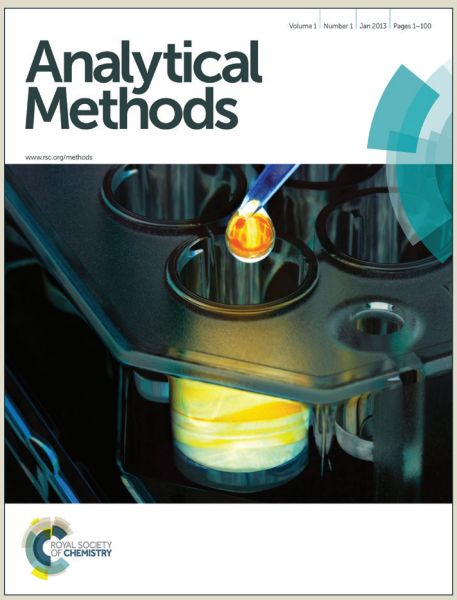
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

PAPER

Correlation Coefficient Optimization in Partial Least-Squares Regression with Application to ATR-FTIR Spectroscopic Analysis

Yifang Chen,[1] Jiemei Chen,[2,*] Tao Pan,[1,*] Yun Han,[1] Lijun Yao[1]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

A wavelength selection method for spectroscopic analysis, named correlation coefficient optimization coupled with partial least-squares (CCO-PLS), was proposed, and was successfully employed for reagent-free ATR-FTIR spectroscopic analysis of albumin (ALB) and globulin (GLB) in human serum. By varying upper bound of correlation coefficient between absorbance and analyte's content, CCO-PLS method achieve multi-band selection. Two PLS-based methods, which used a waveband having positive peaks of first loading vector (FLV) and a combination of positive peaks for correlation coefficient spectrum, were also conducted for comparison. Based on leave-one-out cross validation for CCO-PLS, the appropriate wavebands combinations for ALB and GLB were selected, the root-mean-square error of prediction for validation samples were 1.36 and 1.35 (g L^{-1}) for ALB and GLB, respectively, which were better than the two comparison methods. CCO-PLS method provided a new approach of multi-band selection to achieve high analytical accuracy for the molecular absorption bands that were composed of several spaced wavebands.

Introduction

As we all know, mid-infrared (MIR) spectrum has rich information on the molecular structure and contents of the material, which could be used to analyze biological and biomedical samples. However, the interference caused by the strong absorption of water molecules in the MIR region is the main difficulty in the use of MIR for direct measurement of biological and biomedical samples (e.g., blood). Therefore, sample drying is general need.

Incidentally, in parallel developments of a Fourier transform infrared (FTIR) spectrometers and attenuated total reflection (ATR) techniques[1], the spectroscopic method using an FTIR equipped with an ATR accessory (ATR-FTIR) provides substantial potential as a quantitative tool to directly measure the samples that contain water molecules. The design of an ATR accessory is based on the principle of internal reflection of light. Infrared light that is emitted by a light source through a crystal with a large refractive index can be projected onto the sample surface using a small refractive index; total reflection occurs when the angle of incidence is greater than the critical angle. Actually, an attenuated evanescent wave is formed on the contact surface; not all the infrared light is reflected back; it partially penetrates to a certain depth beneath the surface of the specimen and then returns to the surface. The penetration depth of the evanescent wave is significantly lesser than the optical path of ordinary transmission accessories; thus avoiding the strong absorption of water molecules. The ATR-FTIR technique has been extensively applied in the areas of life sciences, clinical medicine, and others. [2-12]

Partial least-squares (PLS) regression is widely used for comprehensively screening spectroscopic data, extracting information variables and overcoming spectral co-linearity.[13-18] However, wavelength selection is necessary because the prediction effect of PLS is difficult to improve when the signal-to-noise ratio (SNR) of the wavelengths combination is not adequately high. For a complex object with multiple components (e.g., blood), the spectroscopic analysis of some components must mitigate noise disturbance from its other components. Therefore, an appropriate wavelength selection is an important and difficult technical aspect to improve prediction effectiveness, reduce model complexity, and design a specialized spectrometer with a high SNR. Furthermore, appropriate chemometric methods must be utilized for optimizing the wavelengths combination.

Albumin (ALB) and globulin (GLB) contained in human blood both are very important for clinical test and disease diagnosis.[19] A chemical-free and rapid analysis of ALB and GLB using ATR-FTIR spectroscopy has been the focus of previous study.[2] A PLS-based method for wavelength selection, which used PLS regression in a region having positive peaks of first loading vector (FLV), was implemented.[2] The FLV-PLS method was based on an assumption that the FLV in PLS regression corresponds to the first-order approximation of the pure-component spectrum of analyte. And the data of a continuous

[1]*Department of Optoelectronic Engineering, Jinan University, Huangpu Road West 601, Tianhe District, Guangzhou 510632, China.* [2]*Department of Biological Engineering, Jinan University, Huangpu Road West 601, Tianhe District, Guangzhou 510632, China. *E-mail: tpan@jnu.edu.cn (T. Pan) and tchjm@jnu.edu.cn (J. M. Chen), Tel: +86-20-85224379.*

region having the highest positive peaks of FLV were uesd as PLS regression. The selection of initial and end wavenumbers was based on the best prediction accuracy for the corresponding PLS model. FLV-PLS is a single-band screening method, and it cannot easily achieve high analytical accuracy for the object of relative dispersion of absorption bands. ALB and GLB both are the protein molecules in blood; they have multiple absorption bands in the MIR region. Therefore, FLV-PLS method is not sufficient for the analysis of ALB and GLB.

Indeed, the correlation coefficient (R) between the absorbance and the concentration of the analyte (e.g., ALB or GLB) in each wavenumber is also an important factor to consider. Than other spectral region, such as near-infrared (NIR) region, the correlation coefficient values in MIR region are usually higher. A traditional method is based on the data in the positive peaks of correlation coefficient spectrum (CCS),[3-5] and it is usually possible to achieve good prediction effect for the samples that component structure is relatively simple. However, for the complex samples with multi-component, the presence of a lot of interference, the information's in the positive peaks of CCS are insufficient.

It is noteworthy that each wavenumber corresponds to an R value in CCS. Therefore, it could be an interesting attempt to perform wavenumber selection on the basis of R value. In the present study, a method for wavenumber selection, named correlation coefficient optimization coupled with PLS (CCO-PLS), was proposed based on the selection for upper bound of R value. Because different wavenumber may correspond to the same R value, so a range of R value may correspond to the combination of multiple wavebands. CCO-PLS method could be easily selected appropriate combination of several wavebands corresponding to the relative dispersion of the molecular absorption band.

In this study, the ATR-FTIR analysis of ALB and GLB were used as the examples to verify the proposed CCO-PLS method. While the FLV-PLS method and the method that based on the positive peaks of CCS were also conducted for comparison.

Materials and methods

Experimental materials, instruments and measurement methods

A total of 194 human peripheral blood samples were collected, and the informed consent from all subjects was obtained. Experiment was performed in compliance with the relevant laws and institutional guidelines and approved by local medical institution. Let each peripheral blood sample stand for about one hour at room temperature to clot. An appropriate amount of coagulated sample was extracted and centrifuged using centrifuge at 3,000 rpm for 8 minutes. The pale yellow supernatant, namely serum, was obtained via pipette. Given that serum is free from haemocytes and fibrinogen, ALB and GLB in serum are usually used as the clinical reference indicators. The standard clinical analysis methods of ALB and GLB are bromocresol green colorimetry and immune turbidimetry, both of them need reagents.

The ALB and GLB values of the serum samples were measured via a routine clinical method using a Hitachi 7060 automatic biochemical analyzer (Hitachi, Japan). The obtained values were used as the reference values for spectroscopic analysis in modeling and validation process. The statistical analysis of the actual measured ALB and GLB values of the 194 samples is given in Table 1.

