



ELSEVIER

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Multicomponent blood lipid analysis by means of near infrared spectroscopy, in geese

George Bazar^{a,1,*}, Viktoria Eles^a, Zoltan Kovacs^b, Robert Romvari^{a,2}, Andras Szabo^{a,c,3}^a Institute of Food and Agricultural Product Qualification, Faculty of Agricultural and Environmental Sciences, Kaposvar University, 40 Guba Sandor str., 7400 Kaposvar, Hungary^b Department of Physics and Control, Faculty of Food Science, Szent Istvan University, 14-16 Somloi str., 1118 Budapest, Hungary^c Hungarian Academy of Sciences – Kaposvar University "MTA-KE Mycotoxins in the Food Chain" Research Group, 40 Guba Sandor str., 7400 Kaposvar, Hungary

ARTICLE INFO

Article history:

Received 9 December 2015

Received in revised form

13 April 2016

Accepted 19 April 2016

Available online 20 April 2016

Keywords:

Blood serum

NIR spectroscopy

Cholesterol

Lipid

Triglyceride

Laboratory error

ABSTRACT

This study provides accurate near infrared (NIR) spectroscopic models on some laboratory determined clinicochemical parameters (i.e. total lipid (5.57 ± 1.95 g/l), triglyceride (2.59 ± 1.36 mmol/l), total cholesterol (3.81 ± 0.68 mmol/l), high density lipoprotein (HDL) cholesterol (2.45 ± 0.58 mmol/l)) of blood serum samples of fattened geese. To increase the performance of multivariate chemometrics, samples significantly deviating from the regression models implying laboratory error were excluded from the final calibration datasets. Reference data of excluded samples having outlier spectra in principal component analysis were not marked as false. Samples deviating from the regression models but having non outlier spectra in PCA were identified as having false reference constituent values. Based on the NIR selection methods, 5% of the reference measurement data were rated as doubtful. The achieved models reached R^2 of 0.864, 0.966, 0.850, 0.793, and RMSE of 0.639 g/l, 0.232 mmol/l, 0.210 mmol/l, 0.241 mmol/l for total lipid, triglyceride, total cholesterol and HDL cholesterol, respectively, during independent validation. Classical analytical techniques focus on single constituents and often require chemicals, time-consuming measurements, and experienced technicians. NIR technique provides a quick, cost effective, non-hazardous alternative method for analysis of several constituents based on one single spectrum of each sample, and it also offers the possibility for looking at the laboratory reference data critically. Evaluation of reference data to identify and exclude falsely analyzed samples can provide warning feedback to the reference laboratory, especially in the case of analyses where laboratory methods are not perfectly suited to the subjected material and there is an increased chance of laboratory error.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Foie gras as a luxury food product is made of specially fattened duck or goose liver. Total foie gras production of the European Union was around 25,000 t in 2014 covering world's 90% [1]. The fatty goose liver production of Hungary is nearly 2000 t per annum, being a world leader, and the country has old traditions of fattening [2]. During gavage the birds are strongly overfed with

high-fat diets that provides, besides 8–10-fold liver mass increase, severe dyslipidemia [3]. Because of the force-feeding procedure, gavage-based foie gras production is controversial and production or retailed products are denied in several countries, and there is an expressed need for avoiding gavage in foie gras production. Evaluating the effects of gavage or gavage-free feeding alternatives, and production performances of different breeds requires biochemical analytical background [4]. Thus, quick and accurate analysis of body fluids, primarily blood plasma or serum is essential to evaluate health or nutritional status of animals. The most characteristic alteration of serum blood biochemical constituents during hepatic steatosis concerns the blood lipids: total lipid, triglyceride and the lipoproteins, i.e. total and high density lipoprotein (HDL) cholesterol.

In the routine clinicochemical laboratory analysis the current practice is applying automated equipment utilizing enzymatic and direct photometric methods for the metabolite and enzyme measurements, while most ion concentrations are measured with

* Corresponding author.

E-mail address: bazar@agrilab.hu (G. Bazar).¹ Department of Nutritional Science and Production Technology, Faculty of Agricultural and Environmental Sciences, Kaposvar University, 40 Guba Sandor str., 7400 Kaposvar, Hungary.² Institute of Animal Science, Faculty of Agricultural and Environmental Sciences, Kaposvar University, 40 Guba Sandor str., 7400 Kaposvar, Hungary.³ Institute of Diagnostic Imaging and Radiation Oncology, Faculty of Agricultural and Environmental Sciences, Kaposvar University, 40 Guba Sandor str., 7400 Kaposvar, Hungary.

ion-selective electrodes [5]. Because of hyperlipidemia, it is often problematic to get accurate results for blood parameters of fat-tended geese using automats widely applied in human diagnostics. Moreover, some of the analyses are loaded with uncertainty derived from the improper standard substrate used for calibration [6].

The widespread application of near infrared (NIR) spectroscopy for the quick estimation of biochemical components in divergent biological matrices is reasoned by multiple factors, such as the short analysis time, reagent-less work, lack of waste production, and the parallel estimation possibility of multiple components. These advantages provoked the expansion of the NIR technique in clinical chemistry. Possibly one of the earliest approaches was performed by Hall & Pollard [7] describing the estimation of serum proteins, triglycerides and glucose, with the very important warning that purified chemical standard solutions cannot be applied as a basis for quantitative analysis. Instead, to develop calibrations, a real primary sample set with known clinicochemical values is needed. This methodology was utilized by Heise et al. [8] reporting the successful assay of total protein, glucose, total cholesterol, triglyceride, and urea in EDTA-treated human plasma, recommending the use of wide wavelength regions. Domján et al. [9] and Turza et al. [10] made important steps towards simplification of the sample application, namely authors analyzed also whole (anticoagulant treated) blood, Turza et al. [10] even directly in sampling tubes instead of cuvettes. From a physiological aspect it is worth mentioning that the NIR spectroscopy based estimations are not limited to chemically clearly defined molecules (e.g. glucose [10] or urea [7]), but it is also accurate for heterogeneous compounds, like total protein (albumin and globulins [8]), and for highly complex particles, like lipoproteins [9,11]. Petter et al. [11] described NIR method useful as a potential alternative or even supplementary clinical method for the quick determination of low density lipoprotein (LDL) and HDL in human serum.

There are several factors affecting the performance of NIR calibration models, such as sample preparation and presentation to scanning, geometry and operation of the spectrometer, environmental effects, disabilities of mathematical modelling, or the inaccuracy of the reference data. There are tests in use to evaluate most of these effects, but in many cases reference data are being considered as accurate, however this is not always checked.

Due to the multi collinearity of NIR spectral data, and the complexity caused by the overlapping combination and overtone bands, multivariate statistical methods are required for developing quantitative and qualitative models. Evaluation of the multivariate models according to chemometrics reveals the background of calibration and classification procedures, describes the relationship between the spectral data, the chemical composition, and the physical parameters [12]. Thus, through detailed chemometric analysis of NIR data, latent variables, and regression coefficient vectors it is possible to describe the chemical reason why the developed models work. Since model performance may decrease in any case when unreliable reference or spectral data are observed, evaluation of models can provide information on the input data.

Chemical constituents provide complex absorption signals in various parts of the NIR spectrum. Assuming a well-operated benchtop spectrometer, NIR spectra describe the samples very accurately, and the performance of the developed calibration model may be limited mostly by the slight inaccuracies of the reference data [13,14]. In this sense, NIR technique is useful in testing and evaluating the reference data, as errors of NIR models may refer to the errors of the laboratory reference data [15].

The objective of the present study was to develop NIR calibration models for further feeding and housing experiments which can predict geese's blood clinicochemical parameters, using

different spectral regions, and spectra of different measurement modes – i.e. transmission and transreflectance. Furthermore, our aim was to develop an automatable objective method to evaluate the outlier samples in order to increase model accuracy and give feedback to reference analytical laboratory in highly doubtful cases. The biological sample in focus was a highly characteristic, postprandial, high-fat avian blood serum, providing generally large variation (broad range) in the population investigated.

