Journal of Innovative Optical Health Sciences
Vol. 7, No. 4 (2014) 1350060 (9 pages)
© The Authors
DOI: 10.1142/S1793545813500600

World Scientific
www.worldscientific.com

J. Innov. Opt. Health Sci. 2014.07. Downloaded from www.worldscientific.com
by 103.240.126.9 on 10/21/18. Re-use and distribution is strictly not permitted, except for Open Access articles.

# Quantification of glycated hemoglobin indicator HbA1c through near-infrared spectroscopy

Tao Pan*, Minmiao Li, Jiemei Chen† and Haiyan Xue
*Key Laboratory of Optoelectronic Information and Sensing Technologies of Guangdong Higher Educational Institutes*
*Jinan University, Guangzhou 510632, P. R. China*
**tpan@jnu.edu.cn*
†*tchjm@jnu.edu.cn*

Received 31 July 2013
Accepted 23 September 2013
Published 18 November 2013

A new strategy for quantitative analysis of a major clinical biochemical indicator called glycated hemoglobin (HbA1c) was proposed. The technique was based on the simultaneous near-infrared (NIR) spectral determination of hemoglobin (Hb) and absolute HbA1c content (Hb • HbA1c) in human hemolysate samples. Wavelength selections were accomplished using the improved moving window partial least square (MWPLS) method for stability. Each model was established using an approach based on randomness, similarity, and stability to obtain objective, stable, and practical models. The optimal wavebands obtained using MWPLS were 958 to 1036 nm for Hb and 1492 to 1858 nm for Hb • HbA1c, which were within the NIR overtone region. The validation root mean square error and validation correlation coefficients of prediction ($V$-SEP, $V$-$R_{\mathrm{P}}$) were $3.4\,\mathrm{g\,L^{-1}}$ and 0.967 for Hb, respectively, whereas the corresponding values for Hb • HbA1c were $0.63\,\mathrm{g\,L^{-1}}$ and 0.913. The corresponding $V$-SEP and $V$-$R_{\mathrm{P}}$ were 0.40% and 0.829 for the relative percentage of HbA1c. The experimental results confirm the feasibility for the quantification of HbA1c based on simultaneous NIR spectroscopic analyses of Hb and Hb • HbA1c.

*Keywords*: Glycated hemoglobin; HbA1c; NIR spectroscopy; wavelength selection stability.

## 1. Introduction

The incidence of diabetes shows an increasing trend, and has become a serious threat to human health. Thus, improving the accuracy in the diagnosis of diabetes and strengthening prevention measures are significant. In traditional diagnoses and therapeutic monitoring of diabetes, various parameters, such as fasting blood glucose, postprandial blood glucose and oral glucose tolerance, are determined. However, the measured values of these parameters only correspond to the instantaneous blood glucose level. Glycated hemoglobin (Hb), which is an effective

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 3.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

J. Innov. Opt. Health Sci. 2014.07. Downloaded from www.worldscientific.com by 103.240.126.9 on 10/21/18. Re-use and distribution is strictly not permitted, except for Open Access articles.

index of long-term blood glucose level, has been widely used in the diagnosis of diabetes mellitus, as well as in evaluating the effects of therapy on diabetes mellitus. Glycated Hb comprises HbA1 and other Hb-glucose adducts. HbA1 is made up of HbA1a, HbA1b, and HbA1c. HbA1c, which is the major component of HbA1, is formed by a nonenzymatic irreversible process of combination of the aldehyde group of glucose with the amino-terminal valine of the $\beta$-chain of Hb. This process comprises a sequence of nonenzymatic reactions, known as Maillard reactions.[1] HbA1c is a major clinical biochemical indicator of glycated Hb. The clinical value of HbA1c is expressed as a unit of relative percentage, which is equivalent to the ratio of the absolute content of HbA1c and the amount of Hb. In clinical practice, HbA1c is referenced to a nondiabetic range of 4.0% to 6.0% [mean = 5.0%, standard deviation (SD) = 0.5%]. Phenotype-positive patients for diabetes mellitus are those with HbA1c >6.0%.[2]

Near-infrared (NIR) spectroscopy mainly reflects absorption of the overtone and combination of vibrations of a functional group $X$-H (such as C–H, O–H, and N–H). This rapid simple technique is commonly used in the fields of agriculture,[3] food,[4] environment,[5] medicine,[6] and pharmaceuticals,[7] among others.[8]

NIR spectroscopy has been extensively applied in biomedicine. Computational NIR formulations for predicting and quantifying injuries and diseases have become prevalent. Previous studies[9,10] involved the use of very high-fidelity two-dimensional (2D) and three-dimensional (3D) simulations to accurately and efficiently predict and quantify local and global injuries for organs such as the brain and lungs. The researchers were able to (i) noninvasively "numerically penetrate" the tissues and (ii) reconstruct the optical properties of oxygenated and deoxygenated blood in the presence of water. These numerical noninvasive measurements were used to predict the extent and severity of organ hemorrhage/injury because the use of the traditional method for clinical determination is difficult. NIR has significant potential for clinical diagnosis.