Table 1 Statistical analysis of measured ALB and GLB values (g L^{-1}) of 194 human serum samples.

Indicator	Min	Max	Ave	SD
ALB	20.90	54.60	38.97	6.36
GLB	18.70	39.50	27.62	4.41

Note: SD is the abbreviation of standard deviation.

Considering that the high content of tangible part in whole blood can lead to strong absorption in the MIR region and scattering interference, in this study, the spectrum of serum was used for quantitative analysis of ALB and GLB.

Spectra were collected using a VERTEX 70 FTIR Spectrometer (BRUKER Co., Germany) equipped with a KBr beam splitter and a deuterated triglycine sulfate KBr detector. The MIR spectra were obtained from 4,000 to 600 cm^{-1} using a horizontal ATR sampling accessory with a diamond internal reflection element on a zinc selenide crystal (SPECAC Co., UK). Thirty-two scans of symmetrical interferograms at 4 cm^{-1} resolution were added to each spectrum. The instrument was allowed to purge for several minutes prior to the acquisition of spectra for minimizing the spectral contribution from atmospheric water vapor. Each sample (0.08 mL) was measured thrice, and the mean value of the three measurements was used for modeling and validation. The spectra were obtained at a temperature of 25±1°C and 45±1% RH; the time of acquisition an ATR-FTIR spectrum about 1 min.

Modeling and validation process

All the samples were randomly divided into modeling and validation sets. The modeling samples were performed leave-one-out cross validation (LOOCV) for PLS model, and then optimized model parameters, such as wavebands combination, number of PLS factor etc. The randomly selected validation samples, which were not subjected to the modeling optimization process, were used to validate the optimal model.

The specific procedure was the follows. First, 84 of the all 194 samples were randomly selected as the validation set. The remaining 110 samples were used as modeling samples, and were performed the LOOCV for PLS model. Namely, each modeling sample was leaved out to predict, and the remaining 109 samples were used to construct the calibration model. Based on the same process, the prediction values of all 110 modeling samples were calculated. The measured and predicted values for i^{th} sample in the modeling set were denoted as $C_{\mathrm{M},i}$, $\widetilde{C}_{\mathrm{M},i}$, $i=1,2,\cdots,n_{\mathrm{M}}$, $n_{\mathrm{M}}=110$. The mean measured value and mean predicted value of all samples were denoted as $C_{\mathrm{M,Ave}}$, $\widetilde{C}_{\mathrm{M,Ave}}$, respectively. The prediction accuracy was evaluated by the root-mean-square errors and the correlation coefficients of prediction for LOOCV, and denoted as SECV and $R_{\mathrm{P,CV}}$, respectively. The calculation formulas were as the follows:

$$\mathrm{SECV}=\sqrt{\frac{\sum_{i=1}^{n_{\mathrm{M}}}(\widetilde{C}_{\mathrm{M},i}-C_{\mathrm{M},i})^2}{n_{\mathrm{M}}}}, \quad (1)$$

$$R_{P,CV}=\frac{\sum_{i=1}^{n_M}(C_{M,i}-C_{M,Ave})(\widetilde{C}_{M,i}-\widetilde{C}_{M,Ave})}{\sqrt{\sum_{i=1}^{n_M}(C_{M,i}-C_{M,Ave})^2(\widetilde{C}_{M,i}-\widetilde{C}_{M,Ave})^2}}. \quad (2)$$

The model parameters were selected to achieve minimum SECV.

Using all modeling samples as the calibration set, the selected model was then validated against the validation set. The root-mean-square error and correlation coefficients of prediction were then calculated and denoted as SEP and R_P, respectively. The measured and predicted values of k^{th} sample in the validation set were denoted as $C_{V,k}$, $\widetilde{C}_{V,k}$, $k=1,2,\cdots,n_V$, $n_V=84$. The mean measured value and mean predicted value of all validation samples were denoted as $C_{V,Ave}$, $\widetilde{C}_{V,Ave}$, respectively. The calculation formulas are as the follows:

$$SEP=\sqrt{\frac{\sum_{k=1}^{n_V}(\widetilde{C}_{V,k}-C_{V,k})^2}{n_V-1}}, \quad (3)$$

$$R_P=\frac{\sum_{k=1}^{n_V}(C_{V,k}-C_{V,Ave})(\widetilde{C}_{V,k}-\widetilde{C}_{V,Ave})}{\sqrt{\sum_{k=1}^{n_V}(C_{V,k}-C_{V,Ave})^2\sum_{k=1}^{n_V}(\widetilde{C}_{V,k}-\widetilde{C}_{V,Ave})^2}}. \quad (4)$$

FLV-PLS method

The specific steps of FLV-PLS method were as follows: (1) PLS regression was performed based on the matrix of the full spectum and the concentration vector of all modeling samples, and the loading and score matrices were obtained. (2) The loading vector corresponding the maximum value of score, which named the first loding vector (FLV), was extracted. FLV was considered as the first-order approximation of the pure-component spectrum of analyte, and its the highest positive peak was derived. (3) The LOOCV for PLS models was performed in each continuous regions having the highest positive peaks of FLV, and the corresponding SECV and $R_{P,CV}$ were calculated. By varying the initial and end wavenumbers, the optimal waveband with minimum SECV was selected to achieve the best prediction accuracy.

Refered to literature [2], the region 3,009-895 cm^{-1}, except the high water absorption band at 1,697-1,591 cm^{-1}, was used as the full spectral range.

Proposed CCO-PLS method

The correlation coefficient (R) between the absorbance and the concentration of the analyte (e.g., ALB or GLB) in each wavenumber λ is defined by the following equation:

$$R(\lambda)=\frac{\sum_{i=1}^{n}(A_i(\lambda)-A_{Ave}(\lambda))(C_i-C_{Ave})}{\sqrt{\sum_{i=1}^{n}(A_i(\lambda)-A_{Ave}(\lambda))^2\sum_{i=1}^{n}(C_i-C_{Ave})^2}}, \quad (5)$$

where n was number of all samples, n=194; $A_i(\lambda)$, C_i were the absorbance for wavenumber λ and the measured concentration of i^{th} sample, $i=1,2,\cdots,n$; $A_{Ave}(\lambda)$, C_{Ave} were mean absorbance and mean measured concentration for all samples, respectively.

The entire scanning region was set as Λ, then the correlation coefficient spectrum (CCS) can be expressed as the follows

$$R(\Lambda)=\{R(\lambda)\,|\,\lambda\in\Lambda\}. \quad (6)$$

The minimum and maximum values of CCS were set as R_{min} and R_{max}, respectively; a upper bound value of $R(\lambda)$ was arbitrarily set and denoted as R_{upper}, $R_{min}\le R_{upper}\le R_{max}$.Then, the range of R value, [R_{upper}, R_{max}], could corresponded to a combination of multiple wavebands. The proposed CCO-PLS method was based on varying the value of R_{upper} to select a appropriate wavebands combination with optimal prediction effect. The specific steps were as follows:

Step 1 An appropriate step of R value (ε) was set, and the range of R value, [R_{min}, R_{max}], was divided into m equal portions by m +1 nodes.

Step 2 Any one node was selected from all m +1 nodes as a value of R_{upper}. For the range of R value, [R_{upper}, R_{max}], a combination of wavebands was selected according to relationship between wavenumber and R value. The LOOCV for PLS model was employed on the basis of the obtained wavebands combination. The corresponding SECV, $R_{P,CV}$ were calculated.

Step 3 The optimal value of R_{upper} and the corresponding wavebands combination were selected according to minimum SECV.

In this study, the entire scanning region Λ was 4000-600cm^{-1}, and the number of wavenumbers (N) was 1764. In the region Λ, the R_{min} values for ALB and GLB were -0.30, -0.20 respectively, while the R_{max} values were 0.84、0.68 respectively. The step ε was set as 0.01, and the correlation coefficient range, [R_{min}, R_{max}], was divided into 115 equal portions by 116 nodes for ALB, 89 equal portions by 90 nodes for GLB, respectively. The number of PLS factors (F) was set to be $F\in\{1,2,\cdots,30\}$.

