 63 = | 0.0020201 | X | X | X | | | NH4-Oxalate monohydr |
| 99 | 0003_09900 NH4-Oxalate Monohydr_449073/ | 64 = | 0.0020348 | X | X | X | | | NH4-Oxalate monohydr |
| 100 | 0003_09900 NH4-Oxalate Monohydr_449073/ | 65 = | 0.0020326 | X | X | X | | | NH4-Oxalate monohydr |
| 101 | 0003_09900 NH4-Oxalate Monohydr_449073/ | 66 = | 0.0020199 | X | X | X | | | NH4-Oxalate monohydr |
| 102 | 0004_09900 NH4-Oxalate Monohydr_449073/ | 67 = | 0.0020272 | X | X | X | | | NH4-Oxalate monohydr |
| 103 | 0004_09900 NH4-Oxalate Monohydr_449073/ | 68 = | 0.0020121 | X | X | X | | | NH4-Oxalate monohydr |
| 104 | 0004_09900 NH4-Oxalate Monohydr_449073/ | 69 = | 0.0020263 | X | X | X | | | NH4-Oxalate monohydr |
| 105 | 0005_09900 NH4-Oxalate Monohydr_449073/ | 70 = | 0.0020022 | X | X | X | | | NH4-Oxalate monohydr |
| 106 | 0005_09900 NH4-Oxalate Monohydr_449073/ | 71 = | 0.0020404 | X | X | X | | | NH4-Oxalate monohydr |
| 107 | 0005_09900 NH4-Oxalate Monohydr_449073/ | 72 = | 0.0020188 | X | X | X | | | NH4-Oxalate monohydr |
| 108 | 0001_51456 Urea_449278/1-01-006-10 | 73 & | 0.0042226 | X | | X | Oxalic acid | | Urea |
| 109 | 0001_51456 Urea_449278/1-01-006-10 | 74 & | 0.0042696 | X | | X | Oxalic acid | | Urea |
| 110 | 0001_51456 Urea_449278/1-01-006-10 | 75 & | 0.0043015 | X | | X | Oxalic acid | | Urea |
| 111 | 0002_51456 Urea_449278/1-01-006-10 | 76 & | 0.0042500 | X | X | X | Oxalic acid | | Urea |
| 112 | 0002_51456 Urea_449278/1-01-006-10 | 77 & | 0.0042337 | X | X | X | Oxalic acid | | Urea |
| 113 | 0002_51456 Urea_449278/1-01-006-10 | 78 = | 0.0042785 | X | | X | | | Urea |
| 114 | 0003_51456 Urea_449278/1-01-006-10 | 79 = | 0.0042870 | X | | X | | | Urea |
| 115 | 0003_51456 Urea_449278/1-01-006-10 | 80 = | 0.0042150 | X | | X | | | Urea |
| 116 | 0003_51456 Urea_449278/1-01-006-10 | 81 = | 0.0042590 | X | | X | | | Urea |
| 117 | 0004_51456 Urea_449278/1-01-006-10 | 82 = | 0.0042155 | X | | X | | | Urea |
| 118 | 0004_51456 Urea_449278/1-01-006-10 | 83 & | 0.0042671 | X | | X | Oxalic acid | | Urea |
| 119 | 0004_51456 Urea_449278/1-01-006-10 | 84 = | 0.0042658 | X | | X | | | Urea |
| 120 | Total outliers | | | 69 | 59 | 69 | | | |
| 121 | | | | | | | | | |
| 122 | Total not in a cluster (!) | 0 | Residual OK, but not in a cluster | | | | | | |
| 123 | Total identified (*) | 15 | Residual OK, cluster found, predicted is equal original | | | | | | |
| 124 | Total false identified (?) | 0 | Residual OK, cluster found, predicted is NOT equal original | | | | | | |
| 125 | Total not identified, Cluster OK (%) | 0 | Residual too big, cluster found, predicted is equal original | | | | | | |
| 126 | Total not identified, Cluster BAD (&) | 6 | Residual too big, cluster found, predicted is NOT equal original | | | | | | |
| 127 | Total not identified known (-) | 0 | Residual too big, no cluster found, original property is available | | | | | | |
| 128 | Total not identified unknown (=) | 63 | Residual too big, no cluster found, original property is not available | | | | | | |

In case the predicted and original properties are different, it is a wrong identification: spectra 73-77 and 83. These spectra are in a false cluster, but will not be identified, while the residuals are much higher (0.0042) as the allowed residual (0.0002).

**Correct identifications** -case 3- and not identified unknown -case 7- are **less important**, deactivating e.g case 3 and 7 will give a shorter protocol.
Especially cases **1 and 5 are critical**. By case 5 the residual relationship (actual against allowed) should be controlled. **Case 2 is inacceptable**.

In summary all spectra should be listed under the number of “Total identified” or “Total not identified unknown”, no spectra should be in the 5 other types. Printouts of the prediction protocols are part of the validation procedure.

When the result is unsatisfactory, the calibration adjustment should be changed.
In case there is no any calibration, which problems without works, the possibilities of improvement are:

| | |
|---|---|
| not in a Cluster (!) | Add these spectra to the project without allocating them to C- or V-Set. Calculate and observe if the spectra are lying far from the cluster. If not, the spectra must remain in the project (similar spectra) |
| false identified (?) | Add the spectrum that was identified as false to the project and enlarge the calculation (similar spectra) |
| identified (*) | The spectra are OK (known, identified) |
| not identified Clu OK (%) | These spectra are rare, because the known substance spectra are only to be found in one project, eventually add such spectra to the C-Set. |
| not identified Clu BAD (&) | Reducing the Radii Blow Up Factor may help. Otherwise compare max allowed residual with actual residual. When the actual one is 150% or more than the allowed, the calibration can be used. In case the actual residual is slightly higher than the allowed, the spectra should be calibrated in the same project. |
| not identified known (-) | These spectra are rare because the known substance spectra are only to be found in one project, eventually add these to the C-Set. |
| not identified unknown (=) | The spectra are OK (unknown, not identified). |

**Only calibrations without false identification can be used.**

***NOTE***
The warning: incompatible spectra means, that the number of data point or the wavelength range of the calibration and predicted spectra are not matching because of different resolution or different instrument type.

# 3.18 Matrices

## 3.18.1 How to display matrices

Each matrix can be displayed in different plots. Available views are:

- 1D Scatter
- 2D Line
- 2D Scatter
- 2D Combined Scatter
- 3D Scatter
- 3D Surface
- Table

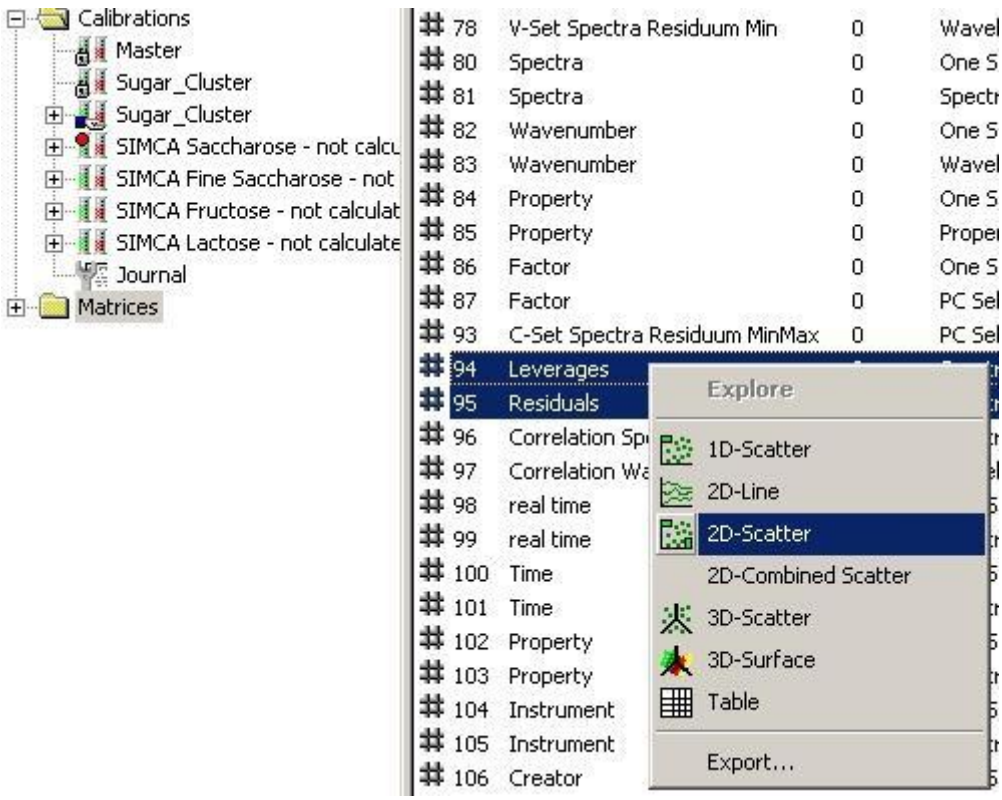
Export:
Export the matrix to a *.dat File. To export the table to excel display the matrix first as table and then export to *.xml (slow) or *.txt (fast) format.

Rename:
Matrices can be temporarily renamed.

**Example : 2D Scatter plot**

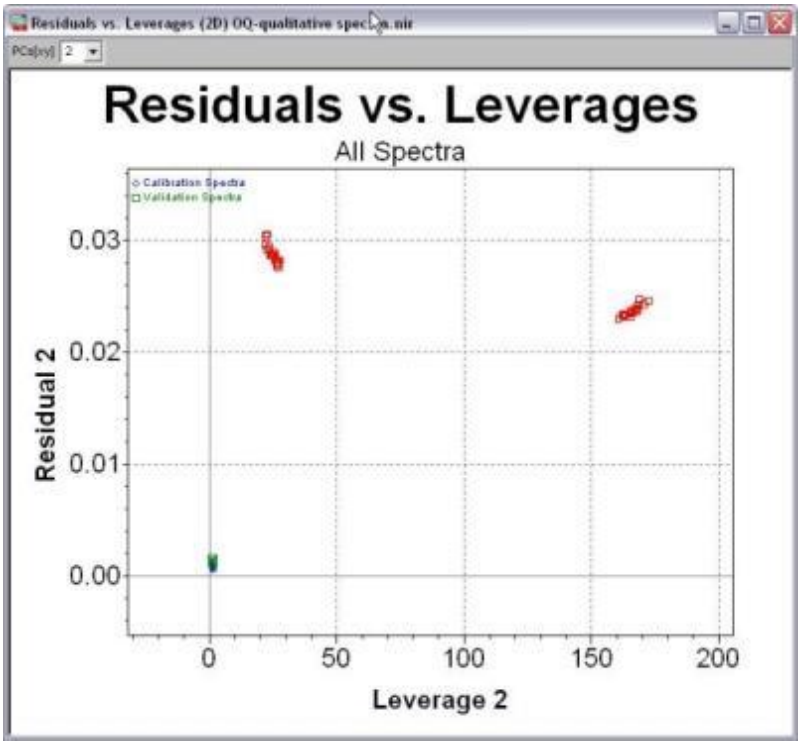
The Coomans plot is part of the SIMCA Overview plot. It is a 2D-Scatter plot of the 2 matrices

- Residual (matrix ID 95)
- Leverages (matrix ID 94)



1. Select **Matrices** in the NIR-Explorer tree
2. then select the two matrices in the NIR-Explorer
3. click the right mouse and then 2D-Scatter

***NOTE***
To save such a created plot a workspace can be saved and reused within other projects.




### 3.18.2 Calibration Residual Limits

| | |
|---|---|
| Description | The min. and max. residual is depending on the number of primary PCs. |
| Use | The maximum allowed residual of the calibration is valid for the application. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 55 |
| Tip | The highest residual of the C-Set is used for the residual limit calculation. |
| Details | The allowed residual = 2 * maximum C-Set residual. 2 is the default of residual blow up, can be edited by the user. |
| Related Topic | Residuals, Primary PCs |

### 3.18.3 Calibration Residual Limits with Blow Up

| | |
|---|---|
| Description | The min. and max. residual is depending on the number of primary PCs. |
| Use | The maximum allowed residual of the calibration is valid for the application. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 56 |
| Tip | The highest residual of the C-Set is used for the residual limit calculation. |
| Details | The allowed residual = 2 * maximum C-Set residual.<br>2 is the default of residual blow up, can be edited by the user. Do not use residual blow up smaller as 1!<br>The minimum residual = minimum C-Set residual /2. It is not used for the application. |
| Related Topic | Residuals, Primary PCs, Blow Up Limits |

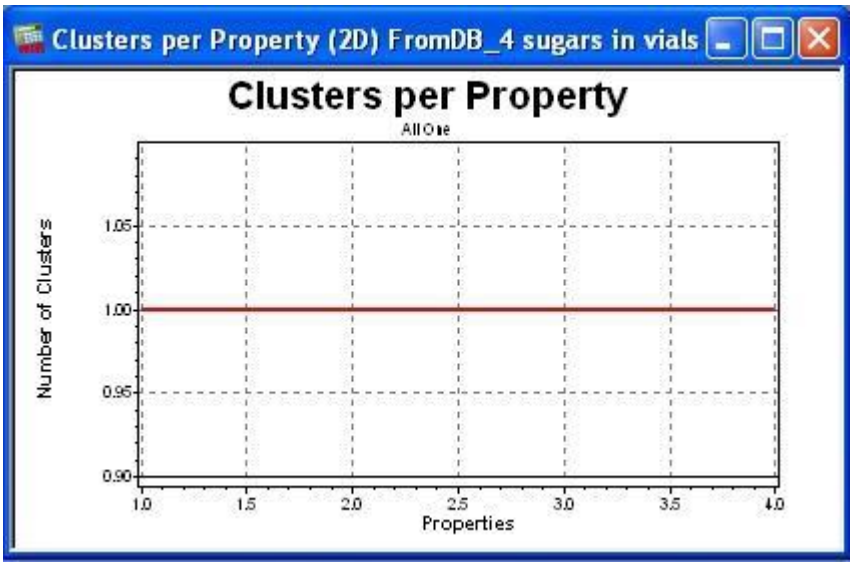
### 3.18.4 Calibration Score Limits

| | |
|---|---|
| Description | Allowed minimum and maximum score values of the calibration. |
| Use | Score outlier detection in the validation set, especially in Cluster calibration. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 45 |
| Tip | Only scores outliers according the secondary PCs are important. |
| Details | |
| Related Topic | Scores |

The **allowed score limits** for the validation: **1.05 x min. and max. scores of the C-set spectra**.
The default blow up limit of 1.05 can be edited by the user.

### 3.18.5 Cluster per Property

| | |
|---|---|
| Description | Number of cluster -connected radii area- per property. |
| Use | Additional information for property separation. |
| Method | Cluster (CLU) / SIMCA |
| Matrices ID | 52 |
| Tip | **The number of cluster per property is always 1 in a good calibration.** |
| Details | Each spectra of the same property should be together and build one connected cluster. |
| Related Topic | Property Box Radii, Property Box Center |

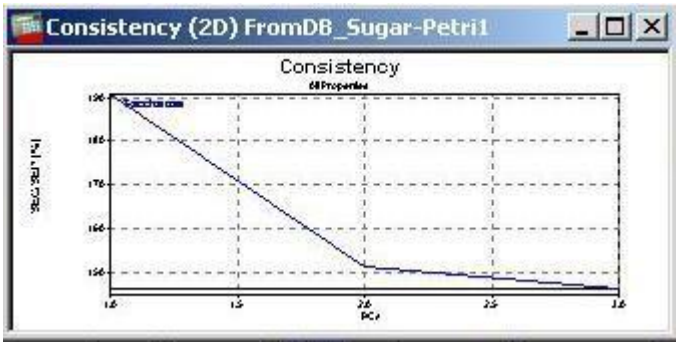



A good calibration has only one cluster for each property

### 3.18.6 Consistency

| | |
|---|---|
| Description | Consistency describes the relation between the standard errors of the calibration and validation sets. |
| Use | Selecting the secondary PCs for a robust calibration. |
| Method | PCR / PLS (only with VS mode) |
| Matrices ID | 33 |
| Tip | SEP- and SEE-value should be small and similar for a stable calibration. With an optimal number of PCs the consistency shows values **around 100** (80-110). |
| Details | Visible in the Overview, in the View: Secondary PCs Selection and under Graphics. |
| Formula | $\text{Consistency} = 100 \times \text{SEE} / \text{SEP}$ |
| Related Topic | C-Set SEE, V-Set SEE (SEP) |

It is possible, that different numbers of PCs are ideal for a certain parameter. In this situation the different secondary PCs should be selected, the calibrations recalculated and compared.
Very low consistency indicates an overfitting (too many secondary PCs used).




### 3.18.7 Correlation Spectra

| | |
|---|---|
| Description | Called also as synchronous 2D correlation spectra. |
| Use | Shows the correlating spectra. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 96 |
| Tip | See Top view or 3D-surface plot. Pretreatments have influence (press F5 after changing pretreatments). |
| Details | Cw = X'X = Transpose(X) * X<br>X = pretreated spectra |
| Related Topic | Correlation Wavelength |

### 3.18.8 Correlation Wavelength

| | |
|---|---|
| Description | Called also as synchronous 2D correlation wavelength. |
| Use | Shows the correlating wavelengths. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 97 |
| Tip | See Top view or 3D-surface plot. Pretreatments have influence (press F5 after changing pretreatments). |
| Details | Cw = XX' = X * Transpose(X)<br>X = pretreated spectra |
| Related Topic | Correlation Spectra |

### 3.18.9 Creator

| | |
|---|---|
| Description | Contains the creator / user index. |
| Use | For 1D-scatter and dependency plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 106 |
| Tip | |
| Details | Creator index is a column vector. |
| Related Topic | |

| | |
|---|---|
| Description | Contains the creator / user index. |
| Use | For 1D-scatter and dependency plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 107 |
| Tip | |
| Details | Creator index is a row vector. |
| Related Topic | |

### 3.18.10 CV C-Set PRESS

| | |
|---|---|
| Description | CV C-Set PRESS is the Predicted Residual Error Sum Square for the samples in case they are in the C-Set of the calibration. |
| Use | To estimate the number of PCs. Compare it with C-Set PRESS of the final calibration. |
| Method | PCR / PLS |
| Matrices ID | 68 |
| Tip | Optimal number of PC, where CV C-Set PRESS has the minimum value. |
| Details | Shows the CV C-Set PRESS in function of the number of PCs. |
| Related Topic | C-Set PRESS |

### 3.18.11 CV Group Index

| | |
|---|---|
| Description | Before CV the spectra can be grouped: the number indicates, to which group a spectrum belongs. |
| Use | To find out, which spectra have e.g. same or similar property values with Property equal or Property segment selection. Using CV each spectra belonging to the same group will be at the same calculation step in V-Set. |
| Method | PCR / PLS |
| Matrices ID | 65 |
| Tip | Can be used for dependency plot or for combined scatter plot. |
| Details | The group number and the spectra belonging to the groups are listed in the CV calibration protocol.<br>Zero as group index means: the spectrum belongs to no any group (unused spectrum: U-Set). |
| Related Topic | Cross Validation Grouping; Custom assign Group to Spectra |

### 3.18.12 CV Leverages

| | |
|---|---|
| Description | Mean of the spectra Leverages during Cross Validation of the V-Set spectra group. |
| Use | To compare with the “Leverages” of the final calibration with all C-Set spectra to find possible outliers. |
| Method | PCR / PLS |
| Matrices ID | 70 |
| Tip | High leverages indicate outliers. |
| Details | For the calculation the primary PCs are taken. |
| Related Topic | Leverages |

### 3.18.13 CV Predicted Property

| | |
|---|---|
| Description | CV Predicted Property is the predicted property value for the sample in case it is left out of the calibration. |
| Use | Compare it with the predicted value of the final calibration and with the original property value. |
| Method | PCR / PLS |
| Matrices ID | 66 |
| Tip | It should be close to the original property value and to the predicted value of the final calibration. |
| Details | The CV predicted property is calculated according the selected secondary PCs. |
| Related Topic | Predicted Property, CV Property Residuum |

### 3.18.14 CV Property Residuum

| | |
|---|---|
| Description | CV Property Residuum is the difference between the original and CV predicted property value. |
| Use | It gives information of outlier sample. |
| Method | PCR / PLS |
| Matrices ID | 67 |
| Tip | It should be as close to zero as possible. |
| Details | Big differences indicate outliers. |
| Related Topic | CV Predicted Property |

### 3.18.15 CV Regression Coefficients[1]

| | |
|---|---|
| Description | It shows as many regression coefficients, as many calculations / groups exist in the CV calibration. Each regression coefficients name has the group name, which was left out. |
| Use | Help for the optimal secondary PCs selection. |
| Method | PCR / PLS |
| Matrices ID | 88 |
| Tip | In case each regression coefficients look similar, the calibration is stable. A group with huge deviation indicates outlier.<br>Select PCs, which have stable regression coefficients (not noisy). |
| Details | [1] means the 1st C-Set property. In general for each application only one C-Set property is allowed, so for each property a separate calibration is necessary (NIRCal could handle more, but NIRWare is concipated with only one property per calibration for quantitative measurements).<br>***NOTE***<br>In the graphics the tooltip shows spectra and number e.g. "spectra (45)", the number corresponds to the CV Group number (ignore the other information). |
| Related Topic | Regression Coefficient, Regression Coefficients [1] |

### 3.18.16 CV Regression Coefficients[1] Mean

| | |
|---|---|
| Description | Mean of CV Regression Coefficients [1] for each wavelength. |
| Use | Used for the CV Regression Coefficients [1] t-test. |
| Method | PCR / PLS |
| Matrices ID | 90 |
| Tip | |
| Details | |
| Related Topic | CV Regression Coefficients[1] t-test |

### 3.18.17 CV Regression Coefficients[1] Sdev

| | |
|---|---|
| Description | Standard Deviation of CV Regression Coefficients [1] for each wavelength. |
| Use | Used for the CV Regression Coefficients[1] t-test. |
| Method | PCR / PLS |
| Matrices ID | 89 |
| Tip | |
| Details | |
| Related Topic | CV Regression Coefficients[1] t-test |

### 3.18.18 CV Regression Coefficients[1] t-test

| | |
|---|---|
| Description | Uncertainty testing in regression model. |
| Use | Used for selecting most relevant and stable wavelength. |
| Method | PCR / PLS |
| Matrices ID | 92 |
| Tip | |
| Details | |
| Related Topic | |

Explanation:

t = Abs( xm ) / sx * sqrt(g)

where
t = CV Regression Coefficients[1] t-Test;
xm = Regression Coefficients[1] (of the final calibration); sx
= CV Regression Coefficients[1] Sdev;
g = number of CV groups.

What are good prediction equation coefficients?

- Where Abs( xm ) is big, there are wavelength with a huge relevance to the prediction equation;
- Where sx is small there are stable wavelengths in the prediction equation.

Due to this 2 points above, consider only at t because it combines the aspect of both, so deselect wavelengths with a small t value.
To get a statistical based limit, look in a t-distribution table, select a alpha i.e. 0.05 and use a degree of freedom of g-1.

Recommended readings:
[Uncertainty testing in PLS regression" A.M.C.Davies, Norwich NIR Consultancy, 75 Intwood Road, Cringleford, Norwich NR4 6AA, UK]
http://www.spectroscopyeurope.com/TD_13_2.pdf

### 3.18.19 CV Regression Coefficients[1] Variance to final Calib

| | |
|---|---|
| Description | Sum of squared differences of the Property Spectrum of the final calibration to each CV CV Regression Coefficients [1] of each group. |
| Use | Used for the CV Regression Coefficients [1] t-test. |
| Method | PCR / PLS |
| Matrices ID | 91 |
| Tip | |
| Details | Sum( Sqr( B - B(g) ) |
| Formular | B8-CV SECV |
| Related Topic | CV Regression Coefficients[1] t-test |

### 3.18.20 CV SECV

| | |
|---|---|
| Description | CV SECV: Standard Error Cross Validated |
| Use | The CV selects the number of secondary PCs, where CV SECV has a minimum. |
| Method | PCR / PLS |
| Matrices ID | 72 |
| Tip | The secondary PCs selection is indicated with red color. |
| Details | The first local minimum of CV SECV is taken. |
| Formula | $SECV = \sqrt{\frac{1}{n-p}\sum (y-Y)^2}$<br>y = reference method value<br>Y = predicted value (of the left out spectra)<br>n = number of batches = **number of C-Set spectra**<br>p = number of coefficients = **number of secondary PCs** |
| Related Topic | SEE |

### 3.18.21 CV Spectra Residuals

| | |
|---|---|
| Description | The spectra residual of the spectra in the V-Set during CV. |
| Use | Compare the CV spectra residual with the residual of final calibration to find out possible outliers. |
| Method | PCR / PLS |
| Matrices ID | 71 |
| Tip | Big CV spectra residual indicates outlier. |
| Details | The number of primary PCs are used for the calculation. |
| Related Topic | Residuals, Primary PCs |

### 3.18.22 CV V-Set PRESS

| | |
|---|---|
| Description | CV V-Set PRESS is the Predicted Residual Error Sum Square for the samples in case they are in the V-Set of the calibration. |
| Use | To estimate the number of PCs. Compare it with C-Set PRESS of the final calibration. |
| Method | PCR / PLS |
| Matrices ID | 69 |
| Tip | Optimal number of PC, where CV V-Set PRESS has the minimum value. |
| Details | Shows the CV C-Set PRESS in function of the number of PCs. |
| Related Topic | C-Set PRESS, CV C-Set PRESS |

### 3.18.23 C-Set BCAP-PRES

| | |
|---|---|
| Description | PRES function of the calibration Set comparable with the DOS based BCAP Software. |
| Use | To estimate the number of factors |
| Method | PCR / PLS |
| Matrices ID | 21 |
| Tip | |
| Details | Shows the C-Set PRES (Predicted Error Sum) in function of the number of PCs. |
| Related Topic | V-Set BCAP Pres |

### 3.18.24 C-Set BIAS

| | |
|---|---|
| Description | **Bias** provides information on the **average deviation** of the predicted values from the true values.<br>BIAS of the calibration set spectra. |
| Use | To compare with the V-Set Bias. |
| Method | PCR / PLS |
| Matrices ID | 17 |
| Tip | The C-Set Bias is **zero** by definition. |
| Details | Available in the calibration protocol. |
| Formula | $Bias = 1/N \bullet \sum (x_n - y_n)$ |
| Related Topic | V-Set BIAS |

### 3.18.25 C-Set Mean Property

| | |
|---|---|
| Description | Average property value of the C-Set spectra. Internally used for meancentering. |
| Use | Compare with the predicted mean property value. |
| Method | MLR / PCR / PLS |
| Matrices ID | 35 |
| Tip | |
| Details | It is available in the calibration protocol (C-Set Orig.mean). |
| Related Topic | Pretreated Property |

### 3.18.26 C-Set Mean Spectrum

| | |
|---|---|
| Description | Mean spectra of the calibration set. |
| Use | Internally used for meancentering. |
| Method | all |
| Matrices ID | 34 |
| Tip | |
| Details | Pretreatments take effect. |
| Formula | $<I(k)> = [I1(k) + I2(k) + \ldots\ldots IN(k)] / N$ |
| Related Topic | Pretreated Spectra |

It is calculated by averaging the intensity values at each wavelength.

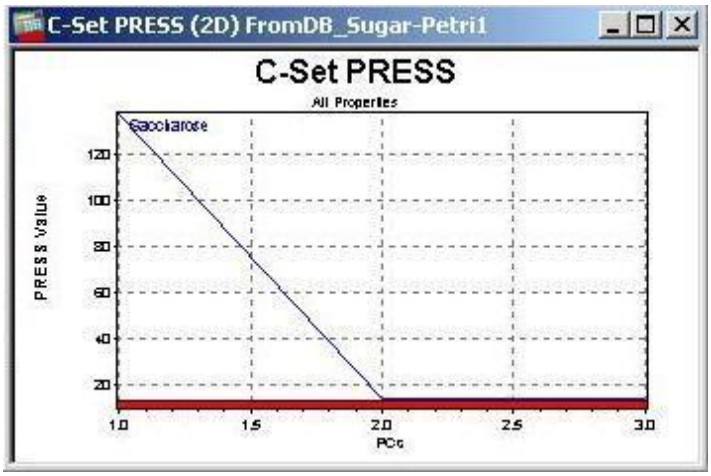
The centre of the new coordinate system is shifted to that point: mean centering.

### 3.18.27 C-Set PRESS

| | |
|---|---|
| Description | Calibration Set: Predicted Residual Error Sum Square. |
| Use | To estimate the number of PCs. Compare it with the V-Set PRESS |
| Method | PCR / PLS |
| Matrices ID | 13 |
| Tip | Graphics / C-Set Statistics / PRESS |
| Details | Shows the calibration set PRESS in function of the number of PCs. |
| Formula | $PRESS = \sum_{n=1}^{N} (x_n - y_n)^2$ |
| Related Topic | V-Set PRESS |

This function calculates the error sum square as a function of the number of PCs. The term „residual“ here refers to the difference between the predicted values and from the original reference values.

The optimum number of PC is always given by the smallest number of PC where the PRESS function for the calibration and for the validation set is approximately **equal and minimal**.
If the error of the prediction diminishes only very slightly by the addition of another PC, it is not worth while to add that PC. This is because higher PCs with little influence will often result in a poorer reproducibility or stability of the calibration.



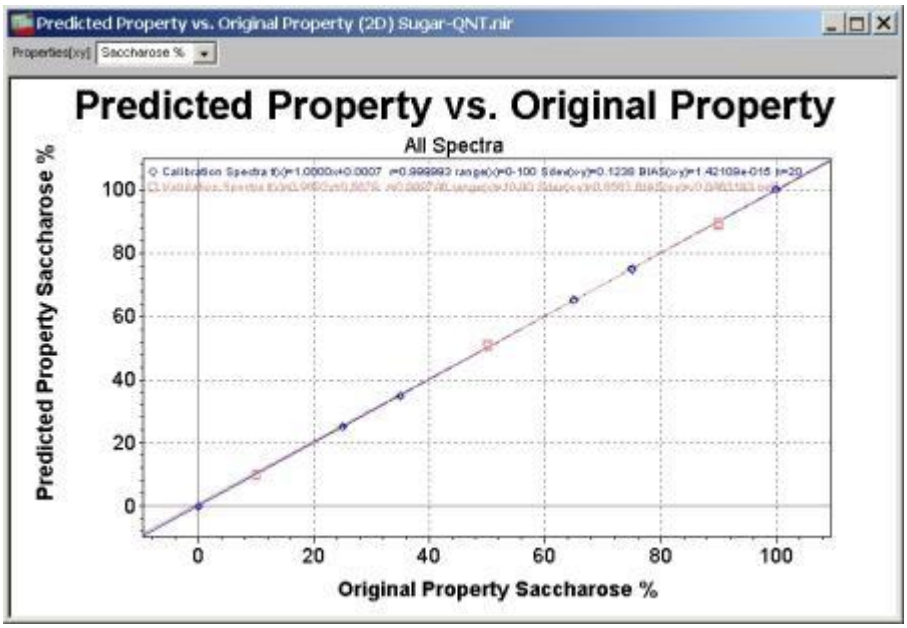

### 3.18.28 C-Set Property Dependencies

| | |
|---|---|
| Description | Regression coefficient between all original C-Set spectra properties. |
| Use | Shows linear dependencies between different properties. |
| Method | MLR / PCR / PLS |
| Matrices ID | 36 |
| Tip | Use a table (grid) to check internal property dependencies |
| Details | Absolute regression coefficient near 1.0 shows that two properties are linearly dependent. Only the property of C-Set Spectra take effect. |
| Related Topic | Original Property |

### 3.18.29 C-Set Regression Coefficient

| | |
|---|---|
| Description | Calibration Set Regression Coefficient of Original Property and Predicted Property (also known as correlation coefficient or Pearson's correlation coefficient) |
| Use | To compare with the V-Set regression Coefficient |
| Method | PCR / PLS |
| Matrices ID | 18 |
| Tip | Should be as close to 1 as possible. |
| Details | Visible on the calibration curve and in the calibration protocol. |
| Formula | $r = \frac{\sum_n (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_n (x_n - \bar{x})^2 \sum_n (y_n - \bar{y})^2}}$ |
| Related Topic | V-Set Regression Coefficient, Original Property, Predicted Property |

The regression coefficient "r" shows how well the predicted values match with the reference values (original property values) on average.
The correlation is rated as acceptable when r > 0.9 is achieved (the error of the conventional reference method goes into the NIR-calibration via the reference values).




The regression curve can be described better with the slope (a) and intercept (b): f(x)= ax + b

### 3.18.30 C-Set Regression Intercept

| | |
|---|---|
| Description | Intercept of the regression line equation of the calibration set. |
| Use | To compare with the V-Set Regression Intercept |
| Method | PCR / PLS |
| Matrices ID | 19 |
| Tip | A value around **zero** is expected for a good calibration. |
| Details | A better description of the C-Set regression coefficient. |
| Related Topic | C-Set Regression Coefficient, C-Set Regression Slope |

### 3.18.31 C-Set Regression Slope

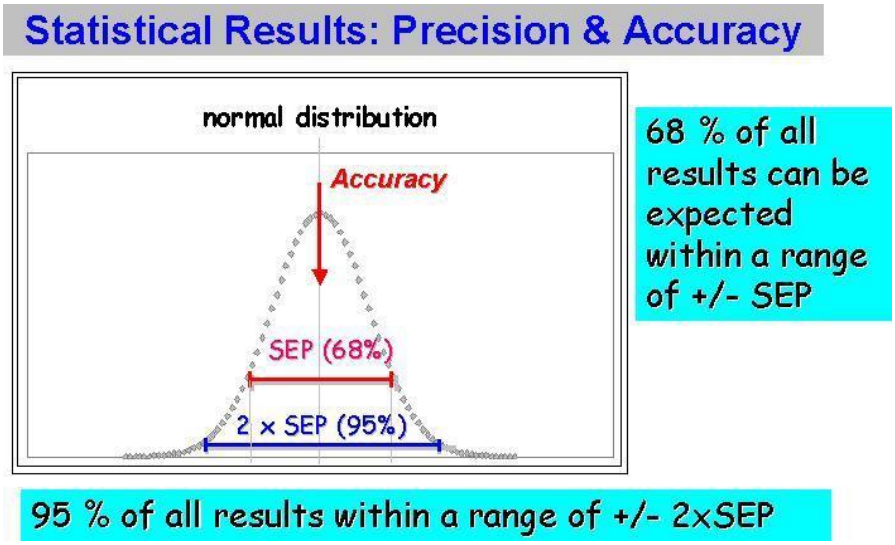
| | |
|---|---|
| Description | Slope of the regression line equation of the calibration set. |
| Use | To compare with the V-Set Regression Slope |
| Method | PCR / PLS |
| Matrices ID | 20 |
| Tip | A value around **one** is expected for a good calibration. |
| Details | A better description of the C-Set regression coefficient. |
| Related Topic | C-Set Regression Coefficient, C-Set Regression Intercept |

### 3.18.32 C-Set SEE (SEC)

| | |
|---|---|
| Description | Standard Error of Estimation: Standard Deviation of the Property Residuum of the C-Set Spectra. |
| Use | To compare with the V-Set SEE (SEP) |
| Method | PCR / PLS |
| Matrices ID | 16 |
| Tip | It should be as small as possible, but comparable with the standard deviation of the lab method. |
| Details | Precision. Available in the calibration protocol. |
| Formula | $SEE = (1/N - 1\sum (x_n - y_n - BIAS)^2)^{1/2}$ |
| Related Topic | V-Set SEE (SEP), Property Residuum |

The SEC and SEP provide the magnitude of the **standard deviation** for the calibration set and the independent validation set. The two values should be **as small as possible**, but they are likely to be comparable with the standard deviation of the conventional laboratory method. With an acceptable calibration, the two values are also roughly equal (Consistency: around 100).

Statistically the expected error is with a probability of 68 % within an interval of ± SEC and with a probability of 95 % within an interval of ± 2 SEC.




### 3.18.33 C-Set Spectra Residuum Limits

| | |
|---|---|
| Description | The min.-max. limit of the C-Set residuum spectra. |
| Use | Judgement of the number of primary PCs and disturbing (not fitted) wavelength range. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 53 |
| Tip | Pretreatments have influence: visible on the Y-axis. |
| Details | Depends only on the number of primary PCs. |
| Related Topic | Spectra Residuum, C-Set Spectra Residuum Max, C-Set Spectra Residuum Min |

### 3.18.34 C-Set Spectra Residuum Max

| | |
|---|---|
| Description | Shows the maximun residuum of the C-Set spectra using different number of PCs in the selected wavelength range. |
| Use | To find the optimum number of primary PCs. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 75 |
| Tip | Small residuum indicates a good reconstruction, but avoid overfitting. |
| Details | The residuum with e.g. PCs 3 shows the residuum spectra using 1-3 PCs for the reconstruction. |
| Related Topic | Spectra Residuum, Primary PCs |

### 3.18.35 C-Set Spectra Residuum Min

| | |
|---|---|
| Description | Shows the minimun residuum of the C-Set spectra using different number of PCs in the selected wavelength range. |
| Use | Find the optimum number of primary PCs. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 76 |
| Tip | Small residuum indicates a good reconstruction, but avoid overfitting. |
| Details | The residuum with e.g. PCs 3 shows the residuum spectra using 1-3 PCs for the reconstruction. |
| Related Topic | Spectra Residuum, Primary PCs |

### 3.18.36 C-Set Spectra Residuum MinMax

| | |
|---|---|
| Description | Shows the maximal residual of the C-Set spectra in dependency of the number of PCs. |
| Use | To find the optimum number of primary PCs. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 93 |
| Tip | Small residuals indicate a good reconstruction, but avoid overfitting. |
| Details | |
| Related Topic | Spectra Residuum, Primary PCs |

### 3.18.37 C-Set Std Orig Property

| | |
|---|---|
| Description | Standard deviation of the original property values of the calibration set spectra (standard deviation of the whole calibration range, X-axis). |
| Use | To compare with the C-Set Std Pred Property |
| Method | PCR / PLS |
| Matrices ID | 14 |
| Tip | Gives information about the size of the original calibration range. |
| Details | It is available in the calibration protocol: C-Set Orig.sdev. |
| Related Topic | C-Set Std Pred Property Original Property |

### 3.18.38 C-Set Std Pred Property

| | |
|---|---|
| Description | Standard deviation of the predicted property values of the calibration set spectra (standard deviation of the whole predicted calibration range, Y-axis). |
| Use | To compare with the C-Set Std orig Property |
| Method | PCR / PLS |
| Matrices ID | 15 |
| Tip | Gives information about the size of the predicted calibration range. |
| Details | It is available in the calibration protocol: C-Set Pred.sdev. |
| Related Topic | C-Set Std Orig Property Predicted Property |

### 3.18.39 C-Set X-PRESS

| | |
|---|---|
| Description | Predicted Residual Sum Squares of the C-Set spectra. |
| Use | See the effect of the number of PCs for reconstructing the C-Set spectra. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 73 |
| Tip | Help for selection of the number of primary PCs. |
| Details | The smallest number of PC, that is still showing changes, should be selected. Compare it with V-Set PRESS. |
| Related Topic | X-PRESS, V-Set X-PRESS |

### 3.18.40 Durbin-Watson Factor Statistics

| | |
|---|---|
| Description | Statistical test for the determination of the linearity. |
| Use | Addition information for statistics like regression slope and intercept. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 62 |
| Tip | The C-Set and V-Set selection has a big influence on the Durbin-Watson value. |
| Details | dw = Durbin-Watson = Sum of consecutive/successive Residual Difference Square / Sum of Residual Square |
| Related Topic | Calibration Protocol Quantitative |

**Definition**: "The Durbin-Watson test is a test for first-order serial correlation in the residuals of a time series regression. A value of 2.0 for the Durbin-Watson statistic indicates that there is no serial correlation. This result is biased toward the finding that there is no serial correlation if lagged values of the regressors are in the regression.
Formally, the statistic is: d=(sum from t=2 to t=T of: (et-et-1)2/(sum from t=1 to t=T of: et2) where the series of "et" are the residuals from a regression.

