

pH value monitoring during human albumin purification with near infrared spectroscopy and chemometrics

Qiaofeng Sun^{*,§}, Zhongyu Sun^{*,¶}, Fei Wang^{*}, Lian Li[†], Ronghua Liu^{*}, Lei Nie^{*},
Jiayue Wang^{*}, Mingyu Wang[‡] and Hengchang Zang^{*,||}

**School of Pharmaceutical Sciences
Shandong University, Wenhua Road 44
Jinan 250012, P. R. China*

*†School of Basic Medical Sciences
Shandong University, Wenhua Road 44
Jinan 250012, P. R. China*

*‡Shandong Taibang Biological Products
Limited Company, No. 14 East Hushan Road*

Taian 271000, P. R. China

§qiaofengsun2015@163.com

¶longyu504@sina.com

||zanghew@126.com

Received 24 August 2017

Accepted 19 October 2017

Published 29 November 2017

Human albumin (HA) is a very important blood product which requires strict quality control strategy. Acid precipitation is a key step which has a great effect on the quality of final product. Therefore, a new method based on quality by design (QbD) was proposed to investigate the feasibility of realizing online quality control with the help of near infrared spectroscopy (NIRS) and chemometrics. The pH value is the critical process parameter (CPP) in acid precipitation process, which is used as the end-point indicator. Six batches, a total of 74 samples of acid precipitation process, were simulated in our lab. Four batches were selected randomly as calibration set and remaining two batches as validation set. Then, the analysis based on material information and three different variable selection methods, including interval partial least squares regression (iPLS), competitive adaptive reweighted sampling (CARS) and correlation coefficient (CC) were compared for eliminating irrelevant variables. Finally, iPLS was used for variables selection. The quantitative model was built up by partial least squares regression (PLSR). The values of determination coefficients (R_c^2 and R_p^2), root mean squares error of prediction (RMSEP), root mean squares error of calibration (RMSEC) and root mean squared error of cross validation (RMSECV) were 0.969, 0.953, 0.0496, 0.0695 and 0.0826, respectively. The paired

||Corresponding author.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

t test and repeatability test showed that the model had good prediction ability and stability. The results indicated that PLSR model could give accurate measurement of the pH value.

Keywords: Near infrared spectroscopy; human albumin; acid precipitation process; pH value; chemometrics.

1. Introduction

As the main plasma protein, Human albumin (HA) plays an important physiological role in human body.¹ It is widely used in the treatment of various diseases.² There are two main sources of HA products on the market. One is prepared by genetic recombination,³ and the other is purified from human plasma.⁴ At present, most of the HA products are isolated from plasma with low temperature ethanol or improved methods which was developed by Professor Edwin J. Cohn in 1940.^{5,6} As the intermediate step of HA production, the acid precipitation is a key operation unit, which will directly affect the final product quality. The pH value is used as the process end-point indicator. When pH reaches the isoelectric point (pI) (4.60–4.70) of HA, it has the lowest solubility and can precipitate in large quantities. The pH value greater or less than the pI will produce positive or negative ions, which will lead to increase the solubility of HA and impurity, affecting the quality of final products. Therefore, it is important to monitor the trajectory of pH value. However, traditional method which is based on the industrial pH meter is not suitable for HA production. The raw material for acid precipitation contains about 40% ethanol,⁷ and now there is almost no available online industrial pH meters in the market. The measurement of pH values by pH meter is off-line and time-consuming, so it is unable to meet the needs of online production. Hence, a new spectral tool was introduced to investigate the feasibility for pH monitoring during the precipitation process.

Near infrared spectroscopy (NIRS) is a spectroscopic technique developed from the near infrared light, mainly reflecting the frequency doubling and combination tune of hydrogen groups.⁸ The changes of pH value were mainly caused by the changes of CH₃COOH. NIRS can detect the signal of CH₃COOH, and then monitoring pH value. NIRS was widely used in many fields because of its advantages of fast, nondestructive and pollution-free, such as pharmaceuticals,^{9,10} food,^{11,12} agriculture,^{13,14}

