Gradus Vol 12, No 1 (2025)
ISSN 2064-8014

GRADUS
GRADUS.KEFO.HU

AN OVERVIEW OF SAMPLE AND SPECTRUM PREPARATION METHODS FOR NEAR-INFRARED SPECTROSCOPY

*Görbe, Mihály*0009-0007-5861-248X*

[1] Department of Science and Technology, GAMF Faculty of Engineering and Computer Science, John von Neumann University, Hungary

https://doi.org/10.47833/2025.1.ENG.009

Keywords:
Near-infrared spectroscopy
NIR spectroscopy
Sample preparation
Spectrum preparation
Spectroscopy calibration

Article history:

Received	10 July 2024
Revised	03 Feb 2025
Accepted	21 Feb 2025

Abstract

Near-infrared (NIR) spectroscopy has increasing significance in various parts of industry, such as food and feed industry, pharmaceutical industry, polymer and semiconductor manufacturing etc. However, the output reliability and precision of NIR spectroscopy depends on the calibration quality, and both measurement accuracy and calibration quality depend on sample and spectrum preparation. This paper gives a short review of the main methods of NIR sample and spectrum preparation methods used in industry and science.

1 Introduction

Near-infrared (NIR) spectral range is part of the electromagnetic spectrum, adjacent to visible light on the longer wavelength side. It is perceptible to humans as radiating heat. This radiation interacts with characteristic groups in organic molecules.

The NIR spectrum of a sample, that is, the reflectance or the transmittance (depending on spectroscopy technique) as the function of wavelength, as can be seen in Fig. 1. contains several peaks, that can be associated with vibrational and rotational energy transitions of characteristic molecular bonds in the substance of interest.

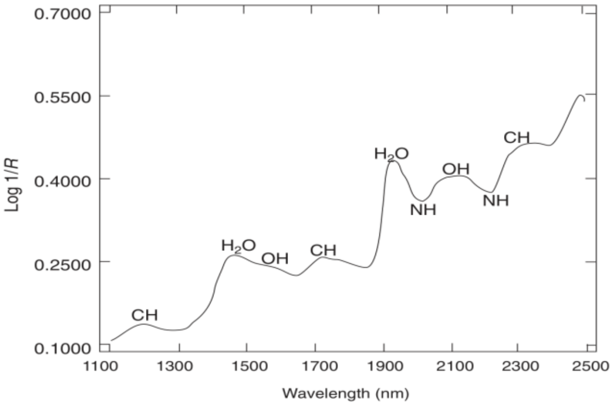

Figure 1. Absorption spectrum of hay sample in the NIR range, with the characteristic wavelengths of main molecular bonds shown [1]

This property of NIR spectra can essentially be utilized for measuring the concentrations of chemical components of the sample, however several other complex properties, that result from component concentrations can be the objective of NIR spectral measurements, like moisture content in logging

---

* Corresponding author.
E-mail address: gorbe.mihaly@nje.hu

residues [2], density of linear low-density polyethylene [3], detection of *E. coli* in lettuce samples [4], berry texture [5], soluble solids content (SSC), pH, firmness and seedlessness [6] of grape, SSC and firmness of pear [7], SSC and dry matter contents (DMC) of apple [8], SSC, pH and firmness of tomatoes [9], maturation levels of apple [10], etc.

In the case of transparent samples, e.g. solutions, it is possible to measure the light transmitted, so transmittance spectrum can be measured. Opaque samples however, cannot transmit infrared light, in this case the light diffusely reflected from the surface can be measured.

Scientific and industrial experience show, that in combination with elaborate spectrum processing and calibration techniques NIR spectroscopy doesn't need, or it needs only very little preparation of the samples [11]. In the case of products, that can be only traded as a whole, it is an invaluable advantage. Indeed, whole fruits can be analyzed [12], with suitable spectrograph, even several ones at a time [13]. Meat can be measured either at the output of the mincer [13], or by putting the ground meat inside of the apparatus [14]. Meat cuts can be analyzed for example by handheld spectrometers, even the fat characterization of live animals is possible [15].

2 Methodology

2.1 Pre-processing of samples

However, there are products and properties, that cannot be measured without pre-processing, or simply the results are better with a pre-treatment. In these cases, the typical operations are homogenization (like grinding, blending and shaking) [16], drying, freeze drying [17], pulverization [18] etc.