Use Durbin-Watson test to assess correlation between adjacent observations.

$$dw = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} (e_i)^2}$$

e = Property residual = (Original property - Predicted property)

The Durbin-Watson test checks for sequential dependence in which each error (and also residual) is correlated with those before and after it in the sequence."

**Interpretation**:

- ranges from 0 (perfect positive correlation) to 4 (perfect negative correlation);
- values from 1.5 (= du) to 2.5 (= 4-du) indicate no serious violation of independence ( for n > 30 )

0_____1___1.5__**2**__2.5___3_____4

**Results of dw:**

- 0 <= dw <= 4 always;
- the distribution of dw is symmetric about 2;
- if successive residuals are positively serially correlated, that is positively correlated in their sequence, dw will be near 0;
- if successive residuals are negatively serially correlated, that is negatively correlated in their sequence, dw will be near 4, so that (4 - dw) will be near 0.

Abbreviation:
dl = dw lower limit;
du = dw upper limit;
k = 1, considering 2 dimensional plots like regression plot;
Residuals from a fittet straigth line Y = b0 + b1* X;
n = number of C-Set resp. V-Set spectra;
alpha = 5% significance, typical;

Limits for dl and du after k, n, alpha are to find in reference:
[Savin, N.E. and White, K.J., "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors", Econometrica, Vol. 45, 1977, pp. 1989-1996.]

Order : "The observations and residuals have a natural order."

The spectra are internal for the Durbin-Watson calculation ascending sorted according the original property values of each properties. By the same property values the original sequence will be kept.

C-Set / V-Set : the dw will be calculated separately. If the V-Set is empty dw = 2 protocoled.

Durbin-Watson test is available in NIRCal:

- the calibration protocol of quantitative calibration has the dw value for C-Set and V-Set;
- the dw values are in the matrices.

### 3.18.41 Eigenvalue

| | |
|---|---|
| Description | Corresponding Eigenvalues for PCs and scores. |
| Use | To select the number of primary PCs. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 5 |
| Tip | The eigenvalue matrix is in diagonal form. |
| Details | High eigenvalue describes high spectral influence. |
| Related Topic | Scores,Primary PCs,Loadings |

Eigenvalue gives the information about the importance of the PCs: high value means high information, small value indicates less information.

Explanation of mathematical correlation of the loadings and eigenvalue using Mahalanobis distance of the PCA (PCR, CLU, SIMCA):

$$FF' = \lambda$$

$$|F_i| = \sqrt{\lambda_i}$$

$$LL' = I$$

$$|L_i| = 1$$

where:

- F factor / loadings matrix
- F` transposed factor / loadings matrix;
- $\lambda$ eigenvalue
- L scores matrix;
- L` transposed score matrix;
- I identity matrix (diagonal matrix with 1).

The equation of the spectra reconstruction using this terminology:

$$X = \vec{1}' \bar{\vec{X}} + L' F + R$$

This is equivalent to the equation given by the **Loadings (links)** explanation.
The results are different using Mahalanobis or Euclidean distance:

| | **Mahalanobis** | **Euclidean** |
|---|---|---|
| FF' | Diagonal matrix with eigenvalue | Diagonal matrix with 1 |
| LL' | Diagonal matrix with 1 | Diagonal matrix with eigenvalue |

The calculation of eigenvalue is different by the PLS method: the eigenvalue are calculated under consideration of the property value variance. Therefore iterative algorithms, like nipals, are used.

### 3.18.42 Instrument

| | |
|---|---|
| Description | Contains the used instrument index. |
| Use | For 1D-scatter and dependency plots. Can be used for Outlier selection. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 104 |
| Tip | |
| Details | Instrument index is a column vector. |
| Related Topic | |

| | |
|---|---|
| Description | Contains the used instrument index. |
| Use | For 1D-scatter and dependency plots. Can be used for Outlier selection. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 105 |
| Tip | |
| Details | Instrument index is a row vector. |
| Related Topic | |

### 3.18.43 Leverages

| | |
|---|---|
| Description | Mahalanobis distance from the center of the score space to each spectra. |
| Use | To find Leverage Outliers. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 32 |
| Tip | |
| Details | All primary PCs are used. The secondary PC selection has no effect on the leverages. |
| Related Topic | Scores |

| | |
|---|---|
| Description | Contains the Leverages against the PCs for the different properties (Cluster). |
| Use | For PC selection. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 94 |
| Tip | |
| Details | |
| Related Topic | Leverages |

### 3.18.44 Loadings or Principal Components

| | |
|---|---|
| Description | Loadings build up the base for reconstructing the spectra together with the scores and the Eigenvalues. |
| Use | Look for the wavelengths activity in the loadings and compare it with the original spectra. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 4 |
| Tip | Mark interesting bands in the wavelengths and have a look to the spectra. The wavelengths selection will also appear in the plot. |
| Details | **Loadings** are called also **Principal Components** or sometimes till called "**factors**" (NIRCal 4 terminology). The maximum available number of PCs calculated is called the **primary PCs**.<br>The selected PCs for predicting properties are called **secondary PCs**. |
| Related Topic | Scores, Eigenvalues |

The loadings are artificial difference spectra. Only principal components with characteristic spectral information should be used.

The formula of spectra reconstruction:

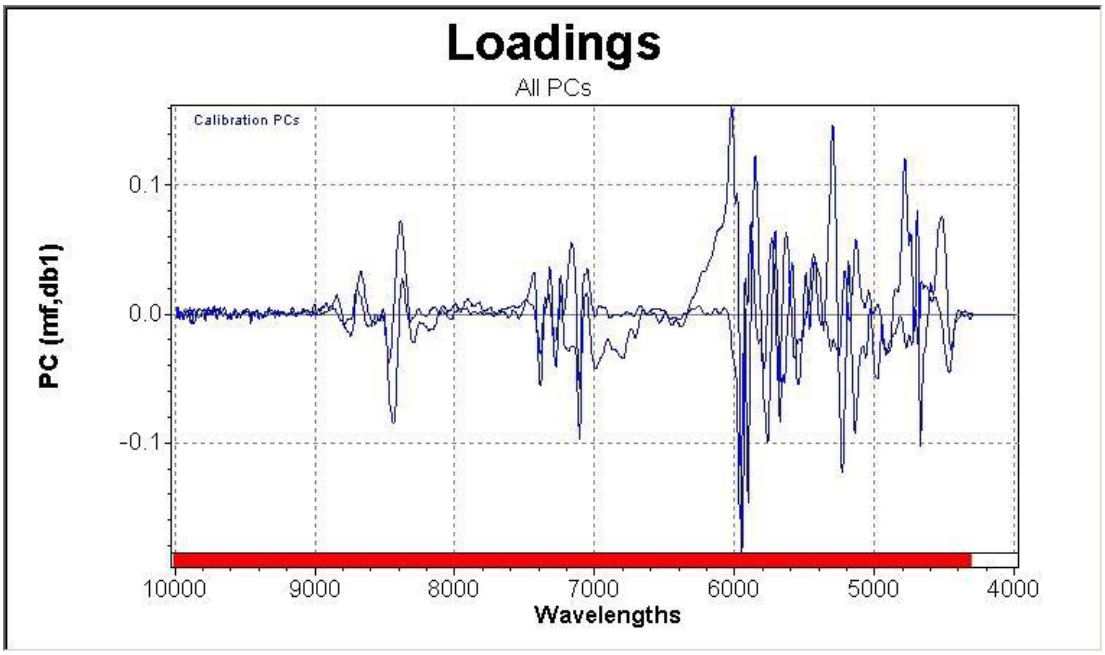
$$I_n(k) = \langle I(k) \rangle + \sum_{i=1}^{\max i} v_{in} \bullet U_i(k) + R_n(k)$$

where:

- I n (k) : spectrum ( n: number of spectra, k: wavenumber);
- <I (k)> : mean spectrum;
- v in : scores (i: number of PC);
- U i (k) : PCs / loadings;
- R n : residuum spectrum.

A more detailed formula for the matrix is: Eigenvalue.
Minimum number of loading is one, the maximum number of PCs is never higher than the number of C-Set spectra or C-Set Wavelengths.

See the PCs under the menubar: **Graphics / Loadings**.




### 3.18.45 Original Property

| | |
|---|---|
| Description | Original property values of all spectra in the project. (Dependent variable) |
| Use | Defines the concentrations contained or classification of the spectra. |
| Method | all |
| Matrices ID | 1 |
| Tip | Editing is possible in NIRWare: Sample Management. |
| Details | Table / Properties / Original |
| Related Topic | Pretreated Property, Predicted Property |

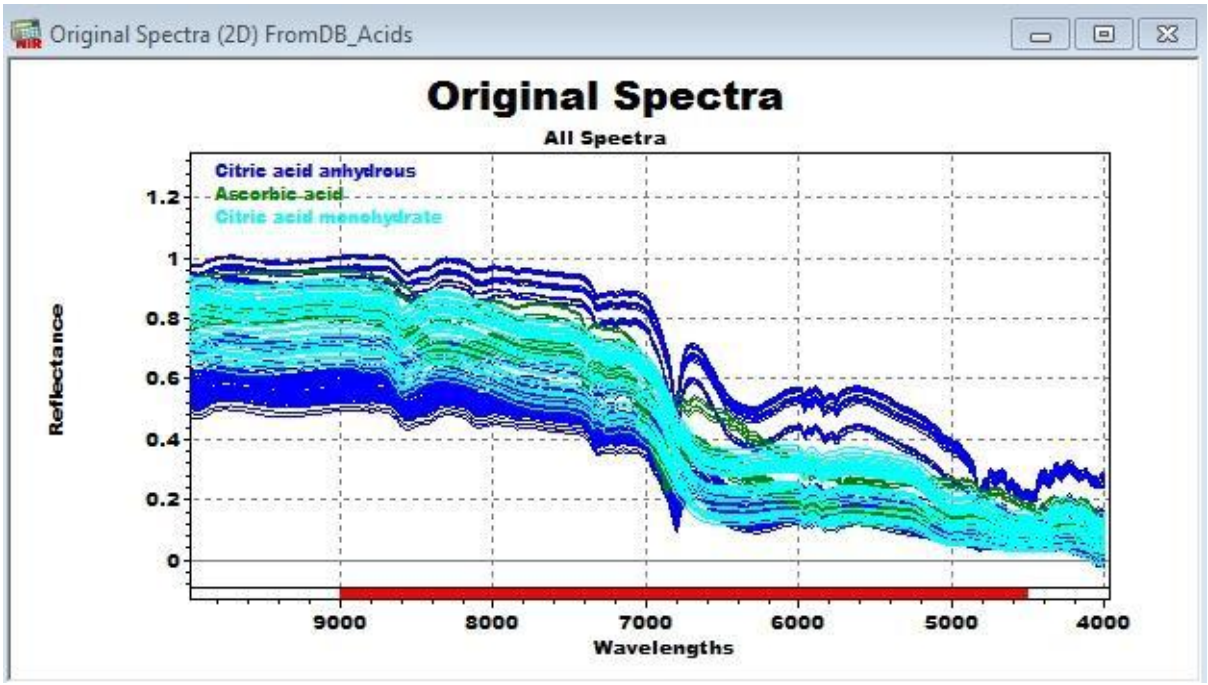
In cluster method each spectra is assigned to exactly one property by setting a 1 (one) to the class, where it belongs, all other properties must have a value of 0 (zero).
In quantitative calibrations the original property values are equal to the results of the laboratory methods referred to.

### 3.18.46 Original Spectra

| | |
|---|---|
| Description | Original Spectra in the project (Independent variable) |
| Use | Overview of all spectra in the project. |
| Method | all |
| Matrices ID | 0 |
| Related Topic | Pretreated Spectra |

Original spectra are plotted as the intensity values (reflectance, transmittance, transflectance) were measured according the wavelengths.

Available in the Menubar: **Graphics / Spectra / Original**.




The values are also available in table form.

### 3.18.47 PCs or Factor

| | |
|---|---|
| Description | Contains the PC index number. |
| Use | For 1D-scatter plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 86 |
| Tip | |
| Details | PC is a column vector. |
| Related Topic | |

| | |
|---|---|
| Description | Contains the PC index number. |
| Use | For 1D-scatter plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 87 |
| Tip | |
| Details | PC is a row vector. |
| Related Topic | |

### 3.18.48 PCR B-Matrix

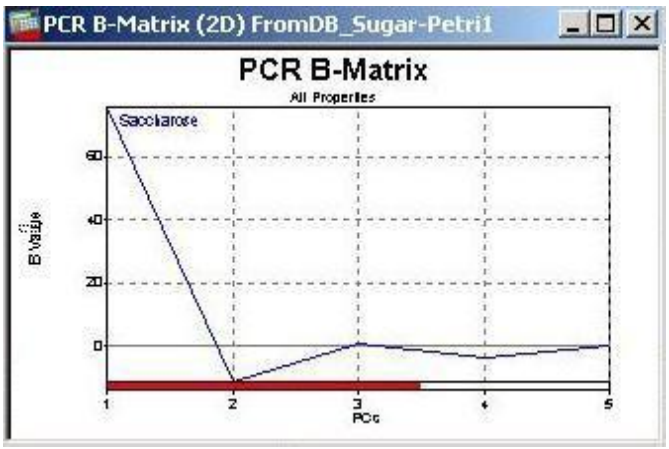
| | |
|---|---|
| Description | Shows the influence of the PCs to the property value. |
| Use | Useful for **secondary PC selection**. |
| Method | PCR |
| Matrices ID | 7 |
| Tip | Select the PCs with **high absolute correlation value** to optimize the prediction. |
| Details | Also known as **correlation coefficient** or **B-matrix**.<br>Depends only from the number of primary factors. The selected secondary PCs take no effect on the PCR B-matrix. |
| Related Topic | Scores, Original Property |

The PCR algorithm makes as first step a Principal Component Analysis and the second step is a Multiple Linear Regression. In the MLR the scores are multiplied with the correlation coefficients: B-values.

Formula of MLR by PCR:

**Property value = ymean + b1 * v1 + b2 * v2 + b3 * v3 +.....**

where: b : correlation coefficient; v : score; 1-2: number of PC.




Here the 3. PC has very small B-value, so only the first 2 PCs are important for the parameter.

***NOTE***
The B-values are **not normalized between -1 and + 1** as usual for correlation coefficients, because the scores are already normalized using the Mahalanobis distance.

### 3.18.49 Predicted Property

| | |
|---|---|
| Description | The predicted property values for all spectra in the project with the activated calibration. (Estimate of dependent variable) |
| Use | Main result of a calibration. |
| Method | all |
| Matrices ID | 10 |
| Tip | The original and predicted property values should be as similar as possible. |
| Details | The prediction depends on the selected secondary PCs. |
| Related Topic | Property Residuum, Original Property |

In cluster method the spectra get a 1 (one) to a class, if it is identified as such (distance and residual limits are fulfilled).
In quantitative calibrations the predicted property values are the results of the NIR methods.

### 3.18.50 Pretreated Property

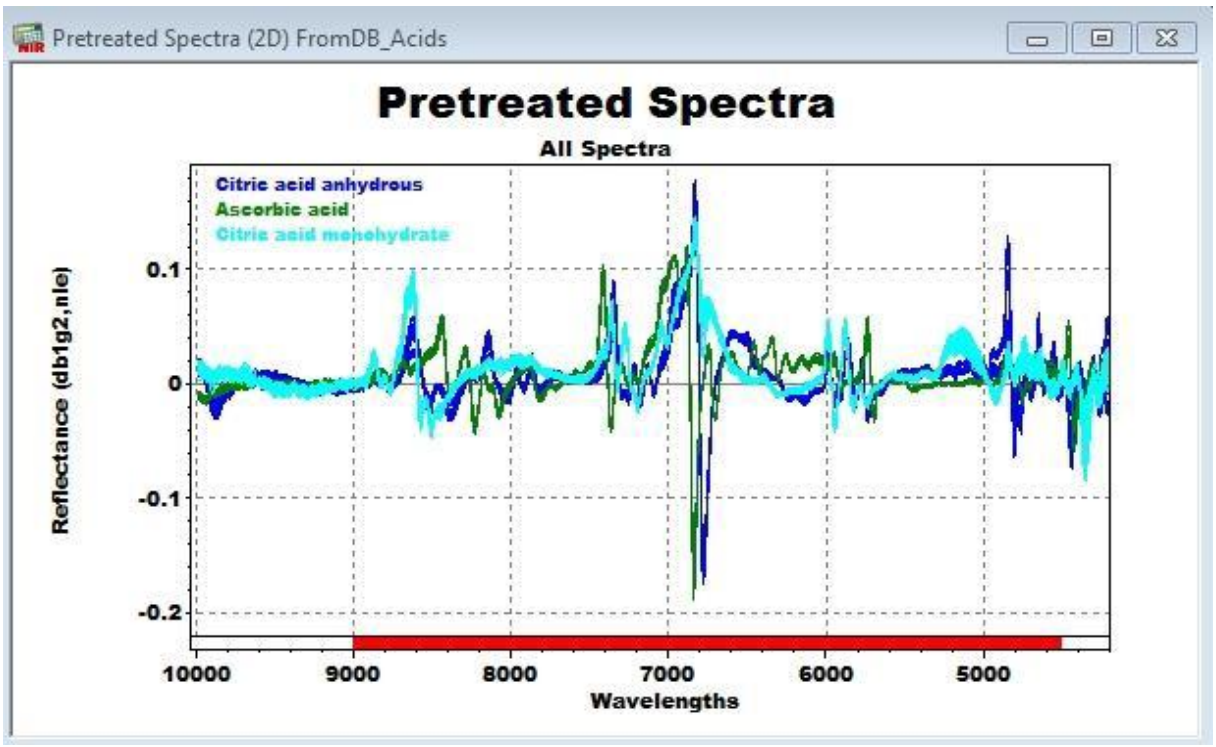
| | |
|---|---|
| Description | A copy of Original Property |
| Method | all |
| Matrices ID | 3 |
| Tip | Editing is not allowed. |
| Details | In a later version there will be special property pretreatments. |
| Related Topic | Original Property |

### 3.18.51 Pretreated Spectra

| | |
|---|---|
| Description | Original spectra with the selected data pretreatments applied. |
| Use | Looks at the effects of the various data pretreatments. |
| Method | all |
| Matrices ID | 2 |
| Tip | Try zooming to see minor effects |
| Details | Without any Pretreatments the Pretreated Spectra are identical to the Original Spectra. Each Pretreatment works directly on the Pretreated Spectra as input and the result are the Pretreated Spectra again. In this way, it is possible to build up a sequence with Pretreatments. |
| Related Topic | Original Spectra, Pretreatments |

Pretreated spectra are plotted as the intensity values after the pretreatment or pretreatment sequence according to the wavelengths.

Available in the menubar: **Graphics / Spectra / Pretreated**.




The applied pretreatments are visible on the intensity axis in different plots related to the spectra intensity (e.g. Spectra Residuum). The used calibration wavelength is marked red.

### 3.18.52 Property

| | |
|---|---|
| Description | Contains the property index. |
| Use | For 1D-scatter plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 84 |
| Tip | |
| Details | Property is a column vector. |
| Related Topic | |

| | |
|---|---|
| Description | Contains the property index. |
| Use | For 1D-scatter plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 85 |
| Tip | |
| Details | Property is a row vector. |
| Related Topic | |

| | |
|---|---|
| Description | Contains the first C-Set Property[1] Values |
| Use | For 1D-scatter plots and dependency plots |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 102 |
| Tip | |
| Details | Property values are column vectors. |
| Related Topic | Residuals, Primary PCs, Blow Up Limits |

| | |
|---|---|
| Description | Contains the first C-Set Property[1] Values |
| Use | For 1D-scatter plots and dependency plots |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 103 |
| Tip | |
| Details | Property values are row vectors. |
| Related Topic | Residuals, Primary PCs, Blow Up Limits |

### 3.18.53 Property Adjacency

| | |
|---|---|
| Description | The minimum Mahalanobis distance in score space between the nearest two C-Set spectra of property A to property B, C, etc.. |
| Use | See neighborhood relationships between the properties for all PC`s.<br>The first nearest distance is reported in the calibration protocol under "Property Separation / Distance". |
| Method | Cluster (CLU) |
| Matrices ID | 49 |
| Tip | Look at it in a 2D-Plot in top view to see hidden effects. |
| Details | Secondary PCs take an effect. Calculation is based on C-Set spectra and C-Set properties. |
| Related Topic | Score Adjacency |

### 3.18.54 Property Box Center

| | |
|---|---|
| Description | The center position (expressed as score) of the smallest possible box in the score space that contains each C-Set spectra of a property. |
| Use | Shows the distribution of the properties in the score space. |
| Method | Cluster (CLU) / SIMCA |
| Matrices ID | 46 |
| Tip | In higher PC dimensions the property box center falls together. These higher PCs are useless for property separation, so remove them from the secondary PC selection. See also the Property Box Radii matrix. |
| Details | Only C-Set spectra and C-Set property are used for calculations. |
| Related Topic | Property Box Radii |

### 3.18.55 Property Box Center Uniformity

| | |
|---|---|
| Description | Uniformity of the distribution of the Property Box Centers. |
| Use | Additional information for PCs selection. |
| Method | Cluster (CLU) / SIMCA |
| Matrices ID | 51 |
| Tip | |
| Details | Estimates the place consumption of all properties per PC.<br>If several properties lay near each other in a PC, the PC becomes a low uniformity value. A uniformity value of 1.0 is ideal. |
| Related Topic | Property Box Radii, Property Box Center, Scores |

### 3.18.56 Property Box Radii

| | |
|---|---|
| Description | The distance from the Property Box Center to each wall of the smallest property box. |
| Use | Shows the score distribution of the properties versus PCs. One of the best matrix to select the secondary PCs for the cluster method. |
| Method | Cluster (CLU) / SIMCA |
| Matrices ID | 47 |
| Tip | Small values (normally bellow 0.1) indicate good reproducibility. To get good separation of the properties select these PCs for the secondary PCs. |
| Details | |
| Related Topic | Property Box Center |

### 3.18.57 Property F-Test backward

| | |
|---|---|
| Description | Significance of PCs starting at the minimum of the SEP function. |
| Use | For secondary principal component selection. |
| Method | PCR / PLS |
| Matrices ID | 37 |
| Tip | |
| Details | Starting at the PCs they set the SEP value to the minimum, take PCs away as long as no significant change for the worse of the SEP value occurs. This process is called backward search operation. |
| Related Topic | Property F-Test forward |

### 3.18.58 Property F-Test forward

| | |
|---|---|
| Description | Significance of additional PCs. |
| Use | For secondary principal component selection. |
| Method | PCR / PLS |
| Matrices ID | 38 |
| Tip | |
| Details | Alternative for proofing the robustness of a calibration with a backward search operation is to start at a certain calibration checking for additional PCs that will improve the SEP value significantly. |
| Related Topic | Property F-Test backward |

### 3.18.59 Property Interference

| | |
|---|---|
| Description | Interference = Property Score Extension / Property Nearest Neighbor |
| Use | Judgement of the influence of the property for the cluster separation. |
| Method | Cluster (CLU) |
| Matrices ID | 61 |
| Tip | Small interference indicates a good influence of the property for the calibration. |
| Details | Property Interference is the relationship between the size of a cluster and distance to the nearest property.<br>The interference values are listed in the calibration protocol. |
| Related Topic | Property Score Extension, Property Nearest Neighbor |

### 3.18.60 Property missing Values

| | |
|---|---|
| Description | Spectra, which have no defined property value (0 = value missing; 1 = value available)<br>***NOTE***<br>This has no relevance in NIRCal file-mode, only in database-mode. |
| Use | Find spectra, which have no property value |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 115 |
| Tip | Edit the property values in Manager Console/ Sample Management |
| Details | Spectra without property values can not be used for the C- and V-Set. |
| Related Topic | Original Property |

Interpretation:

- 1 means: there is defined property value,
- 0 means: missing property value = there is no defined property value.

### 3.18.61 Property Nearest Neighbor

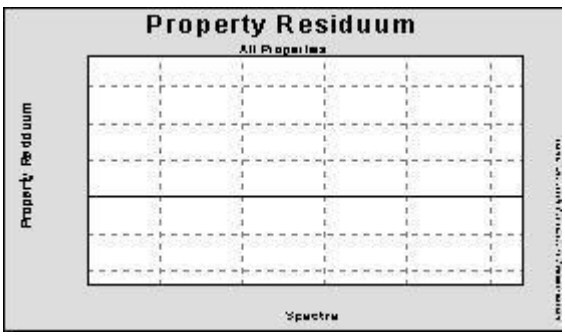
| | |
|---|---|
| Description | The smallest Mahalanobis distance between the property mean to the next property center. |
| Use | Judge the property separation. |
| Method | Cluster (CLU) |
| Matrices ID | 60 |
| Tip | High values indicate good separation. |
| Details | These values together with the next neighbors are listed in the calibration protocol. |
| Related Topic | Scores |

### 3.18.62 Property Residuum

| | |
|---|---|
| Description | The difference between the original property and the predicted property. |
| Use | See where (property, spectra) the major differences are. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 11 |
| Tip | |
| Details | |
| Related Topic | Original Property,Predicted Property |

**Qualitative calibrations**
The predicted property value depends on the cluster, in which the spectrum is located. For the property residuum, only a value of 0 is acceptable. Property residuum zero means all spectra are in the correct cluster.
Property residuum +1 means that a spectrum is outside any cluster: it is not identified.
Property residuum -1 means that a spectrum is in a wrong cluster: it is false identified.

**Quantitative calibrations**
The values should be as small as possible. Property residuum zero means, the predicted NIR value is exactly the same, as the referring lab result. An acceptable range is equal to the standard deviation of the lab method.




### 3.18.63 Property Score Diffusion

| | |
|---|---|
| Description | Property Score Diffusion is the scaled disorder, number of different property spectra according the scores of different PC per property. The scaling is made by the number of spectra of a property. |
| Use | Useful for secondary PCs selection in higher dimensional space (> 3). |
| Method | Cluster (CLU) |
| Matrices ID | 58 |
| Tip | It is not depending on the size of the score scattering, only the number of interchanging of the properties are important. Shows, which property is mixed and shuffeled with other property spectra. This plot is very informative in Top View. Dark means low diffusion, white means high diffusion. Select the PCs, by which more than one dark area (property) appears. Two dark areas means: these two properties are separated with that PC. Positioning the mouse over the area, the property name is visible in the Status bar. |
| Details | Small disorder values indicate a clear separation for the property of a given PC.<br>Property Score Diffusion is calculated only for the C-Set spectra. |
| Related Topic | Scores,Secondary PCs, Property Score Disorder |

Algorithm (unpublished) by Roman Bossart 18. June 1998

### 3.18.64 Property Score Disorder

| | |
|---|---|
| Description | Property Score Disorder: number of different property spectra according the scores of different PC. |
| Use | Useful for secondary PCs selection in higher dimensional space (> 3). |
| Method | Cluster (CLU) |
| Matrices ID | 57 |
| Tip | The smallest number of property score disorder = (number of selected property -1). This value indicates a clear separation of the different properties. |
| Details | Small disorder values indicate a clear separation (how many times the color is changing with the score of a PC). Property Score Disorder is calculated only for the C-Set spectra. |
| Related Topic | Scores, Secondary PCs, Property Score Diffusion |

**Disorder = Sum (Property Score Diffusion) * (number of C-Set spectra) / (number of property)**

### 3.18.65 Property Score Extension

| | |
|---|---|
| Description | Property Score Extension: the maximal distance between two C-Set spectra of the same property. |
| Use | Useful for secondary PCs selection in higher dimensional space (> 3). |
| Method | Cluster (CLU) / SIMCA |
| Matrices ID | 59 |
| Tip | Small values indicates good reproducibility and repeatability. PCs with small extension are prefered for secondary PC selection. |
| Details | Property Score Extension is calculated only for the C-Set spectra. The extension values are listed in the calibration protocol. |
| Related Topic | Scores, Property Score Disorder, Property Score Diffusion |

### 3.18.66 Property Wavelength Regression

| | |
|---|---|
| Description | Contains all regression coefficients between each wavelength of all the calibration set spectra and the property value. |
| Use | In a 2D-plot very useful to optimize the wavelength selection for all properties. Especially useful for quantitative methods. |
| Method | all but not SIMCA |
| Matrices ID | 44 |
| Tip | Absolute values near 1.0 shows strong wavelength dependencies with the corresponding property. |
| Details | Dependencies:<br>- pretreatments<br>- C-Set spectra<br>- Calibration property |
| Related Topic | Property Spectra |

### 3.18.67 Quant Calib. Statistics

| | |
|---|---|
| Description | Contains all statistics of the quantitative calibration protocol. |
| Use | Use it as a fast or short protocol. |
| Method | MLR / PCR / PLS |
| Matrices ID | 42 |
| Tip | Open this matrix as Table (Grid). |
| Details | Summary of the relevant statistics matrices. |
| Related Topic | |

### 3.18.68 Q-Values

| | |
|---|---|
| Description | Value between 0 and 1. 0 means a bad 1 would be a good calibration. But 1 cannot be reached.<br>Q-Values depending on the number of PCs.<br>There is a warning limit for the stored calibration:<br>0.8 for qualitative<br>0.6 for quantitative |
| Use | Select the number of secondary PC according the highest Q-Value. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 64 |
| Tip | In generell: the highest the Q-Value the better the calibration. |
| Details | The Q-Value is in the calibration protocol, see explanation: LINK: Calibration Protocol Qualitative<br>It is used for selection of the best calibrations for the automatic Calibration wizard. |
| Related Topic | Quant Calib. Statistics, Cluster per Property, Spectra Residuals too big |

Explanation of the Q-Value calculation of quantitative calibration:

| Value | Weights | Formula for value | General term | Aspects of a good calibration | Detail |
|---|---|---|---|---|---|
| wNum NotC | 10 | Number of C-Set spectra with Residual too big | Rejection of known | Rejection count of known should be 0 | |
| wNum NotV | 1 | Number of V-Set spectra with Residual too big | Rejection of unknown | Rejection count of unknown should be 0 | |
| wQa | 2 | Abs (SEE-SEP)/ (Abs(SEP)+1.0) | Relative Consistency | C-Set and V-Set should have similar low standard error of prediction | BIAS corrected |
| wQb | 2 | Abs (VsetBIAS)/ Abs(Range) | Weighted BIAS | The absolute V-Set BIAS should be low | |
| wQc | 1 | 1-VsetRegr | Validity | The V-Set regression coefficient should be near 1 | |
| wQd | 1 | Abs( CsetRegr-VsetRegr) | Comparability | C-Set and V-Set should have similar high regression | |
| wQe | 1 | SEP / Abs(Range) | Precision | The V-Set standard error of prediction should be low | BIAS corrected |
| wQf | 1 | Abs(RSS)/Abs(Range) | Weighted Accuracy | The V-Set Residual Sum of squares= RSS should be low | not BIAS corrected |

**Q-Value V5:**

**Q = 1 / ( 1 + sum (weights * value))**

***NOTE***

The weights are changed between the NIRCal 4.21 Q-Value calculation weights, so the NIRCal 5 Q-Values are differing against NIRCal 4.21.

The Q-Value calculation is **not changed for the Cluster** calibration between NIRCal 4.21 and NIRCal 5.

Explanation of the Q-Value calculation of qualitative calibration:

| General term | Aspects of a good calibration | Weights |
|---|---|---|
| C-Set false identified | Should be zero (indicates wrong user settings) | 10 |
| C-Set not identified | Should be zero (indicates wrong user settings) | 10 |
| V-Set false identified | Should be zero (indicates wrong sample or calibration) | 5 |
| V-Set not identified | Should be zero (residual or distance problem) | 1 |
| Cluster Index "Number of Clusters" | "Number of C-Set Properties": should be zero | 1 |
| Property Uniformity | Should be small: the spectra spreading in the clusters is uniform | 1 |
| Property Interference | Should be small: the clusters are independent from each other | 0.1 |

$$Q-Value = \frac{1}{1+\sum_{i=1}^{n} w_i v_i}$$

w : weights
v : value
n : number of aspects

The Q-Value is normally high for CV and SIMCA calibrations, while a quantitative calibration with Cross Validation has empty V-Set (Weighted Accuracy = 0), a qualitative calibration using SIMCA has no neighboring substance (single property calibration: property interference = 0).

### 3.18.69 Real time

| | |
|---|---|
| Description | Contains the real time of the spectra creation time in seconds. |
| Use | For 1D-scatter and dependency plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 98 |
| Tip | |
| Details | Real time is a column vector. |
| Related Topic | |
| Description | Contains the real time of the spectra creation time in seconds. |
| Use | For 1D-scatter and dependency plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 99 |
| Tip | |
| Details | Real time is a row vector. |
| Related Topic | |

### 3.18.70 Reconstructed Spectra

| | |
|---|---|
| Description | Reconstructed pretreated spectra with the number of primary PCs. |
| Use | See how good the reconstruction works. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 8 |
| Tip | It should be as similar to the pretreated spectra as possible. |
| Details | Reconstructed spectra are plotted as the intensity values after the used pretreatments according the wavelengths after the reconstruction of the PCA using the primary PCs. |
| Related Topic | Pretreated Spectra Spectra Residuum |

Available in the menubar: **Graphics / Spectra / Reconstructed**.

A reconstructed spectrum is the sum of [mean spectrum and the sum of PCs multiplied with the scores values].

### 3.18.71 Residuals

| | |
|---|---|
| Description | Contains the residual value against the PCs for the different spectra. |
| Use | For selection of the primary PCs. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 95 |
| Tip | Select the smallest number of PCs, where the residuals are similar small and do not change anymore. |
| Details | Residual is the sum of squares of the residuum spectrum. Residuum spectrum = pretreated spectrum - reconstructed spectrum. |
| Related Topic | Spectra Residuum, X-PRESS |

### 3.18.72 Regression Coefficients [1]

| | |
|---|---|
| Description | It shows the regression coefficients in function of the PCs. Each regression coefficients name has the name of the PC. |
| Use | Help for the optimal secondary PCs selection. |
| Method | PCR / PLS |
| Matrices ID | 79 |
| Tip | Select PCs, which have stable regression coefficients (not noisy). |
| Details | [1] means the 1st C-Set property. In general for each quantitative application only one C-Set property is allowed, so for each property a separate calibration is necessary (NIRCal could handle more in file-mode, but NIRWare is conceived with only one property per calibration for quantitative measurements). |
| Related Topic | Regression Coefficient, CV Regression Coefficients[1] |

### 3.18.73 Regression Coefficients / Property Spectra (CLU)

| | |
|---|---|
| Description | Spectra of the property. This matrices is called Regression Coefficient for quantitative calibration and Property Spectra for Cluster calibrations. |
| Use | Useful for wavelength selection. |
| Method | PCR / PLS / Cluster (CLU) |
| Matrices ID | 31 |
| Tip | Select wavelengths with high absolute values. |
| Details | **Cluster** method calculates the mean spectra of **all C-Set spectra with the same property**. |
| Related Topic | Loadings, Property Wavelength Regression |

**In quantitative calibration the Regression Coefficients is used for building up the prediction equation:**

Y = (X-X mean) * P + Y mean

Where:

| | |
|---|---|
| Y | predicted property value |
| X | pretreated spectrum |
| X mean | Mean of C-Set spectra ("C-Set Mean Spectrum") |
| Y mean | Mean Property value of the C-Set spectra ("C-Set Mean Property") |
| P | Regression Coefficients |

The Regression Coefficients can be opened as Excel table in the Regression Coefficients plot pressing "G".
The value of "Y mean" can be found in the Matrices: "C-Set Mean Property".

### 3.18.74 Sample

| | |
|---|---|
| Description | Contains the sample index number. |
| Use | For 1D-scatter plots. Can be used for Outlier selection. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 113 |
| Tip | Spectra from the same sample get the same sample index. |
| Details | Sample is a column vector. |
| Related Topic | |

| | |
|---|---|
| Description | Contains the sample index number. |
| Use | For 1D-scatter plots. Can be used for Outlier selection. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 114 |
| Tip | Spectra from the same sample get the same sample index. |
| Details | Sample is a row vector. |
| Related Topic | |

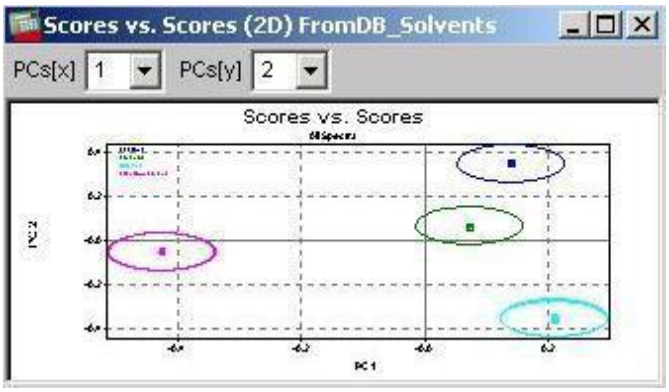
### 3.18.75 Scores

| | |
|---|---|
| Description | Each spectra is placed in the n-dimensional score space. The position of the spectra is given by the n-dimensional coordinate of the scores. |
| Use | Similar spectra are placed near each other. Look for clustering effects and for Outliers. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 6 |
| Tip | Looks very nice in a 3D scatter plot or 2D scatter plots. |
| Details | Vin: score of the i PC and n spectrum |
| Formula | Reconstruction of a spectrum $In(k) = <I(k)> + \sum vin \bullet Ui\,(k) + Rn$ |
| Related Topic | Loadings |

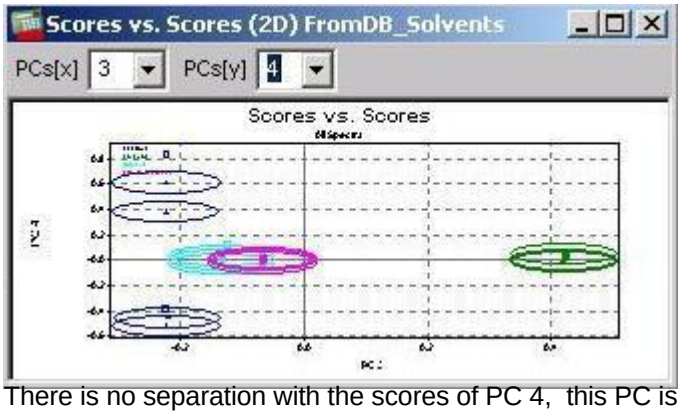
Scores are the weightings of each PC after the pretreated spectrum has been transformed by PCA. A score is the portion of a PC used for the spectra reconstruction. Each spectrum has different scores for each primary factors.

