

Multivariate Calibration Quick Guide

Table Of Contents

1.	HOW TO CREATE CALIBRATION MODELS	1
1.1.	Introduction into Multivariate Calibration Modelling	1
1.1.1.	Preparing Data	1
1.2.	Step 1: Calibration Wizard Steps Overview	3
1.3.	Step 2: Entering General Information	4
1.4.	Step 3: Calibration Model and appropriate Parameters	5
1.5.	Step 4: Spectra Selection.....	6
1.5.1.	Add / Remove spectra from Calibration list	7
1.5.2.	Statistical Spectra	8
1.5.3.	Auto Validation	9
1.5.4.	Cross Validation	9
1.5.5.	Editing Labels.....	10
1.6.	Step 5: Preprocessing - Applying Mathematical Operations	13
1.7.	Step 6: Definition of relevant Spectral Ranges (Variable Selection).....	17
1.8.	Step 7: Multivariate Factor Analysis.....	19
1.9.	Final Step: Calibration Model Results– Ready for Review and interactive Optimization	20
1.9.1.	Prediction Plot	21
1.9.2.	Residuals Plot	22
1.9.3.	Loadings Plot.....	22
1.9.4.	Scores Plot.....	23
1.9.5.	2D and 3D Factor plots.....	23
1.9.6.	Spectra Selection	24
1.10.	Saving the Calibration Model	25
1.11.	Expected Results.....	25
2.	PREDICTION OF UNKNOWN SAMPLES	27

1. How to create Calibration Models

This Quantify Wizard quick guide was prepared to help users to get started in using the quantification module of the software and its utilities. This guide contains very basic information only, and by no means intends to replace the user's manuals!

In this quick guide, we will develop a calibration for iodine value in epoxidized soybean oil. Iodine value is the measurement of the number of double bonds in an organic compound.

1.1. Introduction into Multivariate Calibration Modelling

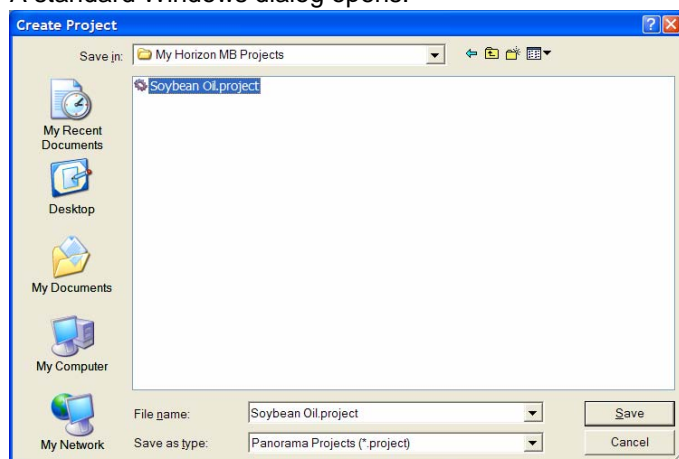
This guide is designed to give a short overview of setting up a new calibration model. It is not an introduction into Chemometrics nor does it discuss in detail the effect of special parameters, influencing the overall calibration module.

1.1.1. Preparing Data

To start a new calibration you need a project containing your own sample files.

Creating a project is simple, click **Project->New**.

A standard Windows dialog opens:



Enter "Soybean Oil" as **File name** and press the **Save** button.

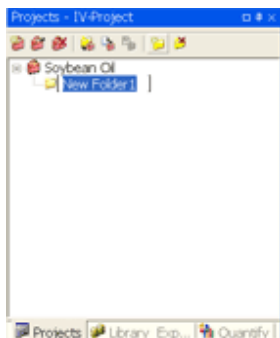
The new project appears in the **Projects** pane of the application (top right corner).



Multivariate Calibration Quick Guide

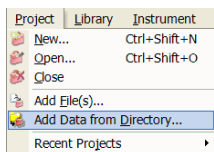
Optionally, you may create new folders in your project to organize data.


Click the **New Folder** icon in the project explorer toolbar. A new folder is inserted automatically.



Rename the folder node into **"Spectra"**

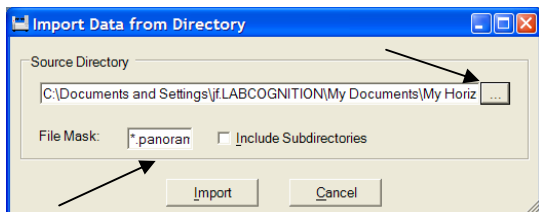
Up to now the project contains no data. Select the function **Add Data from Directory** from the Projects menu



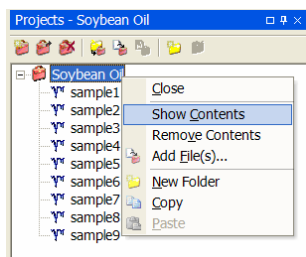
and navigate to the data source directory using the  button.

Make sure the File mask contains your favorite data format (e.g. *.spc).

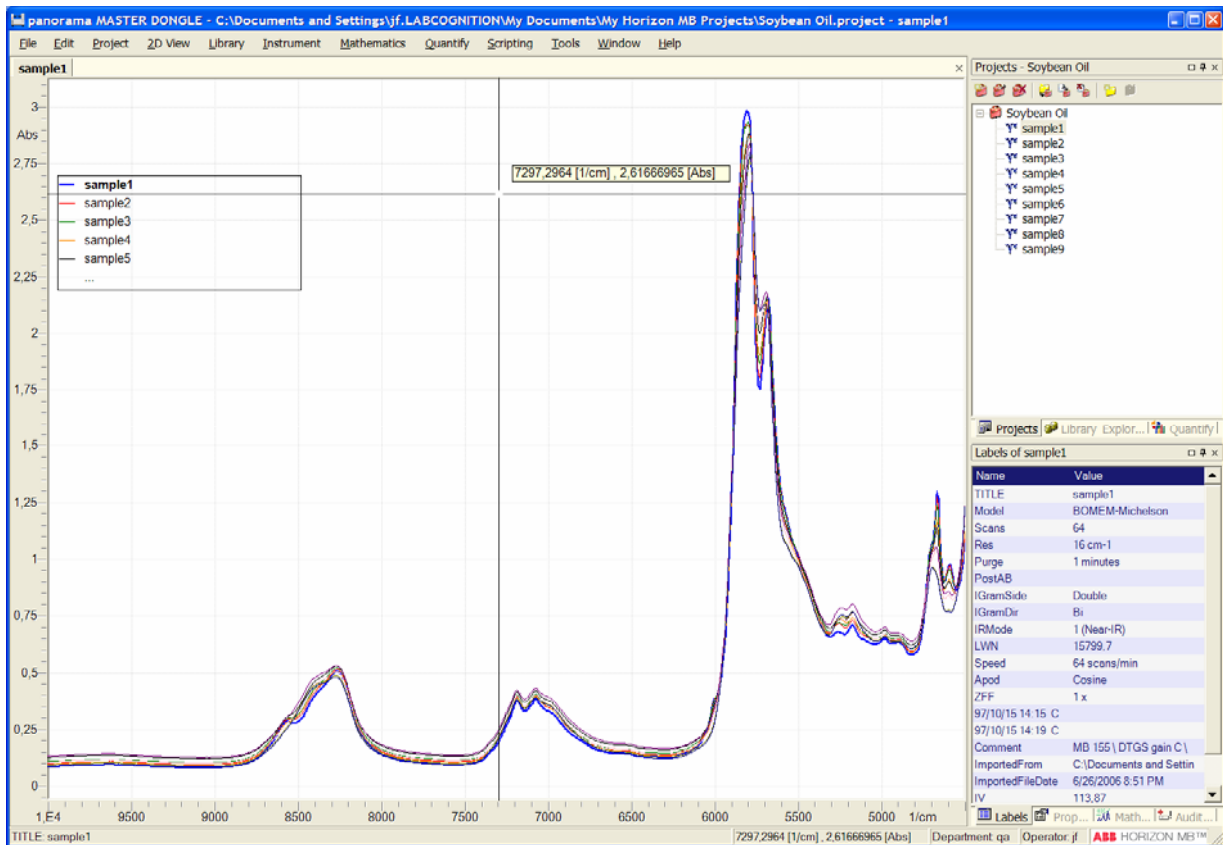
Click **Import** to initiate data import with the defined extension into the project.



All loaded spectra can now easily be displayed with the command **Show Contents**, available from the projects right mouse context menu.



As you can see the spectra are now displayed merged in a single data view window within the main workspace. A legend is automatically shown on the left side with the actual selected spectrum marked bold face.



You are now ready to setup the calibration model.

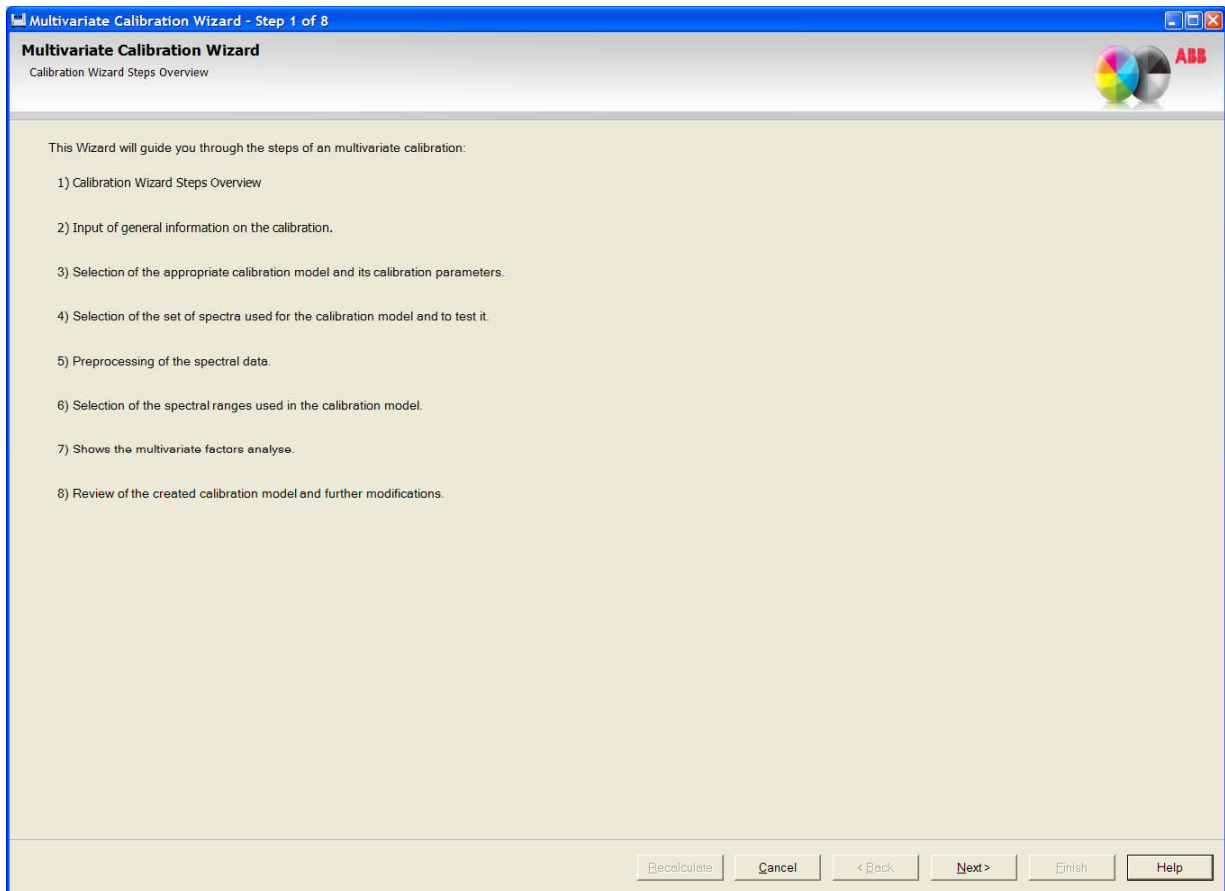
Select the **Soybean Oil project** node in the Project explorer.

