Improved stability of blood glucose measurement in humans using near infrared spectroscopy

C.F. So [a,*], Joanne W.Y. Chung [b], Maggie S.M. Siu [a] and Thomas K.S. Wong [a]

[a] *Centre for Integrative Digital Health, School of Nursing, The Hong Kong Polytechnic University, Kowloon, Hong Kong*
[b] *Department of Health and Physical Education, The Hong Kong Institute of Education, Kowloon, Hong Kong*

Abstract. Near infrared (NIR) spectroscopy has become a promising technique for blood glucose monitoring. However, an appropriate model of spectral response in humans is yet to be determined because of the reliability problem. In this study, 48 subjects were recruited. The subjects' left forearms were scanned using near infrared spectroscopy to obtain NIR spectra. Simultaneously, a blood sample of glucose was drawn. A new method based on Monte Carlo approach is applied for partial least squares (PLS), named as PLS_{MC}, is proposed. A large numbers of models are built from calibration subsets which are randomly selected from the whole calibration set in order to minimize the noises. It is then determining the mean value over the models with high correlation and small prediction errors. The results show that the method can enhance the stability of PLS model. Also, the performance of the PLS_{MC} shows more accurate prediction results as compared with conventional PLS.

Keywords: Blood glucose, partial least squares, near infrared, prediction model, spectroscopy

1. Introduction

Currently, about 220 million people have diabetes mellitus (DM) worldwide [22,23]. Due to the world's increasingly ageing populations, increasingly unhealthy diets, sedentary lifestyles and obesity, it is estimated that the prevalence of DM will double by 2030 [21,23]. Diabetics need to have their blood glucose levels closely monitored, with lancing devices, more than three times per day to avoid serious complications. Nevertheless many people find experiences that involve sharp objects and blood at best unpleasant. Diabetic patients also frequently suffer pain and are at risk of infection. Thus, the comfort advantage of a non-invasive technology is easily understood and for this purpose many researchers are working hard to develop a non-invasive method for blood glucose monitoring [1,8,10,13,17,19]. Indeed, this will help make monitoring a more pleasant routine, resulting in more accurate administering of DM drugs and better management of DM.

Many researchers are developing a variety of non-invasive methods that monitor blood glucose [1, 8,10,13,19]. Near infrared (NIR) spectroscopy has become a promising technique for blood glucose monitoring among those potential non-invasive approaches. However, an appropriate model of spectral response in humans is yet to be determined. The key component of NIR spectral quantitative analysis

---

*Corresponding author: C.F. So, Centre for Integrative Digital Health, School of Nursing, The Hong Kong Polytechnic University, Kowloon, Hong Kong. E-mail: hscfso@polyu.edu.hk.

is dependent on multivariate calibration methods that require enough and useful calibration spectra and sufficient spectral data points to allow analytical information to be extracted from spectra accurately. The quality of a multivariate calibration model mainly depends on the quality of both response and independent variables. Undoubtedly, partial least squares (PLS) [11,15] is the most common use of multivariate calibration procedure to build the quantitative model in NIR spectroscopy. It is because PLS attempts to maximize covariance between the response and independent variables. However, the performance of the PLS model will be severely deteriorated in the presence of outliers [4,14]. In practice, analytical data matrix generally includes unexpected experimental errors or measurement noise and normally contains hundreds of samples in the NIR spectra. Hence, outliers are possibly existed in such a large data matrix.

The fact that outliers can deteriorate the quantitative model derived from a normal PLS analysis has motivated the development of algorithms that are less affected by outliers [7,9,16]. These algorithms should be effective for extraction of data for use in process modeling. Although various PLS methods [9,12,16] have been developed for the problem of outliers, most of these methods use a single quantitative model to predict the relationship between NIR spectra and blood glucose concentration. A single prediction model from PLS analysis may underestimate or overemphasize some features and even worse ignore some important characteristics contained in the largely complex NIR spectra. The prediction results obtained by a single prediction model may be unstable or correlated with spurious spectral variance particularly when the training set is comparatively small. In this regard, a new approach to enhance both stability and performance of PLS prediction model is required.