The scores are visible in a 2 or 3 dimensional scatter plot.

In qualitative calibration, the separate scores determine the number of secondary PCs.



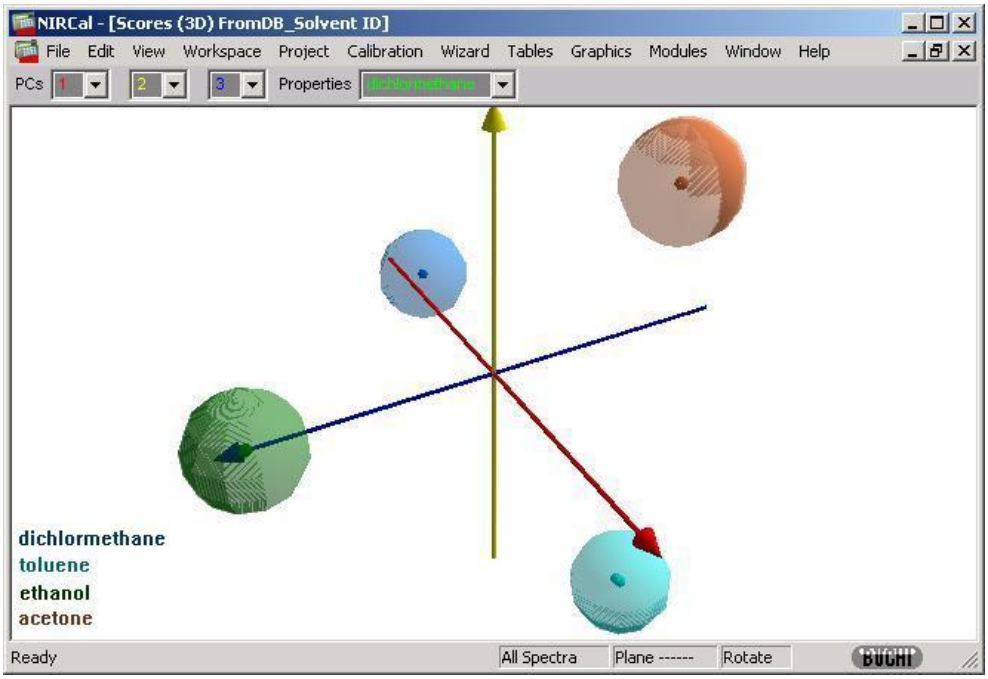

There is a clear separation of the 4 different properties spectra according to the PC 1 and 2.




There is no separation with the scores of PC 4,  this PC is necessary for the acetone spectra.

Here PC 1, 2 and 3 are secondary PCs, the PC 4 is not a secondary PC.

The separation with 3 selected calibration PCs is visible in a 3 D graphics.



### 3.18.76 Score Adjacency

| | |
|---|---|
| Description | Distance in score space between each spectra against all others. |
| Use | The neighborhood relationships between the spectra for all PCs at once.<br>For special interest with the Cluster method. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 40 |
| Tip | Look at it in a 2D-Plot in top view to see hidden effects. |
| Details | Secondary PCs take an effect. |
| Related Topic | Scores, Leverage |

### 3.18.77 SEP Generalized Cross Validation

| | |
|---|---|
| Description | The determination of the SEP (Standard Error of Prediction) with a Cross Validation (CV) is very time consuming.<br>The SEP value determined with the GCV (Generalized Cross Validation) process is from theoretically side at least equally good as the SEP of a conventional CV process. |
| Use | Select the number of PC, where the SEP Generalized Cross Validation has a minimum. |
| Method | MLR / PCR / PLS |
| Matrices ID | 39 |
| Tip | While SEP Generalized Cross Validation does not always have a minimum, it is not always optimal to use this for the secondary PCs selection. |
| Details | See also: Gene H. Golub, Michael Heath, and grace Wahba. Generalized crossvalidation as a method for choosing a good ridge parameter. Technometrics, 21(2):215-223,1979. |
| Related Topic | V-Set SEE (SEP) |

Formula:

$$SEP_{GCV} = \frac{\sqrt{n * CSetPRESS(a)}}{n - a}$$

where:

- n : number of calibration spectra
- a : number of secondary PCs

### 3.18.78 Spectra

| | |
|---|---|
| Description | Contains the spectra index number. |
| Use | For 1D-scatter plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 80 |
| Tip | |
| Details | Spectra is a column vector. |
| Related Topic | |

| | |
|---|---|
| Description | Contains the spectra index number. |
| Use | For 1D-scatter plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 81 |
| Tip | |
| Details | Spectra is a row vector. |
| Related Topic | |

### 3.18.79 Spectra Property Index

| | |
|---|---|
| Description | Each spectra is assigned to exactly that property index with the 1 (one) in the Original Property matrix.<br>The property index value is in the range from 0 to the number of properties minus 1. |
| Use | Use it to select all spectra with the same property (see Tip). |
| Method | Cluster (CLU) / SIMCA / PCR / PLS / MLR |
| Matrices ID | 48 |
| Tip | Open it as 2D-plot,<br>- PopUp Menu -> Options -> Show All Values,<br>- Set in PopUp Menu -> Transpose Data.<br>- Set in PopUp Menu -> Option -> Mouse Window Select.<br>Select with the mouse the number 0.<br>Then all spectra with the first property are selected. Save this configurated plot in a workspace i.e. as "Select spectra by property Plot". |
| Details | The following special cases are defined by codes (negative numbers)<br>-1 : the spectrum is assigned to more than one property<br>-2 : the spectrum is not assigned to a property (unknown identity)<br>-3 : the property values of the spectrum seem to be for quantitative use and not for qualitative (cluster). |
| Related Topic | Original Property |

### 3.18.80 Spectra Radii

| | |
|---|---|
| Description | Each C-Set spectra is covered by a sphere. Its center, the spectra scores and one radius for all secondary PCs dimensions define the sphere. |
| Use | The spheres define the space where the well known spectra with the defined property can be. |
| Method | Cluster (CLU) / SIMCA |
| Matrices ID | 50 |
| Tip | Good views are:<br>- Graphics -> Scores -> 3D-Scatter<br>- Graphics -> Scores -> Multi 2D-Scatter<br>The spectra Radii depend on the **secondary PCs selection**. |
| Details | There are 3 Radii calculation formula and the Radii Blow Up limit can be also changed by the user. |
| Related Topic | Property Box Radii, Property Box Center |

The Radii Blow Up limit and the Radius calculation formula (1 and 2: for Cluster, 3: only for SIMCA) can be changed under:

Calibration -> Parameter -> Blow Up Limit -> Radii and Radius formula.

All radii can be blowed up by a constant scale. The scale can be in the range from near zero to 1000 (zero makes no sense!).

See: Cluster (CLU)

***NOTE***
Changing the calculation formula and blow up parameter can modify the calibration in a dramatic way. An optimum between robustness and selectivity of the calibration should be found.

### 3.18.81 Spectra Residuals

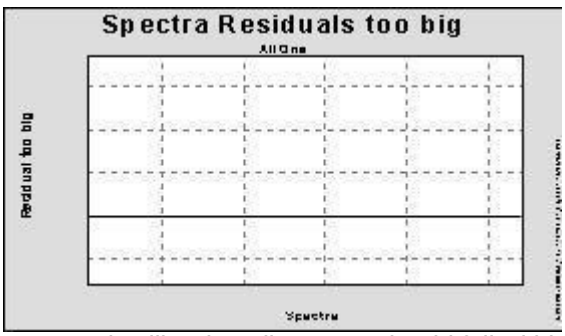
| | |
|---|---|
| Description | The RMSD (Root Mean Squared Deviation) of the spectra residuum. |
| Use | To look at residual outlier. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 41 |
| Tip | Special tip to optimize the calibration |
| Details | The secondary PCs take no effect. |
| Formula | $In(k) = <I(k)> + \sum vin \bullet Ui\ (k) + Rn$ |
| Related Topic | X-PRESS, Spectra Residuum |

Indicates the difference of the pretreated spectra and the reconstructed spectra with the number of primary PCs.

For the application the **allowed residual** is in most cases: **2x C-Set maximum residual**. The default blow up limit of 2 can be adjusted by the user.

### 3.18.82 Spectra Residuals too big

| | |
|---|---|
| Description | Number of spectra with too big residual. |
| Use | Additional information for primary PCs selection or for outlier search. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 54 |
| Tip | **The number of spectra with residual too big should be 0.** |
| Details | The allowed residual is = 2* max. C-Set residual. Using this calculation only V-Set spectra can be out of the limit.<br>2 is the default of residual blow up, can be edited by the user. |
| Related Topic | Residuals, Primary PCs |



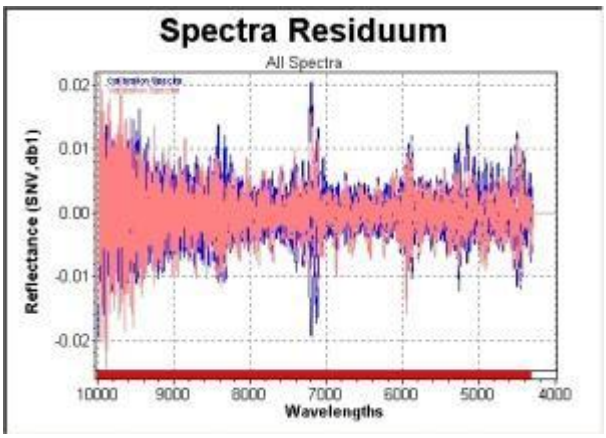

In a good calibration all spectra should fall within the allowed residual limits, so it should be zero in this window.

### 3.18.83 Spectra Residuum

| | |
|---|---|
| Description | Difference spectra between the pretreated spectra and the reconstructed spectra. |
| Use | See where (wavelength, spectra) the major differences are. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 9 |
| Tip | To see that the C-Set spectra are not overfitted open this plot twice. In the first set the visibility to the C-Set spectra, in the other set the visibility to the V-Set spectra. The two sets should be similar. |
| Details | |
| Related Topic | Loadings, Pretreated Spectra, X-PRESS, Reconstructed Spectra |

The spectra residuum shows the difference spectrum for all pretreated spectra and the corresponding reconstructed spectra calculated with the number of primary principal components. The size of these difference spectra should be roughly the same for C- and V-Set, when the correct number of primary

principal components has been selected. These difference spectra should have only noisy signal, but no unfitted peaks.




### 3.18.84 Spectrum Nearest Neighbor

| | |
|---|---|
| Description | Spectrum Nearest Neighbor is the smallest distance in the score plots between two spectra. |
| Use | Check, which spectra have big deviation. Similar small distances indicate good repeatability and reproducibility. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 63 |
| Tip | The Spectrum Nearest Neighbor vs. Spectra scatter plot can help to reduce redundancy with deleting some spectra with very small distance or to find outliers with selecting spectra with high distance. |
| Details | The distance is calculated according the selected secondary PCs. |
| Related Topic | Scores |
| Description | Spectrum Nearest Neighbor is the smallest distance in the score plots between two spectra (distance vs. PC). |
| Use | Check, which spectra have big deviation. Similar small distances indicate good repeatability and reproducibility. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 126 |
| Tip | |
| Details | The distance is calculated according the selected secondary PCs. |
| Related Topic | Scores |

### 3.18.85 Spectrum Nearest Neighbor Index

| | |
|---|---|
| Description | Spectrum Nearest Neighbor Index contains the spectrum index of the nearest spectrum in the calibration. Best shown as 1D-Scatter-Plot for selected spectra. This matrix is related to "Spectrum Nearest Neighbor" which contains the Mahalanobis Distance between nearest spectra pairs. |
| Use | Check, which spectra have big deviation. Similar small distances indicate good repeatability and reproducibility. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 124 |
| Tip | The Spectrum Nearest Neighbor Index vs. Spectra scatter plot can help to reduce redundancy with deleting some spectra with very small distance or to find outliers with selecting spectra with high distance. |
| Details | The distance is calculated according the selected secondary PCs. |
| Related Topic | Scores |
| Description | Spectrum Nearest Neighbor Index contains the spectrum index of the nearest spectrum in the calibration. Best shown as 1D-Scatter-Plot for selected spectra. This matrix is related to "Spectrum Nearest Neighbor" which contains the Mahalanobis Distance between nearest spectra pairs (spectra vs. PC). |
| Use | Check, which spectra have big deviation. Similar small distances indicate good repeatability and reproducibility. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 125 |
| Tip | |
| Details | The distance is calculated according the selected secondary PCs. |
| Related Topic | Scores |

### 3.18.86 Time

| | |
|---|---|
| Description | Contains the spectra creation time as an increasing index |
| Use | Dependency plots, 1D scatter plots |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 100 |
| Tip | |
| Details | |
| Related Topic | Residuals, Primary PCs, Blow Up Limits |

| | |
|---|---|
| Description | Contains the spectra creation time as an increasing index |
| Use | Dependency plots, 1D scatter plots |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 101 |
| Tip | |
| Details | |
| Related Topic | Residuals, Primary PCs, Blow Up Limits |

### 3.18.87 V-Set BCAP-PRES

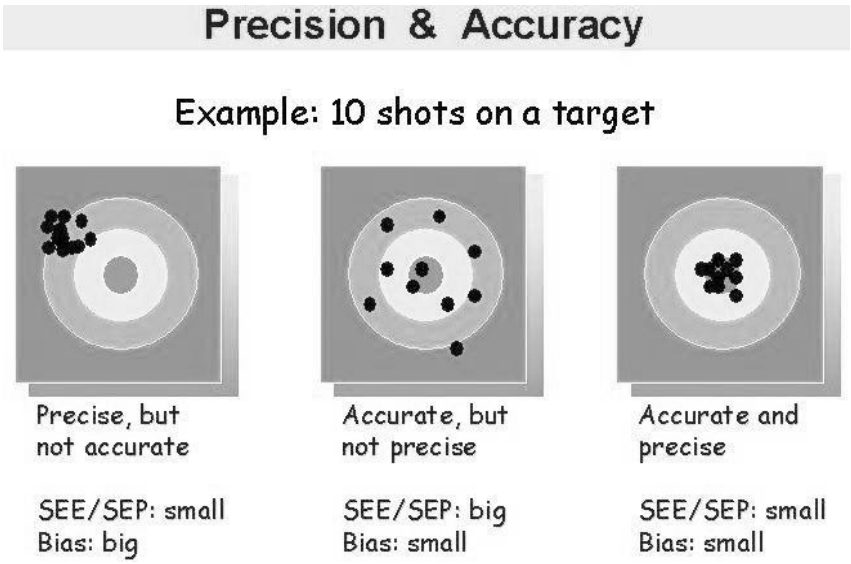
| | |
|---|---|
| Description | PRES function of Validation Set comparable with the BCAP software. |
| Use | To estimate the number of PCs. |
| Method | PCR / PLS |
| Matrices ID | 30 |
| Tip | |
| Details | Shows the V-Set PRES (Predicted Error Sum) in function of the number of PCs. |
| Related Topic | C-Set BCAP PRES |

### 3.18.88 V-Set BIAS

| | |
|---|---|
| Description | BIAS of the validation set spectra. |
| Use | Judge the V-Set results. |
| Method | PCR / PLS |
| Matrices ID | 26 |
| Tip | Should be as close to zero as possible. |
| Details | This is a value for accuracy. Available in the calibration protocol. |
| Formula | $Bias = 1/N \bullet \Sigma (x_n - y_n)$ |
| Related Topic | C-Set BIAS |

The **V-Set Bias** provides information on the **average deviation** of the predicted values from the true values. This value gives information on a systematic deviation of the calibration and therefore should be **as close to zero as possible**.

The C-Set Bias is zero by definition.

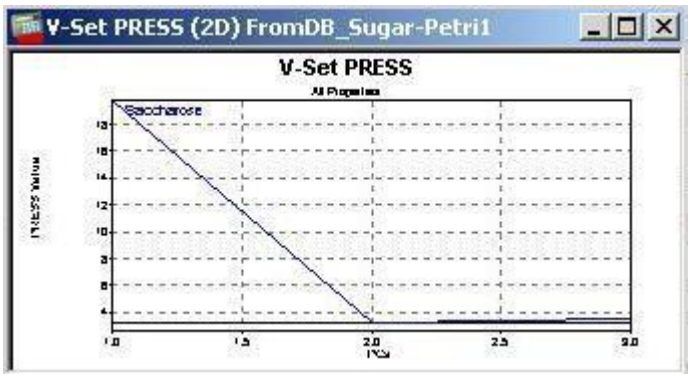



### 3.18.89 V-Set PRESS

| | |
|---|---|
| Description | Validation Set: Predicted Residual Error Sum Square |
| Use | To estimate the number of PCs. Compare it with the C-Set PRESS. |
| Method | PCR / PLS |
| Matrices ID | 22 |
| Tip | Graphics / V-Set Statistics / PRESS |
| Details | Shows the validation set PRESS in function of the number of PCs. |
| Formula | $PRESS = \sum_{n=1}^{N} (x_n - y_n)^2$ |
| Related Topic | C-Set PRESS |

This function calculates the error sum square as a function of the number of PCs. The term „residual“ here refers to the difference between the predicted values and the reference values.

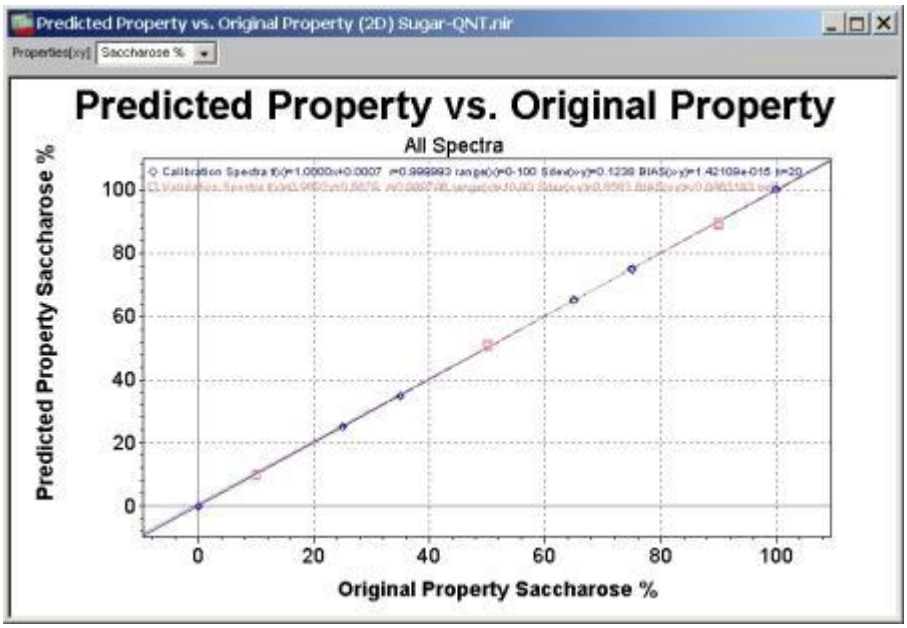
The optimum number of PC is always given by the smallest number of PC where the PRESS function for the calibration and for the validation set is approximately **equal and minimal**.
If the error of the prediction diminishes only very slightly by the addition of another PC, it is not worth while to add that PC. This is because higher PCs with little influence will often result in a poorer reproducibility or stability of the calibration.




### 3.18.90 V-Set Regression Coefficient

| | |
|---|---|
| Description | Validation Set Regression Coefficient of Original Property and Predicted Property (also known as correlation coefficient or Pearson's correlation coefficient) |
| Use | To compare with the C-Set regression Coefficient |
| Method | PCR / PLS |
| Matrices ID | 27 |
| Tip | Should be as close to 1 as possible. |
| Details | Visible on the calibration curve and in the calibration protocol. |
| Formula | $r = \frac{\sum_n (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_n (x_n - \bar{x})^2 \sum_n (y_n - \bar{y})^2}}$ |
| Related Topic | C-Set Regression Coefficient, Original Property, Predicted Property |

The regression coefficient "r" shows, how well the predicted values match with the reference values (original property values) on average.
The correlation is rated as acceptable when r > 0.9 is achieved (the error of the conventional reference method goes into the NIR-calibration via the reference values).




The regression curve can be described better with the slope (a) and intercept (b): f(x)= ax + b

### 3.18.91 V-Set Regression Intercept

| | |
|---|---|
| Description | Validation Set Regression Intercept of Original Property and Predicted Property |
| Use | To compare with the C-Set Regression Intercept |
| Method | PCR / PLS |
| Matrices ID | 28 |
| Tip | A value around **zero** is expected for a good calibration. |
| Details | A better description of the V-Set regression coefficient. |
| Related Topic | V-Set Regression Coefficient, V-Set Regression Slope |

### 3.18.92 V-Set Regression Slope

| | |
|---|---|
| Description | Slope of the regression line equation of the validation set. |
| Use | To compare with the C-Set Regression Slope |
| Method | PCR / PLS |
| Matrices ID | 29 |
| Tip | A value around **one** is expected for a good calibration. |
| Details | A better description of the V-Set regression coefficient. |
| Related Topic | V-Set Regression Coefficient, V-Set Regression Intercept |

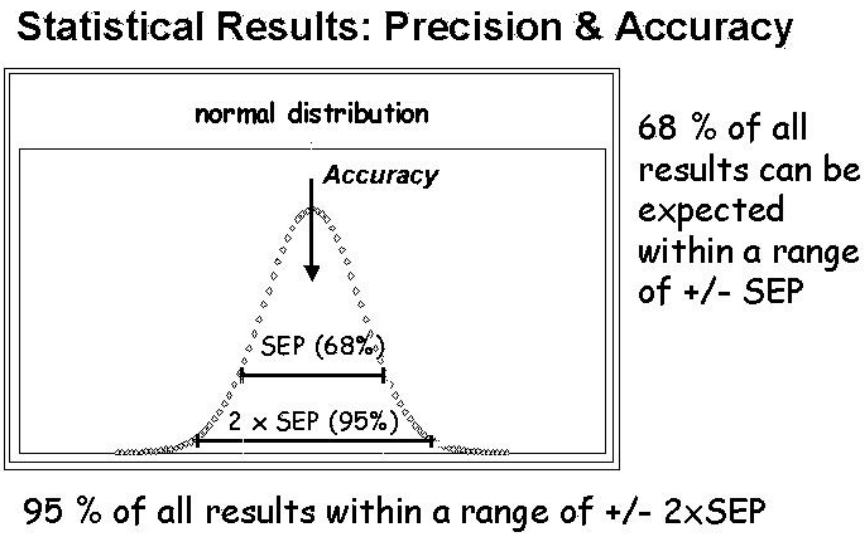
### 3.18.93 V-Set SEE (SEP)

| | |
|---|---|
| Description | Standard Error of Prediction: Standard Deviation of the Property Residuum of the V-Set Spectra. |
| Use | To compare with the C-Set SEE (SEC) |
| Method | PCR / PLS (only with VS mode) |
| Matrices ID | 25 |
| Tip | It should be as small as possible, but comparable with the standard deviation of the lab method. |
| Details | Precision. Available in the calibration protocol. |
| Formula | $SEP = (1/N - 1 \sum (x_n - y_n - BIAS)^2)^{1/2}$ |
| Related Topic | C-Set SEE (SEP) |

The SEE and SEP provide the magnitude of the **standard deviation** for the calibration set and the independent validation set. The two values should be **as small as possible**, but they are likely to be comparable with the standard deviation of the conventional laboratory method. With an acceptable calibration, the two values are also roughly equal (Consistency: around 100).

Statistically the expected error is with a probability of 68 % within an interval of ± SEP and with a probability of 95 % within an interval of ± 2 SEP.

If the SEP is significantly higher than the SEE, the calibration is overfitted (too many secondary PCs are selected).




### 3.18.94 V-Set Spectra Residuum Max

| | |
|---|---|
| Description | Shows the maximum residuum of the V-Set spectra using different number of PCs in the selected wavelength range. |
| Use | To find the optimum number of primary PCs. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 77 |
| Tip | Small residuum indicates a good reconstruction, but avoid overfitting. |
| Details | The residuum with e.g. PCs 3 shows the residuum spectra using 1-3 PCs for the reconstruction. |
| Related Topic | Spectra Residuum, Primary PCs |

### 3.18.95 V-Set Spectra Residuum Min

| | |
|---|---|
| Description | Shows the minimum residuum of the V-Set spectra using different number of PCs in the selected wavelength range. |
| Use | To find the optimum number of primary PCs. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 78 |
| Tip | Small residuum indicates a good reconstruction, but avoid overfitting. |
| Details | The residuum with e.g. PCs 3 shows the residuum spectra using 1-3 PCs for the reconstruction. |
| Related Topic | Spectra Residuum, Primary PCs |

### 3.18.96 V-Set Std Orig Property

| | |
|---|---|
| Description | Standard deviation of the original property values of the validation set spectra (standard deviation of the whole validation range, X-axis) |
| Use | To compare with the V-Set Std Pred Property |
| Method | PCR / PLS |
| Matrices ID | 23 |
| Tip | Gives information about the size of the original validation range. |
| Details | It is available in the calibration protocol (V-Set Orig.sdev). |
| Related Topic | V-Set Std Pred Property |

### 3.18.97 V-Set Std Pred Property

| | |
|---|---|
| Description | Standard deviation of the predicted property values of the validation set spectra (standard deviation of the whole predicted validation range, Y-axis). |
| Use | To compare with the V-Set Std Orig Property |
| Method | PCR / PLS |
| Matrices ID | 24 |
| Tip | Gives information about the size of the predicted validation range. |
| Details | It is available in the calibration protocol (V-Set Pred.sdev). |
| Related Topic | V-Set Std Orig Property |

### 3.18.98 V-Set X-PRESS

| | |
|---|---|
| Description | Predicted Residual Sum Squares of the V-Set spectra. |
| Use | See the effect of the number of PCs for reconstructing the V-Set spectra. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 74 |
| Tip | Help for selection of the number of primary PCs. |
| Details | The smallest number of PC, that is still showing changes, should be selected. Compare it with C-Set PRESS. |
| Related Topic | X-PRESS, C-Set X-PRESS |

### 3.18.99 Wavenumber

| | |
|---|---|
| Description | Contains the wavenumber index. |
| Use | For 1D-scatter plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 82 |
| Tip | |
| Details | Wavenumber is a column vector. |
| Related Topic | |

| | |
|---|---|
| Description | Contains the wavenumber index. |
| Use | For 1D-scatter plots. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA / MLR |
| Matrices ID | 83 |
| Tip | |
| Details | Wavenumber is a row vector. |
| Related Topic | |

### 3.18.100 X-PRESS

| | |
|---|---|
| Description | Predicted Residual Sum Squared of all Spectra over all PCs. |
| Use | See the effect of the numbers of PCs for reconstructing the spectra. |
| Method | PCR / PLS / Cluster (CLU) / SIMCA |
| Matrices ID | 12 |
| Tip | Help for selection of the number of primary PCs. |
| Details | The smallest number of PC, that is still showing changes, should be selected. |
| Related Topic | Spectra Residuals,Spectra Residuum,Loadings |

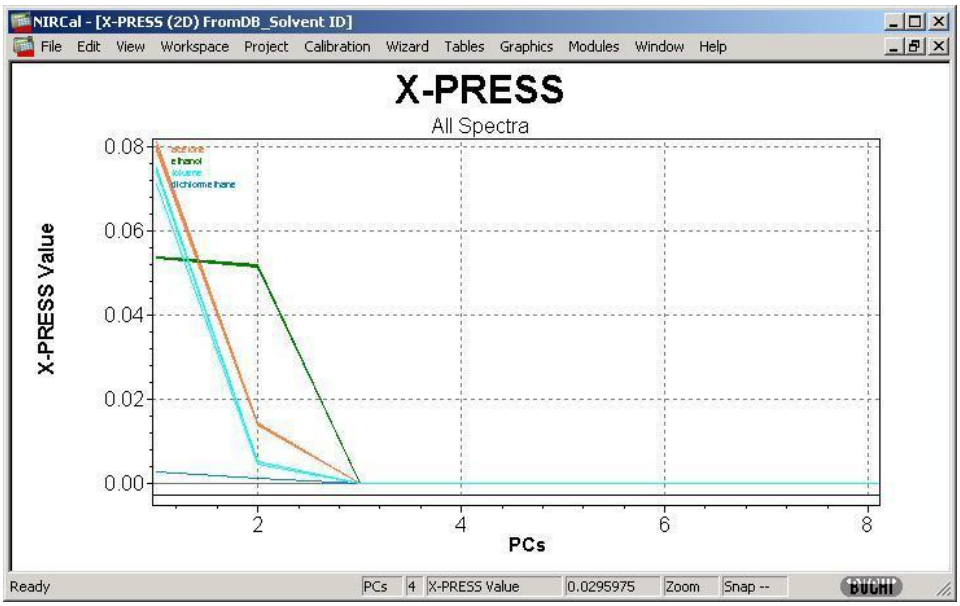
The X-PRESS function shows from which PC on an additional PC does not improve the spectra reconstruction.

The X-PRESS function sums up the intensity values that are left after reconstruction with each PC.

Interpretation: difference between pretreated spectrum and reconstructed spectrum will be summed up and squared. It is with the:

1. PC: X-PRESS 1= [pretreated spectrum-(mean spectrum + score 1 x PC 1)]2
2. PC: X-PRESS 2= [pretreated spectrum-(m.s. + score 1 x PC 1+ score 2 x PC 2)]2
3. PC: X-PRESS 3= [pretreated spectrum-(m.s. + score 1 x PC 1+ score 2 x PC 2 + score 3 x PC 3)]

The primary principal components should be set at the point where no significant changes in the X-PRESS value can be determined. The PC, which has no effect on the X-PRESS, normally has very small or only noisy information (see: Loadings).




Here: 1-3 PCs are important.

### 3.18.101 Y-Scores

| | |
|---|---|
| Description | Each spectra is placed in the n-dimensional y-score space. The position of the spectra is given by the n-dimensional coordinate of the y-scores. |
| Use | For PLS2 (multi property calibration). |
| Method | PLS |
| Matrices ID | 111 |
| Tip | Compare with the Scores. |
| Details | Calculated for the C-Set (V-Set Y-Scores are 0). |
| Related Topic | Scores, Y-Loadings |

Matrices of the final PLS calculation:

- Spectra are decomposed to Scores and Loadings.
- Properties are decomposed to Y-Scores and Y-Loadings.
- Y-Scores are predicted from the Scores and Properties are predicted from the Y-Scores.

### 3.18.102 y-loadings

| | |
|---|---|
| Description | Y-Loadings build up the base for reconstructing the property together with the Y-Scores. |
| Use | For PLS2 (multi property calibration). |
| Method | PLS |
| Matrices ID | 112 |
| Tip | |
| Details | For one C-Set property (PLS1) the Y-Loadings are constant 1. |

| Related Topic | Loadings, Y-Scores |
|---|---|

Matrices of the final PLS calculation:

- Spectra are decomposed to Scores and Loadings.
- Properties are decomposed to Y-Scores and Y-Loadings.
- Y-Scores are predicted from the Scores and Properties are predicted from the Y-Scores.

# 3.19 NIRCal Configuration Options

## 3.19.1 Configuration Dialog

Menu: Edit / Options

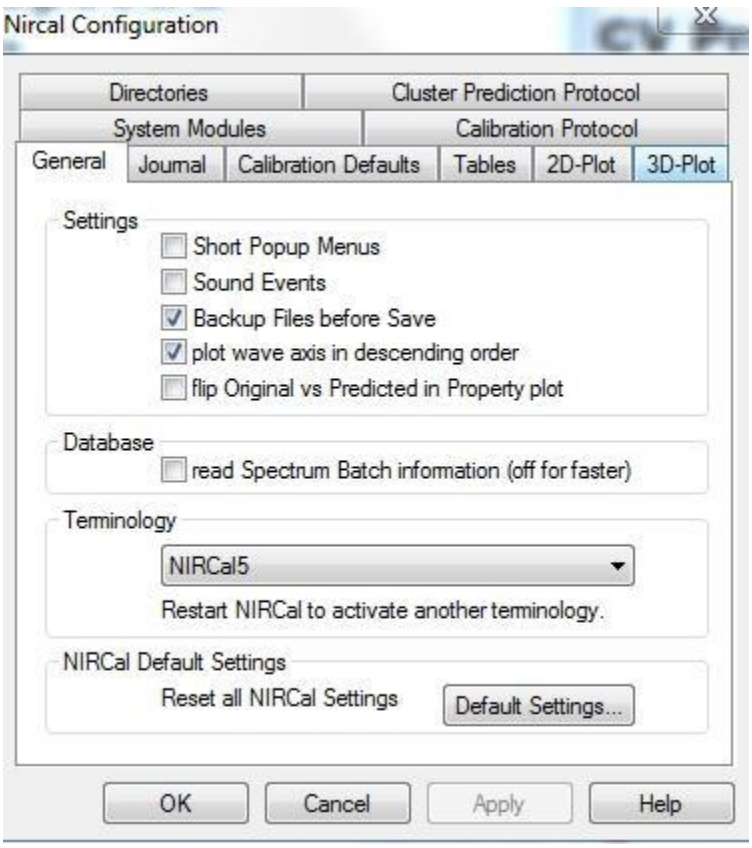
Icon: 

Different NIRCal functions can be configured here individually for each NIRCal Windows user account.

Configuration Dialog Tabs:

- General
- Journal
- Calibration Defaults
- Tables
- 2D-Plots
- 3D-Plots
- System Modules
- Calibration Protocols
- Directories
- Cluster Prediction Protocol

## 3.19.2 General



Settings

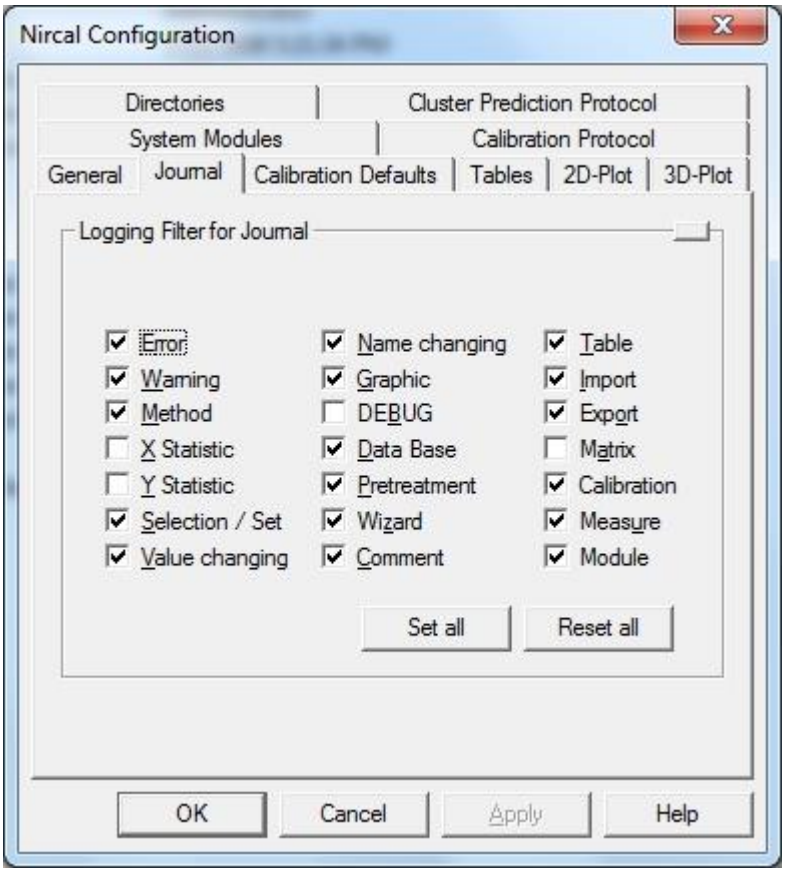
| | |
|---|---|
| Short Popup Menus | Displays short pop-up menus (right mouse button).<br>It is recommended to deactivate so that the long, extended pop-up menu is available. |
| Sound Events | Activates acoustic messages/sounds about the actual status of NIRCal (e.g. loading spectra, calculating PLS, ..... ) |
| Backup Files before Safe | Create a backup copy before a **file** is saved.<br>It is recommended to activate backup, which guarantees that a copy of each project is available. The backup file has the extension *.ni_. Rename the file extension to *.nir to be able to use it as a project. |
| plot wave axis in descending order | If it is activated, the plots show the wavenumber from 10'000 to 4'000 cm-1, if it is deactivated the plots show the wavenumber from 4'000 to10'000 cm-1 (like NIRCal 4). |
| flip Original vs Predicted in Property plot | Exchanges the axis in the plot "Predicted Property versus Original Property" |
| Database | If the batch information is not read, the spectra are loaded quicker from the database. |
| Terminology | Select NIRCal4 or NIRCal5 from the drop down list.<br>NIRCal5 uses the Marten-Naes terminology (scores, loadings/PC), NIRCal4 uses the Malinowski terminology (loadings, factors) |

NIRCal Default Settings

| | |
|---|---|
| Reset all NIRCal Settings | Resets all NIRCal settings to the default values or settings. |

### 3.19.3 Journal

Defines the Logging Filters for the journal.




The selected topics will be documented in the project journal. It is an important feature for the audit trail therefore the selection should be done carefully.