The computer algorithms for the methods discussed above were designed using the MATLAB 7.6 version software.

Results and discussion

The ATR-FTIR spectra of the 194 human serum samples in the entire scanning region 4,000-600 cm^{-1} are shown in Fig. 1. For comparison, the spectrum of distilled water is represented by a dotted line in Fig.1. Fig. 1 had a red dotted line and 194 black solid lines, where the red dotted line was the spectrum of distilled water, and the black solid lines were the spectra of 194 samples of human serum. Fig. 1 shows that the water molecules had an absorption within the range of 3,290 and 1,637 cm^{-1}, the absorption of ALB and GLB composition and other blood components mainly appeared within the MIR fingerprint region 1,700-900 cm^{-1}, and different samples had significantly different ranges of absorption in the fingerprint region.

For comparision, the PLS models for ALB and GLB were first established on the basis of the entire scanning region. The prediction effects (SECV, $R_{P,CV}$) are summarized in Table 2. The results show that the prediction effects for ALB and GLB both are lower. Moreover, the number of wavenumbers employed was 1,764, so the models were very complex. Therefore, wavenumber selection is necessary.

Table 2 Prediction effects of PLS models with the entire scanning region (4000-600 cm^{-1}) for ALB and GLB (g L^{-1}).

Indicator	N	F	SECV	$R_{P,CV}$

ALB	1764	13	1.60	0.959
GLB	1764	13	1.67	0.926

Note: *N*: number of wavenumbers; *F*: number of PLS factor.

PLS models based on the positive peaks of CCS

The correlation coefficient spectra (CCS's) for ALB and GLB are shown in Fig. 2. The results show that the high correlation coefficient mainly appeared within the MIR fingerprint region 1,700-900 cm^{-1}. The positive peaks of CCS's in 1,700-900cm^{-1} region are shown in Fig. 3. For ALB, the five wavenumbers were 1,655, 1,545, 1,396, 1,296 and 1,173 cm^{-1}. For GLB, the eight wavenumbers were 1,614, 1,570, 1,526, 1,452, 1,398, 1,227, 1,155 and 1,070 cm^{-1}. The PLS models for ALB and GLB were established within the combination of the wavenumbers corresponding to the positive peaks, respectively. The prediction effects results (SECV, $R_{P,CV}$) are summarized in Table 3. The results show that the prediction accuracy for ALB and GLB both are lower than the entire scanning region. Therefore, the information's in the positive peaks of CCS are insufficient.

Table 3 Prediction effects of PLS models based on the positive peaks of CCS for ALB and GLB (g L^{-1}).

Indicator	*N*	*F*	SECV	$R_{P,CV}$
ALB	5	4	2.28	0.915
GLB	8	7	2.69	0.792

Note: *N*: number of wavenumbers; *F*: number of PLS factor.

FLV-PLS Method

The FLV's of ALB and GLB are displayed in Fig. 4. The FLV for ALB showed the same the positive peaks as those for GLB. Their the highest peaks are both 1,547 cm^{-1}. The initial and end wavenumbers of analytical wavebands were slected form 3,009-1,549 cm^{-1} and 1,545-895 cm^{-1} respectively.

According to the minimum SECV, the optimal wavebands for ALB and GLB were selected, and the corresponding prediction effects are summarized in Table 4. The result show that the optimized waveband were 1,580-1,113 cm^{-1} for ALB and 1,566-1,151 cm^{-1} for GLB, which are within the high absorption peaks of protein between 1,591 and 1,111 cm^{-1}. And the optimal FLV-PLS models were better than above two methods.

Table 4 Optimal FLV-PLS waveband and prediction effects for ALB and GLB (g L^{-1}).

Indicator	Waveband (cm^{-1})	*N*	*F*	SECV	$R_{P,CV}$
ALB	1580-1113	245	7	1.45	0.966
GLB	1566-1151	218	12	1.50	0.941

Note: *N*: number of wavenumbers; *F*: number of PLS factor.

CCO−PLS method

According to the method proposed above, the SECV's corresponding to each upper bound of correlation coefficient R_{upper} for ALB and GLB were caculated, and are shown in Fig. 5. The result show that the optimal values of R_{upper} were 0.72 for ALB and 0.59 for GLB. The corresponding wavebands combinations were 1,666-1,643 & 1,578-1,499 & 1,408-1,387 & 1,177-1,171 (cm^{-1}) for ALB and 1,643-1501 & 1,402-1,389 & 1,252-1,200 & 1,167-1,132 & 1,119-1,109 & 1,097-1,005 (cm^{-1}) for GLB, are shown in Fig. 6. The prediction effects are shown in Table 5.

It is worth noting that, the wavebands combinations selected by CCO-PLS located in the spectral regions 1,666-1,171 cm^{-1} for ALB and 1,643-1,005 cm^{-1} for GLB. Within 1,666-1,171 cm^{-1}, there are four absorption peaks of reference spectrum of ALB, [20] where three of them (1,656, 1,541 and 1,398 cm^{-1}) are included in the waveband combinations for ALB selected by CCO-PLS. While within 1,643-1,005 cm^{-1}, there are six absorption peaks of reference spectrum of GLB,[21] which are 1,638, 1,585, 1550, 1535, 1518, and 1242 cm^{-1}. All of them are included in the waveband combinations for GLB selected by CCO-PLS. The results indicate that the waveband combinations obtained by CCO-PLS are in almost consistency with the reference spectrum of ALB and GLB.

Table 5 Prediction effects of CCO-PLS method for ALB and GLB (g L^{-1}).

Indicator	*N*	*F*	SECV	$R_{P,CV}$
ALB	100	7	1.34	0.971
GLB	240	12	1.33	0.953

Note: *N*: number of wavenumbers; *F*: number of PLS factor.

The result shows that both SECV and $R_{P,CV}$ values of CCO−PLS models were improved compared with FLV-PLS model for two indicators. It is worth noting that, the selected wavebands 1,666-1,643 for ALB and 1,643-1,501 for GLB shown in Fig. 6, were wholly or partially within of the high water absorption band at 1,697-1,591 cm^{-1}, which were not subjected to the modeling process in FLV-PLS method. The better prediction effects of CCO-PLS method indicated that removing the high water absorption band from analytic region is not wise. Actually, referring to the reference spectra of ALB and GLB, the high water absorption band (1,697-1,591 cm^{-1}) contains one absorption peak of ALB (1,656 cm^{-1}) and four absorption peaks of GLB (1,680, 1,674, 1,683 and 1,638 cm^{-1}). Therefore, the high water absorption band cannot be removed from analytic region.

On the other hand, Fig.5 shows that the wavenumbers corresponding to low *R* values were rejected, and the selected wavenumbers form a wavebands combination rather than a continuous waveband for each indicator. Therefore, CCO-PLS method could be easily selected appropriate combination of several spectral wavebands corresponding to multiple molecular absorption bands, and overcome the restrictions of single-band screening method, such as FLV-PLS. In fact, the protein molecules, such as albumin and globulin, contain a lot of functional groups having rich absorption information on MIR region, and their absorption bands were usually the relative dispersion. Therefore, as a multi-band slection method, CCO-PLS represented high applicability for quantitative analysis of albumin and globulin.

Model validation

The randomly selected validation samples, which were excluded from the modeling optimization process, were used to validate the optimal FLV-PLS and CCO-PLS models. The predicted ALB and GLB values of the validation samples were then calculated. Fig. 7 and 8 show the relationship between the predicted and clinically measured values of the 84 validation samples for ALB and GLB

obtained with the optimal FLV-PLS and CCO-PLS models, respectively. The validation effects (SEP and R_P) are summarized in Table 6. The results indicate that the two methods both have good prediction accuracy and very high correlation for the clinically measured values. The error of CCO-PLS method for clinically measured values was obviously smaller than that of FLV-PLS method. The another advantage of CCO-PLS method is the performance of multi-band screening, it could be used in quantitative analysis of the object that relative dispersion of the molecular absorption bands.