Choose **New Multivariate Calibration** from the Quantify menu. The calibration wizard opens and guides you through the steps of setting up a calibration model.

1.2. Step 1: Calibration Wizard Steps Overview

The calibration steps overview page informs you on all relevant steps for multivariate calibration modelling. No user interaction is required during this step.

Multivariate Calibration Quick Guide



NOTE: This page is only available when you set up a new model! It is not shown when editing calibrations.

Press **Next** to proceed.

1.3. Step 2: Entering General Information

Step 2 is meant to enter a descriptive name for the calibration model. This name is used later on to identify the calibration model within the project and on reports.

NOTE: within a project the name of the calibration model must be unique!

Optionally, you might give some more details to the purpose of your calibration in the Description text box.

Multivariate Calibration Wizard - Step 2 of 8

Multivariate Calibration Wizard
Input of general information on the calibration.

Name of the Calibration:
Multivariate

Description:

Additional information on the multivariate calibration.

Recalculate Cancel < Back Next > Finish Help

Press **Next** to proceed.

1.4. Step 3: Calibration Model and appropriate Parameters

Step 3 plays a key role. At this point you need to define on which kind of model the calibration will be calculated. Click the **Model** list box to select a model e.g.:

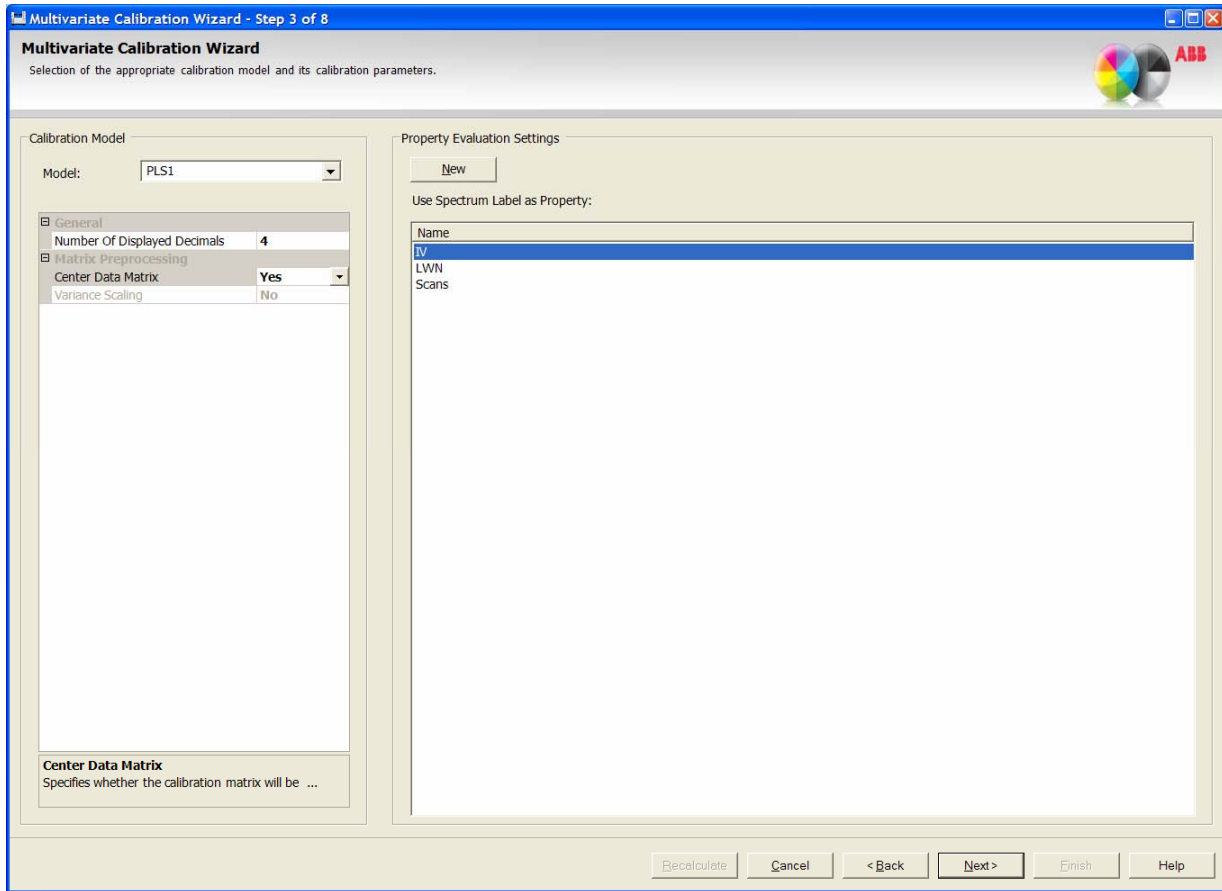
- MLR
- PLS1
- PLS2
- SIMPLS

For quantification, we choose in the **PLS-1** model.

As additional parameters use

Number of Decimals: 4

Center Data Matrix: Yes.



Quantization using statistical evaluation does not work without numeric sample property values, e.g. concentrations. Such values will be stored in so called **Labels** attached to spectral data. A list of all applicable labels is shown in the **Property Evaluation Settings** area.

Select the appropriate sample property to be calibrated in the list. In this example we choose the **IV** label.

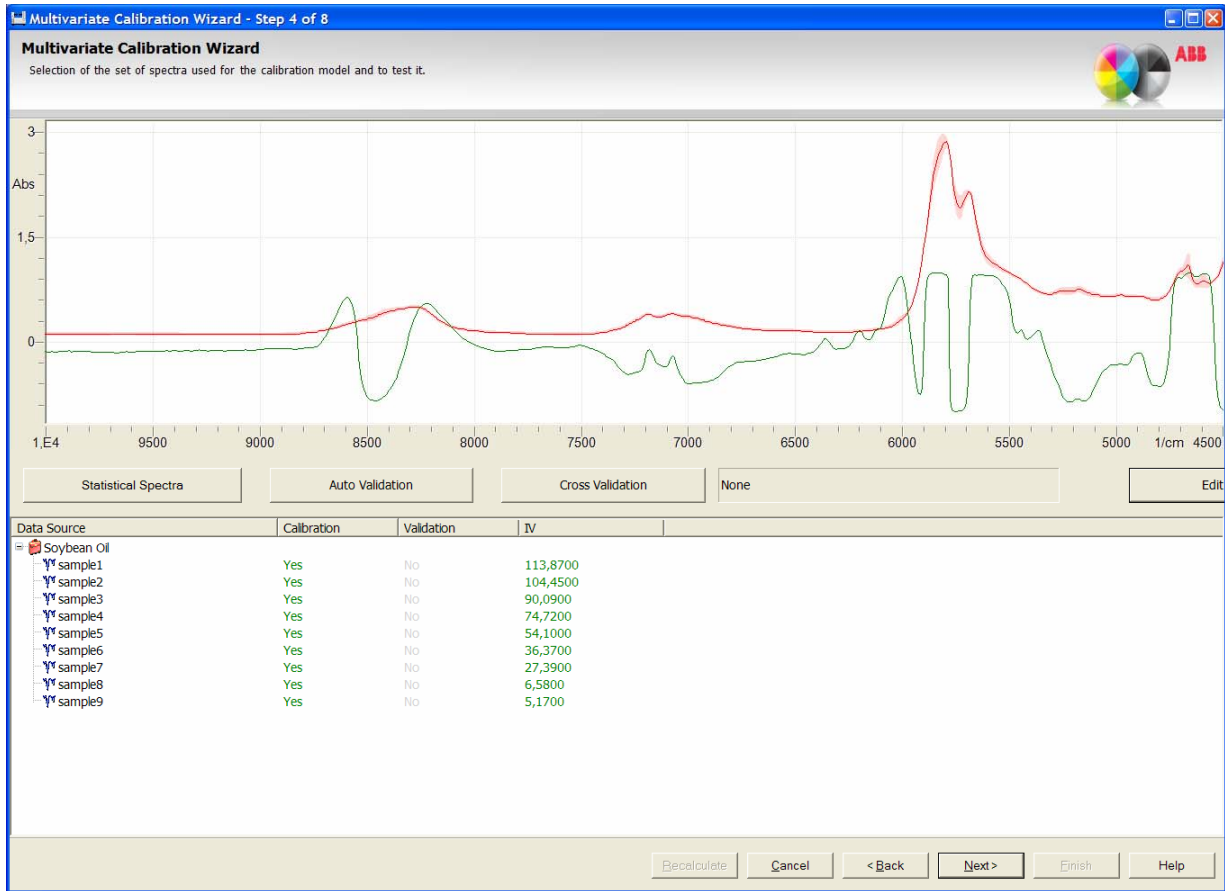
NOTE: New labels can be created easily. Just click the **New** button to prepare a new label. Filling labels with concentration values is described in the next chapter.

Press **Next** to proceed.

1.5. Step 4: Spectra Selection

After definition of the calibration model it is important to select an appropriate set of calibration spectra. In this step a set of independent validation spectra might be assigned too.

NOTE: All spectra joined in the project are available here.



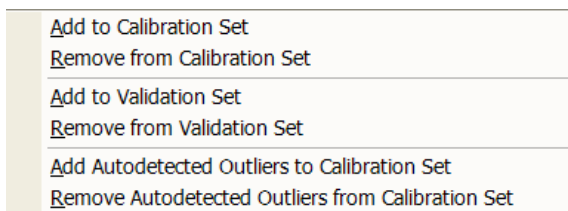
For convenience the actual calibration spectra selection is shown in the upper spectral area together with some statistical information like correlation (green line) and average (red line). A sample, marked in the Data Source grid below the spectra, is displayed as a blue line inside the average band.

