Near-Infrared Spectroscopy: Quantitative analysis according to ASTM E1655

Alexander Kadenkin

Near-Infrared spectroscopy (NIRS) is a widely used analytical technique for quantitative analysis of various products in research and industrial applications. This white paper summarizes the workflow of the development of quantitative methods according to ASTM E1655.

Metrohm International Headquarters; Herisau, Switzerland

Metrohm
NIRSystems

Metrohm White Paper

Introduction

Near-Infrared Spectroscopy (NIRS) is a widely used analysis technique for qualitative and quantitative analysis in various industry sectors. It is often used for qualitative analysis in terms of incoming raw material inspection, because it enables the possibility to verify rapidly the identity and the quality of the delivered materials. NIRS is able to ensure that the right product is introduced into the production chain. However, the main application field of NIRS is quantitative analysis. The number of possible applications is unlimited. This can be quantitative determination of quality parameters in, e.g., final products or intermediates such as polymer samples, diesel, gasoline, and lubricating oils. Because of the instrumental setup, the analyzer can be used offline in the quality control lab, placed atline close to the production plant or integrated inline or online into an existing production process in order to improve and monitor the process in real time.

One of the hurdles for the introduction of NIRS is quite often the method development, which at first glance appears to be complex. Like various other analysis techniques, NIRS is not an absolute but a relative analysis technique. The NIR instrument needs to be calibrated using samples with known analyte content (chemical properties, e.g., moisture content or physical properties, e.g., viscosity). This is comparable to other analytical techniques such as gas chromatography, X-ray fluorescence and other techniques [1]. In NIRS, samples with known analyte content are measured and the mathematical relationship between the sample's absorbance and the reference values is calculated. Afterwards, the mathematical model can be applied for the analysis of the spectra of unknown samples in order to estimate content of one or multiple analytes simultaneously in routine analysis.

However, NIR spectra are dominated by broad overlapping peaks, which limits the possibilities of the classical univariate regression [2]. The state of the art solution for this problem is the use of advanced calibration strategies, a so called multivariate analysis. Here the mathematical model is calculated using a part of the spectrum or the whole spectrum instead of using the intensity of a single peak. This enhances the analytical figures of merit of the method and enables new application fields.

The development of multivariate methods (chemometrics) is quite often supported by dedicated software, which guides the user through different steps like sample selection, model development, validation etc. Quite often, this procedure appears to be confusing and it is not clear what is really required for the next steps. Questions like «How many samples are needed?» or «Which analytical figures of merit are important?» remain unanswered.

The related information can be found in scientific literature or different norms. One of the typical and clear guidelines for the development of quantitative multivariate methods for NIRS is ASTM E1655 «Standard Practices for Infrared Multivariate Quantitative Analysis» [3]. As mentioned in the second sentence of this practice, it can be applied for near-infrared spectroscopy or mid-infrared spectroscopy. As defined by ASTM the word practice «underscores a general usage principle» [4]. Therefore, it can be used for all types of quantitative applications in the area of NIRS. The workflow described in the standard practice is shown in **Figure 1** and this whitepaper summarizes different steps, which are indispensable for the successful development of multivariate methods according to ASTM E1655.
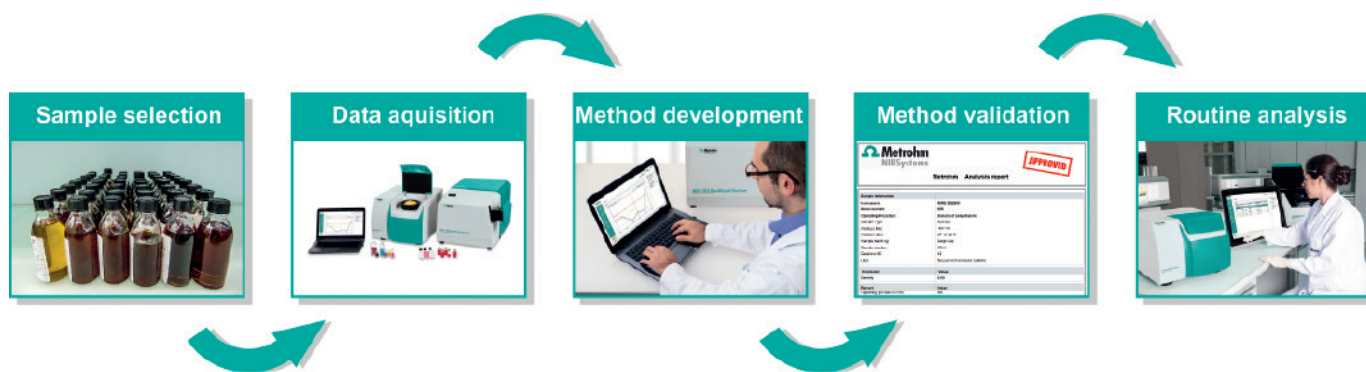

Figure 1: Workflow for the method development.