Table 6 Validation effects for FLV-PLS and CCO-PLS method for ALB and GLB (g L^{-1}).

Indicator	SEP	R_P
FLV-PLS method		
ALB	1.51	0.979
GLB	1.52	0.955
CCO-PLS method		
ALB	1.36	0.981
GLB	1.35	0.965

Conclusions

Most of the wavelength selection methods are appropriate algorithms implemented within a continuous waveband. Because of the difficulties on the algorithm, an effective method for multi-band selection is still limited. Based on high correlation between MIR spectroscopy and component content, the correlation coefficient optimization coupled with PLS (CCO-PLS) method was proposed for multi-band selection, and was successfully applied to the ATR-FTIR analysis of albumin and globulin in human serum. We believe that the methodological framework and the algorithm platforms can be used for other analytical objects, especially for the complex systems.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 61078040), the Science and Technology Project of Guangdong Province of China (No. 2012B031800917), and the Science and Technology Project of Guangzhou of China (No. 2014Y2-00002).

References

1 F. M. Mirabella and N. J. Harrick, *Internal reflection spectroscopy: review and supplement*, Harrick Scientific Corporation, New York, USA, 1985.
2 Y. J. Kim and G. Yoon, *Appl. Spectrosc*, 2002, **56**, 625−632.
3 K. Nakanishi, A. Hashimoto, T. Pan and T. Kameoka, *Appl. Spectrosc*, 2003, **57**(12), 1510−1516.
4 E. Diessel, S. Willmann, P. Kamphaus, R. Kurte, U. Damm and HM. Heise, *Appl. Spectrosc*, 2004, **58**(4), 442−450.
5 T. Pan, A. Hashimoto, M. Kanou, K. Nakanishi and T. Kameoka, *Bioprocess Biosyst. Eng*., 2003, **26**, 133−139.
6 T. Pan, A. Hashimoto, M. Kanou, K. Nakanishi and T. Kameoka, *Japan Journal of Food Engineering*, 2004, **5**(1), 22−31.
7 E. Diessel, P. Kamphaus, K. Grothe, R. Kurte, U. Damm and HM. Heise, *Appl. Spectrosc*, 2005, **59**(4), 442−451.
8 J. M. Chen, T. Pan and X. D. Chen, *Opt. Precis. Eng*., 2006, **14**(1), 1−7.
9 M. Pleitez, H. von Lilienfeld-Toal and W. Mantele, *Spectrochim. Acta A*, 2012, **85**(1), 61−65.
10 J. H. Wan, P. L. Tian, H. Yin, Y. Han, X. C. Wei and T. Pan, *Clin. Biochem*., 2013, **46**, 128−132.
11 S. Kino, Y. Tanaka and Y. Matsuura, *J. Biomed. Opt*., 2014, **19**(5), 057010.
12 X. L. Long, G. S. Liu, T. Pan and J. M. Chen, *J. Biomed. Opt*., 2014, **19**(8), 087004.
13 Y. P. Du, Y. Z. Liang, J. H. Jiang, R. J. Berry and Y. Ozaki, *Anal. Chim. Acta*., 2004, **501**, 183−197.
14 H. Z. Chen, T. Pan, J. M. Chen and Q. P. Lu, *Chemometr. Intell. Lab*., 2011, **107**, 139−146.
15 T. Pan, Z. H. Chen, J. M. Chen and Z. Y. Liu, *Anal. Methods*, 2012, **4**, 1046−1052.
16 Z. Y. Liu, B. Liu, T. Pan and J. D. Yang, *Spectrochim. Acta A*, 2013, **102**, 269−274.
17 T. Pan, J. M. Liu, J. M. Chen, G. P. Zhang and Y. Zhao, *Anal. Methods*, 2013, **5**, 4355−4362.
18 T. Pan, M. M. Li and J. M. Chen, *Appl. Spectrosc*., 2014, **68**, 263−271.
19 V. Menon, T. Greene and X. L. Wang, *Kidney. Int*., 2005, **68**(2), 766−772.
20 H. Malonga, J.F. Neauly, and H. Arakawa, *DNA Cell Bio*., 2006, **25**(1), 63−68.
21 F.M. Wasacz, J.M. Olinger, and R.J. Jakobsen, *Biochemistry*., 1987, **26**(5), 1464−1470.

Figure legend

Figure. 1 ATR-FTIR spectra of 194 human serum samples for (a) entire scanning region 4,000-600 cm^{-1} and (b) MIR fingerprint region 1,700-900 cm^{-1}

Figure. 2 Correlation coefficient spectra for (a) ALB and (b) GLB.

Figure. 3 Correlation coefficient spectra at 1,700-900cm^{-1} region which showed the positive peaks for (a) ALB and (b) GLB.

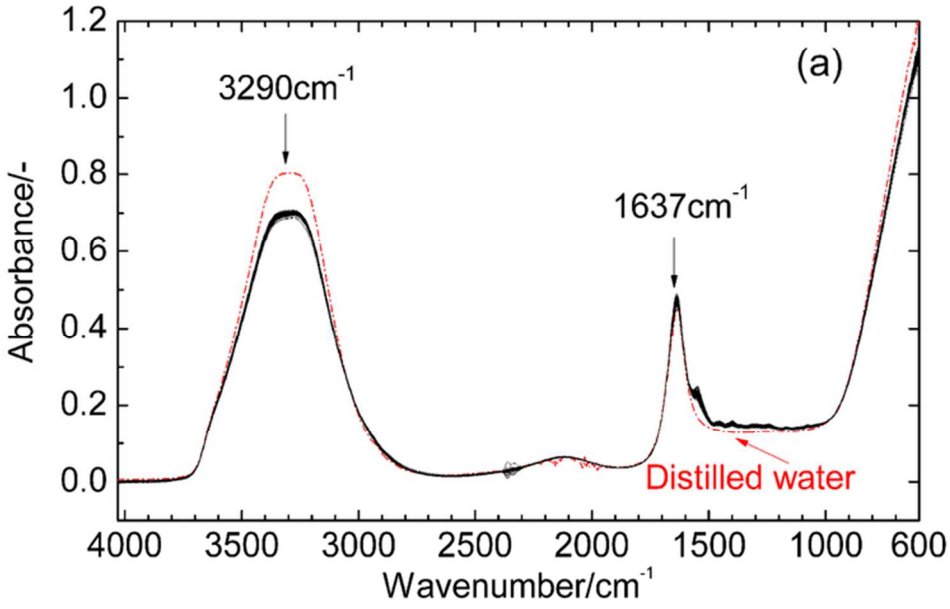
Figure. 4 First loading vector for ALB and GLB (Water absorption band 1,697-1,591 cm^{-1} is not shown).

Figure. 5 SECV corresponding to each upper bound of correlation coefficient for (a) ALB and (b) GLB.