In this paper, an improved stability and performance of PLS prediction model derived from Monte Carlo simulation [3,6,18] is presented. Monte Carlo method is used to rate and analyze NIR spectra by simulating a large number of PLS models from calibration subsets which are randomly selected from the whole calibration set. The reason is that, comparatively outliers will be significantly less than that of normal samples, by which the PLS models can minimize the adverse effect caused by the outliers. It is then determining the mean value over the models with small prediction errors. The advantage of Monte Carlo method over other techniques increases as the sources of uncertainty increase, it is therefore suitable for predicting blood glucose concentration since glucose measurement in humans is subject to a number of confounders. The results show that Monte Carlo method can improve both stability and performance of PLS prediction model of the NIR spectrometry.

2. Algorithm of PLS$_{MC}$

In linear least squares model, the prediction **y** is computed by the equation:

$$\mathbf{y} = \mathbf{Xb} + \varepsilon, \tag{1}$$

where $\mathbf{X}$ is a $n \times p$ matrix contains p dependent variables of n samples, $\mathbf{b}$ is a $p \times 1$ vector of regression coefficients obtained from PLS analysis and ε is the model offset.

In this work, a large number of PLS models with different calibration subsets selected by the Monte Carlo method are produced. The stability of the corresponding coefficient is calculated by using the regression coefficients of these models. The procedure has advantages of reducing dependence on single model and it can be expected that the stability of the prediction results will be greatly enhanced. Moreover, the effect of the outlier can be minimized with such a large number of PLS models because the outliers can deteriorate the prediction errors of the PLS models. Therefore, the PLS models with

smallest prediction errors are most likely to be assumed as the model without the outliers. Based on this assumption, a simple criterion is proposed to identify the outliers and the procedures of this approach can be described as below.

The raw spectra will be randomly divided into 2 parts – calibration set $\mathbf{X}_c$ and prediction set $\mathbf{X}_p$. The prediction set will be used to evaluate the robustness of the prediction model because it will not involve in training the model. A large number of subset, i.e., 80% of the samples in the calibration set will randomly select for calibration, and therefore a large number of PLS models are constructed by using these subsets from the whole calibration set. After that, these PLS models will be validated by the prediction set, with those provide smallest root mean square error (RMSE) of prediction, i.e., 5% of the overall models, will be selected and taking the averaged prediction of these models as the final model. The RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{k} \sum_{i=1}^{k} (\mathbf{y}_i - \tilde{\mathbf{y}}_{i,k})^2}, \tag{2}$$

where $\mathbf{y}_i$ and $\tilde{\mathbf{y}}_{i,k}$ are the measured and predicted concentration of the ith sample and k is the size of calculation set.

And the averaged PLS prediction model is calculated as:

$$\bar{\mathbf{b}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{b}_i, \tag{3}$$

where N is the number of 5% of the overall models, $\mathbf{b}_i$ is the prediction model of the smallest 5% RMSE of prediction of the overall models.

The main advantage of this approach is to obtain an enough number of PLS models without outliers for the purpose of outlier detection. Finally, the performance of this averaged PLS model is assessed by means of the root mean square error of prediction (RMSE$_p$).

3. Materials and methods

We assembled the set of instruments as the ASD LabSpec® NIR spectrometer with internal halogen light source and optical probe to measure the NIR spectra responses of each subject. The detector used in the study was graded index InGaAs photodiodes. Figure 1 shows the schematic illustration of apparatus for obtaining a NIR spectral scan of a left forearm of a person. The absorbance responses of the spectrum obtained were averaged 10 times for each subject.