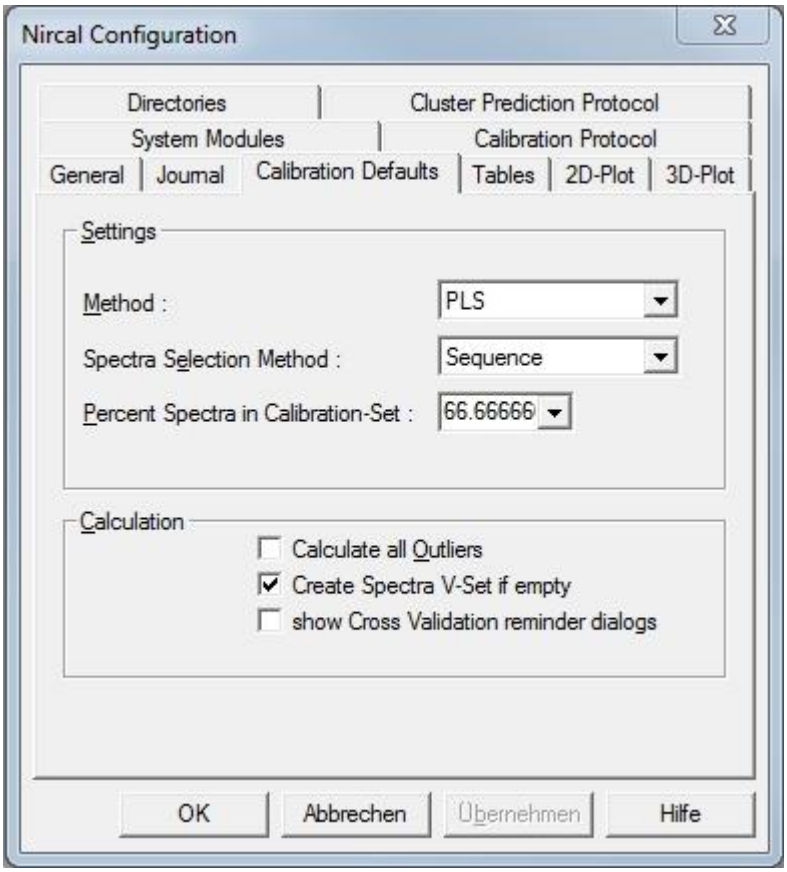
***NOTE***
The small buttons on the top right of a frame resets all values of the frame to the **default setting**.

***NOTE***
**This feature was used in the file-mode**. In database mode there is as **System Logger** in the Administrative Tools of the NIRWare Management Console.

***NOTE***
At the beginning of the calibration phase, **it is suggested to deactivate all entries** using the “Reset all” button, then activate Error and Warning. After the validation, activate all filters by clicking on the small box on the top right corner.

### 3.19.4 Calibration Defaults

Edits the default parameters for calibrations.




Settings

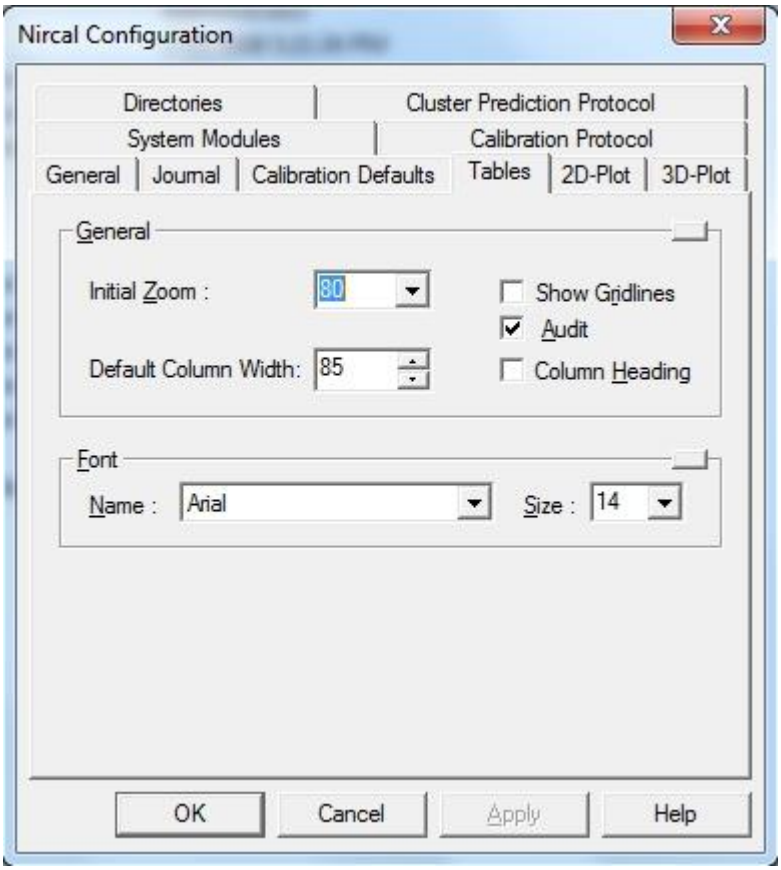
| | |
|---|---|
| Method | CLU or MLR, PCR, PLS. Default: PCR |
| Spectra Selection Method | Monte Carlo or Sequence. Default: Sequence |
| Percent Spectra in Calibration Set | between 0% and 100%. Default: 66.66% (2/3) |

Calculation

| | |
|---|---|
| Calculate all Outliers | Directs to calculate all outliers. Default: off<br>The parameters for the Outlier detection can be defined in menu Project/Outlier detection. |
| Create Spectra V-Set if empty | In case the V-Set is empty, the system create it with the chosen Spectra Selection Method. Default: on. |
| show Cross Validation reminder dialogs. | The reminder dialog is visible before the cross validation, warning the user that the calculation with a high number of spectra takes too much time. Default: on<br><br>See : Cross Validation |

### 3.19.5 Tables

Sets/changes the parameters for the Table window.




***NOTE***
The small buttons on the top right of a frame resets all values of the frame to the default setting.

General

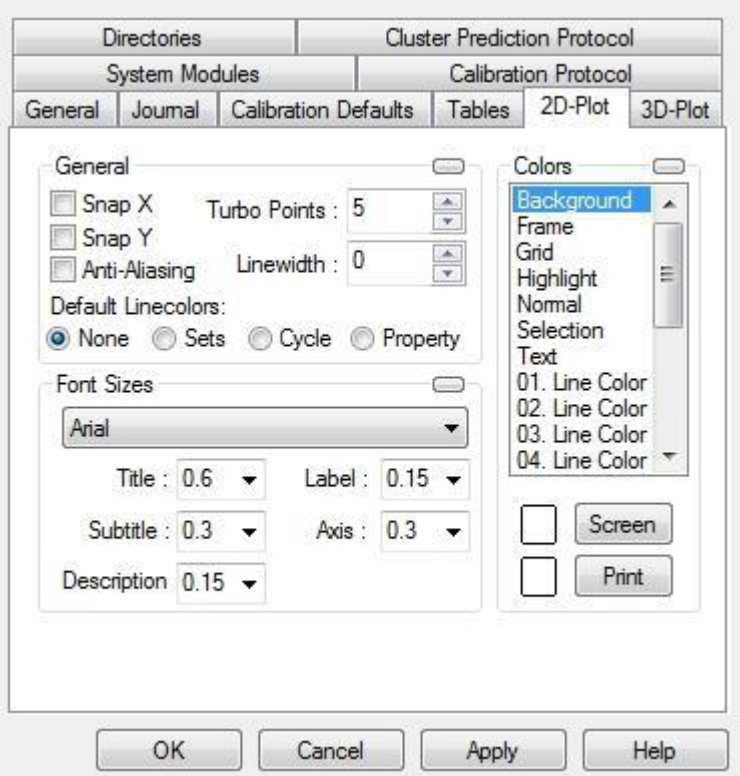
| | |
|---|---|
| Initial Zoom | default 80% |
| default Column Width | default 3250 |
| Show Gridlines | Enable / Disable, default OFF |
| Audit | Enable / Disable, default ON |
| Column Heading | Fixes the column heading, default OFF |

Font

| | |
|---|---|
| Name | default Arial |
| Size | default 10 |

### 3.19.6 2D-Plots

Sets/changes the parameters for a 2D graphic window.




***NOTE***
The small buttons on the top right of a frame resets the all values of the frame to default setting.

General

| | |
|---|---|
| Snap X | Snaps the zooming so that there is always a value at the beginning and at the end of the x-axis, default OFF. |
| Snap Y | Snaps the zooming so that there is always a value at the beginning and at the end of the y-axis, default OFF. |
| Anti-Aliasing | The new graphic feature for plots, default: off. |
| Turbo Points | Defines the step width of point for displaying a spectrum in an 2D-Plot. Use this to speed up the display operation for your 2D-Plots, default 5. To switch a Graphic into Turbo Mode choose Options / Turbo Mode. |
| Linewidth | Defines the width of the lines use for displaying or printing spectra/vectors in the 2D-Line-Plot.<br><br>**Recommendation**: Change the value only for printing spectra (e.g. to 5) with thicker lines. Afterwards reset the values to 0 (zero). Otherwise the display of graphs slows down considerably. |
| Default Linecolors None | Disables displaying 2D-Line-Plots using different colors. |
| Default Linecolors Sets | Enables displaying 2D-plots using different colors for C- and V-Set selection. |
| Default Linecolors Cycle | Enables displaying 2D-plots using all possible different line colors cycling around the spectra. |
| Default Linecolors Property | Enables displaying 2D-plots using different colors for different properties. |

Suggested default Line Colors:

- Select **Property for qualitative** calibrations or
- Select **Sets for quantitative** calibrations. (02 Line: C-Set, 03 Line: V-Set)

Line Colors: 02 Line Color = 1st color; 03 Line Color = 2nd color.

Font Sizes

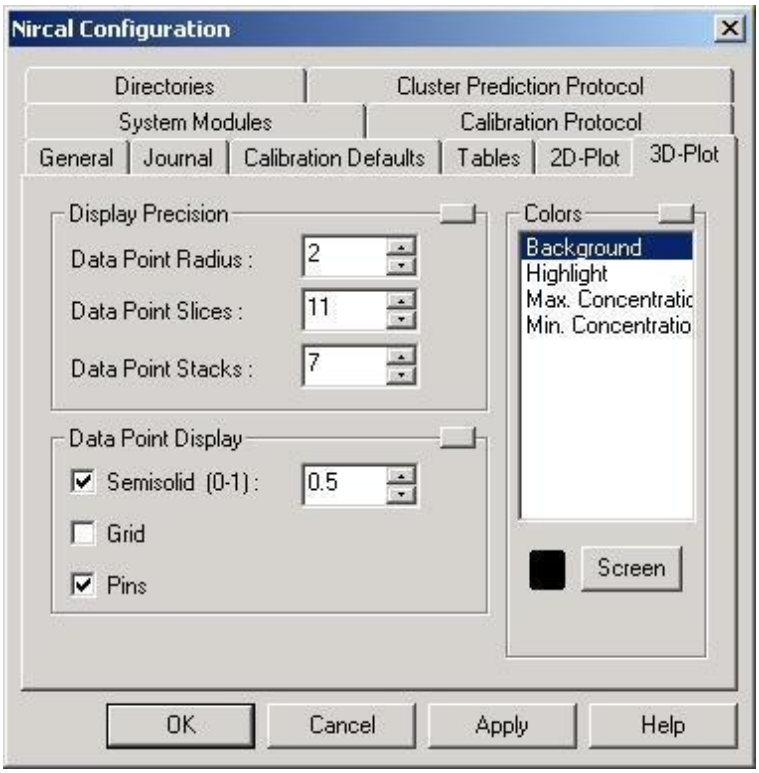
| | |
|---|---|
| Font Drop Down List | Defines the Font used in any 2D-Plot, default Arial |
| Title | Defines the size of the font used for titles, default 0,6. |
| Subtitle | Defines the size of the font used for subtitles, default 0,3. |
| Description | Defines the size of the font used for descriptions, default 0,15. |
| Label | Defines the size of the font used for labels, default 0,15. |
| Axis | Defines the size of the font used for axis description, default 0,3. |

Colors

| | |
|---|---|
| Screen | Defines the screen color for the selected entry in the list above. The selected color is displayed in the box on the left of the button. |
| Print | Defines the printing color for the selected entry in the list above. The selected color is displayed in the box on the left of the button. |

### 3.19.7 3D-Plots

Sets/changes the parameters for 3D graphical windows.




Note: the small buttons on the top right of a frame resets the all values of the frame to default setting.

Display Precision

| | |
|---|---|
| Data Point Radius | Defines the radius in generalized units for displaying data points in the 3D-Plot, default 2. |
| Data Point Slices | Defines the number of slices in generalized units for displaying data points in the 3D-Plot, default 11 |
| Data Point Stacks | Defines the number of stacks in generalized units for displaying data points in the 3D-Plot, default 7 |

Data Point Display

| | |
|---|---|
| Semisolid (0-1) | Displays the "spheres" semisolid (transparent), default: off. It is suggested to switch to on with 0.5. |
| Grid | Displays the "spheres" in a grid, default: off. |
| Pins | Displays a pin for each spectra, default: on. |

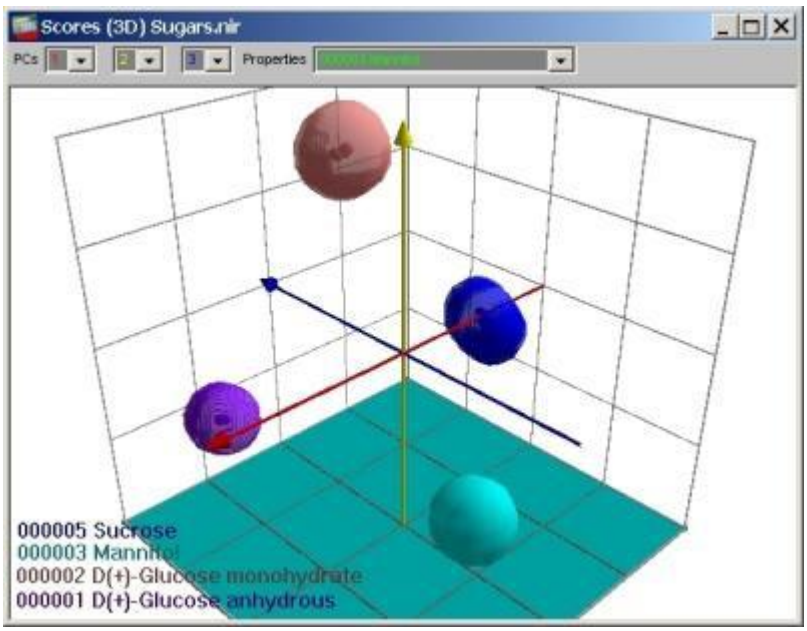
Colors

| | |
|---|---|
| Screen | Defines the screen color for the selected entry in the list above. The selected color is displayed in the box on the left of the button. |

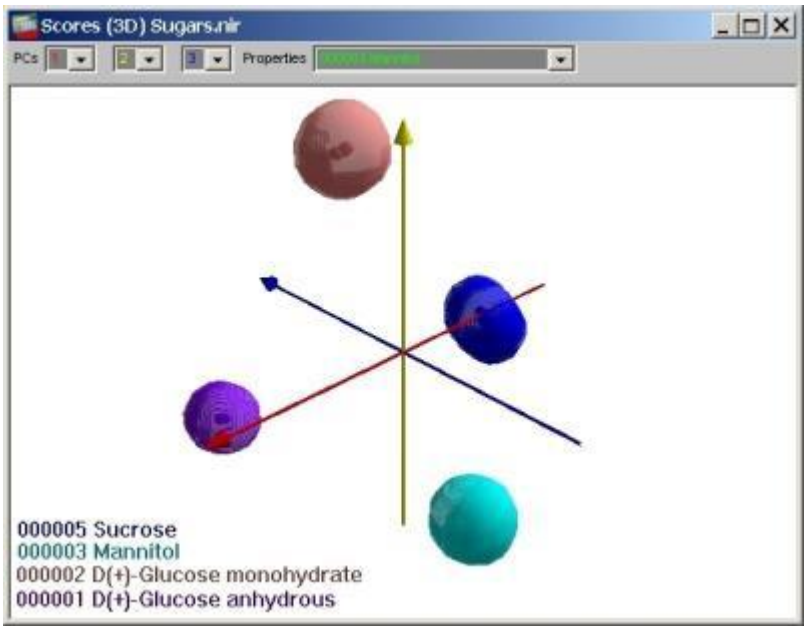
Displays the "spheres" as a grid.

Boxes can be toggled by tipping B on the keyboard

3D with boxes

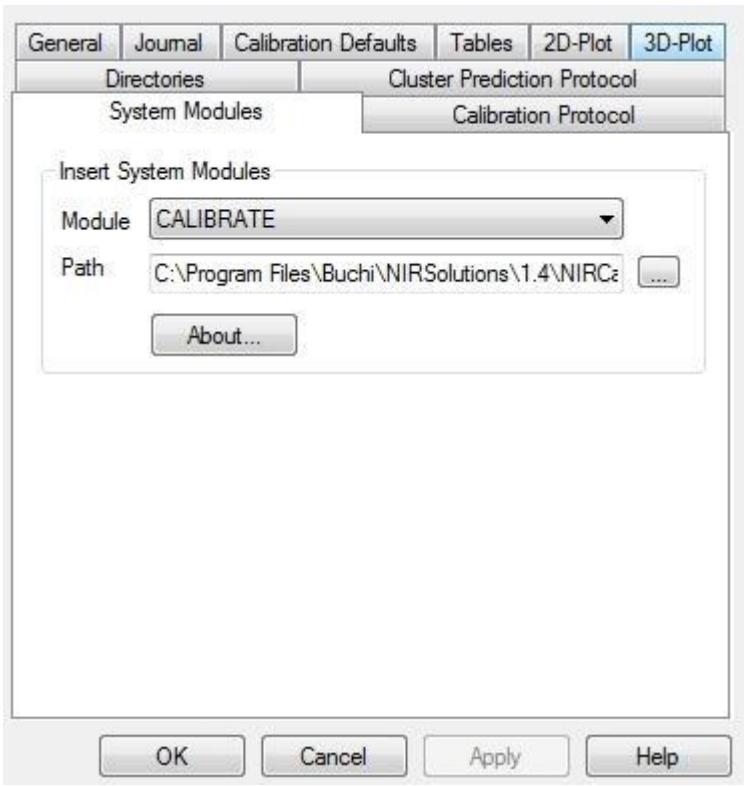


3D without boxes




## 3.19.8 System Modules

Sets the path for the system modules.



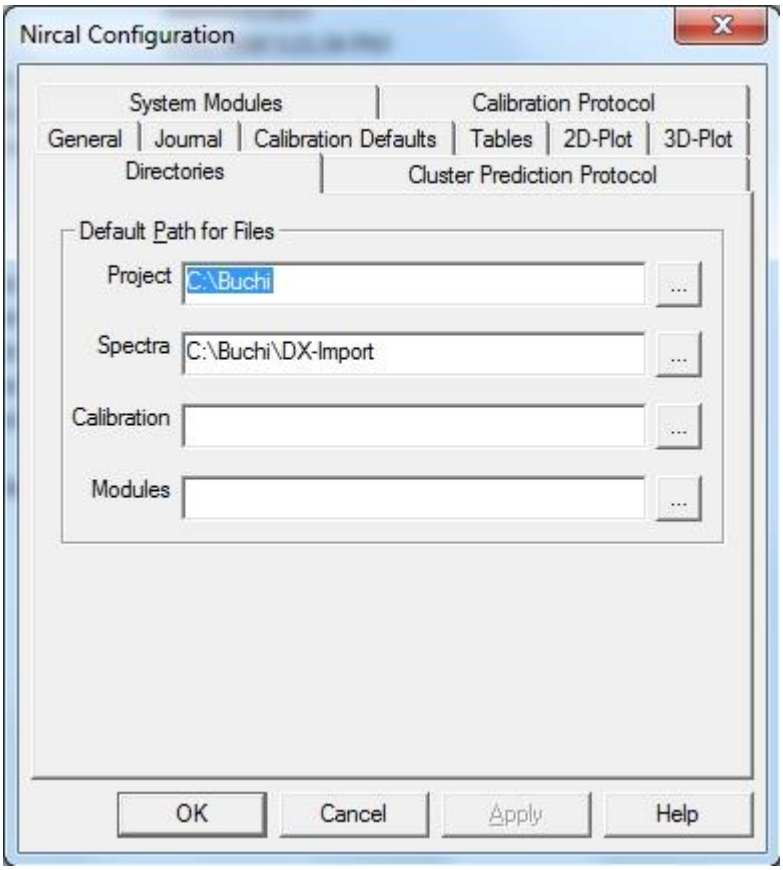

***NOTE***
**The path should be changed only for your own** created or modified **modules: it is not recommended to use it.**

## 3.19.9 Calibration Protocols

The content of the full calibration protocol can not be defined, this field **has no function** anymore.

### 3.19.10 Directories

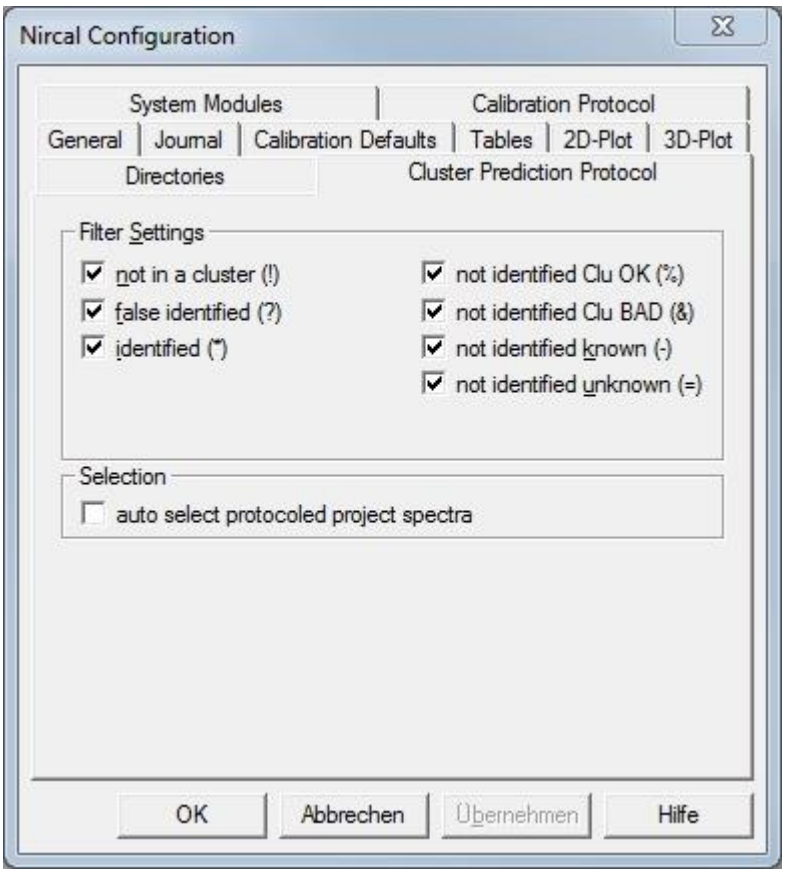
The default paths for various files can be set individually.




***NOTE***
**It is NOT suggested to change the path for Modules and Data Exchange Filters!**

### 3.19.11 Cluster Prediction Protocol

Sets/changes the filter for the Prediction Protocol of Cluster and SIMCA calibrations.




Filter Settings

| | |
|---|---|
| not in a cluster (!) | residual is within limits, but the distance is higher than allowed (spectrum is not in a cluster), default: on. |
| false identified (?) | residual is within limits, distance is smaller than allowed (spectrum is in a cluster), the predicted and original property name are different: this spectrum is identified as false, default: on. |
| identified (*) | residual is within limits, distance is smaller than allowed, the predicted and original property names are the same: this spectrum is identified correctly, default: off. |
| not identified CLU OK (%) | residual is not within limits, distance is smaller than allowed, the predicted and original property names do not match: this known spectrum is not identified, default: on. |
| not identified CLU BAD (&) | residual is not within limits, distance is smaller than allowed, the predicted and original property names are not the same: this spectrum is not identified, but it is in a known cluster, default: on. |
| not identified known (-) | residual is not within limits, the distance is not fulfilled, but the property is known in the calibration, default: on. |
| not identified unknown (=) | residual and distance is not fulfilled, the property is not available in the calibration, default: on. |

With the shown settings, only correct identified spectra are not listed in the prediction protocol.
By deactivating further categories, the list of spectra in the prediction protocol are reduced.

| Case | Group | Symbol | Residual | Distance | Name/Property |
|---|---|---|---|---|---|
| 1 | not in a cluster | ! | OK | not OK | not OK |
| 2 | false identified | ? | OK | OK | not OK |
| 3 | identified | * | OK | OK | OK |
| 4 | not identified CLU OK | % | not OK | OK | OK |
| 5 | not identified CLU BAD | & | not OK | OK | not OK |
| 6 | not identified known | - | not OK | not OK | OK |
| 7 | not identified unknown | = | not OK | not OK | not OK |

Especially cases **1 and 5 are critical**. By case 5 the residual relationship (actual against allowed) should be controlled. **Case 2 is inacceptable.**
**Correct identifications** -case 3- and not identified unknown -case 7- are less important.

Selection

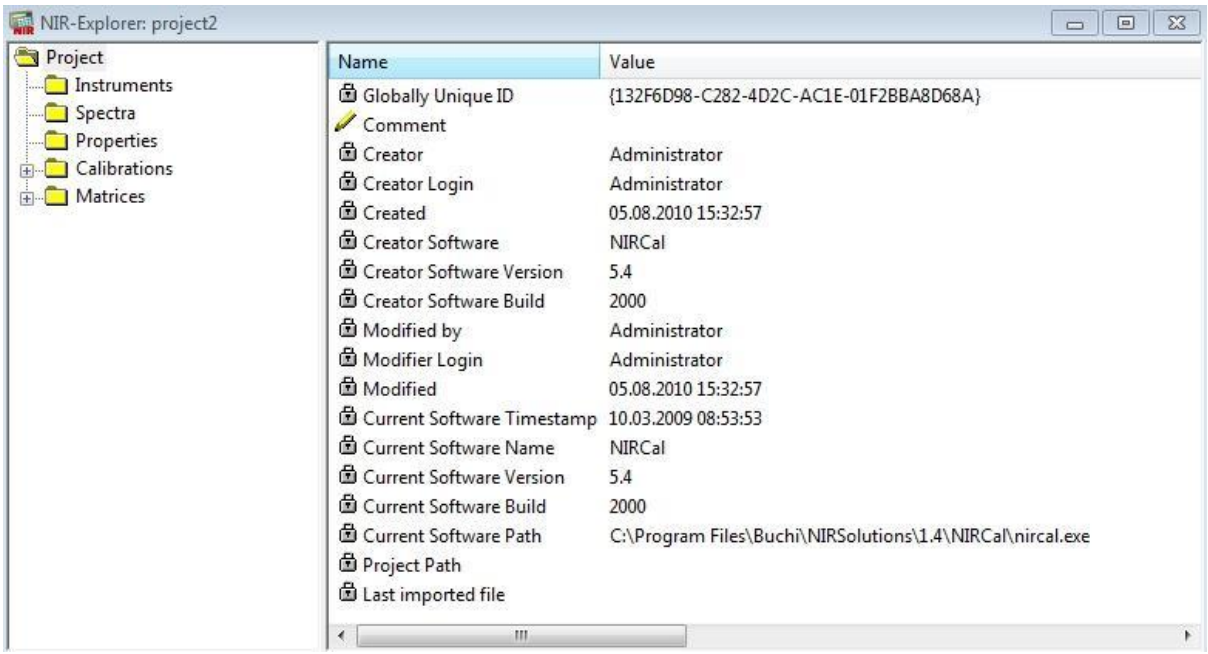
auto select protocoled project spectra — Default OFF. With ON the protocoled spectra are automatically selected, can help to identify outliers in the project (with "Predict Project" and "Multi Predict Project".
It is suggested to switch on by "predict projects" and "multi predict projects"

# 3.20 Menu commands

## 3.20.1 File menu

### New Project


Creates a new NIRCal Project.



**Close Project**
Closes the project.

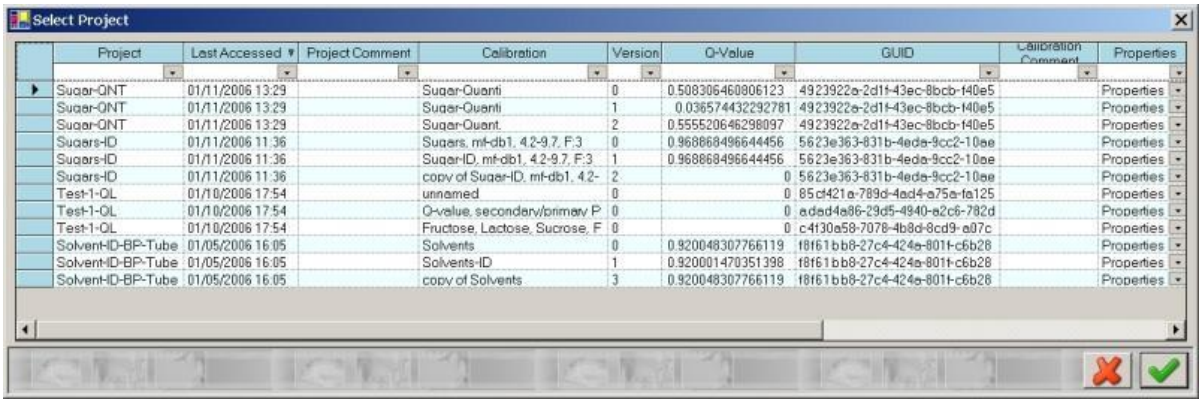
The following message appears:


It is recommended to press „Yes".

**Database**

***Open Project***
Menu: File / Database / Open Project...
Icon:

Opens a NIRCal Project from the NIRWare Database.

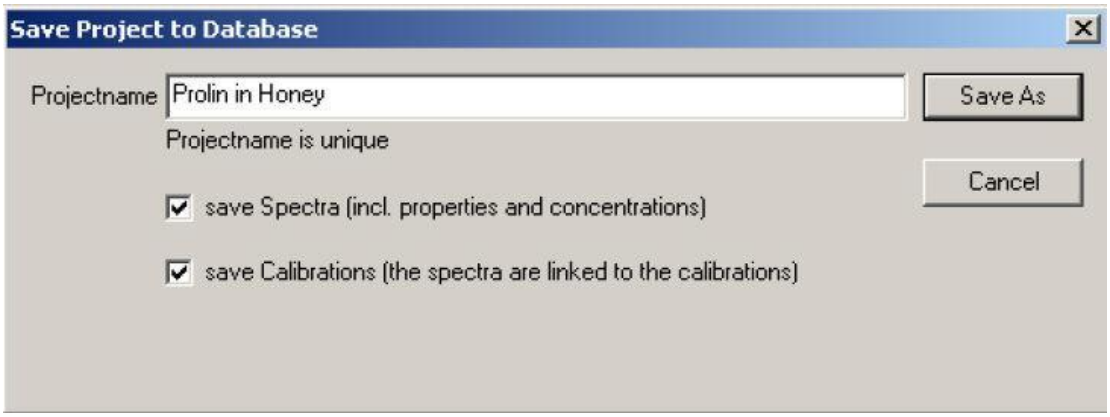


***Save Project***

Menu: File \ Database \ Save Project As...

Icon: 

Saves the active project to the NIRWare Database.

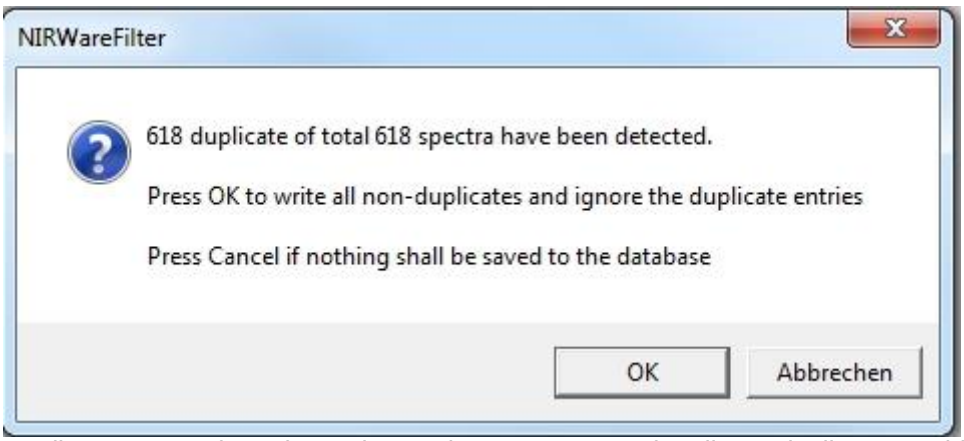



This function is needed to save a NIRCal-Project (opened from file *.nir) to the NIRWare Database.

Saving the spectra with the setting : **[x] save Spectra (incl. properties and concentrations)** will force a duplicate spectra check. (Note: This works because all NIRCal written files include the spectrum GUID's (also in JCAMPDX) and this GUID is also used in the NIRWare database.)

After saving the project to the database, the dialog to assign the Project to an application automatically appears.

*Duplicate spectrum GUID*




Duplicate spectra have been detected. Press OK to write all non-duplicates and ignore the duplicate entries. Press Cancel if nothing shall be saved to the database.

***Save Project As...***

Saves the active project to the NIRWare Database with a new Name


This function is needed:

- To save a project if the spectra are loaded from the database;
- To create a copy of a NIRCal-Project (opened from database) with another name. Not recommended to use.

Note: the calibrations are not copied; they are linked to the project. To copy the calibration use Lifecycle/Copy.

***NOTE***

Use a unique project name with Latin letters (a-z) only.

***Import latest Application spectra***

Menu: File / Database / Import latest Application spectra...

Icon: 

Automatically imports all spectra from the database to NIRCal which are assigned to the Application. See Manage Application...

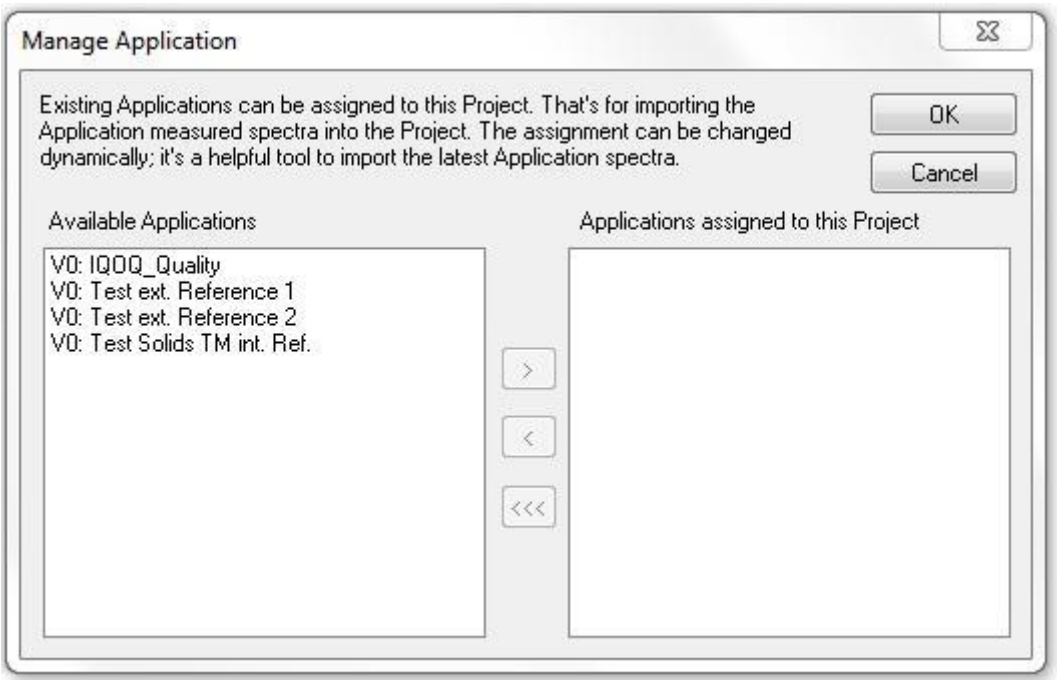
***Manage Application...***

Menu: File / Database / Manage Application...

Icon: 

An existing application in the database can be assigned to a project. New measured spectra (NIRWare Operator) from an application assigned to a NIRCal project can automatically be imported to this project. This is useful for recalibration.

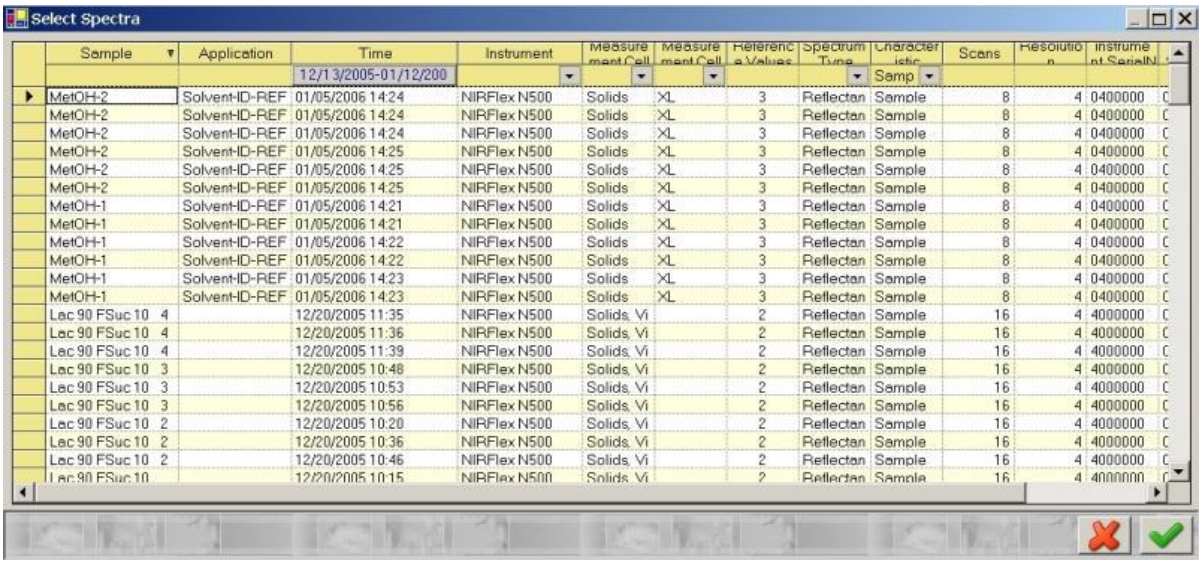
See Import Spectra to assigned NIRCal Project




This Dialog also appears when a NIRCal Project is saved to DB.