2. Materials and methods

2.1. Animals and blood sampling

Grey Landes type ganders were examined at 11th, 17th and 20th days within the *a priori* planned and legislated force-feeding protocol, performed at the experimental farm in a closed stable of the Kaposvár University by a professional animal caretaker. The force-feeding diet contained 4.91 g ether extract /100 g dry matter (DM), 12.78 g crude protein /100 g DM, 2.33 g crude ash /100 g DM and 12.4 MJ / kg metabolizable energy. The force-feeding was performed three times per day, until satiation, being a routine protocol. The investigation period was chosen due to the high blood concentration of the analytes in study. The same thirty geese were tested on day 11 and 17, while additional 15 geese were involved on day 20 (n=30+30+45). Blood samples were drawn from the *v. brachialis* after cutaneous local analgesia into 13 ml centrifuge tubes with 22 G sterile needles. Blood was left to clot at 35 °C, centrifuged on 1500 rpm for 10 min, before serum separation. Serum was siphoned and stored in 5 ml Eppendorf tubes at –70 °C until analysis. Animals after the force-feeding were sold for a professional goose abattoir and were processed industrially.

The study was carried out in strict accordance with the animal welfare protocol for experiments at Kaposvár University, and all efforts were made to minimize suffering. The protocol was approved by the Somogy County Animal Health and Food Chain Safety Directorate under the allowance number: XV-I-31/446-10/2012, in accordance with the concerning EU legislations.

2.2. Clinicochemical analysis

Clinicochemical analysis was performed on an automated equipment (Hitachi 917) in a single analytical run. Triglyceride (TG) concentration was measured on the basis of the Trinder reaction [16], total cholesterol (TCh) by the enzymatic method of Röschlau et al. [17], and high-density lipoprotein cholesterol (HDL-Ch) by the phosphotungstic acid/MgCl₂ precipitation method. Reagent kits for TG, TCh, and HDL-Ch were purchased from Human Diagnostics Ltd. (Wiesbaden, Germany). The total lipid (TL) content of serum was determined spectrophotometrically (Shimadzu UV 160 A equipment), with the sulfo-phospho-vanillin method and measured in 10 mm path length optical glass cuvettes [6]. There was no replicate measurement of reference clinicochemical data at the single blood serum samples.

2.3. Near infrared spectrum acquisition

Each geese blood serum sample was scanned using a FOSS NIRSystems 6500 spectrometer (FOSS NIRSystems, Inc., Laurel, MD, USA). Absorbance spectra (log₁/T for transmission mode, and log₁/R for transreflectance mode, further on referred to as transmission spectra and transreflectance spectra, respectively) were recorded with WinISI II, version 1.5 software (InfraSoft International LLC, State College, PA, USA). Spectra were recorded with a spectral increment of 2 nm in the 400–2498 nm interval. Acquisition of transmission spectra was performed with a 1 mm open-top

cuvette. Transflection spectra were measured in reflectance mode, using a camlock cell with gold-plated reflector back, in which the layer thickness was 0.1 mm. Required quantity of sample was 1 cm³ and 0.15 cm³ in transmission and transflection measurement modes, respectively. Temperature of cuvette holder and inserted samples was not controlled during scanning, but unfrozen samples in closed Eppendorf tubes were kept at constant ambient room temperature (25 °C) for at least one hour prior to scanning. Samples (n=105) were scanned in random order, once in each measurement modes.

2.4. Data evaluation

Blood clinicochemical reference data were checked for outliers using descriptive boxplot analysis, and the group-dependent statistical analysis of the parameters were performed with one-way analysis of variance (ANOVA) – when ANOVA results indicated, pairwise significant differences were calculated using *post hoc* Tukey test (IBM SPSS Statistics 20 for Windows).

The Unscrambler 9.7 (CAMO Software AS, Oslo, Norway) chemometrics software was applied for evaluation of the recorded NIR spectra. Various wavelength intervals were applied during the evaluations, representing 1st overtone absorption regions of OH bonds of water (1300–1600 nm), 1st overtone absorption regions of CH bonds of lipids (1650–1800 nm), or wider range containing absorption regions of both OH and CH bonds (1150–1800 nm). Long wavelength NIR region (1800–2500 nm) was not used for data evaluation because the absorbance spectra were truncated as small amount of low energy light passed the samples and stray light dominated the signals of the distorted interval. Visible and short wavelength NIR region (400–1150 nm) was neglected because of the high level of non-linear spectral variation caused by the light scattering of the blood serum samples.

Because of the physiological conditions of the animals (i.e. high incidence of hyperlipidemia), blood serum samples showed considerable optical differences. The turbid samples resulted in different optical density with increased light scattering at a high variance. In order to remove the observed linear distortions of the NIR spectra, and to reveal the underlying signals of the overlapping peaks, 2nd order difference spectra were calculated using the Norris gap derivative function with 5-point gap. Moving average smoothing with 9 points was applied on each spectrum before calculating the derivatives. The smoothing window and the gap size for differentiation was comparable with the NIR bandwidths of the investigated biological absorbers measured with the applied spectrometer having nominal bandpass of 10 nm [18]. Standard Normal Variate correction was used on the applied intervals of the 2nd derivative spectra to reduce the remaining additive (multiplicative) effects of scatter [15,19]. Raw and pretreated spectra were also applied for multivariate data analyses. Single application of pretreatments was tested against the combination of those, as described above. Best results were achieved with the combined application of the spectral pretreatments, as described above, and only these results will be shown further on.

Principal component analysis (PCA) was applied to describe multidimensional patterns of the NIR dataset and to discover outliers [20]. Principal component regression (PCR) and partial least squares regression (PLSR) were used for the calibration models on individual blood clinicochemical parameters [15]. The main difference of the two quantitative methods lays in the calculation of the latent variables (LVs), i.e. LVs of PCR are optimized for describing the variance of X (spectral) variables only, while LVs of PLSR are optimized for describing the variance of both X and Y (reference) variables, simultaneously. Each model was tested with leave-one-out cross-validation, when one sample was excluded from the calibration set, and the generated model was tested on

this retained sample. The procedure was repeated iteratively to ensure that all the samples were applied once for validation. Precision and accuracy of the quantitative models were evaluated by the coefficient of determination (R²), root mean squared error (RMSE) and systematic error was indicated by the bias value. The respective measures were calculated for calibrations, and cross-validations. The number of LVs applied for each model was increased until the cross-validation error started increasing. As a rule of thumb, maximum number of LVs was determined as 10% of total number of the samples in the model.

2.5. Methods for Y-residual outlier selection

Selection methods were applied to identify and eliminate extreme outliers that were materially decreasing the model performance and samples being far from the models (having large Y-residual), presumably because of a false reference data.

In the first method (method-1, “residual vs. reference-SD”), PCR was calculated on the complete database in the given spectral regions. PCR calibrations were favored instead of PLSR, because latent variables of PCR were not influenced by the false reference data, thus, models were better indicators of reference errors. Residuals were defined as the difference of the laboratory value and the predicted value of cross-validation. Those samples were recognized as outliers where the absolute residual was bigger than the standard deviation of the laboratory data. Finally, PLSR was calculated without outliers.

In the second method (method-2, “residual vs. residual-SD”) the residuals were calculated the same way as in method-1. The standard deviation of the residuals (residual-SD) was calculated, and a sample was identified as an outlier if its absolute residual was higher than one and a half residual-SD [15]. (1.5 × SD was an empirical value in order to keep the 5–10% selection rate.) Finally, PLSR was calculated without outliers.

Method-1 can be useful if the reference method is highly inaccurate, but the overall SD of values is relatively low. In this case, it is evident, that the model residuals will be large, because chemometrics will be unable to model the uncertain values, and these large residuals would make the “residual to residual-SD” method unreasonably permissive. However, the SD of the inaccurate reference values should not be necessarily high, thus, it can be good to cut limit based on the SD of the reference value.

