Copyright © 2018 Tech Science Press CMC, vol.57, no.2, pp.261-268, 2018

Band Selection Method of Absorption Peak Perturbance for the FTIR/ATR Spectrum Analysis

Jun Xie[1], Chong Wang[1], Jiaxiang Cai[2] and Fuhong Cai[1, *]

Abstract: The rapid quantification method of human serum glucose was established by using the Fourier transform infrared spectroscopy (FTIR) and attenuated total reflection (ATR). By the subtracted spectra between glucose aqueous solution and de-ionized water, absorption peaks are calculated in fingerprint area. Based on these absorption peaks and multiple linear regression (MLR) model, discrete band selection method of absorption peaks disturbance model (APDM) was developed. 5 absorption peaks 1150 cm^{-1}, 1103 cm^{-1}, 1078 cm^{-1}, 1034 cm^{-1}, 991 cm^{-1} were found in fingerprint area. Used these absorption peaks to establish absorption peaks disturbance model, the optimal wavelength combinations are 1140 cm^{-1}, 1096 cm^{-1}, 1084 cm^{-1}, 1030 cm^{-1}, 993 cm^{-1}, the corresponding C-RMSEP and C-R$_P$ are 1.164 mmol/L and 0.828 respectively. The results show that the optimal prediction effect of APDM was obviously better than the one of the Partial least squares (PLS) model, and the complexity of the optimal model is reduced greatly also. The results also provide a theoretical basis for design of small and portable human serum glucose spectrometer.

Keywords: Band selection, FTIR, ATR, serum glucose.

1 Introduction

In the field of analytical chemistry, infrared spectroscopy is an effective analysis technology for determination of structure of matter and materials. And it is also an online, real-time, *in-situ* determination of the quantitative analysis methods. It has the advantages of short measurement time, high accuracy, non-destructive testing and continuous determination. Fourier transform infrared spectroscopy (FTIR) and attenuated total reflection (ATR) infrared spectroscopy [Rios-Corripio, Rojas-Lopez and Delgado-Macuil (2012); Saguer, Alvarez, Sedman et al. (2013); Anjos, Campos, Ruiz et al. (2015); Engel, Postma, Peufflik et al. (2015); Gurbanov, Bilgin and Severcan (2016)], have wide application in the measurement of the quality of agricultural products and food, cell metabolism measurement, enzyme activity analysis. For example, data from FTIR/ATR could quantify corn syrup, high fructose syrup, and inverted sugar in Mexican honey when optimal calibrations were performed with partial least squares (PLS) [Tenenhaus, Esposito, Chatelinc et al. (2004)].

[1] The Mechanical and Electrical Engineering College, Hainan University, Haikou, Hainan, 570228, China.

[2] Department of Industrial Systems Engineering & Management, National University of Singapore, Singapore.

[*] Corresponding Author: Fuhong Cai. Email: caifuhong@zju.edu.cn.

CMC. doi:10.32604/cmc.2018.03669 www.techscience.com/cmc

Human blood is a complex multi ingredients system. Measuring blood components by FTIR/ATR technology needs to overcome many difficulties. Prediction accuracy by the spectra of various components in the blood content have not reached the level of clinical application [Ostrovsky, Zelig, Gusakova et al. (2013); Feng, Dupont and Twigg (2016)]. Spectral modeling optimization of human blood spectrum is one of the most important research fields.

In physics, the absorption bands of glucose are determined. However, human serum is a complex system with multiple components, the absorption bands inevitable is changed by the absorption interference of other unknown components.

Wave band selection is necessary because the prediction effect of model to improve when the signal-to-noise ratio (SNR) of the wave band is relatively low. The spectroscopic analysis of a single blood component requires mitigating the interference of other components and noise. Improving prediction effect, reducing model complexity and designing specialized spectrometers with high SNR are all important. Therefore, appropriate chemometric methods are necessary for wave band optimization [Xie, Zhang, Li et al. (2017); Parab, Srivastava, Samui et al. (2014)].

Partial least squares (PLS) regression is a statistical method that bears some relation to principal components regression. It finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Because both independent variables and dependent variables are projected to new spaces, the PLS family of methods are known as bi-linear factor models [Tenenhaus, Esposito, Chatelinc et al. (2004)]. PLS has been most widely used in the spectral analysis of model calibration, such as screening spectroscopic data, extracting information variables and overcoming spectral co-linearity.