clinic¹⁵ and so on. NIRS has been used as an important process analytical technology (PAT) tool for critical process monitoring,^{16,17} quality inspection of final products.¹⁸ Prieto *et al.*¹⁹ concluded that pH measurement was feasible for meat products with NIRS. Wang *et al.*²⁰ demonstrated the feasibility for the determination of moisture content in human coagulation factor VIII freeze-drying process with NIRS. Zhang *et al.*²¹ used NIRS to determine the IgG content in the process of alcohol precipitation, and established a quantitative analysis model. NIRS was introduced to monitor the ethanol precipitation process of fraction I+II+III supernatant in HA separation by Li *et al.*²² It has been used in the food field for pH determination.²³ Monitoring the pH value trajectory during acid precipitation process in fraction IV (FIV) supernatant by NIRS has not been reported yet. In this paper, the production conditions were simulated in the laboratory, and the measurements of pH values by NIRS were studied.

The objective of this study is to monitor the variation of the pH value during the acid precipitation process by NIRS. It has been established a quantitative model for pH measurement, and the rapid and effective judgment of end point has been made.

2. Materials and Methods

2.1. Materials and equipment

Filtrated supernatant FIV and acetate buffer (pH 4.0) were provided by Shandong Taibang biological products co., LTD. Anhydrous ethanol with chromatographically pure was from Shandong Yuwang industrial co., LTD. Antaris II FT-NIR spectrometer (American Thermo Fisher scientific company) with glass tubular (Ø 4 × 50 mm Kimble Chase, Germany) was used to collect the NIR spectra. Low temperature reaction apparatus (Zhengzhou Great Wall industry & trade co., LTD.) was used to simulate the precipitation process in our lab. Gas chromatograph (Agilent Technologies company) and pH meter (Germany Sartorius company) was

used to determine the ethanol content and pH value, respectively. MATLAB 2013b (Mathworks, USA) and PLS_Toolbox toolkit (Eigenvector Research, USA) were used for data processing.

2.2. Methods

2.2.1. Acid precipitation process

Six batches acid precipitation processes were simulated under laboratory conditions. Specific process was as follows: 100 mL filtrated supernatant FIV were taken in a round flask which were placed in a low temperature reaction apparatus. Acetate buffer was added every 4 min. In the first nine times, 0.2 mL acetate buffer was added each time, 0.8 mL acetate buffer was added every time in the last six times. Each sample contained 1.2 mL and centrifuged for the spectral collection, the determination of ethanol content and pH value. Each precipitation process could obtain 16 samples.

2.2.2. Determination of ethanol content and pH value

The content of ethanol in the samples was determined by gas chromatography. The samples were diluted with deionized water to ethanol content of 10% based on the results by gas chromatography, and then the pH value of samples was measured using pH meter at room temperature.

2.3. Near infrared spectra collection

The spectra were acquired in transmission mode by Antaris II Fourier-transform near infrared spectrometer. The samples were loaded in a cuvette with 4 mm path length. Each spectrum was the average of 32 scans with resolution 8 cm^{-1} , the wave number ranges from $10,000$ to 4000 cm^{-1} . All samples were collected with air as reference.

2.4. Spectral pre-treatments and variables selection

Near infrared raw spectrum generally require pre-treatment to extract effective information, eliminate irrelevant information and noise interference. In this study, pre-treatment methods such as first derivative (FD), second derivative (SD) and orthogonal signal

correction (OSC)²⁴ were investigated. OSC²⁵ showed that how the variances of partial least squares (PLS) can be used to achieve orthogonal signal correction that is as close as possible to a given Y -vector. Thus, it ensures that the signal correction removes as little information as possible. The strict orthogonality is obtained when the number of X -variables exceed the number of observations.

In the study, in order to select the optimal effective information interval, the variables were investigated from two aspects. One of them was analysis of material information. In this part, the solution spectrum and spectral absorbance value were investigated. Because the samples were diluted with normal saline originally, the normal saline spectrum was investigated. The absorbance values vary widely at different wavelengths, which had a large effect on the quality of the spectrum. The other was the analysis of variable selection methods. The main methods for variable selection including interval partial least squares regression (iPLS), competitive adaptive reweighted sampling (CARS),²⁶ correlation coefficient (CC), Monte Carlo-uninformative variable elimination (MC-UVE)²⁷ and randomization test (RT).²⁸ In this paper, iPLS, CARS and CC methods were investigated to extract the informative variables related to pH value. The iPLS²⁹ method is compared to full-spectrum partial least-squares and this method including forward iPLS (FiPLS) and reverse iPLS (RiPLS). CC³⁰ is represented as the degree of linear association between two variables. The greater the correlation coefficient, the more information the variable contains. And variables were selected based on CC.

