mlsl-01-2025

White Paper

Calibration of the mylab NIR-Analyzer in Food Analysis

Schewtschenko O, Meuche F., Muecke A, Wietstock S, and Heise J*

fzmb GmbH, Forschungszentrum für Medizintechnik und Biotechnologie; Geranienweg 7;
99947 Bad Langensalza; Germany

*Corresponding Author: jheise@fzmb.de

Introduction

Near infrared spectroscopy (NIRS)

NIR spectroscopy (near infrared red spectroscopy) is a commonly used technology for fast and reliable analysis for a wide range of sample types. The ability of NIR spectroscopy to provide accurate results relies on both the quality of obtained NIR spectrum, as well as the quality of the calibration. The NIR spectrum is obtained from a reflectance or transmission analysis of an illuminated sample. The chemical groups (e.g. hydroxy-, sulfhydryl-, amino-, methin-, carboxy-groups) of the sample matter provide characteristic adsorption properties in the NIR spectrum. NIR spectroscopy (NIRS) determines differences in how NIR energy is specifically transmitted or reflected from a sample. As NIRS is a secondary technique its calibration depends on a primary reference method in order to determine individual chemical components.

NIRS calibration model & parameter value prediction

An NIRS calibration model is a mathematical correlation between the raw NIR spectroscopy data from samples and their determined component of interest by the reference method. It pairs light absorbance or reflectance spectra with accurate laboratory reference values to train predictive chemometric algorithms. A robust NIRS calibration requires a collection of approximately 100 samples representative for future unknown samples. In general, the smaller the number of samples in the dataset, the less accurate the model can predict the parameters of unknown samples. However, this application notes compares the capability of different sizes of datasets for calibration models to predict different parameters in food samples.

mylab NIR-Analyzer instrument

The mylab NIR-Analyzer represents the third generation in development by fzmb GmbH. It is designed for NIR spectroscopy of liquid and solid compounds. Its measuring field with a diameter of 55 mm allows for the complete scan of a standard petri dish and the analysis of inhomogeneous samples without the need of a sample rotator.

Materials and Methods

Sample preparation

Solid sample constituents such as fish, meat and cheese were homogenized using a blender. 100-200 ml of the homogenized sample was transferred to a petri dish (accessory of the mylab NIR-Analyzer) and NIR spectroscopy was performed according to the manufacturer's manual.

Set-up of the mylab NIR-Analyzer calibration model

For the calibration of the mylab NIR-Analyzer the primary reference analyses were performed by the accredited Prueflabor-fuer-Lebensmitteluntersuchungen (www.lebensmittellab.de) according to §64 German Food & Feed Code LFBG.

The calibration of the analytical parameters of the mylab NIR-Analyzer was based on neural networks combined with the GPR algorithm (Gaussian Process Regression). For the training of the neural network each calibration data set (e.g. meat, fish, yeast, diary, flour) was divided into a training set and a test set. The training data were used to build the model. The test data set was used to estimate the quality of the regression models using the calculated RMSEP (Root Mean Square Error of Prediction) and coefficient of determination (R^2).

- Coefficient of determination $\boldsymbol{R^2}$
  R^2 is the degree of linear correlation between prediction $Y_{i,M}$ and reference values $Y_{i,R}$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_{i,R} - Y_{i,M})^2}{\sum_{i=1}^{n}(Y_{i,R} - \overline{Y_R})^2}$$

  with $\overline{Y_R}$ - mean value of the reference values
  The higher the coefficient of determination, the better the correlation between the variance in the concentration and spectral data. Low values of R^2 lead to generally poor analysis results.

- **RMSEP** = **R**oot **M**ean **S**quare **E**rror of **P**rediction
  The RMSEP is the root mean square of the prediction error during external validation. This value represents a quantitative measure of the accuracy with which the samples of the test set were analyzed.

$$RMSEP = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{i,R} - Y_{i,M})^2}$$

Results

Large data set for calibration (>500 samples)

A set of sample data of more than 500 samples ensure high overlaps of the predicted values from NIR spectroscopy and the values obtained from the reference analysis. The scatter plots (Figure 1) derived from 750 different samples from meat and meat products show the content of fat and protein in meat samples by NIRS analysis (mylab NIR-Analyzer) and the corresponding reference analysis.