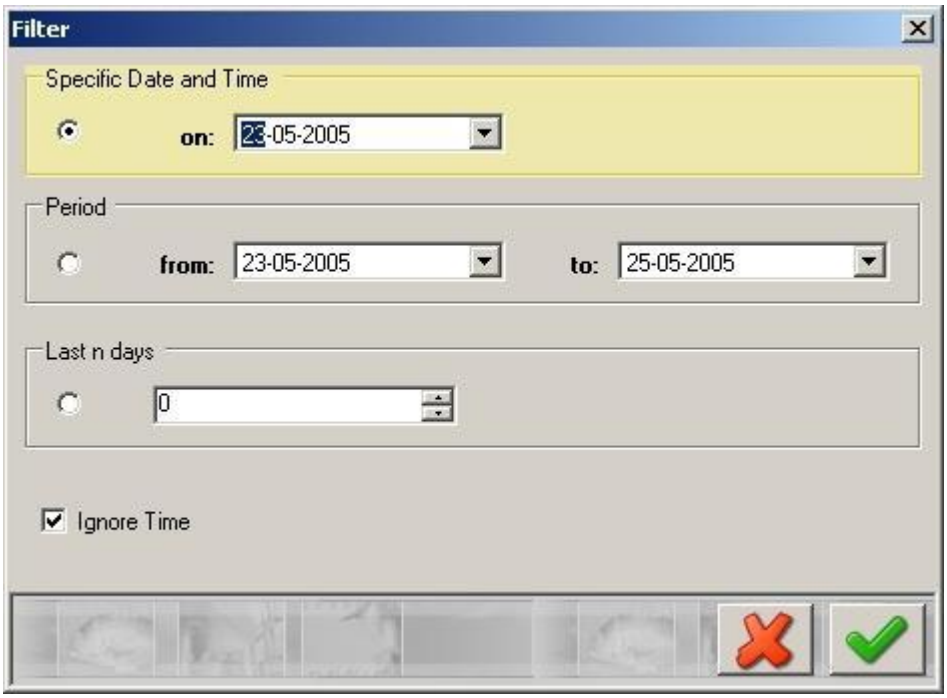
***Search and Import Spectra...***

 Opens the dialog to search and import spectra from the NIRWare Database to NIRCal.
The list displays per default only spectra from last 30 days [Time = e.g. from 05.01.2005 to 4.02.2005] and with an assigned reference value [Reference Values >0].




To remove the default filter and display all spectra within the database :

- click on the grey cell below [Time] and following dialog appears




- change the filter settings and click ok.
  Clicking Cancel disables the time filter completely. Same effect when clicking the grey cell below [Time] with the right mouse.

- click in the cell below [Reference Values] , delete >0 and press ENTER

Filtering options:

- Time Filter, see above
- Drop Down list for [Instrument], [Measurement Cell], [Measurement Cell Option], [Spectrum Type], [Characteristic]
- Text for [Sample], [Application]; e.g. edit "sug" in the cell below [Application] will display all spectra measured with applications, e.g. "sugar" or "sugars" etc.
- Digit for [all others]; e.g. edit "<10" in the cell below [Scans] will display all spectra scanned with less then 10 scans (<,>,<>,=,>=,<=)

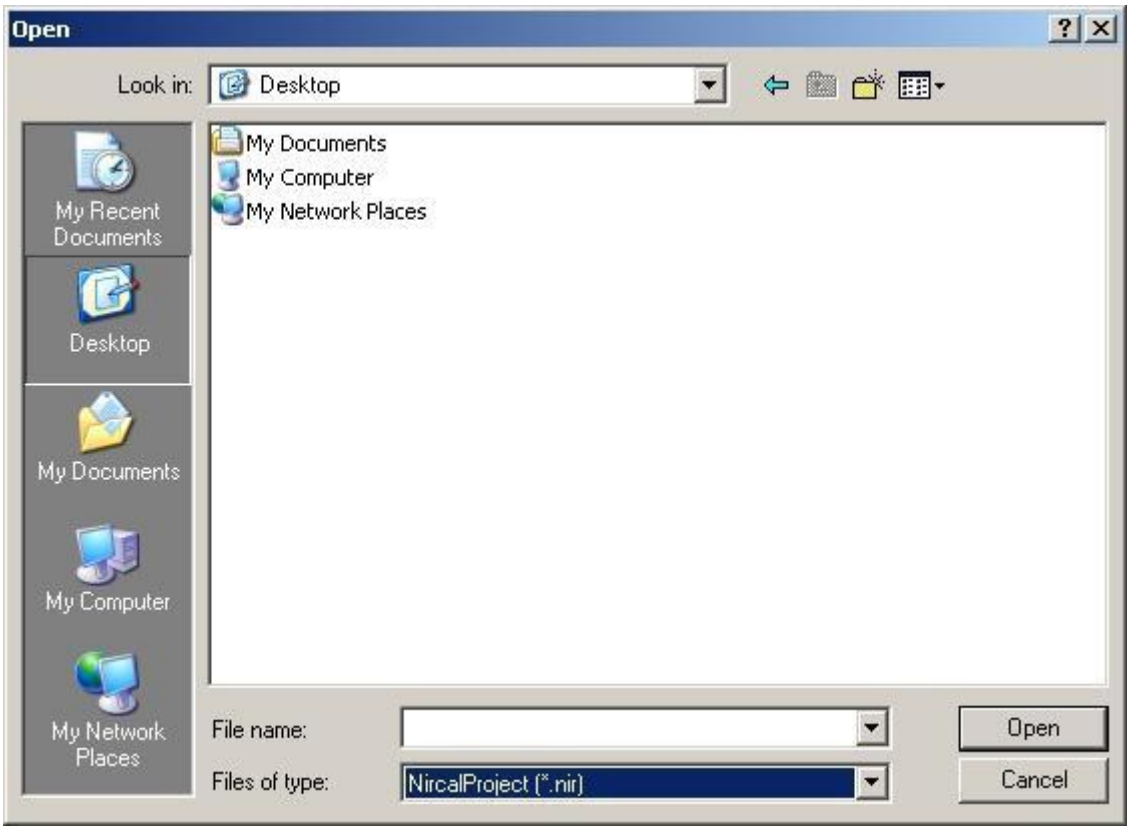
**Database Options**

***Delete Project***

Not used projects can be deleted. Delete the existing calibrations inside the project first with “Lifecycle/Delete”, otherwise the calibrations are still existing in the database.

**Import**

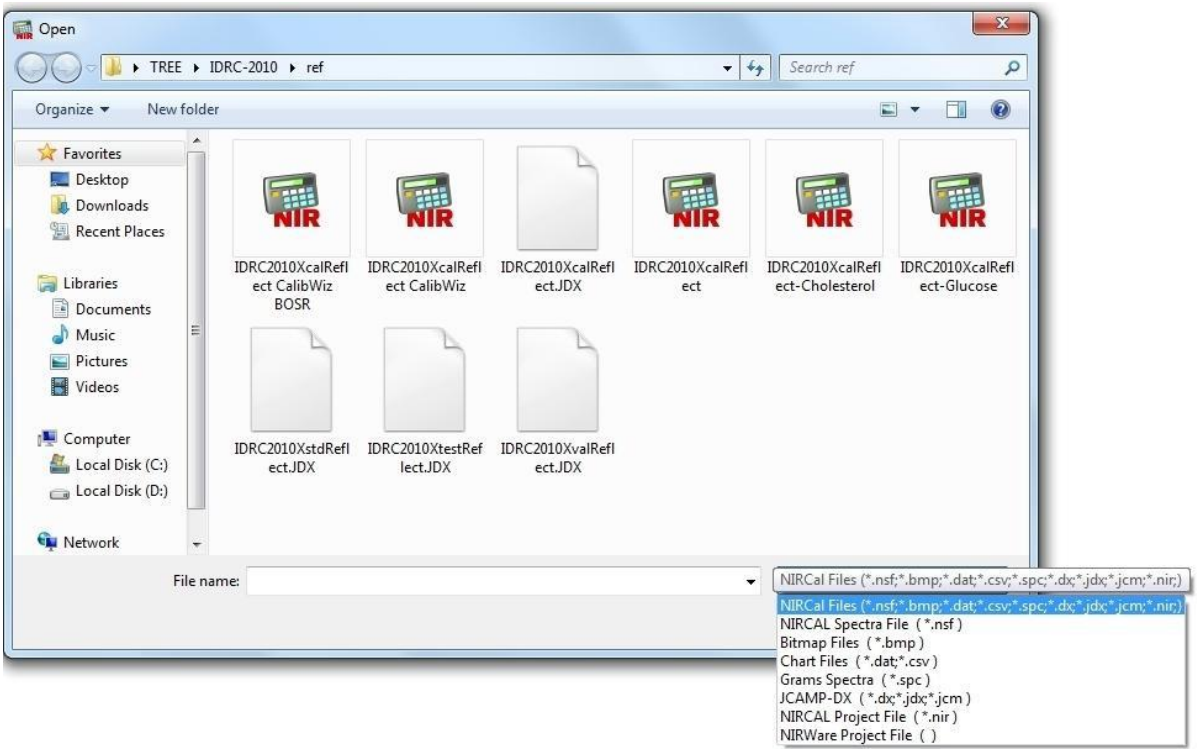
***Import Project***
Imports NIRCal Project Files (*.nir)




***NOTE***
BCAP Series can be loaded adjusting "Files of type" to "all files" (*.*) and selecting the spectra file "...Sxx"

***Import Spectra***

Imports Spectra from a File.

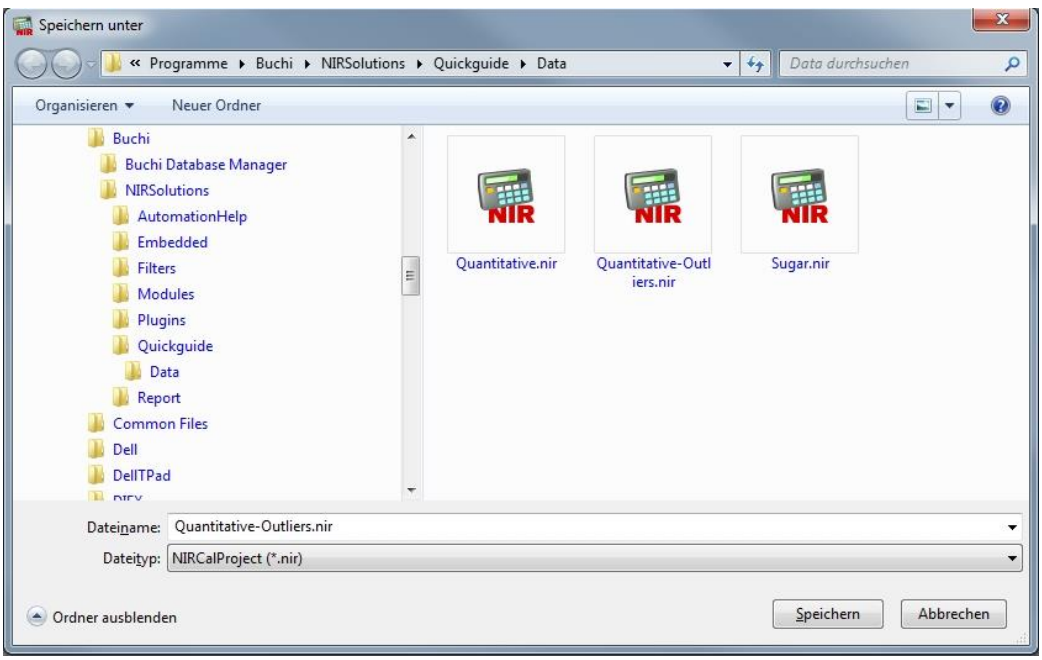



Several data formats can be imported.

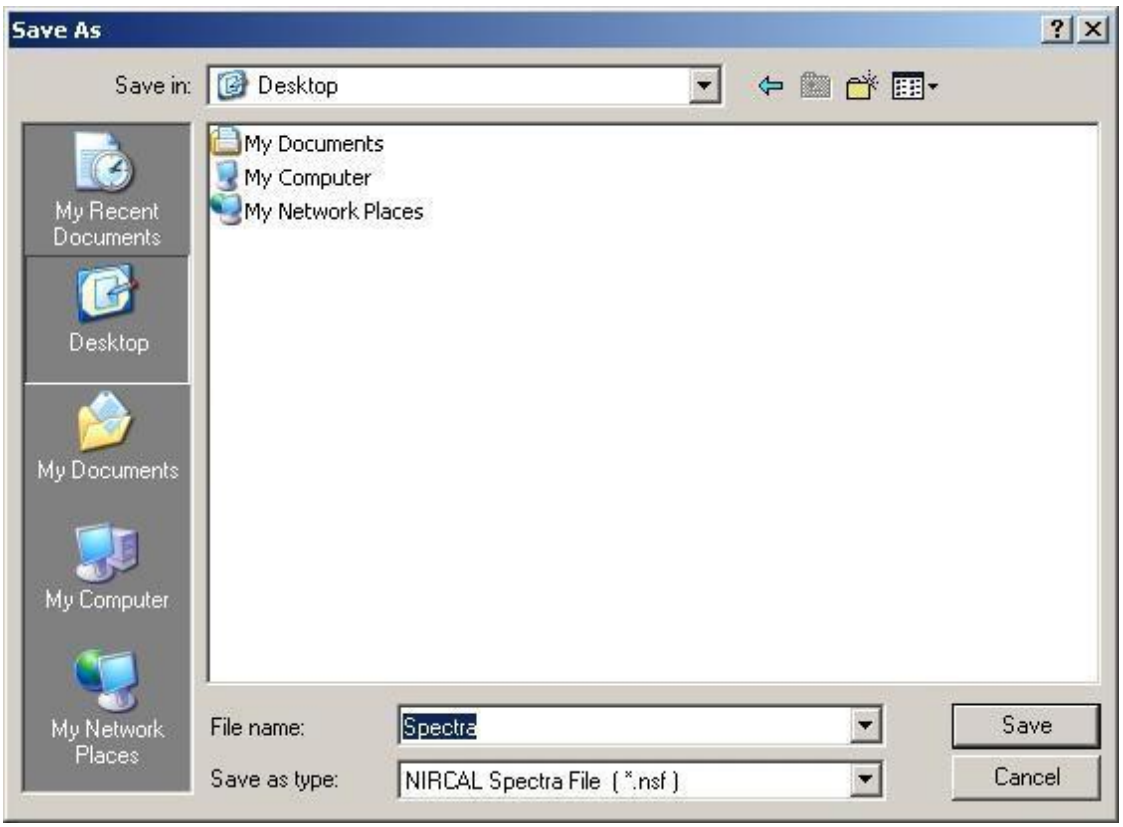
**Export**

***Export Project***

Save a NIRCal Project as *nir File. The question: “Store the project into the database before?” appears. It is better to export without storing (the calibration will be closed and will not be editable).



***Export Spectra***

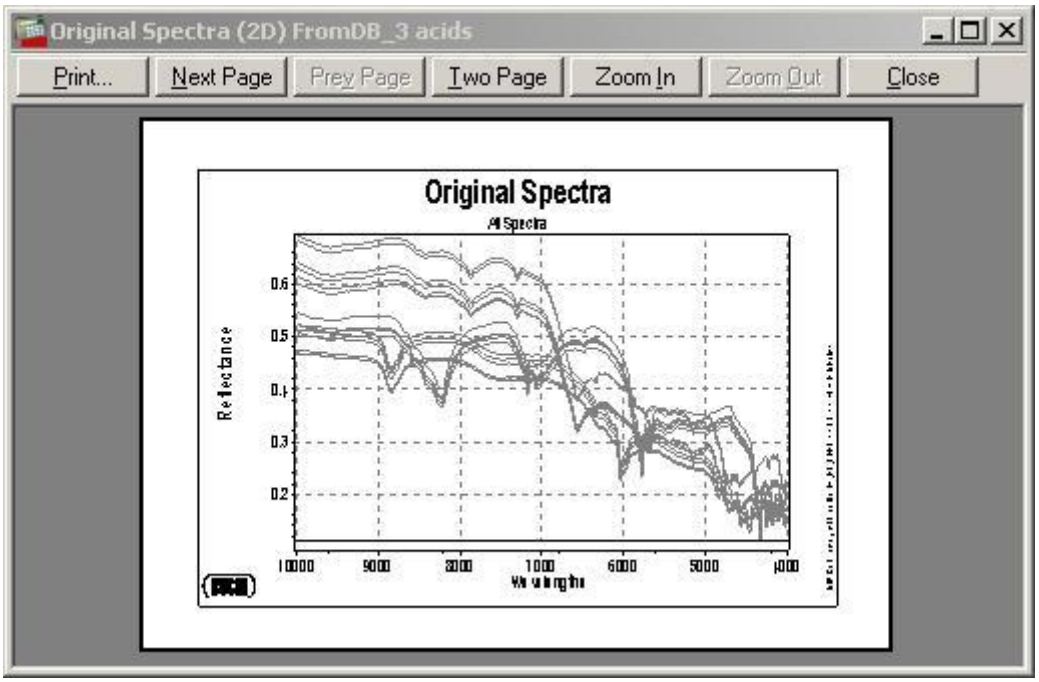



**Print**

Prints the current view.

**Print Preview**

Print Preview of the current view.



**Print Setup**
Opens dialog to setup the default printer.


**Exit**
Closes NIRCal incl. all opened NIRCal-Projects.

## 3.20.2 Edit menu

**Undo**
Undos text editing

**Cut**
Ctrl+X

**Copy**
Copies the active object to the clipboard

View Type Clipboard Object
2D-Plots Metafile
3D-Plots Bitmap
Tables ASCII-Text cells separated by TAB

**Paste**
Pastes text from the clipboard into the actual document.

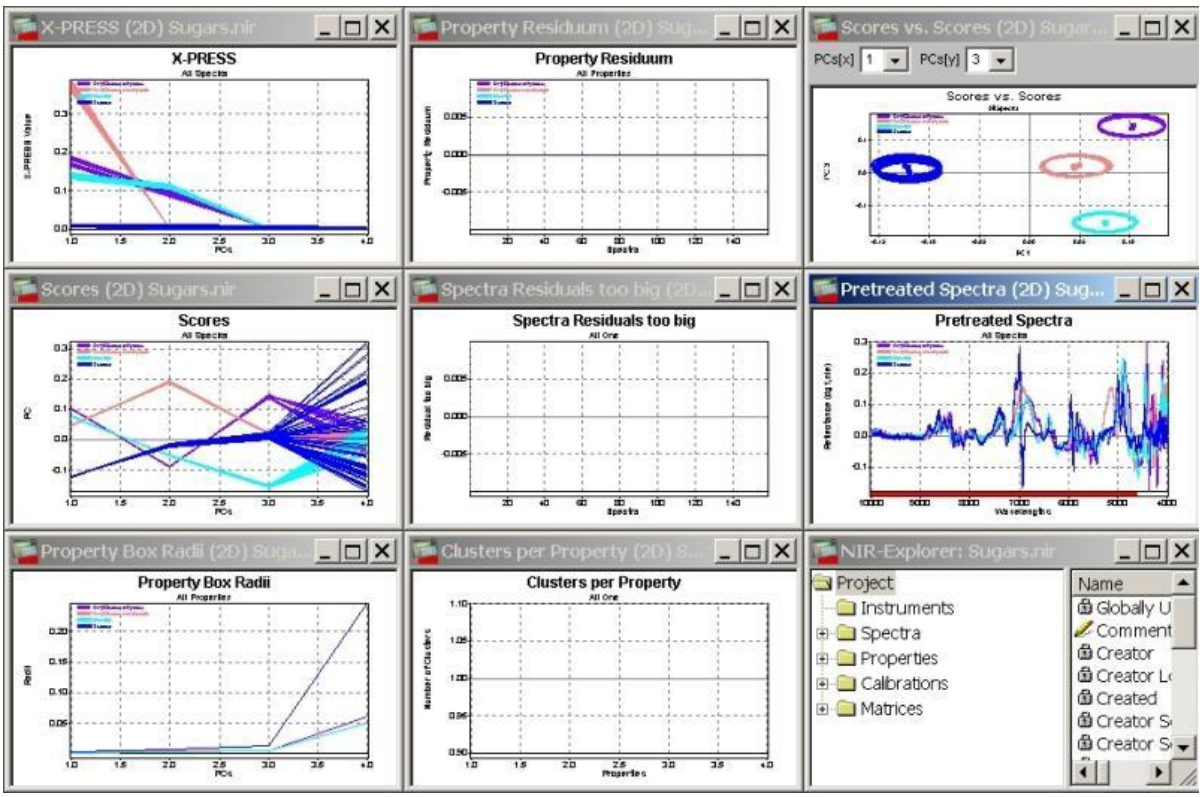
**Options**
Opens the Dialog Box to change NIRCal Configuration Settings.

### 3.20.3 View menu

**Overview**

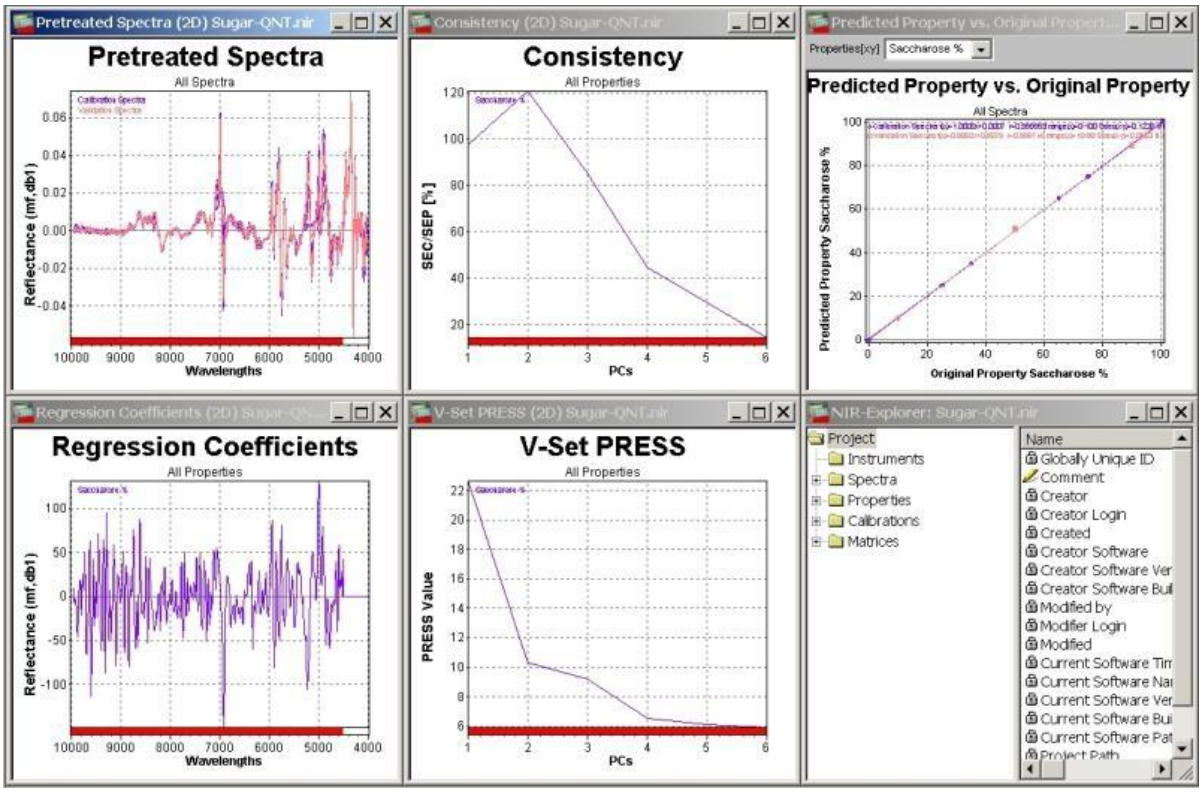
 Recalculates the calibration and opens several graphics for an overview of the calibration. Depending on the selected calibration method the most useful graphics are opened.

**Example: Overview of a Cluster calibration**




For detailed information see the description of the corresponding matrices:
X-PRESS, Scores, Property Box Radii, Property Residuum, Spectra Residuals too big, Cluster per Property, Pretreated Spectra

**Example: Overview of a Quantitative calibration**



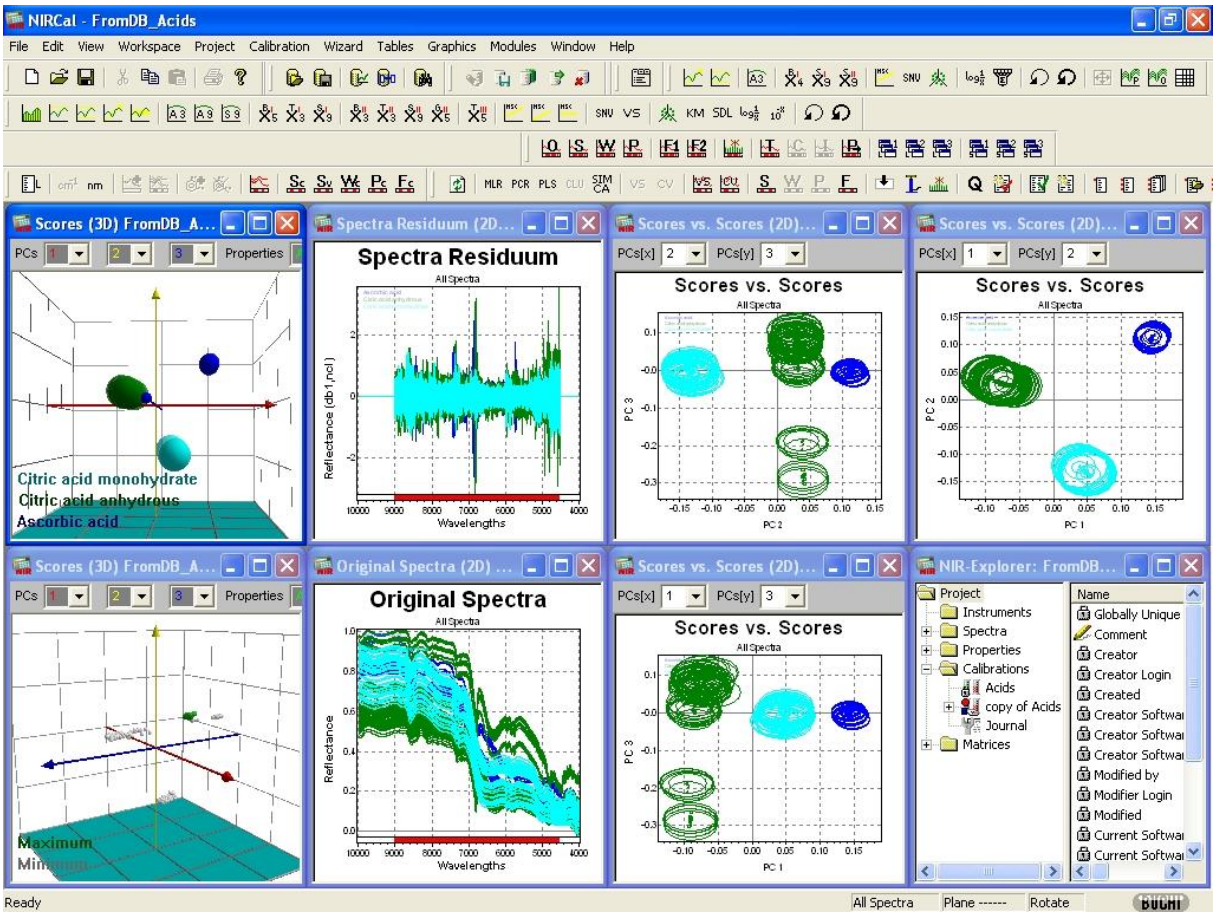

For detailed information see the description of the corresponding matrices:
Pretreated Spectra, Regression Coefficients / Property Spectra (CLU), Consistency, V-Set PRESS, C-Set Regression Coefficient

### Spectra Selection

 Opens a few windows for spectra selection:

You can select spectra in the plot by double clicking on the displayed spectra or changing the selection mode to "Mouse Window Select" using the popup menu of the window.

Example of a Spectra Selection Plot

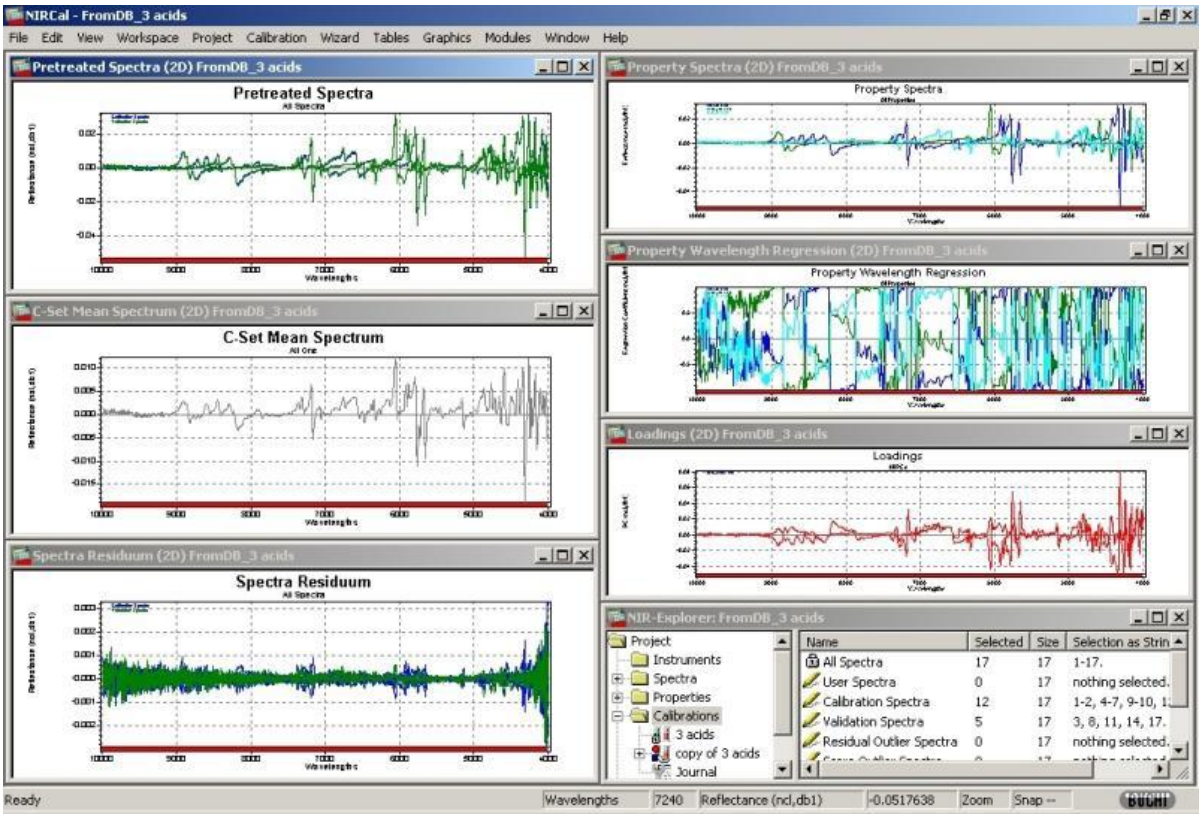


**Wavelength Selection**

 Opens a few windows for wavelength selection:

You can select wavelengths in the plots by changing the selection mode to x axis selection and then marking a region in the plot keeping the left mouse button pressed.

Example of a Wavelength Selection Plot

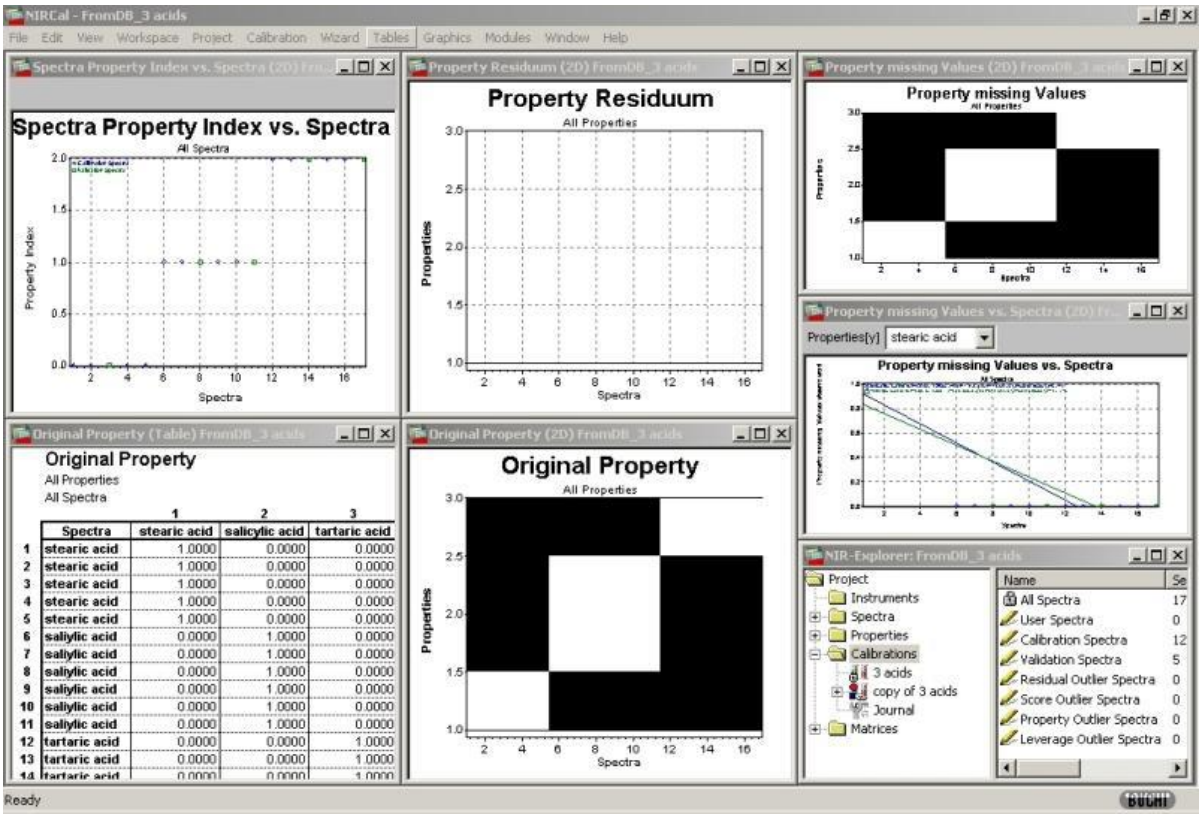


### Property Selection


Opens a few windows for property selection:

You can select property in the plot by double clicking on the displayed property or changing the selection mode to "Mouse Window Select" using the popup menu of the window.

