NEAR-INFRARED SPECTROSCOPY

Pavel Matějka

Introduction

Near-infrared spectroscopy (or near-infrared (NIR) spectrometry) is a method of molecular spectroscopy, that uses spectral range of near-infrared radiation, i.e. a range of wavelengths 800 – 2500 nm, that is equivalent to range of wavenumbers 12500 – 4000 cm^{-1}. NIR region is adjacent to the visible range at short wavelengths; at long wavelengths it borders the mid-infrared region. The limits are not given strictly and different values can be found in literature; they can be derived from possibilities of spectrometers to cover a specific range or they can be related to the types of quantum transitions observed.

Absorption of radiation in the NIR region is caused usually by quantum transitions between vibration levels of molecules[1], namely to combination transitions (modes)[2] and overtones[3], not to fundamental modes, which play a dominant role in mid-infrared range (MIR). Both combination modes and overtones are markedly less probable than fundamental modes, therefore absorption of radiation in NIR region is (using the same pathlength) significantly (usually two to three orders of magnitude) weaker than in MIR region. Hence, the transmission cells with longer pathlengths (typically several millimetres) are frequently used. An assignment of absorption bands to individual combination modes and overtones is rather difficult. Hence, the interpretation of spectra based on the concept of characteristic bands of functional groups (applied to spectra in MIR region) is not used commonly, though some types of groups (e.g. amides, esters) can be identified quite easily. Furthermore, it is possible to specify characteristic ranges, where are dominant bands of (a) combination modes (ca. 4000 – 5300 cm^{-1}), (b) first overtones (ca. 4600 – 7300 cm^{-1}), (c) second overtones (ca. 6000 – 10000 cm^{-1}) and (d) third overtones (ca. 8800 – 14500 cm^{-1}). From the point of view of qualitative information it is possible to compare measured spectra of pure substances with library spectra (some commercial libraries are available, user libraries can be created). Hence, precise and reliable identification of substances is obvious. For example, spectral libraries of polymers and pharmaceutically important chemicals are installed in the student's

[1] For several types of substances and materials we can observe in NIR region transitions between specific electronic quantum states. It is the case of some coordination compounds of transition metals or lanthanoides. Such types of transitions are omitted both in this text and corresponding laboratory exercise.

[2] Combination modes represent simultaneous excitations of several vibration modes (energy of a given combination mode is equal to sum of energies of fundamental transitions of corresponding vibration modes).

[3] Ovetone corresponds to excitation of particular vibration mode to higher excited level. First overtone is thus observed at ca. double energy of corresponding fundamental mode, second overtone at ca. triple, third at ca. quadruple value. For correct description it is necessary to consider anharmonicity of vibration modes.

laboratory. NIR spectra are very often sorted and classified using various chemometric approaches. NIR spectra are used frequently for quantitative analysis, namely for complex samples in various technological branches, for example petrochemical, pharmaceutical, paper-and-pulp or food-processing industries. In many cases it is possible to determine multiple components from a single spectrum without any necessity to separate complex mixtures. In many cases the spectra can be collected using on-line systems directly in an industrial process. NIR spectroscopy is classified as a process analytical method (suitable for PAT – process analytical technology), when the speed of analysis including the possibility of continual on-line monitoring in process plan (at production line) is emphasized. For example, it is possible to determine simultaneously contents of fats, proteins, lactose and urea in milk and dairy products (in various stages of their processing), or contents of ethanol and saccharides in alcoholic drinks (or during their production at various stages of running fermentative processes). In recent years NIR spectroscopy is applied in environmental monitoring or in medicinal chemistry. The measurement itself is quite fast[4], frequently non-destructive and it does not require any special tedious sample preparation and/or treatment. Hence, the use of chemicals, consumption of single-use analytical sets and generation of environmentally hazardous waste are minimized. It is possible to measure samples in glass and several other transparent containers. Water absorbs strongly in some limited parts of NIR region, nevertheless, relatively diluted aqueous solutions can be analyzed satisfactorily. The processing and evaluation of measured spectral data is much more time-consuming and labour-intensive than the measurement itself.

Measurement Techniques of NIR spectra

It is possible to measure NIR spectra as an attenuation of the radiant flux either after passage of light through the sample (transmission measurements) or after reflection of radiation (reflection techniques). Within the reflection techniques the most frequently used in NIR spectroscopy are based on the principle of diffuse reflection[5]. In the case of diffuse reflection the incident radiation is reflected from the surface of individual small particles of powdered samples. This approach is applied frequently for analysis of samples in pharmaceutical industry or in industrial production of feed for the agriculture.

Transmission measurements are used predominantly for analysis of liquids, pulpous samples and polymer foils. Liquid samples can be measured in cells manufactured from some types of special optical glasses (INFRASIL, SUPRASIL), which possess high transmission in the whole NIR region. The pathlength of cells is

[4] To collect a single spectrum one minute is a sufficient time, several seconds can be satisfactory in some cases.

[5] Diffuse reflection principle of FTIR spectra measurement is frequently named by abbreviation DRIFT.

usually from 1 mm to 10 mm. The choice of an optimal cell is related both to concentration of an analyte in solution and to the specific optical properties of the solvent used.

Besides the measurements of samples in sample compartment of the spectrometer, NIR spectra are measured often remotely using fibre optics with various types of probes placed directly in system studied, e.g. in chemical or biotechnological reactor.

NIR spectrometer

Both dispersion spectrometers and Fourier-transform (FT) ones are used in NIR region. Dispersion spectrometers cover usually NIR region together with visible one (sometimes the UV range is also accessible). FT spectrometers are designed frequently to cover NIR range and mid-infrared range (MIR), the extension to far infrared range (FIR) is optional. Both dispersion and FT spectrometers are commonly single-beam spectrometers. Hence, to obtain absorption spectral characteristics of the sample itself a collection of reference spectrum is necessary (so called background spectrum). A cheap tungsten lamp can be used as a radiation source for NIR spectrometry, but currently a more efficient tungsten-halogen lamp is applied widely to cover both near-infrared and visible regions. The optical elements for NIR region are manufactured from high-quality quartz glass. Furthermore, durable quartz fibre optics with resistant transmission and/or reflection probes are used for remote sampling and NIR spectra collection.