Multiple linear regression (MLR), also known as inverse least square method, is a commonly used method in quantitative analysis of infrared spectra [Tenenhaus, Esposito, Chatelinc et al. (2004)]. MLR is simple, easy to understand and fast. The key of MLR in spectral analysis is how to choose the appropriate wavelength. Usually calculating all the wavenumbers combination of the whole spectral region is impossible because of very large workload. In this work, discrete band selection method of absorption peaks disturbance model (APDM) was developed to reduce the huge amount of computation.

Firstly, by the subtracted spectra between glucose aqueous solution and de-ionized water, absorption peaks were calculated in fingerprint area. Secondly, based on these absorption peaks and multiple linear regression (MLR) model, discrete band selection method of absorption peaks disturbance model (APDM) was developed. Finally, the independent samples set was tested for model verification. As a comparison, the whole spectral region PLS model of NIRS analysis of serum glucose was also established.

2 Experimental and methods

2.1 Experimental materials, instruments and measurement methods

One hundred and eighty-six human serum samples were collected. Glucose concentration of the samples was measured by routine clinical method with BC-3000Plus automatic blood cell analyzer (Shenzhen Mairea Company). Glucose concentration ranged from 4.09-17.82 mmol/L, the mean values and standard deviations were 6.33 mmol/L and 2.31

mmol/L, respectively. The instrument used for the experiment was a VERTEX 70 FTIR spectrometer (BRUKER Company) equipped with a KBr beam splitter and a deuterated triglycine sulfate KBr detector. The scanning scope of the spectrum was 4000-600 cm^{-1} with a horizontal ATR sampling accessory with a diamond internal reflection element on a ZnSe crystal (SPECAC Company, 45° angle of incidence, 3 times reflective). Each sample was measured three times and the mean value of the three measurements was used for modeling. The spectra were measured at 22°C±1°C and 41%±1% RH.

2.2 Model evaluation indicators and division method for sample sets

Sixty-two samples were randomly selected from all 186 samples as the validation set; the remaining 124 samples were grouped as the modeling set. The modeling set was divided into similar calibration set (84 samples) and prediction set (40 samples) for a total of 20 times. For each division *i*, modeling root mean square error of prediction and modeling correlation coefficients of prediction were denoted as M-SEP_i and M-$R_{P,i}$, respectively. Their mean value and standard deviation for all divisions were denoted as M-SEP_{Ave}, M-$R_{P,Ave}$, M-SEP_{Std}, and M-$R_{P,Std}$, respectively, serving as the basis for discussing the prediction accuracy and stability of the modeling. The model parameters were selected according to the minimum M-SEP_{Ave}. Finally, the selected model was validated in the validation set, and the validation root mean square error of prediction and validation correlation coefficients of prediction were calculated and denoted by V-SEP and V-R_P, respectively.

The similarity of the sample sets was defined by the predictive bias of cross-validation (leave-one-out mode) for PLS model according to on the whole spectral region. The predictive bias of each sample was calculated and is called Partial Least Square Cross-validation Predictive Bias, denoted by PLSPB. If the mean value and the standard deviation of PLSPB in calibration set were close to those in prediction set, the calibration set and the prediction set were defined similar.

2.3 APDM method

Infrared spectrum of material is the reflection of its molecular structure, and the absorption peaks in the spectrum correspond to the vibration of the groups in the molecule. The various groups of molecules, such as O-H, N-H, C-H, C=C, C-OH and C-C, all have specific infrared absorption regions. The other parts of the molecule have less influence on their absorption position. Generally, this kind of absorption band is called group frequency, and its location is also known as characteristic absorption peak. Each group of molecules has some unique infrared signature absorption peaks.

Absorption spectra of a single component in human blood will vary because of complexity of human blood. Therefore, the absorption peaks based on a single component also have some small disturbances. If the absorption peaks are still used for modeling, the model effect will not be very good. So, in the complex human blood system, these absorption peaks have to change. In this work, an absorption peak perturbance model (APDM) was developed to adapt the perturbance of human blood spectra.