To the best of our knowledge, a quantification method for screening preliminary diabetes indicators (i.e., HbA1c) using NIR spectroscopy has not been developed. Hb is a macromolecule that contains various $X$-H functional groups, which have significant absorption in the NIR region. Chemical-free and rapid analysis of Hb using NIR spectroscopy has been the focus of previous studies.[11–13] The Hb glycation process (Maillard reactions) involves some functional groups that contain hydrogen; thus, NIR spectroscopy can be used to obtain information about glycated Hb.

However, as a relative percentage of total Hb, HbA1c and spectral absorbance do not fulfill Beer's law because the absolute content (Hb • HbA1c) of the samples with the same relative percentage (HbA1c) may vary when the amount of total Hb is different. The spectral absorption caused by molecular vibration also varies. Therefore, the relative percentage of HbA1c cannot be measured directly using NIR spectroscopy.

Indirect measurement of HbA1c is considered. Given that Hb levels can be measured by NIR spectroscopy,[11–13] we assume that Hb • HbA1c can also be measured using the same method. With the simultaneous measurement of the two indicators, the predictive value of HbA1c can be obtained. In this study, we performed experiments to confirm the feasibility of simultaneous quantitative analysis of Hb • HbA1c and Hb using NIR spectroscopy.

Partial least square (PLS) regression analysis is used for comprehensive screening of spectroscopic data, extracting information variables, and overcoming spectral co-linearity. However, waveband selection is necessary because improving the prediction capability of PLS is difficult when the signal-to-noise ratio (SNR) of the waveband is not adequately high. Human blood is a complex system with multiple components. The spectroscopic analysis of a single blood component requires mitigation of the interference of other components and noise. Appropriate wavelength selection is an important albeit difficult, technical aspect. Improving the effectiveness of spectral prediction, reducing method complexity, and designing specialized spectrometers with high SNR are important. Therefore, appropriate chemometric methods are necessary for wavelength optimization. Moving window PLS (MWPLS) is an effective method for waveband selection.[14–16]

Differences in the partitioning of calibration and prediction sets can result in fluctuations in the predictions and parameters, thus yielding unstable results. In this study, a portion of samples was randomly selected as a validation set, which was not subjected to the modeling optimization process. The remaining samples were used as the modeling set. Based on the varied partitioning of the calibration

J. Innov. Opt. Health Sci. 2014.07. Downloaded from www.worldscientific.com by 103.240.126.9 on 10/21/18. Re-use and distribution is strictly not permitted, except for Open Access articles.

and prediction sets in the modeling set, the MWPLS method was improved by considering the stability. Using NIR spectroscopy with improved MWPLS method, simultaneous quantitative analysis of Hb and Hb • HbA1c with wavelength selection was performed. Thus, quantification of HbA1c was achieved. The divisions for the calibration and prediction sets were based on certain similarities to avoid evaluation distortion.

## 2. Materials and Methods

### 2.1. *Experimental materials, instruments, and measurement methods*

A total of 224 samples of human peripheral blood were collected and placed in 0.2% ethylenediamine tetraacetic acid-containing tubes. The Hb values of the samples were measured using a BC-3000Plus automatic blood cell analyzer (Shenzhen Mairui, China). The measured Hb values ranged from 77 to 156 g L$^{-1}$. The mean value and the SD were 131.2 and 14.3 g L$^{-1}$, respectively. The HbA1c values of the samples were measured by high-pressure liquid chromatography analysis method using an ADAMS$^{TM}$ A1c HA-8160 automatic glycated Hb analyzer (ARKRAY, Japan). The measured HbA1c values ranged from 4.6% to 10.8%, with mean value and SD of 6.40% and 0.84%, respectively. The obtained Hb • HbA1c values ranged from 4.2 to 15.0 g L$^{-1}$, with mean value and SD of 8.42 and 1.58 g L$^{-1}$, respectively. Based on the cut-off values of HbA1c (6.0%), 84 negative and 140 positive samples were obtained.

Scattering and noise disturbance may occur when light passes through the samples because the peripheral blood samples are highly viscous. Furthermore, Hb is contained within an erythrocyte, and the noise from the cell membrane should be overcome to determine Hb levels. Thus, the accuracy of spectral analysis of peripheral blood samples may decrease.

We compared the effects of peripheral blood and hemolysate samples on the prediction capability of the method in our previous work.[17] The peripheral blood samples were configured to 2×, 3×, 4×, 5×, and 6× dilute hemolytic solution samples. Six sample groups (including the group of peripheral blood samples) were obtained, and then calibration and prediction models for Hb were established in each group based on two wavebands (4000 to 600 cm$^{-1}$ and 1800 to 800 cm$^{-1}$). The results showed that the group with the peripheral blood samples had significantly low prediction accuracy, whereas the prediction for each group of hemolysate samples was close to each other. This result suggests that peripheral blood has high noise disturbance, so hemolysate samples were used. Given that some weak information corresponding to glycated Hb may be lost at low concentration, 2× dilute hemolytic solutions were adopted in this study.

The peripheral blood samples were diluted with distilled water to rupture the erythrocytes and obtain hemolysate samples. A volume of blood samples was mixed with an equal volume of distilled water. The samples were then used for spectrometric measurement.