There are installed two FTIR spectrometers Nicolet NEXUS (Thermo USA) in the laboratory. Their configuration enables to collect spectra either in NIR or in MIR region. They are equipped with two manually exchangeable, automatically recognized beamsplitters (KBr – mainly designed for MIR region with large extension to NIR and CaF_2 mainly intended for NIR region with significant extension to MIR), two software exchangeable radiation sources (ceramic, refractory composite source Everglo for MIR with extension to NIR and tungsten halogen lamp for NIR with certain expansion to both MIR and visible region) and two software exchangeable detectors (DTGS for MIR with extension to NIR and very sensitive InGaAs for NIR). The proper beamsplitter has to be inserted manually before starting the software OMNIC. All other basic settings are adjusted automatically based on the predefined “experiments“. An “experiment” appropriate to the sample compartment accessory used has to be selected from the menu immediately after the start-up of OMNIC. Finally, the settings can be optimized to the specific problem solved in menu “Collect – Experiment Setup…“. In the windows (tabs) of “Experiment Setup” the diagnostic tools are available to check the instrument settings and functionality of individual parts (source, interferometer, detector, electronics) of the instrument.

Quantitative analysis

NIR spectroscopy is predominantly used in quantitative analysis. Generally, in the case of transmission measurements in any optical spectroscopy the fundamental law is Lambert-Beer equation. For every individual component marked **i** of a complex mixed sample a following relation can be written:

$$A_{\lambda,\mathbf{i}} = \boldsymbol{\varepsilon}_{\lambda,\mathbf{i}}\, b\, c_{\mathbf{i}} \qquad (1),$$

where $A_{\lambda,\mathbf{i}}$ is the contribution of **i**-th component to the total absorbance A_λ observed at given wavelength λ, $\boldsymbol{\varepsilon}_{\lambda,\mathbf{i}}$ is a molar absorption coefficient of **i**-th component at given wavelength λ, b is pathlength of absorbing medium, $c_{\mathbf{i}}$ is concentration of **i**-th component in a mixture. The total absorbance A_λ at given wavelength λ is a sum of absorption contributions from all m independent components of the system studied:

$$A_\lambda = \sum_{i=1}^{m} A_{\lambda,\mathbf{i}} \qquad (2).$$

Considering the broadness of bands v NIR region for pure substances and taking into account the possibilities of overlaps of bands of individual components and mutual interactions of components varying upon changes of their concentrations affecting finally the shape of corresponding absorption bands, the simple principle of Lambert-Beer's law is not fulfilled in most cases of use of NIR spectroscopic method. For calibration in NIR spectroscopy it is necessary to develop calibration models using advanced multivariate chemometric algorithms, which require large sets of standards (commonly more than 30 calibration samples). The set of calibration samples has to be sufficiently representative; it should cover the whole expected or anticipated variability range of sample's characteristics, which should be quantitatively analyzed. Not only the variation of analytes quantities but also the changes of other physical and chemical factors have to be considered. Preparation of an appropriate set of calibration (and validation) samples requires careful planning of an experiment. Multiple linear regression (MLR) was the primary algorithm for development of calibration model, but now two advanced regression methods are mainly used in chemometric software packages, i.e. the principal component regression (PCR) and the method of partial least squares (PLS). In both these regression methods the broad spectral regions or complete NIR spectra are used in calibration model instead of individual values of absorbance in specified maxima of selected bands. The aim is to find out an appropriate relation between multidimensional spectral information (represented by matrix of values of absorbance in selected spectral ranges for set of spectra of calibration samples) and composition of samples (represented by matrix of values of concentrations of individual monitored

analytes in the set of calibration samples). In some cases the absorption bands are too broad, therefore spectra for calibration models have to be pre-processed, for example the first or second derivatives of experimental spectra are calculated and used. Experimental parameters of spectra collection and the way of their processing have to be preserved in all steps from calibration, through validation to analysis of unknown samples.

The development of robust calibration models and their proper validation are the key, quite time-consuming steps of quantitative analysis based on NIR spectroscopy. Not only preparation of large set of calibration samples, but also selection of appropriate set of validation samples should be considered. Furthermore, various approaches to preliminary spectral data processing have to be tested. Finally, various regression methods have to be performed and compared and an optimization of their parameters is necessary based on a set of verification steps and testing of outliers. A set of validation measurements is essential for assessment of performance characteristics of calibration model. Undepreciated validation is based on analysis of independent set of samples with known composition using the developed calibration model. For preliminary validation of calibration model various faster approaches are applied. So called cross-validation procedures (e.g. random, systematic or full) based on common “leave-one-out” algorithm are frequently used.