Each spectral absorption peak has a small perturbation around itself. The subtracted spectra between glucose aqueous solution and deionized water were illustrated as the example in Fig. 1. Only 5 of the absorption peaks are drawn. The main steps are as follows. At first, the absorption peaks of the difference spectrum of glucose solution and deionized water were calculated, assuming that the number of absorption peaks is P. Secondly, any combination of the absorption peaks were calculated. Thirdly, in each combination, the absorption peaks can swing freely in a certain range. At last, all the models base on the combination of perturbation peaks were calculated, and the optimal model was selected.

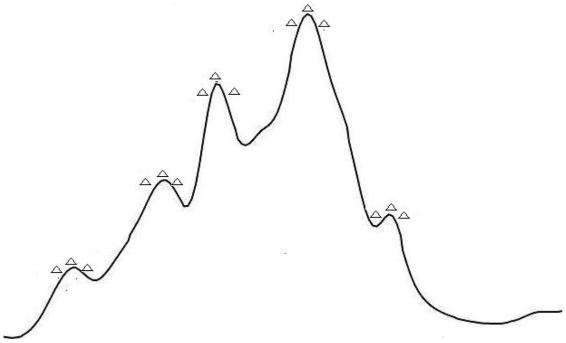

Figure 1: Perturbation of absorption peaks

The number of wavelength combinations was calculated as follows. Assuming that the total number of absorption peaks is *P*, and the maximum perturbation number around the peak is *u*. Thus, the corresponding number of each peak is *2u+1*. The number of models is defined by the Eq. (1):

$$c(u,k)=C_P^k(2u+1)^k \tag{1}$$

if there are k peaks. The total number m of model is defined by the Eq. (2):

$$m=\sum_{k=1}^{p}c(u,k)=\sum_{k=1}^{P}C_P^k(2u+1)^k \tag{2}$$

The global optimal band combination was obtained by calculation all the models. The above algorithm platform was built up by us by python software version 3.4.

3 Results and discussion

FTIR/ATR spectra of 186 serum samples were shown in Fig. 2. Spectra are heavily overlapping each other. And it is impossible to distinguish them.

The band 2400-2300 cm^{-1} is a blind area of VERTEX 70 FTIR spectrometer and the curves are abnormal. But that will not affect the research because spectra near 2350 cm^{-1} are often eliminated. There is a strong absorption of water molecules near the band 1637 cm^{-1}.

One hundred and twenty-four samples were used as the calibration set for modeling, the remaining 62 samples which are completely independent of the modeling samples were used to test the model. Chemical value distributions of the calibration set and prediction set are indicated in Tab. 1 which shows that the range of the prediction set is within the range of the calibration set.

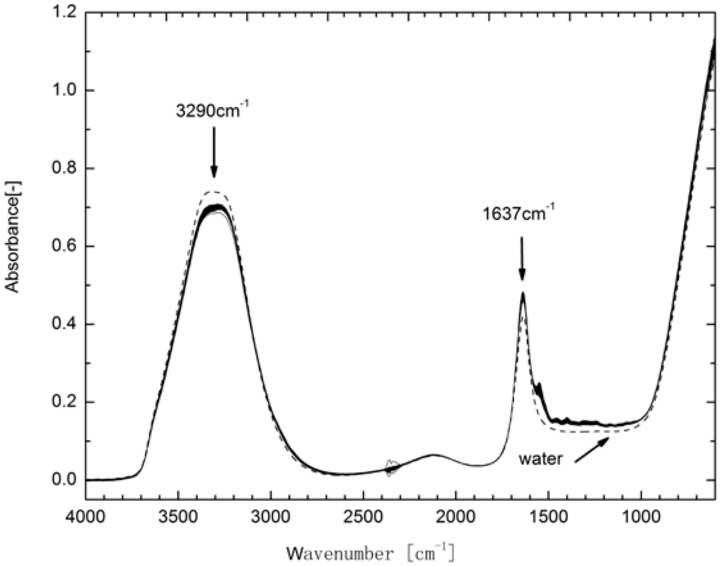

Figure 2: FTIR/ATR spectra of 186 human serum samples

Table 1: Distribution of chemical values in the calibration set and the prediction set (Unit: mmol/L)