Method-2 is an applicable alternative if the reference method provides accurate data which – with good chemometrics – can result low residuals in general. This option is beneficial when the reference value naturally moves in a wide range, and SD of the reference value in the sample set is high, or can increase anytime because of some new samples having accurate but extreme reference data. In this case, the “residual to reference-SD” method would be unreasonably permissive.

In the present study, results of both methods were summarized, and feedback to reference laboratory was based on the combined evaluations.

3. Results and discussion

3.1. Results of chemical analysis

Two blood serum samples (one from day 17 and 20) were prepared with late cell separation and contained hemoglobin because of hemolysis, which resulted in obvious outlier clinicochemical and spectral data. Further two extreme outlier HDL-Ch values were omitted from the dataset, based on being beyond the whiskers (one and a half the interquartile range) in the boxplot analysis. These concerning data were not included in further

Table 1
Results of clinicochemical analyses on blood serum constituents showing differences among the different stages of force-feeding.

| Sampling | TL (g/l) | | | TG (mmol/l) | | | TCh (mmol/l) | | | HDL-Ch (mmol/l) | | |
|-------------|----------|--------------------|------|-------------|-------------------|------|--------------|-------------------|------|-----------------|-------------------|------|
| | n | Mean | SD | n | Mean | SD | n | Mean | SD | n | Mean | SD |
| Day 11 | 30 | 5.57 ^{ab} | 0.86 | 30 | 2.97 ^a | 0.87 | 30 | 3.44 ^a | 0.41 | 29 | 2.00 ^a | 0.34 |
| Day 17 | 29 | 4.52 ^b | 1.07 | 29 | 1.56 ^b | 0.84 | 29 | 3.76 ^b | 0.31 | 29 | 2.37 ^b | 0.27 |
| Day 20 | 44 | 6.29 ^a | 2.56 | 44 | 3.04 ^a | 1.55 | 44 | 4.08 ^b | 0.87 | 43 | 2.82 ^c | 0.62 |
| All samples | 103 | 5.57 | 1.95 | 103 | 2.59 | 1.36 | 103 | 3.81 | 0.68 | 101 | 2.45 | 0.58 |

a,b,c: mean difference of constituent values is significant at the 0.05 level among sampling days.
TL: total lipid. TG: triglyceride. TCh: total cholesterol. HDL-Ch: HDL cholesterol.

descriptive statistics, ANOVA and *post hoc* tests, and NIR analyses.

The results of ANOVA and *post hoc* Tukey tests are given in Table 1. TG values showed significant ($p < 0.05$) decrease until the middle of the force-feeding period, then the measured parameters increased until the end of the experiment. TL showed a similar trend. TCh and HDL-Ch levels showed traceable increasing trend along the experiment. The standard deviation of all investigated clinicochemical parameters increased till the end of the experiment, showing the highly different physiological response of the animals on the force-feeding process. These results are in harmony with previous findings [3].

There were strong correlations observed between TL and TG values ($r=0.76$), or TCh and HDL-Ch values ($r=0.85$), while weak or moderate correlation was observed between TL–TCh, TL–HDL-Ch, TG–TCh, or TG–HDL-Ch ($r=0.25$; 0.26 ; 0.38 ; 0.41 , respectively).

3.2. Basic spectral evaluation

Raw and second derivative spectra of the transmission and transmittance measurements are plotted in Fig. 1. The baseline differences and multiplicative effect of light scattering along the whole spectrum caused by high lipid content of samples are prominent in both measurement modes. The applied spectral pre-treatments decreased these effects and also highlighted the overlapping absorption bands in the 1350–1550 and 1650–1800 nm regions.

All models and tests discussed below were calculated on both transmission and transmittance datasets, but outlier selection will be presented only on transmission data in detail, because results of the two measurement modes were highly similar. Comparison of the final results of the two measurement modes achieved with the outlier-selected datasets will be shown at the end of the evaluations.

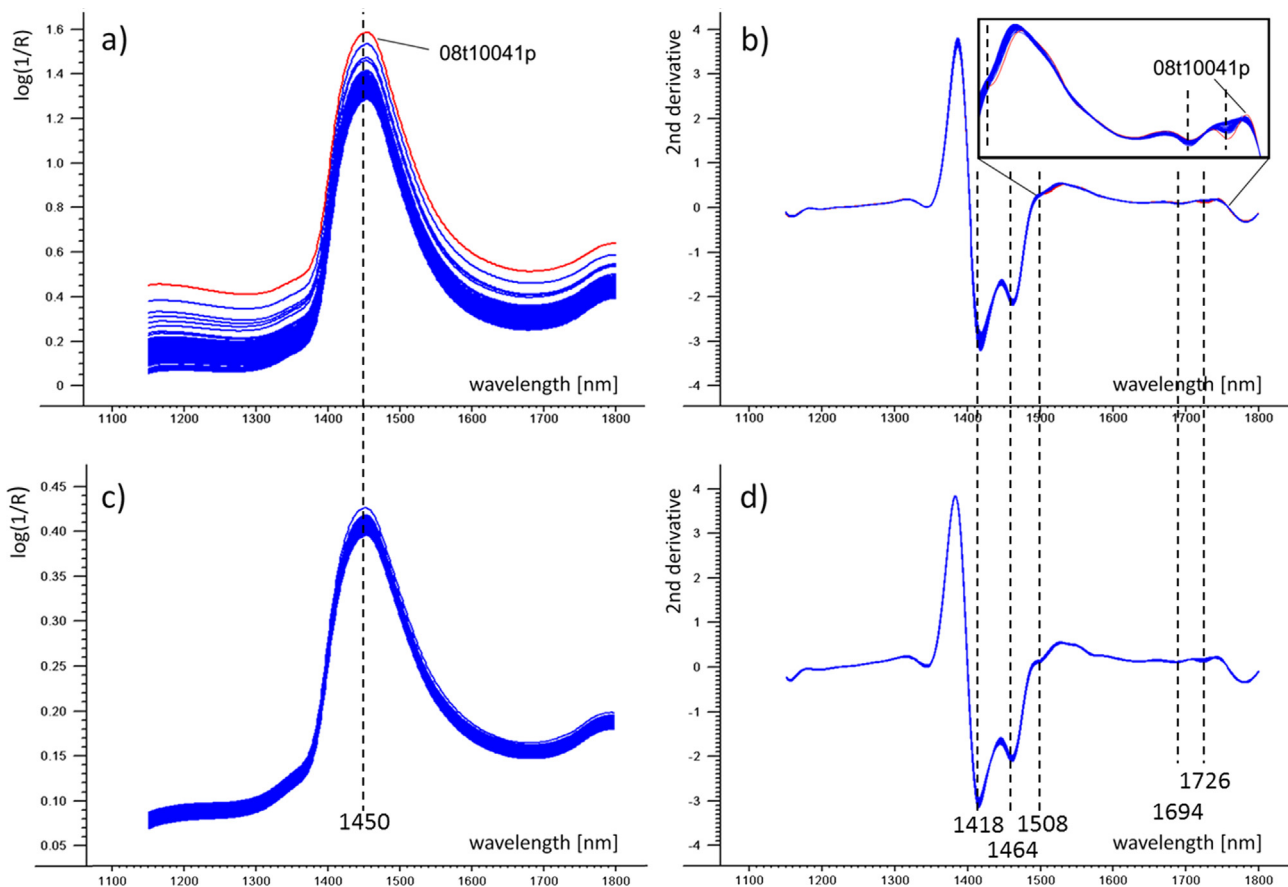


Fig. 1. Apparent absorbance spectra of the investigated goose serum samples. Raw (a, c) and pretreated (b, d) (smoothing (9 points), 2nd derivative (gap = 5 points), and SNV correction) absorbance spectra of the investigated blood serum samples measured in transmission (a, b) or transmittance (c, d) modes ($n=103$), with the indication of an outlier spectrum.

