Net analyte signal with floating reference theory in non-invasive blood glucose sensing by near-infrared spectroscopy

Wanjie Zhang (张婉洁), Rong Liu (刘 蓉)*, Wen Zhang (张 雯), Jiaxiang Zheng (郑加祥), and Kexin Xu (徐可欣)

State Key Laboratory of Precision Measuring Technology and Instrument, Tianjin University, Tianjin 300072, China

**Corresponding author: rongliu@tju.edu.cn*

Received December 19, 2011; accepted March 14, 2012; posted online June 15, 2012

Based on the floating reference theory, a new method for extracting the net analyte signal (NAS) is proposed. The noise background subspace is spanned by spectra at the floating radial reference point, and then, the spectra at the measurement point are projected on the subspace. Thereafter, the glucose concentrations in intralipid solutions are investigated through Monte Carlo simulation and experiments, and the partial least squares (PLS) models with and without NAS analysis are built. The root mean square errors of calibration and prediction reach to 28.87% and 27.33%, respectively. The results confirm the existence of information induced by glucose concentration variations as well as the validity of the floating reference theory.

OCIS codes: 300.0300, 070.0070.
doi: 10.3788/COL201210.083002.

With the improvement of people's living standard and quality of life, diabetes mellitus has become a serious health problem around the world. Self-monitoring and tight glycemic control are recognized as the primary goals in the management of diabetes to reduce the incidence of complications. Therefore, many diabetics often have to measure their blood glucose concentration via an invasive method, such as using a portable glucose meter, which is painful and has the risk of infection. Transcutaneous detection of blood glucose will thus make diabetic patients' lives better.

Near-infrared (NIR) spectroscopy is fast and does not need medical disposable materials. It has been proposed as a valid means for noninvasively measuring *in vivo* blood glucose. NIR spectroscopy probes the overtone and combined vibrational transitions of the C-H and O-H of glucose molecules. Combined with *a priori* knowledge of the spectra and analyte concentration, multivariate calibration methods such as partial least squares (PLS) regression or principal component regression are applied to build models for determining the glucose concentrations of unknown samples.

To date, many groups have reported encouraging results for glucose predictions via NIR spectroscopy[1−3]. However, the glucose level in human tissues is low, and changes in the physiological range are very small. Moreover, vibrational transition bands are typically broad, leading to highly overlapping bands between the glucose and other interferents, such as water, fat, and proteins. The glucose signal of human skin in the NIR spectra is very weak compared with that from other skin tissue components or body temperature change. Thus, chance temporal correlation may occur easily from unreasonable experimental and analytical processes[4,5]. In addition, the most widely used PLS regression algorithm, which is designed to accentuate all signals, may result in functional models based on some chance correlation factors. Moreover, these models may not distinguish or enhance glucose-specific spectral variations from complex spectra. Hence, proving that glucose concentration predictions are based on the spectral features of glucose rather than on spurious correlations is difficult[6].

For chance correlations caused by time-dependent sampling, Arnold *et al.*[7] used a phantom model to demonstrate that non-invasive human NIR spectra and calibration models could not predict glucose values when the glucose assignments were random. However, apparently, functional models are obtained when glucose assignments are made in a nonrandom, time-dependent manner. The net analyte signal (NAS) calibration vectors were compared with PLS calibration vectors, and the chance temporal correlations between assigned glucose concentrations and some uncontrolled experimental parameters were confirmed[8]. Similar experimental validations via Raman spectroscopy have been conducted[9,10]. In Ref. [11], random sampling, background correction, and a careful experimental design could be used to remove/minimize chance correlation.

Our group investigated the diffuse reflectance light intensity change based on glucose concentration variations in specific radial positions, and the results showed two particular radial points: the reference and measurement points. The floating-reference theory was then proposed[12−14], as shown in Fig. 1. The floating-reference point is insensitive to glucose concentration variations, and its existence has already been demonstrated by Yang *et al.*[14]. On the other hand, the spectra at the measurement point have maximum sensitivity. The existence of the floating-reference point has been demonstrated via Monte Carlo simulations of intralipid solution, skin models, and a series of *in vitro* experiments.