	Min	Max	Mean	Standard deviation
Calibration set	4.09	17.82	6.34	2.34
Prediction set	4.25	16.44	6.32	2.26

There is no significant difference between the spectral of glucose solution and that of human blood, so the subtracted spectra between glucose aqueous solution and de-ionized water were as the reference, and the optimal model can be got also. The subtracted spectra were shown as Fig. 3. In the finger spectral region 1200-900 cm^{-1}, 5 peaks are 115 cm^{-1}, 1103 cm^{-1}, 1078 cm^{-1}, 1034 cm^{-1}, 991 cm^{-1}. These peaks are correlated with the C-O Stretching vibration and deformation. Although there are also absorption peaks near 3100 cm^{-1}, 2900 cm^{-1} and 1400 cm^{-1}, the intensity is all much smaller. Only the 5 absorption peaks are considered when establishing the discrete combination model.

The prediction effect, stability of the optimal model of each wavenumber combination is summarized in Tab. 2 and Tab. 3. There are too many combinations by 29 glucose absorption peak. On the other hand, the APDM greatly reduces the computation while getting better model effect. The APDM model does find a very good wavelength combination around the five absorption peaks of the differential spectrum, and it is necessary and effective to perturbation at the wavelength of the absorption peak.

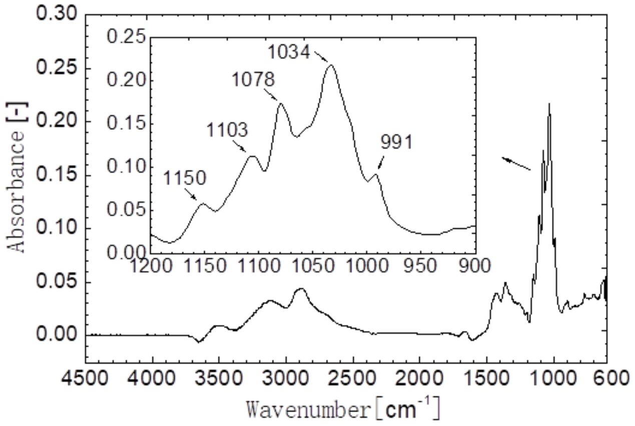

Figure 3: The subtracted spectra between glucose aqueous solution and de-ionized water

Table 2: Comparison of absorption peaks combination models

Number of wavenumber	Number of model	M-SEP$_{Ave}$ (mmol/L)	M-RP$_{Ave}$	Combination of wavenumber (cm^{-1})
1	5	2.021	0.502	1034
2	10	1.628	0.717	1103,1034
3	10	1.612	0.724	1150,1103,1034
4	5	1.613	0.724	1150,1103,1078,1034
5	1	1.626	0.719	1150,1103,1078,1034,991

Table 3: Comparison of APDM models

Number of wavenumber	Number of model	M-SEP$_{Ave}$ (mmol/L)	M-RP$_{Ave}$	Combination of wavenumber (cm^{-1})
1	55	1.846	0.455	1034
2	1210	1.282	0.787	1096,1084
3	13310	1.216	0.811	1140,1096,1084
4	73205	1.164	0.828	1140,1096,1084,1030
5	161051	1.165	0.827	1140,1096,1084,1032,993

Used these absorption peaks to establish absorption peaks disturbance model, the optimal wavelength combinations are 1140 cm^{-1}, 1096 cm^{-1}, 1084 cm^{-1}, 1030 cm^{-1}, 993 cm^{-1}, the corresponding C-RMSEP and C-RP are 1.164 mmol/L and 0.828 mmol/L respectively.

The optimal prediction effect of absorption peaks disturbance model is obviously better than the one of the PLS model. The numbers of wavelengths adopted are only 5, and the complexity of the optimal model is reduced greatly. The results also provide a theoretical basis for design of small and portable human serum glucose spectrometer.

The model based on the combination bands 1140 cm^{-1}, 1096 cm^{-1}, 1084 cm^{-1}, 1030 cm^{-1}, 993 cm^{-1} was validated by the validation set. The V-SEP and V-R_P are 1.164 mmol/L and 0.828 mmol/L, respectively, which showed a rather good validation effect.