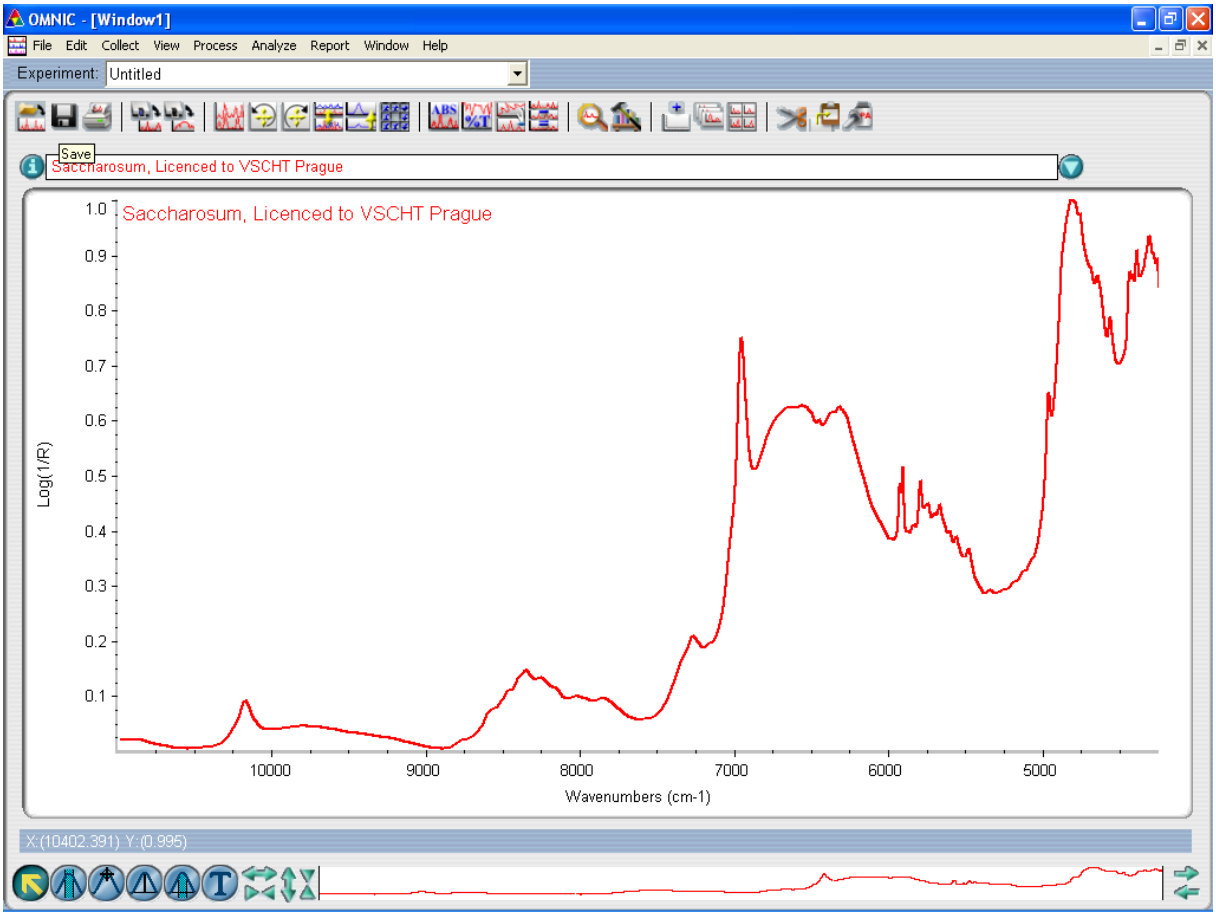
Tasks

1. Measure spectra of samples of water, acetone and stock aqueous solution of sucrose in cells with pathlength 1, 2 and 5 mm. Verify on selected bands the dependence of absorbance values on the pathlength of absorption media.
2. From the stock solution of sucrose prepare a set of calibration solutions, appropriate for correct development of calibration models, allowing determination of the content of sucrose in aqueous medium.
3. Measure spectra of the set of calibration solutions in an appropriate cell. Two spectra per solution have to be collected with repetitive refilling of the cell.
4. Develop a calibration model for determination of sucrose in aqueous solutions, study the influence of selected spectral range, regression method and number of factors (principal components) on quality of calibration model (half of all measured spectra should be used for calibration, the second half is assigned for validation).
5. Used the developed model for analysis of spectra of several given unknown samples.

Measurement of NIR spectra

NIR spectra are collected using the software OMNIC. The basic setup of the instrument and computer for measurements, the selection of appropriate

“experiment” and some other parameters have to be done by teacher. FTIR spectrometer Nicolet NEXUS is a single-beam spectrometer, thus the reference single-beam spectrum („background“) using a command “Collect Background” has to be collected prior any measurement of samples. After insertion of the cell with the sample in the sample holder placed in sample compartment of the spectrometer the sample spectrum is recorded using the command “Collect Sample”. Measured spectrum has to be saved immediately using the icon “Save” to minimize any risk of data loss. It is advisable, to fulfil the “Spectrum Title“ precisely before measurement to use it finally as a file-name. The content of “Spectrum Title” has to identify easily every measured sample, number of repetitive measurement, the pathlength of the cell and the operator person. The unique identification of every spectral data is essential for development of calibration model without mistakes, for any assignment of spectra to calibration or validation group, and for the appropriate assignment of all known quantitative characteristics. The standard rules for data description (encoding) will be specified at the beginning of your work. Every measured data of a student have to be saved in one designated directory. The original data have to be maintained unchanged in this directory. For further processing and data evaluation you should create duplicates of spectra by copying them to another (working) directory. From this (working) directory you’ll open data for development of calibration models or in the case of other requisite tasks, to minimize any risk of random overwriting of original measured data with modified data.