In the case of pharmaceutical industry, the pre-processing can and must be especially precisely specified, complex and time-consuming. In an application reported [19], a sample has to be shaken at minimum for 40 minutes, first for 30 minutes, followed by a NIR spectrum measurement, and then for another 10, when a spectrum comparison follows. The 10-minutes shaking must be repeated, until the last two spectra are identical.

In most cases the optimal sample pre-processing procedure must be determined empirically, even scientific papers report about comparison of the pre-processing steps tested, for example when compressing logging residuals to given densities in order to measure residual moisture [2], or when fish feed samples are mixed with a non-protein nitrogen compound at pre-determined ppm levels [20].

Keeping samples at a regulated temperature is always advisable, due to the temperature dependence of the spectral peaks of hydrogen bond [21][22], and because the optical properties of NIR device components depend on temperature too [3].

2.2 Pre-processing of spectra

The next preparation step in NIR spectroscopy is the preprocessing of spectra by mathematical means. This is necessary first because the recorded spectra of even the same sample differ slightly, especially in the case of reflective measurement, because of the randomly placed scattering centers. This unavoidable effect can be reduced by scattering correction methods [23], see Fig. 2. for the illustration of steps.

Typically, the baseline of the spectrum has an overall tendency, that is unnecessary to be included while evaluating [24]. This can also be removed by baseline correction.

The result now contains information that can already be qualitatively seen by a skilled human operator, however quantitatively it is still a challenge to separate the different peaks, and that's where peak enhancement by differentiation comes into the picture.

Depending on the noise in spectra, naïve numerical differentiation would add very high noise levels, and that's why widespread methods use polynomial fitting in a running data window, and the polynomials fitted are used to give the derivatives.

The steps above only provide general guidelines, the actual steps of spectrum pre-processing in a successful NIR analysis method must be however empirically determined.