1.5.1. Add / Remove spectra from Calibration list

When setting up a new experiment all spectra are included for calibration automatically. They are set **Calibration = Yes**. But it's easy to change selection. There are two ways to do it:

Double click on the **Yes** value to toggle to **No**.

Alternatively, mark the spectrum and press the **right mouse** button. A context menu opens.



Select **Add .to Calibration Set** or **Remove from Calibration Set** accordingly.

This only changes the status for one spectrum. If you like to change multiple spectra at once, do the following:

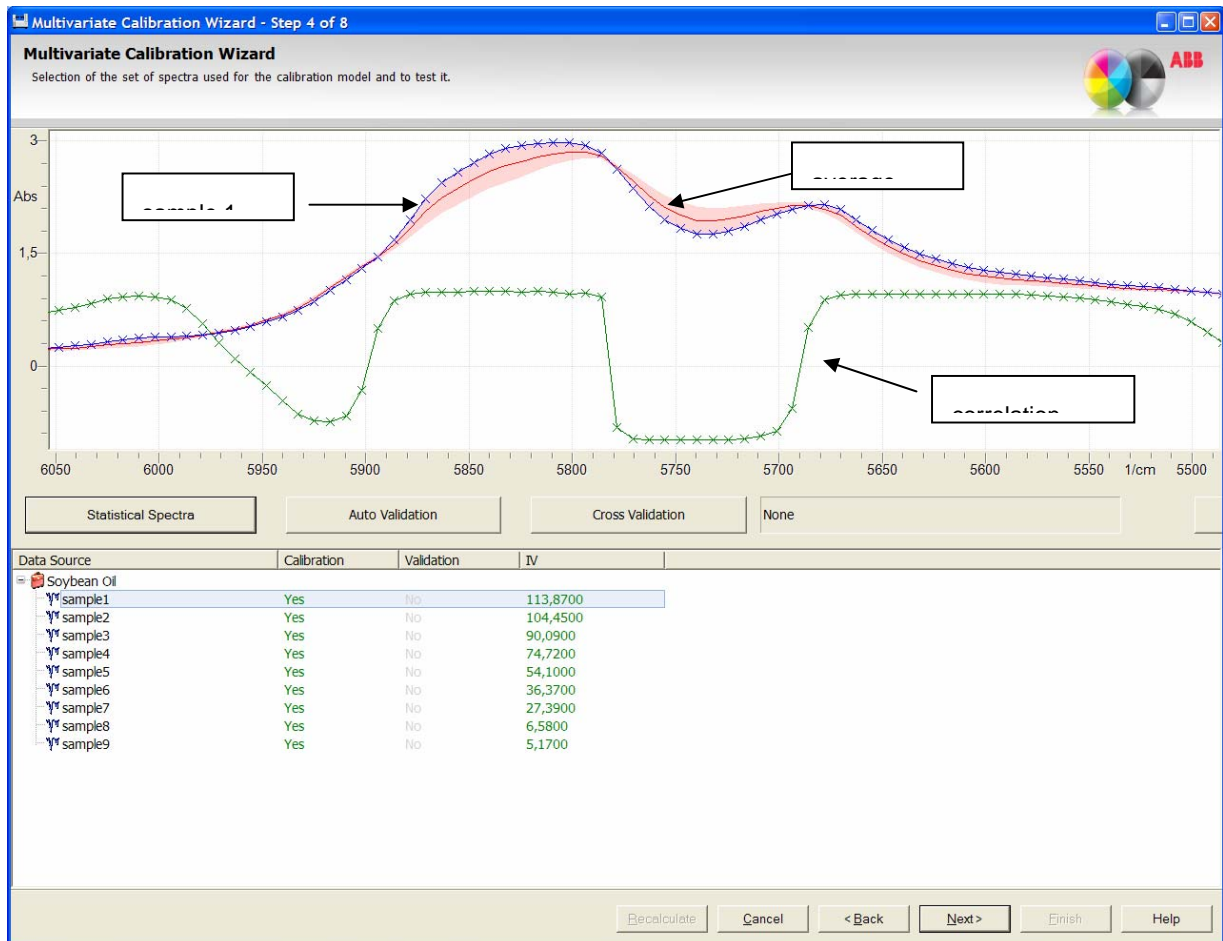
Selection of a multiple items in the list is identical with typical Windows standard file selection. Keep **CTRL** or **SHIFT** key pressed and select with the **left mouse** button.

Multivariate Calibration Quick Guide

The click the **right mouse** button to open the context menu. Choose the operation to change all selected items.

NOTE: The context menu is also available on folders. In this case all items in a folder are changed accordingly.

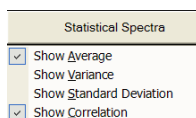
Follow the same procedure to setup spectra being used for validation. This might be an independent set of spectra or the same.



The spectra table in the lower part of the window also lists the numeric content of the calibrated label which has been selected previously. In this example the label **IV** is used and shows concentration values for each spectrum.

Between the spectral object and the data grid you will find some Function buttons

1.5.2. Statistical Spectra

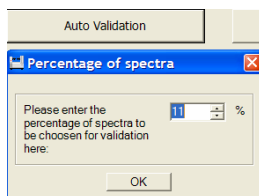


Show Average	Displays the average over all spectra selected for calibration.
Show Variance	Displays the variance over all spectra selected for calibration.
Show Standard Deviation	Displays the standard deviation over all spectra selected for calibration.
Show Correlation	Displays the correlation between spectra and property value over all spectra selected for calibration.

The **Correlation** curve is typically the most important information. It reveals highly correlated spectral regions which should be taken into account as calibration range. This helps you to identify prominent spectral regions that are highly correlated to the property under investigation. In general highly correlated regions will have a R^2 close to 1 or -1.

NOTE: Statistical information strongly depends on spectrum selection. Statistical data is updated automatically, when changing the list of spectra used in the calibration set.

1.5.3. Auto Validation

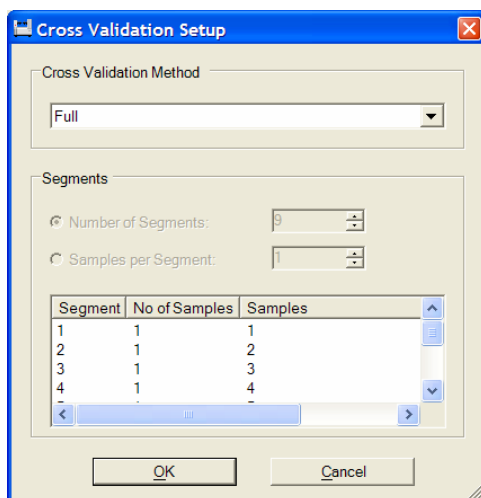


This function defines a percentage of spectra, randomly chosen for automatic validation.

NOTE: The Auto Validation is not used in this example.

1.5.4. Cross Validation

Cross validation is used to test the robustness of a calibration model. All spectra included in the calibration set are assigned a cross validation segment. By leaving out particular segments in model calculation the influence of the group of spectra on the whole model can be estimated.



Several cross validation methods are available from the drop down list:

- None

Multivariate Calibration Quick Guide

- Full (Leave one out)
- Random
- Systematic

If a **Cross Validation Method** is chosen a respective column is inserted into the Data Source table in the lower part of the wizard window. It shows the cross validation segment assignment. In our example we chose the method Full, known as the method Leave-one-out too.

Data Source	Calibration	Validation	Cross Validation Segment	IV
[-] Soybean Oil				
[-] sample1	Yes	No	1	113,8700
[-] sample2	Yes	No	2	104,4500
[-] sample3	Yes	No	3	90,0900
[-] sample4	Yes	No	4	74,7200
[-] sample5	Yes	No	5	54,1000
[-] sample6	Yes	No	6	36,3700
[-] sample7	Yes	No	7	27,3900
[-] sample8	Yes	No	8	6,5800
[-] sample9	Yes	No	9	5,1700

For convenience the description of the method is also displayed in the read-only field next to the button.

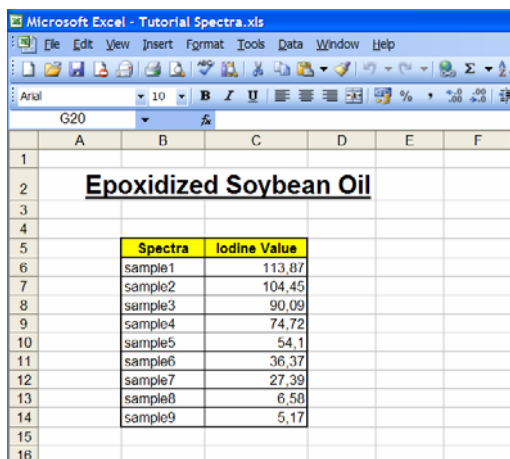
1.5.5. Editing Labels

Editing the property value of a single sample is easily done within the data source table.

Simply **click the value** with the **left mouse** button to enter the field.

On a batch level with several samples there is more convenient way using the label editor of the software. The procedure is described in the following.

At first open or create a list of concentrations e.g. in MS-Excel.