Calibration set selection

The analytical performance of any application in routine analysis depends on the properties of the samples used for the calibration development. For simple applications with only a few chemical components, samples can be prepared by simple mixing of different components. A typical application where such a calibration strategy can be used is the determination of moisture and denaturant content in ethanol used for gasoline blending.

More complex applications such as quality control of gasoline or polymers should be developed using real samples. Furthermore, sample variation expected in the final routine analysis, e.g., particle size in case of polyethylene analysis or geographical origin in case of petrochemical analysis, should be included in the calibration set. Further typical influence factors and applications, where such factors can have an influence on the final model performance, are summarized in **Table 1**.

One specific case should be mentioned separately. Because of possible interferences of different NIR peaks it is not sufficient to vary only the content this analyte only. A typical example where such a situation can occur is the determination of the hydroxyl number in polyols. As recommended by ASTM D6342, two primary wavelength ranges should be used for the method development: 1 380-1 500 nm and 2 000-2 300 nm [5]. However, in this spectral regions the interference with water absorption bands at 1 450 and 1 930 nm can occur. Therefore, it is essential to include the variation of the moisture content in the calibration set, if it varies in the final routine application. This should be done even if the moisture content is not the analyte of interest.

For applications for which it is not clear, if an NIR method can be developed, ASTM recommends to conduct a feasibility study with 30-50 samples prior to the development of the final method. Here, the concentration range for the analyte of interest should be at least 5 times higher than the reproducibility of the reference method.

For applications where it is generally known that an adequate NIR method can be developed, the calibration should contain not less than 24 samples after the outlier detection, if the number of latent variables is 3 or lower. The term «latent variables» will be explained later in section «Model development». For complex applications with more than three latent variables, such as quality control of petrochemical products, ASTM recommends to use not less than 6*(number of latent variables + 1) samples. In an ideal case, the range of the analyte should be higher than the range expected in the final routine application.

As mentioned above, NIR is a relative analysis technique and therefore the determination of the reference values of the calibration samples is required for the successful calibration development. The standard practice recommends different workflows for the estimation of the reference values and reproducibility and repeatability of the reference method. The simplest case is when the used reference method is an established ASTM method, such as ASTM D2699-16e1 «Standard Test Method for Research Octane Number of Spark-Ignition Engine Fuel» [6]. Here, the repeatability and reproducibility is already included in the final NIR method and the operator can proceed with the method development.

Table 1: **Influence factors and typical applications, where such factors can have an impact on the model performance.**

Influence factor	Example of application
Geographical origin	Mobile fuel screening origin
Manufacturer	Quality control of lubricating oils
Particle size	Analysis of polymer pellets
Sample temperature	Liquid samples like palm oil
Moisture content	Determination of hydroxyl number
Sample presentation	Quality control of natural samples e.g. medical herbs
Process conditions	Atline quality control

Spectral data acquisition

The instrument should be installed and operated in accordance with the instructions of the instrument manufacturer. The standard practice mentions different types of NIR instruments, which can be used. However, as mentioned in the chapter 7, «For most infrared quantitative applications involving complex matrices, it is a general consensus that scanning-type instruments [...] provide the greatest performance, due to the stability and reproducibility of modern instrumentation and to the greater amount of spectral data provided for computer interpretation». As an example of the scanning-type instruments, a dispersive instrument like those provided by Metrohm is mentioned.

The performance of the used analyzer should be monitored periodically using a well-defined procedure. Possible procedures are described in ASTM E275 [7], USP Chapter <1119> [8], or European Pharmacopeia [9] and are beyond the scope of this white paper.

The standard practice describes different modes and adjustable parameters of the data acquisition. It recommends to measure liquid samples in transmission mode, whereas solids should be measured in diffuse reflectance mode. Essential for the method development is to keep all adjustable parameters constant for the measurement of calibration and validation sets as well as for routine analysis. Such influential parameters can be temperature, sampling technique, and others.