In this paper, PLSR³¹ method was used to establish quantitative analysis models for determination of pH value in acid precipitation process, and the predictive ability and repeatability of models were evaluated.

3. Results and Discussion

3.1. pH value

The results of pH value are shown in Table 1. Since the pH of the initial sample is far from the end point, samples from 0 to 2 in the first to third batches, samples from 0 to 4 in the fourth batch and 0 to 3 in fifth and sixth batches did not measure the pH values, and finally there were 74 samples in total.

Table 1. The pH values of acid precipitation process.

Sample number	First	Second	Third	Fourth	Fifth	Sixth
3	5.49	5.55	5.43	—	—	—
4	5.34	5.32	5.31	—	5.43	5.38
5	5.24	5.30	5.21	5.28	5.34	5.25
6	5.15	5.23	5.14	5.17	5.25	5.18
7	5.09	5.11	5.07	5.06	5.17	5.11
8	5.04	5.08	5.00	5.03	5.09	5.06
9	4.99	5.02	4.96	4.98	5.05	5.01
10	4.94	4.89	4.84	4.85	4.92	4.86
11	4.80	4.80	4.75	4.76	4.79	4.77
12	4.74	4.70	4.67	4.65	4.71	4.68
13	4.65	4.62	4.62	4.63	4.64	4.60
14	4.60	4.61	4.55	4.57	4.58	4.60
15	4.55	4.56	4.50	4.52	4.54	4.55

3.2. Near infrared spectra

The raw spectra of 74 samples are shown in Fig. 1. The spectral differences between samples were small, so it is necessary to use the chemometrics for processing of spectral data and extracting the effective information.

The principal component scores are shown in Fig. 2. As can be seen from Fig. 2, samples 12 and 13 of first batch were not within 95% confidence limits, away from the sample data set. Samples 12 and 13 of first batch were abnormal samples and were excluded from the sample set. Four batches including the second, third, fourth and fifth batches (49 samples) were selected randomly as calibration set and remaining two batches including the first, sixth batches (23 samples) as validation set.

3.3. Spectral pre-treatment results

In this study, the effects of different pre-treatment methods on the model were compared with RMSEP as the evaluation index. PLSR model results are shown in Table 2. After pre-treatments, R_c^2 and R_p^2 were increased, RMSEC, RMSECV and RMSEP decreased significantly, indicating that pre-treatment could improve the robustness of the model. The results of models parameters indicated that the results of different pre-processing methods were different, and the order of pre-processing methods had a great influence on the model results. The results of the derivative processing and OSC processing were obviously better. The reason may be that OSC is a pre-processing method based on concentration matrix (Y value). Spectral information and concentration matrix had a high correlation. The best pre-processing method was FD combined with Savitzky–Golay (SG)³² smoothing filter with 15 data points with OSC. The model results after pre-processing are shown in Fig. 3.

3.4. Variable selection

The average spectrum of the samples and normal saline spectrum after the second derivative pre-treatment are shown in Fig. 4. Figure 4(b) showed that it had obvious absorption peaks at the wave number 5400–5200 cm^{-1} and 7350–6950 cm^{-1} , which reflected O–H combination tune absorption and the double-frequency absorption in water molecules. A similar absorption peak can be found