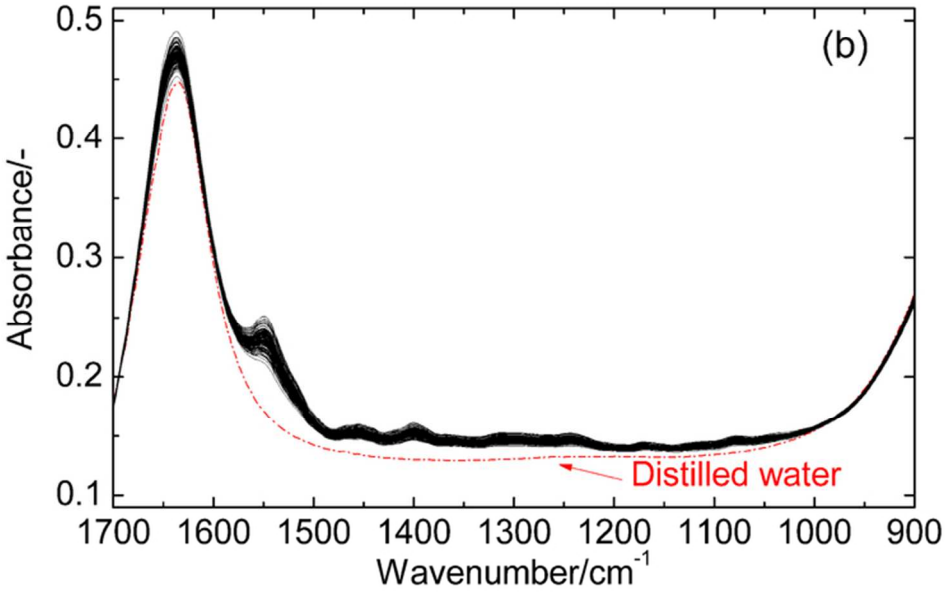
Figure. 6 Distribution of wavebands combinations with CCO-PLS method for (a) ALB and (b) GLB.

Figure. 7 Relationship between predicted and measured values of validation samples with FLV-PLS method for (a) ALB and (b) GLB.

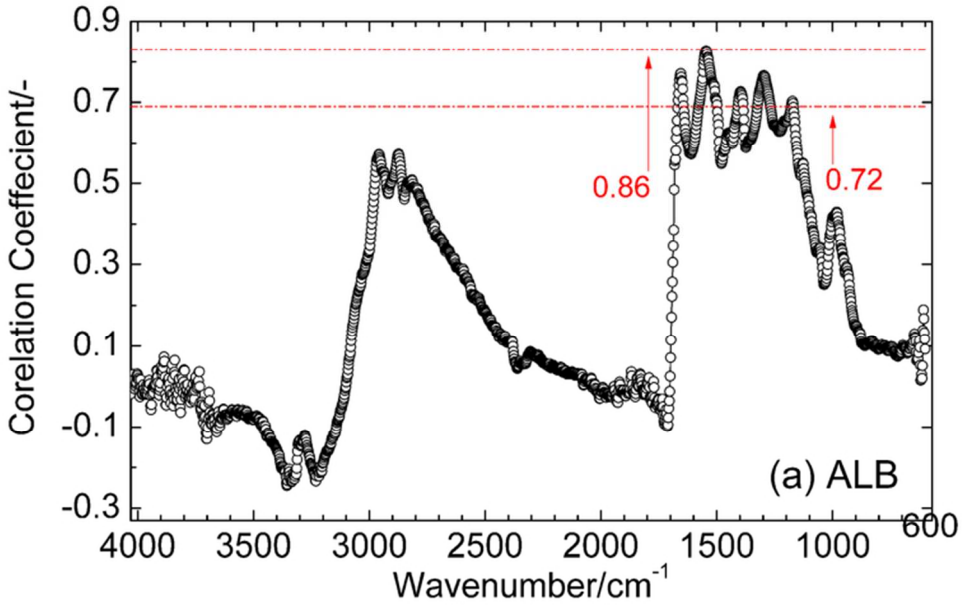
Figure. 8 Relationship between predicted and measured values of validation samples with CCO-PLS method for (a) ALB and (b) GLB.



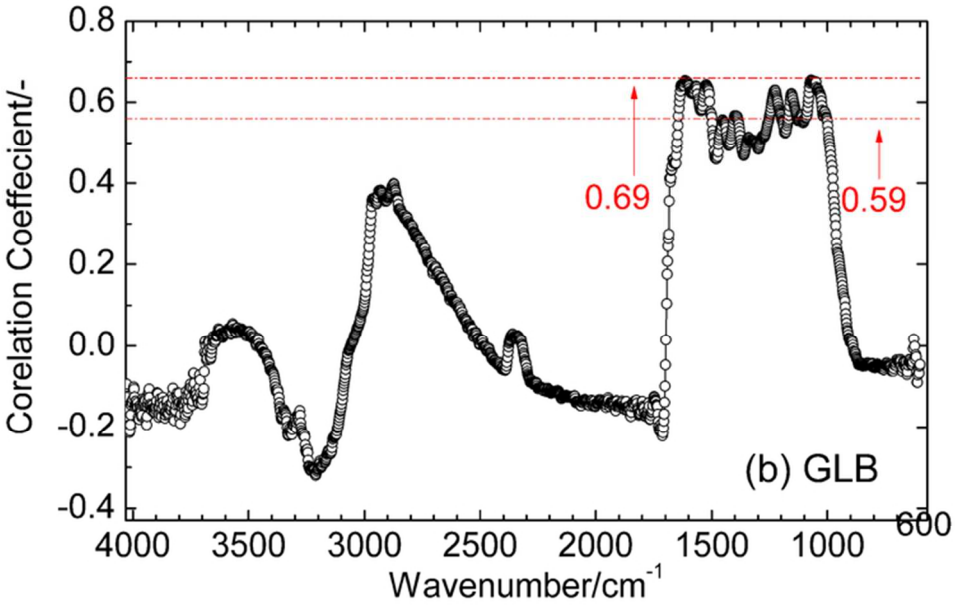
66x44mm (300 x 300 DPI)



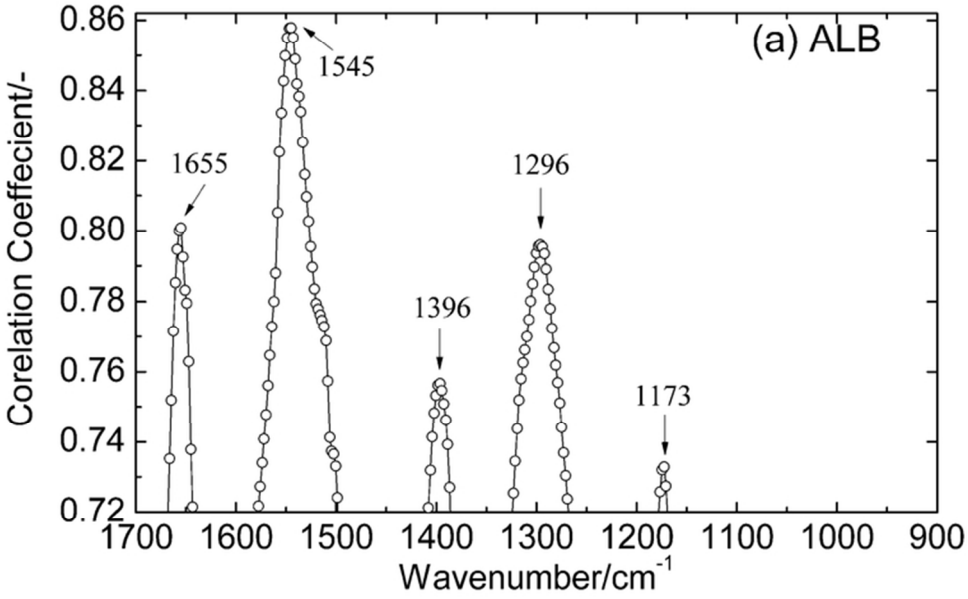
66x44mm (300 x 300 DPI)