Model development

This part is quite often confusing for the end user, because the quantitative method is based on multivariate regression. However, this part is usually less problematic when using a dedicated software like Metrohm Vision Air complete software for the method development, which guides the operator through all steps.

Prior to the method development, different preprocessing techniques can be applied on the spectra in order to remove redundant information, e.g., spectral background. This can dramatically improve the analytical performance of the method. Generally, different preprocessing techniques should be applied iteratively and compared with each other. The optimal preprocessing can be selected based on the analytical figures of merit.

Additionally, the operator can select the spectral region with information related to the change of the constituent values. Such a selection of spectral regions can dramatically improve the analytical figures of merit of the method such as accuracy, precision, and robustness. A typical example of the wavelength selection is described in ASTM D6342, which recommends two specific wavelength ranges for the determination of the hydroxyl number in polyols instead of using the whole spectral range [5].

Three different mathematical algorithms for quantitative modelling are mentioned in the standard practice: multilinearregression analysis (MLR), principal component regression (PCR), and partial-least squares regression (PLS). The best choice is quite often the PLS regression because the estimated analytical figures of merit are usually better than in case of MLR or PCR [2].

Generally, there are two different types of PLS regression. PLS1 regression estimates a mathematical relationship between one analyte and spectral data and is more specific. In contrast, PLS2 regression calculates simultaneously the relationship between multiple analytes and spectra. It is important to mention that for calibration development according to ASTM E1655, only PLS1 should be used. Different specific PLS1 models can be later combined in an operating procedure for routine analysis in order to estimate multiple parameters simultaneously.

The PLS regression method performs a reduction of dimensionality of the data matrix and the regression simultaneously. The basic principle of dimensionality reduction is demonstrated in **Figure 2**. The plot on the left side represents the data in the original three dimensional coordinate system. The PLS algorithm estimates the maximal variance in the original data set and defines a new coordinate system. The first axis in the new coordinate system is given by the direction of the maxi-

mal variance in the original data set. The new axis is called factor or latent variable. For the provided examples it is shown on the left side with a red arrow called «factor 1». All further axes of the new coordinate system are orthogonal to the first one as it is demonstrated on the right side. The dimensionality of the transformed data set is lower. In the example shown, the original information is reduced to two dimensions, factor 1 and factor 2. The algorithm uses this reduced data set for the determination of a mathematical relationship between the spectra and corresponding reference values.
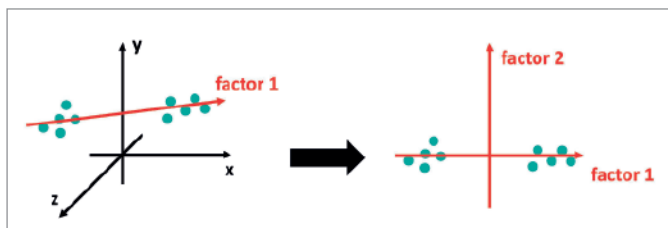

Figure 2: Principle of PLS regression.

Figure 2 explains the PLS regression on a simple three dimensional case. In the area of NIR spectroscopy the dimensionality of the original data set is higher. It is defined by the number of spectral data points in the spectrum or spectral region used. When applying PLS, the spectral original data set with hundreds or even thousands of data points is reduced to usually not more than 10 dimensions.

The residual variance in the original data set unaccounted by these factors forms the residual matrix, which consists of irrelevant information. An example of such irrelevant information are spectral peaks not related to the changes of the analyte content like peaks from the matrix.

The dimensionality of this new coordinate system is defined by the number of so called latent variables or factors. The number of factors has a direct impact on the analytical figures of merit. The operator can select the optimal number of factors. The selection is a critical point and there are no rules, which can be applied for each application. In general, with the increasing number of factors the standard error of the calibration (SEC) decreases (Equation 1).