The screenshot shows a Microsoft Excel window titled 'Tutorial Spectra.xls'. The spreadsheet has columns A through F and rows 1 through 16. The title 'Epoxydized Soybean Oil' is centered in row 2. A table is located in the range C6:C14, with the following data:

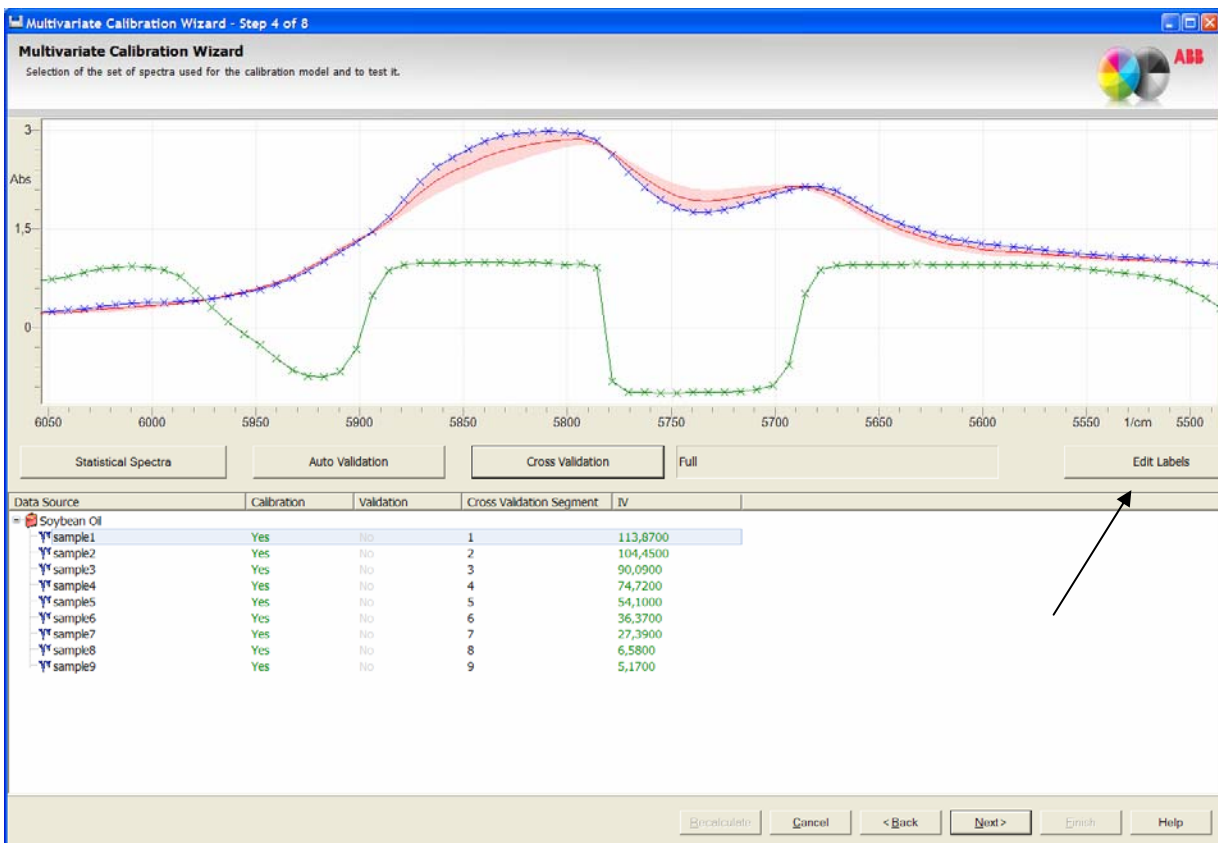
Spectra	Iodine Value
sample1	113,87
sample2	104,45
sample3	90,09
sample4	74,72
sample5	54,1
sample6	36,37
sample7	27,39
sample8	6,58
sample9	5,17

Select the numeric value range starting with the cell containing the sample1 property (here it's C6 to C14) and copy the selection into the clipboard using **Edit->Copy** or **CTRL-C** keys. Alternatively use the context menu clicking the right mouse button and choose **Copy**.

Epoxidized Soybean Oil

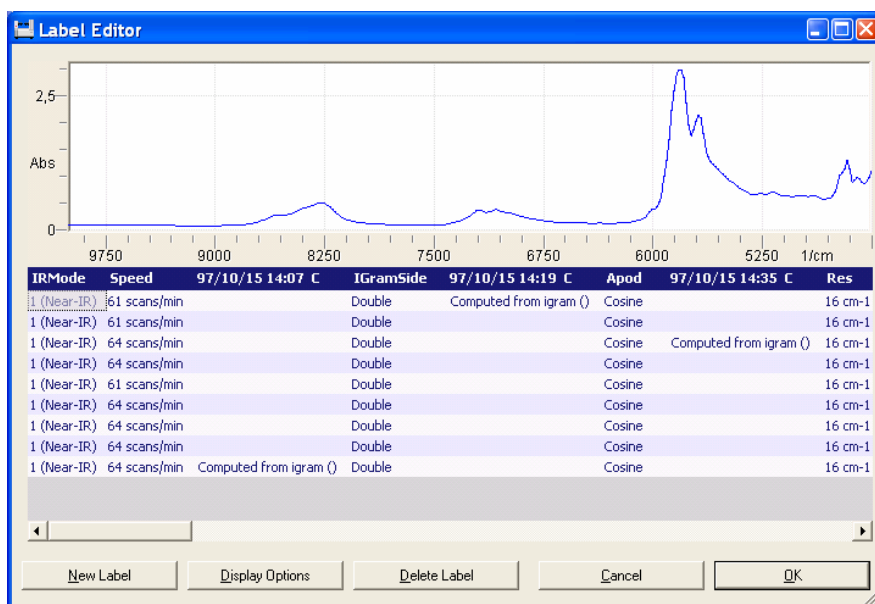
Spectra	Iodine Value
sample1	113.87
sample2	104.45
sample3	90.09
sample4	74.72
sample5	54.1
sample6	36.37
sample7	27.39
sample8	6.58
sample9	5.17

Toggle back to the calibration wizard pane and press the **Edit Labels** button to open the **Label Editor**.

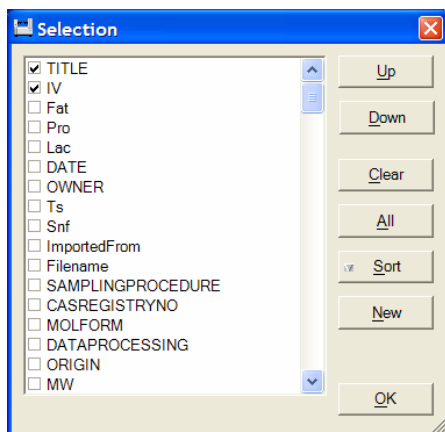


The Label Editor dialog is opened.

Multivariate Calibration Quick Guide

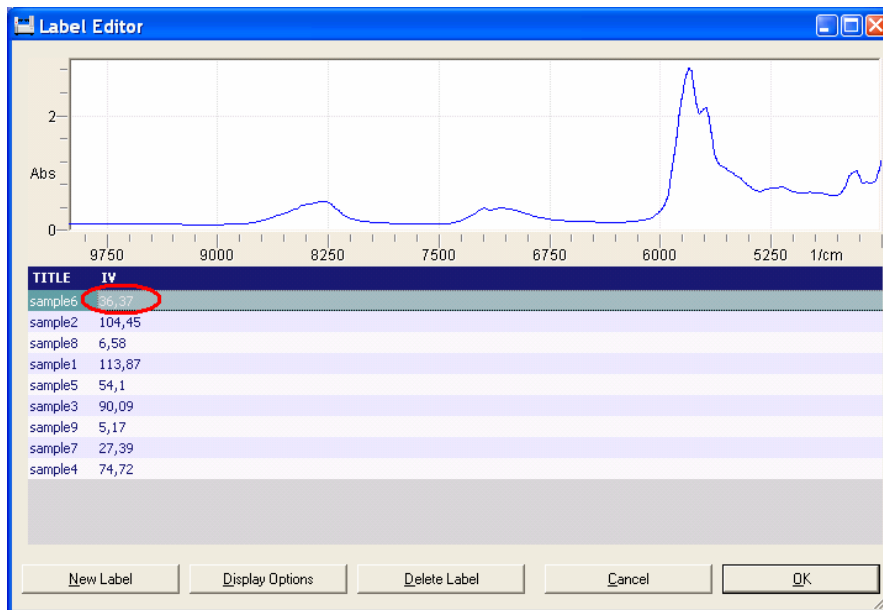


Now change the display options for displayed columns in the lower data table. Just click the **Display Options** button.



For a convenient overview you may reduce the number of displayed labels by pressing the **Clear** button and then selecting the label(s) of interest, in our example **TITLE** and **IV**. Finally click **OK** to leave the dialog.

Select the first cell of the column **IV** (indicated by red circle) and paste the selected values from the Excel sheet with **CTRL-V**.



Close the dialog with **OK**.

Values are updated automatically.

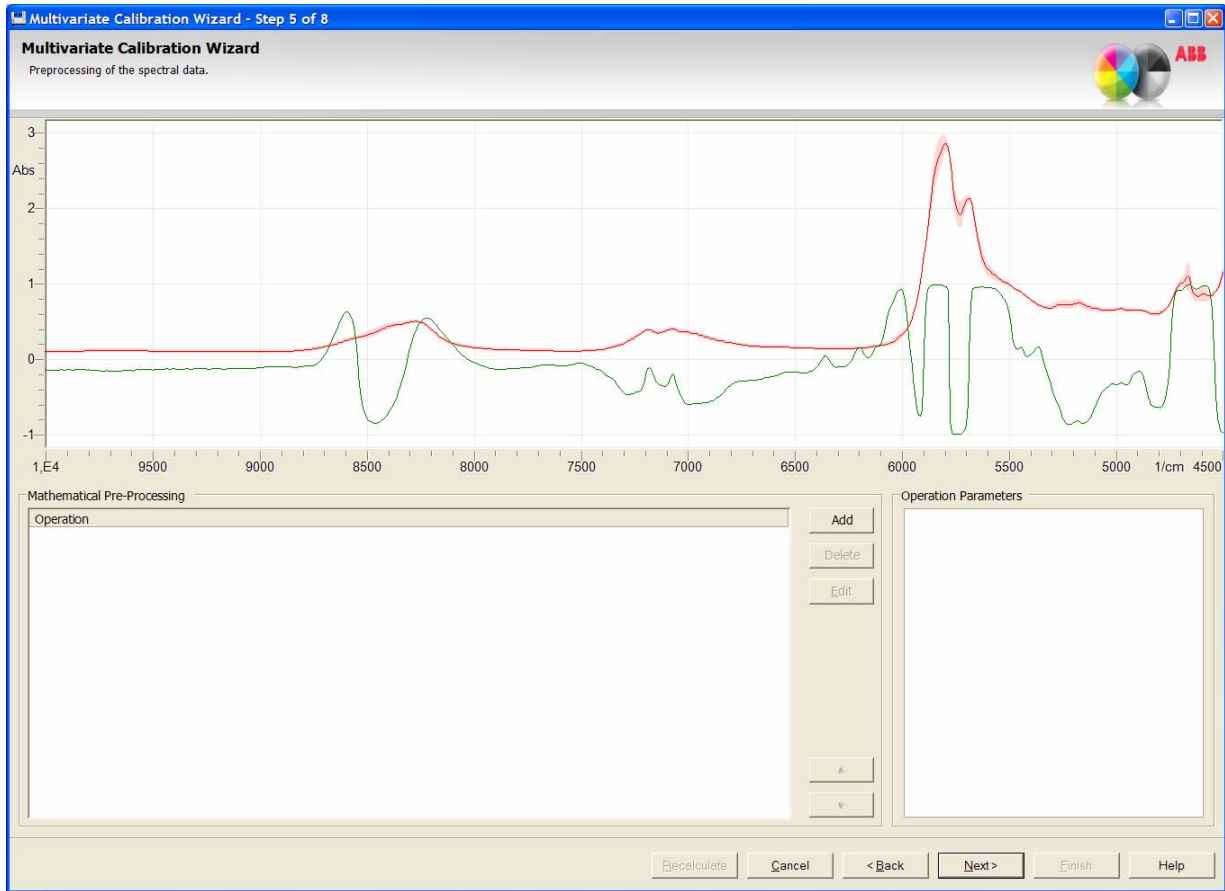
NOTE: If no concentration values are available for a spectrum it cannot be selected for calibration set.