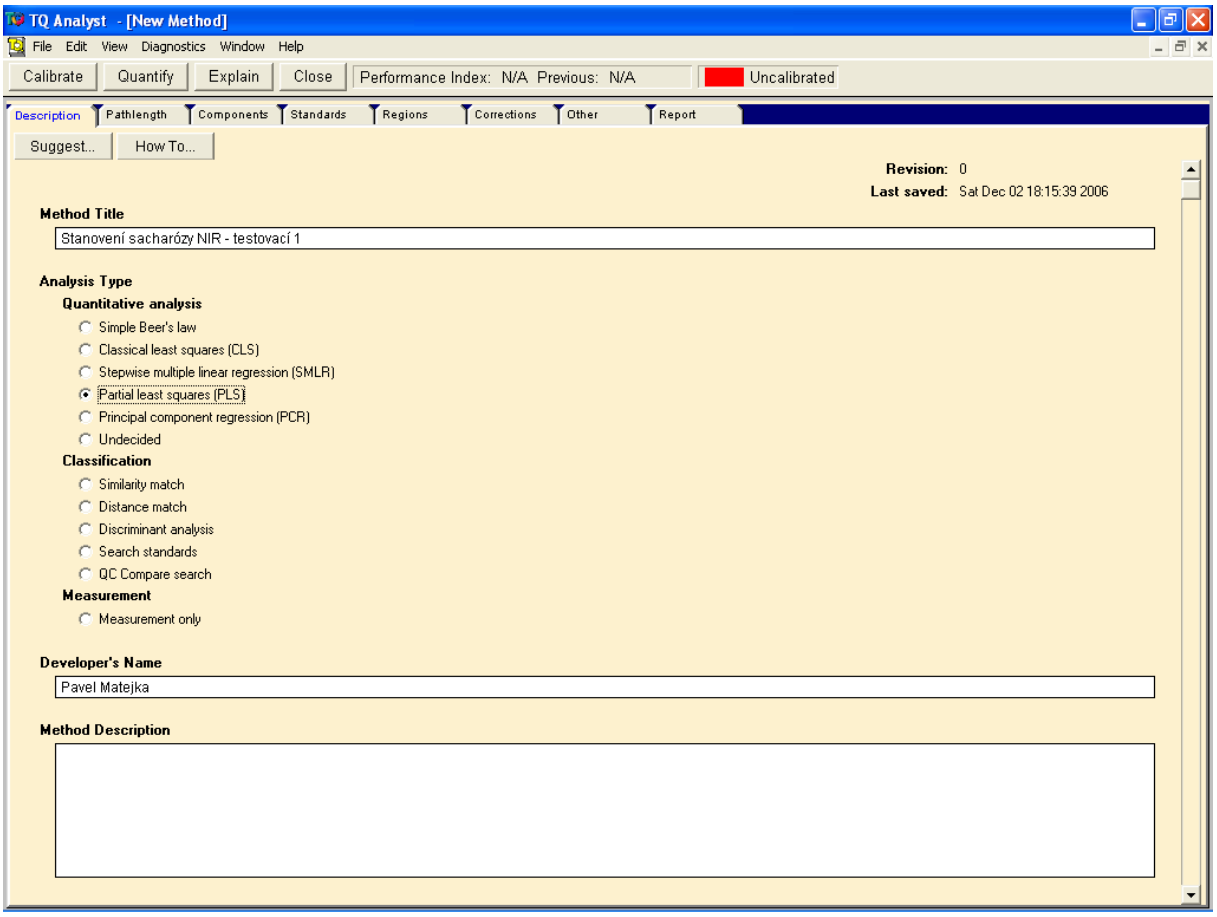
The starting window of program OMNIC is a standard window with menu bar, icon bar, experiment bar selection and a spectral frame.

Development of calibration model

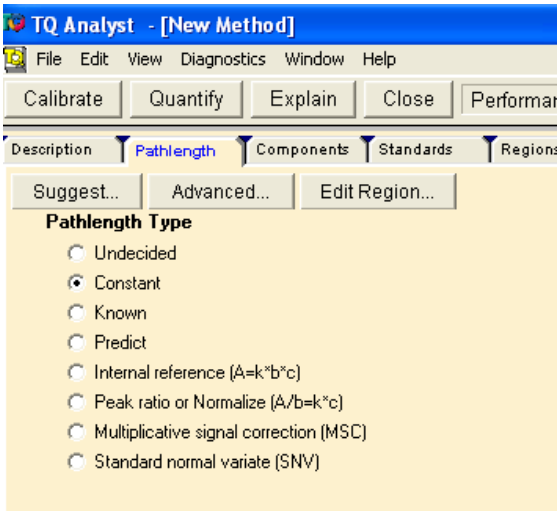
The key step of quantitative analysis based on NIR spectroscopy is careful development and optimization of a calibration model. The development of calibration models will be performed in a program TQ Analyst, that cooperate closely with the program OMNIC. No data-format conversion of saved spectral data is necessary.

The use of software TQ Analyst is user-friendly, besides the standard menu ”Help” you can find help in a default button bar using the button „Explain“ that opens context help related to selected tab. In the course of development of calibration model you have to save it after processing of any tab to minimize any risk of data loss of recorded information. When you modify any parameters of your model, save it as a different file using different filename; a consecutive numbering of models is recommended. In any time you can see on the main bar the status of your model that means “Uncalibrated” or “Calibrated” method. The calibration is performed using a button “Calibrate” after fulfilment of values in all tabs. On the main screen you can see date and time of last savings of the model and the sequential index of revision. Detailed information on the development status is thus available. The aim is to learn

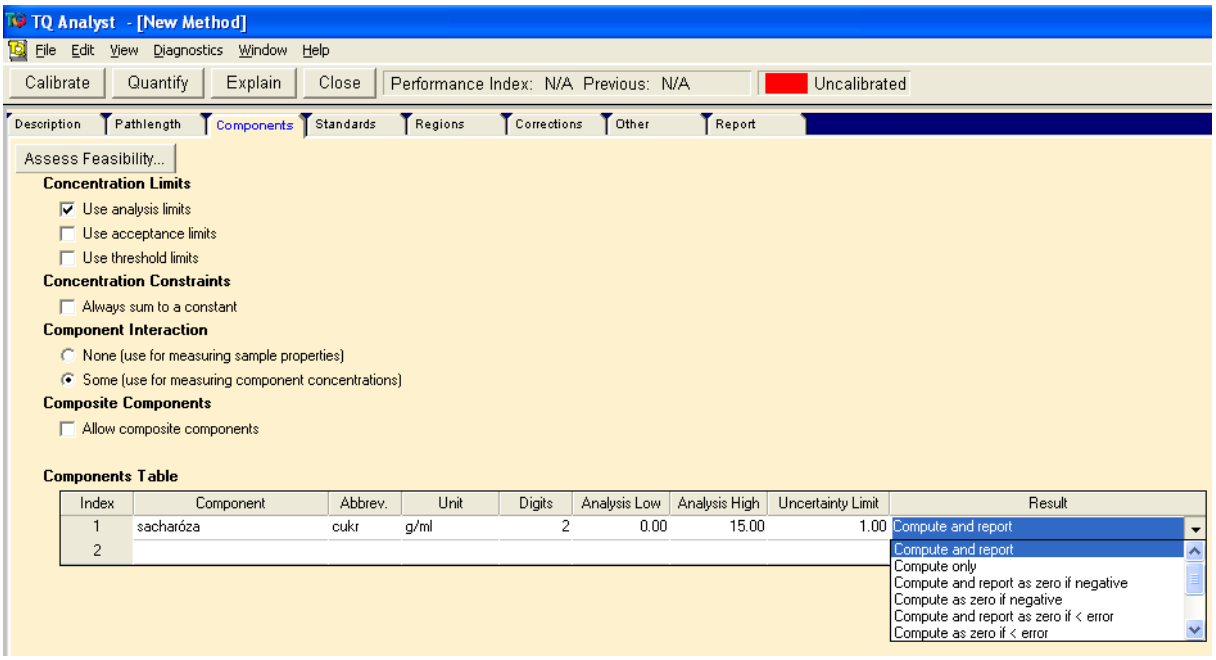
the principles of calibration model creation. As the time of laboratory exercise is limited the number of evaluated samples is quite small and the number of calibration and validation spectra is very limited causing lower reliability of calibration models and sometimes quite small values of performance characteristics of computed models.