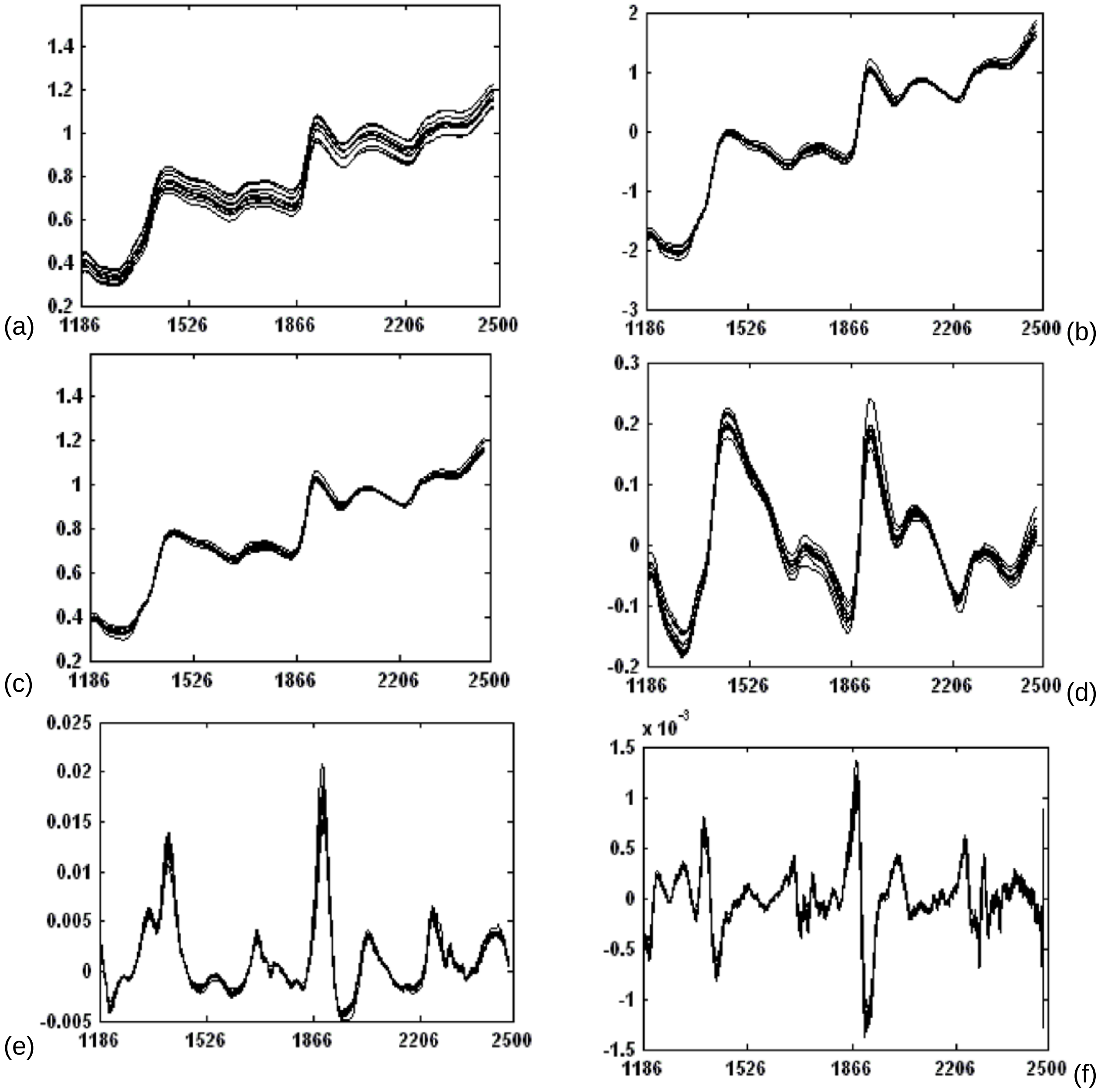

Figure 2. (a) Raw absorbance spectra (b) After standard normal variate correction (SNV) (c) After multiplicative scatter correction (MSC) (d) After linear baseline correction (e) First derivative spectra (f) Second derivative spectra (window size 5 and third order polynomial for (e) and (f)) [24]

In the case of a moisture determination problem in logging residues [2] second derivative preprocessing using a Savitzky–Golay filter [25] with 11 points and a quantic polynomial was found to be the best. Interpolated second derivative pre-processing can be successfully chosen for SSC and DMC determination in apple [8]. There are complicated project results, e.g. when at a three qualities (firmness, pH, SSC) prediction problem for Japanese table grapes first derivative preprocessing was most suitable for firmness estimation, multiplicative scatter correction combined with first derivative preprocessing for pH measurement and multiplicative scatter correction combined with second derivative preprocessing for SSC measurement [6]. There are works that report on systematic performance comparison of various pre-processing possibilities [4][6][7][9].

The raw data amount in the resulting normalized spectra then should be reduced in quantity into fewer characteristic factors. A widespread reduction method is the principal component analysis [26]. Typically 2-3 principal components (PCs) are enough to satisfactorily capture the essence of the spectra, however mostly only experience can show if two PCs are enough [3,4,9], three are needed [2][8], etc.

An actively researched field is the application of artificial neural networks (ANNs) for NIR spectrum evaluation, that can transform the data amount in spectra directly into the quantities/qualities of interest. There are works that report on the performance of various network parameters [7][9].

2.3 Calibration sample set preparation

NIR spectral analysis methods need calibrated evaluation of the pre-processed spectra. The calibration makes the instruments, the sample treatment procedures, the sample holders, the numerical methods work together and give a correct result about the substances of interest in the current project. During calibration, spectra recorded are compared against the results of reference methods (see Fig. 3.), and in the end a successful validation is needed. It is not unusual, that a valid calibration model doesn't give the former accurate values any more, if for example elements of instrumentation or method steps are changed. [24]