To proceed to the next wizard step press the **Next** button.

1.6. Step 5: Preprocessing - Applying Mathematical Operations

In particular cases it might be necessary to preprocess spectral data, e.g. normalize, calculate a derivative or do a thickness correction.

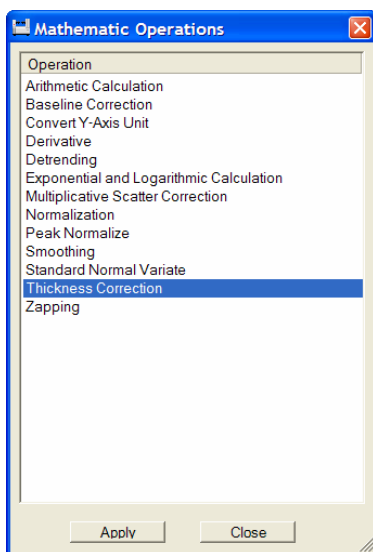
Multivariate Calibration Quick Guide



All applicable mathematical operations provided by the software are available here and can be used for pre-processing.

Press the **Add** button to open the **Mathematic Operations** dialog.

Select an operation, e.g. in our case the **Thickness Correction** and confirm with **Apply**.



Click **Close** to leave the dialog.

Applied mathematical operations are added to the list with default parameters. The number of selectable operations is not limited; therefore it is possible to combine several operations and change their order. Now parameters need to be updated to satisfy your needs.

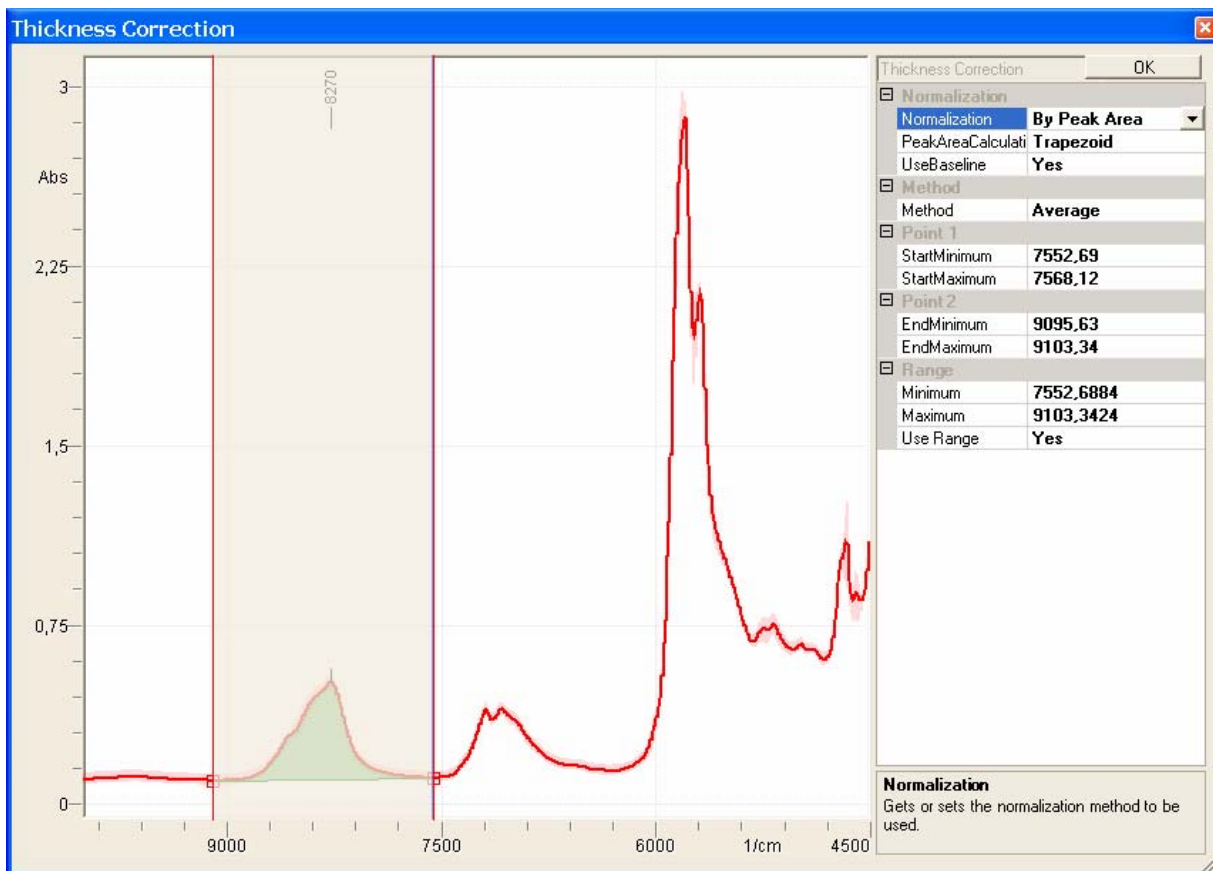
The **Thickness Correction** is used to do the path length correction, or normalization. Normalization (thickness path length correction) is done by drawing a baseline under a peak and integrating the peak over this baseline. This area then divides the entire spectrum. Often this type of path length correction is called the “internal standard” method. One requirement for this method is that there must be an isolated band in every spectrum that arises from a constituent that doesn’t vary in concentration in all samples.

Therefore, by normalizing the entire spectrum to the intensity of the band, the path length variation is effectively removed.

After selection is carried out you can optimize the parameters individually. Select the mathematical operation **Thickness Correction** and press the **Edit** button to get into the interactive mode.



A new window opens, showing calibration data and statistical spectra. Several manipulation parameters are available on the right hand similar to the list of parameters in the Mathematics tab in the main software.



First setup the **Normalization** parameter. Select the option **By Peak Area** from the drop down list. Set the **Peak Area Calculation** to Trapezoid and **Use Baseline** to **Yes**.

Next set the **Baseline Correction** method to **Average**.

The baseline of the peak considered in thickness correction is defined by two baseline points. Each point is allowed to be located in a user defined range. Ranges need to be set up in the **Point1** and **Point2** sections accordingly. The peak in between Point1 and Point2 is used for thickness correction.

Enter the values into the parameter section on the right hand of the window. Another option is moving (and resizing) the vertical baseline border lines in the data view on the left. Move the mouse pointer over the line, press and keep pressed the **left mouse** button and **move the line** to the desired position. The process of fine tuning the intervals should be done manually in the editable parameters section

In this example, the left baseline point will be the average of all points in the frequency interval **9105-9095 cm^{-1}** , and the right baseline point will be the average of all points in the range from **7565-7555 cm^{-1}** . This means, the area of the peak between 9100-7560 cm^{-1} will be used for thickness correction. These limits are typical for applications involving surfactants, polyols, oils, and fats.

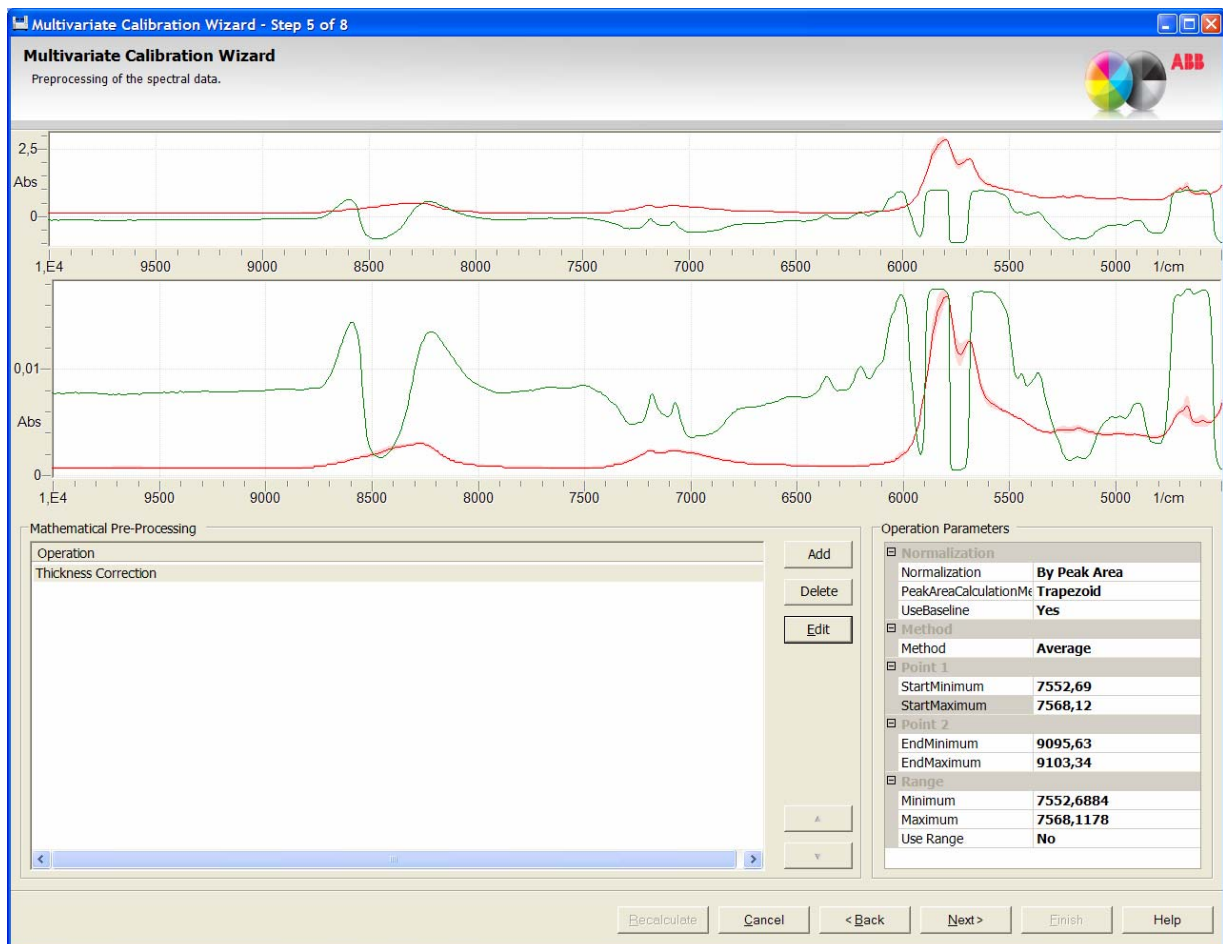
In addition to the normalization described above, a **baseline correction** can be performed in advance. This is an optional operation. To setup the baseline range, do the following:

Set the **Use Range** value to **Yes**.