Since the colour of skin maybe one of factors that affect the NIR spectroscopy measurement, therefore in this study, we focused on the East Asia group people that are person of Asian descent to avoid the different combination of the skin colour, thereby affecting the model validation. In this experiment, all spectra were collected in the wavelength range 1100–1400 nm for the reason of its deeper penetration into tissue and the overtone band of glucose. Figure 2 shows an example of the NIR spectra with wavelength range 1100–1400 nm. Moreover, another reason to choose this wavelength range is that it avoids the absorption peak of water at around 1440 nm since water is always a critical component for NIR spectra [2,20,24].
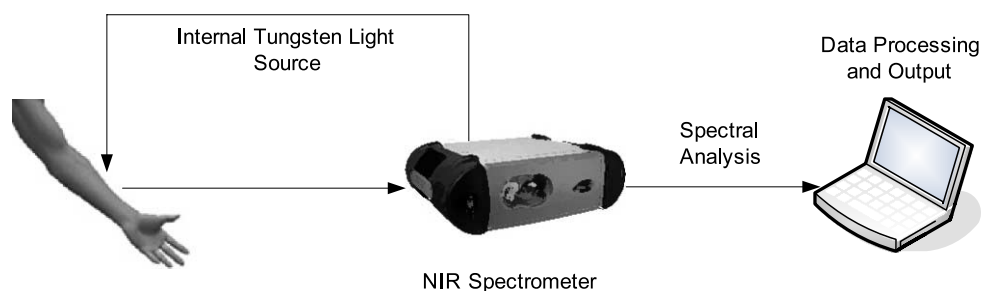
Fig. 1. Schematic illustration of apparatus for obtaining a NIR spectral scan of a left forearm. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/SPE-2011-0507.)
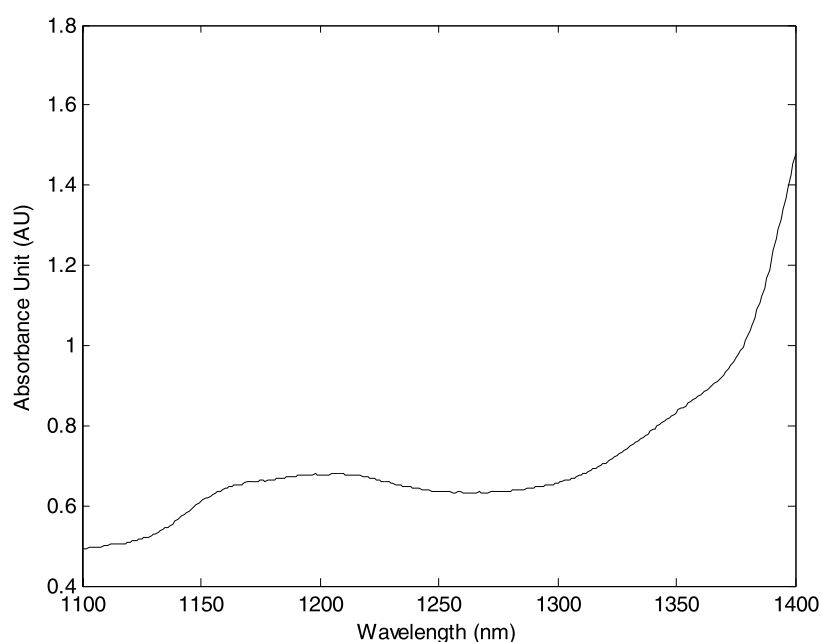

Fig. 2. An example of the measured NIR spectra.

48 adults participated in the study voluntarily. Among them, 23 (47.9%) suffered from DM. For those suffering from DM, their situation is to some extent well controlled. This is a group of community subjects and thus this is different from that hospitalization DM subjects whose blood glucose level is not under controlled. Prior to the data collection, we explained the aim and procedure of the study to the subjects and obtained their written consent. They were informed that they could withdraw from the study without penalty. The night before the study, each subject was required to fast for 12 h after dinner (they were only allowed to drink plain water during the fasting period). Each subject attended our clinic the next morning for the study. All subjects with DM history were required to have one more blood taking 30 min after they had their breakfast. 2 ml of venous blood were drawn for blood glucose measurement. The subject's left forearms (mid-way between the radial side of the first wrist crease and the medial side of the elbow) were scanned using near infrared spectroscopy to obtain NIR spectra. The study was approved by the Human Subjects Ethics Committee of The Hong Kong Polytechnic University.