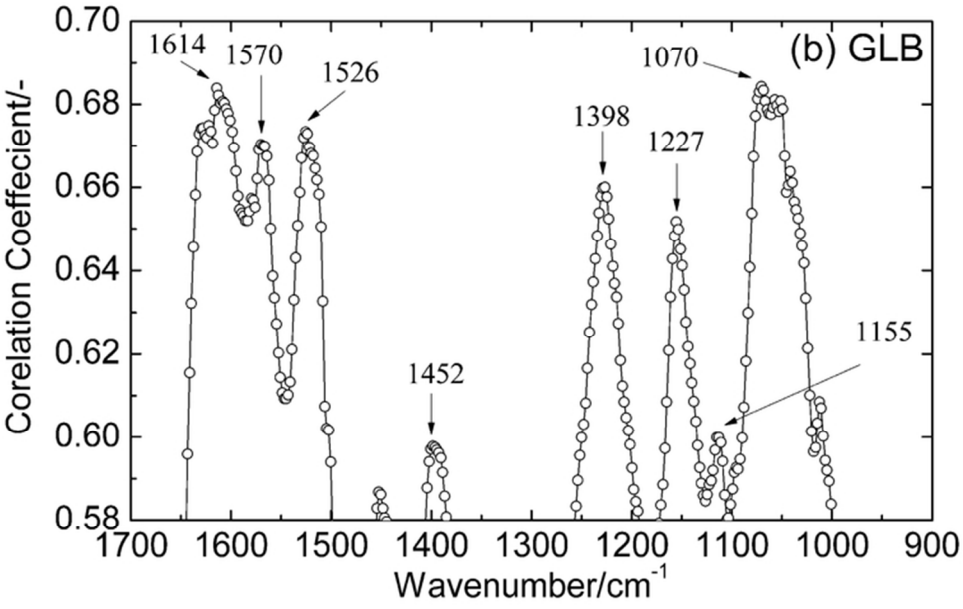
66x44mm (300 x 300 DPI)



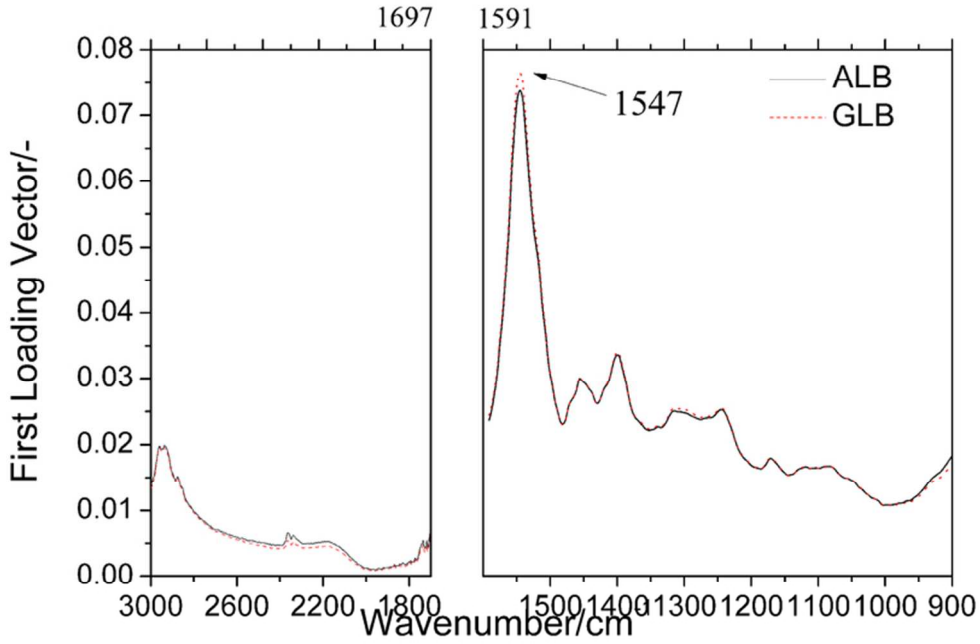
66x44mm (300 x 300 DPI)



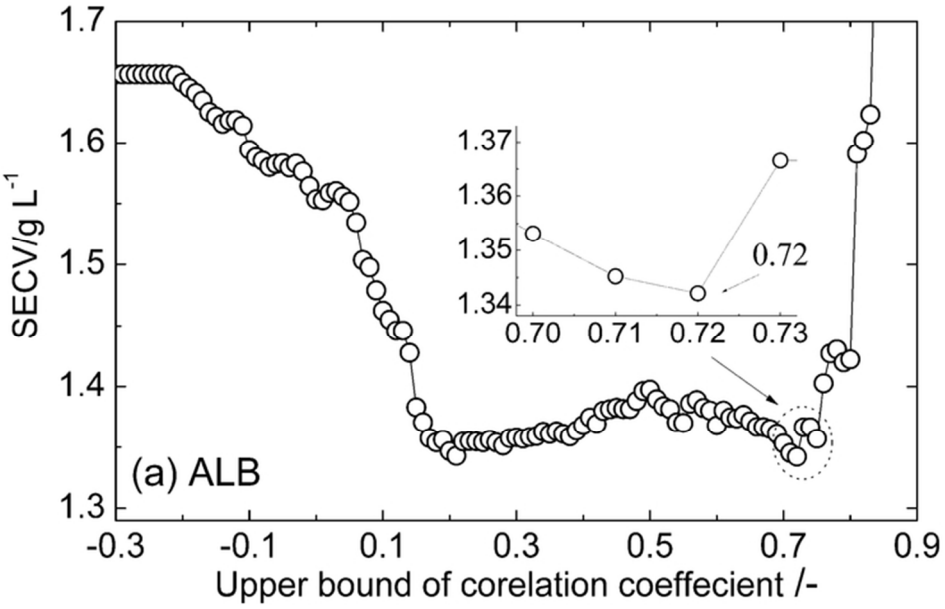
66x44mm (300 x 300 DPI)



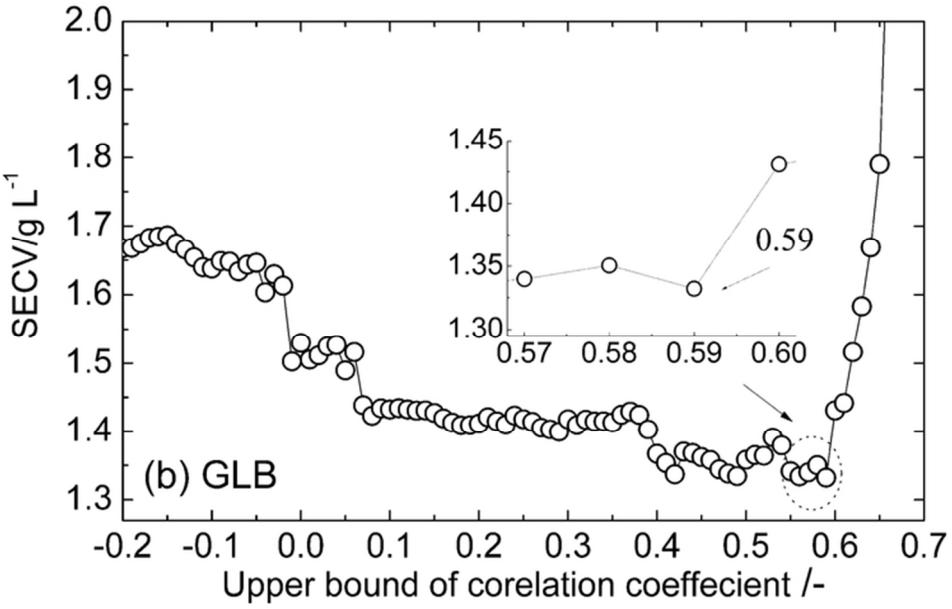
66x44mm (300 x 300 DPI)