The spectroscopy instrument used was an XDS Rapid Content$^{TM}$ Liquid Grating Spectrometer (FOSS, Denmark) equipped with a transmission accessory and a 2-mm cuvette. The scanning scope of the spectrum was 780 to 2498 nm with a 2-nm wavelength interval. The scanning scope included the overall NIR region. Wavebands of 780 to 1100 nm and 1100 to 2498 nm were employed for Si and PbS detection, respectively. Each sample (0.8 mL) was measured thrice, and the mean value of the three measurements was used for modeling. The spectra were obtained at 25 ± 1°C and 45 ± 1% relative humidity.

### 2.2. *Sample set division and model optimization frame*

Some of the samples were randomly selected as the validation set. The remaining samples were used as the modeling set. The modeling set was divided many times into similar calibration and prediction sets. Calibration and prediction were performed for each division $i$. The root mean square error for the calibration and prediction were denoted as $M$-SEC$_i$ and $M$-SEP$_i$, respectively, and their corresponding correlation coefficients were denoted as $M$-$R_{C,i}$ and $M$-$R_{P,i}$. Based on all divisions, the mean value and SD of the root mean square error and correlation coefficients for the prediction were further calculated and denoted as $M$-SEP$_{Ave}$, $M$-$R_{P,Ave}$, $M$-SEP$_{Std}$, and $M$-$R_{P,Std}$. These values were used to analyze the prediction accuracy and modeling stability. The equation $M{-}\mathrm{SEP}^{+} = M{-}\mathrm{SEP}_{Ave} +$

J. Innov. Opt. Health Sci. 2014.07. Downloaded from www.worldscientific.com by 103.240.126.9 on 10/21/18. Re-use and distribution is strictly not permitted, except for Open Access articles.

$M-\mathrm{SEP}_{\mathrm{Std}}$ was used as a comprehensive indicator of prediction accuracy and stability. Smaller $M$-$\mathrm{SEP}^{+}$ indicated higher accuracy and stability of the model. The model parameters were selected based on the minimum $M$-$\mathrm{SEP}^{+}$. The selected model was then revalidated against the validation set. The randomly selected validation samples, which were not subjected to the modeling optimization process, were regarded as the prediction set, whereas the original modeling set was used as calibration set. The validation root mean square error and validation correlation coefficients of prediction were then calculated and denoted as $V$-SEP and $V$-$R_P$, respectively.

The measured HbA1c value, which does not meet Beer's law, was the relative percentage of HbA1c to the total Hb. However, the absolute content (Hb • HbA1c) of HbA1c corresponded to the molecular structure and molecular vibration, which should be used in the spectroscopic analysis. Quantitative analyses of Hb and Hb • HbA1c were performed independently using the same modeling process. A total of 74 samples (29 negative samples and 45 positive samples) were randomly selected from a total of 224 samples; these selected samples were used as the validation set. The remaining 150 samples (55 negative samples and 95 positive samples) were used as the modeling set. The modeling set was divided into similar calibration (80 samples) and prediction (70 samples) sets 100 times.

When the division for the calibration and prediction sets is random, the modeling process may appear as a distorted evaluation because of contingency. For example, a randomly generated calibration set comprises samples with low Hb values, whereas the prediction set may comprise samples with high Hb values. Improving the prediction under such conditions is difficult, and often yields incorrect models. Given these considerations, the modeling set should be divided into a calibration set and a prediction set with certain similarity. In this study, the similarity of sample sets was defined using the measured values. When the mean value and SD of the measured values in the calibration set were close to those in the prediction set, the two sets could be regarded as similar. The mean values and SDs of the Hb values in the calibration, prediction, and entire modeling sets were denoted as $\mathrm{Hb}_{\mathrm{C,Ave}}$, $\mathrm{Hb}_{\mathrm{C,Std}}$, $\mathrm{Hb}_{\mathrm{P,Ave}}$, $\mathrm{Hb}_{\mathrm{P,Std}}$, $\mathrm{Hb}_{\mathrm{Ave}}$, and $\mathrm{Hb}_{\mathrm{Std}}$, respectively. The similarity degree was defined as follows:

$$\alpha = \max\left\{\frac{|\mathrm{Hb}_{\mathrm{C,Ave}} - \mathrm{Hb}_{\mathrm{P,Ave}}|}{\mathrm{Hb}_{\mathrm{Ave}}}, \frac{|\mathrm{Hb}_{\mathrm{C,Std}} - \mathrm{Hb}_{\mathrm{P,Std}}|}{\mathrm{Hb}_{\mathrm{Std}}}\right\} \times 100\%. \quad (1)$$

Similarity is high when the value of $\alpha$ is small. According to this definition, all of the modeling samples were randomly divided into calibration and prediction sets for a sufficient number of times using a computer program, and then 100 divisions that satisfy $\alpha < 5\%$ were retained for modeling.

### 2.3. *Optimization frame of the MWPLS method*

The spectrum of complex systems (such as blood samples) includes multiple absorptions of various components, which means that the spectral absorption that corresponds to each wavelength is not purely dependent on a single component. Achieving a desirable analytical effect using the spectral absorption peaks of some specific components for modeling is usually difficult because of too much interference. Conversely, the use of statistics and chemometric methods, according to the prediction capability, is a convenient method in determining the modeling analytical waveband. This method may not determine the absorption peaks of relevant components, but it can obtain an appropriate analytical waveband for the combined absorption of all components.