4. Results and discussions

There were 37 females and 11 males, with a mean age of 52.44 (SD 13.37). The mean blood glucose range was 7.89 (SD 3.76). We used NIR spectra (wavelength 1100–1400 nm) and blood glucose levels (BGL) ranging from 3.7–16.8 mmol/l in this experiment. The total number of samples was 68 and the data were divided into two sets, one of which (60% of the total number of samples, 40 subjects) was randomly selected for calibration purposes, in which PLS was performed to maximize the correlation between the predicted and reference values; the remaining data set (40% of the total number of samples, 28 subjects) was then fitted to the calibration model for prediction. Each time, a subset of calibration samples (32 out of 40 samples) was taken for training. All the results are averaged over 20 runs.

Firstly, we determine the number of the latent variable (LV) of PLS model because it is an important parameter in the modeling. Figure 3 shows the variation of $RMSE_p$ with the number of LV of the PLS_{MC} and conventional PLS. In this simulation, 1000 PLS models are built and the optimal 5% of the overall models (50 models) are selected for the PLS_{MC}. Same calibration sets and prediction sets are used in both models for the sake of fairness. The figure shows that the $RMSE_p$ of PLS_{MC} has a descending trend with the increase of the number of LV while stabilize after LV larger than 10. On the other hand the conventional PLS performs worse when the number of LV increases. The results confirmed the use of Monte Carlo approach for PLS can well perform the conventional PLS. In order to simplify the analysis and to make the model as fewer complexes as possible, the number of LV will be set equal to 10 for the following simulations.

Secondly, we determine the number of PLS models for the use of PLS_{MC} which decides the stability and accuracy of the model. When the number of PLS models is too small, the stability and accuracy of the model may be affected because it may either overemphasize or underestimate some aspects, and ignore many important features contained in such a large NIR data set. However, if the number of PLS models is
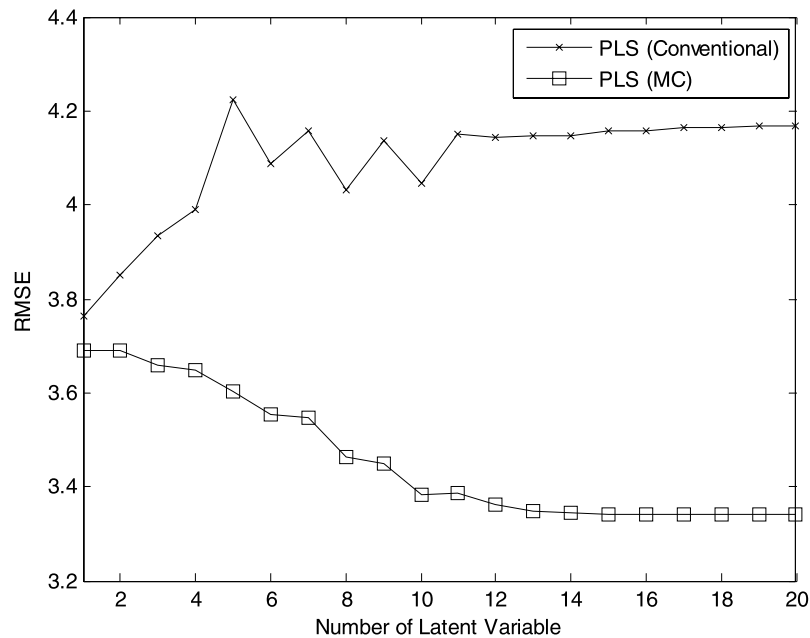

Fig. 3. Variation of $RMSE_p$ with the number of LV.