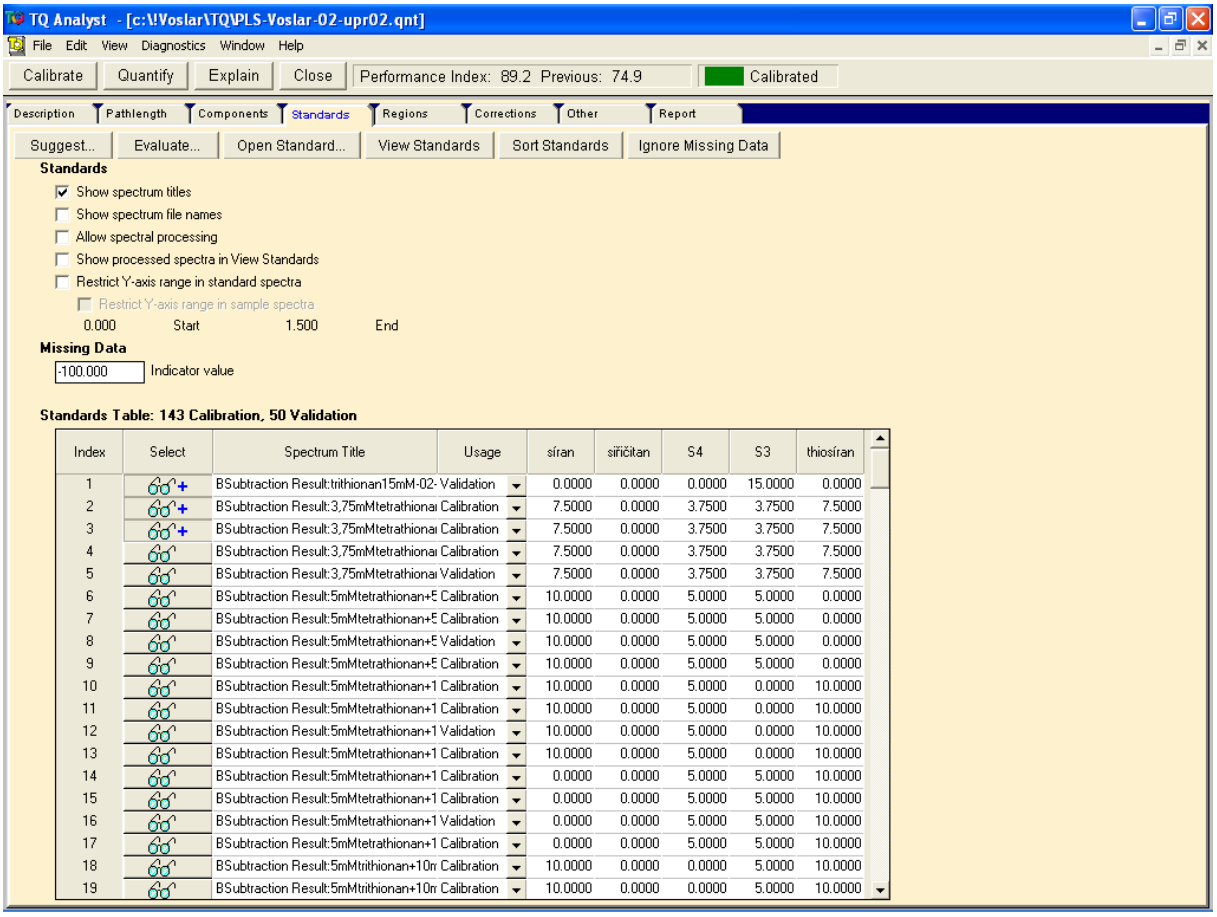
On the initial tab “Description” you should fulfil the method title, type of regression method (analysis type) and your name. Then you switch to the tab “Pathlength” and select the appropriate pathlength type. Select “Constant” for measurement in transmission cell.