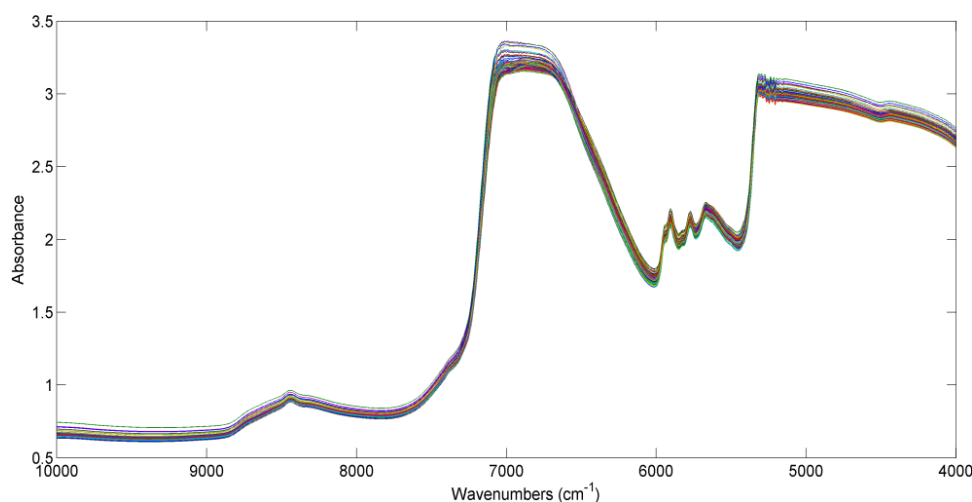


Fig. 1. The raw spectra of samples.

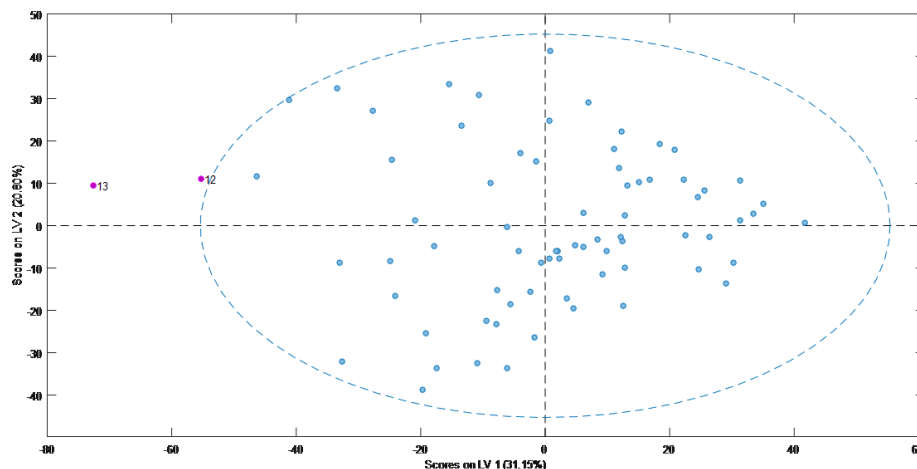


Fig. 2. Principal component score scatter plot of PC1 versus PC2.

Table 2. Model results of different pre-processing methods.

Methods	R_c^2	R_p^2	RMSEC	RMSECV	RMSEP
None	0.533	0.734	0.1927	0.2054	0.1899
OSC	0.845	0.704	0.1109	0.1284	0.2121
FD ^a	0.966	0.912	0.0521	0.0996	0.1210
SD ^a	0.981	0.834	0.0389	0.0902	0.1605
OSC+FD ^a	0.963	0.695	0.0544	0.1069	0.2212
OSC+SD ^a	0.983	0.821	0.0369	0.0906	0.1513
FD ^a +OSC	0.984	0.910	0.0353	0.0872	0.1055
SD ^a +OSC	0.987	0.871	0.0321	0.0908	0.1383

^aThe spectral derivative is subjected to SG 15-point smoothing at the same time.

in Fig. 4(a), indicating that the two peaks represented information of normal saline. To eliminate the interference of the background information, the spectrum was processed by removing the

absorption bands of 5400–5200 cm^{-1} and 7350–6950 cm^{-1} .

From the raw spectra in Fig. 1, the absorbance values at 5341–4227 cm^{-1} and 7112–6510 cm^{-1} were found to be above 2.80 and the band of 10,000–9000 cm^{-1} was about 0.70. It can be seen that the absorbance difference was large. The effects of high and low absorbance of the models were investigated with RMSEP as an index.