too large, uninformative features may be contained in the model and even worsen the prediction result. Hence, the variation of the $RMSE_p$ of the number of PLS models is investigated. Figure 4 shows the $RMSE_p$ obtained with various number of PLS models from 20 to 80,000. For each simulation, 5% of overall models that provided smallest $RMSE_p$ are selected and averaged the model to develop the PLS_{MC} model. The number of models selected and the standard error of the $RMSE_p$ for the simulation trails are shown in Table 1. Simulation results shown that, the $RMSE_p$ is large for the case of 1 model selected out of 20 PLS models, then with the number of PLS models increase, the $RMSE_p$ becomes smaller. The $RMSE_p$ is steadily saturated at 1000 PLS models with little ripple even for the large number of PLS models (i.e., 80,000) is calculated. It can be seen that, the fewer the PLS models, the effect of outlier is
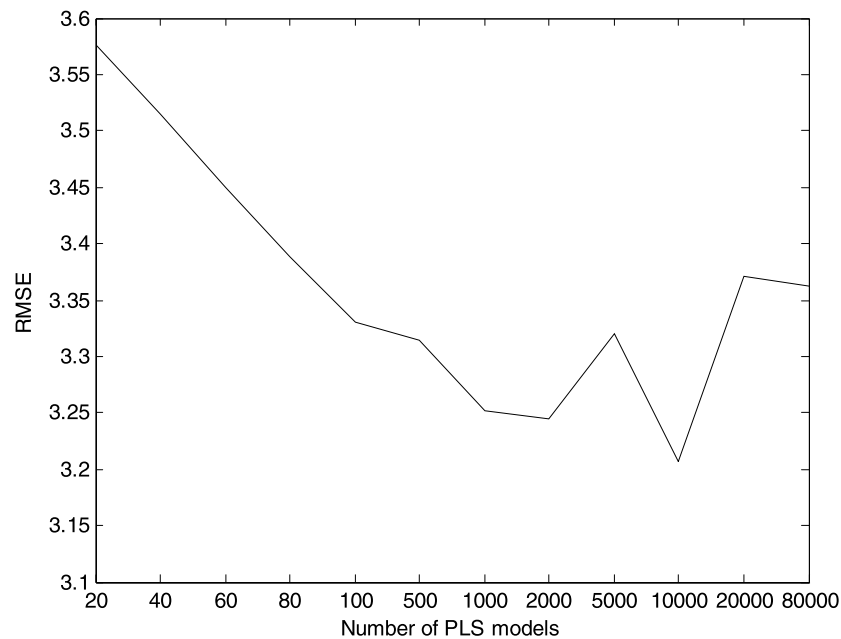

Fig. 4. Variation of $RMSE_p$ with the number of PLS models.

Table 1

$RMSE_p$ and the mean value obtained by various PLS models

No. of PLS models	No. of models selected	$RMSE_p$ (σ)
20	1	3.5759 (0.3663)
40	2	3.5142 (0.4521)
60	3	3.4495 (0.4634)
80	4	3.3882 (0.4834)
100	5	3.3309 (0.3011)
500	25	3.3147 (0.3166)
1000	50	3.2515 (0.4001)
2000	100	3.2449 (0.3122)
5000	250	3.3199 (0.3178)
10,000	500	3.2075 (0.3035)
20,000	1000	3.3713 (0.4290)
80,000	4000	3.3629 (0.4027)

more significant and thus deteriorated the performance. On the other hand, when too large PLS models are used, uninformative features will affect the prediction results because it is unavoidable to include outlier for the calibration sets. Therefore, in this study it is suggested to use 2000 PLS models for the use of PLS_{MC}.