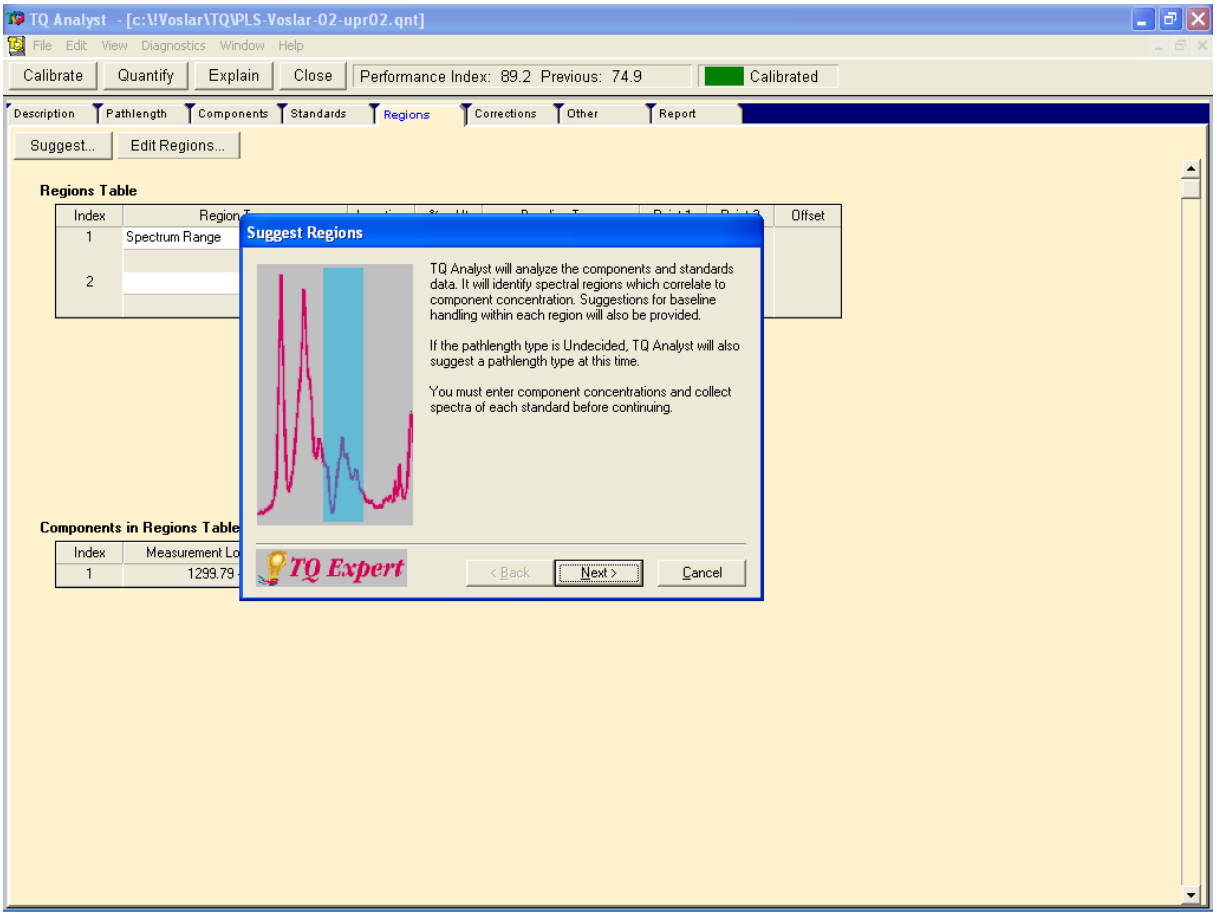


In the next step you switch to the tab “Components“, where it is absolutely necessary to fulfil precisely the table named “Components Table”. You have to give a concrete name of determined component, its appropriate abbreviation, units of the quantity used for quantitative analysis, lower and upper limits of quantification, acceptable limits of uncertainty and the way of presentation of computed results. In this table is crucial to specify each of the determined analytes, because the fulfilment of this table influences critically the view content of other tabs needed for calibration.



The next tab “Standards” should contain information on all standards spectra measured and on composition of all standards samples. Using the button “Open Standard…” you can select and open spectra of measured samples, which you want to use for either calibration or validation. Then in a table you have to specify usage of the corresponding spectrum and you have to fulfil carefully data on known composition of corresponding calibration/validation sample.

