

Optimal Partner Wavelength Combination Method Applied to NIR Spectroscopic Analysis of Human Serum Globulin

Yun Han (✉ hanyun0539@163.com)

Guangdong Ocean university

Yun Zhong

Zhanjiang No.2 High School Hai Dong

Huihui Zhou

Guangdong Ocean University

Xuesong Kuang

Guangdong Ocean University

Research article

Keywords: Optimal partner wavelength combination, Near-infrared spectroscopy, Human serum globulin

Posted Date: March 13th, 2020

DOI: <https://doi.org/10.21203/rs.2.20972/v2>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Chemistry on May 24th, 2020. See the published version at <https://doi.org/10.1186/s13065-020-00689-z>.

Abstract

Human serum globulin (GLB), which contains various antibodies in healthy human serum, is of great significance for clinical trials and disease diagnosis. In this study, the GLB in human serum was rapidly analyzed by near infrared (NIR) spectroscopy without chemical reagents. Optimal partner wavelength combination (OPWC) method was employed for selecting discrete information wavelength. For the OPWC, the redundant wavelengths were removed by repeated projection iteration based on binary linear regression, and the result converged to stable number of wavelengths. By the way, the convergence of algorithm was proved theoretically. Moving window partial least squares (MW-PLS), a well-performed wavelength selection method, was also performed for comparison.

The optimal models were obtained by the two methods, and the corresponding root-mean-square error of cross validation and correlation coefficient of prediction (SECV, R_{PCV}) were $0.813 \text{ g} \cdot \text{L}^{-1}$ and 0.978 with OPWC combined with PLS (OPWC-PLS), and $0.804 \text{ g} \cdot \text{L}^{-1}$ and 0.979 with MW-PLS, respectively. The two methods achieved almost the same good results. However, the OPWC only contained 28 wavelengths, so it had obvious lower model complexity. Thus it can be seen that the OPWC-PLS has great prediction performance for GLB and its algorithm is convergent and rapid. The results provide important technical support for the rapid detection of serum.

Introduction

Near infrared (NIR) spectroscopy is a green and developing analytical technique, which has been widely used in life sciences,¹⁻⁷ agricultural products and food,⁸⁻¹¹ soil,¹²⁻¹⁴ and other fields.^{15,16} For NIR spectroscopic analysis of complex system, wavelength selection is necessary and difficult. So far, many methods including continuous mode and discrete mode of wavelength selection have been successfully used in NIR spectroscopy analysis, but a general and effective method has not been found. Moving window partial least squares (MW-PLS) is a method that uses a moving window whose position and size can be changed to identify and select continuous wavebands in terms of the prediction effect, and such waveband can correspond to absorption of specific functional groups.^{13,15,16-19} Therefore, the MW-PLS is a method with good chemical interpretation. However, it can be seen from the papers [17-19], as a traversal algorithm for continuous wavebands, all possible continuous bands are screened, this method is time-consuming when encountering a large dataset.

Serum globulin (GLB), which is synthesized by human monocyte-phagocyte system, contains various antibodies in the serum of healthy people, so it can enhance the body's resistance to prevent infection. It is mainly used for immunodeficiency diseases as well as prevention and treatment of viral infections and bacterial infections such as infectious hepatitis, measles, chickenpox, mumps and herpes zoster. In addition, it can also be used in asthma, allergic rhinitis, eczema and other endogenous allergic diseases. Therefore, the GLB in human serum is very important for clinical trials and disease diagnosis. In previous studies,^{20,21} FTIR/ATR spectroscopy was used for determination of GLB. The study found that for blood index, the NIR has higher quantitative analysis accuracy than the FTIR/ATR spectroscopy.^{6,22} The

experimental results show that the molecular absorption information of GLB can be captured by NIR spectroscopy without reagent.

Optimal partner wavelength combination (OPWC) is a method of selecting discrete information wavelength by iteration. For the method, the best partner of each wavelength in a predetermined wavelength region was determined based on binary linear regression (BLR), and a partner wavelength subset (PWS) was obtained; then the best partner of each wavelength in the PWS was obtained with the same method. The iterative process may be continued until convergence was met, and the last obtained wavelength subset was called OPWC. On the basis of the OPWC, PLS model was established. In order to make full use of the samples, the leave-one-out cross validation (LOOCV) was adopted.

Because human serum is a complex multi-component system and the absorption interference of other components is very complex, it is difficult to extract the characteristic information of GLB. Therefore, OPWC-PLS method was employed to remove redundant wavelength and establish a high precision quantitative model. MW-PLS, which is a popular wavelength selection method, was also performed for comparison. Experimental results showed that the OPWC-PLS has great prediction performance and the algorithm is convergent and rapid.

Materials And Methods

Due to technical limitations, the Materials and Methods section is only available as a download in the supplementary files section.