Enter the range 9100-7560 cm^{-1} into the **Maximum** and **Minimum** fields accordingly.

NOTE: When entering values into parameter fields, they will be updated automatically to match the closest data point position.

Click **OK** to transfer current parameter settings to the main **Mathematics Operation Parameters** box.



As you can see the wizard pane now shows two different spectral data views. On top, the original data is shown and below a preview of converted data is visible.

Click **Next** to proceed.

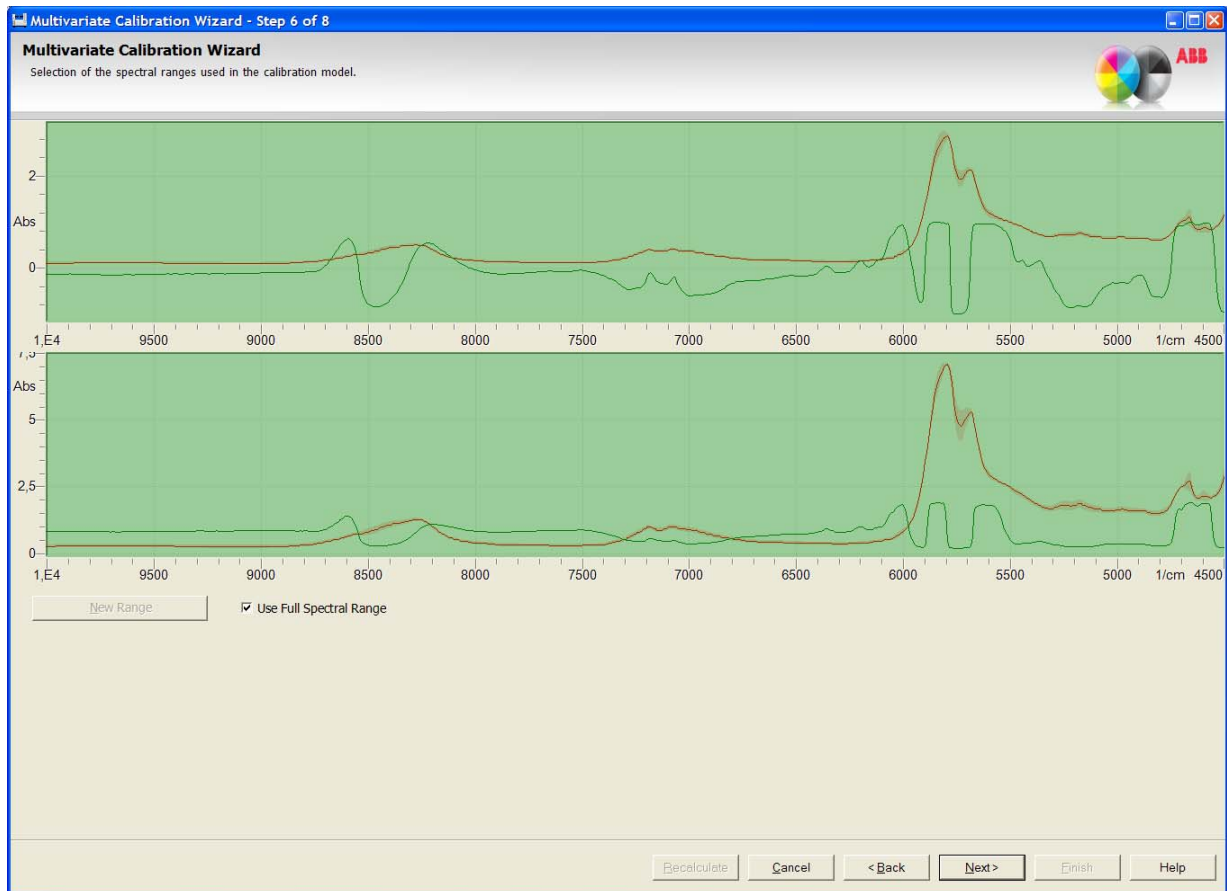
1.7. Step 6: Definition of relevant Spectral Ranges (Variable Selection)

Selection of significant variables for calibration is another important step. In this wizard window you see the original data view on top and below converted objects.

Statistical information like correlation (green line) and variance (light red shape) and some advice from the software will help you to make your decision.

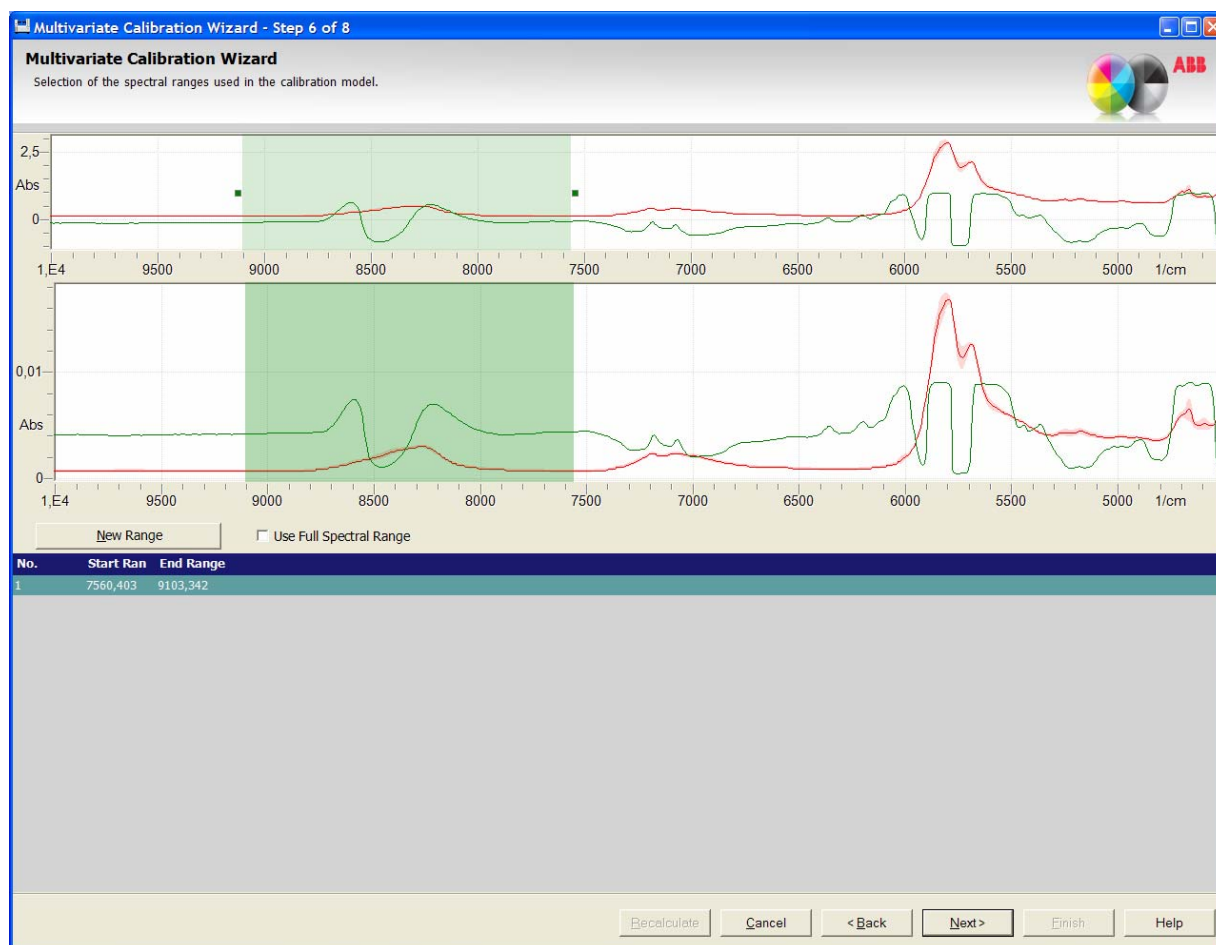
You can choose either the full spectral range or select user defined ranges conveniently. Default settings are always Use Full Spectral Range.

Multivariate Calibration Quick Guide



User defined spectral ranges can be adjusted easily either graphically with drag and drop or numerically in the table. Uncheck the **Use Full Spectral Range** flag to enable user defined ranges.

Enter the new range information in the grid. Use as **Start = 7560 cm⁻¹** and as **End = 9100 cm⁻¹**.



Determination of iodine value is done in the CH second overtone region of the spectrum, in the region $9100\text{-}7560\text{ cm}^{-1}$. In this example, the calibration region is the same one used for normalization, but it is not always the case.

The calibration model parameters are now ready.

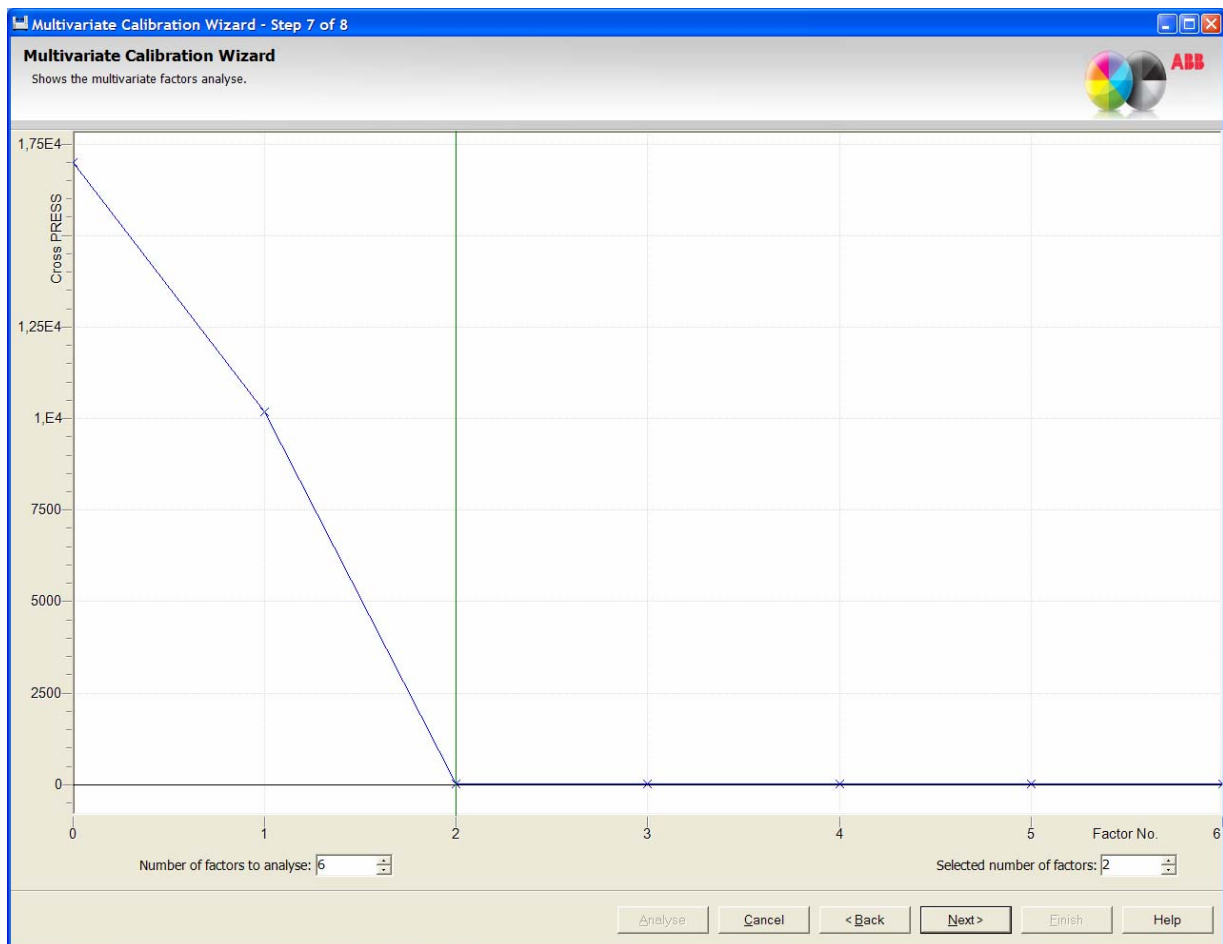
Click **Next** to proceed.