On the other hand, with the increasing number of factors, the systematic error of the model increases and the robustness of the calibration decreases. This effect is called overfitting. It can be demonstrated on the example of the application for the determination of the hydroxyl value. **Figure 3** shows a relationship between different analytical figures of merit and the number of factors. The standard error of calibration decreases continuously with the increasing number of factors. However, standard errors of cross-validation (SECV) and prediction (SEP) show slightly different behavior. With the increasing number of factors, the gap between SEC and SECV and SEC and SEP increases indicating the increase of the systematic error.
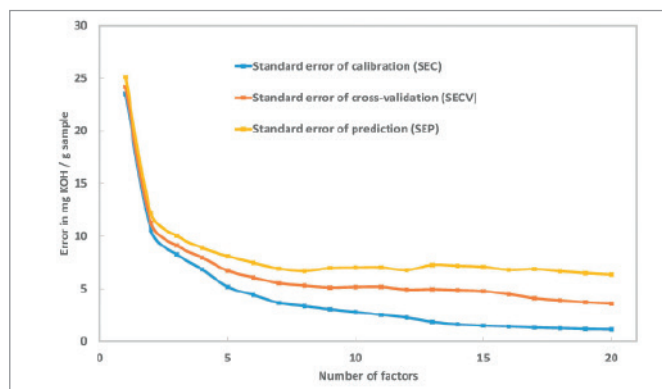

Figure 3: Influence of the number of factors on the standard errors on the example of the application for the determination of the hydroxyl value.

The ASTM practice recommends to use the model with the lowest standard error of cross-validation. However, if more than one model have similar SECV values, the one with the lowest number of factors should be used.

Additionally, it should be mentioned, that the number of factors has a direct impact on the number of samples used for the calibration and validation. After the development of the first model, the user needs to double-check if the criteria for the number of samples are still fulfilled.

Outlier detection
Essential for the development of the calibration according to ASTM E1655 is the outlier detection. It should be applied during calibration as well as during routine analysis.

Two different types of outliers should be identified during the model development. The first type is a spectral outlier. Such outlier can be caused by a sample, which represents an extreme composition to the reminder of the calibration set, e.g. one polyethylene terephthalate (PET) sample that was erroneously added to the calibration set for the quality control of polyethylene (PE). On the other hand, it can be a sample measured under different experimental conditions, e.g., a jet fuel sample measured in a cuvette with a different diameter. Such samples or spectra have a very high leverage on the model and can lead to erroneous predictions when included into the calibration set. The detection of such outliers is based on so called leverage statistics and is implemented in Metrohm Vision Air Complete Software as Mahalanobis distance.

The second type of outlier is observed when the difference between the predicted value and reference value (residual) is significant. Such outliers can arise, e.g., by erroneous reference values. Here, the ASTM norm allows different possibilities such as outlier detection based on standardized residuals. If the sample is recognized as an outlier based on residuals, the norm recommends to repeat the measurement of the reference value and the NIR spectrum in order to find the source of the error.

Validation

After the successful model development, the model needs to be validated using an independent and representative validation set. According to the ASTM practice, separate validation samples should span 95% of the calibration range for each parameter. The samples should be uniformly distributed over the whole concentration range. Furthermore, the number of validation samples should be at least 20, if the number of factors is 5 or lower. Otherwise, the number of samples should be 4* (number of factors + 1). In addition, the sample set should contain all variations expected in routine analysis.

The most important analytical figure of merit estimated during validation is the standard error of prediction (SEP), also called standard error of validation (SEV) or standard error of estimate (SEE). The SEP value should be comparable to the SEC and SECV values estimated before. Additionally, the averaged bias for the validation has to be determined and it significance should be estimated using t test.

It is important to demonstrate the agreement between the model and the reference method. 95% of values by NIR must fall within the interval:

[(measured NIR value - reproducibility) < reference value < (measured NIR value + reproducibility)].

The number of samples used for the estimation of the precision should be equal to the number of factors used in the model development. The samples should span 95% of the concentration range in order to estimate the precision over the whole concentration range. Each sample should be measured at least 6 times. For each sample the average and the standard deviation should be calculated. Finally, the significance of the standard deviation should be evaluated using $\chi 2$ test.

Routine analysis

After its successful validation, the method can be used for routine analysis. However, it is desirable to test the performance of the analyzer and the model on a regular basis. The performance of the analyzer can be monitored according to the instructions of the instrument manufacturer. The performance of the model can be controlled using quality control samples, which should have similar absorbance behavior as the calibration samples. In addition, the QC samples should be measured under the same instrumental conditions as the calibration samples. The complete procedure of the performance monitoring is described in ASTM Practice D6122 [10].

Based on the results of the quality control, it might be necessary to update the model using a slope and bias adjustment in order to improve the model performance. In this case, the model needs to be fully revalidated.