In this study, the modeling process was based on MWPLS method, which was used to search for an analytical waveband, instead of spectral absorption peaks for modeling. In searching for the analytical waveband, we traversed all of the wavebands and compared their prediction effectiveness to obtain the optimal waveband. The method fully considered the absorbance of other compounds in finding the most suitable analytical waveband for modeling.

In the MWPLS method, consecutive spectral data on $N$ adjacent wavelengths were designated as windows. By moving the window and changing its size, PLS models of all of the windows in the entire spectral collecting region were established, and the optimal analytical wavebands were selected. Given the position and length of the wavebands and the PLS factor, the search parameters were set as follows: starting wavelength and its serial number ($B$),

J. Innov. Opt. Health Sci. 2014.07. Downloaded from www.worldscientific.com by 103.240.126.9 on 10/21/18. Re-use and distribution is strictly not permitted, except for Open Access articles.

number of wavelengths ($N$), and number of PLS factors ($F$).[18–20] The range of parameters $B$, $N$, and $F$ can be set based on the actual condition. The PLS models were established for all of the combinations ($B$, $N$, and $F$). The corresponding $M$-SEP$_{\mathrm{Ave}}$, $M$-$R_{\mathrm{P,Ave}}$, $M$-SEP$_{\mathrm{Std}}$, $M$-$R_{\mathrm{P,Std}}$, and $M$-SEP$^{+}$ were then calculated.

PLS can comprehensively screen spectroscopic data and extract information variables. The number of PLS factors $F$, which corresponds to the number of integrated variables of the sample, is a major parameter. The selection of a suitable $F$ is both necessary and difficult.[21,22] In this study, $F$ was selected by considering the number of divisions for the calibration and prediction sets. Thus, the optimized $F$ exhibited stability and practicality. All wavebands corresponded to a unique combination of parameters ($B$ and $N$). The optimal $F$ of the corresponding PLS model was selected according to the minimum $M$-SEP$^{+}$. The global optimal model was further selected by comparing the minimum $M$-SEP$^{+}$ of all wavebands.

Instrument design typically involves some restrictions, such as costs and material properties, as well as position and number of wavelengths. In some instances, the demand of actual conditions cannot be met by the global optimal waveband. Therefore, local optimal wavebands that correspond to different positions and number of wavelengths are necessary. In evaluating the predictive effects of different starting wavelengths, $B$ was fixed and the other parameters ($N$ and $F$) were arbitrarily changed. The local optimal model corresponding to $B$ was selected based on the minimum $M$-SEP$^{+}$. In evaluating the predictive effects of different numbers of wavelength, $N$ was fixed and the other parameters ($B$ and $F$) were arbitrarily changed. The local optimal model corresponding to $N$ was also selected based on the minimum $M$-SEP$^{+}$.

The search range for the MWPLS method covered the overall scanning region from 780 to 2498 nm with 860 wavelengths. $B$ was set as $B \in \{1, 2, \ldots, 860\}$, and $F$ was set as $F \in \{1, 2, \ldots, 20\}$. To reduce the workload and maintain representativeness, $N$ were set as follows:

$$N \in \{1, 2, \ldots, 100\} \cup \{110, 112, \ldots, 200\} \cup \{210, 220, \ldots, 500\} \cup \{520, 540, \ldots, 860\}.$$

The computer platform was constructed using MATLAB 7.6 software.

## 3. Results and Discussion

### 3.1. *Waveband optimization using the MWPLS method*

The NIR spectra of the 224 human hemolysate samples on the overall scanning region (780 to 2498 nm) are shown in Fig. 1. The PLS models for Hb and Hb • HbA1c were first established based on the overall scanning region (780 to 2498 nm). The results of the prediction accuracy and stability are summarized in Table 1. The results show that the predicted values and clinically measured values obtained from the 100 divisions were highly correlated for Hb, but not for Hb • HbA1c. Moreover, the number of adopted wavelengths was 860, and the model had high complexity. To improve the prediction accuracy and reduce the complexity, waveband optimization was further performed using the MWPLS method. Based on the minimum $M$-SEP$^{+}$, the optimal MWPLS models for Hb and Hb • HbA1c were selected. The corresponding parameters $B$, $N$, and $F$ and the prediction effects $M$-SEP$_{\mathrm{Ave}}$, $M$-$R_{\mathrm{P,Ave}}$, $M$-SEP$_{\mathrm{Std}}$, $M$-$R_{\mathrm{P,Std}}$, and $M$-SEP$^{+}$ are summarized in Table 2. The results show that the optimal starting wavelength and $N$ were respectively 958 nm and 40 for Hb, and 1492 nm and 184 for Hb • HbA1c. The corresponding wavebands were 958 to 1036 nm for Hb, and 1492 to 1858 nm for Hb • HbA1c, which were within the NIR overtone region. The model complexity significantly decreased. Tables 1 and 2 show that the $M$-SEP$_{\mathrm{Ave}}$ and $M$-SEP$^{+}$ of the optimal MWPLS models were significantly lower than those of the overall scanning region for the two indicators. Thus, the prediction accuracy and stability of the optimal

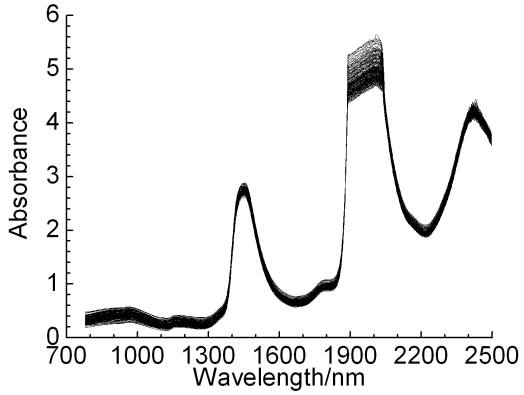


Fig. 1. NIR spectra of 224 human hemolysate samples in overall scanning region (780 to 2498 nm).