Finally, a PLS_{MC} model was constructed based on previous findings. The training procedure was repeated 2000 times, the number of LV will be set equal to 10 and the predictor matrix was obtained with the mean of the training results. Based on the findings, a model is developed to predict the glucose value using PLS analysis. The correlation of the calibration set in cross-validation was $R = 0.72$. To further determine the accuracy of the estimation of blood glucose level using this model, Clarke Error Grid Analysis (EGA) was employed [5]. Table 2 shows the comparison between conventional PLS and PLS_{MC} and the results of PLS_{MC} are shown in Fig. 5. The estimated glucose values were plotted against the referenced glucose values. 100% of the data points fell within Zone A and Zone B. Moreover, the

Table 2

A comparison of results obtained by PLS and PLS_{MC}

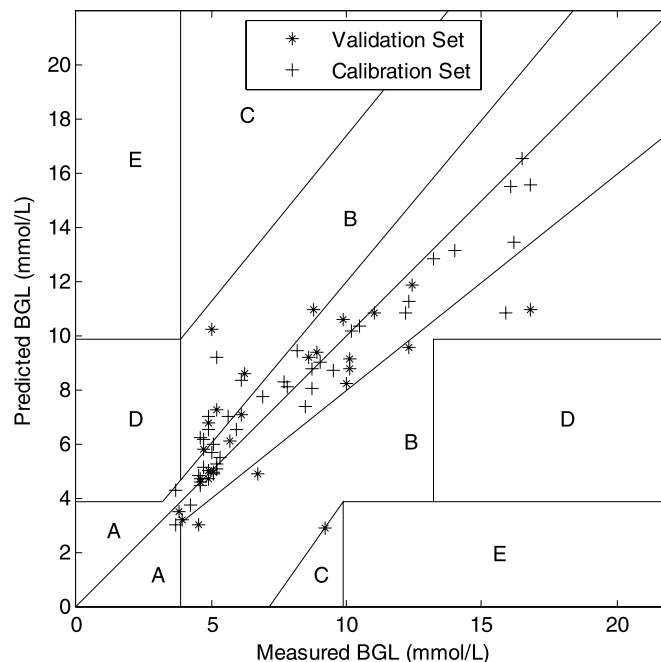|	PLS	PLS_{MC}
$RMSE_p$	3.06	2.26
R	0.58	0.72
EGA		
A	50	49
B	14	19
C	2	0
D	2	0
E	0	0

Fig. 5. Calibration and validation results of Clarke Error Grid Analysis ($+$ is calibration set and $*$ is validation set), $R = 0.72$.

RMSE$_p$ is smaller than the conventional PLS which have 25% improvement. The correlation coefficient is also have 24% improvement which proven that the Monte Carlo approach can not only stabilize the conventional PLS model, also improve the performance of the prediction model. The results of these experiments and field trials support the feasibility of measuring blood glucose level using near infrared (NIR) spectroscopy.

Though we have developed an improved stability models, we need to interpret the results with caution. The limitation inherent in the study is the impossibility of using a higher range spectrometer (wavelength range 2000–2500 nm), which is due to the high temperature generated by the light source. The heat may cause harm and be intolerable to human subjects. Thus, there is a possibility that we lost the higher spectra portion for analysis. Moreover, given the low signal to noise ratio, the power of the statistics can be further improved with a better filtering technique for noise and an increase in sample size.

5. Conclusion

The wavelength selection based on the NIR absorption bands was discussed and selected. We have presented a new method, Monte Carlo approach, to average the PLS model result, the model provides stable and accurate results for non-invasive blood glucose estimation. The blood glucose level (BGL) ranging from 3.7–16.8 mmol/l using PLS$_{MC}$ analysis is presented. The method was found to produce a model providing accurate and computation for the wide spread of glucose readings in this study.