NAS was first proposed by Lorber[15] in 1986 as a part of the pure component spectrum orthogonal to all other

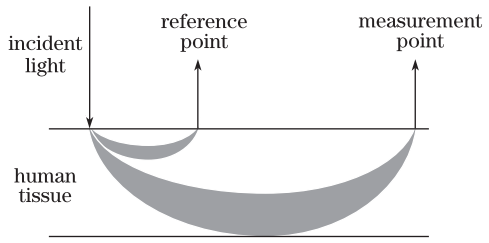

Fig. 1. Diagram of the floating-reference.

coexisting constituents. The key step in NAS analysis is to find a background subspace. Thus, the quality of the background subspace constructed from the noise signal determines the accuracy of the analyte signal extraction[16]. In the system based on the floating-reference theory, the spectra measured at the reference and measurement points cover the same interferences. However, no glucose concentration variation is included at the reference point. Therefore, the spectra matrix at the reference point may act as the background subspace to reduce the influence of interference on glucose determination.

In this letter, the spectra matrix at the reference point is used to substitute for the projection matrix to calculate the NAS

$$\mathbf{r}^{\mathrm{net}} = (\mathbf{I} - \hat{\mathbf{R}}_{\mathrm{F}}\hat{\mathbf{R}}_{\mathrm{F}}^{+})\hat{\mathbf{r}}, \tag{1}$$

where $\hat{\mathbf{r}}$ is the measured spectra after preprocessing at J wavelengths, $\hat{\mathbf{R}}_{\mathbf{F}}$ is the reconstructed background matrix, the superscript "+" symbolizes the Moore–Penrose pseudoinverse, and $\mathbf{I}$ is the identity matrix.

$\mathbf{r}^{\mathrm{net}}$ is free from interferences. Thus, it can be replaced by a scalar representation, e.g., its Euclidean norm, without loss of information[17]. The Euclidean norm of NAS nasnorm($\mathbf{r}^{\mathrm{net}}$) is proportional to the concentration of component c

$$\mathrm{nasnorm}(\mathbf{r}^{\mathrm{net}}) = k_1 c + k_0, \tag{2}$$

where k_1 and k_0 are constants.

In this case, NAS is only related to the measurement spectra and background subspace constructed from the interference signals and does not need glucose concentration reference value. It can avoid errors caused by the glucose difference from different sites or the physiological lag between glucose in the blood and interstitial fluid compartments[18]. Moreover, the chance correlations between the analyte of interest and other interferents, instrument noise, instrument drift, change of measurement conditions, and sample status variation can be significantly eliminated.

To validate the effectiveness of the new method, we performed a Monte Carlo simulation[19] on the 10% intralipid solution. The light propagation process was simulated, and the distribution of diffuse reflectance at different radial distances was obtained by initializing the optical parameters. The incident photon number was 1×10^8, and glucose concentration varied from 0 to 2 000 mg/dL with an interval of 100 mg/dL. The wavelength range was 1 200 to 1 600 nm with an interval of 20 nm. Under the illumination of an infinite ultrafine point light source, 21 groups of diffuse reflectance spectra induced by glucose concentration variations at different radial distances can be obtained.

After being corrected by the pure 10% intralipid solution, the variation in the diffuse reflectance spectra after adding glucose (50, 1 000, 1 500, and 2 000 mg/dL) in the 1 320 and 1 480 nm distances are shown in Figs. 2(a) and (b).

Figure 2 shows point A where the variation of diffuse reflectance spectra is equal to zero, indicating insensitivity to glucose, and point B where the variation of diffuse reflectance spectra has maximum sensitivity. Similar results can be observed at other wavelengths. Point A is defined as the radial floating reference. In other words, at different wavelengths, the reference and measurement points have slight differences in radial distances from the light source. Then, we constructed the background space spanned by the diffuse reflectance wavelength spectra from point A at different wavelengths and glucose concentrations, and the spectra at point B were projected on it. The regression curve between the Euclidean norm of NAS and glucose concentrations was established, as shown in Fig. 3.