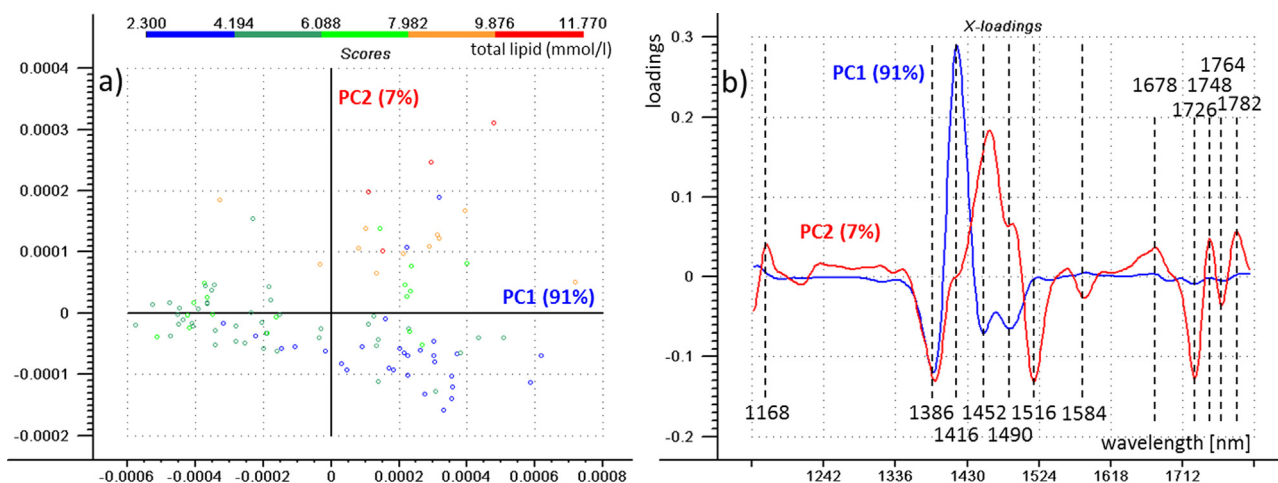


Fig. 2. Result of PCA performed on the whole set of pretreated absorbance spectra acquired in transmission mode ($n=103$). Score plot (a) showing the pattern of the dataset in the function of PC1 and PC2 – scores are colored according to total lipid content, as color coding is shown on upper color bar. The respective loading vectors (b) highlight the prominent spectral regions responsible for the principal components. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 2 shows the results of PCA performed on the pretreated blood serum transmission spectra. In the score plot (Fig. 2(a)) scores are colored according to the total lipid content, and good separation of samples with various lipid content can be seen along PC2. Accordingly, loading vector of PC1 (Fig. 2(b)) shows the dominance of water (OH 1st overtone) in the 1300–1600 nm spectral region, at 1386, 1416, 1452, and 1490 nm [21], while in the case of PC2 also the absorbance bands of lipids (CH 1st overtone) at 1726, 1764 nm [22] appear with high importance. PC1 representing water bands may refer to slight changes in temperature of samples during scanning and contains less information on variables in question since score plot does not show any separation by total lipid content along PC1. According to PCA, this fluctuation in water spectral regions caused the greatest variance of the spectra, while proportional changes of the different blood lipid fractions were secondary in spectral information.

3.3. Outlier samples

A sample can be outlier on the basis of either NIR or reference data, or both. In general, it is much harder to realize the mistakes of measurement or sample preparation in the reference methods than those in NIR spectroscopy. If a sample is only a spectral outlier but does not have an outlier clinicochemical value, or vice versa, it can be assumed that the outlier data is caused by a failure of measurement. Regarding biological samples, it is common to have extreme measurement results that are different from the bulk of the samples, both in chemical and NIR data. This does not necessarily mean that those values have to be false. Indeed, keeping those outlier samples describing sampling variability may be useful as a source of increasing model robustness if future samplings will result similar individuals.

In the presented case, outlier selection must also be done carefully, since the biggest variance of spectral information is caused by temperature effect that is basically not subject of interest. It is not a good solution to eliminate samples that seem to be outliers in PCA as this assumption may lay on the temperature difference. On the contrary, if calibration models were able to ignore this effect and weight those information that are related to the investigated constituents, then keeping most of the samples with various temperature effects will increase the robustness of the dataset, preparing it for future samples having similar defects driven by acquisition variability. Acknowledging the harmless odd spectra provides the possibility for not identifying similarly odd

and harmless future spectra as PCA outliers simply because of a largely varying but indifferent property. If those samples were deleted, one must be aware that there may be future samples similar to the deleted outliers, and the new samples will be outliers again – basically without any reason. There are targeted correction methods like EMSC [23] that can help to eliminate such known and unwanted effects, but the aim of the present study was out of this scope.

In the case of reference data, a mistake in measurement can be detected only if replicate measurements of single samples were recorded, or if the difference of correctly and erroneously recorded data is so large and determinant that the resulted value appears as an outlier in descriptive statistics. Still, an outlier among the reference data might mean a sampling or measurement failure or an unusual sample. It is very hard to decide without replicate measurement which of the above mentioned cases is experienced. Mostly a slight mistake in reference measurement or a mislabeling provides a fully acceptable, but still false value. If there was no mistake and the sample was just unusual, and presumably such data can appear in the future again, it must be retained in the model to increase robustness, otherwise it has to be omitted.

The same stands for the NIR spectra. All outliers cannot be eliminated based on the NIR spectra, because most of the times these types of samples can appear in the future practice again. Thus, elimination would result weakening of our method in regards of future challenges. However, it is highly unlikely to make an unregistered failure during scanning a sample and recording a false spectrum – be it because of erroneous sample preparation and presentation to the instrument, or because of an instrument flaw. Most of the times, mistakes do reflect well in the entire spectrum, or at some parts of it, and these scans will appear as clear outliers in the multidimensional space of PCA. Of course, these alterations could also be covered e.g. by continuous mistakes of sample preparation, when a large amount of the sample scans are false and no obvious outlier will indicate any mistakes.

The influence plot of the PCA model calculated on the transmission spectra is shown in Fig. 3(a). Regarding the two axes of the plot, residual X-variance measures how well samples fit the model determined by the components, and the leverage indicates how strong influence each observation has on the PCA model [24]. The plot shows some spectral outliers that are far from the rest of the samples, especially the one having high values on both axes. Generally, these are either samples with extreme chemical composition and NIR data, or results of false spectrum acquisition. In

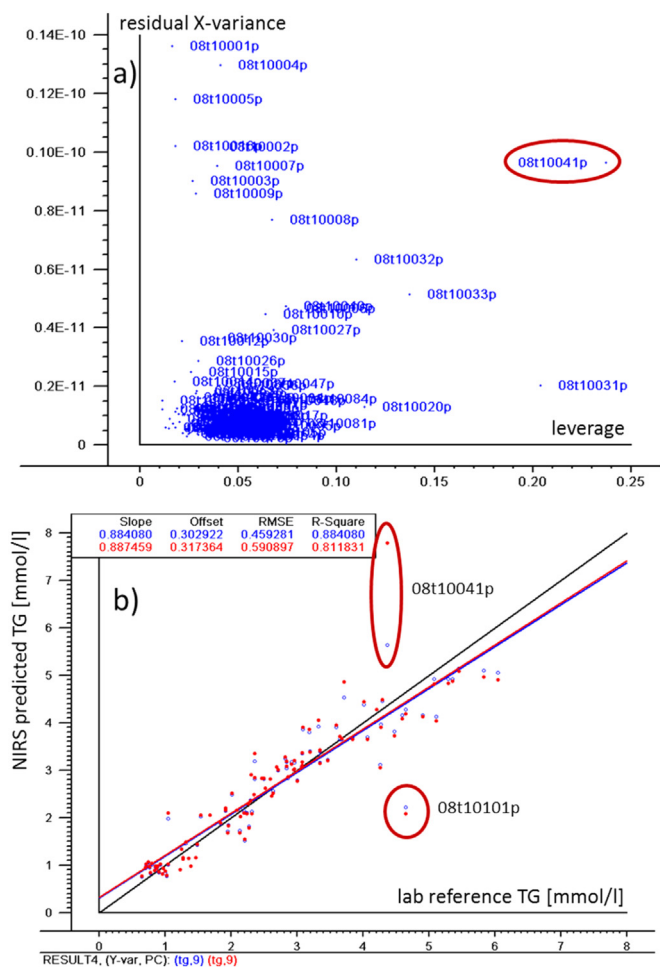


Fig. 3. Picking up outlier samples based on PCA and PCR. The influence plot of PCA (a) performed on the 1150–1800 nm interval of the absorbance spectra of the serum samples acquired in transmission mode, showing peculiar samples far from the rest of the set ($n=103$). Results of PCR (b) fitted on the same NIR spectra and triglyceride (TG) content with indication of samples where relation of NIR and chemical data has been broken.