Table 1. Prediction accuracy and stability of PLS models based on the overall scanning region for Hb and Hb • HbA1c.

| Indicator | Waveband (nm) | $N$ | $F$ | $M$-SEP$_{\mathrm{Ave}}$ (g L$^{-1}$) | $M$-SEP$_{\mathrm{Std}}$ (g L$^{-1}$) | $M$-$R_{\mathrm{P,Ave}}$ | $M$-$R_{\mathrm{P,Std}}$ | $M$-SEP$^{+}$ (g L$^{-1}$) |
|---|---|---|---|---|---|---|---|---|
| Hb | 780–2498 | 860 | 6 | 4.3 | 0.4 | 0.961 | 0.007 | 4.7 |
| Hb • HbA1c | 780–2498 | 860 | 8 | 1.29 | 0.09 | 0.669 | 0.051 | 1.38 |

Table 2. Prediction accuracy and stability of PLS models based on the optimal MWPLS wavebands for Hb and Hb • HbA1c.

| Indicator | Waveband (nm) | $N$ | $F$ | $M$-SEP$_{\mathrm{Ave}}$ (g L$^{-1}$) | $M$-SEP$_{\mathrm{Std}}$ (g L$^{-1}$) | $M$-$R_{\mathrm{P,Ave}}$ | $M$-$R_{\mathrm{P,Std}}$ | $M$-SEP$^{+}$ (g L$^{-1}$) |
|---|---|---|---|---|---|---|---|---|
| Hb | 958–1036 | 40 | 11 | 3.7 | 0.3 | 0.971 | 0.005 | 4.0 |
| Hb • HbA1c | 1492–1858 | 184 | 7 | 0.64 | 0.07 | 0.931 | 0.015 | 0.71 |

MWPLS models were significantly improved in both indicators, specifically for Hb • HbA1c.

The optimal $M$-SEP$^{+}$, which corresponds to each starting wavelength and $N$ for the two indicators, are shown in Figs. 2 to 5. For the Hb indicator, Figs. 2 and 3 show the minimum $M$-SEP$^{+}$ achieved for the starting wavelength of 958 nm and $N = 40$. For the Hb • HbA1c indicator, Figs. 4 and 5 show the minimum $M$-SEP$^{+}$ achieved for the starting wavelength of 1492 nm and $N = 184$. The results indicate that these parameters had the best prediction accuracy and stability. These data may serve as valuable reference for designing the splitting system of spectroscopic instruments. Some local optimal models whose predictions are close to those of the global optimal model are still a good choice. These models address the restrictions, such as costs and material properties, as well as the position and number of wavelengths in instrument design.

## 3.2. *Model validation*

The randomly selected validation samples, which were excluded in the modeling optimization process, were used to validate the optimal MWPLS models (958 to 1036 nm for Hb, and 1492 to 1858 nm for

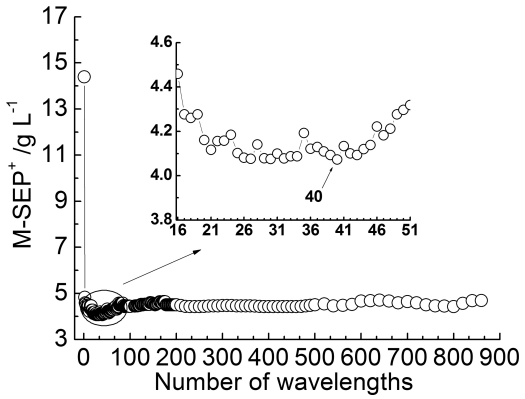


Fig. 3. Optimal $M$-SEP$^{+}$ values corresponding to each number of wavelength for Hb.

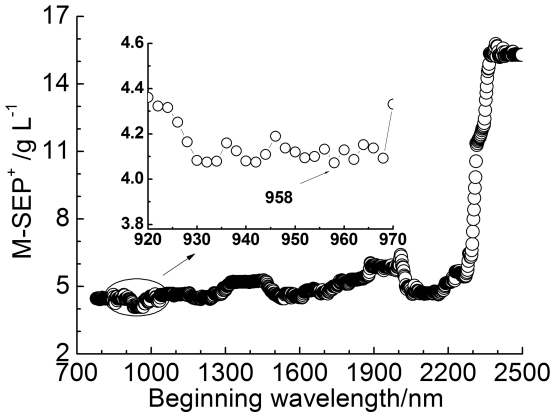


Fig. 2. Optimal $M$-SEP$^{+}$ values corresponding to each starting wavelength for Hb.