Results And Discussion

3.1. Results with MW-PLS

The NIR spectra of 230 human serum samples in the scanning area (780-2498 nm) were shown in Figure 1. As can be seen from the figure, absorption at about 2000 nm and 2400 nm has obviously strong noise. In order to obtain satisfactory results, wavelength selection must be carried out to overcome noise interference. For comparison, PLS model of the full spectrum region was first established. The corresponding SECV and R_{PCV} were 1.423 g L⁻¹ and 0.935, respectively.

MW-PLS method was performed to optimize waveband and improve prediction accuracy. Depending on minimum SECV value, the optimal MW-PLS model was selected out. The corresponding waveband was 1504 to 1820 nm, located in the long-NIR region (1100 to 2498 nm). Prediction effects (SECV and R_{PCV}) and parameters of the above two methods were summarized in Table 2. The results showed that the predicted values were highly correlated with clinical measurements for the two methods, and comparing with optimal PLS model in full spectrum region, the optimal MW-PLS model achieved better prediction effect with fewer wavelengths.

3.2. Results with OPWC-PLS

The OPWC method was performed for screening information wavelength based on the steps mentioned in section 2.4. Firstly, 104 best partners for all 860 wavelengths were determined according to the results of LOOCV-BLR analysis, and PWS⁽¹⁾ with 104 wavelengths was obtained. Thus, the number of wavelengths was greatly reduced after the first projection. The correspondence between all 860 wavelengths and their best partners was shown in Figure 2. As shown in the figure, some wavelengths had the same best partner, such as the 2156 nm and 2190 nm as best partners of other wavelengths appeared 3 and 8 times, respectively, so projection was not a one-to-one mapping function in the whole spectral region. Obviously, was a subset of and the projection continues.

Based on the corresponding relationship determined above, the best partner of was easy to be selected, and the PWS⁽²⁾ was obtained. Repeated the same process for PWS⁽²⁾, and PWS⁽³⁾ was obtained. As the projection progresses, the number of wavelengths decreased gradually until the number of wavelengths for PWS⁽⁶⁾ no longer changed. The PWS⁽⁶⁾ was the OPWC and it had only 28 wavelengths. Figure 3 showed the 28 wavelengths and their best partners. As the figure showed, the 28 wavelengths are divided into 14 groups, and the two wavelengths in each group are the best partners for each other.

Based on PLS, the LOOCVs were performed for every PWS, and the corresponding minimum SECV value and number of wavelengths ($N^{(s)}$) used are shown in Figure 4. As shown in the figure, the $N^{(s)}$ and minimum SECV values have almost the same trend. After the first projection, both of them decrease rapidly, and the remaining wavelengths are more important, so as the number of projections increases, they slowly decrease. This may be due to the removal of a large amount of noise and background information from the original spectrum after the first projection, so both the $N^{(s)}$ and minimum SECV values decrease rapidly. The partner wavelength subset of the original spectrum contains less redundant information, so the $N^{(s)}$ and minimum SECV values decrease slowly in the later projection iteration.

3.3 Comparison of the two methods

Screening the information wavelengths of GLB in the human serum of a multi-component complex system is difficult and complicated. The wavelengths selected by the OPWC-PLS and MW-PLS methods, which correspond to the information of GLB, were shown in Figure 5. As indicated in Figure 5, the wavelengths selected by the OPWC method have a wider distribution range and partially coincides with the wavelengths selected by MW-PLS. This may be because the local characteristics of MW-PLS method make some wavelengths cannot be detected, which reflects the complexity of NIR model optimization and the commonness and difference of different methods.

Figure 6 showed the relationship between the predicted and measured GLB values based on the MW-PLS and OPWC-PLS methods, respectively. The prediction effect and corresponding parameters N and F were summarized in Table 2. The SECV and R_{PCV} were 0.813 g L⁻¹ and 0.978 with OPWC-PLS, and 0.804 g L⁻¹ and 0.979 with MW-PLS, respectively. The results show that, like MW-PLS, the prediction effect of OPWC-PLS was also obviously better than that of the whole spectrum PLS, and the OPWC is an effective method for screening wavelengths. The phenomenon conveys that better prediction results can be

achieved with fewer wavelengths. Thus one can conclude that it is very necessary to first perform wavelength selection before building a calibration model. The two methods had achieved almost the same good prediction results (SECV and $R_{P, CV}$). However, the optimal OPWC-PLS model adopted only 28 wavelengths, while the other adopted 159 wavelengths. Therefore, the OPWC method has great prediction performance for wavelength selection.

The differences in prediction of the OPWC-PLS and MW-PLS methods for GLB illustrate that MW-PLS can achieve higher prediction accuracy, but it is time-consuming and employs more wavelengths, while OPWC-PLS can achieve similar prediction results with MW-PLS in less time. In addition, MW-PLS, as a continuous wavelength screening method, is more suitable for determining the object with relatively concentrated molecular absorption bands; while OPWC-PLS, as a discrete wavelength screening method, may be more suitable for determining the object with relatively fragmented molecular absorption bands.