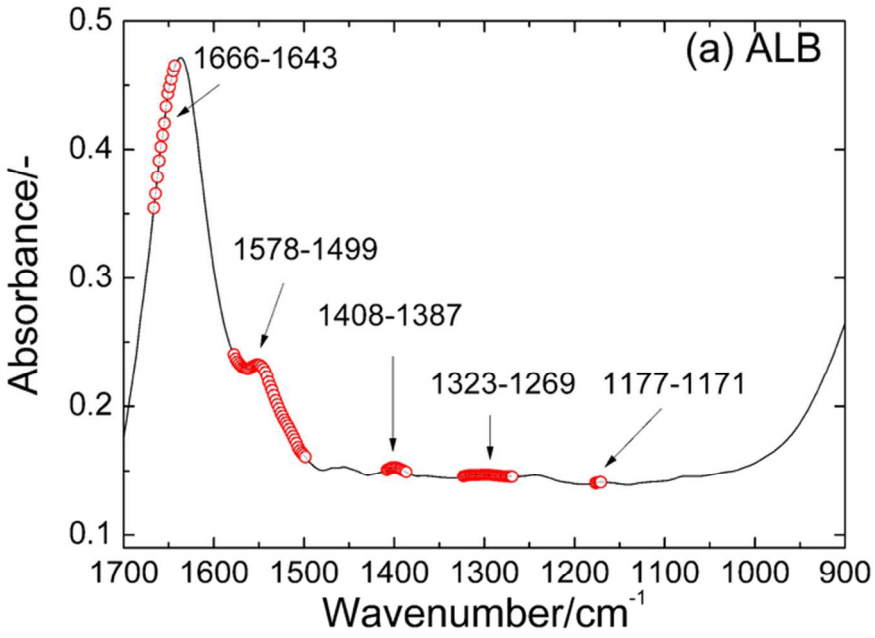
70x49mm (300 x 300 DPI)



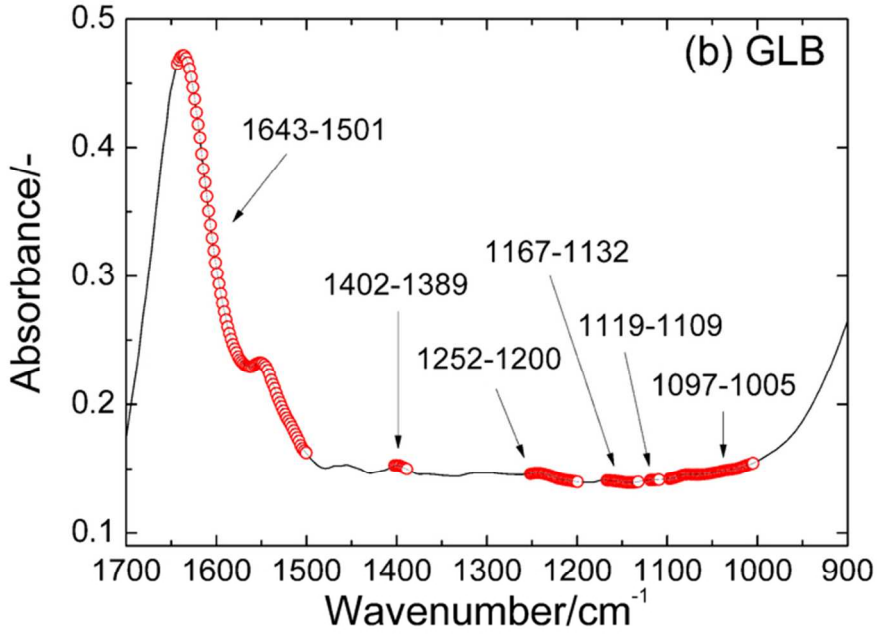
66x44mm (300 x 300 DPI)



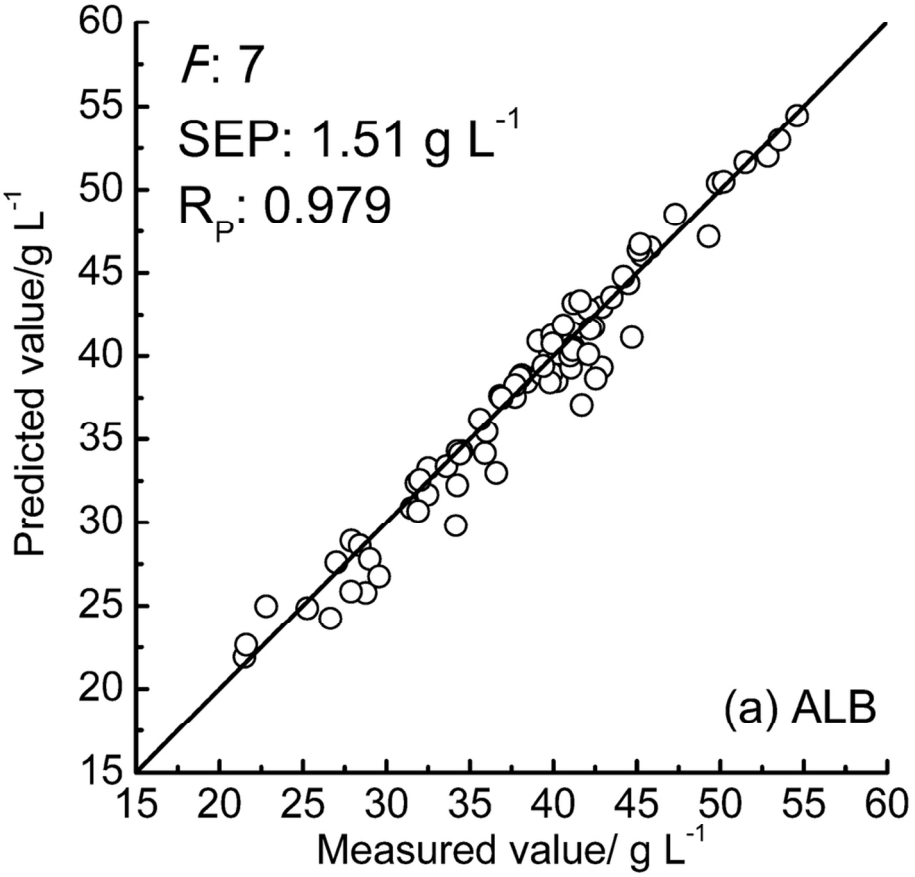
66x44mm (300 x 300 DPI)



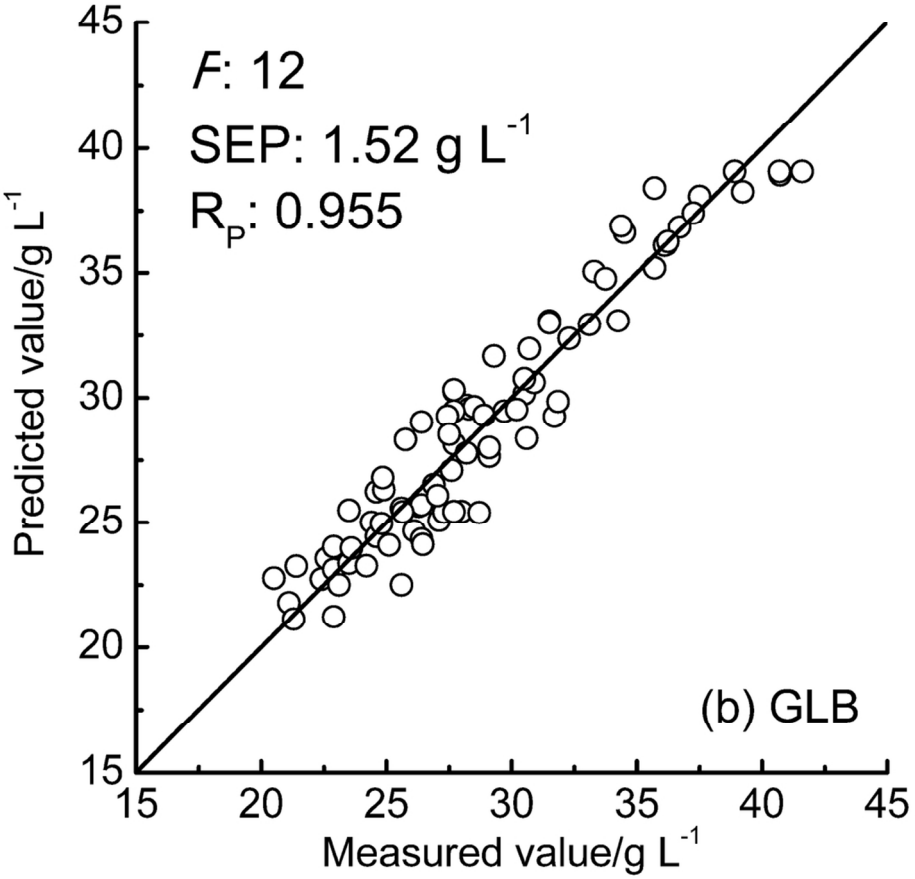
70x49mm (300 x 300 DPI)