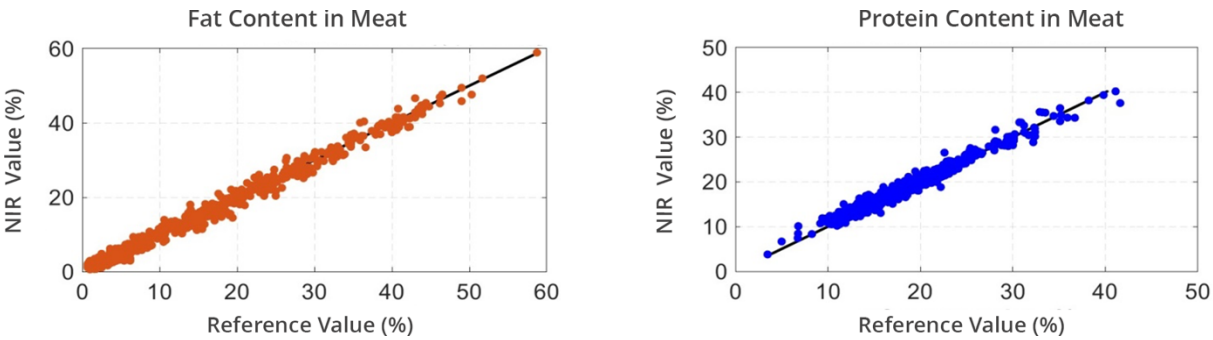

Fig. 1. Scatter plots of reference analysis values from 750 meat samples and predicted values of NIRS-analysis using the mylab NIR-Analyzer.

Compact data set for calibration (75-150 samples)

Figure 2 show scatter plots of fat and protein content of samples from cheese. The data set contained 120 samples.

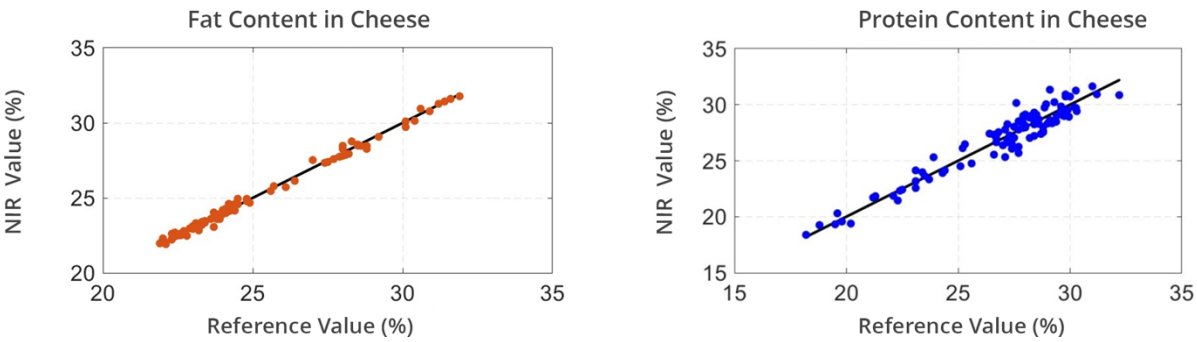

Fig. 2. Scatter plots of reference analysis values from 120 different cheese samples and predicted values of NIRS-analysis for fat and protein content using the mylab NIR-Analyzer.

Custom-specific data set for calibration

Figure 3 show scatter plots of fat and salt content of only 52 samples from fish (Blue Whiting, Herring, Mackerel). The NIR spectra were recorded by the customer using the mylab NIR-Analyzer. The reference values were provided by the customer. The data set (NIR spectra and reference values) were transferred to fzmb GmbH, and the calibration was processed.