Conclusion

The change of GLB content in human serum has important reference value for clinical trial and disease diagnosis. In this study, the OPWC-PLS method was employed for rapid analysis of GLB based on NIR spectroscopy. MW-PLS, a well-performed variable selection method, was also employed for comparison. The optimal OPWC-PLS model adopted 28 wavelengths, and corresponding SECV and $R_{P, CV}$ were 0.813 g L⁻¹ and 0.978, respectively. The optimal MW-PLS model adopted 159 wavelengths, and corresponding SECV and $R_{P, CV}$ for were 0.804 g L⁻¹ and 0.979, respectively. The results indicate that, both methods achieved satisfactory prediction results, and OPWC-PLS achieved almost the same prediction effect as MW-PLS with faster speed and fewer wavelengths. Therefore, OPWC is an efficient approach for information wavelength selection.

The predicted GLB values obtained by MW-PLS and OPWC-PLS were highly correlated with the measured values obtained by traditional methods. Compared with traditional method, the method based on NIR spectroscopy has the merits of rapidity, simplicity and no chemical reagent. Therefore, the results have important reference value for the rapid determination of GLB. In addition, the wavelengths selected by the two methods are partially the same, reflecting the commonness and difference of different methods.

Declarations

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Youth Innovation Talents Project of Colleges and Universities in Guangdong Province (No. Q18285) and Guangdong Ocean University Scientific Research Start-up Funding for the Doctoral Program (No. R17057).

Authors' contributions

HY analyzed the spectral data of human serum samples and optimized the wavelength model, and was a major contributor in writing the manuscript. ZY and ZHH carried out the spectrum experiment. KXS performed model validation. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Abbreviations

GLB: Globulin; NIR: Near infrared; OPWC: Optimal partner wavelength combination; MW-PLS: Moving window partial least squares; SECV: Root-mean-square error of cross validation of prediction; R_{PCV} : Correlation coefficient of prediction; BLR: Binary linear regression; PWS: Partner wavelength subset; LOOCV: Leave-one-out cross validation; SD: Standard deviation.

References

1. Chen JM, Peng LJ, Han Y, et al. A rapid quantification method for the screening indicator for β -thalassemia with near-infrared spectroscopy. *Acta A*. 2018, 193: 499-506.
2. Han Y, Pan T, Zhou HH, Yuan R. ATR-FTIR spectroscopy with equidistant combination PLS method applied for rapid determination of glycosylated hemoglobin. *Methods*. 2018, 10:3455-3461.
3. Yao LJ, Tang Y, Yin ZW, et al. Repetition rate priority combination method based on equidistant wavelengths screening with application to NIR analysis of serum albumin. *Intell. Lab. Syst.* 2017, 162: 191-196.
4. Han Y, Chen JM, Pan T, Liu G Determination of glycosylated hemoglobin using near-infrared spectroscopy combined with equidistant combination partial least squares. *Chemom. Intell. Lab. Syst.* 2015, 145: 84-92.
5. Lee Y, Lee S, In JY, et al. Prediction of plasma hemoglobin concentration by near-infrared spectroscopy. *Korean Med. Sci.* 2008, 23:674-677.
6. Pan T, Liu JM, Chen JM, et al. Rapid determination of preliminary thalassaemia screening indicators based on near-infrared spectroscopy with wavelength selection stability. *Methods*. 2013, 5(17): 4355-4362.

7. Yao LJ, Lyu N, Chen JM, et al. Joint analyses model for total cholesterol and triglyceride in human serum with near-infrared spectroscopy. *Acta A*. 2016, 159: 53-59.
8. Lyu N, Chen JM, Pan T, et al. Near-infrared spectroscopy combined with equidistant combination partial least squares applied to multi-index analysis of corn. *Infrared Physics & Technology*. 2016, 76: 648-654.
9. Guo HS, Chen JM, Pan T, et al. Vis-NIR wavelength selection for non-destructive discriminant analysis of breed screening of transgenic sugarcane. *Methods*. 2014, 6(10): 8810-8816.
10. Chen JY, Iyo C, Kawano S. Effect of multiplicative scatter correction on wavelength selection for near infrared calibration to determine fat content in raw milk. *Near Infrared Spec*. 2002, 10(4): 301-307.
11. Liu ZY, Liu B, Pan T, et al. Determination of amino acid nitrogen in tuber mustard using near-infrared spectroscopy with waveband selection stability. *Acta A*. 2013, 102: 269-274.
12. Pan T, Li MM, Chen JM. Selection method of quasi-continuous wavelength combination with applications to the near-infrared spectroscopic analysis of soil organic matter. *Spectrosc*. 2014, 68(3): 263-271.
13. Pan T, Han Y, Chen JM, et al. Optimal partner wavelength combination method with application to near-infrared spectroscopic analysis. *Intell. Lab. Syst*. 2016, 156: 217-223.
14. Chen JM, Pan T, Liu GS, et al. Selection of Stable Equivalent Wavebands for Near-Infrared Spectroscopic Analysis of Total Nitrogen in Soil. *J. Innov. Opt. Health Sci*. 2014, 7(4):1-9.
15. Pan T, Chen ZH, Chen JM, et al. Near-Infrared Spectroscopy with Waveband Selection Stability for the Determination of COD in Sugar Refinery Wastewater. *Anal. Methods*. 2012, 4 (4):1046-1052.
16. Li HD, Liang YZ, Xu QS, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Chim. Acta*. 2009, 648: 77-84.
17. Jiang JH, Berry RJ, Siesler HW, et al. Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data, *Chem*. 2002, 74:3555-3565.
18. Du YP, Liang YZ, Jiang JH, et al. 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*. 2004, 501(2): 183-191.
19. Li HD, Liang YZ, Xu QS, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration, *Chim. Acta*. 2009, 648: 77-84.
20. Chen YF, Chen JM, Pan T, et al. Correlation coefficient optimization in partial least squares regression with application to ATR-FTIR spectroscopic analysis. *Methods*. 2015, 7: 5780-5786.
21. Kim YJ, Yoon G. Multicomponent assay for human serum using mid-infrared transmission spectroscopy based on component-optimized spectral region selected by a first loading vector analysis in partial least-squares regression. *Spectrosc*. 2002, 56(5):625-632.
22. Long XL, Liu GS, Pan T, et al. Waveband selection of reagent-free determination for thalassemia screening indicators using Fourier transform infrared spectroscopy with attenuated total reflection.