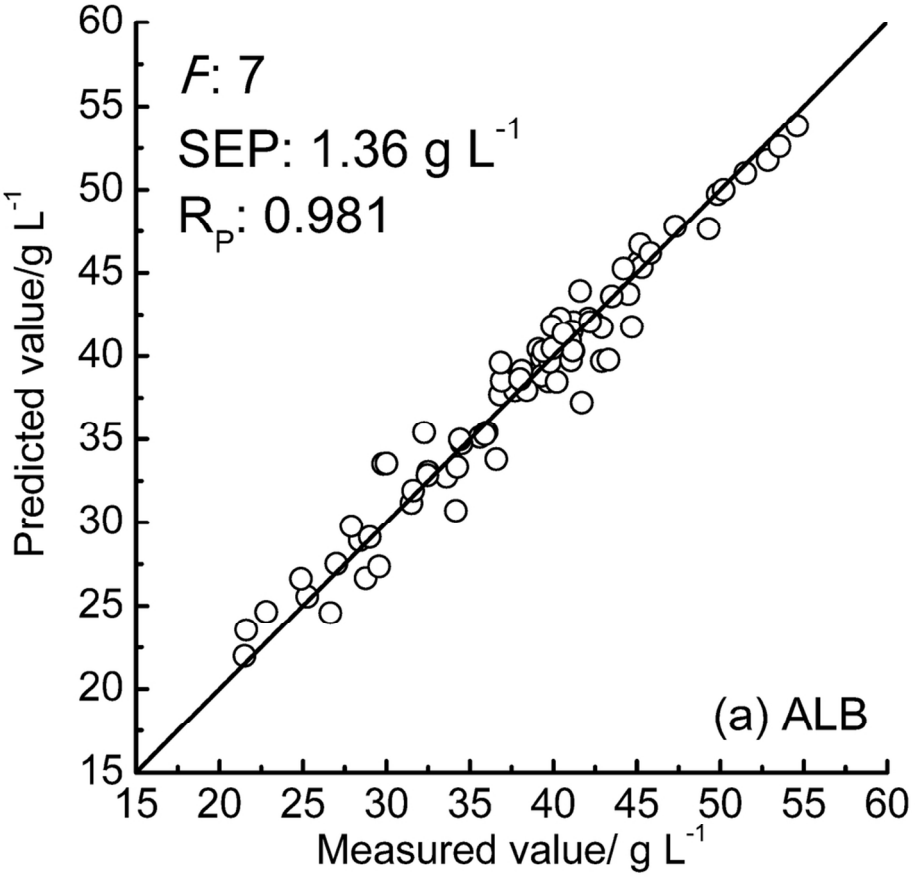
70x49mm (300 x 300 DPI)



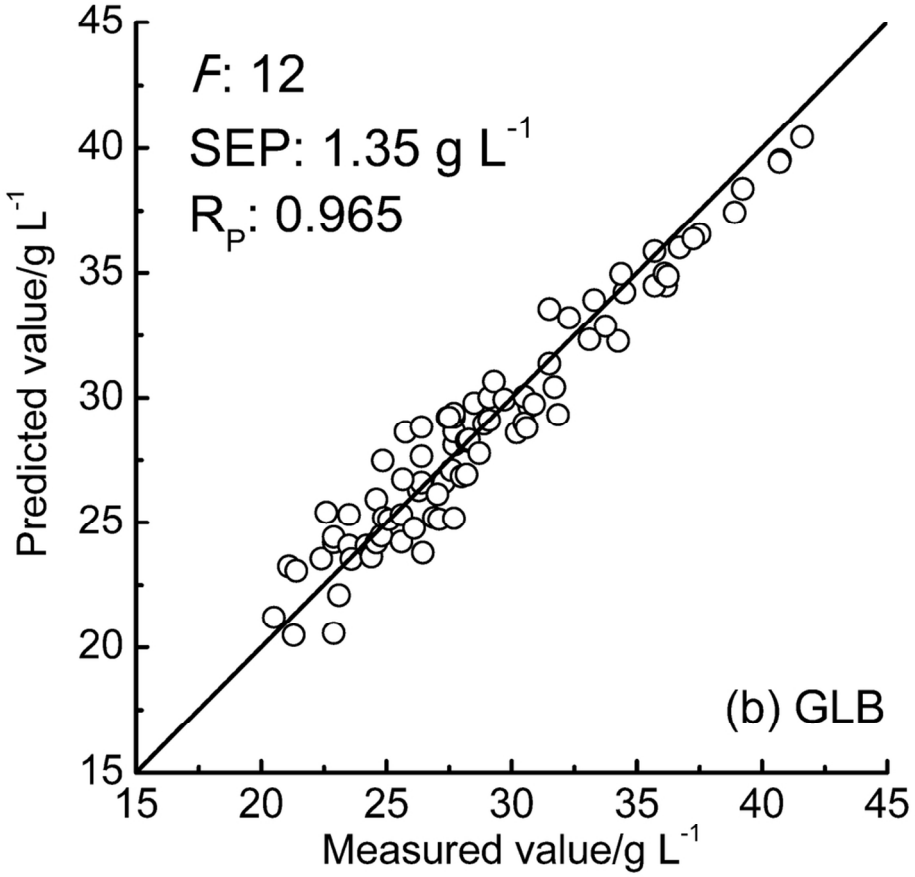
99x99mm (300 x 300 DPI)



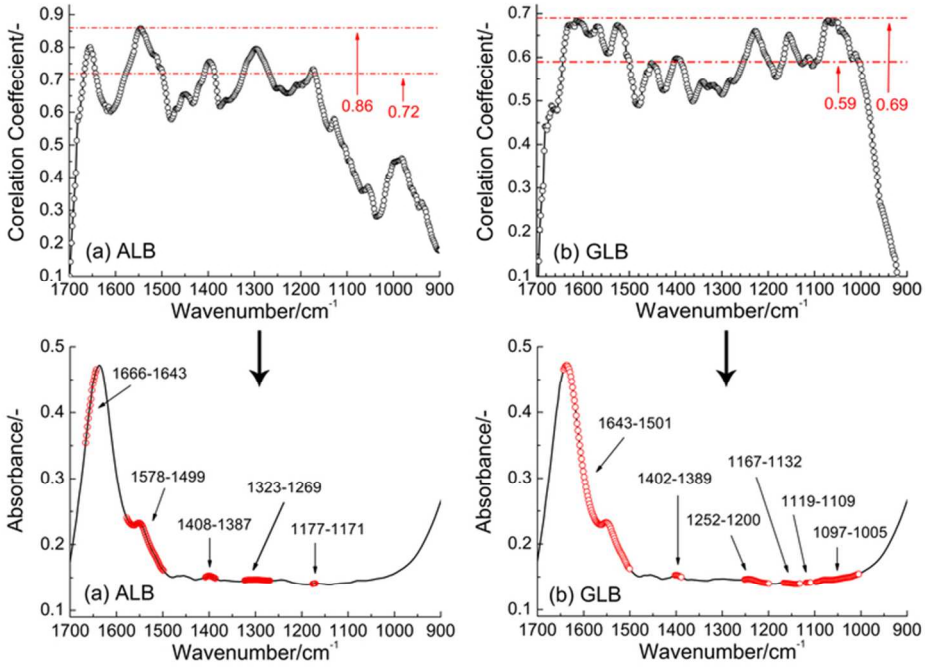
99x99mm (300 x 300 DPI)



99x99mm (300 x 300 DPI)



99x99mm (300 x 300 DPI)



70x50mm (300 x 300 DPI)