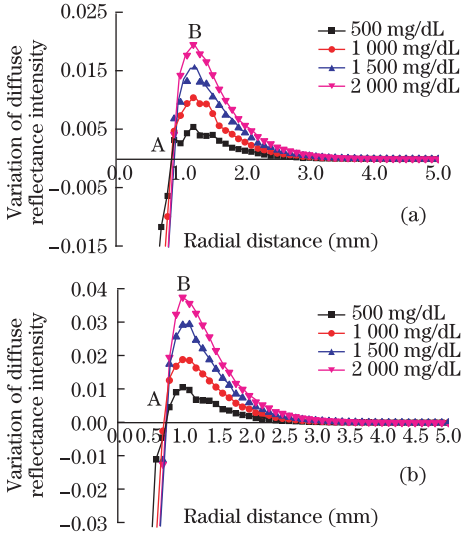

Fig. 2. (Color online) Variation of diffuse reflectance intensity in 10% intralipid with different glucose concentrations in (a) 1 320- and (b) 1 480-nm distances (in simulation).

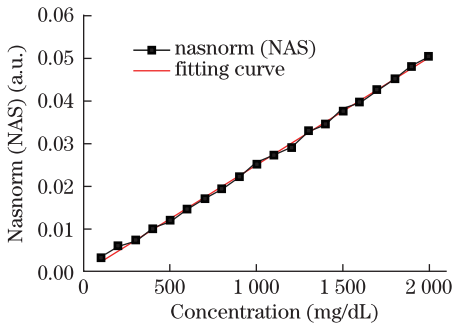

Fig. 3. (Color online) Regression curve between Euclidean norm of NAS and glucose concentration (in simulation).

The results of the NAS analysis of the simulation curves presented a good linear relationship between the Euclidean norm of NAS and glucose concentration (correlation coefficient r=0.999).

For the experimental validation, we developed a NIR diffuse reflectance measurement system based on an acousto-optic tunable filter. Spectra at the reference and measurement points at the wavelength region of 1 110−1 600 nm can be obtained simultaneously using a dual optical fiber probe. All the fiber bundles, which were composed of densely arranged single root quartz fibers (outer diameter: 0.1 mm), were integrated into one end of the probe (C). Meanwhile, the other end was divided into three probes, corresponding to the incident fiber bundle and two concentric fiber bundle loops. The incident fiber bundle (O) was in the probe center, and its radius was 1 mm. The other two loops were used for receiving backward diffuse reflectance light. The inner and outer radii of the reference point fiber bundle were 1.3 and 1.7 mm, respectively, and those of the measurement point fiber bundle were 1.7 and 2.1 mm, respectively. The schematic diagram is shown in Fig. 4.

The 10% intralipid solution with different glucose concentrations was used to simulate the human tissue, with different glucose concentrations from 100 to 2 000 mg/dL at 100 mg/dL intervals. Therefore, the dataset with 20 samples induced by glucose concentration variations and the corresponding NIR spectra can be acquired.

During the measurement, the spectra of the intralipid with and without glucose were obtained alternately, and the samples were measured randomly. The spectra of each sample at the reference and measurement points were collected simultaneously. After performing principle component decomposition on the spectra, we obtained the reconstructed measured spectra matrix $\hat{\mathbf{r}}$ and background spectra matrix $\hat{\mathbf{R}}_{\mathrm{F}}$ to calculate the NAS of glucose.

The residuals of the dataset were inspected, and the samples with glucose concentrations of 400, 1 700, 1 800, and 1 900 mg/dL were considered spectral outliers that had to be removed. Then, 16 calibration samples were set aside to build the models, which were created from the remainder of the dataset. The regression curve between the Euclidean norm of NAS and glucose concentrations was established after removing the singular points, as shown in Fig. 5.