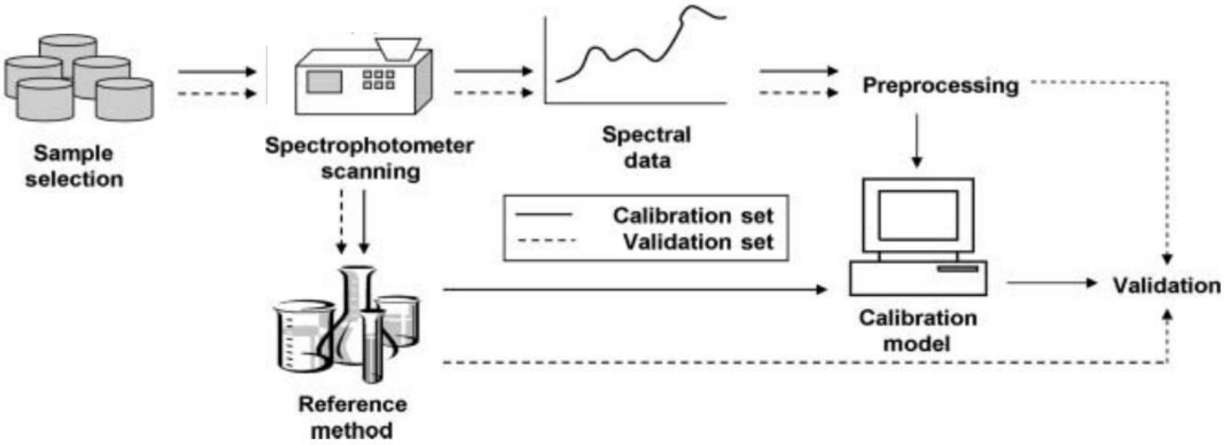

Figure 3. Scheme of NIR calibration [24]

The composition of the calibration sample set is crucial, it has to represent all the cases expected to evaluate well, e.g. with respect to the concentration range. For example, in the sample sets at pharmaceutical industry, the concentration of the analyte substances (e.g. the active compounds) must be of uniform distribution, of 3 or more different levels (5 is most typical), between the nominal concentration ±25 % and uncorrelated with each other. At the same time the concentration of the non-analyte substances (e.g. the excipients) must be uncorrelated with each other and any active compounds too [19]. Keeping the same guidelines is harder in the case of agriculture or food industry, where samples cannot be manufactured systematically, in contrast they have to be collected.

Collecting calibration samples for agriculture and food industry “in the wild” in a manner that approximately covers intervals of variables of interest, is implemented in ways like measuring fruits at different pre-determined times after harvest in order to calibrate grape berries texture [5] and cherry tomato quality [9] estimation, to calibrate Japanese table grape [6] and Korla fragrant pear [7] SSC measurement; measuring fruits of after pre-determined storage lengths and storage methods (air and controlled atmosphere) to calibrate apple SSC and DMC measurement [8]; even intentionally contaminating lettuce samples with pre-determined dilutions of *E. coli* solution in order to detect contamination [4]. Often it is crucial however to collect sample sets, that are orthogonal in the quantities/qualities of interest and the factors not treated important: like measuring grape berries from all possible locations in a bunch [6], measuring fruits from random tree individuals and random canopy layers [7].

It is especially important, when several substance concentrations are to be measured and calibrated, that there are elaborate mathematical considerations [27] to keep sample set sizes at feasible levels, that otherwise would grow exponentially with the number of quantities of interest.

2.4 Outliers detection

Outliers in calibration data sets may influence the resulting model adversely. Outliers can be a result of faulty sample preparation, instrument malfunction, incorrect sample identification, etc. Some of these errors can be corrected, and then the sample will be no longer an outlier, and it can contribute to calibration. Trying to measure a sample, that lies out of the sample set of a calibration may give unreliable results. By all means detecting outliers is of high importance, and is a notable and far reaching sub-field in mathematical statistics.

Just to give a feeling about how a possible and graphic outlier detection method works, a plot of the calibration set can be seen on Fig. 4. The method under calibration is *E. coli* detection by NIR spectroscopy [4]. The Q-residuals are plotted vs. the Hotelling's T^2 parameter.