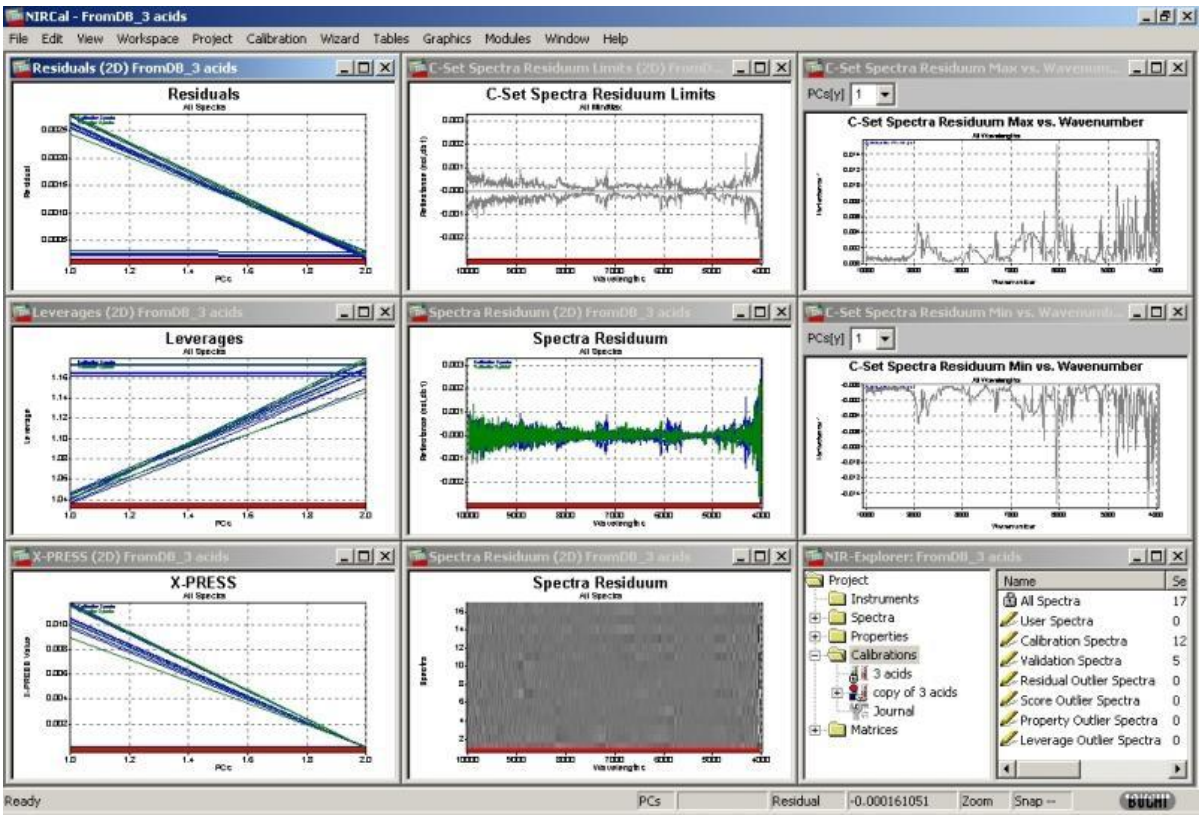
Example of a Property Selection Plot




| | Spectra | 1 stearic acid | 2 salicylic acid | 3 tartaric acid |
|---|---|---|---|---|
| 1 | stearic acid | 1.0000 | 0.0000 | 0.0000 |
| 2 | stearic acid | 1.0000 | 0.0000 | 0.0000 |
| 3 | stearic acid | 1.0000 | 0.0000 | 0.0000 |
| 4 | stearic acid | 1.0000 | 0.0000 | 0.0000 |
| 5 | stearic acid | 1.0000 | 0.0000 | 0.0000 |
| 6 | salicylic acid | 0.0000 | 1.0000 | 0.0000 |
| 7 | salicylic acid | 0.0000 | 1.0000 | 0.0000 |
| 8 | salicylic acid | 0.0000 | 1.0000 | 0.0000 |
| 9 | salicylic acid | 0.0000 | 1.0000 | 0.0000 |
| 10 | salicylic acid | 0.0000 | 1.0000 | 0.0000 |
| 11 | salicylic acid | 0.0000 | 1.0000 | 0.0000 |
| 12 | tartaric acid | 0.0000 | 0.0000 | 1.0000 |
| 13 | tartaric acid | 0.0000 | 0.0000 | 1.0000 |
| 14 | tartaric acid | 0.0000 | 0.0000 | 1.0000 |

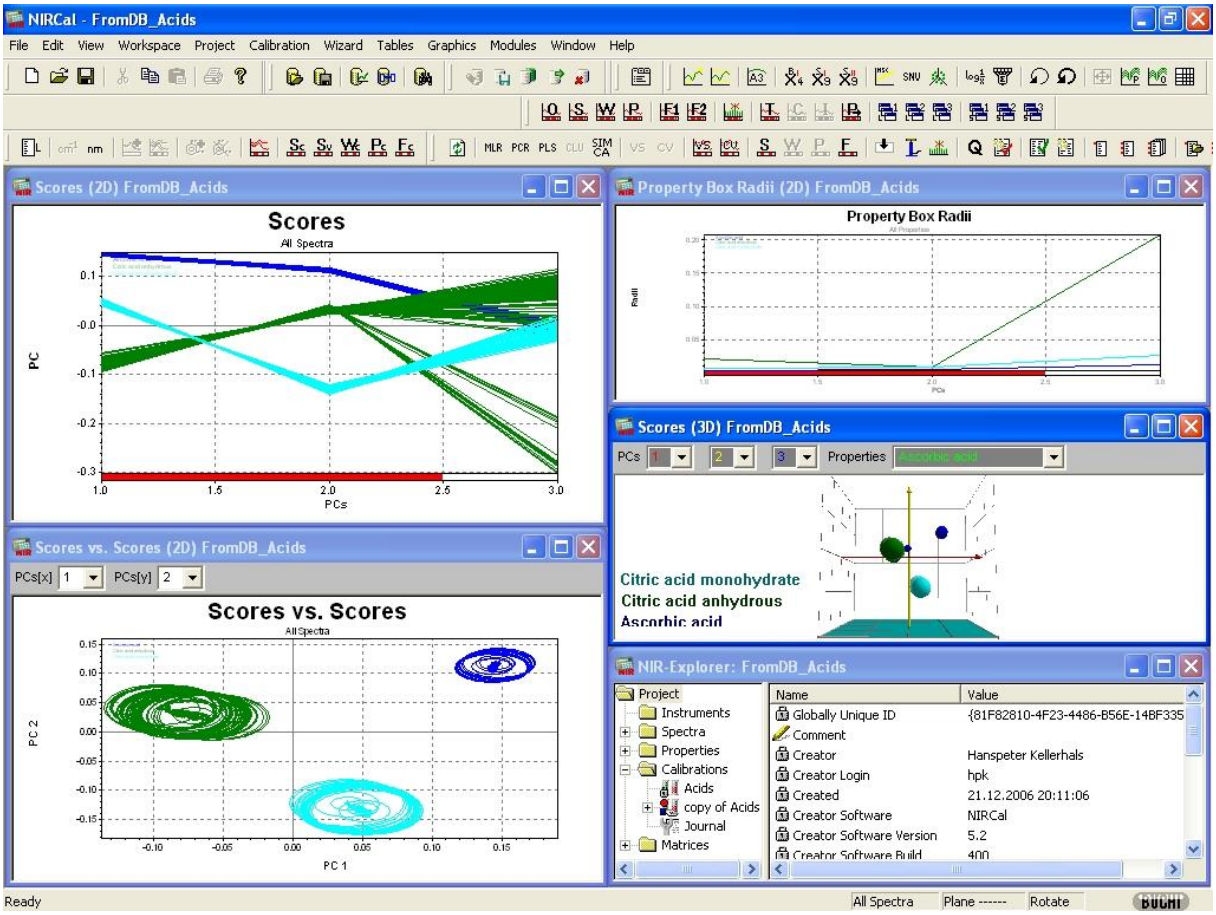
### Primary PCs (Factor) Selection

Example of a Primary PCs Selection Plot



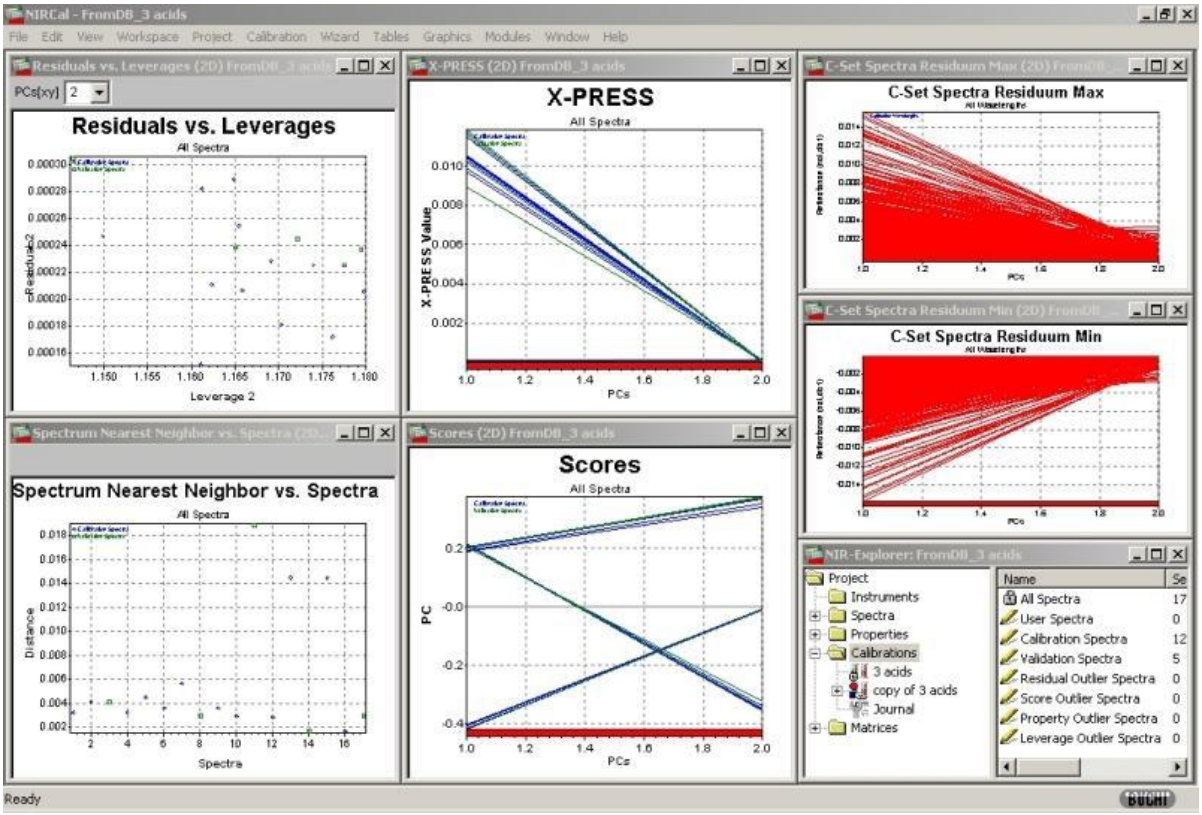
### Secondary PCs (Factor) Selection

Example of a Secondary PCs Selection Plot



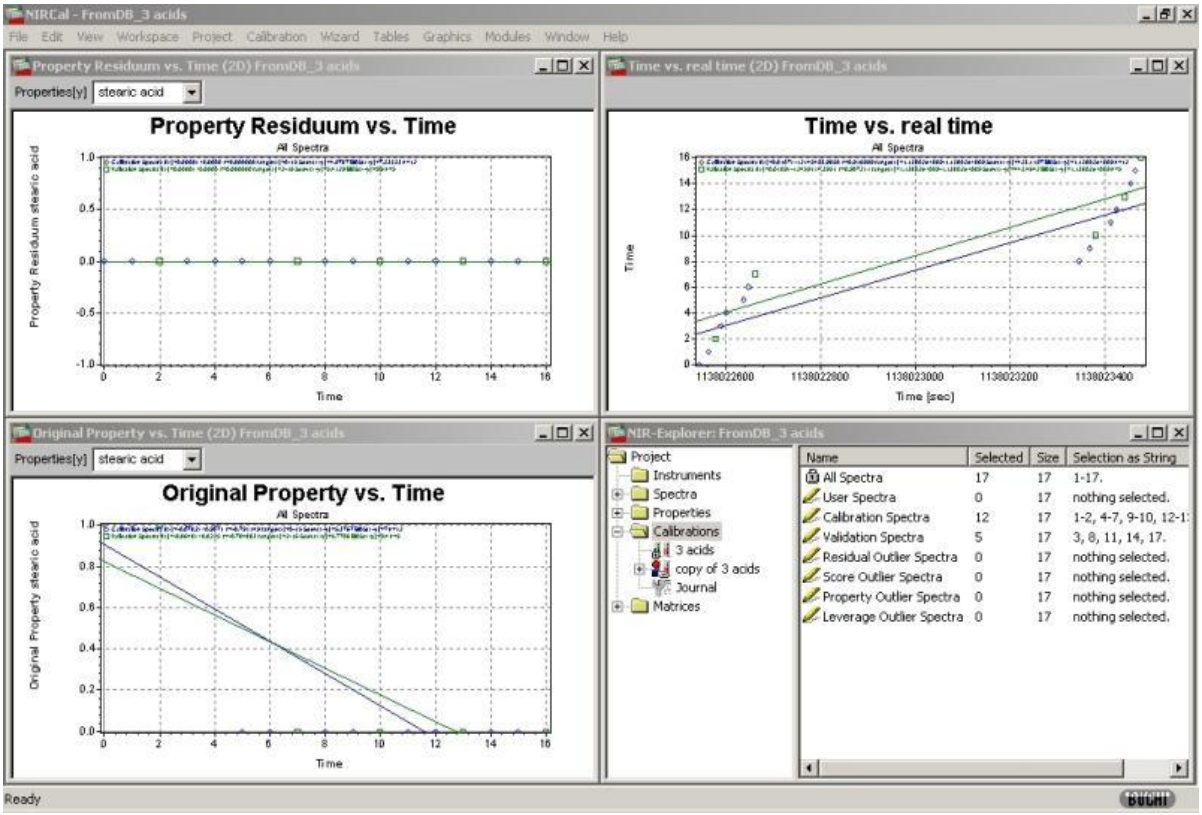
**Outlier Detection**

Example of an Outlier Selection Plot



### Time Dependency

Example of a Time Dependency Plot

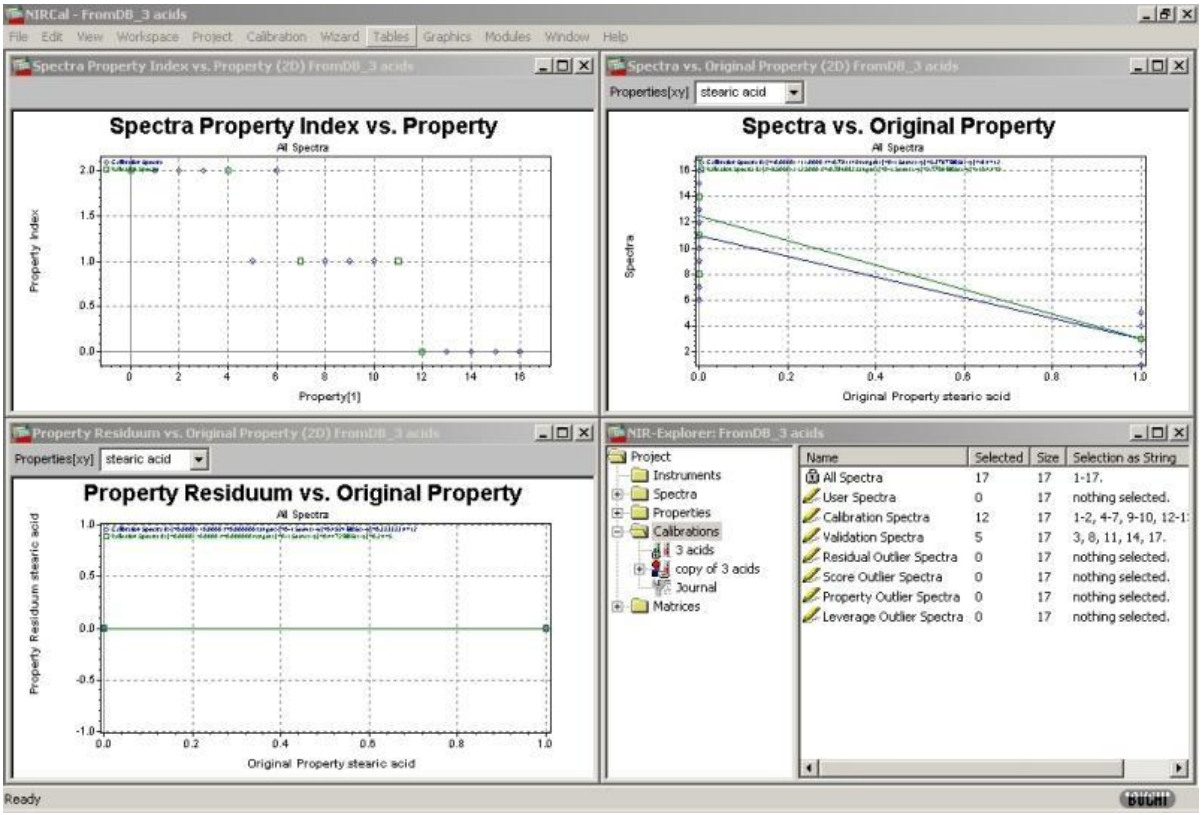



### Creator Dependency

### Instrument Dependency

### Property Dependency

Example of a Property Dependency Plot




### Custom

For customizing workspaces with the plugin System Module "WIZARD PLOTS". If no plugin is installed the menu item is gray.

**Toolbars**
Toggles the display of each toolbar. Same can be done using the right mouse in the toolbar region.


***Show all Toolbars***
Displays all toolbars.

***Hide all Toolbars***
Hides all toolbars.

***Status Bar***
Displays/hides the status bar at the bottom of the NIRCal window.

Depending on the active window or the called function the status bar contains different information.

**Toolbars save positions**
Saves the actual set of displayed toolbars and their position.

**Toolbars load positions**

Displays and repositions all toolbars saved as user toolbar positions.

See Toolbars save positions

## 3.20.4 Workspace

**Workspace**

It is possible to create ones own workspace. Open a set of graphics and/or protocols and arrange them as wished. Then Save the workspace.

It is also possible to save and load workspaces (1-3) with Icons in the Toolbar "Wizard Window"

**Save Workspace**

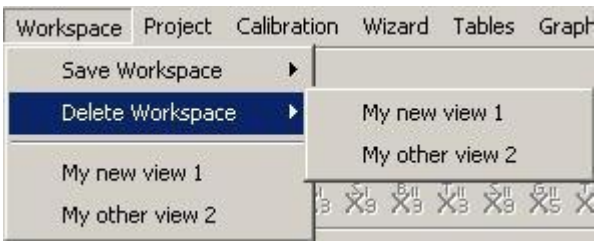
***New***

Creates a new workspace based on the actual opened windows.


**Delete Workspace**

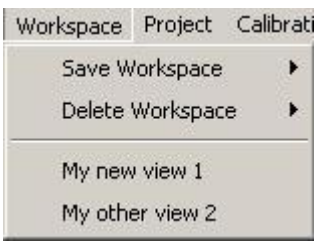
***List Workspace***

To delete a workspace select it from the list




**List Workspace**

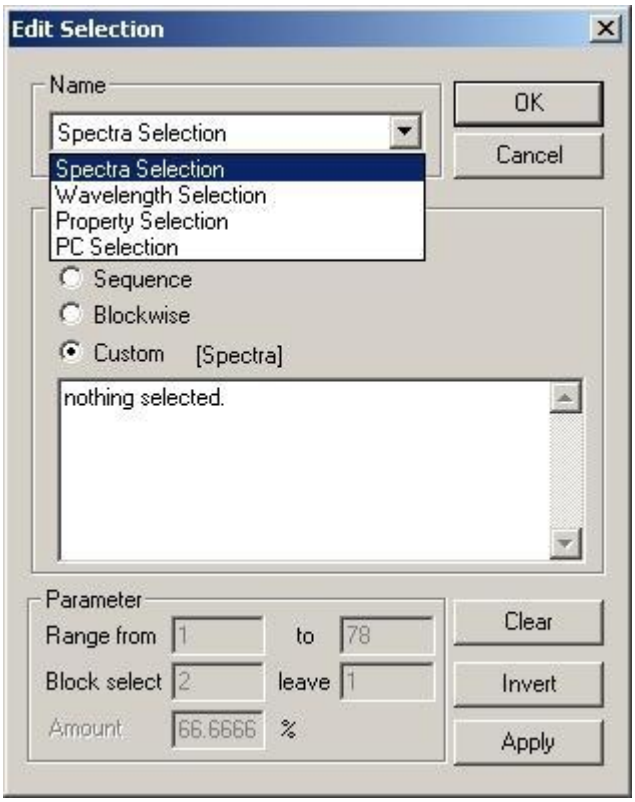
Select a workspace from the list:



## 3.20.5 Project

**Edit Selections**

 Opens the Dialog Box to edit selections.




**Load Selection from**

***Spectra C-Set***

 Load selection from the Spectra C-Set.

***Spectra V-Set***

 Load selection from the Spectra V-Set.

***Wavelengths C-Set***

 Load selection from the Wavelength C-Set.

***Properties C-Set***

 Load selection from the Properties C-Set.

***PCs C-Set***

 Load selection from the PCs C-Set.

**Change X-Unit to**

***Wavenumbers***


cm-1

***Wavelength***


nm

**Spectra**

***Add empty spectra***


Only possible in NIRCal file-mode.

***Delete selected spectra***


Deletes the red selected spectra from the project. To remove spectra from the calibration, they can be deleted in the calibration data set selection C-Set and V-Set.

**Property**

***Add property***


Only possible in NIRCal file-mode.

***Delete selected property***


Only possible in NIRCal file-mode.

**Journal**

***Show***

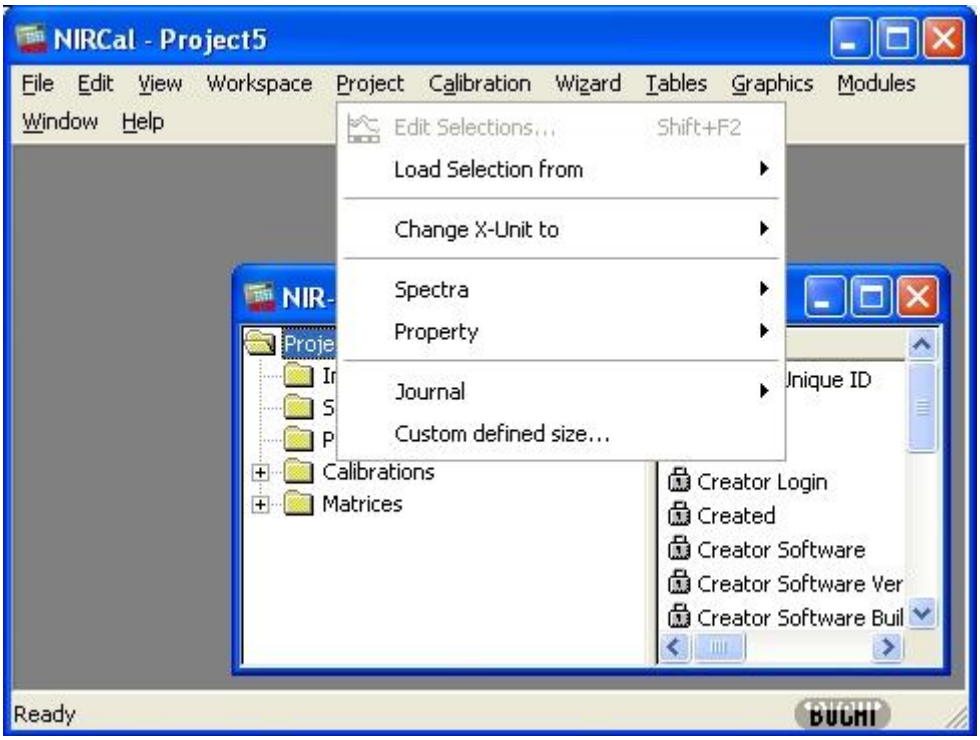

Displays the project journal.

***Logging Filter***

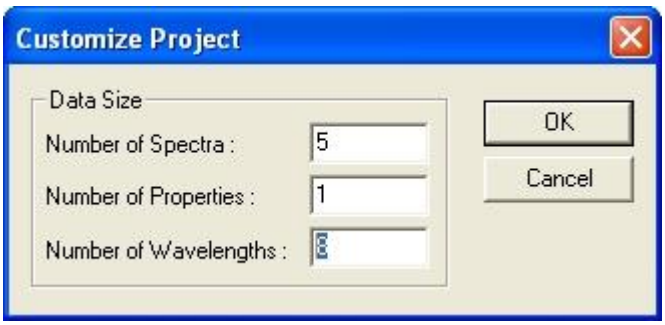
Opens Dialog to change the filter.

**Custom defined size**

It is possible to use NIRCal as a tool for multivariate analysis of any type of data. For this purpose the dimensions of the data matrices have to be defined and prepared for input.
Open a new project (Menu File / New) and select “Custom defined Size” in the Menu Project:

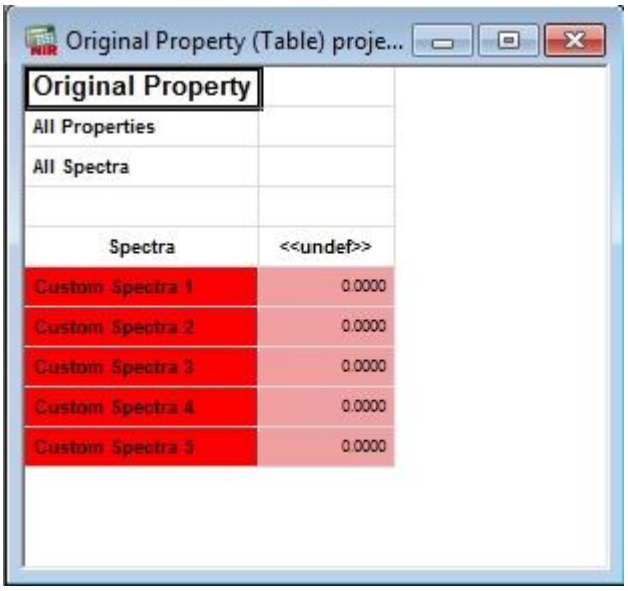


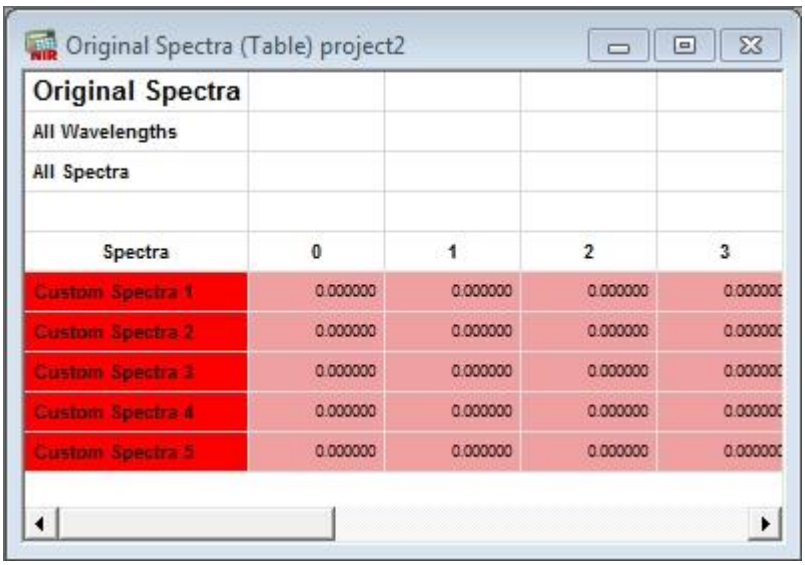

Define the size of your data in the appearing window:




***NOTE***
Since NIRCal normally is used for the analysis of spectra, the terminology might be unusual for the statistician. The following terms are equivalent:

Number of spectra = Number of samples
Number of Properties = Number of dependent variables (y-values)
Number of Wavelengths = Number of independent variables (x-values)
Once the dimensions of the data matrices have been entered, corresponding tables are opened and are available for edition:




The easiest way to enter data is via copy and paste (e.g. from EXCEL files).
Custom defined size: NIRCal works in file-mode which allows to resize the spectra and property afterwards as well. The given waves or variables are constant after Custom defined size.

***NOTE***
X-variables have to be entered in transposed format if the number of variables is smaller than the number of samples.
Under Modules / Settings Advanced.../ Custom Data, the transpose (tilt) of the property and spectra grid can be adjusted.

## 3.20.6 Calibration

**New**

Creates a new calibration.

**Calibration Wizard**

 Starts the Calibration Wizard.

For further details see: Calibration Wizard

**Default Parameter**


Opens the dialog to edit the calibration default settings.

For further details see: Calibration Defaults

**Method**

***Principal Component Regression***

Changes Calibration Method to Principal Components Regression

***Partial Least Squares Regression***



Changes Calibration Method to Partial Least Squares

***Multiple Linear Regression***



Changes Calibration Method to Multiple Linear Regression

***Cluster***



Changes Calibration Method to Cluster

***SIMCA***

Changes Calibration Method to SIMCA

**Validation Method**

***Validation Set***



Changes Validation Method to Validation Set

***Cross Validation***



Changes Validation Method to Cross Validation

**Parameter**

***Calculation***


Opens Dialog to Edit Calculation Parameters

***Blow Up Limits***


Opens Dialog to Edit Blow Up Parameter

***Lifecycle***

***Edit***

 Switches the lifecycle state of the calibration to **Editing.**
The Calibration-Icon changes  so that the calibration settings can be changed.

See Calibration Handling for details regarding the use of the lifecycle model on calibrations. More detailed general information regarding the NIRWare Lifecycle can be found in the NIRWare Software Manual.

***Save***

 Switches the lifecycle state of the calibration to **Saved.**
The Calibration-Icon changes to  so that the calibration settings are locked.

See Calibration Handling for details regarding the use of the lifecycle model on calibrations.

***Copy***

 Creates a copy of the calibration in the NIRWare Database. This enables the possibility to edit the copy of an **approved** Calibration.

See Calibration Handling for details regarding the use of the lifecycle model on calibrations.

***Next***

 Sets the calibration into the next lifecycle state. e.g. from **created to approved**.

See Calibration Handling for details regarding the use of the lifecycle model on calibrations. More detailed general information regarding the NIRWare Lifecycle can be found in the NIRWare Software Manual.

***Delete***

 Deletes a calibration from the lifecycle.

## 3.20.7 Wizard

**Calibration Wizard**

 Starts the Calibration Wizard.

For further details see: Calibration Wizard

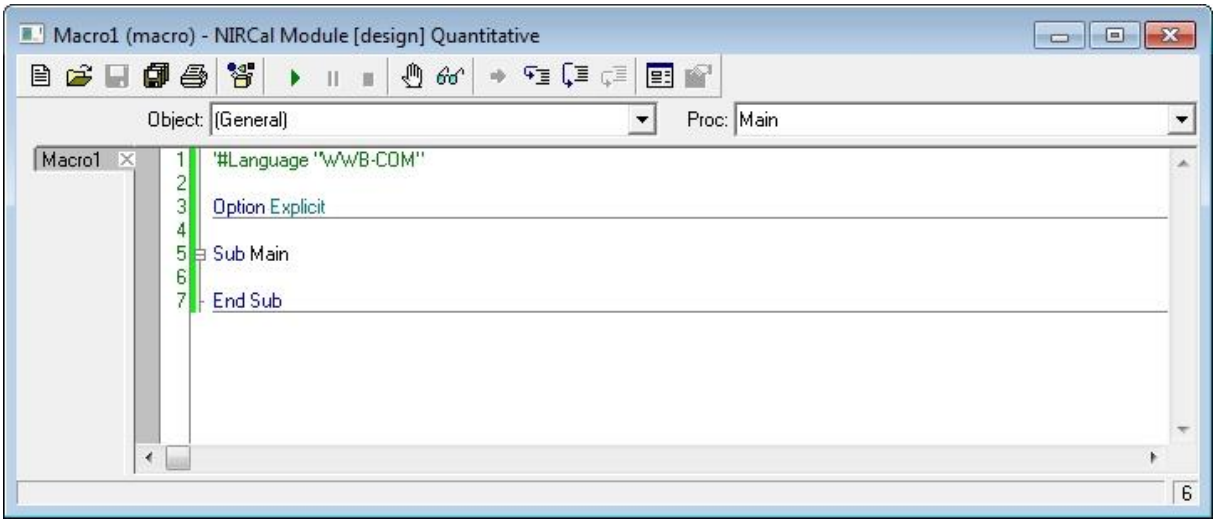
**Toolbox**

 Starts the Toolbox.

## 3.20.8 Modules

### Edit Module

Opens the macro-editor for entering BASIC macro code.




Click F1 for the Help File.

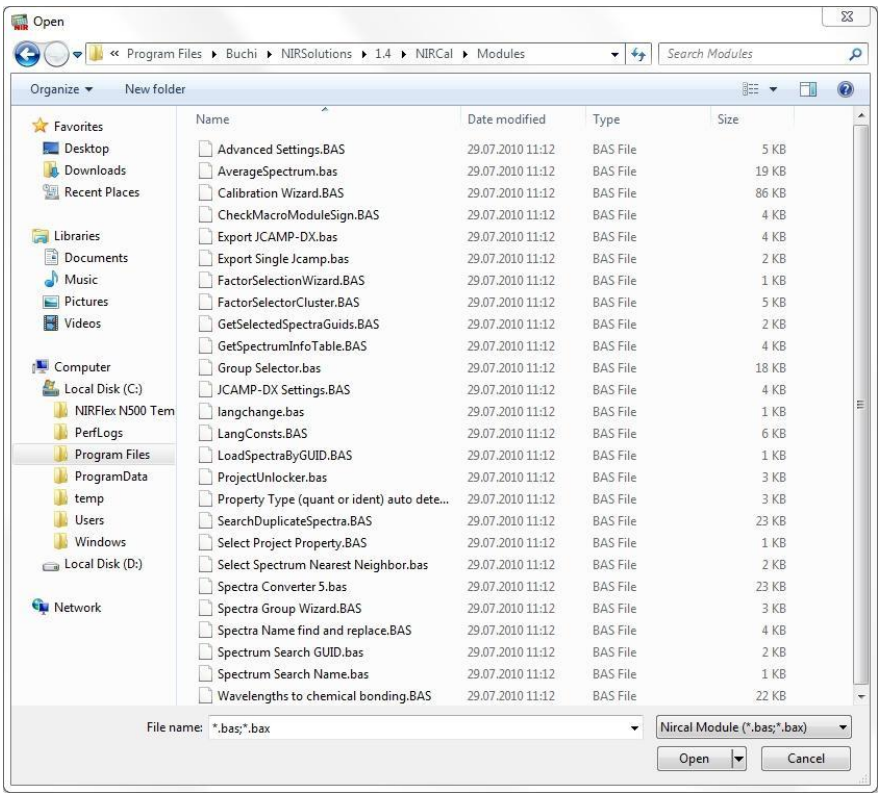
### Run Module

Opens the dialog to select NIRCal Module (*.bas) to be executed.

### Organize Menu

#### *Add*

Opens a dialog to select a module which will be added to the modules list. The modules are saved in the following folder: C:\Program Files\Buchi\NIRSolutions\Modules



***Edit***

Select a module from the module list to be opened with the Module Editor.

***Remove***

Select a module that shall be removed from the module list.

**Modules List**

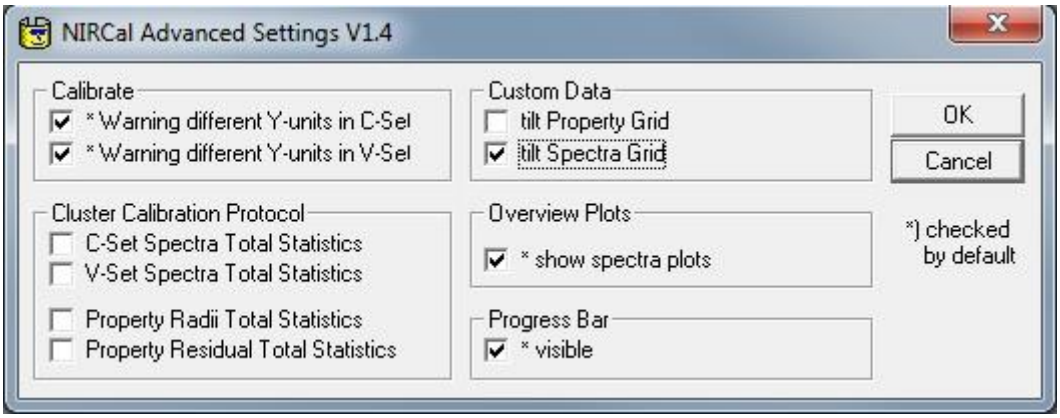
Default list of  modules. The list can be adapted by Organize Menu.

*Chemical Bonding*

Translates the wavelengths to chemical bondings. See mouse cursor position in the status bar.

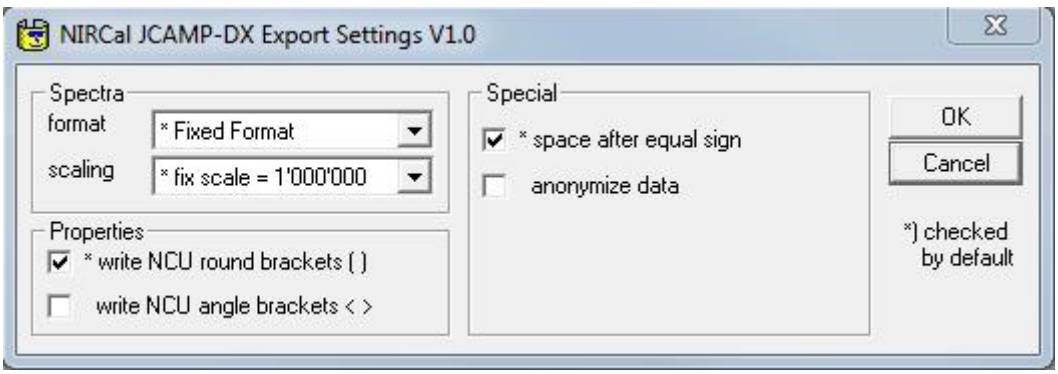
*Settings Advanced*

Opens the dialog to change advanced NIRCal settings.




If there are spectra with **different Y-units** (BCAP and NIRCal spectra in the same project), the warning should be activated (remove check).
The spectra plot can be switched off in the Overview plot (calculation is quicker for huge data set).

*Setting JCAMP-DX Export*




*Spectra Converter*

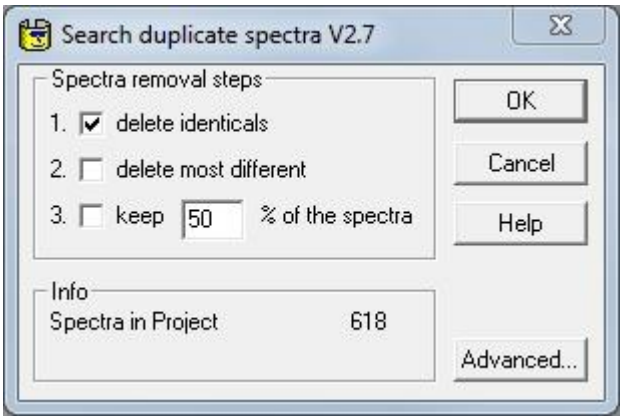
Converts spectra to the format of the NIRFlex N-500, e.g. spectra of NIRFlex-400 or NIRLab N-200. See detailed information under Convert and Import spectra from other instruments to a database.

*Spectra Duplicate Search*

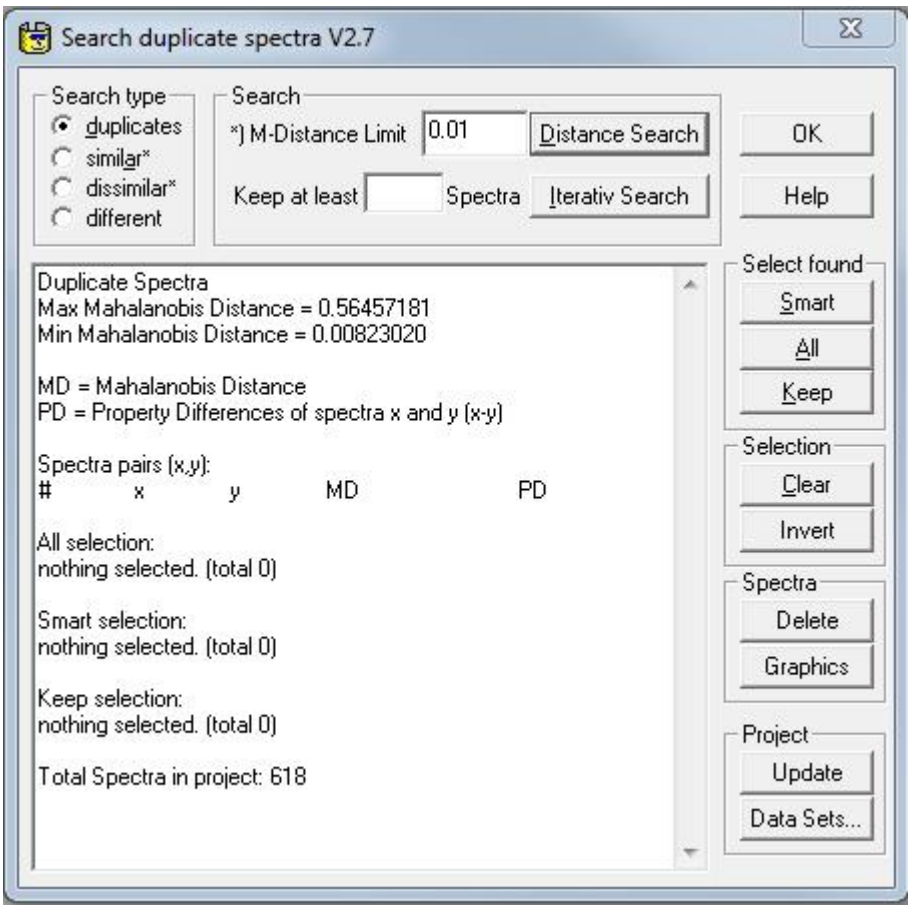
Spectra removal steps:

1. delete identicals: spectra pairs with a Mahalanobis distance of zero in secondary PCs space.
2. delete most different: from the spectra pair that has the highest distance, that spectrum with the higher leverage will be removed. Used for outlier removal.

3. keep x% of the spectra: an iterative similar search is used to find as much spectra to thin out the amount of spectra.