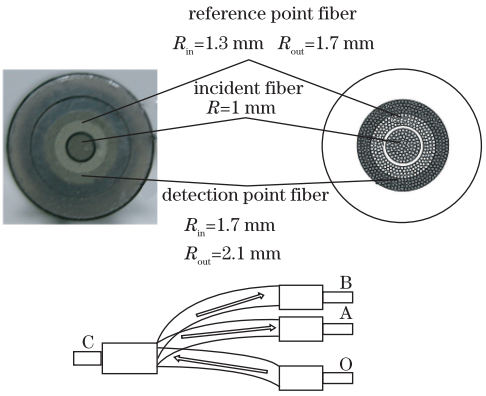

Fig. 4. Photograph and schematic diagram of the dual optical fiber probe.

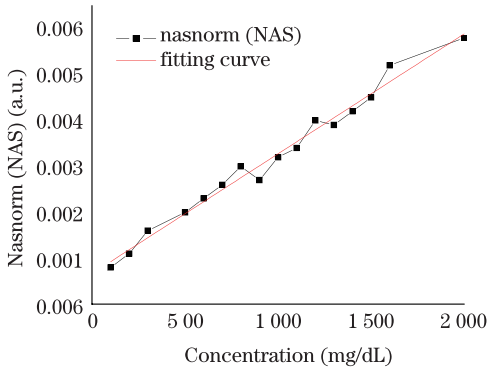

Fig. 5. Regression curve between Euclidean norm of NAS and glucose concentration (*in vitro* experiment).

Figure 5 shows that the results of the *in vitro* experiments are consistent with those of the simulation. A good linear relationship between Euclidean norm of NAS and glucose concentration (r=0.992) was observed, and the standard deviation was 0.00137 a.u.

After removing the outliers, we built the PLS regression models based on the calibration set spectra of the original measurement and the spectra with the NAS analysis included. In our investigations, the experimental dataset, consisting of 16 samples, was split into calibration (15) and prediction (1), that is, no independent prospective prediction set was obtained because of the limited number of samples. Thus, a leave-one-out full cross-validation protocol was employed to estimate the models built from the samples. The prediction accuracy of the models was evaluated in terms of the root mean square errors of calibration (RMSEC) and prediction (RMSEP), which were calculated using Unscrambler 9.7 (CAMO, Norway). The calibration and validation results are listed in Table 1.

The correlation coefficients for the calibration and cross-validation of the two models had no obvious difference. However, after NAS analysis, RMSEC and RMSEP decreased to 71.226 and 82.836 mg/dL from 100.135 and 113.989 mg/dL, respectively. The precision of the prediction errors for calibration and validation was improved by 28.87% and 27.33%, respectively.

In conclusion, a novel method based on the floating-reference theory is proposed for the extraction of NAS. The background subspace is spanned by the spectra at the floating-reference point. Then, the spectra at the measurement point are projected on it to calculate the NAS of glucose, which can effectively avoid the influence of chance temporal correlations. Moreover, the Monte Carlo simulation and *in vitro* experiments of the 10%

Table 1. Calibration and Validation Results of *in vitro* Experiment

Calibration Sets	Correlation Coefficient for Calibration	RMSEC (mg/dL)	Correlation Coefficient for Cross-validation	RMSEP (mg/dL)
1	0.981	100.135	0.977	113.989
2	0.991	71.226	0.978	82.836

1 means the calibration sets based on the original spectra at the measurement point; 2 means the calibration sets based on the spectra after NAS analysis.

ntralipid solution with different glucose contents are conducted to verify the effectiveness of the method. The results show that the NAS of glucose in the experiment is consistent with that in the simulation. The Euclidean norm of NAS is highly correlated with the glucose concentration, and the prediction error of the multivariate model can be improved, indicating that the spectra at the measurement point are glucose specific and useful in extracting the characteristic information of glucose based on the floating-reference method. Multiple factors, including fluctuations of temperature, physiological glucose dynamics, and skin heterogeneity in human subjects, are known to introduce curved effects in the relationship between the analyte concentrations and the spectra. A nonlinear regression algorithm, such as support vector machines, is introduced to enhance the robustness and prediction accuracy of the calibration model[20−23]. To evaluate the benefits of the new method, we design and conduct experiments containing nonlinear factors, which will be the focus of our future work.