this case, leverage outliers are samples with high lipid content and spectra are influenced by the dispersed lipids. Calculating the PCR calibration using these samples can result model deficiencies, i.e. low determination coefficient (R^2) and high error (RMSE), and instability, which is shown by the great difference in calibration (blue) and cross-validation (red) results (Fig. 3(b)).

Two samples (circled in Fig. 3(b)) appear extreme far from the target line in the PCR and elimination of these samples will significantly increase the model performance. There are other samples representing the case of having uncommon spectrum (could be identified as PCA outliers in Fig. 3(a)), but they have correct chemical reference values and there is good relationship found between spectral and chemical data – the samples are well fitted on the calibration and cross-validation line, and those are simply extreme samples on the periphery of the investigated range (presumably cases of hyperlipidemia), or have some variations in temperature that causes considerable amount of spectral variation and impacts the 1st PC (Fig. 2(b)) resulting high leverage values, but this spectral variation is irrelevant for the concerning calibrations.

Regarding the two outliers in Fig. 3(b), sample 08t10041p is also an outlier in the PCA influence graph, having high leverage and high residual X-variance values what indicates its odd character (Fig. 3(a)). The spectrum of this sample is indicated in Fig. 1. When this sample being spectrally different from the bulk of the

samples was kept in the calibration dataset, it was possible to build a model that was able to consider its extremity (blue dot is relatively close to the regression line). Though, in cross-validation, when this sample was left out of the model and calibration was tested using its outlier spectrum, the predicted value was very far from the measured value (red dot). This sample is unlike the others (others did not capture the variation this sample caused), so, in comparison with the bulk of the dataset, this exact sample is a spectral outlier. A more interesting case is represented by sample 08t10101p which was not a spectral outlier in Fig. 3(a), was not an outlier based on the raw clinicochemical data, but appeared as an outlier in the PCR model (both in calibration, blue dot, and validation, red dot). Since the well-defined relation of the spectral data and chemical composition was broken in this latter case, it could be assumed that this sample needed attention as it may have had false reference data combined with a correct NIR spectrum.

In the first case, it was possible to slightly rotate the space of latent variables to the direction of the odd spectrum (sample 08t10041p) without breaking the whole model. In the second case, trying to tailor the model to sample 08t10101p with false reference data would have considerably decreased the performance at the other samples, thus, the model ignored this sample.

Since presumably the rest of the samples have fairly correct reference data and NIR spectra, the calibration model accentuates those spectral regions that are in relation with the reference data in question. Having false reference data, these spectral regions (and mainly any other regions) provide no useful information for the approximation, thus, it is nearly impossible to fit a falsely analyzed sample into a perfect model.

The error of the reference method can originate from a measurement or sampling mistake, or from the inaccuracy of the specific laboratory analytical method. Inaccuracy is always present in every analytical method, and most of the times it is fully acceptable. It is intended to avoid mistakes, and decrease inaccuracy. The question is, how much the permissible inaccuracy is, since this will define the accuracy of any further models using these inaccurate reference data.

The false reference data are easily recognizable and can appear as sample 08t10101p, but the inaccurate measurements will provide not much different values than the correct ones, so these will appear much closer to the model, probably with increasing distance at one or two ends of the scale – as a nature of the analytical model's accuracy.

3.4. Improving results with outlier selection

Considering all the above described reasons, outlier selection methods 1 and 2, based on Y-residuals (see Materials and Methods section), were applied to develop an objective procedure for cleaning the dataset for NIR and reference outliers that caused significant drop of model performance. Elimination of no more than 5–10% of the samples was aimed in order to keep the robustness of the models. Outlier selection methods were performed in three spectral intervals (1150–1800 nm, 1300–1600 nm, 1650–1800 nm).

Table 2 shows the cross-validation results of the original PLSR models using all samples, while Table 3 summarizes the cross-validation results of the PLSR models performed on the datasets that were cleaned using the described outlier selection methods (method-1 and 2). Both methods provided greatly improving results when compared to the original dataset: decreasing RMSE and bias, increasing R^2 . The number of applied latent variables, and the difference between calibration and cross-validation results have also decreased (not shown here), indicating the increased robustness of the models.

Table 2

Results of the PLSR models built on the clinicochemical constituents of the serum samples using various ranges of the transmission NIR spectra of all samples.

| Constituent | range [nm] | n | n _{LV} | R ² _{CV} | RMSE _{CV} | bias _{CV} |
|--------------------------|------------|-----|-----------------|------------------------------|--------------------|--------------------|
| Total lipid (TL) | 1150–1800 | 103 | 5 | 0.697 | 1.080 | 0.000 |
| | 1300–1600 | 103 | 5 | 0.627 | 1.200 | 0.002 |
| | 1650–1800 | 103 | 2 | 0.695 | 1.085 | 0.003 |
| Triglyceride (TG) | 1150–1800 | 103 | 6 | 0.814 | 0.587 | 0.019 |
| | 1300–1600 | 103 | 6 | 0.767 | 0.658 | 0.021 |
| | 1650–1800 | 103 | 2 | 0.745 | 0.688 | 0.006 |
| Total cholesterol (TCh) | 1150–1800 | 103 | 7 | 0.751 | 0.343 | −0.001 |
| | 1300–1600 | 103 | 3 | 0.481 | 0.496 | 0.003 |
| | 1650–1800 | 103 | 4 | 0.741 | 0.350 | −0.001 |
| HDL cholesterol (HDL-Ch) | 1150–1800 | 101 | 4 | 0.708 | 0.314 | 0.001 |
| | 1300–1600 | 101 | 4 | 0.573 | 0.379 | 0.003 |
| | 1650–1800 | 101 | 4 | 0.704 | 0.316 | 0.003 |

n: number of samples.

n_{LV}: number of latent variables.

R²_{CV}: determination coefficient of cross-validation.

RMSE_{CV}: root mean squared error of cross-validation.

bias_{CV}: systematic error (mean residual) of cross-validation.

The predictability of cholesterol (TCh and HDL-Ch) was less than that of TL and TG, which may indicate either the lack of sufficient information on cholesterol content in the NIR spectra, or less accurate reference values for cholesterol measurement than for TL and TG. According to the performance data of the applied clinicochemical reference methods, the relationship of absorbance and concentrations can be exactly described ($R^2 > 0.999$) with linear models in all tests, in the concerning concentration intervals. Reproducibility tests of TG, TCh, and HDL-Ch gave coefficient of variance ($CV\% = (SD / \text{mean}) \times 100$) values as follows: 1.57, 1.79, and 4.01, respectively. This shows slight advantage of the reference methods of TG over those of TCh and HDL-Ch.