The threshold range was set between 2.25 and 3.25 with a gradient of 0.10 to select variables after the investigation of high absorbance. In RMSEP results, the absorbance during 2.25 to 3.25 is shown in Fig. 5. The variables had little effect on the model from 2.25 to 2.80. However, when the absorbance was between 2.80 and 2.89, the RMSEP values of the model decreased rapidly, indicating that the variables of this interval could have an important

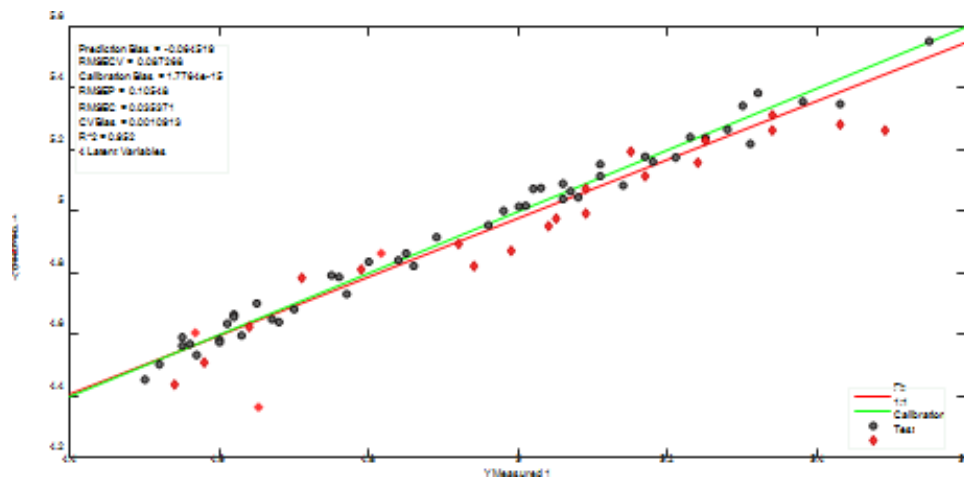


Fig. 3. Model results after FD+SG smoothing (15 data points, second polynomial order) + OSC pre-treatment.

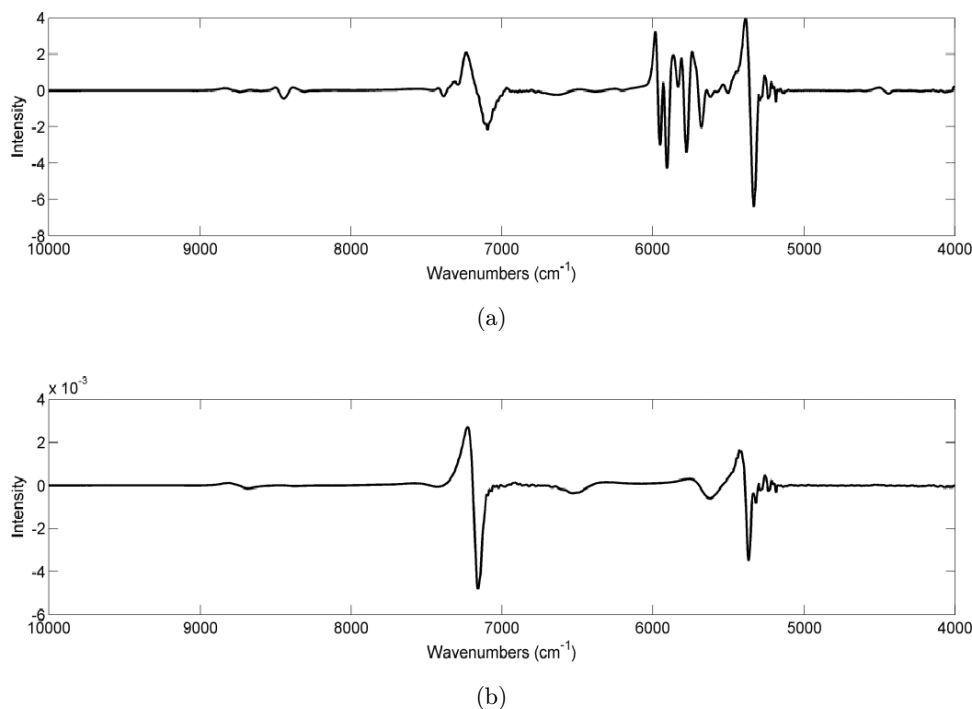


Fig. 4. The average spectrum of samples (a) The spectrum of normal saline and (b) (Second derivative pre-processing).