This work was supported by the State Key Program of the National Natural Science Foundation of China (No. 60938002) and the National Natural Science Foundation of China (No. 30900275).

References

1. C. Bai, T. L. Graham, and M. A. Arnold, Anal. Lett. **41,** 2773 (2008).
2. H. Chung, M. A. Arnold, M. Rhiel, and D. W. Murhammer, Appl. Spectrosc. **50,** 270 (1996).
3. K. Maruo, M. Tsurugi, M. Tamura, and Y. Ozaki, Appl. Spectrosc. **57,** 1236 (2003).
4. K. Maruo, T. Oota, M. Tsurugi, T. Nakagawa, H. Arimoto, M. Tamura, Y. Ozaki, and Y. Yamada, Appl. Spectrosc. **60,** 441 (2006).
5. J. Qu and B. C. Wilson, J. Biomed. Opt. **2,** 319 (1997).
6. J. T. Olesberg, L. Liu, V. V. Zee, and M. A. Arnold, Anal. Chem. **78,** 215 (2006).
7. M. A. Arnold, J. J. Burmeister, and G. W. Small, Anal. Chem. **70,** 1773 (1998).
8. M. A. Arnold, L. Liu, and J. T. Olesberg, J. Diabetes Sci. Technol. **1,** 454 (2007).
9. N. C. Dingari, I. Barman, G. P. Singh, J. W. Kang, R. R. Dasari, and M. S. Feld, Anal. Bioanal. Chem. **400,** 2871 (2011).
10. J. Lipson, J. Bernhardt, U. Block, W. R. Freeman, R. Hofmeister, M. Hristakeva, T. Lenosky, R. McNamara, D. Petrasek, D. Veltkamp, and S. Waydo, J. Diabetes Sci. Technol. **3,** 233 (2009).
11. R. Liu, W. Chen, X. Gu, R. K. Wang, and K. Xu, J. Phys. D: Appl. Phys. **38,** 2675 (2005).
12. Y. Luo, "Tissue optics basis and application in noninvasive sensing of blood glucose concentration with near-infrared spectroscopy" (in Chinese) PhD. Thesis (Tianjin University, 2006).
13. K. Xu and W. Chen, "Method of component concentration detection based on reference wavelength" U.S. patent US 2011/0131021 A1 (2011).
14. Y. Yang, W. Chen, Z. Shi, and K. Xu, Chin. Opt. Lett. **8,** 421 (2010).
15. A. Lorber, Anal. Chem. **58,** 1167 (1986).
16. R. Liu, W. Chen, X. Gu, and K. Xu, Nanotechnol. Precis. Eng. (in Chinese) **6,** 207 (2008).
17. A. Lorber, K. Faber, and B. R. Kowalski, Anal. Chem. **69,** 1620 (1997).
18. I. Barman, C. Kong, G. P. Singh, R. R. Dasari, and M. S. Feld, Anal. Chem. **82,** 6104 (2010).
19. L. Wang, S. L. Jacques, and L. Zheng, Comput. Meth. Programs Biomed. **47,** 131 (1995).
20. I. Barman, C. Kong, N. C. Dingari, R. R. Dasari, and M. S. Feld, Anal. Chem. **82,** 9713 (2010).
21. U. Thissen, B. Üstün, W. J. Melssen, and L. M. C. Buydens, Anal. Chem. **76,** 3099 (2004).
22. F. Wülfert, W. Th. Kok, and A. K. Smilde, Anal. Chem. **70,** 1761 (1998).
23. X. Yu, Y. Wang, G. Wei, P. Zhang, and X. Long, Chin. Opt. Lett. **9,** 051201 (2011).