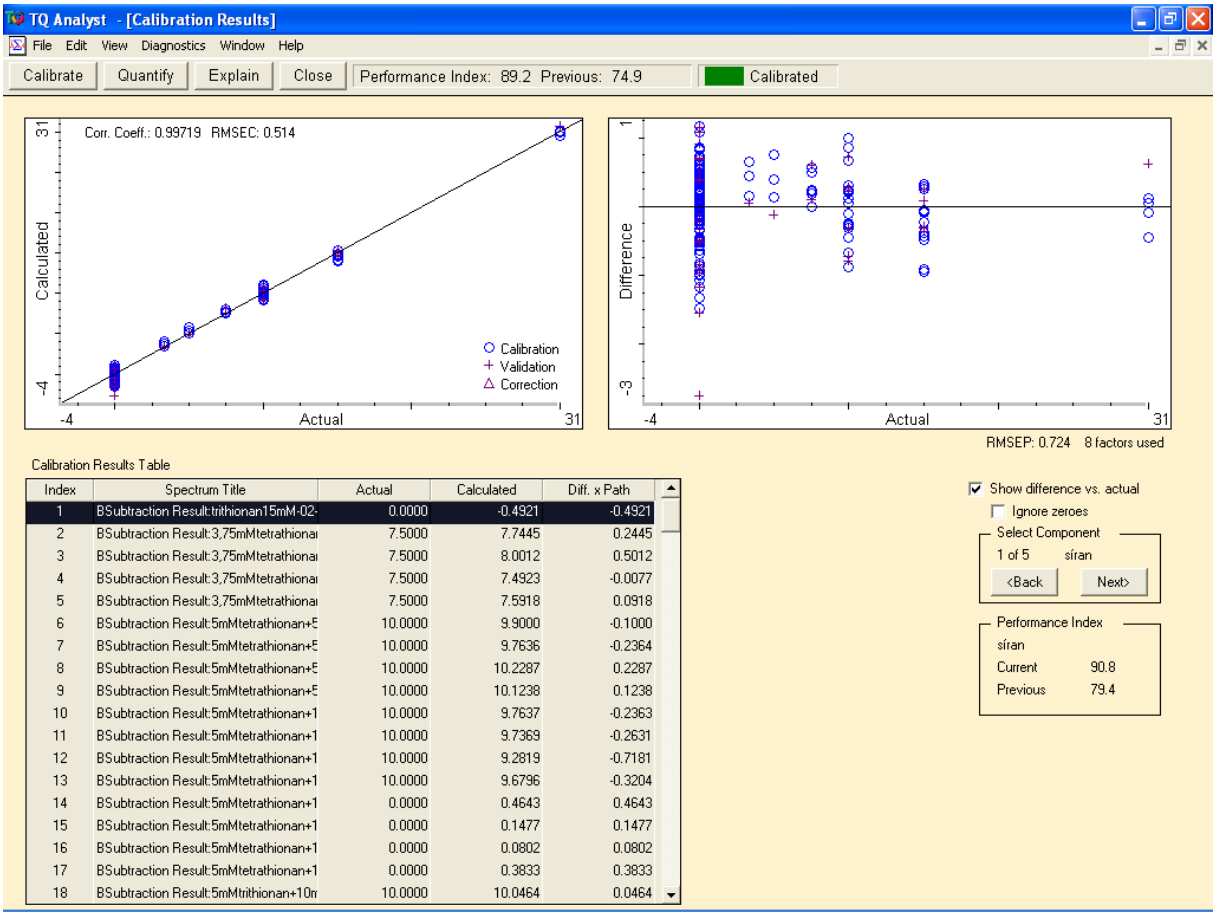

TQ Analyst - [c:\!Voslar\TQ\PLS-Voslar-02-upr02.qnt]
File Edit View Diagnostics Window Help
Calibrate Quantify Explain Close Performance Index: 89.2 Previous: 74.9 Calibrated
Description Pathlength Components Standards Regions Corrections Other Report
Suggest... Evaluate... Open Standard... View Standards Sort Standards Ignore Missing Data
Standards
Show spectrum titles
Show spectrum file names
Allow spectral processing
Show processed spectra in View Standards
Restrict Y-axis range in standard spectra
Restrict Y-axis range in sample spectra
0.000 Start 1.500 End
Missing Data
-100.000 Indicator value
Standards Table: 143 Calibration, 50 Validation
Index Select Spectrum Title Usage síran siřičitan S4 S3 thiosíran
1 BSubtraction Result:trithionan15mM-02- Validation 0.0000 0.0000 0.0000 15.0000 0.0000
2 BSubtraction Result:3,75mMtetrathionat Calibration 7.5000 0.0000 3.7500 3.7500 7.5000
3 BSubtraction Result:3,75mMtetrathionat Calibration 7.5000 0.0000 3.7500 3.7500 7.5000
4 BSubtraction Result:3,75mMtetrathionat Calibration 7.5000 0.0000 3.7500 3.7500 7.5000
5 BSubtraction Result:3,75mMtetrathionat Validation 7.5000 0.0000 3.7500 3.7500 7.5000
6 BSubtraction Result:5mMtetrathionan+5 Calibration 10.0000 0.0000 5.0000 5.0000 0.0000
7 BSubtraction Result:5mMtetrathionan+5 Calibration 10.0000 0.0000 5.0000 5.0000 0.0000
8 BSubtraction Result:5mMtetrathionan+5 Validation 10.0000 0.0000 5.0000 5.0000 0.0000
9 BSubtraction Result:5mMtetrathionan+5 Calibration 10.0000 0.0000 5.0000 5.0000 0.0000
10 BSubtraction Result:5mMtetrathionan+1 Calibration 10.0000 0.0000 5.0000 0.0000 10.0000
11 BSubtraction Result:5mMtetrathionan+1 Calibration 10.0000 0.0000 5.0000 0.0000 10.0000
12 BSubtraction Result:5mMtetrathionan+1 Validation 10.0000 0.0000 5.0000 0.0000 10.0000
13 BSubtraction Result:5mMtetrathionan+1 Calibration 10.0000 0.0000 5.0000 0.0000 10.0000
14 BSubtraction Result:5mMtetrathionan+1 Calibration 0.0000 0.0000 5.0000 5.0000 10.0000
15 BSubtraction Result:5mMtetrathionan+1 Calibration 0.0000 0.0000 5.0000 5.0000 10.0000
16 BSubtraction Result:5mMtetrathionan+1 Validation 0.0000 0.0000 5.0000 5.0000 10.0000
17 BSubtraction Result:5mMtetrathionan+1 Calibration 0.0000 0.0000 5.0000 5.0000 10.0000
18 BSubtraction Result:5mMtrithionan+10r Calibration 10.0000 0.0000 0.0000 5.0000 10.0000
19 BSubtraction Result:5mMtrithionan+10r Calibration 10.0000 0.0000 0.0000 5.0000 10.0000


TQ Analyst - [c:\!Voslar\TQ\PLS-Voslar-02-upr02.qnt]
File Edit View Diagnostics Window Help
Calibrate Quantify Explain Close Performance Index: 89.2 Previous: 74.9 Calibrated
Description Pathlength Components Standards Regions Corrections Other Report
Suggest... Edit Regions...
Regions Table
Index Region Offset
1 Spectrum Range
2
Components in Regions Table
Index Measurement Lo
1 1299.79
Suggest Regions
TQ Analyst will analyze the components and standards data. It will identify spectral regions which correlate to component concentration. Suggestions for baseline handling within each region will also be provided.
If the pathlength type is Undecided, TQ Analyst will also suggest a pathlength type at this time.
You must enter component concentrations and collect spectra of each standard before continuing.
TQ Expert
< Back Next > Cancel