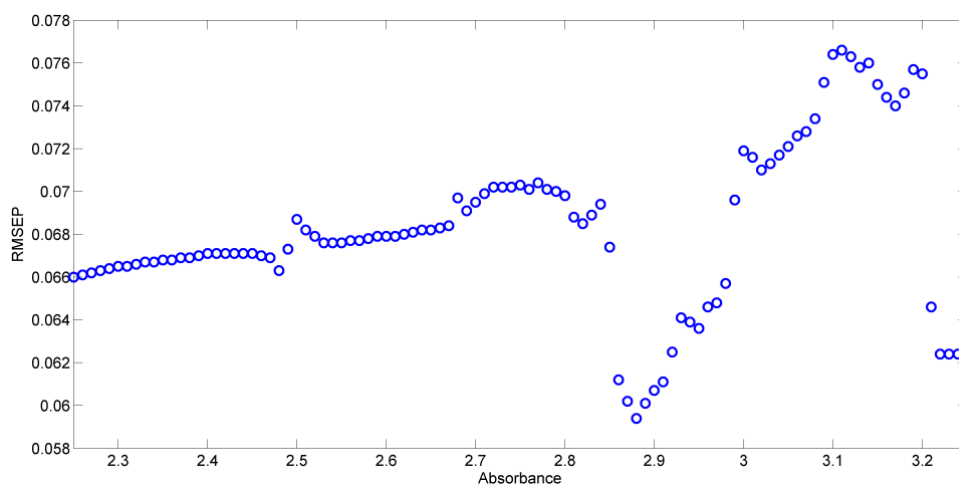


Fig. 5. The results of RMSEP values with different high absorbance thresholds.

influence on the establishment of the model. When the threshold was between 3.2 and 3.25, RMSEP values fell again, but there was no value below the threshold of 2.88. So the high absorbance limit was 2.88. The absorbance threshold was set during 0.63 to 1.50 to select variables with an absorbance above the threshold after the study of low absorbance. The results of RMSEP values when absorbance was during 0.63 to 1.50 are shown in Fig. 6. As can be seen from Fig. 6, RMSEP values showed that there

were two rising intervals and two stable intervals. RMSEP reached the minimum value when the absorbance limits was 0.63, indicating that all low absorbance variables should be involved in the establishment of model.

In summary, the normal saline spectrum and the absorbance of the spectrum were investigated. Eventually, the three spectral ranges including 4613–4000 cm^{-1} , 6538–5404 cm^{-1} and 10,000–7355 cm^{-1} in total of 1142 variables were extracted.

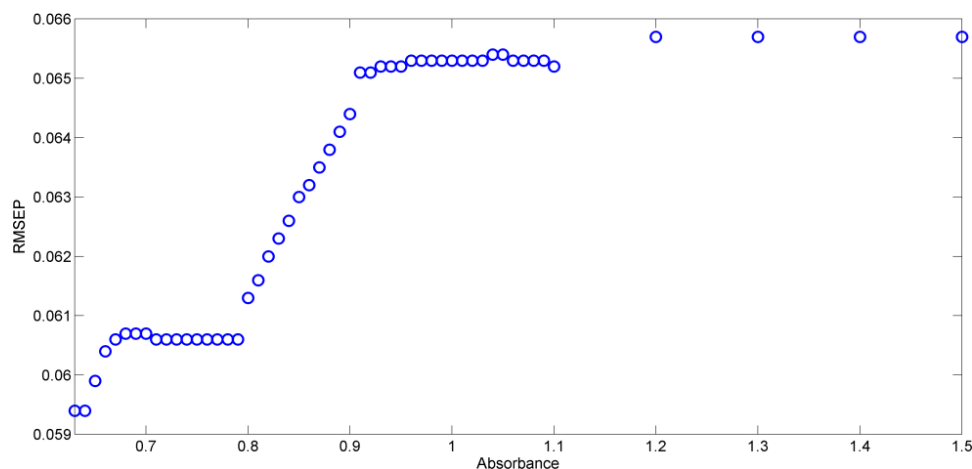


Fig. 6. The results of RMSEP values with different low absorbance thresholds.