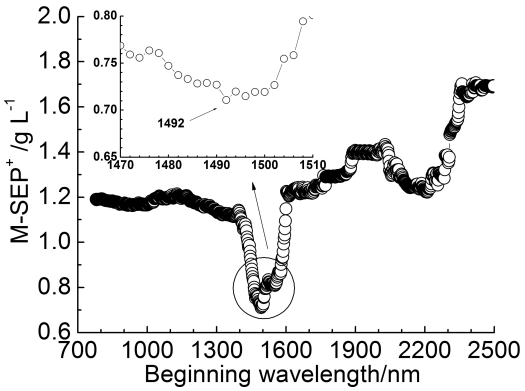


Fig. 4. Optimal $M$-SEP$^{+}$ values corresponding to each starting wavelength for Hb • HbA1c.



J. Innov. Opt. Health Sci. 2014.07. Downloaded from www.worldscientific.com by 103.240.126.9 on 10/21/18. Re-use and distribution is strictly not permitted, except for Open Access articles.

J. Innov. Opt. Health Sci. 2014.07. Downloaded from www.worldscientific.com by 103.240.126.9 on 10/21/18. Re-use and distribution is strictly not permitted, except for Open Access articles.

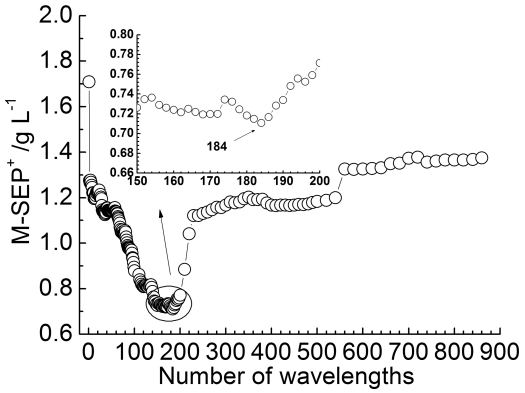


Fig. 5. Optimal $M$-SEP$^+$ values corresponding to each number of wavelength for Hb • HbA1c.

Hb • HbA1c). The PLS regression coefficients were calculated using the spectral data and clinically measured values in the modeling set based on the corresponding parameters. The predicted values of the validation samples were then calculated using the obtained regression coefficients and spectra of the validation samples.

Figures 6 and 7 show the relationship between the predicted and clinically measured values of 74 validation samples for Hb and Hb • HbA1c. Table 3 shows the validation effects ($V$-SEP and $V$-$R_{\mathrm{P}}$). The results indicate that the two models had high validation effects. Hb and Hb • HbA1c prediction values of the samples are close to those of the clinically measured values. Satisfactory validation effects were achieved for the random validation

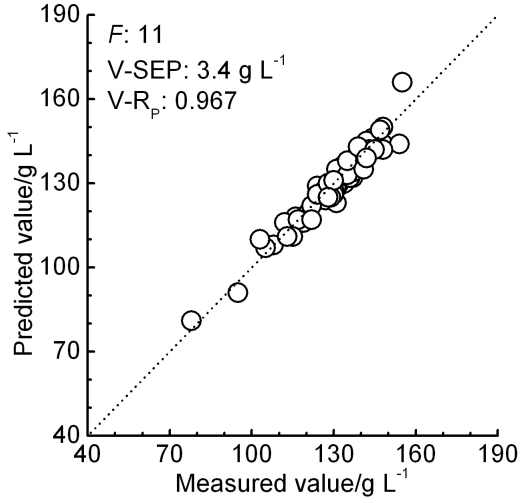


Fig. 6. Relationship between the predicted and the measured values of the validation samples at 958 to 1036 nm for Hb.

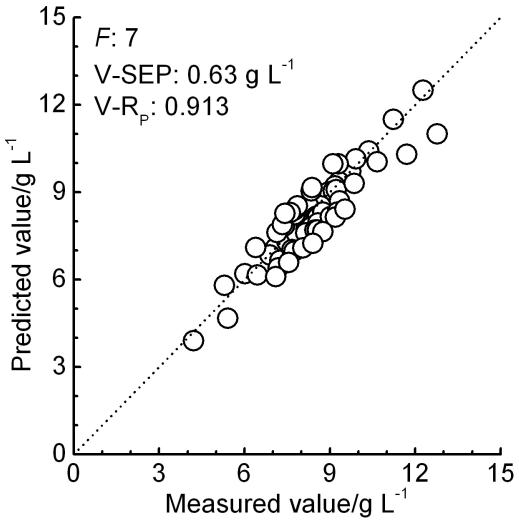


Fig. 7. Relationship between the predicted and the measured values of the validation samples at 1492 to 1858 nm for Hb • HbA1c.

samples because stability was considered in the modeling optimization process.

The predicted values of HbA1c was further calculated based on the predicted Hb and Hb • HbA1c values. The corresponding $V$-SEP and $V$-$R_{\mathrm{P}}$ were 0.40% and 0.829 for HbA1c. The relationship between the predicted and clinically measured values of the 74 validation samples for HbA1c is shown in Fig. 8. The results show that the predicted values and clinically measured values were also highly correlated for HbA1c. The experimental results confirm the feasibility for the quantitative analysis of HbA1c based on the simultaneous NIR spectroscopic analyses of Hb and Hb • HbA1c.