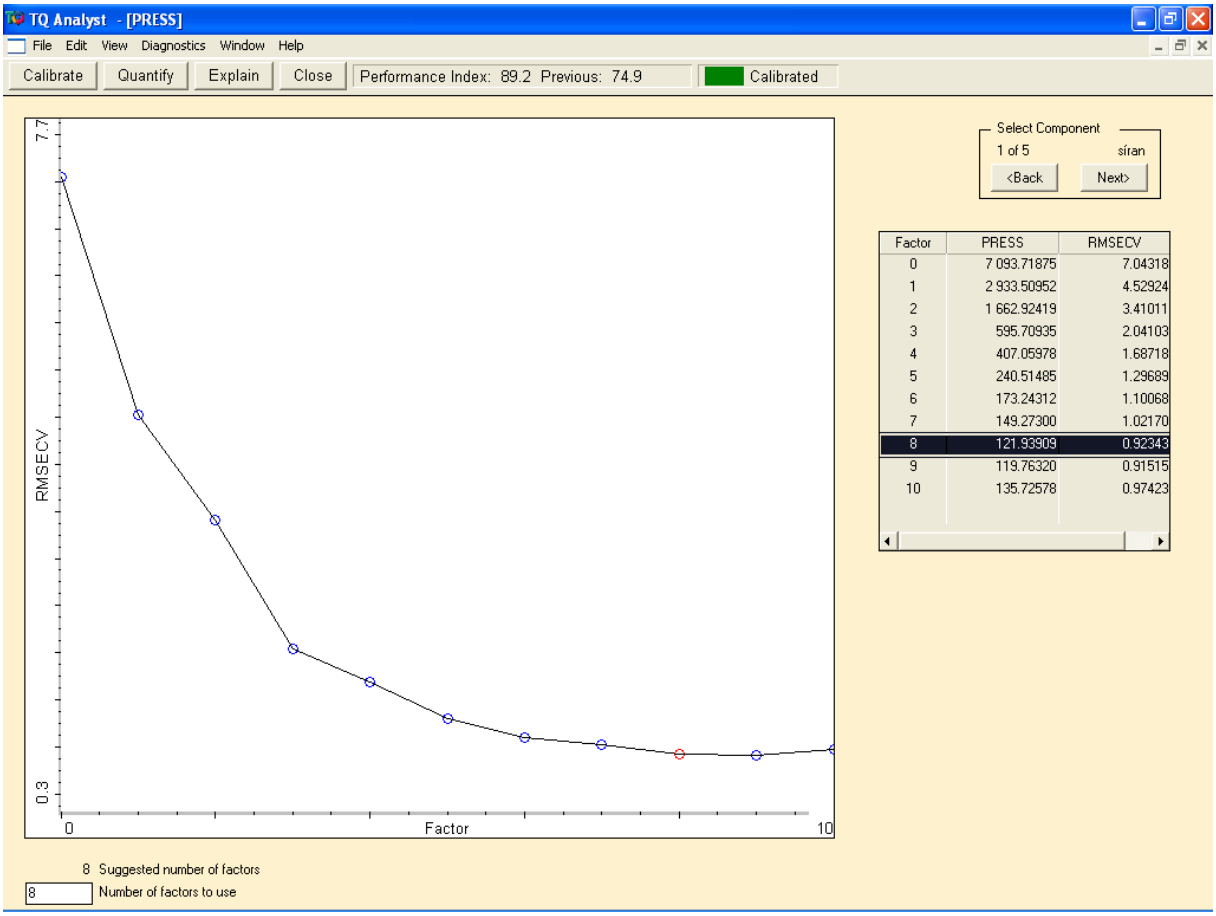
The next tab “Regions“ is related to selection of appropriate spectral regions, which will be processed in the course of calculation of calibration model. For the first suggestion you can use the wizard “TQ Expert“, that is started clicking on button “Suggest…“. Let the “Expert” to explore the whole range of recorded NIR spectra and after completion of the wizard the suggested regions occurred in the table “Regions Table”. Clicking on the button “Edit Regions…“ you can open a window “Region Selection“, where you can interactively modify the preselected regions, view the calculated statistical spectra and models of spectra of pure components (substances) “Pure Component Spectra“. Any acceptable changes are saved clicking on the button “Save“ prior closing the window.

When you specify all determined components, used spectra of standards and their composition and appropriate spectral regions, all data necessary for calibration are rounded up. Now you can try to calibrate your model clicking on button “Calibrate“. After successful calculation of the quantitative model a window ”Calibration Results“ appears, where you can find graphical and tabular form of calibration results, a comparison of calculated and actually given values shows the quality of your calibration model both for calibration and validation spectra. You can recognize outlying measurement/sample and afterwards you try to recalculate calibration after removing outlier from “Standards Table”. An important criterion is the value of performance characteristics called ”Performance Index“, that can take values in an interval from minimal (negative) value –100 (inefficient, very unsuitable model) to maximal (positive) value +100 (highest-rating model). Changes of this value depend on the selection of spectral range, data pre-processing and modification of the list of calibration and validation standards.

For advanced development and optimization of calibration model various tools accessible in menu “Diagnostics“ could be used.



In the case of development of calibration models based on regression methods PCR and PLS the most important diagnostic tool labelled with abbreviation PRESS (“predicted residual sum of squares“) is used to optimize the number of PCs/factors. The values of “sum of squares” are plotted against the numbers of principal component used (PC – “principal component“) in the case of PCR. The number of PLS factors is a similar variable in the case of PLS regression. Optimal number of PCs/factors is found close to minima of the PRESS curve. In this rangem selection of lower number of PCs/factors is recommended, because high number of used PCs/factors lowers the quality of prediction for different measurements from the calibration set. High number of PCs/factors means that the calibration model comprises large contribution of noise from analyzed calibration data; thus the model describes more “perfectly” the calibration data but including all errors, mistakes and random effects contained. Low number of PCs/factors means that the relevant information is not satisfactorily used. Number of accesible PCs/factors je influenced by the number of calibration measurements. The higher number of calibration samples and their spectra, the higher number of PCs/factors can be used.



From the window of PRESS diagnostics you can see the shape of PRESS curve and you should compare it with the theoretical presumptions (curve with distinct minimum, different shapes mean that the model is poor). The number of factors to be tested is apparent (in the case on the figure promising number of factors is from seven to ten) and the automatically suggested number of factors is marked (in this case 8).

Questions

1. Specify the wavelength/wavenumber range typical for near-infrared (NIR) region.
2. Specify quantum transitions which cause absorption of radiation in NIR region.
3. Specify the usage of libraries of NIR spectra.
4. Describe the key analytical application of NIR spectra.
5. Describe various regression methods used for calibration models of NIR spectra.
6. Compare calibration and validation measurements. What is the purpose of both calibration and validation?