Table 3. The PLSR models results of different variable selection methods.

Methods	R_c^2	R_p^2	RMSEC	RMSECV	RMSEP
4613–4000, 6538–5404, 10,000– 7355 cm^{-1} ,	0.946	0.914	0.0653	0.0914	0.1350
CARS(150)	0.887	0.876	0.0946	0.1078	0.1058
CC	0.983	0.894	0.0363	0.0817	0.1782
FiPLS	0.969	0.953	0.0496	0.0695	0.0826
RiPLS	0.983	0.818	0.0367	0.0761	0.2575

The above wave numbers contained the combination tune absorption and double-frequency absorption of C–H, double-frequency absorption of C=O (RCOOH). The results of the PLSR models are shown in Table 3.

Beside the above analysis, three classic variable selection methods were also used to extract variables. Including CARS, CC, FiPLS, RiPLS. The results of PLSR model with different variable selection methods³⁰ are shown in Table 3. The number of variables decreased and the results of model became better when variable selection methods were adopted. The FiPLS method was superior to other methods, so it was used as the optimal variable selection methods. The results of selected variables are shown in Fig. 7.

3.5. Model optimization and evaluation

3.5.1. Model optimization results

Pre-treatment by FD combined with SG smoothing filter with 15 data points with OSC, 600 variables were selected by FiPLS for the final model

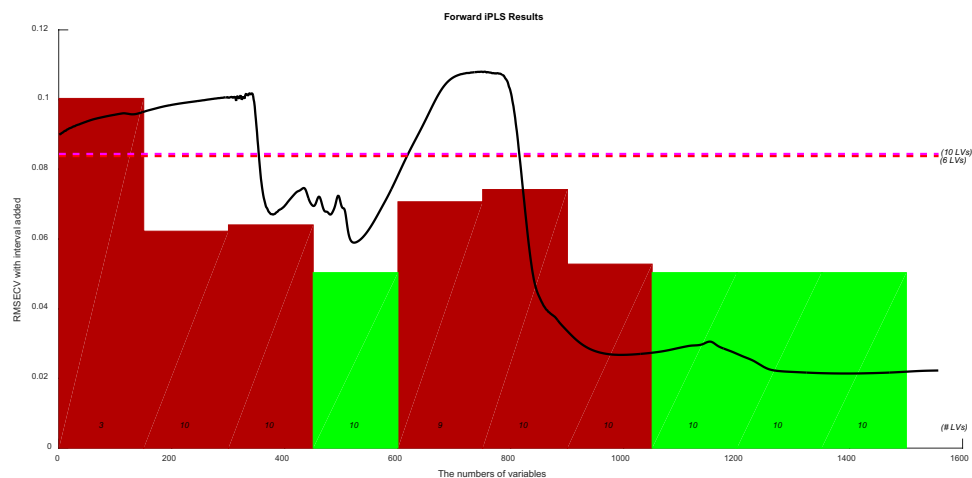


Fig. 7. The results of variables selection.

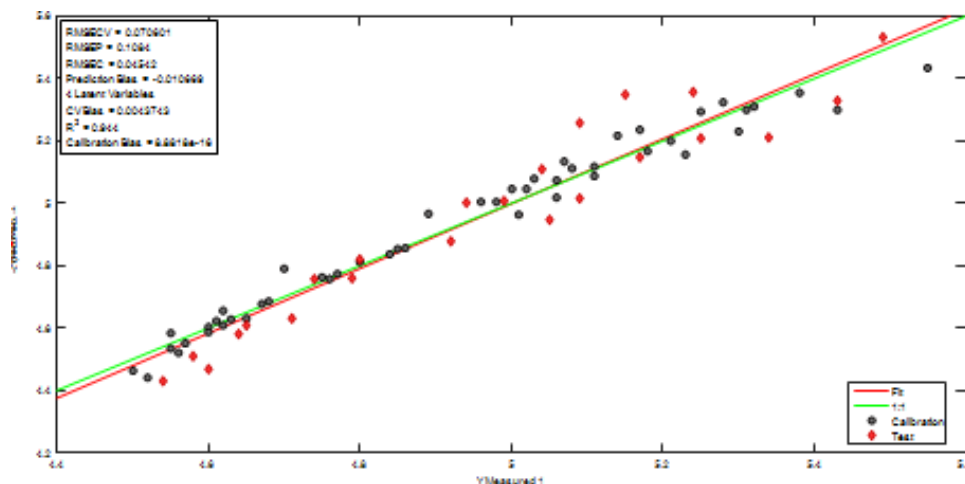


Fig. 8. The PLSR model after SG filter with 15 data points smoothing and FD and OSC pre-treatment and variable selection.