References

[1]   M.A. Arnold and G.W. Small, Noninvasive glucose sensing, *Analytical Chemistry* **77**(17) (2005), 5429–5439.
[2]   D.A. Burns and E.W. Ciurczak, *Handbook of Near Infrared Analysis*, 3rd edn, CRC Press/Taylor & Francis, Boca Raton, FL, 2008.
[3]   D. Chen, W.S. Cai and X.G. Shao, A strategy for enhancing the reliability of near-infrared spectral analysis, *Vibrational Spectroscopy* **47** (2008), 113–118.
[4]   D. Chen, X.G. Shao, B. Hu and Q.D. Su, A background and noise elimination method for quantitative calibration of near infrared spectra, *Analytica Chimica Acta* **511** (2004), 37–45.
[5]   W.L. Clarke, D. Cox, L.A. Gonder-Frederick, W. Carter and S.L. Pohl, Evaluating clinical accuracy of systems for self-monitoring of blood glucose, *Diabetes Care* **10** (1987), 622–628.
[6]   A. Doucet, N.D. Freitas and N. Gordon, *Sequential Monte Carlo Methods*, Springer, New York, 2001.
[7]   J.A. Fernandez Pierna, F. Wahl, O.E. de Noord and D.L. Massart, Methods for outlier detection in prediction, *Chemometrics and Intelligent Laboratory Systems* **63** (2002), 27–39.
[8]   I. Fine and L. Shvartsman, Method for non-invasive optical measurements of blood parameters, United States Patent 6587704, July 1, 2003.
[9]   M.I. Griep, I.N. Wakeling, P. Vankeerberghen and D.L. Massart, Comparison of semirobust and robust partial least squares procedures, *Chemometrics and Intelligent Laboratory Systems* **29** (1995), 37–50.
[10] H.M. Heise, A. Bittner and R. Marbach, Clinical chemistry and near infrared spectroscopy: technology for non-invasive glucose monitoring, *Journal of Near Infrared Spectroscopy* **6** (1998), 349–359.
[11] A. Hoskuldsson, PLS regression methods, *Journal of Chemometrics* **2** (1988), 211–228.
[12] S. Kasemsumran, Y.P. Du, K. Maruo and Y. Ozaki, Improvement of partial least squares models for *in vitro* and *in vivo* glucose quantifications by using near-infrared spectroscopy and searching combination moving window partial least squares, *Chemometrics and Intelligent Laboratory Systems* **82** (2006), 97–103.
[13] K. Maruo, M. Tsurugi, J. Chin, T. Ota, H. Arimoto and Y. Yamada, Noninvasive blood glucose assay using a newly developed near-infrared system, *Journal of Selected Topics in Quantum Electronics* **9** (2003), 322–330.
[14] C.R. Mittermayr, H.W. Tan and S.D. Brown, Robust calibration with respect to background variation, *Applied Spectroscopy* **55** (2001), 827–833.
[15] G. Paul and R.K. Bruce, Partial least-squares regression: a tutorial, *Analytica Chimica Acta* **185** (1986), 1–17.

[16] R.J. Pell, Multiple outlier detection for multivariate calibration using robust statistical techniques, *Chemometrics and Intelligent Laboratory Systems* **52** (2000), 87–104.

[17] R. Raghavachari, *Near-Infrared Applications in Biotechnology*, Marcel Dekker, New York, 2001.

[18] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, New York, 2004.

[19] M.R. Robinson, R.P. Eaton, D.M. Haaland, G.W. Koepp, E.V. Thomas, B.R. Stallard et al., Noninvasive glucose monitoring in diabetic patients: a preliminary evaluation, *Clinical Chemistry* **38** (1992), 1618–1622.

[20] L.G. Weyer and S.C. Lo, Spectra-structure correlations in the near-infrared, in: *Handbook of Vibrational Spectroscopy*, Vol. 3, Wiley, Chichester, UK, 2002, pp. 1817–1837.

[21] WHO, Screening for type 2 diabetes, Report of a World Health Organization and International Diabetes Federation meeting, World Health Organization, Geneva, 2003, available at: http://www.who.int/diabetes/publications/en/screening_mnc03.pdf.

[22] WHO, Diabetes programme: fact sheets N°312, World Health Organization, Geneva, 2011, available at: http://www.who.int/mediacentre/factsheets/fs312/en/.

[23] S. Wild, G. Roglic, R. Sicree, H. King and A. Green, Global prevalence of diabetes: estimates for the year 2000 and projections for 2030, *Diabetes Care* **27** (2004), 1047–1053.

[24] J. Workman, *Handbook of Organic Compounds: NIR, IR, Raman, and UV-Vis Spectra Featuring Polymers and Surfactants*, Academic Press, San Diego, CA, 2001.