Tables

Table 1 Statistical analysis of measured GLB values of 230 samples

Indicator	Number	Min	Max	Mean	SD
GLB(g L ⁻¹)	230	18.70	41.60	27.47	73.95

Table 2 Prediction effects of three methods

Methods	Adopted wavelengths (nm)	<i>N</i>	<i>F</i>	SECVR _{P,CV}
PLS	780-2498	860	151.42	30.935
MW-PLS	1504-1820	159	100.80	40.979
OPWC-PLS	1410, 1534, 1536, 1538, 1542, 1676, 1678, 1698, 1732, 1734, 1738, 1742, 1744, 1746, 1750, 1870, 2128, 2132, 2218, 2220, 2222, 2228, 2254, 2258, 2306, 2310, 2318, 2340	28	7	0.8130.978

Figures

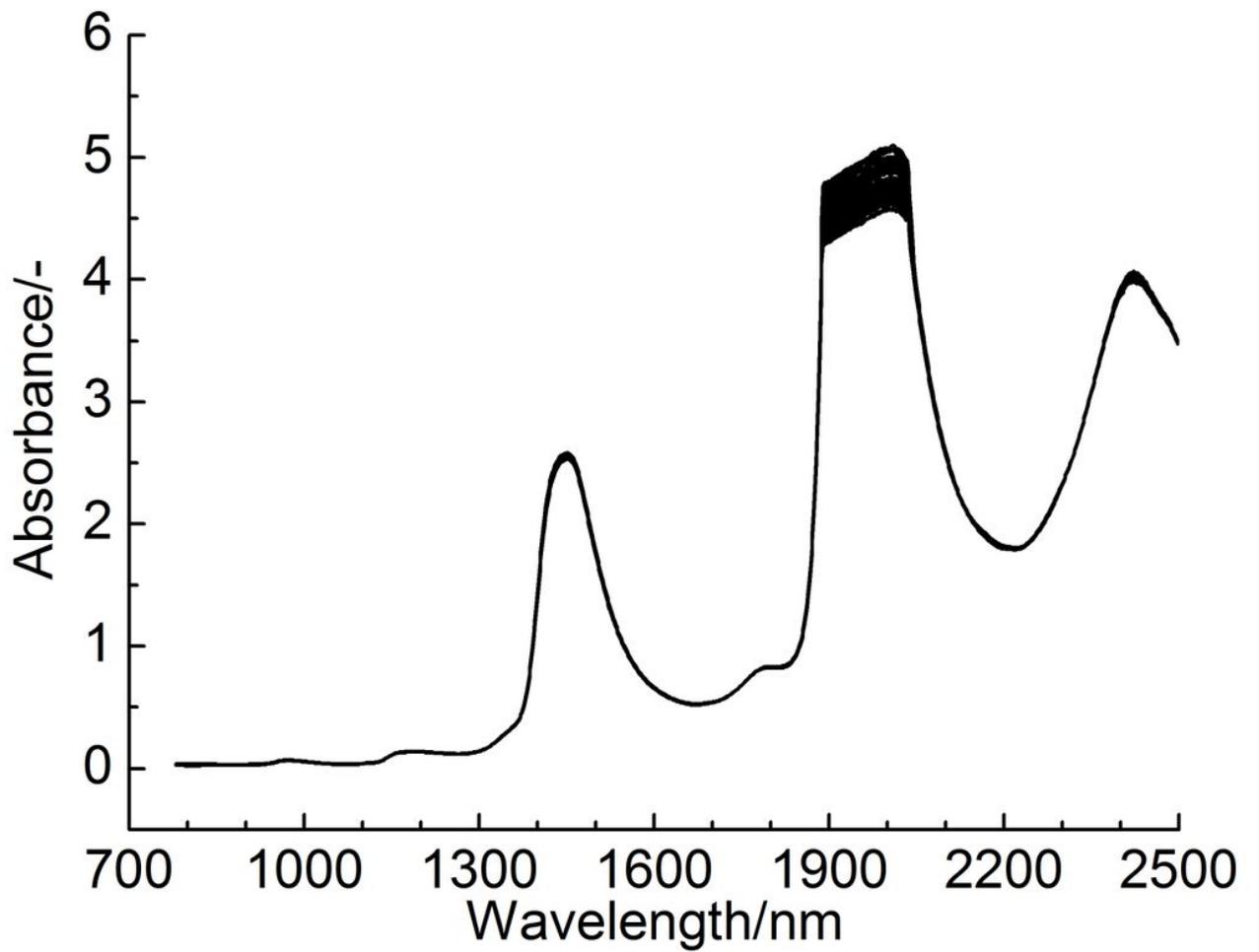


Figure 1

NIR spectra of 230 human serum samples in the scanning area (780-2498 nm).

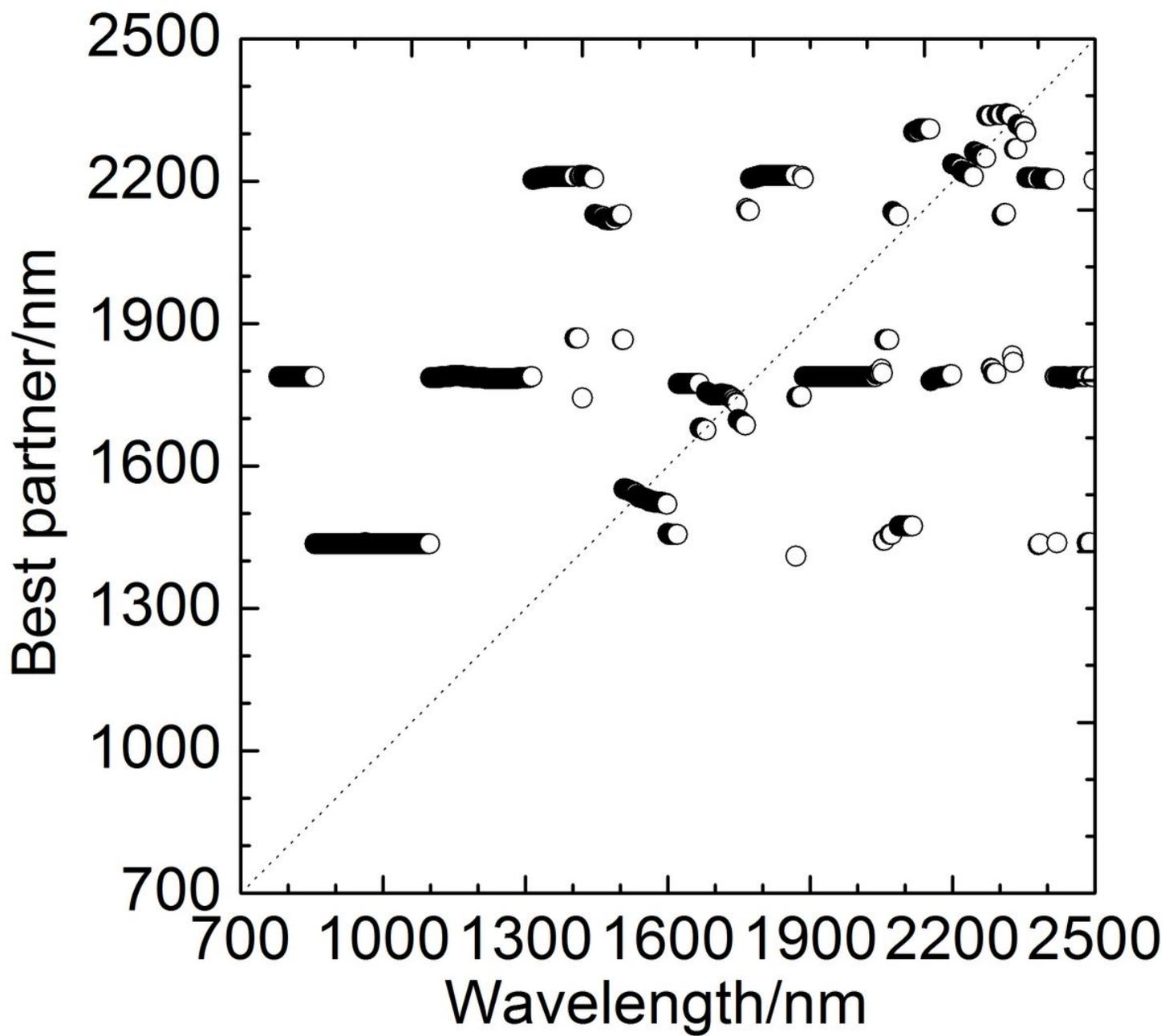


Figure 2

Best partners of 860 wavelengths in the full spectrum region.