Figs. 4 and 5 depict the results of PLSR models of TG and HDL-Ch constituents, respectively, also presenting the loading vectors

Table 3

Results of the PLSR models built on the clinicochemical constituents of the serum samples using various ranges of the transmission NIR spectra after different outlier selection methods.

| const. | range [nm] | Method-1 (residual vs. reference-SD) | | | | | Method-2 (residual vs. residual-SD) | | | | |
|--------|------------|--------------------------------------|-----------------|------------------------------|--------------------|--------------------|-------------------------------------|-----------------|------------------------------|--------------------|--------------------|
| | | n | n _{LV} | R ² _{CV} | RMSE _{CV} | bias _{CV} | n | n _{LV} | R ² _{CV} | RMSE _{CV} | bias _{CV} |
| TL | 1150–1800 | 100 | 4 | 0.883 | 0.642 | 0.004 | 97 | 4 | 0.901 | 0.578 | 0.002 |
| | 1300–1600 | 96 | 5 | 0.840 | 0.730 | −0.002 | 94 | 5 | 0.846 | 0.699 | −0.003 |
| | 1650–1800 | 98 | 2 | 0.892 | 0.596 | 0.002 | 96 | 2 | 0.912 | 0.541 | 0.001 |
| TG | 1150–1800 | 100 | 4 | 0.909 | 0.400 | 0.002 | 95 | 5 | 0.942 | 0.312 | 0.004 |
| | 1300–1600 | 98 | 6 | 0.910 | 0.393 | 0.005 | 93 | 5 | 0.926 | 0.350 | 0.000 |
| | 1650–1800 | 100 | 3 | 0.895 | 0.436 | 0.001 | 97 | 3 | 0.896 | 0.412 | 0.001 |
| TCh | 1150–1800 | 93 | 7 | 0.778 | 0.278 | 0.001 | 88 | 7 | 0.823 | 0.246 | 0.004 |
| | 1300–1600 | 89 | 3 | 0.660 | 0.310 | 0.000 | 92 | 3 | 0.593 | 0.339 | 0.002 |
| | 1650–1800 | 96 | 4 | 0.720 | 0.284 | 0.001 | 88 | 4 | 0.769 | 0.241 | 0.001 |
| HDL-Ch | 1150–1800 | 96 | 4 | 0.808 | 0.243 | 0.000 | 94 | 4 | 0.822 | 0.229 | −0.002 |
| | 1300–1600 | 92 | 4 | 0.721 | 0.283 | 0.000 | 92 | 4 | 0.721 | 0.283 | 0.000 |
| | 1650–1800 | 95 | 4 | 0.792 | 0.249 | 0.000 | 88 | 4 | 0.850 | 0.213 | 0.001 |

n: number of samples.

n_{LV}: number of latent variables.

R²_{CV}: determination coefficient of cross-validation.

RMSE_{CV}: root mean squared error of cross-validation.

bias_{CV}: systematic error (mean residual) of cross-validation.

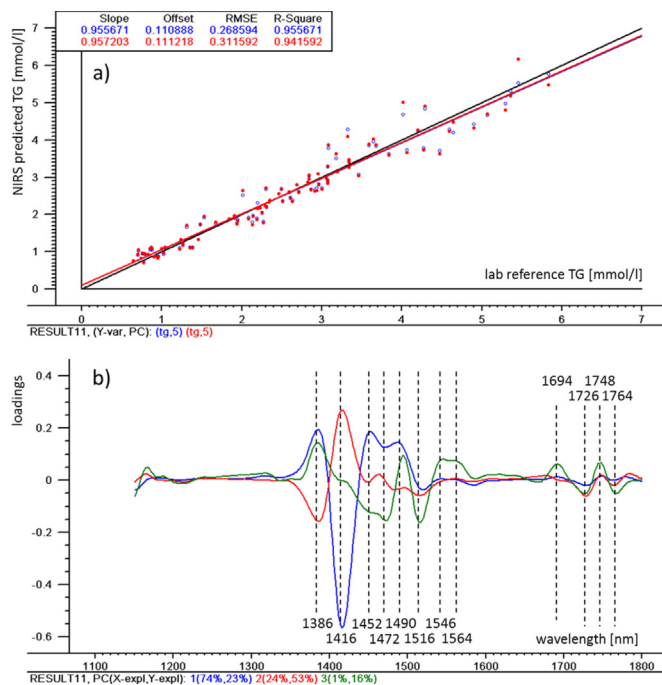


Fig. 4. Results of the PLSR model developed on the transmission NIR spectral data and the laboratory reference values of triglyceride, after method-2 outlier selection ($n=95$). Regression Y-fit (a) of calibration (blue) and cross-validation (red). The loading vectors of the first three latent variables (b) indicating the variance explained by the individual loadings in brackets (percentage of explained variance of spectral data, percentage of explained variance of reference data). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the first three latent variables. Since the higher values of loading refer to the importance of the related spectral region, calibration on TG is mainly dominated by information contained in the water absorption region (1st overtone of stretching vibrations of hydroxyl group; region of 1416 nm representing weakly H-bonded OH, region of 1490 nm representing bands of strongly H-bonded

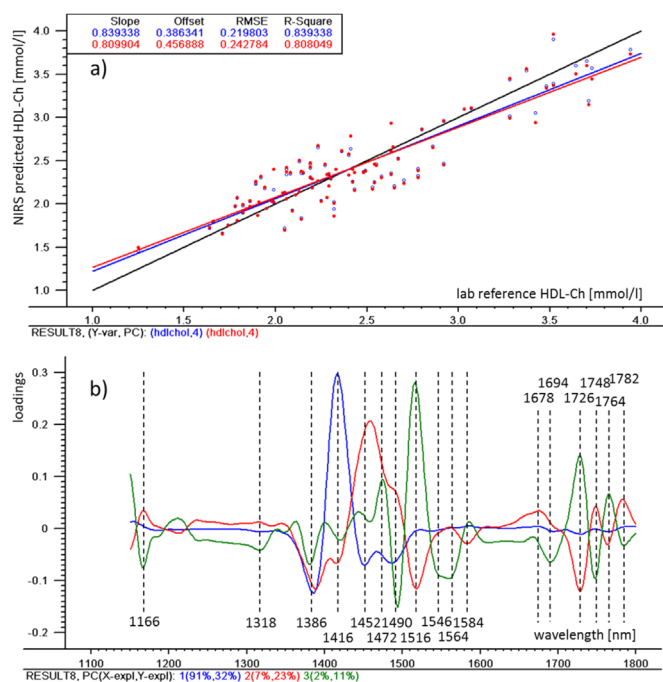


Fig. 5. Results of the PLSR model developed on the transmission NIR spectral data and the laboratory reference values of HDL cholesterol, after method-1 outlier selection ($n=96$). Regression Y-fit (a) of calibration (blue) and cross-validation (red). The loading vectors of the first three latent variables (b) indicating the variance explained by the individual loadings in brackets (percentage of explained variance of spectral data, percentage of explained variance of reference data). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

OH) [21] and less by the lipid region (1st overtone of asymmetric and symmetric stretching vibrations of methylene group ($-\text{CH}_2-$) at 1724 and 1764 nm, respectively) [22]. In the case of models on cholesterol concentration, the impact of lipid spectral region is increased. This tendency can also be followed in Table 3 (and later in Table 5 for the results calculated on absorbance spectra acquired in transmittance mode), namely the water region provides similar results to fat or total spectral regions in the case of TL and TG, but gives considerably weaker results for TCh and HDL-Ch.

The improvement in model stability after outlier selection is shown by the decrease of offset (Y-crossing) and the increase of slope of the regression line (Fig. 4(a) compared to Fig. 3(b)), i.e. the actual Y-fit got closer to the targeted optimal Y-fit after the outlier selection.

3.5. Feedback to reference methods

A total of six selection rounds were performed with the two methods using three wavelength intervals. Samples which were excluded as outliers repeatedly, at least three times out of the six rounds, even with different methods using various spectral regions, were rated as erroneous. The large residuals can indicate errors in reference measurements or samples with a different relationship between the spectrum and reference data [24], i.e. either the reference data or the spectra were imperfect. Table 4 lists the doubtful samples per investigated constituents, also indicating the number of warnings.