4 Conclusions

By the FTIR/ATR difference spectra between glucose aqueous solution and de-ionized water, absorption peaks are calculated in fingerprint area. Based on these peaks and multiple linear regression (MLR) model, discrete band selection method of absorption peaks disturbance model (APDM) was developed. Used these absorption peaks to establish absorption peaks disturbance model, the optimal wavelength combinations are 1140 cm^{-1}, 1096 cm^{-1}, 1084 cm^{-1}, 1030 cm^{-1}, 993 cm^{-1}, the corresponding C-RMSEP and C-RP are 1.164 mmol/L and 0.828 mmol/L respectively. The results show that APDM model is effective and concise, and it can also serve as the theoretical basis for design of small and portable human serum glucose spectrometer.

Acknowledgements: This work was financially supported by the Natural Science Foundation of Hainan Province (417087), the Key Research and Development Program of Hainan Province (ZDYF2018007), the Science and Research Project of Hainan Province Education Department (No. Hnky2015-1), and Research Fund for Advanced Talents of Hainan University (No. kyqd1577).

References

Anjos, O.; Campos, M. G.; Ruiz, P. C.; Antunes, P. (2015): Application of FTIR-ATR spectroscopy to the quantification of sugar in honey. *Food Chemistry*, vol. 169, no. 169, pp. 218-223.

Bekiari, V.; Avramidis, **P.** (2014): Data quality in water analysis: Validation of combustion-infrared and combustion-chemiluminescence methods for the simultaneous determination of total organic carbon (TOC) and total nitrogen (TN). *International Journal of Environmental Analytical Chemistry*, vol. 94, no. 1, pp. 65-76.

Engel, J.; Postma, G. J.; Peufflik, **I. V.; Blanchet, L.; Buydens, L. M. C.** (2015): Pseudo-sample trajectories for variable interaction detection in dissimilarity partial least squares. *Chemometrics and Intelligent Laboratory Systems*, vol. 146, no. 1, pp. 89-101.

Feng, C.; Dupont, V.; Twigg, M. V. (2016): Temperature-programmed reduction of nickel steam reformingcatalyst with glucose. *Applied Catalysis A: General*, vol. 527, pp. 1-8.

Gurbanov, R.; Bilgin, M.; Severcan, F. (2016): Restoring effect of selenium on the molecular content, structure and fluidity of diabetic rat kidney brush border cell membrane. *Biochimica et Biophysica Acta*, vol. 1858, no. 4, pp. 845-854.

Ostrovsky, E.; Zelig, U.; Gusakova, I.; Ariad, S.; Mordechai, S. et al. (2013): Detection of cancer using advanced computerized analysis of infrared spectra of peripheral blood. *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 343-353.

Parab, S.; Srivastava, S.; Samui, P. Murthy, A. R. (2014): Prediction of fracture parameters of high strength and ultra-high strength concrete beams using Gaussian process regression and Least squares support vector machine. *Computer Modeling in Engineering & Sciences*, vol. 101, no. 2, pp. 139-158.

Rios-Corripio, M. A.; Rojas-Lopez, M.; Delgado-Macuil, R. (2012): Analysis of adulteration in honey with standard sugar solutions and syrups using attenuated total reflectance-fourier transform infrared spectroscopy and multivariate methods. *CYTA-Journal of Food*, vol. 10, no. 2, pp. 119-122.

Saguer, E.; Alvarez, P. A.; Sedman, **J.; Ismail, A. A.** (2013): Study of denaturation/aggregation behavior of whole porcine plasma and its protein fractions during heating under acidic pH by variable-temperature FTIR. *Food Hydrocolloids*, vol. 33, no. 2 pp. 402-414.

Tenenhaus, M.; Esposito, V.; Chatelinc, Y. M.; Lauro, C. (2004): PLS path modeling. *Computational Statistics & Data Analysis*, vol. 48, no. 3, pp. 159-205.

Xie, J.; Zhang, H.; Li, J.; Cai, F. (2018): Window subtracted wave band selection method for the FTIR ATR spectrum analysis. *Progress in Electromagnetic Research M*, vol. 68, pp. 53-59.