HbA1c is the main preliminary screening indicator for diabetes mellitus. The classification of the negative and positive samples can be observed using the 2D diagram of Hb and HbA1c. Among the 74 validation samples, 29 are negative and 45 are positive. Figure 9 shows the corresponding 2D diagrams for the predicted values of the 74 validation samples. The corresponding specificity and sensitivity were 90% and 96%, respectively.

Table 3. Validation effects of the optimal MWPLS models for Hb and Hb • HbA1c.

| Indicator | Waveband (nm) | $N$ | $F$ | $V$-SEP (g L$^{-1}$) | $V$-$R_{\mathrm{P}}$ |
|---|---|---|---|---|---|
| Hb | 958–1036 | 40 | 11 | 3.4 | 0.967 |
| Hb • HbA1c | 1492–1858 | 184 | 7 | 0.63 | 0.913 |

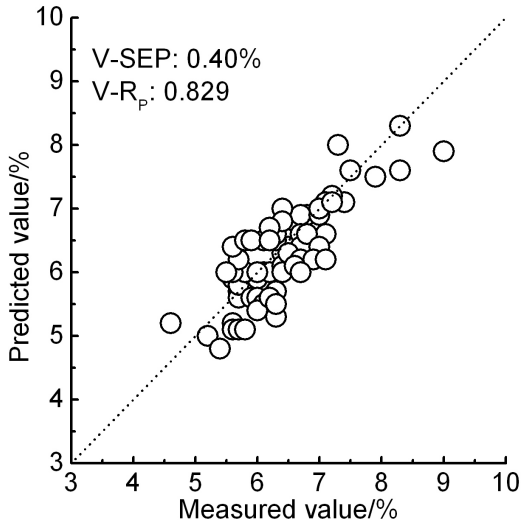


Fig. 8. Relationship between the predicted and the measured values of the validation samples for HbA1c.

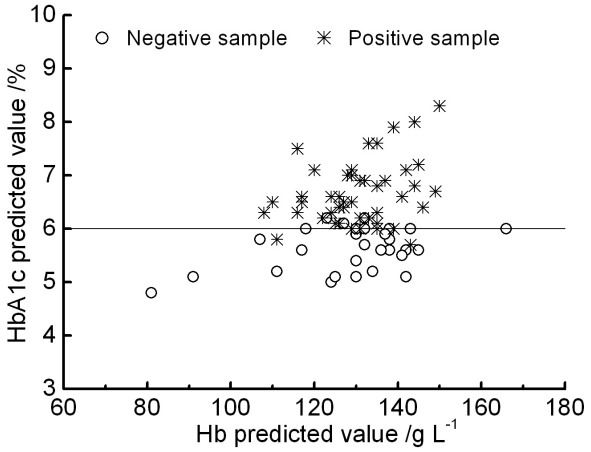


Fig. 9. The 2D diagrams (Hb and HbA1c) for the negative and positive NIR predicted values of validation samples.

The prediction errors were distributed mainly around the cut-off line (HbA1c = 6%). The neighboring region of the cut-off line is blurry, and a few prediction errors in this region are understandable. The results also showed the feasibility for screening samples that are negative and positive for diabetes mellitus.

## 4. Conclusion

A rapid simultaneous determination method for Hb and Hb • HbA1c based on NIR spectroscopy was developed. Quantification of the relative percentage HbA1c was achieved using this method. Wavelength selections were accomplished using the improved MWPLS method for stability. Each model was established using a modeling approach based on randomness, similarity, and stability to obtain objective, stable, and practical models.

The optimal wavebands screened using MWPLS were 958 to 1036 nm for Hb and 1492 to 1858 nm for Hb • HbA1c, which were within the NIR overtone region. $V$-SEP and $V$-$R_P$ were 3.4 g L$^{-1}$ and 0.967 for Hb, respectively, and the corresponding values for Hb • HbA1c were 0.63 g L$^{-1}$ and 0.913. The two models achieved high validation effects. The predicted values of HbA1c were calculated based on the spectral predicted Hb and Hb • HbA1c values. The corresponding $V$-SEP and $V$-$R_P$ were 0.40% and 0.829. The results show that the predicted values and the clinically measured values were also highly correlated for HbA1c. The experimental results confirm the feasibility for the quantitative analysis of HbA1c based on simultaneous NIR spectroscopic analyses of Hb and Hb • HbA1c. This study also provides valuable references for designing specialized spectrometers.

## Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 61078040) and the Science and Technology Project of Guangdong Province (No. 2012B031800917).

## References

1. O. S. Zhernovaya, V. V. Tuchin, I. V. Meglinski, "Monitoring of blood proteins glycation by refractive index and spectral measurements," *Laser Phys. Lett.* **6**, 460–464 (2008).
2. American Diabetes Association, "Standards of medical care for patients with diabetes mellitus," *Diabetes Care* **26**, 33–50 (2003).
3. R. Welle, W. Greten, B. Rietmann, S. Alley, G. Sinnaeve, P. Dardenne, "Near-infrared spectroscopy on chopper to measure maize forage quality parameters online," *Crop Sci.* **43**, 1407–1413 (2003).
4. J. Y. Chen, H. Zhang, R. Matsunaga, "Rapid determination of the main organic acid composition of raw Japanese apricot fruit juices using near-infrared spectroscopy," *J. Agric. Food Chem.* **54**, 9652–9657 (2006).
5. A. C. Sousa, M. M. L. M. Lucio, O. F. Bezerra, G. P. S. Marcone, A. F. C. Pereira, E. O. Dantas, W. D. Fragoso, M. C. U. Araujo, R. K. H. Galvao,



J. Innov. Opt. Health Sci. 2014.07. Downloaded from www.worldscientific.com by 103.240.126.9 on 10/21/18. Re-use and distribution is strictly not permitted, except for Open Access articles.