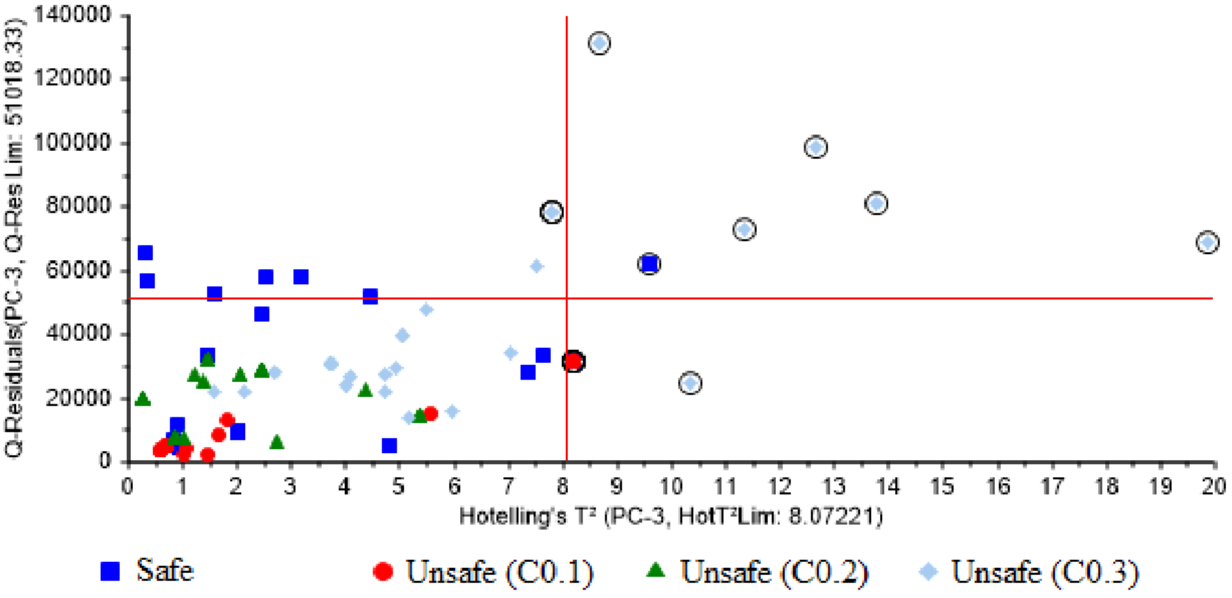

Figure 4. Calibration set plotted for outlier detection [4]

The red lines on the plot mean 5 % level of significance. Data points with circles around are recognized as outliers based on being beyond one or both of the red lines.

3 Summary

Near-infrared spectroscopy has the well-known advantages of being non-invasive and having rapid analytical capabilities. However, the accuracy and reliability of NIR spectroscopy heavily depend on meticulous calibration and precise sample and spectrum preparation. This paper reviewed the main methods of NIR sample and spectrum preparation. Whatever progression the future holds in this field, as now, the optimal procedure steps in the solutions of specific measurement tasks will likely have to be researched and fine-tuned later on by experienced human scientists and engineers.

Acknowledgment

This work is supported by GINOP_PLUSZ-2.1.1-21-2022-00144 *Development of near-infrared spectroscopic instruments, methods and models suitable for on-site measurement of ensiled feed* project.

References

[1] D.J. Givens et al. (eds.), Forage Evaluation in Ruminant Nutrition, CABI Publishing, 2000, ISBN 0-85199-344-3.
[2] S. Hwang et al., "Investigation of NIR spectroscopy and electrical resistance-based approaches for moisture determination of logging residues and sweet sorghum," BioResources, vol. 18, no. 1, pp. 2064-2082, 2023. DOI: https://doi.org/10.15376/biores.18.1.2064-2082

[3] M. Watari et al., "On-line monitoring of the density of linear low-density polyethylene in a real plant by near-infrared spectroscopy and chemometrics," Appl. Spectrosc., vol. 58, no. 2, pp. 248-255, Feb. 2004. DOI: https://doi.org/10.1366/000370204322843011

[4] S. Rahi et al., "Visible/near-infrared spectroscopy as a novel technology for nondestructive detection of Escherichia coli ATCC 8739 in lettuce samples," in Proceedings of The World Conference on Engineering and Technology, 2019. DOI: https://doi.org/10.33422/worldcet.2019.10.285

[5] T. Basile et al., "Use of artificial neural networks and NIR spectroscopy for non-destructive grape texture prediction," Foods, vol. 11, p. 281, 2022. DOI: https://doi.org/10.3390/foods11030281

[6] C. Kanchanomai et al., "Non-destructive analysis of Japanese table grape qualities using near-infrared spectroscopy," Horticulture, Environment, and Biotechnology, vol. 61, 2020. DOI: https://doi.org/10.1007/s13580-020-00256-4

[7] X. Yu et al., "Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging," Postharvest Biology and Technology, vol. 141, pp. 39-49, 2018. DOI: https://doi.org/10.1016/j.postharvbio.2018.02.013

[8] Y. Zhang et al., "Non-destructive prediction of soluble solids and dry matter contents in eight apple cultivars using near-infrared spectroscopy," Postharvest Biology and Technology, vol. 151, pp. 111-118, 2019. DOI: https://doi.org/10.1016/j.postharvbio.2019.01.009

[9] L. Feng et al., "Nondestructive detection of postharvest quality of cherry tomatoes using a portable NIR spectrometer and chemometric algorithms," Food Anal. Methods, vol. 12, pp. 914-925, 2019. DOI: https://doi.org/10.1007/s12161-018-01429-9