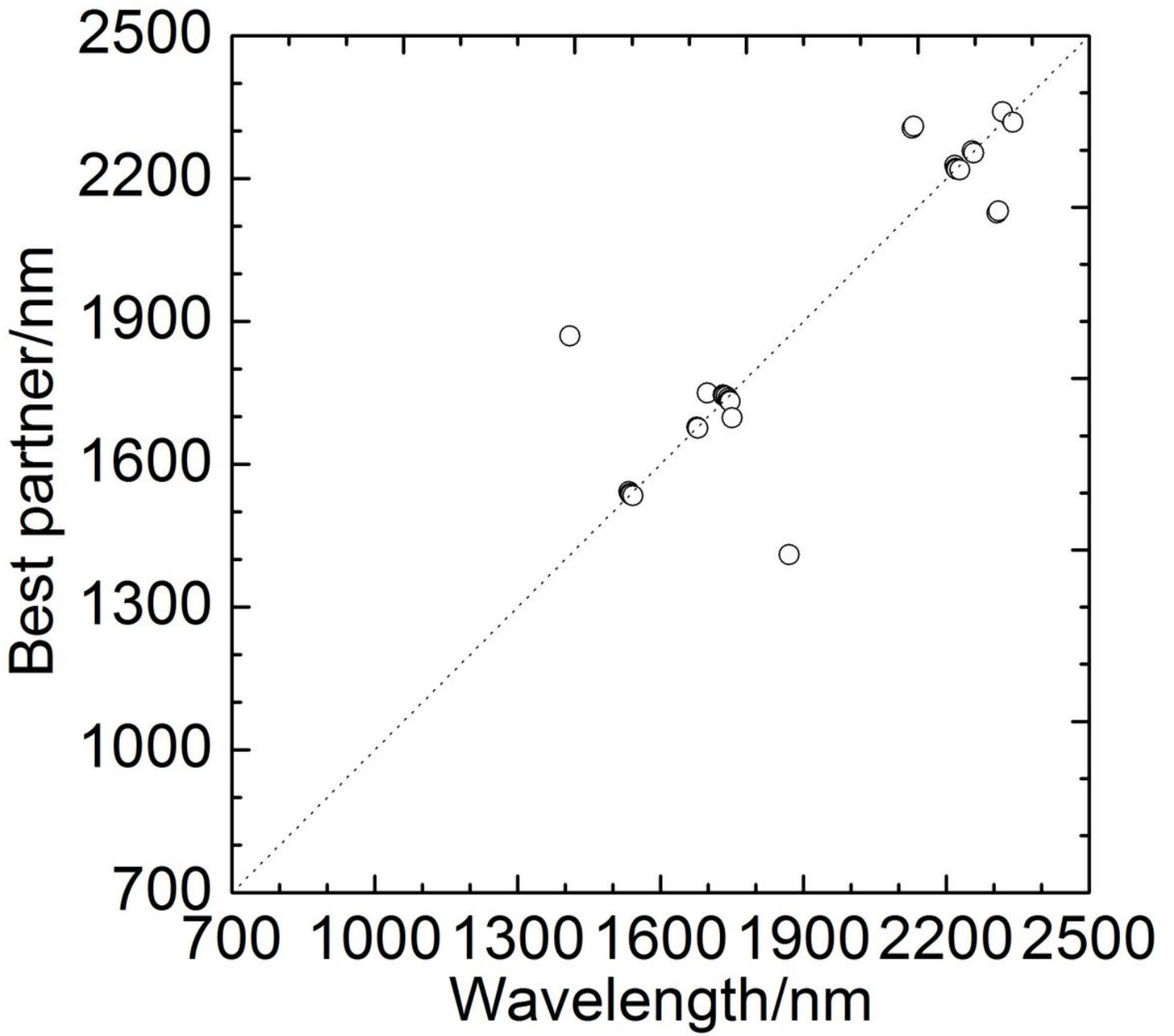


Figure 3

Best partners of the selected 28 wavelengths.