It may be necessary to update the model with new data collected in routine analysis in order to improve the performance or to increase the range of application. In this case the model needs to be redeveloped using the previously described procedure. The updated model can be partially revalidated using the same validation set as for the primary model. However, it is important to add new samples to the validation set according to the previously described conditions (outlier detection, representative samples). Furthermore, the percentage of the samples added to the validation set must be at least as large as the percentage of the samples added to the calibration set.

Metrohm White Paper

Summary

The present white paper summarizes different steps needed for the successful development of quantitative methods using NIRS according to ASTM E1655 (**Figure 4**). Additional information is available in the standard practice itself. The development of quantitative NIR methods according to ASTM E1655 can be supported by your local Metrohm representative.
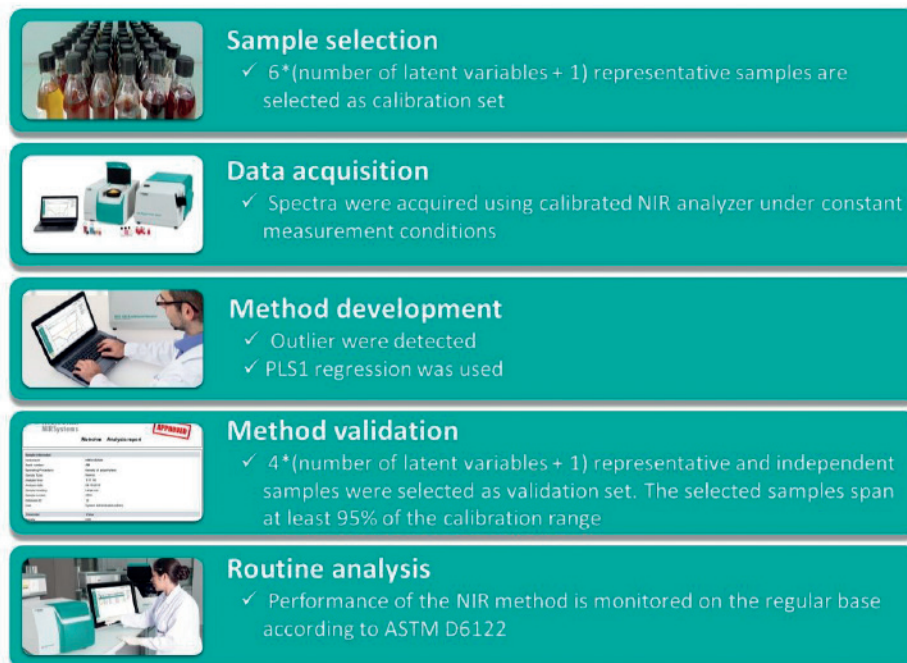
Figure 4: Different steps and key requirements for the successful development of quantitative methods according to ASTM E1655.

References

[1] Encyclopedia of Analytical Chemistry, John Wiley and Sons, 2014.

[2] D.A. Burns, E.W. Ciurczak, Handbook of Near-Infrared Analysis, CRC Press, 2007.

[3] ASTM E1655-05(2012), Standard Practices for Infrared Multivariate Quantitative Analysis, ASTM International, West Conshohocken, PA, 2012, **www.astm.org**

[4] R. Wlhelm, «Know your type of standards», ASTM, 2000.

[5] ASTM D6342-12(2017)e1, Standard Practice for Polyurethane Raw Materials: Determining Hydroxyl Number of Polyols by Near Infrared (NIR) Spectroscopy, ASTM International, West Conshohocken, PA, 2017, **www.astm.org**

[6] ASTM D2699-16e1, Standard Test Method for Research Octane Number of Spark-Ignition Engine Fuel, ASTM International, West Conshohocken, PA, 2016, **www.astm.org**

[7] ASTM E275-08(2013), Standard Practice for Describing and Measuring Performance of Ultraviolet and Visible Spectro-photometers, ASTM International, West Conshohocken, PA, 2013, **www.astm.org**

[8] United States Pharmacopeia 39, 2016.

[9] European Pharmacopoeia (Ph. Eur.), 9th Edition, 2016.

[10] ASTM D6122-15, Standard Practice for Validation of the Performance of Multivariate Online, At-Line, and Laboratory Infrared Spectrophotometer Based Analyzer Systems, ASTM International, West Conshohocken, PA, 2015, **www.astm.org**

Contact: Alexander Kadenkin: alexander.kadenkin@metrohm.com

Ω Metrohm