When you click Advanced, the following dialog opens:

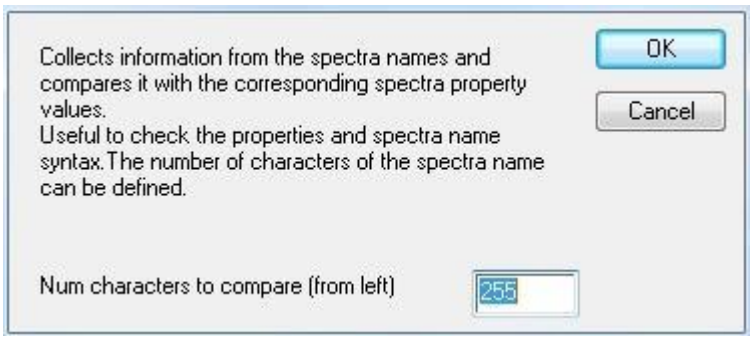


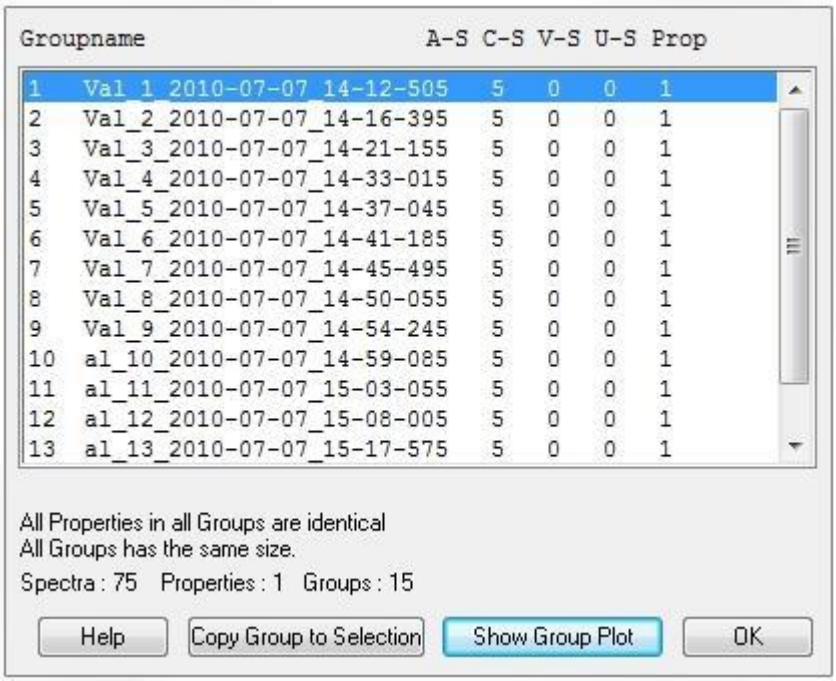

The following fields, areas and buttons are available:

Search type: defines the search logic.
Search: defines the search criteria and method.
Distance Search: The Mahalanobis Distance -MD- Limit depends on the data and the calibration settings. Each search trial shows the Min and Max Mahalanobis Distance to get an idea about the limit to choose.
Iterative Search: Often it is easier to define the number of spectra that should be found instead of an M-Distance Limit.
Select found (Smart, All, Keep): Selects the found spectra in NIRCal (red selection).
Selection (Clear, Invert): Helpful tool to manipulate the red selection.
Clear: Clears the selection.

Invert: Inverts the selection.
Spectra (Delete, Graphics)
Delete: Deletes selected spectra from the project. Typically the smart selection can be deleted to shrink the amount of calibration spectra.
Graphics: Opens a pretreated spectra plot to visualize the selection.
Project (Update, Data Sets...)
Update: Updates the calibration calculation. Used when data sets are edited or spectra are deleted.
Data Sets: Opens the Edit Data Sets dialog to change the calibration selections. Used to copy and paste textual selections from the search results.

*Spectra Group*




*Spectra Info Table to clipboard*


The spectra information can be copied into Excel.

*Spectra Nearest Neighbor Select*

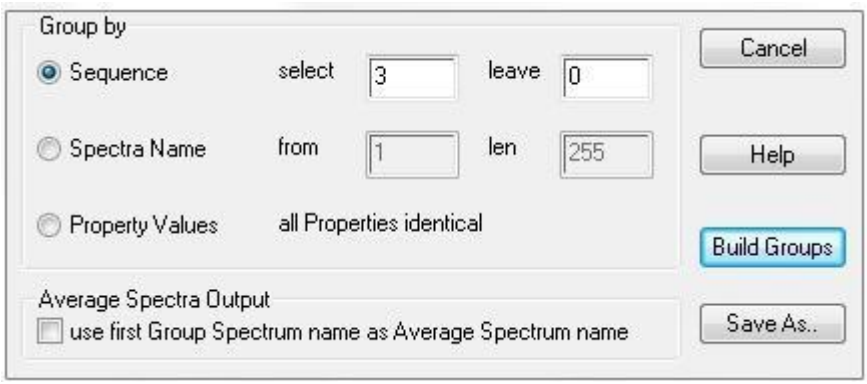
The Matrix "Spectrum Nearest Neighbor Index": (NIRExplorer / Matrices / "Spectrum Nearest Neighbor Index") contains the spectrum index of the nearest spectrum to each spectrum in the project. The best is to show it as 1D-Scatter-Plot for selected spectra. This matrix is related to "Spectrum Nearest Neighbor" which contains the Mahalanobis Distance between nearest spectra pairs. As referred in the ASTM E1655-05 "16.4.8 Nearest Neighbor Distance". **Note**: "Spectrum Nearest Neighbor Index" and "Spectrum Nearest Neighbor" depends on the secondary PC-selection.
**Note**: In most cases, the project has multiple repetitive measurements. Then the picture shows a nearly upward diagonal with some spikes. The reason is that repetitive measurements are the most similar and successively stay together.
**Note**: pair-wise distances are not always symmetric. A can be nearest to B and B can be nearest to C (and to A).

*Spectra Select by QL Property*

All spectra belonging to the marked qualitative property will be selected.

***NOTE***
These modules are installed but not added to the Modules menu.

*Average spectrum*




Average Spectra can be calculated and saved to file (nsf). The project is not affected, it is only the data source.
To define which spectra will be averaged they can be defined as so called Groups:

- **Sequence** block wise sequences
- **Spectra Name** same spectra name
- **Property Values** same property values

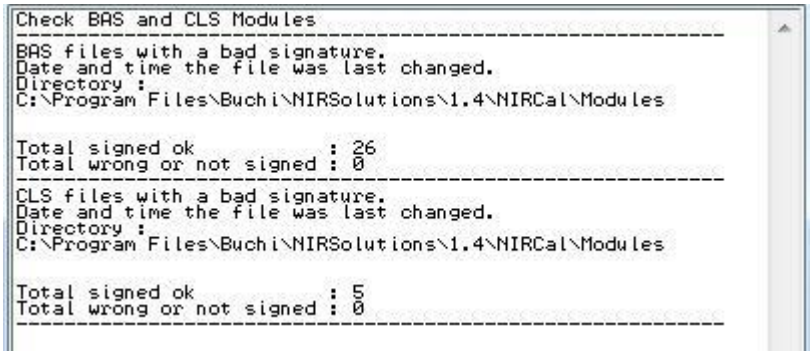
The stored Average Spectra contains the following additional information:

- Comment i.e. Average of 12 Spectra
- Description contains all the spectra GUID's of the averaged Spectra

Instrument Serial -1 as marking of a Average Spectrum

*Check Module Signature*

Choose the folder: C:\Program Files\Buchi\NIRSolutions\Modules



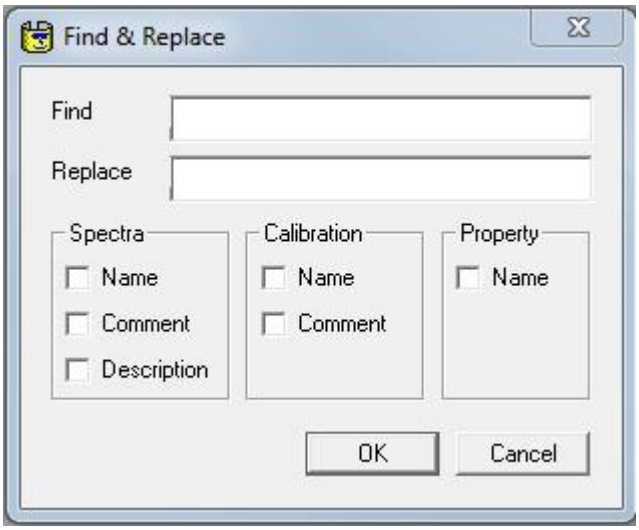

Can be used to check the software for original content.

Checks the Digital Signature in NIRCal Macro Modules. A detailed log lists all Modules that are not correctly signed or modified. The modification date/time is also listed.

*Spectra Name find and replace*

***NOTE***
This module is installed but not added to the Modules menu.

To replace the name is only possible in **File mode**, not allowed in projects stored already in the database.




## 3.20.9 Windows

**Cascade**
All windows will be opened in a cascade.

**Tile Horizontally**
All windows will be opened tiled horizontally.

**Tile Vertically**
All windows will be opened tiled vertically.

**Close All**
Closes all opened windows except the NIR-Explorer.

**Fit All Windows**
Fits the content of every window to default. E.g. the zooming factor.

Related to View Fit which fits only the content of the active window.

**Toggle Plot Size**
Toggles the plot size of every window.

Related to Large Plots which changes the plot size of only the active window.

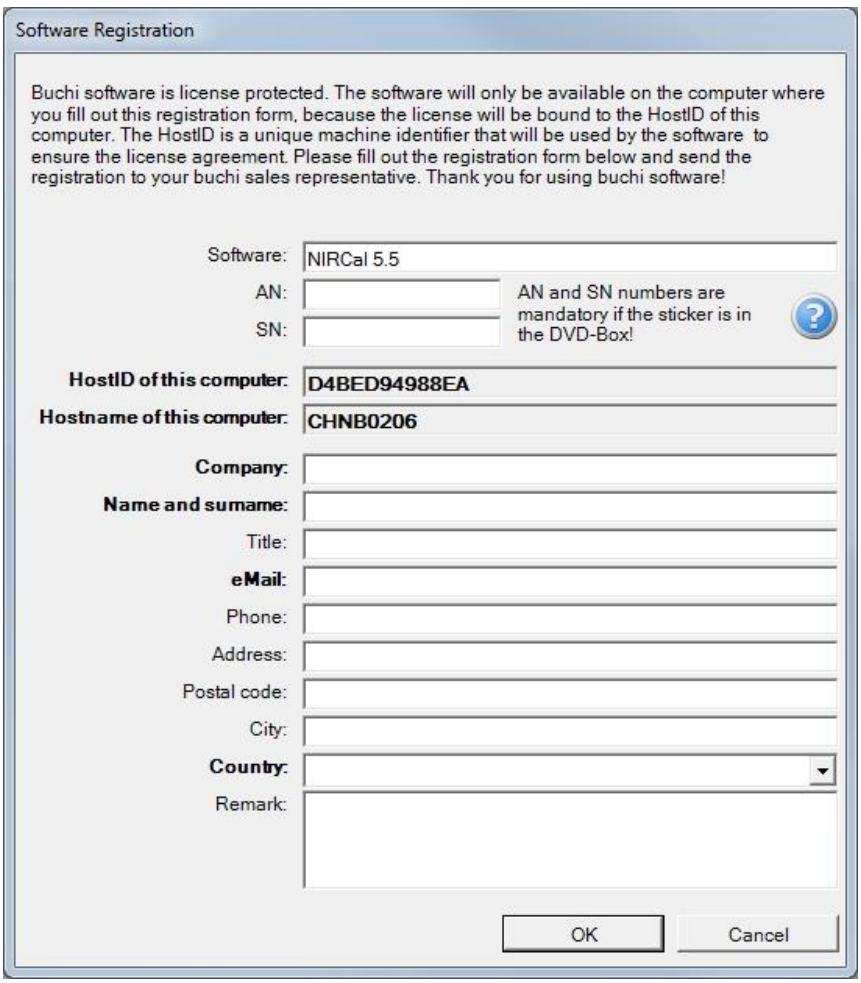
**Arrange Icons**

**Windows List**
List of all opened window. The active window is marked with a checkmark.

### 3.20.10 Help

**Index**
Opens this Online Help.

**Software Registration**
The registration file is created here. Please fill in all necessary input.

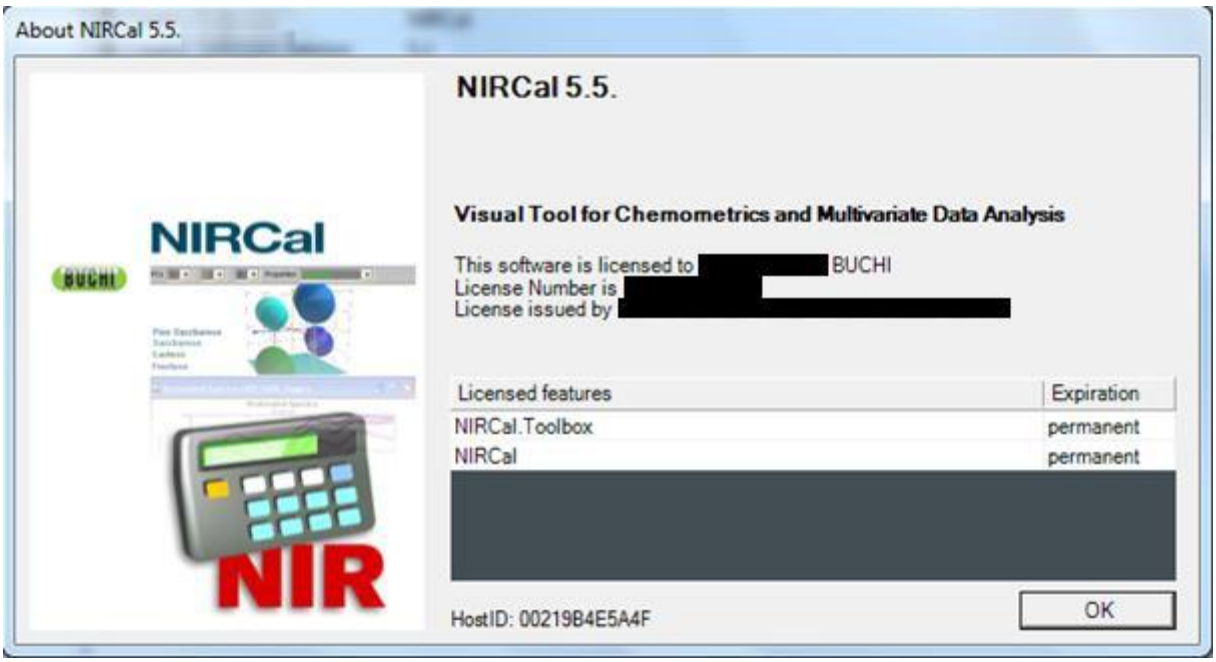


**Import License**

Import the licence file: “BuchiLic-550021-…” for NIRCal and “BuchiLic-550041-…” for NIRCal toolbox.

**About NIRCal BASIC**

**About NIRCal**




# 3.21 Context menu (right mouse)

## 3.21.1 Add Selection to Set

Add the objects currently selected in the plot to the chosen set.
CAUTION: This action will change the contents of your validation or calibration set if you add data to these sets.

## 3.21.2 Alpha-Blending

2D-Plots support Alpha-Blending for better visualization and review of a huge amount of data.

## 3.21.3 Anti-Aliasing

2D-Plots support Anti-Aliasing for better graphic quality.

## 3.21.4 Autorotate

| | |
|---|---|
| Menu: | right mouse / Autorotate |
| Keyboard shortcut | A |

Toggles the autorotation status of the plot On/Off. The plot will slowly rotate around the yellow axis.

This menu point is active in the 3D-Scatter or 3D-Surface plot.
Make sure this feature is turned off while working with a calibration, because the rotation needs a lot of processor time.

Tip: Zoom or rotate the coordinate system around its center while the left mouse button is pressed.

Keyboard:

| | |
|---|---|
| left / right | rotation around the y-axis |
| up / down | rotation around the x-axis |
| + / - | zoom in / zoom out |

### 3.21.5 Choose [x][y] Independent

Applies to 2D scatter plots only. This means that the value of each cell in one column or row of matrix 1 is plotted against the value of the same cell in any column or row of matrix 2. If you open the "Original Property" plot and transpose it you will be able to plot one property against another and check for dependencies.

Related Topic: Choose [xy] Independent

### 3.21.6 Choose [xy] Independent

Applies to 2D scatter plots only. This means that the value of each cell in one column or row of matrix 1 is plotted against the value of the same cell in the same column or row of matrix 2. In the "True against Predicted" plot this means that the same property is displayed on x and y axis.

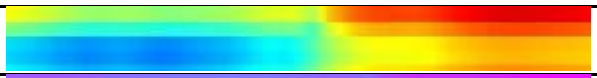
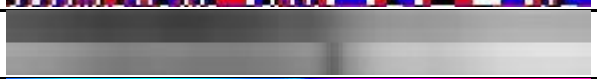
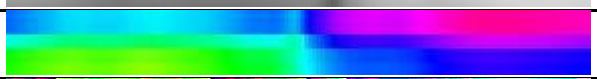
Related Topic: Choose [x][y] Independent

### 3.21.7 Clear Selection

Removes all objects from the current selection.
Changes to the current selection do not affect your calibration or validation set.

### 3.21.8 Colors

The following color models can be used with the top view plot and in 3D surface plot.

| | | |
|---|---|---|
| nirBone | Grayscale with a light touch of blue. |  |
| nirHot | Low values are black, intermediate red and high values are yellow |  |
| nirJet | Contrast enhancing color model. |  |
| nirCool | Low values are ice blue, intermediate blue and high values are pink. |  |
| nirCopper | Monochrome graduation with a coppertone. |  |
| nirFlag | Extremely contrast enhancing color model. |  |
| nirGray | Gray scale palette. |  |
| nirHSV | Contrast enhancing color model. |  |
| St.palette | |  |

### 3.21.9 Copy Plot to Clipboard

Copies the plot to the clipboard.

### 3.21.10 Copy Selection To

Delete the contents of the chosen set and replace it with the objects currently selected in the plot.
CAUTION: This action will change the contents of your validation or calibration set if you copy data to these sets.

Add Selection to Set, Load Selection from Set, Remove Selection from

### 3.21.11 Edit Selection

Opens the Edit Selection Dialog.
Note: Changes to the current selection do not effect your calibration or validation set.

### 3.21.12 Export Table

Export the table as Excel sheet.

### 3.21.13 Fix Table Titles

Fixes the row and table titles for better scrolling.

### 3.21.14 Flip X-Axis

Flips the plot x-axis. Draws the x-axis from lower to upper or vice versa.

### 3.21.15 Flip Y-Axis

Flips the y-axis. Draws the y-axis from lower to upper or vice versa.

### 3.21.16 Invert Selection

Toggles the selected objects to be unselected and vice versa. Use this feature if you have copied the selected objects into the calibration set and want to copy the now unselected objects into the validation set.
Changes to the current selection do not affect your calibration or validation set.

### 3.21.17 Load Selection from Set

The current selection will be replaced by the objects contained in the chosen set.
Changes to the current selection do not affect your calibration or validation set.

### 3.21.18 Large Plots

Menu: Options > Large Plots
Keyboard Shortcut L and Ctrl+L for all plots

The axis labels are not displayed and the graphic area of the plot is expanded.
To change this parameter for all plots at the same time use the "Toggle Plot Size" command from the "Window" menu.

### 3.21.19 Mouse Select

Handling the selection status of objects using the mouse.
Select one corner of the region containing the objects or part of the objects that you want to change the selection status and open a rubber banded rectangle keeping the left mouse button pressed.
When you release the mouse button the objects contained will change there status depending on the keyboard keys pressed:

Button status:

- no button: copy the object(s) to the selection the current selection is deleted
- shift: add the object(s) to the current selection
- control: toggle the selection status of the object(s)
- shift + control: remove the object(s) from the current selection

### 3.21.20 Mouse Select Lasso

2D-Scatter plots allow with Lasso to create free shape selection of the data. Keyboard: S.

### 3.21.21 Mouse X-Axis Select

Menu: Options > Mouse X-Axis Select
Keyboard : Q

Edit the selection displayed in the small window above the x-Axis.
The selection displayed in this window applies to the other side of the matrix (see Transpose). In a spectra plot the wavelength selection will be displayed in that window by default.
Select one corner of the region that corresponds to your x value and open a rubber banded rectangle keeping the left mouse button pressed. When you release the mouse button the x axis region contained will change its selection status depending on the keyboard keys pressed:

Button status:

- no button: copy the object(s) to the selection the current selection is
- deleted shift: add the object(s) to the current selection
- control: toggle the selection status of the object(s)
- shift + control: remove the object(s) from the current selection

### 3.21.22 Mouse Zoom

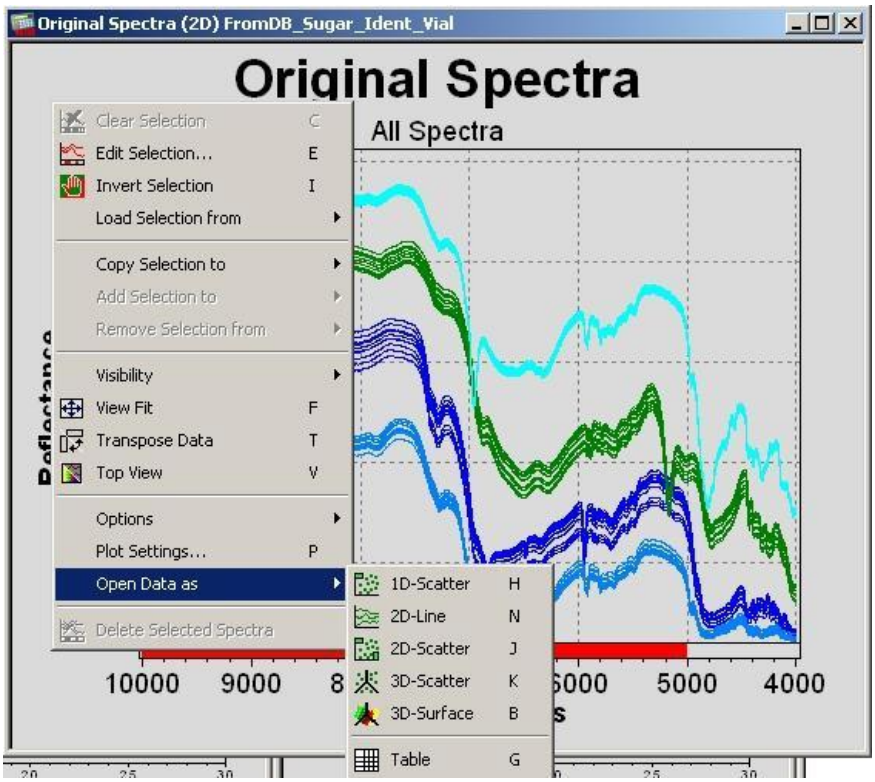
Zoom into a plot using the mouse. Keyboard: Z
Select one corner of the detail that you want to enlarge and open a rubber banded rectangle keeping the left mouse button pressed. When you release the mouse button the plot will be updated with the new view rectangle.

### 3.21.23 Number format

Adjust number of decimal places.
Default depends on the matrix shown.

### 3.21.24 Open data as

Whilst display of a plot it is possible to switch to another visualization type:

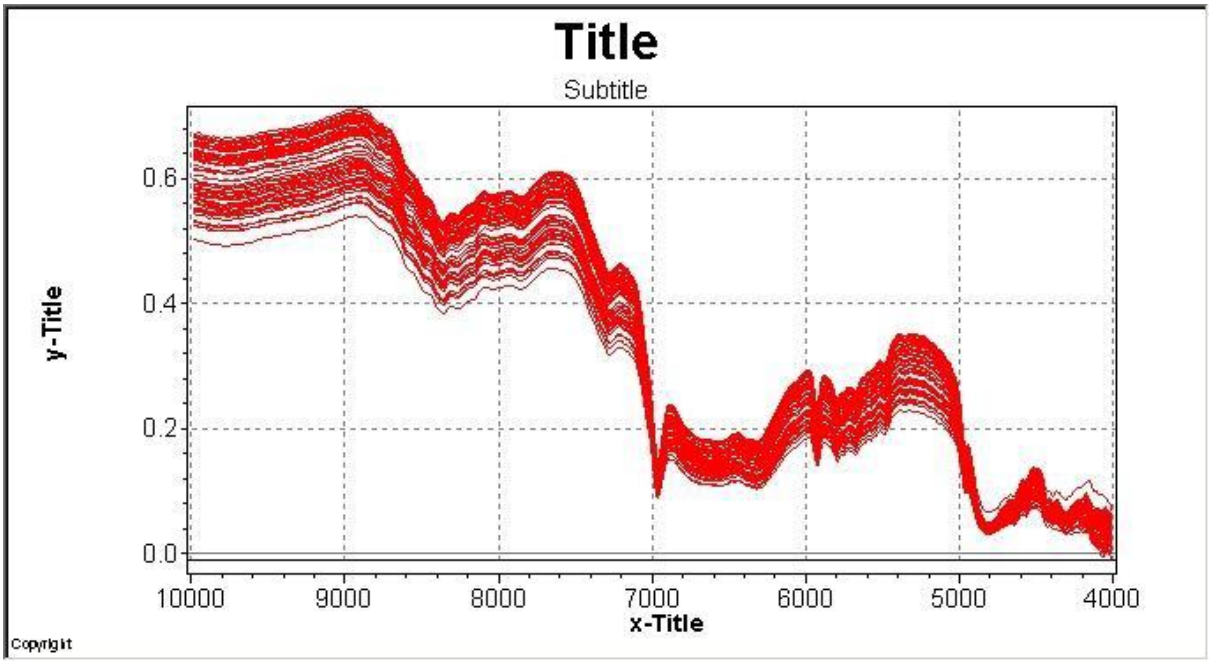


### 3.21.25 Plot Defaults

The Nircal Configuration dialog will be opened. The default parameters for all views can be edited in that dialog. See also 2D-Plots.

### 3.21.26 Plot Settings

The titles and zoom of a plot can be edited before printing.




### 3.21.27 Regression Draw Line

A regression line will be drawn for each set that you activate here. The corresponding regression formula will be displayed as a legend.

### 3.21.28 Regression Copy Equation

The regression formula will be copied to the clipboard and can be pasted to any text or spreadsheet application.

### 3.21.29 Remove Selection from

Add the objects currently selected in the plot to the chosen set.

***NOTE***
This action will change the contents of your validation or calibration set if you copy data to these sets.

### 3.21.30 Snap to Y-Grid

The scale of the y axis is adjusted in such a way, that the upper and lower corners are labeled.
The default for this value can be set in the 2D-plot Tab of the Nircal Configuration Dialog.

### 3.21.31 Snap to X-Grid

The scale of the x axis is adjusted in such a way, that the left and right corners are labeled.
The default for this value can be set in the 2D-plot Tab of the Nircal Configuration Dialog.

### 3.21.32 Show Plane

Menu: Options > Show Plane
Available in Plot : 3D-Scatter

Planes for all 3 dimension can be displayed:

- Blue and Yellow Axis, X-Plane
- Red and Blue Axis, Y-Plane
- Red and Yellow Axis, Z-Plane

### 3.21.33 Show Labels

Show numbers at all points of the plot.

### 3.21.34 Show Line

Show Scatter as Line.

### 3.21.35 Show Selection Colors

Line Coloring by Set.

### 3.21.36 Show Cycle Colors

The color table is used to color each object. When the color table ends, the colors cycle start from the start of the color table again.

### 3.21.37 Show Property Colors

If the spectra have qualitative properties the coloring shows the property relation.

### 3.21.38 Transpose Data

Toggles between the two possibilities of connecting the values of a matrix with lines. If you look at a spectra plot this would mean that NirCal either connects all wavelengths for each spectrum or all spectra for each wavelength.

This powerful feature allows you to select certain wavelengths, just as you would select spectra.

### 3.21.39 Turbo Mode

Menu: Options > Turbo Mode
Keyboard : M

Speeding up the display.
When large quantities of spectra are displayed, it is advisable to change the display mode of the view to "turbo". A defined number of wavelengths will be skipped.
To edit the number of wavelengths to be skipped use the 2D-Plot parameter dialog.

### 3.21.40 Undo Last Pretreatment

Menu: Undo Last

Use
Remove the last pretreatment from the list of pretreatments attached to the active calibration.

### 3.21.41 Undo Sequence of Pretreatments

Menu: Undo Sequence

Use
Remove all pretreatments from the list of pretreatments attached to the active calibration.

### 3.21.42 View Fit

The current plot is resized in such a way, that all objects are inside the visible region.

If you have applied a visibility (e.g. calibration set) to the view, it is possible that certain objects not contained in that set (e.g. validation spectra) will not be visible.

### 3.21.43 Visibility

Restrict the visibility of the objects displayed in the plot to the chosen set.
REMARK: no data will be changed or deleted by this action.

***NOTE***
Use the shortkeys Ctrl+0,Ctrl+1,...,Ctrl+7 to change between the sets. The visibility set is shown as subtitle in the plot.

## 3.22 Toolbar

### 3.22.1 Calibration Toolbar


Updates opened views and recalculates,

Changes Calibration Method to Multiple Linear Regression

Changes Calibration Method to Principal Components Regression

Changes Calibration Method to Partial Least Squares

Changes Calibration Method to Cluster

Changes Calibration Method to SIMCA

Changes Validation Method to Validation Set

Changes Validation Method to Cross Validation


Opens Dialog Box to Edit Data Sets


Opens Dialog Box to Edit Cross Validation Groups


Creates a spectra selection


Creates a wavelength selection


Creates a property selection


Opens Dialog PCs Selection Wizard (Factor)


Opens Dialog to edit Calculation Parameters


Opens Dialog to edit Blow Up Parameter


Opens Dialog for Outlier Detection


Calculates Q-Value and shows Protocol


Opens Dialog Box for the Calibration Wizard


Shows the saved Calibration Protocol

 Calculates the active calibration and shows the Full Calibration Protocol

 Predict Project

 Application Predict Project

 Multipredict Project

 Predict external

 Application predict external

 Multi Predict external

 Predict external GUID file

 Multi Predict external GUID file

 Application Predict external GUID file

### 3.22.2 Database Toolbar

 Open NIRCal Project from Database

 Save NIRCal Project to Database

 Import all Spectra belonging to this Projects Applications

 Assign Applications to the Project

 Search and Import Spectra from Database

 Delete a project from the database

### 3.22.3 File & Edit Toolbar

 Creates a new NIRCal Project

 Loads a project from the database

 Saves NIRCal Project

 Cut selected data

 Copy selected data

 Paste selected data

 Print

 Info / About NIRCal

### 3.22.4 Lifecycle

See Calibration Handling for details regarding the use of the lifecycle model on calibrations. More detailed general information regarding the NIRWare Lifecycle can be found in the NIRWare Software Manual.

 Edit calibration

 Save calibration

 Copy calibration

 Next state

 Delete calibration

### 3.22.5 Modules

 Edit module

 Run module

### 3.22.6 Options Toolbar

 Opens Dialog to change NIRCal Configuration

### 3.22.7 Pretreatments Toolbar

 Normalization by Closure

 Normalization by Maxima

 Normalization between 0 and 1

 Normalization to Unit Length

 Normalization by Sdev

 Smoothing Average 3 points

 Smoothing Average 9 points

 Smoothing Savitzky-Golay 9 points

 1st derivative BCAP 5 points

 1st derivative Taylor 3 points

 1st derivative Savitzky-Golay 9 points

 2nd derivative BCAP 3 points

 2nd derivative Taylor 3 points

 2nd derivative Savitzky-Golay 9 points

 2nd derivative Taylor 3 points Segment5 Gap5 3rd

 derivative Taylor 5 points


Normalization Multiplicative Scatter Correction Full


Normalization Multiplicative Scatter Correction Amplification


Offset Multiplicative Scatter Correction


Normalization Standard Normal Variate


Normalization Variance Scaling


Offset Mean Centering


Transformation Kubelka Munk


Second Derivative / Logarithm


Transformation Absorbance


Transformation Absorbance inverse


Undo Last


Undo Sequence

### 3.22.8 Pretreatments Advanced Toolbar


Squares each data point in the spectra


Divides one through each data point: Reciprocal 1/s


Subtract DC. Uses C-Set Selection


Shifts negative offset of each spectra to zero. Uses C-Set Selection


Add a Constant to each datapoint


Subtract Spectrum


Divide by Spectrum


Linear Filter

### 3.22.9 Pretreatments Minibar Toolbar


Normalization to Unit Length


Normalization between 0 and 1


Smoothing Average 3 points


1st derivative BCAP 5 points


1st derivative Savitzky-Golay 9 points


2nd derivative Savitzky-Golay 9 points


Multiplicative Scatter Correction Full


Normalization Standard Normal Variate


Offset Mean Centering


Transformation Absorbance


Linear Filter


Undo Last


Undo Sequence


Fits the hole plot to the window


Opens Pretreated Spectra Graphic in new window


Opens Original Spectra Graphic in a new window


Opens Original Property Table in a new window

### 3.22.10 Pretreatments Gap2


Smoothing Average 3 Points Gap2


Smoothing Average 9 Points Gap2


Smoothing Savitzky-Golay 9 Points Gap2


1st Derivative BCAP 5 Points Gap2


1st Derivative Taylor 3 Points Gap2


1st Derivative Savitzky-Golay 9 Points Gap2


2nd Derivative BCAP 3 Points Gap2


2nd Derivative Taylor 3 Points Gap2


2nd Derivative Savitzky-Golay 9 Points Gap2


2nd Derivative Taylor 3 Points (Segment5 Gap5) Gap2


3rd Derivative Taylor 5 Points Gap2


Undo Last


Undo Sequence

### 3.22.11 Project Toolbar


Opens Project Journal


Change X-Unit to [$cm^{-1}$]


Change X-Unit to [nm]


Add spectra to Project


Delete selected spectra


Add property to project


Delete selected property


Opens Edit Selection Dialog


Load selection from C-Set spectra


Load selection from V-Set spectra


Load selection from C-Set wavelengths


Load selection from C-Set properties


Load selection from C-Set factors

### 3.22.12 Search Import Export Toolbar


Import spectra file to project.


Export spectra from project to file.


Search spectra in file and import to project.

### 3.22.13 Window Toolbar


Fit all windows


Cascade all windows


Tile all windows horizontally


Tile all windows vertically


Close all windows, except NIRCal-Explorer

### 3.22.14 Wizard Toolbar


Starts the Calibration Wizard.


Opens Dialog to Edit Calculation Parameters


Toolbox

### 3.22.15 Wizard Workspace


Shows an overview of the most significant plots


Shows plots for spectra selection


Shows plots for wavelength selection


Shows plots for property selection


Shows plots for primary PCs (factor) selection


Shows plots for secondary PCs (factor) selection


Shows plots for outlier detection


Shows time dependency plots


Shows creator dependency plots


Shows instrument dependency plots


Shows property dependency plots


Loads all plots saved as workspace 1


Loads all plots saved as workspace 2


Loads all plots saved as workspace 3


Saves all opened plots as workspace 1


Saves all opened plots as workspace 2


Saves all opened plots as workspace 3

### 3.22.16 Zoom & Select Toolbar


Zoom window (Z)


Data selection (S)


Data selection (S), Lasso selection in 2D-Scatter plots


x-Axis selection (Q)


View Fit (F)


Large plot (L)


Snap x-Axis (X)


Snap y-Axis (Y)


Flip x-Axis


Flip y-Axis


Transpose data (T)


Top view (V)


Front view (V)


2-D Line plot (N)


1-D Scatter plot (H)


2-D Scatter plot (J)


3-D Scatter plot (K)


3-D Surface plot (B)


Table grid (G)


Clear selection (C)


Invert selection (I)


Edit selection (E)


Line coloring by set (1)


Line coloring by cycle (2)


Line coloring by property (3)


Copy selection to User set

Visibility to all sets (Ctrl+0)

Visibility to user set (Ctrl+1)

Visibility to calibration set (Ctrl+2)

Visibility to validation set (Ctrl+3)

Visibility to residual outlier set (Ctrl+4)

Visibility to loading outlier set (Ctrl+5)

Visibility to property outlier set (Ctrl+6)

Visibility to leverage outlier set (Ctrl+7)

# 4 Appendix

## 4.1 Keyboard Shortcuts

To following keys are used together with the NIRCal selections:

| Button | Status |
|---|---|
| no button | copy the object(s) to the selection the current selection is deleted |
| shift | add the object(s) to the current selection |
| control | toggle the selection status of the object(s) |
| shift + control | remove the object(s) from the current selection |
| A | 2D: Anti-Aliasing (a), 3D: Autorotate |
| B | 2D: Alpha-Blending (b), 3D: Show Box |
| C | Clear Selection |
| E | Edit Selection |
| F | View Fit |
| G | View Data as table |
| H | View Data as 1D-Scatter plot |
| I | Invert Selection |
| J | View Data as 2D-Scatter plot |
| K | View Data as 3D-Scatter plot |
| L | Large Plot |
| N | View Data as 2D-Line plot |
| P | Plot Settings |
| Q | Mouse X-Axis Select |
| R | 2D: Show Labels (r), 3D: Mouse Rotate |
| S | Mouse Select, Mouse Select Lasso |
| T | Transpose data |
| V | Toggle between Top View and Front View |
| W | Show Line |
| Y | Snap to Y-Grid |
| X | Snap to X-Grid |
| Z | Mouse Zoom |
| 0 | Show Legend |
| 1 | Show selection Colors |
| 2 | Show Cycle Colors |
| 3 | Show Property Colors |
| Alt+M+N | Start "Spectra &Nearest Neighbor Select" module |
| Ctrl+0 | Visibility All-Set |
| Ctrl+1 | Visibility User-Set |
| Ctrl+2 | Visibility Calibration-Set |
| Ctrl+3 | Visibility Validation-Set |
| Ctrl+4 | Visibility Residual Outlier Spectra |
| Ctrl+5 | Visibility  Score Outlier Spectra |
| Ctrl+6 | Visibility Property Outlier Spectra |
| Ctrl+7 | Visibility Leverage Outlier Spectra |
| Ctrl+A | Select all |
| Ctrl+C | Copy selection to clipboard |
| Ctrl+E | Edit module |
| Ctrl+F | Fit all windows |
| Ctrl+L | Toggle Plot Size |
| Ctrl+N | New Project |
| Ctrl+O | Open Project from file |

| Button | Status |
|---|---|
| **Ctrl+r** | **Show full labels** |
| Ctrl+S | Save Project |
| Ctrl+T | Fix Table titles during scrolling |
| Ctrl+V | Paste selection |
| Ctrl+X | Cut selection |
| Ctrl+Z | Undo |
| Ctrl+Shift+M | Import latest Application spectra from DB |
| Ctrl+Shift+S | Save Project |
| Ctrl+Alt+R | Run module |
| F5 | Update |
| F8 | Show Calibration Protocol |
| Ctrl+F5 | Overview Plot |
| Shift+F5 | Calculate all |

## 4.2 Tips and Tricks

1. Fixing table titles
2. Plot / Graphic zooming
3. Zooming in and out
4. NIR-Explorer show all entries
5. Workspaces
6. working with Selections and Visibility (Part 1)
7. working with Selections and Visibility (Part 2)
8. working with Selections - Merge, Intersect, Exclusive (Part 3)
9. working with Selections (Part 4)
10. working with Selections (Part 5)
11. customize derivatives and smooting pretreatments
12. working with Selections - "Undo" (Part 6)

**Fixing table titles**
Open a table, i.e. Property Table, press Ctrl-T.
This will fix the table titles and header while scrolling down a large table.