[10] R. Pourdarbani et al., "Automatic non-destructive video estimation of maturation levels in Fuji apple (Malus Malus pumila) fruit in orchard based on colour (Vis) and spectral (NIR) data," Biosystems Engineering, vol. 195, pp. 136-151, 2020. DOI: https://doi.org/10.1016/j.biosystemseng.2020.04.015

[11] R. Pandiselvam et al., "Recent advancements in NIR spectroscopy for assessing the quality and safety of horticultural products: A comprehensive review," Front. Nutr., DOI: https://doi.org/10.3389/fnut.2022.973457

[12] M.F.R. Pahlawan et al., "Development of a low-cost modular VIS/NIR spectroscopy for predicting soluble solid content of banana," IOP Conf. Ser.: Earth Environ. Sci., vol. 644, p. 012047, 2021. DOI: https://doi.org/10.1088/1755-1315/644/1/012047

[13] H. Huang et al., "Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review," Journal of Food Engineering, vol. 87, pp. 303-313, 2008. DOI: https://doi.org/10.1016/j.jfoodeng.2007.12.022

[14] V. Babushkin, "Application of NIR and FTIR Foss solutions for food analysis," Foss, Samara, 2016.

[15] H.I. Kademi et al., "Applications of miniaturized and portable near infrared spectroscopy (NIRS) for inspection and control of meat and meat products," Food Reviews International, pp. 1-20, 2018. DOI: https://doi.org/10.1080/87559129.2018.1514624

[16] E. Taira et al., "Networking system employing near infrared spectroscopy for sugarcane payment in Japan," J. Near Infrared Spectrosc., vol. 21, pp. 477-483, 2013. DOI: https://doi.org/10.1255/jnirs.1081

[17] G. Bázár, "Application possibilities of near infrared spectroscopy in the qualification of pork, meat products and lard," PhD thesis, Kaposvár University, 2011.

[18] H. Yan et al., "Handheld near-infrared spectroscopy: State-of-the-art instrumentation and applications in material identification, food authentication, and environmental investigations," Chemosensors, vol. 11, p. 272, 2023. DOI: https://doi.org/10.3390/chemosensors11050272

[19] M. Blanco et al., "Influence of the procedure used to prepare the calibration sample set on the performance of near infrared spectroscopy in quantitative pharmaceutical analyses," Analyst, vol. 126, pp. 1129-1134, 2001. DOI: https://doi.org/10.1039/b102090k

[20] S.K. Bhatti, "Using near-infrared reflectance spectroscopy (NIRS) for qualitative determination of undesirable chemical component of high nitrogen content in protein raw material used for fish feed," Master's thesis, University of Bergen, 2018.

[21] F. Despagne et al., "Neural networks in multivariate calibration," Analyst, vol. 123, pp. 157R-178R, 1998. DOI: https://doi.org/10.1039/A805562I

[22] H. Maeda et al., "Near infrared spectroscopy and chemometrics studies of temperature-dependent spectral variations of water: Relationship between spectral changes and hydrogen bonds," J. Near Infrared Spectrosc., vol. 3, pp. 191-201, 1995. DOI: https://doi.org/10.1255/jnirs.69

[23] N.K. Afseth et al., "Extended multiplicative signal correction in vibrational spectroscopy, a tutorial," Chemom. Intell. Lab. Syst., vol. 117, pp. 92-99, 2012. DOI: https://doi.org/10.1016/j.chemolab.2012.03.004

[24] L.E. Agelet et al., "A tutorial on near infrared spectroscopy and its calibration," Crit. Rev. Anal. Chem., vol. 40, no. 4, pp. 246-260, 2010. DOI: https://doi.org/10.1080/10408347.2010.515468

[25] A. Savitzky et al., "Smoothing and differentiation of data by simplified least squares procedures," Anal. Chem., vol. 36, no. 8, pp. 1627-1639, 1964. DOI: https://doi.org/10.1021/ac60214a047

[26] T. Naes et al., Selection of Samples for Calibration, in A User-Friendly Guide to Multivariate Calibration and Classification, NIR Publications, Chichester, UK, 2002, Ch. 15, pp. 191-195. DOI: https://doi.org/10.1255/978-1-906715-25-0

[27] R.G. Brereton, "Introduction to multivariate calibration in analytical chemistry," Analyst, vol. 125, no. 11, pp. 2125-2154, 2000. DOI: https://doi.org/10.1039/b003805i