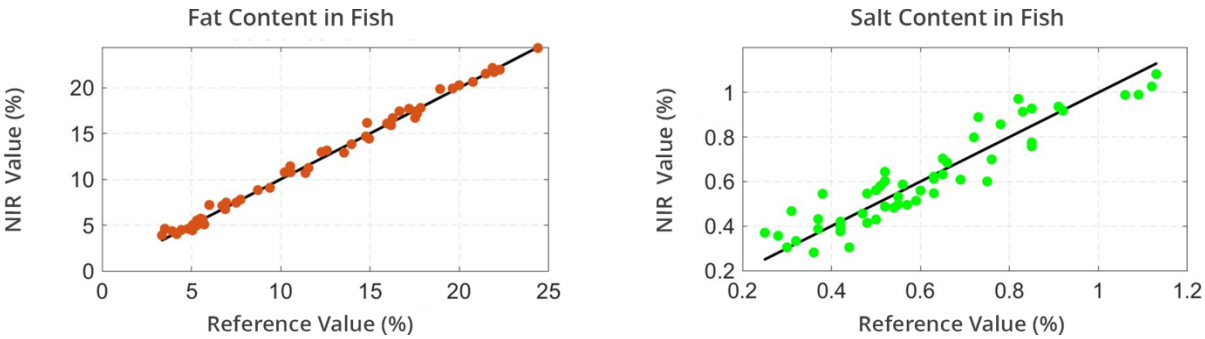

Fig. 3. Scatter plots of reference analysis values from 52 different fish samples and predicted values of NIRS-analysis using the mylab NIR-Analyzer.

Training the mylab NIR-Analyzer for Analysis of Meat and Meat Products

The mylab NIR-Analyzer was trained on 750 samples of meat and meat products to analyze the following parameters: fat, saturated fats, dry matter, sugar**, ash, protein, MPDCP (meat protein devoid of connective tissue protein), water activity (aw-value), pH value, and salt.

The reference analyses were conducted by the accredited Prueflabor-fuer-Lebensmitteluntersuchungen as mentioned earlier Table 1 provides an overview of the range of values for the "Meat and Meat Products" dataset, including the deviations and coefficients of determination for NIRS measurements compared to reference analysis.

Tab. 1. Range value of the data set “Meat and Meat Products”, deviation of the values of obtained parameters by NIRS compared to reference analysis and the certainty of the measurement.

Parameter	Fat	Saturated Fatty Acids	Dry Matter	Sugar**
Range Value* of the Data Set	0.50 - 58.73 g /100 g	0.50 - 58.73 g / 100 g	0,50 - 58.73 g / 100 g	0.50 - 58.73 g / 100 g
Deviation (RMSEP)	1.34 g / 100 g	0.58 g / 100 g	1.00 g / 100 g	0.09 g / 100 g
Coefficients of Determination (R^2)	0.99	0.98	0.99	0.99

Parameter	Protein	Activity of Water	pH-Value	Salt
Range Value* of the Data Set	0.30 - 41.60 g /100 g	0.65 - 1.00 g /100 g	4.40 - 6.92 g /100 g	0.10 - 5.22 g /100 g
Deviation (RMSEP)	1.03 g / 100g	0.01 g / 100g	0.20 g / 100g	0.20 g / 100g
Coefficients of Determination (R^2)	0.97	0.97	0.75	0.95

*Range Value: minmal and maximal values fo the data corresponding data set.

** Sugar: total amount of Saccharose, Glucose, Fructose, Galactose and Lactose according to EU directive No.1169/2011.

Training the mylab NIR-Analyzer for the Analysis of Flour and Cereals

The mylab NIR-Analyzer was trained using 100 flour samples derived from various cereals, including rye, buckwheat, barley, and wheat. The following parameters were analyzed: total protein, gluten, ash, moisture, and sedimentation value. Table 2 presents the range of values for the dataset, the deviation of the parameters obtained via NIRS compared to the reference analysis, and the measurement accuracy.

Tab. 2. Range value of the data set “Flour and Cereals”, deviation of the values of obtained parameters by NIRS compared to reference analysis and the certainty of the measurement.

Parameter	Total Protein	Wet Gluten	Ash
Range Value* of the Data Set	11.7 - 12.9 g /100 g	28.4 - 32.9 g /100 g	0.53 - 0.81 g /100 g
Deviation (RMSEP)	0.081 g / 100g	0.480 g / 100g	0.026 g / 100g
Coefficients of Determination (R^2)	0.91	0.88	0.99

Parameter	Sedimentation Value	Moisture
Range Value* of the Data Set	28.4 - 32.9 g /100 g	13.1 - 15.3 g /100 g
Deviation (RMSEP)	0.480 g / 100g	0.099 g / 100g
Coefficients of Determination (R^2)	0.88	0.97

Training the mylab NIR-Analyzer for the Analysis of Dairies

The mylab NIR-Analyzer was trained using 120 cheese samples, 140 butter samples, and 50 mozzarella samples to determine the following parameters: fat, protein, dry matter, and salt. Table 3 shows the range of values for the cheese dataset, the deviation of the parameters obtained via NIRS compared to the reference analysis, and the measurement accuracy.