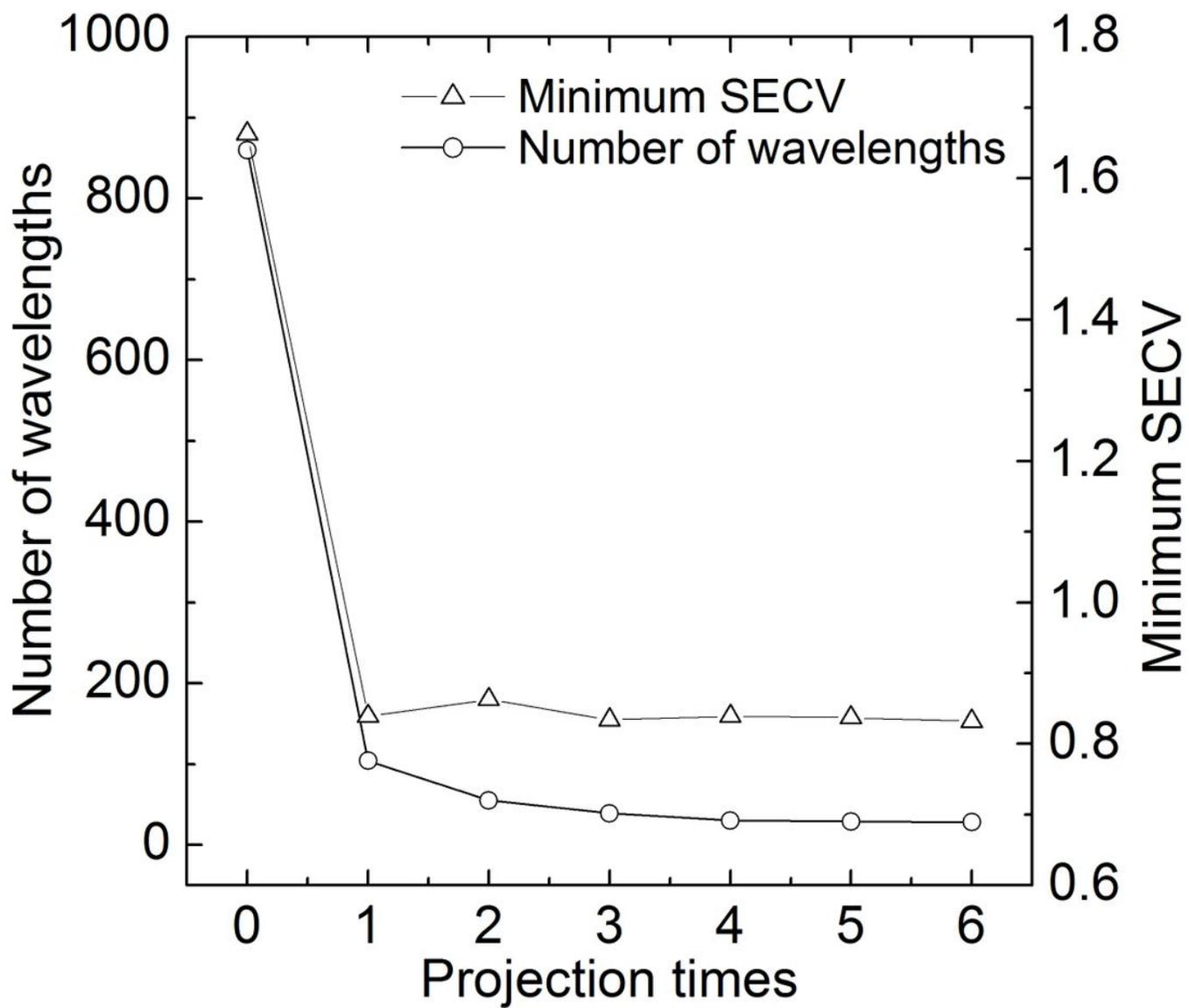


Figure 4

Number of wavelengths and minimum SECV value for each projection.