**Plot / Graphic zooming**
After a zoom-in or a pretreatment change, the plot can be zoomed to fit all the data in the view with the keyboard quick key F (View Fit).
If more than one plot has to be fitted, then use Ctrl+F which fits all plot views. It can be found in the menu Window->Fit All Windows.
Also there is the Toggle Plot Size. If many Views are open the Ctrl+L can also be helpful to get more place on the screen for the data.

**Zooming in and out**
You can zoom plots and tables in/out by holding the Ctrl key down and scroll up/down on the mouse.
By the way this trick also works in other applications in a similar way! (Word, Excel, Internet Explorer...)

**NIR-Explorer show all entries**
To open all collapsed entries in the NIR-Explorer tree, click on the tree's root on "Project" and press the * (asterisk) key on the numberpad.

**Workspaces**
When you want to see a set of other interesting plots, you can customize it. Simply open as many plots and tables as you like and arrange them in your favourite position on the screen. Try out and explore new plots by viewing Matrices from the NIR-Explorer.
Save your customizing with View->Save Workspace Save->New... enter i.e. "quantitative overview".
Load it (project independent) with View->Load Workspace-> ...

To have faster access to your workspaces, there are 3 custom workspace icons on the toolbar "Wizard Workspace". L1,L2,L3 for loading it and S1,S2,S3 for saving it before.

**Working with Selections and Visibility (Part 1)**
In the NIR-Explorer -> Calibrations -> (any non-persistent calibration) -> Data Sets there are 4 selection set types:

- Spectra
- Wavelengths
- Properties
- PC

In each NIRCal plot, below the plot title, the selection name that defines the visibility of the data displayed is visible. Typically the "All"-Set is used as default, i.e. "All Spectra" in the "Original Spectra" plot.

In the plot context menu, choose Visibility->"Calibration Spectra" to change the visibility for that plot to the C-Set spectra only. The plot subtitle changes immediately to "Calibration Spectra" and all non C-Set spectra disappear from in the plot.

You can open the same plot multiple times and change the visibility individually for each plot to any selection.

As mentioned in Hint#5, you can save them to your own workspaces.

**Working with Selections and Visibility (Part 2)**
(please read Part 1 first)
Now we want to get a visibility, for example the first 10 spectra in the project. How do we do that ?

Have a look at the available Spectra Selections in the NIR-Explorer:

- All Spectra : a selection for ALL the spectra
- User Spectra : that is the users notepad for spectra selections
- Calibration Spectra : C-Set Spectra (Calibration Input)
- Validation Spectra : V-Set Spectra (Calibration Input)
- Residual Outlier Spectra : (Calibration Output)
- Loading Outlier Spectra : (Calibration Output)
- Property Outlier Spectra : (Calibration Output)
- Leverage Outlier Spectra : (Calibration Output)

To create a selection that contains the first 10 spectra, we can edit the "User Spectra" by double clicking on "User Spectra"
Note: Only the actual active calibration can be edited.
In the "Edit Selection" dialog choose "Custom", press "Clear" and overwrite "nothing selected" with "1-10", then press "apply" to see if your input is accepted. Press OK, go to the "Original Spectra" plot and set Visibility->User Spectra.

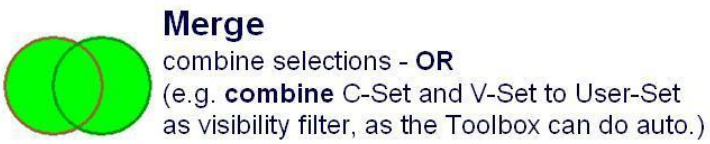
That's it. You will see only the first ten spectra from our project.

By the way, Visibility settings are also possible in Table views!

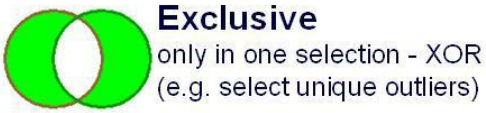
**Working with Selections - Merge, Intersect, Exclusive (Part 3)**
(please read Part 1 and 2 first)
How to create complex selections by merging, inverting and intersections.

NIR-Explorer -> Calibrations -> (any non-persistent calibration) -> Data Sets -> Spectra
On the right side of the NIR-Explorer several items can be selected, i.e. selection of "User Spectra" and "Calibration Spectra". Hold down the Ctrl key while selecting or deselecting with the right mouse button. Then the context menu will have more options to copy the selection i.e. under the menu "Copy to Selection".
There are three set calculation methods available (if more than one set is selected):

- Merge : combines sets
- Intersect : keeps only the overlapping of the sets
- Exclusive : keeps only the non overlapping of the sets








"Copy to Selection" means that the result of the set calculation is copied or added by "Add to Selection" or removed by "Remove from Selection" to the selection.

The "Selection" is the temporarily visualised red selection in the plots and tables.
This selection is not available in the NIR-Explorer, it is only temporarily visualised and can be used for calculations with selection sets. In the plot context menu there are functions to work with that selection:

- "Clear Selection C"
- "Invert Selection I"
- "Edit Selection E"
- "Load Selection from->"

**Working with Selections (Part 4)**
Each matrix has two selection types, i.e. "Original Spectra" owns the "Spectra Selection" and the "Wavelength Selection". The selection currently used in a plot is defined by Option->Mouse "Mode":

- "Mouse Select S" -> i.e. "Original Spectra"
- "Mouse X-Axis Select Q" -> i.e. "Wavelength Selection"

Let's try this key sequence slowly to see the effects on the selection. Watch what happens when you invert a selection.
Starting on "Original Spectra" plot and press the quick keys:

1. I : to invert the selection
2. I : to invert the selection
3. Q : to work on the X-Axis
4. I : to invert the (other) selection
5. I : to invert the (other) selection
6. S: to work on the "lined"-data
7. I : to invert the selection
8. I : to invert the selection

Note: You can freely use "Selection" and "User Selection" for working with the selection sets with no impact to the calibration results. This works with all selection types, also for the Wavelength selections.

**Working with Selections (Part 5)**
To edit all the selections in one dialog, go to main menu Calibration->Change Data Sets->Edit Data Sets... or get the same thing via the Calibration Toolbar "Edit Data Sets".
This is the same dialog as in the Calibration Wizard "Data Sets..." In that dialog the selection can be chosen from the name drop down list. So it's easy to copy and paste selection text in the custom mode (default for that dialog) between the selections.

**Customize derivatives and smoothing pretreatments**
The pretreatment "Savitzky Golay 2nd derivative (15 pt. 2nd order)", as we can see in [1] that the coefficients for 2nd derivative 15 points are:
91,52,19,-8,-29,-48,-53,-56,-53,-48,-29,-8,19,52,91and the norm scale is: 6188

***NOTE***
Due to errors in that [1] paper, use this coefficients:
91,52,19,-8,-29,-44,-53,-56,-53,-44,-29,-8,19,52,91

-44 instead of -48
and the norm scale is: 6188

To use this in the **Linear Filter** pretreatment enter this :
**91,52,19,-8,-29,-44,-53,-56,-53,-44,-29,-8,19,52,91,6188**
that is in the form
oefficient1, coefficient2, ... , coefficient n , norm scale

In the Calibration Protocol the customized coefficients for Linear Filter pretreatment are listed.
To edit and change the coefficients, simply remove the pretreatment and add Linear Filter again.

References:
[1] A. Savitzky and M.J.E. Golay "Smoothing and differentiation of data by simplified least squares procedures" Anal. Chem., 36, 1627-1639 (1964)

**Working with Selections - "Undo" (Part 6)**
Losing a well executed selection by making a mistake is very frustrating. To undo a selection change open the project journal via main menu **Project->Journal->Show**. Scroll to the end of the journal by using the mouse or press **Ctrl+End**.
All recent changes are logged.
The old overwritten selection can be found and can be **copied** as text and **pasted** in the custom edit mode of the selection.

***NOTE***
Logging depends on the Logging Filter settings Edit->Options->Journal.

## 4.3 Conventions / Terminology

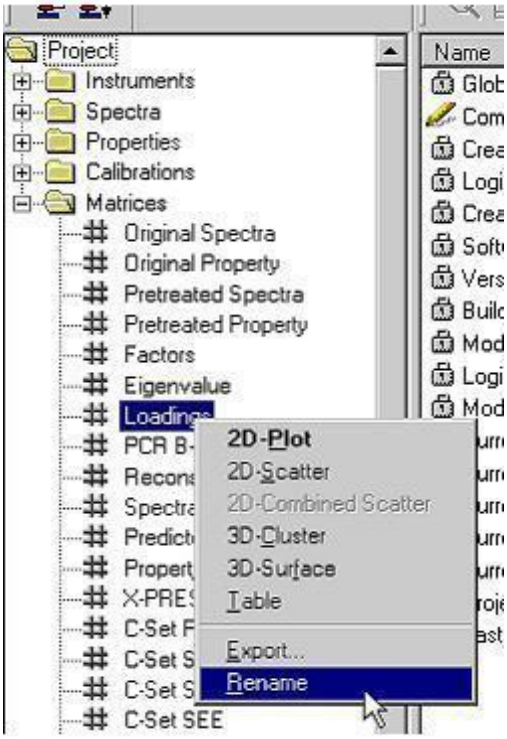
NIRCal uses the name conventions for chemometrics terms according to Malinowski. This notation has been used since the first chemometrics package BCAP for the NIRVis spectrometers. In order to be consistent and to guarantee a smooth transfer from BCAP to NIRCal this notation has been kept for NIRCal as well (see NIRCal 4.21 user manual, Appendix 2).

| Malinowski Notation | Martens / Naes Notation |
|---|---|
| Factors | Loadings / PC's (Principal Components) |
| Loadings | Scores |

| | **Matrices** | | | | **Matrix Dimensions** | | | |
|---|---|---|---|---|---|---|---|---|
| NIRCal 4 | Spectrum | Property | Factor | Loading | Factors | Wave-lengths | Spectra | Properties |
| NIRCal 5 | Spectrum | Property | Loading | Score | PC's | Wave-lengths | Spectra | Properties |
| Unscram-bler | x | y | Loading | Score | PC's | x-variables | Samples | y-variables |

### 4.3.1 Temporary Settings

Open the folder Matrices in the NIR-Explorer. Right clicking the mouse on the appropriate matrix name opens a pop-up menu. Clicking on Rename enables the user to change the name of the matrix.




### 4.3.2 Permanent Settings

In order to change the names permanently it is necessary to perform the changes in the registry settings.
HKEY_CURRENT_USER\Software\Buchi\NIRCAL\Terminology
When performing the changes (including All Factors, etc.) the name convention will be changed permanently and will be used for the NIR-Explorer, tables and graphics.

## 4.4 Characteristic absorbance table

This absorbance table is published for food. It can be used also for chemicals. Because the matrix is different the absorptions can be slightly different.

| Wavelength [nm] | Wavenumber [$cm^{-1}$] | Chemical bonds | Product Example |
|---|---|---|---|
| 1000 | 10000 | O-H str.second overtone | ArOH |
| 1015 | 9852 | 2x C-H str.+3x C-H def. | CH3 |
| 1020 | 9804 | 2x N-H str.+2x amide I | protein |
| | | N-H str.second overtone | ArNH2 |
| 1030 | 9709 | | RNH2 |
| 1037 | 9643 | 2x C-H str.+2x C-H def.+(CH2)n | oil |
| 1053 | 9497 | | CH2 |
| 1060 | 9434 | N-H str.second overtone | RNH2 |
| 1080 | 9259 | 2x C-H str.+2x C-C str. | benzene |
| 1097 | 9116 | | cyclopropane |
| 1143 | 8749 | C-H str.second overtone | aromatic |
| 1152 | 8681 | | CH3 |
| 1170 | 8547 | | HC=CH |
| 1195 | 8368 | | CH3 |
| 1215 | 8230 | | CH2 |

| Wavelength [nm] | Wavenumber [$cm^{-1}$] | Chemical bonds | Product Example |
|---|---|---|---|
| 1360 | 7353 | 2x C-H str.+ C-H def. | CH3 |
| 1395 | 7168 | | CH2 |
| 1410 | 7092 | O-H str.first overtone | ROH |
| 1415 | 7067 | 2x C-H str.+ C-H def. | CH2 |
| 1417 | 7057 | | aromatic |
| 1420 | 7042 | O-H str.first overtone | ArOH |
| 1430 | 6993 | | sucrose,starch |
| 1440 | 6944 | 2x C-H str.+ C-H def. | CH |
| 1446 | 6916 | | aromatic |
| 1450 | 6897 | O-H str.first overtone | starch,H2O |
| 1460 | 6849 | N-H str.first overtone | CONH2 |
| 1471 | 6798 | | CONHR |
| 1480 | 6757 | O-H str.first overtone (intramol.H-bond) | glucose |
| 1483 | 6743 | N-H str.first overtone | CONH2 |
| 1490 | 6711 | | CONHR |
| | | N-H str.first overtone (intramol.H-bond) | CONH2 |
| | | O-H str.first overtone (intramol.H-bond) | cellulose |
| 1492 | 6702 | N-H str.first overtone | ArNH2 |
| 1500 | 6667 | | NH |
| 1510 | 6623 | | protein |
| 1520 | 6579 | O-H str.first overtone | CONH2 |
| | | N-H str.first overtone (intramol.H-bond) | ROH |
| 1528 | 6545 | O-H str.first overtone (intramol.H-bond) | starch |
| 1530 | 6536 | N-H str.first overtone | RNH2 |
| 1533 | 6523 | C-H str.first overtone | CH |
| 1540 | 6494 | O-H str.first overtone (intramol.H-bond) | starch |
| 1570 | 6369 | N-H str.first overtone | CONH- |
| 1580 | 6329 | O-H str.first overtone intermol.H-bond) | starch,glucose |
| 1620 | 6173 | C-H str.first overtone | CH2 |
| 1645 | 6079 | | R-CH-CH |
| 1660 | 6024 | | cis-RCH=CHR |
| 1685 | 5935 | | aromatic |
| 1695 | 5900 | | CH3 |
| 1705 | 5865 | | CH3 |
| 1725 | 5797 | | CH2 |
| 1740 | 5747 | | SH |
| 1765 | 5666 | | CH2 |
| 1780 | 5618 | | cellulose |
| 1820 | 5495 | O-H srt.+2x C-O str. | cellulose |
| 1900 | 5263 | O-H str.+2x C-O str. | starch |
| | | C=O str.second overtone | -CO2H |
| 1908 | 5241 | O-H str.first overtone | POH |
| 1920 | 5208 | C=O str.second overtone | CONH |
| 1940 | 5155 | O-H str.+ O-H def. | H2O |
| 1950 | 5128 | C=O str.second overtone | -CO2R |
| 1960 | 5102 | N-H asym.str.+ amide II | CONH2 |

| Wavelength [nm] | Wavenumber [$cm^{-1}$] | Chemical bonds | Product Example |
|---|---|---|---|
| 1980 | 5051 | | protein |
| 2000 | 5000 | 2x O-H def.+ C-O def. | starch |
| | | N-H sym.str.+amide II | CONH2,CONHR |
| 2030 | 4926 | C=O str.second overtone | CONH2 |
| 2050 | 4878 | N-H sym.str.+ amide II N-H | protein |
| | | asym.str.+ amide II O-H | CONH2 |
| 2080 | 4808 | str.+ O-H def. | ROH,sucrose,starch |
| 2100 | 4762 | 2x O-H def.+ 2x C-O str | starch |
| 2110 | 4739 | N-H sym.str.+ amide III | CONH2,CONHR |
| 2132 | 4690 | N-H str.+ C=O str. | amino acid |
| 2140 | 4673 | =C-H str.+ C=C str. | HC=CH |
| 2150 | 4651 | 2x amide I+ amide III | CONH2 |
| 2160 | 4630 | | CONHR |
| 2180 | 4587 | | protein |
| 2190 | 4566 | CH2 asym.str.+ C= str. | HC=CH |
| 2200 | 4545 | C-H str.+ C=O str. | -CHO |
| 2242 | 4460 | N-H str.+ NH3+ def. | amino acid |
| 2252 | 4440 | O-H str.+ O-H def. | starch |
| 2276 | 4394 | O-H str.+ C-C str. | starch |
| 2280 | 4383 | C-H str.+ C-H def. | CH3 |
| 2294 | 4359 | N-H str.+ C=O str. | amino acid |
| 2310 | 4329 | N-H str.+ C-H def. | CH2 |
| 2323 | 4305 | C-H str.+ C-H def. | CH2 |
| 2336 | 4281 | | cellulose |
| 2347 | 4261 | CH2 sym.str.+ =CH2 def. | HC=CHCH2 |
| 2352 | 4252 | C-H def.second overtone | cellulose |
| 2380 | 4202 | O-H def.second overtone | ROH |
| 2461 | 4063 | C-H str.+ C-C str. | starch |

# 5 Software installation

## 5.1 Software installation

### 5.1.1 Important Notes

The latest NIRWare and NIRCal cannot be installed together with previous versions. Should there be an existing installation of NIRWare and NIRCal it will be removed during installation. An existing NIRWare Database from a prior installation will be upgraded.
-> **Backup before installation!**

A user with administrator rights on the operating system is required to install the software and then as well to change following settings within the NIRWare Suite after installation of the SW:

- changing the IP-address of NIRFlex N-500
- change of the SQL Server path
- backup and restore the database to the default folder in ...\\program files\...
- import of the licenses

### 5.1.2 System requirements

**Hardware:**

| | |
|---|---|
| Processor: | Intel Core i3 or higher and 1.4 GHz or faster |
| RAM: | 4GB or higher |
| HDD: | 20GB free hard disk space |
| DRIVE: | DVD-ROM drive |
| Network Adapter: | 1 x 100 Mbit/s LAN (2x 100 Mbit/s LAN recommend) |
| Display Resolution: | 1280x1024 |

**Operating System:**

- Windows 7 Professional / Enterprise / Ultimate (64-bit) or
- Windows 10 Pro / Enterprise 2016 LTSB (64-bit)

- Firewalls need to be stopped during installation (Note: We recommend not to connect to the internet without protection from a fire wall.)

### 5.1.3 Other requirements

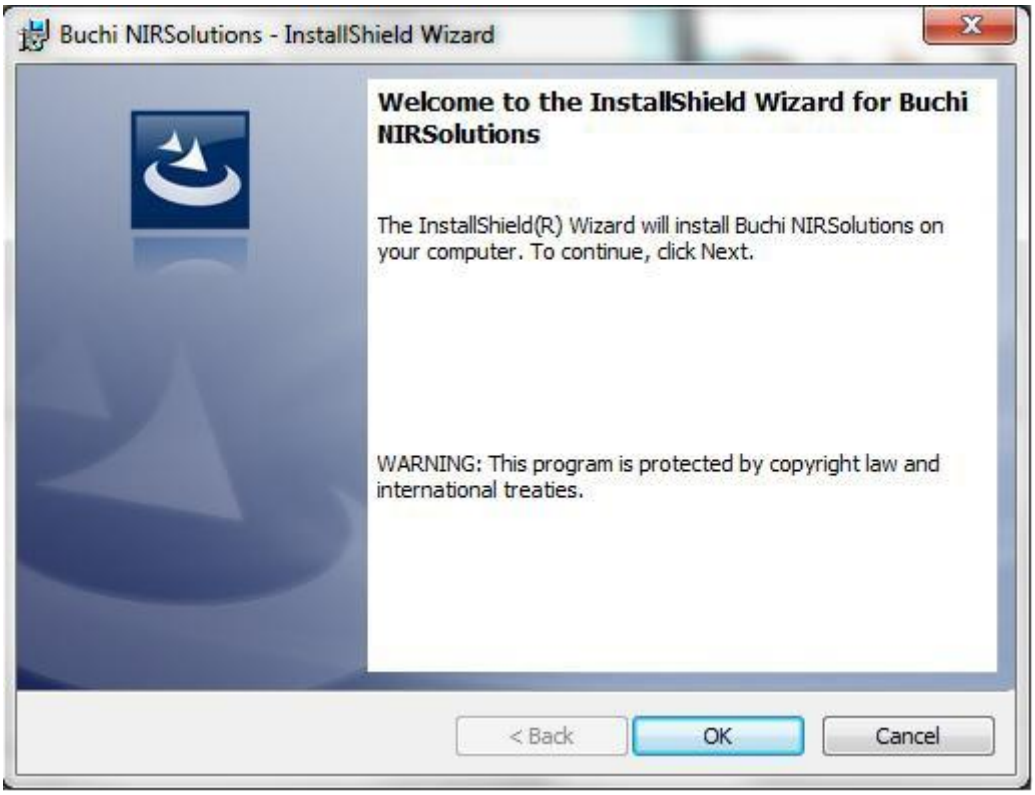
- Administrator rights (on the OS) required for installation

### 5.1.4 Installation procedure

- Insert the Installation CD. In the opening window, click on "Install NIRWare..."


Installation
NIR Software Suite
BUCHI
Operation Manual
License Agreement
Install NIR Software Suite
Product Website
Exit

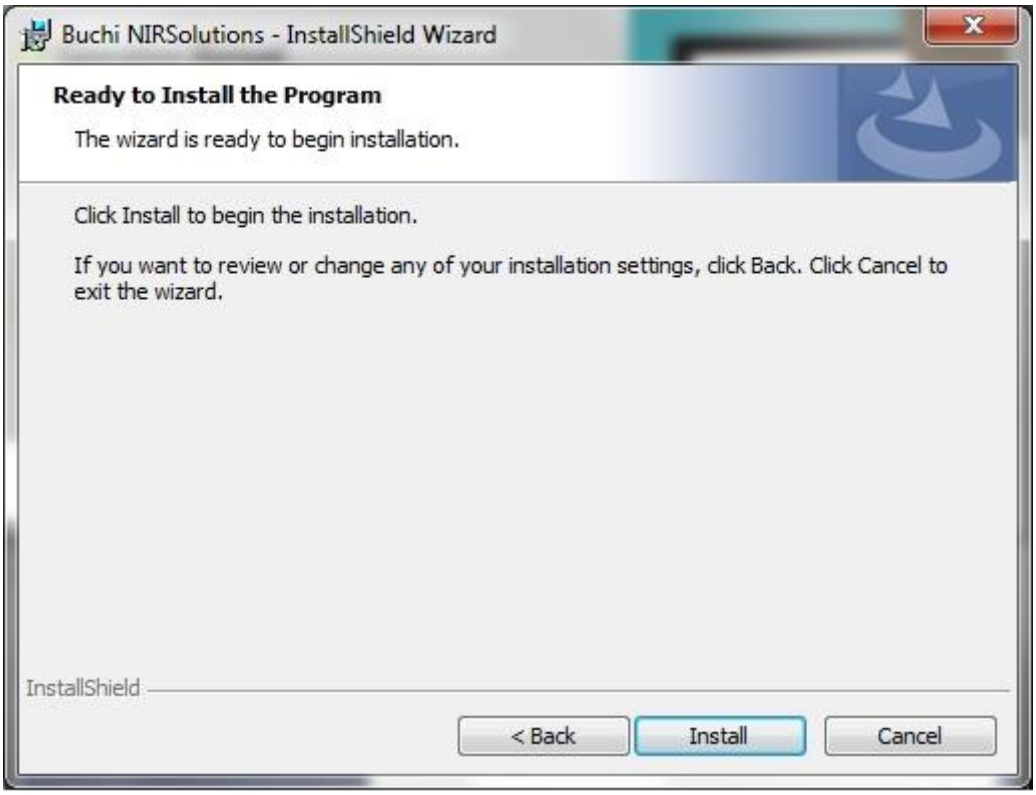
The InstallShield Wizard is opened.



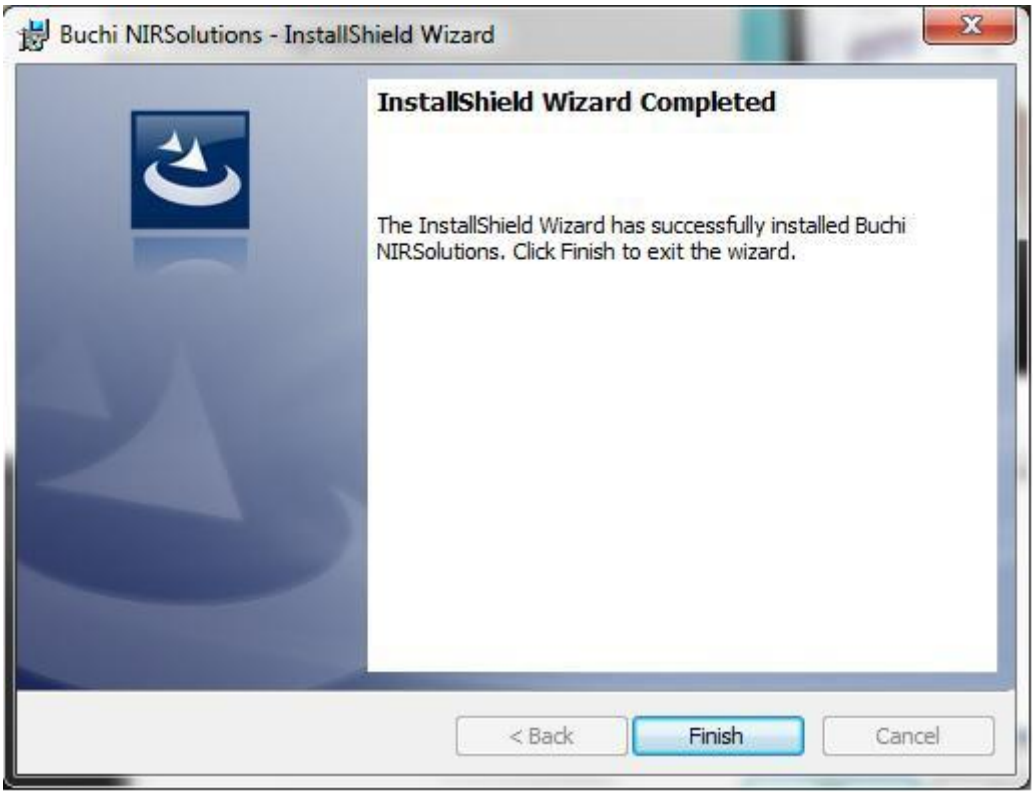

- Click OK


- Accept the license agreement to continue the installation and click Next to continue.


- We recommend to select the Complete installation. Then click Next to continue.




- Click Install to Install to start the installation process.




- Click Finish to exit the Wizard.

The applications are stored by default under C:\Program Files\Buchi\NIRSolutions\ and can be opened with Start > Programs > Buchi > NIRSolutions.

The following modules are available:

- Management Console
- Operator
- Buchi Service Manager
- NIRCal

## 5.2 Software update

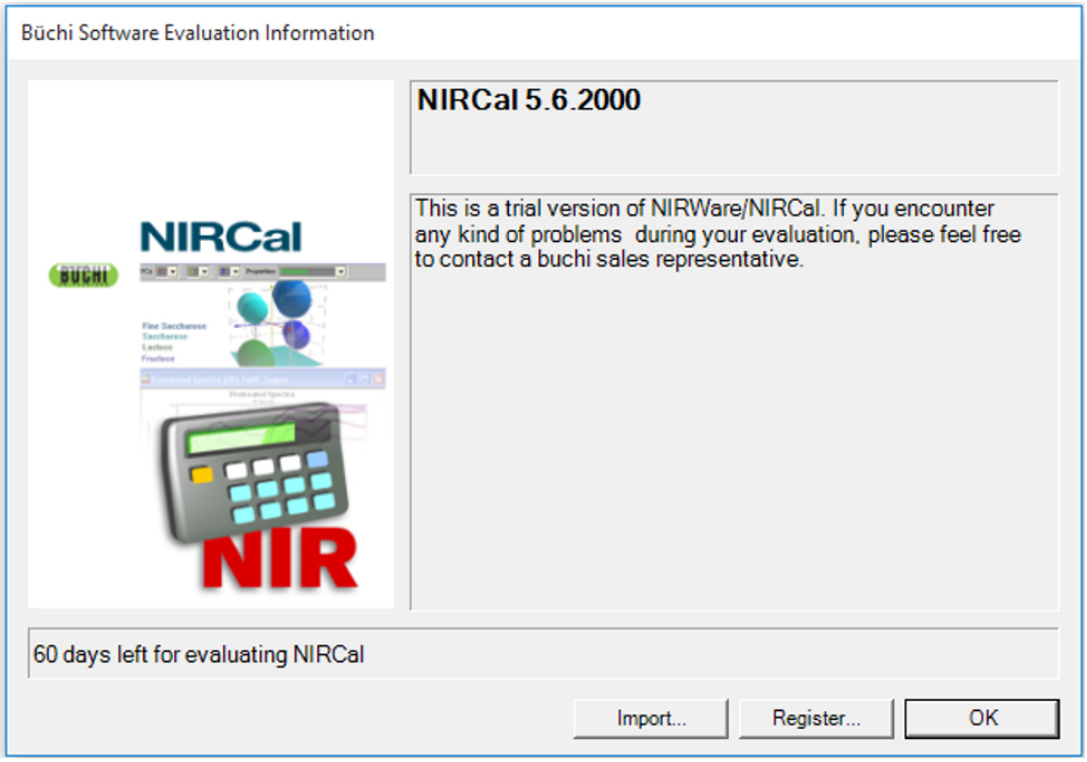
Software update over internet is discontinued since NIRWare 1.4.

## 5.3 Software licenses

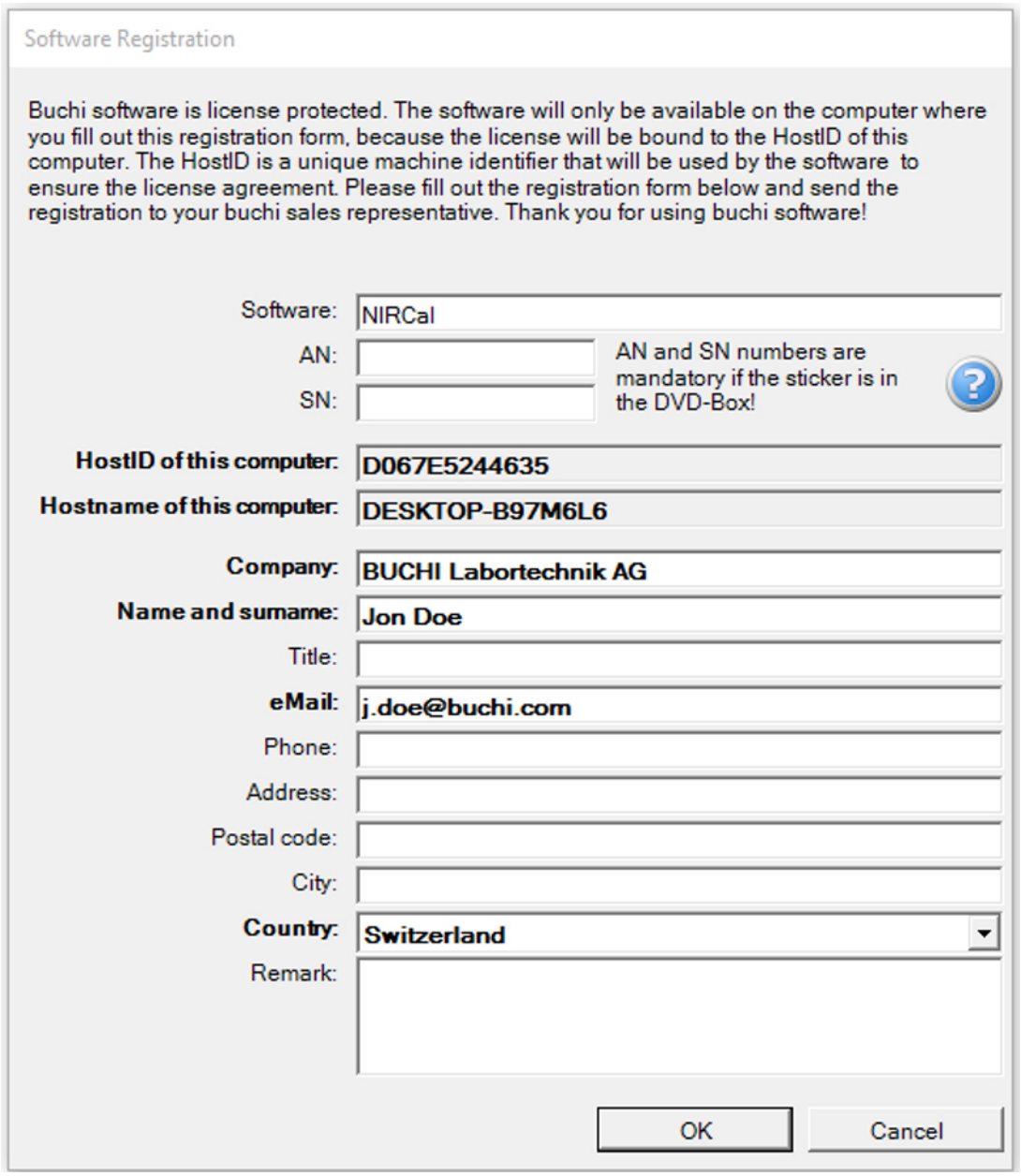
After Installation of NIRWare or NIRCal the software can be used in DEMO mode for 60 days. Within these 60 days it is recommended to register the software (NIRCal and NIRCal Toolbox) and apply for a license.

### 5.3.1 Software Registration

The registration form for NIRCal and Toolbox can be saved as an *.xml file and needs to be sent to your Buchi contact person, e.g. as an email attachment.
The form can be opened using the "NIRCal" --> Menu: Help > Software Registration...” or clicking the button "Register..." on the startup of a trial version.




The Software Registration dialog opens.




Fill it out and send it to your Buchi sales representative.

### 5.3.2 Activating the License

The license are generated based on the information provided in the registration form. To activate the license in NIRCal select Help > Import License.

***NOTE***

- *The user needs administrator rights on the OS to import the license file.*
- *NIRCal5, NIRWare Operator and also licenses for pre-calibrated applications are node locked. This means that these SW Licenses are bound to the PC (Host-ID) from where the registration file was filled out.*

# BUCHI Affiliates:

## Europe

Switzerland/Austria

BÜCHI Labortechnik AG
CH – 9230 Flawil
T +41 71 394 63 63
F +41 71 394 64 64
buchi@buchi.com
www.buchi.com

Benelux

BÜCHI Labortechnik GmbH
Branch Office Benelux
NL – 3342 GT Hendrik-Ido-Ambacht
T +31 78 684 94 29
F +31 78 684 94 30
benelux@buchi.com
www.buchi.com/bx-en

France

BUCHI Sarl
FR – 94656 Rungis Cedex
T +33 1 56 70 62 50
F +33 1 46 86 00 31
france@buchi.com
www.buchi.com/fr-fr

Germany

BÜCHI Labortechnik GmbH
DE – 45127 Essen
T +800 414 0 414 0 (Toll Free)
T +49 201 747 49 0
F +49 201 747 49 20
deutschland@buchi.com
www.buchi.com/de-de

Italy

BUCHI Italia s.r.l.
IT – 20010 Cornaredo (MI)
T +39 02 824 50 11
F +39 02 575 12 855
italia@buchi.com
www.buchi.com/it-it

Russia

BUCHI Russia/CIS
Russia 127287 Moscow
T +7 495 36 36 495
russia@buchi.com
www.buchi.com/ru-ru

United Kingdom

BUCHI UK Ltd.
GB – Oldham OL9 9QL
T +44 161 633 1000
F +44 161 633 1007
uk@buchi.com
www.buchi.com/gb-en

Germany

BÜCHI NIR-Online
DE – 69190 Walldorf
T +49 6227 73 26 60
F +49 6227 73 26 70
nir-online@buchi.com
www.nir-online.de

## America

Brazil

BUCHI Brasil Ltda.
BR – Valinhos SP 13271-200
T +55 19 3849 1201
F +55 19 3849 2907
brasil@buchi.com
www.buchi.com/br-pt

USA/Canada

BUCHI Corporation
US – New Castle, DE 19720
T +1 877 692 8244 (Toll Free)
T +1 302 652 3000
F +1 302 652 8777
us-sales@buchi.com
www.buchi.com/us-en

## Asia

China

BUCHI China
CN – 200052 Shanghai
T +86 21 6280 3366
F +86 21 5230 8821
china@buchi.com
www.buchi.com/cn-zh

India

BUCHI India Private Ltd.
IN – Mumbai 400 055
T +91 22 667 75400
F +91 22 667 18986
india@buchi.com
www.buchi.com/in-en

Indonesia

PT. BUCHI Indonesia
ID – Tangerang 15321
T +62 21 537 62 16
F +62 21 537 62 17
indonesia@buchi.com
www.buchi.com/id-in

Japan

Nihon BUCHI K.K.
JP – Tokyo 110-0008
T +81 3 3821 4777
F +81 3 3821 4555
nihon@buchi.com
www.buchi.com/jp-ja

Korea

BUCHI Korea Inc.
KR – Seoul 153-782
T +82 2 6718 7500
F +82 2 6718 7599
korea@buchi.com
www.buchi.com/kr-ko

Malaysia

BUCHI Malaysia Sdn. Bhd.
MY – 47301 Petaling Jaya,
Selangor
T +60 3 7832 0310
F +60 3 7832 0309
malaysia@buchi.com
www.buchi.com/my-en

Singapore

BUCHI Singapore Pte. Ltd.
SG – Singapore 609919
T +65 6565 1175
F +65 6566 7047
singapore@buchi.com
www.buchi.com/sg-en

Thailand

BUCHI (Thailand) Ltd.
TH – Bangkok 10600
T +66 2 862 08 51
F +66 2 862 08 54
thailand@buchi.com
www.buchi.com/th-th

# BUCHI Support Centers:

South East Asia

BUCHI (Thailand) Ltd.
TH-Bangkok 10600
T +66 2 862 08 51
F +66 2 862 08 54
bacc@buchi.com
www.buchi.com/th-th

Middle East

BÜCHI Labortechnik AG
UAE – Dubai
T +971 4 313 2860
F +971 4 313 2861
middleeast@buchi.com
www.buchi.com

Latin America

BUCHI Latinoamérica
S. de R.L. de C.V.
MX – Mexico City
T +52 55 9001 5386
latinoamerica@buchi.com
www.buchi.com/es-es

We are represented by more than 100 distribution partners worldwide.
Find your local representative at: www.buchi.com