PCA was performed using all samples to identify spectral outliers. It had no reason to eliminate all the samples that were identified as outliers in the PCA leverage evaluations, as most of them were suitable for the calibration models. These were typically samples from the 20th day of gavage, having high fat content. Such samples could appear in the future tests again, thus, it was

Table 4

List of doubtful samples identified with presumably false reference data because of the broken relation of NIR spectra and reference values.

| Investigated constituent | Samples having presumably false data (warning frequency) | |
|--------------------------------|--|---------------|
| total lipid(TL) | 08t10015p (4) | 08t10041p*(6) |
| | 08t10027p (6) | 08t10046p (3) |
| | 08t10033p*(6) | 08t10070p (4) |
| triglyceride(TG) | 08t10027p (3) | 08t10076p (6) |
| | 08t10031p*(3) | 08t10101p (6) |
| | 08t10041p*(6) | |
| total cholesterol(TCh) | 08t10002p (4) | 08t10031p*(4) |
| | 08t10003p*(3) | 08t10034p (3) |
| | 08t10006p*(6) | 08t10041p*(5) |
| | 08t10011p (3) | 08t10049p (3) |
| | 08t10013p (4) | 08t10053p (3) |
| | 08t10016p*(6) | 08t10086p (3) |
| HDL cholesterol(HDL-Ch) | 08t10009p (6) | 08t10041p*(6) |
| | 08t10014p (4) | 08t10084p (3) |
| | 08t10022p (6) | 08t10099p (6) |
| | 08t10036p (3) | 08t10101p (6) |

The frequency of warning incidences within the six selection rounds (two selection methods with three wavelength intervals) is shown in bracket behind each sample number; samples having spectra identified as outlier in PCA leverage evaluations are marked with asterisk (*).

reasonable to keep these samples in the model to increase robustness. However, asterisk indicates those samples in Table 4 which were identified as outliers both in the above described selection methods and in the PCA leverage evaluations. The clinicochemical reference data of these samples were most probably correct, but the samples represented extremities (hyperlipidemia) with highly different NIR spectra (with strong scattering because of dispersed lipid) that were not applicable for the developed chemometric models. Elimination of these samples from the calibrations seemed to be reasonable, but not on the basis of false reference data.

Based on the selection methods, 27 samples in 33 cases were marked because of extreme distance from the regression model. Out of these, there were six samples identified as spectral outliers in PCA, covering ten cases. Accordingly, 23 of the 410 reference data were rated as inaccurate based on the NIR regressions.

3.6. Comparison of spectrum acquisition modes

Comparing results achieved with NIR absorbance spectra of transmission (Table 3) and transmittance (Table 5) acquisition modes, in the case of almost all constituents, more accurate (larger R^2 and lower RMSE) and more robust (less LV) models were obtained with transmission data. Not only cross-validation results, but also independent model validations provided better results for transmission data (Table 6). Although these evidences showed that transmission data held more information on the aimed parameters than transmittance data did, it must be noted, that the results achieved with the two datasets were comparable and competitive. Transmittance measurement technique can be an important and very useful alternative in such cases when only extremely small sample quantities ($< 100 \mu\text{l}$) are available for spectroscopic analysis. In this trial, TL and TG were measured with high accuracy using transmittance NIR data, while TCh and HDL-Ch were measured with moderate accuracy (Tables 5 and 6).

Table 5
Results of the PLSR models built on the clinicochemical constituents of the serum samples using various ranges of the transreflectance NIR spectra after different outlier selection methods.

| const. | range [nm] | Method-1 (residual vs. reference-SD) | | | | | Method-2 (residual vs. residual-SD) | | | | |
|--------|------------|--------------------------------------|-----------------|------------------------------|--------------------|--------------------|-------------------------------------|-----------------|------------------------------|--------------------|--------------------|
| | | n | n _{LV} | R _{CV} ² | RMSE _{CV} | bias _{CV} | n | n _{LV} | R _{CV} ² | RMSE _{CV} | bias _{CV} |
| TL | 1150–1800 | 92 | 5 | 0.890 | 0.606 | 0.006 | 91 | 5 | 0.890 | 0.588 | 0.005 |
| | 1300–1600 | 90 | 5 | 0.812 | 0.772 | 0.004 | 91 | 5 | 0.802 | 0.792 | 0.003 |
| | 1650–1800 | 96 | 1 | 0.893 | 0.618 | −0.004 | 93 | 1 | 0.911 | 0.560 | −0.005 |
| TG | 1150–1800 | 95 | 4 | 0.860 | 0.512 | 0.004 | 90 | 4 | 0.874 | 0.472 | 0.003 |
| | 1300–1600 | 91 | 6 | 0.827 | 0.544 | 0.005 | 86 | 5 | 0.858 | 0.480 | 0.004 |
| | 1650–1800 | 93 | 1 | 0.869 | 0.482 | 0.000 | 92 | 1 | 0.875 | 0.471 | 0.000 |
| TCh | 1150–1800 | 88 | 9 | 0.683 | 0.297 | −0.007 | 86 | 10 | 0.769 | 0.286 | −0.001 |
| | 1300–1600 | 78 | 8 | 0.403 | 0.344 | −0.004 | 87 | 5 | 0.219 | 0.421 | 0.000 |
| | 1650–1800 | 86 | 6 | 0.689 | 0.303 | −0.002 | 84 | 5 | 0.690 | 0.289 | −0.006 |
| HDL-Ch | 1150–1800 | 87 | 8 | 0.734 | 0.296 | 0.005 | 84 | 8 | 0.752 | 0.274 | 0.003 |
| | 1300–1600 | 78 | 7 | 0.534 | 0.331 | −0.002 | 82 | 7 | 0.536 | 0.343 | −0.001 |
| | 1650–1800 | 78 | 5 | 0.701 | 0.275 | −0.005 | 81 | 5 | 0.684 | 0.299 | −0.002 |

n: number of samples.

n_{LV}: number of latent variables.

R_{CV}²: determination coefficient of cross-validation.

RMSE_{CV}: root mean squared error of cross-validation.

bias_{CV}: systematic error (mean residual) of cross-validation.

Table 6
Results of the independent validations of the PLSR models built on the clinicochemical constituents of the serum samples using various ranges of the NIR absorbance spectra acquired in transmission and transreflectance modes. Samples having outlier spectral data or doubtful reference value were not included in calibration models or validation test sets. Number of samples in validation: 20.

| Const. | range [nm] | Transmission mode | | | | | Transreflectance mode | | | | |
|--------|------------|-------------------|-----------------|-----------------------------|-------------------|-------------------|-----------------------|-----------------|-----------------------------|-------------------|-------------------|
| | | n | n _{LV} | R _p ² | RMSE _p | bias _p | n | n _{LV} | R _p ² | RMSE _p | bias _p |
| TL | 1150–1800 | 77 | 5 | 0.854 | 0.662 | −0.199 | 73 | 4 | 0.792 | 0.783 | −0.074 |
| | 1300–1600 | 77 | 6 | 0.819 | 0.735 | −0.178 | 73 | 4 | 0.681 | 0.970 | −0.061 |
| | 1650–1800 | 77 | 2 | 0.864 | 0.639 | −0.233 | 73 | 1 | 0.862 | 0.638 | −0.062 |
| TG | 1150–1800 | 78 | 3 | 0.966 | 0.232 | 0.061 | 74 | 6 | 0.861 | 0.448 | −0.029 |
| | 1300–1600 | 78 | 4 | 0.951 | 0.277 | 0.117 | 74 | 6 | 0.828 | 0.499 | −0.042 |
| | 1650–1800 | 78 | 3 | 0.920 | 0.356 | 0.031 | 74 | 2 | 0.833 | 0.492 | −0.009 |
| TCh | 1150–1800 | 69 | 4 | 0.766 | 0.262 | −0.031 | 65 | 9 | 0.683 | 0.338 | −0.097 |
| | 1300–1600 | 69 | 4 | 0.609 | 0.340 | 0.003 | 65 | 4 | 0.463 | 0.440 | −0.105 |
| | 1650–1800 | 69 | 5 | 0.850 | 0.210 | −0.063 | 65 | 7 | 0.697 | 0.330 | −0.047 |
| HDL-Ch | 1150–1800 | 75 | 6 | 0.793 | 0.241 | −0.007 | 71 | 7 | 0.632 | 0.339 | −0.043 |
| | 1300–1600 | 75 | 6 | 0.681 | 0.299 | 0.022 | 71 | 5 | 0.484 | 0.401 | 0.061 |
| | 1650–1800 | 75 | 4 | 0.778 | 0.249 | −0.017 | 71 | 5 | 0.571 | 0.366 | 0.010 |

n: number of samples in the calibration model.

n_{LV}: number of latent variables.