Tab. 3. Range value of the data set “Cheese”, deviation of the values of obtained parameters by NIRS compared to reference analysis and the certainty of the measurement.

Parameter	Fat	Protein	Dry Matter	Salt
Range Value* of the Data Set	0.03 - 37.00 g /100 g	18.2 - 30.30 g /100 g	29.7. - 61.00 g /100 g	1.29 - 2.20 g /100 g
Deviation (RMSEP)	0.80 g / 100g	0.53 g / 100g	0.90 g / 100g	0.09 g / 100g
Coefficients of Determination (R^2)	0.99	0.98	0.99	0.88

Training the mylab NIR-Analyzer for the Analysis of Fish (According to specific customer demands)

The mylab NIR-Analyzer was customized to meet the specific demands of a customer. The goal was to determine fat, water, and salt content in three different fish species (Blue Whiting, Herring, and Mackerel) using NIRS. The customer provided spectral data from 52 samples obtained with their mylab NIR-Analyzer, along with corresponding reference values. The calibration was performed using neural networks combined with the GPR algorithm (Gaussian Process Regression).

Table 4 presents the range of values for the fish dataset as provided by the customer, the deviation of the parameters obtained via NIRS compared to the reference analysis, and the measurement accuracy.

Tab. 4. Range value of the data set "Fish", deviation of the values of obtained parameters by NIRS compared to reference analysis and the certainty of the measurement.

Parameter	Fat	Water	Salt
Range Value* of the Data Set	3.35- 24.39 g /100 g	57.27 - 78.60 g / 100 g	0.25 - 1.13 g / 100 g
Deviation (RMSEP)	0.65 %	0.67 %	0.07 %
Coefficients of Determination (R^2)	0.99	0.99	0.89

Training the mylab NIR-Analyzer for the Analysis of Yeast (According to specific customer demands)

A mylab NIR-Analyzer was trained specific to the demands from a customer. 50 samples were analyzed in a small range value of the data set. The task was to determine the content of dry matter, fat, Trehalose (a natural sugar composed of 2 glucose molecules), phosphate and nitrogen (N_2) water and phosphorus pentoxide (P_20_5).

Table 5 shows the range value of the data set for Fish samples as provided by the customer, the deviation of the values of obtained parameters by NIRS compared to reference analysis and the certainty of the measurement.

Tab. 5. Range value of the data set "Yeast", deviation of the values of obtained parameters by NIRS compared to reference analysis and the certainty of the measurement.

Parameter	Dry Matter	Fat	Trehalose	P_2O_5	N_2
Range Value* of the Data Set	95.25 - 96.1 /100 g	1.16 - 1.51 g / 100 g	16.7 - 20.9 g / 100 g	2.27 - 2.76 g / 100 g	6.69 - 7.98g / 100 g
Deviation (RMSEP)	0.08 g / 100 g	0.03 g / 100 g	0.24 g / 100 g	0.04 g / 100 g	0.12 g / 100 g
Coefficients of Determination (R^2)	0.77	0.86	0.91	0.87	0.89

Discussion and Conclusions

This study demonstrates that the number of samples in the calibration dataset does not necessarily determine the quality of parameter predictions using the mylab NIR-Analyzer. For instance, a condensed dataset of only 52 fish samples achieved the same coefficient of determination for fat content prediction ($R^2 = 0.99$) as a much larger dataset containing 750 samples (as demonstrated for meat and meat products).

Consequently, even a smaller number of calibration samples may allow for highly precise determination of specific product parameters. However, this accuracy depends on several factors, such as the type of parameter being analyzed (e.g., salt, sugar, or pH value). In addition, the complexity of the NIR analysis for the given substance, as some parameters do not produce distinct spectral signals but instead cause slight wavelength shifts (e.g., due to variations in salt content). Therefore, selecting an appropriate calibration dataset and analytical conditions is crucial for obtaining reliable prediction results.

We therefore recommend contacting our experts in advance. This will allow you to discuss in detail whether and how the mylab NIR-Analyzer should be optimally trained for your analysis.