1.8. Step 7: Multivariate Factor Analysis

The data plot appearing on screen is called the **PRESS** (Predicted Residual Error Sum of Squares) plot. It gives you an indication of the model error vs. the number of factors.

Within the wizard the software proposes a number of analyzed factors. This number has to be confirmed or modified interactively. This is done either by moving the vertical line or increasing / decreasing the values using the spin boxes.

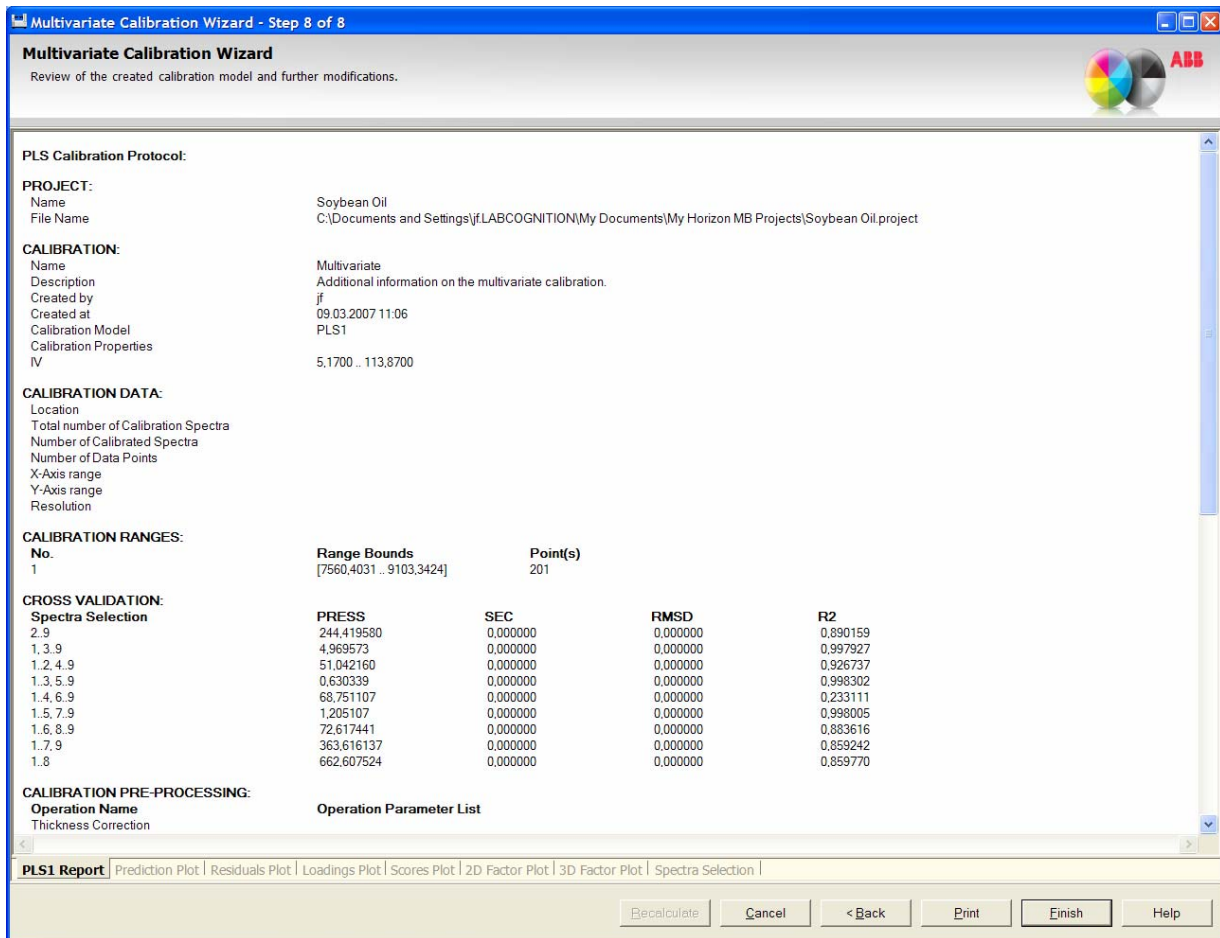
The green vertical line indicates the automatically proposed number of factors, which should be considered the optimum. In the present case, the number of factors at the minimum is two. But keep in mind, this number can be changed as often you want.



Click **Next** to proceed.

1.9. Final Step: Calibration Model Results– Ready for Review and interactive Optimization

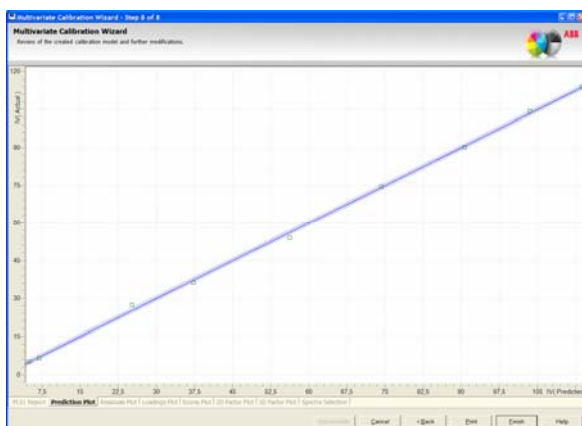
The final step presents a summary of evaluation results on multiple screens. Our example shows the **PLS Calibration Report** first. The report contains an overview over all previous settings including calculation results like basic project information, spectrum selection, factor selection, PRESS values, prediction results, calibration statistics and others.



Some graphical representations of results are available too:

1.9.1. Prediction Plot

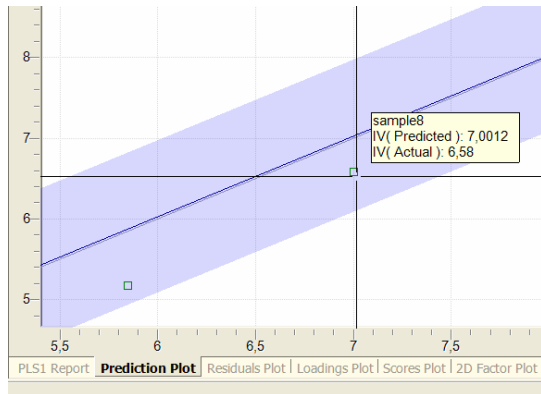
This is the calibration curve. It displays the plot of Predicted vs. Actual values.



The screen shows three different types of information:

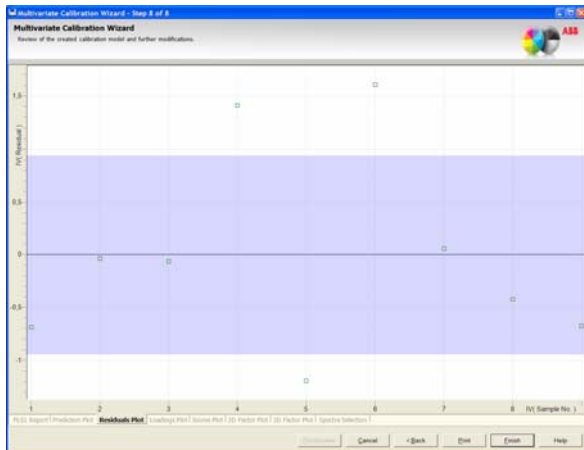
- The light gray diagonal line is the indicator for identity where predicted values were identical with actual ones.

- The blue shape defines the confidence interval.
- A small square defines one sample. Moving the mouse pointer over a square shows a tooltip with plot specific information.



1.9.2. Residuals Plot

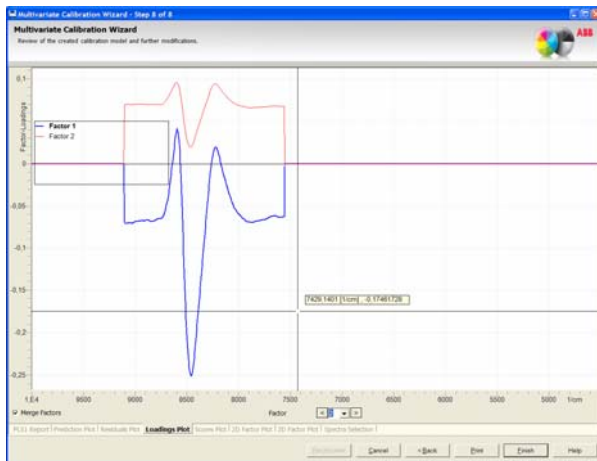
That's the graphical display of the **Concentration Residuals**, it shows the difference between the actual values and the PLS predicted values.



- The small vertical line is the indicator for identity where the difference between predicted and actual value is zero.
- The blue shape again defines the confidence level
- A small square defines one sample. Moving the mouse pointer over a square shows a tooltip with plot specific information.