R_p²: determination coefficient of prediction.

RMSE_p: root mean squared error of prediction.

bias_p: systematic error (mean residual) of prediction.

3.7. Independent model validation

Twenty samples of those identified having good spectral and reference data were randomly selected and used for independent validation tests for models generated without these samples. The validation results are summarized in Table 6 both for transmission and transreflectance datasets. Although bias at TL and TG increased considerably, other indicators of the independent validation performance remained at the level experienced in the cross-validation. This proved the applicability of the achieved models.

4. Conclusions

Quantitative NIR spectroscopy as a correlative technique is based on the reference data obtained from certified chemical or physical measurements. Having a sufficient database with NIR spectra and reference data on constituents, it is possible to predict the composition of unknown samples of the same sort with high accuracy, by using their NIR spectra solely. In this study, accurate models were presented on multiple clinicochemical parameters of blood serum of force-fed geese, applying NIR absorbance spectra

acquired in transmission and transreflectance measurement modes.

NIR technique provides not only a quick, cost effective, non-hazardous alternative method for analysis of several constituents based on one single spectrum of each sample, but it offers the possibility for looking at the laboratory reference data critically, and evaluating them to identify and exclude falsely analyzed samples. In the case of doubtful samples identified with presumably false reference data because of the broken relation of NIR spectra and reference values, NIR spectroscopy, as a process monitoring approach, can provide warning feedback to the reference laboratory. This can be especially useful in relation to analyses where laboratory methods are not perfectly suited to the subjected material and there is an increased chance of laboratory error (e.g. analyzing goose serum with automats designed for human samples).

Accuracy of NIR spectroscopy is highly influenced by the reliability (precision and accuracy) of the laboratory methods providing reference data. Applying the described selection methods for reference samples, the accuracy of NIR spectroscopic predictions can be increased significantly, without losing robustness of models.

Acknowledgements

Authors are grateful to David Kamocsa for his assistance at data acquisition. The study was supported by the Hungarian Academy of Sciences – Kaposvár University “MTA-KE Mycotoxins in the Food Chain” Research Group.

References

- [1] The foie gras sector in a few figures, 2015. (http://www.eurofoiegras.com/en/page/figures_p135/) (accessed May 7, 2015).
- [2] J. Kozák, Marketing and production of fattened goose and duck liver [in Hungarian], *Gazdálkodás* 55 (2011) 309–316.
- [3] L. Locsmándi, G. Hegedüs, G. Andrásy-Baka, F. Bogenfürst, R. Romvári, Following the goose liver development by means of cross-sectional digital imaging, liver histology and blood biochemical parameters, *Acta Biol. Hung.* 58 (2007) 35–48.
- [4] S. Davail, G. Guy, J.-M. André, D. Hermier, R. Hoo-Paris, Metabolism in two breeds of geese with moderate or large overfeeding induced liver-steatosis, *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 126 (2000) 91–99.
- [5] M.L. Bishop, E.P. Fody, L.E. Schoeff, *Clinical Chemistry: Principles, Techniques, and Correlations*, Lippincott Williams & Wilkins, Philadelphia, USA, 2013.
- [6] J.A. Knight, S. Anderson, J.M. Rawle, Chemical basis of the sulfo-phospho-vanillin reaction for estimating total serum lipids, *Clin. Chem.* 18 (1972) 199–202.
- [7] J.W. Hall, A. Pollard, Near-infrared spectrophotometry: a new dimension in clinical chemistry, *Clin. Chem.* 38 (1992) 1623–1631.
- [8] H.M. Heise, R. Marbach, A. Bittner, T. Koschinsky, Clinical chemistry and near infrared spectroscopy: multicomponent assay for human plasma and its evaluation for the determination of blood substrates, *J. Infrared Spectrosc.* 6 (1998) 361–374.
- [9] G. Domján, K.J. Kaffka, M. Jákó, I.T. Vályi-Nagy, Rapid analysis of whole blood and blood serum using near infrared spectroscopy, *J. Infrared Spectrosc.* 2 (1995) 67–78.
- [10] S. Turza, M. Kurihara, S. Kawano, Near infrared analysis of whole blood and plasma in blood-collecting tubes, *J. Infrared Spectrosc.* 14 (2006) 147–153.
- [11] C. Petter, N. Heigl, R. Bakry, G. Bonn, A. Ritsch, C. Huck, Quantification of low-density and high-density lipoproteins in human serum by Material Enhanced Infrared Spectroscopy (MEIRS), *Curr. Med. Chem.* 16 (2009) 4601–4608.
- [12] A.M.C. Davies, What is and what is not chemometrics, *Spectrosc. Eur.* 24 (2012) 33.
- [13] P. Mortensen Paash, Myth: a partial least squares calibration model can never be more precise than the reference method, *NIR News* 25 (2014) 20–22.
- [14] D.B. Coates, Is near infrared spectroscopy only as good as the laboratory reference values? An empirical approach, *Spectrosc. Eur.* 4 (2002) 24–26.
- [15] T. Naes, T. Isaksson, T. Fearn, T. Davies, *A user friendly guide to multivariate calibration and classification*, NIR Publications, Chichester, UK, 2002.
- [16] P. Trinder, Determination of glucose in blood using glucose oxidase with an alternative oxygen acceptor, *Ann. Clin. Biochem.* 6 (1969) 24–27.
- [17] P. Röschlau, E. Bernt, W. Gruber, Enzymatische Bestimmung des Gesamt-Cholesterins im Serum [in German], *Clin. Chem. Lab. Med.* 12 (1974) 403–407.
- [18] K.H. Norris, Instrument Design. Interactions among instrument bandpass, instrument noise, sample-absorber bandwidth and calibration error, *NIR News.* 9 (1998) 3.
- [19] T. Fearn, The interaction between standard normal variate and derivatives, *NIR News.* 19 (2008) 16.
- [20] I.A. Cowe, J.W. McNicol, The use of principal components in the analysis of near-infrared spectra, *Appl. Spectrosc.* 39 (1985) 257–266.
- [21] V.H. Segtnan, Š. Šašić, T. Isaksson, Y. Ozaki, Studies on the structure of water using two-dimensional near-infrared correlation spectroscopy and principal component analysis, *Anal. Chem.* 73 (2001) 3153–3161.
- [22] J. Workman, L. Weyer, *Practical Guide to Interpretive Near-Infrared Spectroscopy*, CRC Press, Taylor & Francis Group, LLC, Boca Raton, FL, USA, 2007.
- [23] H. Martens, E. Stark, Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy, *J. Pharm. Biomed. Anal.* 9 (1991) 625–635.
- [24] CAMO Software, *unscrambler manual methods - Google keresés, CAMO Software, As, Norway, 2006.* (http://www.camo.com/downloads/U9.6_pdf_manual/The_Unscrambler_Methods.PDF) (accessed April 4, 2016).