J. Innov. Opt. Health Sci. 2014.07. Downloaded from www.worldscientific.com by 103.240.126.9 on 10/21/18. Re-use and distribution is strictly not permitted, except for Open Access articles.

"A method for determination of COD in a domestic wastewater treatment plant by using near-infrared reflectance spectrometry of seston," *Anal. Chim. Acta* **588**, 231–236 (2007),

6. K. H. Hazen, M. A. Arnold, G. W. Small, "Measurement of glucose and other analytes in undiluted human serum with near-infrared transmission spectroscopy," *Anal. Chim. Acta* **371**, 255–267 (1998).
7. R. W. Bondi, B. Igne, J. K. Drennen, "Effect of experimental design on the prediction performance of calibration models based on near-infrared spectroscopy for pharmaceutical applications," *Appl. Spectrosc.* **66**, 1442–1453 (2012).
8. Y. Ozaki, "Near-infrared spectroscopy-its versatility in analytical chemistry," *Anal. Sci.* **28**, 545–563 (2012).
9. R. Kannan, A. J. Przekwas, "A near-infrared spectroscopy computational model for cerebral hemodynamics," *Int. J. Numer. Methods Biomed. Eng.* **28**, 1093–1106 (2012).
10. R. Kannan, A. J. Przekwas, "A computational model to detect and quantify a primary blast lung injury using near-infrared optical tomography," *Int. J. Numer. Methods Biomed. Eng.* **27**, 13–28 (2011).
11. V. N. Istvan, J. K. Karoly, M. J. Janos, G. Éva, D. Gyula, "Application of near infrared spectroscopy to the determination of haemoglobin," *Clin. Chem. Acta* **264**, 117–125 (1997).
12. Y. Lee, S. Lee, J. Y. In, S. H. Chung, J. H. Yon, "Prediction of plasma hemoglobin concentration by near-infrared spectroscopy," *J. Korean Med. Sci.* **23**, 674–677 (2008).
13. X. Q. Shan, L. G. Chen, Y. Yuan, C. S. Liu, X. L. Zhang, Y. Sheng, F. Xu, "Quantitative analysis of hemoglobin content in polymeric nanoparticles as blood substitutes using Fourier transform infrared spectroscopy," *J. Mater. Sci.* **21**, 241–249 (2010).
14. J. H. Jiang, R. J. Berry, H. W. Siesler, Y. Ozaki, "Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data," *Anal. Chem.* **74**, 3555–3565 (2002).
15. Y. P. Du, Y. Z. Liang, J. H. Jiang, R. J. Berry, Y. Ozaki, "Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares," *Anal. Chim. Acta* **501**, 183–191 (2004).
16. H. Z. Chen, T. Pan, J. M. Chen, Q. P. Lu, "Waveband selection for NIR spectroscopy analysis of soil organic matter based on SG smoothing and MWPLS methods," *Chemometr. Intell. Lab. Syst.* **107**, 139–146 (2011).
17. H. Yin, T. Pan, P. L. Tian, Y. Han, X. C. Wei, Q. J. Zhang, J. Y. Fang, Q. Zhong, H. Feng, "The rapid quantitative analysis for the human blood hemoglobin applied through the FTIR/ATR spectrum," *Chin. J. Spectroscopy Laboratory* **26**, 431–436 (2009).
18. T. Pan, Z. H. Chen, J. M. Chen, Z. Y. Liu, "Near-infrared spectroscopy with waveband selection stability for the determination of COD in sugar refinery wastewater," *Anal. Methods* **4**, 1046–1052 (2012).
19. T. Pan, Z. T. Wu, J. M. Chen, "Waveband optimization for near-infrared spectroscopic analysis of total nitrogen in soil," *Chinese J. Anal. Chem.* **40**, 920–924 (2012).
20. Z. Y. Liu, B. Liu, T. Pan, J. D. Yang, "Determination of amino acid nitrogen in tuber mustard using near-infrared spectroscopy with waveband selection stability," *Spectrochim. Acta A* **102**, 269–274 (2013).
21. J. Xie, T. Pan, J. M. Chen, H. Z. Chen, X. H. Ren, "Joint optimization of savitzky-Golay smoothing models and partial least squares factors for near-infrared spectroscopic analysis of serum glucose," *Chin. J. Anal. Chem.* **38**, 342–346 (2010).
22. T. Pan, J. M. Liu, J. M. Chen, G. P. Zhang, Y. Zhao, "Rapid determination of preliminary thalassaemia screening indicators based on near-infrared spectroscopy with wavelength selection stability," *Anal. Methods* **5**, 4355–4362 (2013).