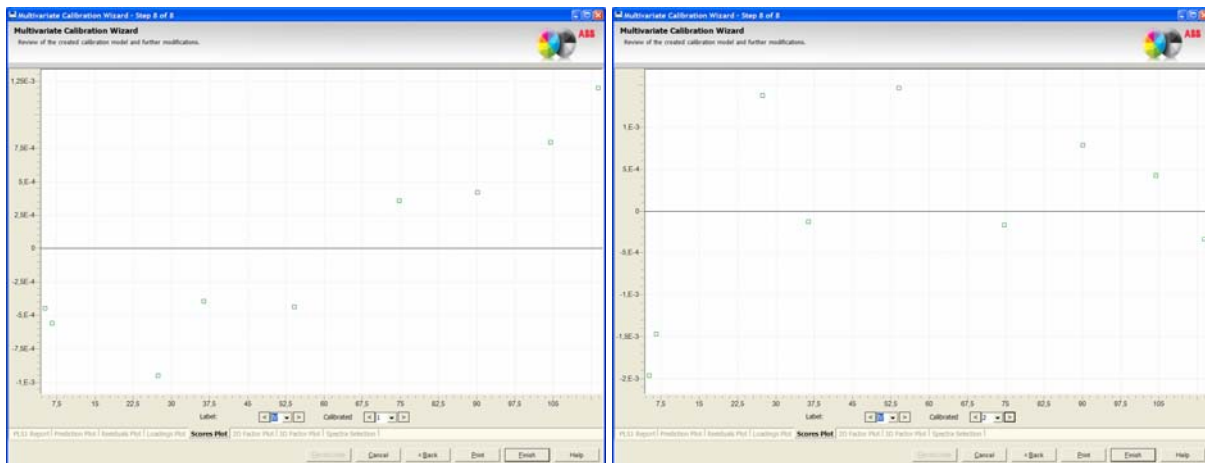
1.9.3. Loadings Plot

The loadings plot is the plot for X-Factors. These factors allow the reconstruction of the spectra based on defined algorithms:



1.9.4. Scores Plot

The model scores can be plotted in several 2D plots, whereby the factors can be chosen. Both views show that all samples are well distributed; no aggregation or clouding can be observed. This indicates that there might be no outlier.

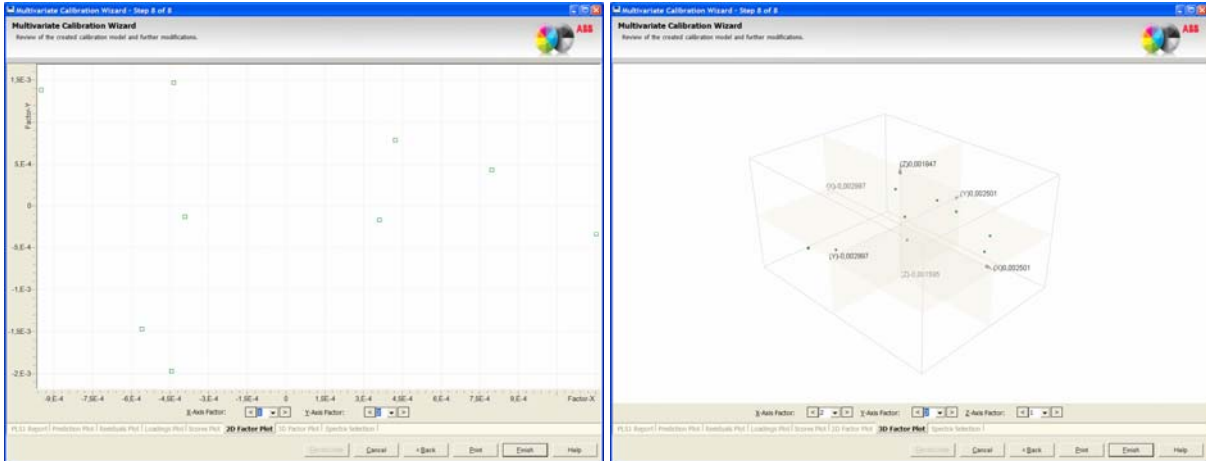


1.9.5. 2D and 3D Factor plots

Factors can be displayed either in a 2D or 3D representation. Outlying samples can be easily determined.

NOTE: the 3D view only shows relevant results, if at least 3 factors have been chosen for the model.

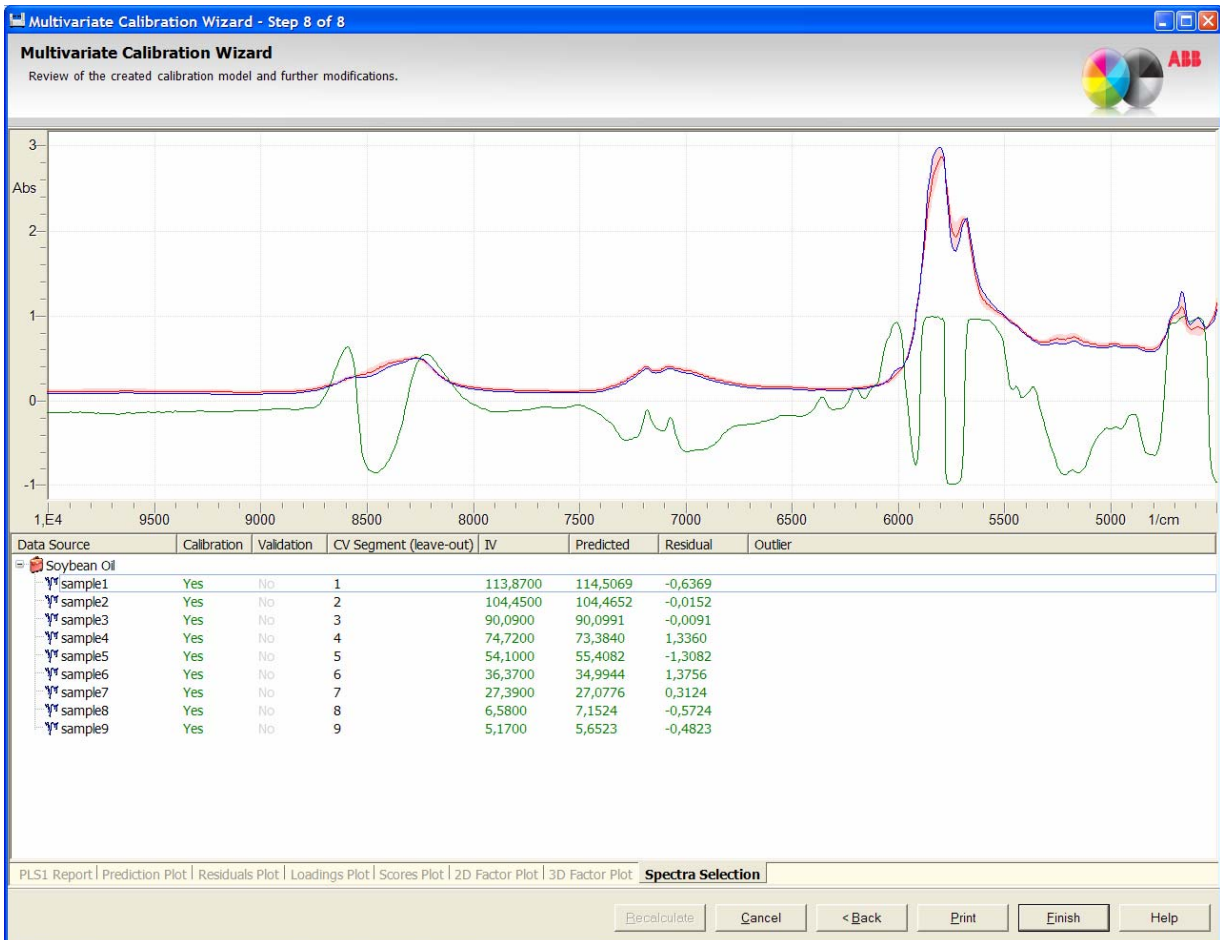
Multivariate Calibration Quick Guide



1.9.6. Spectra Selection

The final tab allows interactive modification of the calibration model. You may choose to remove one or more samples from the calibration model. Simply mark the sample of interest, press the **right mouse** button and select **Remove from Calibration Set** from the pop-up menu.

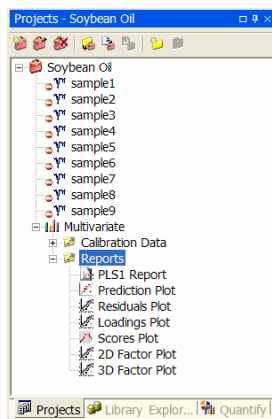
NOTE: If any setting or value has been changed, the **Recalculate** button becomes active whilst the Finish button is becomes disabled.



1.10. Saving the Calibration Model

When all parameters are set and you are satisfied with the results press **Finish** to save the calibration and leave the wizard.

A new calibration node with the calibration name is created in the Soybean Oil project.



The project is automatically updated on disc.

1.11. Expected Results

In the following an excerpt of the expected results is shown.

HORIZON MB

PLS Calibration Protocol:

PROJECT:

Name	IV-Project
File Name	C:\Documents and Settings\ks.LABCOGNITION\My Documents\My HORIZON MB Projects\IV-Project.project

CALIBRATION:

Name	PLS1-CV_full-No_TC
Description	PLS1 Calibration with full Cross Validation (Leave one out method) Investigated Range: 9100-7560 4 Factors
Created by	
Created at	
Calibration Model	PLS1
Calibration Properties	

Multivariate Calibration Quick Guide

IV 5,17000000 .. 113,87000000

CALIBRATION DATA:

Location

Total number of Calibration Spectra 9

Number of Calibrated Spectra 9

Number of Data Points 715

X-Axis range 4497,66845703 .. 10005,96191406 [1/cm]

Y-Axis range 0,08238769 .. 0,53330553 [Abs]

Resolution 7,71469672 [1/cm]

CALIBRATION RANGES:

No.	Range Bounds	Point(s)
1	[7560,40305429 .. 9103,34239799]	201

The following table lists the calibration results.

FACTORS SELECTION:

Factor Count	PRESS	SECV	R2
1	1524,72071	19,5238362	0,886502196
2	6,45249348	1,27008794	0,999519687
3	4,58619216	1,07076984	0,999658611
4	2,62092612	0,809463731	0,999804902

CALIBRATION RESULTS:

Spectrum Name	Actual	Predicted	Residual
---------------	--------	-----------	----------

sample1	113,87	114,013416	-0,14341305
sample2	104,45	103,734365	0,715632412
sample3	90,09	90,6419508	-0,55195451
sample4	74,72	74,5906981	0,129303161
sample5	54,1	55,0369288	-0,93693033
sample6	36,37	35,7955613	0,574437646
sample7	27,39	26,8130219	0,576977476
sample8	6,58	6,47929969	0,10070023
sample9	5,17	5,63475312	-0,46475304

2. Prediction of unknown Samples

There are several ways of predicting unknown samples

- a) Auto Evaluation with a loaded calibration model
- b) Evaluation based on a user defined calibration model
- c) Evaluate with all calibration models opened in projects.