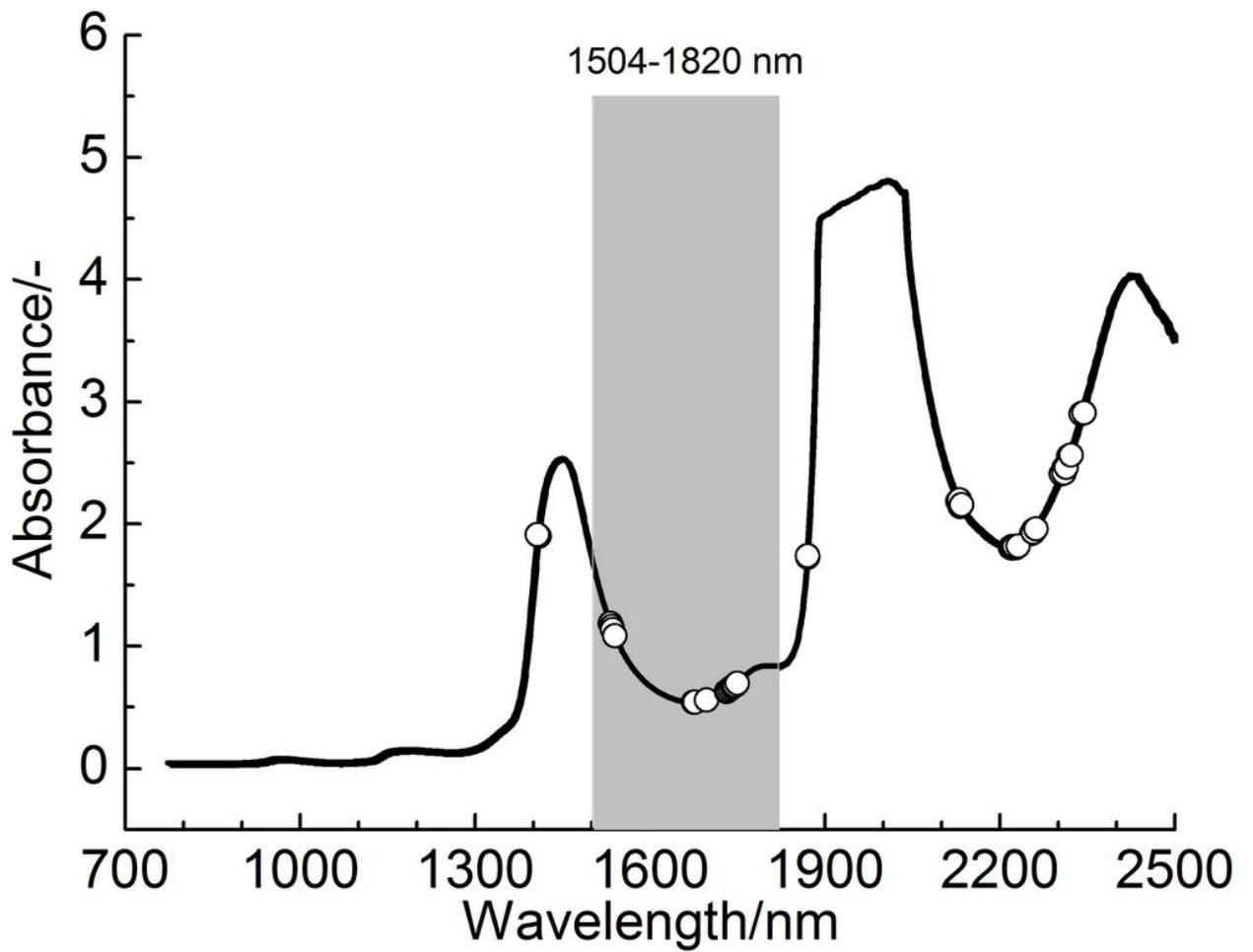


Figure 5

Position of the selected wavelengths with MW-PLS and OPWC-PLS located the average spectrum.

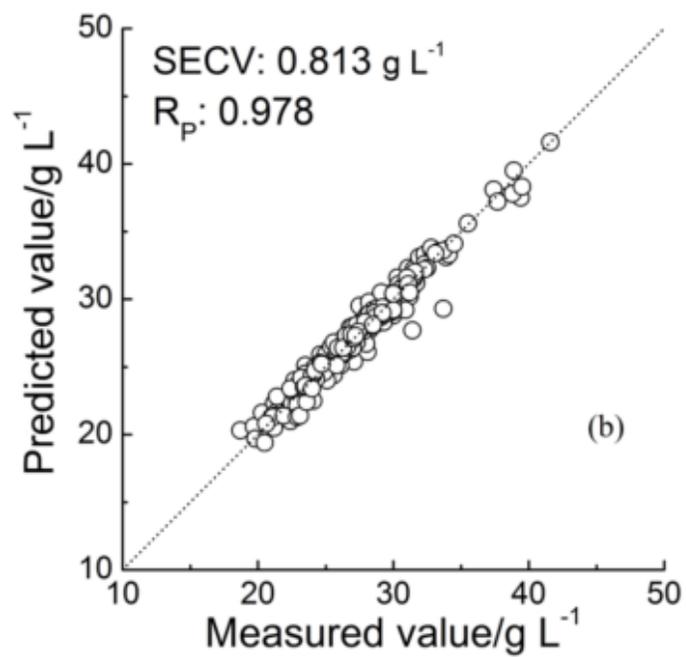
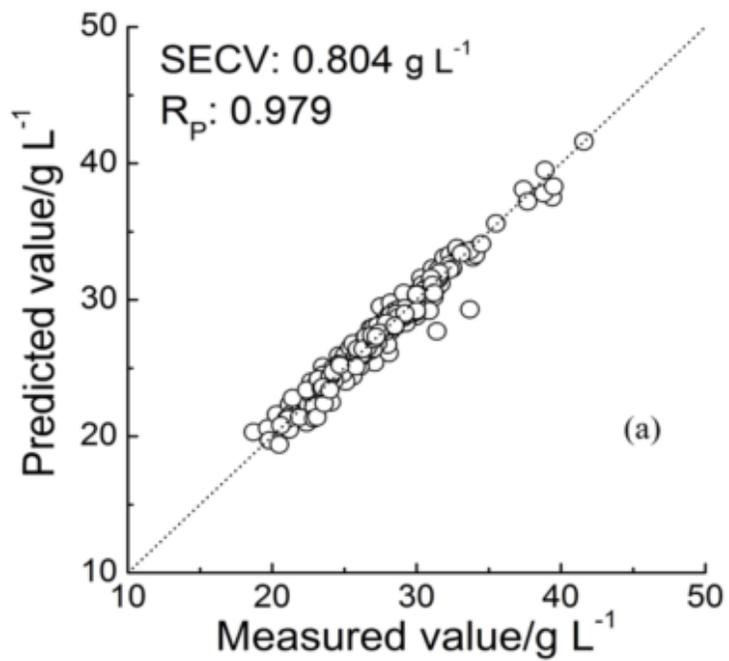


Figure 6

Relationship between the predicted values and measured values of GLB based on (a) MW-PLS and (b) OPWC-PLS methods.