establishment. The best PLSR model are shown in Fig. 8, the results of the specific parameters were as below: $R_c^2 = 0.969$, $R_p^2 = 0.953$, $RMSEC = 0.0496$, $RMSECV = 0.0695$, $RMSEP = 0.0826$.

3.5.2. Model prediction ability evaluation

The pH values of samples which were calculated by the model and measured by the pH meter were used to test the paired *t*-test to evaluate the prediction ability of the model. The statistical results of the paired *t*-test are shown in Table 4. From the table, we found that the average values were the same. There were no significant differences between the two results. NIRS is effective in the determination of pH during acid precipitation.

3.5.3. Repetitive inspection

Three samples were selected which could represent high and low values of pH values, each sample was continuously measured for six spectra. The pH values were predicted by PLSR model. The reproducibility of the near infrared model was investigated. Repetitive inspection result is shown in Table 5. The results showed that predicted values

Table 5. The results of the repetitive studies.

Number	06-4	04-7	06-13
1	5.22	5.07	4.51
2	5.21	5.04	4.54
3	5.18	5.04	4.52
4	5.25	5.00	4.48
5	5.24	4.98	4.52
6	5.29	5.00	4.47
Mean	5.23	5.02	4.51
SD	0.038	0.034	0.027
RSD (%)	0.72	0.67	0.59

were identical with true values. RSD values were far less than 5%, which confirmed the better reproducibility of the model.

4. Conclusion

In this paper, near-infrared quantitative analysis models of pH values during acid precipitation were established, which were used for the determination of pH values and the judgment of the end point of acid precipitation process. Under the laboratory conditions, six batches of acid precipitation were simulated and sampled. The samples were diluted with deionized water to ethanol content of 10% based on the measured results by gas chromatography, and then pH values were measured by pH meter. To make the model more reliable, the calibration set and the validation set were divided according to the batches. Four batches were selected randomly as calibration set and remaining two batches as validation set. The PLSR model was

Table 4. The results of paired *t* test.

Methods	Samples	Mean	SD	<i>t</i> test (0.05)	
				<i>t</i>	<i>P</i>
pH meter	23	4.98	0.27	0.186	0.746
NIRS	23	4.98	0.28		

established by selecting the pre-treatment methods and extracting the valid variables. In terms of variables extraction, the variables were investigated from two aspects. One of them was material analysis. The other was the investigation of three different variable selection methods, including iPLS, CARS and CC method. The FiPLS method was chosen as the best variables selection method. The results of best PLSR model was as follows: $R_c^2 = 0.969$, $R_p^2 = 0.953$, RMSEC = 0.0496, RMSECV = 0.0695, RMSEP = 0.0826. The model was evaluated and the results showed that the model had good prediction ability and repeatability.

This study was carried out under laboratory conditions and provided a theoretical analysis for the application. NIRS is used for the determination of pH and the accurate and effective judgment at the end of acid precipitation in industrial production, which needs further study.

Conflict of Interest

The authors declare that there are no conflicts of interest related to this paper.

Acknowledgement

The authors are grateful to the financial support of the Major Special Project of National Science and Technology (No. 2014ZX09508003-001-003) and the supply of Supernatant FIV of Shandong Taibang Biological Products Limited Company. Qiaofeng Sun and Fei Wang contributed equally to this work and should be regarded as co-first authors.

References

1. J. Rozga, T. Piatek, P. Malkowski, "Human albumin: Old, new, and emerging applications," *Ann. Transpl.* **18**, 205–217 (2013).
2. J. Boldt, "Use of albumin: An update," *Br. J. Anaesth.* **104**, 276–284 (2010).
3. K. Kobayashi, "Summary of recombinant human serum albumin development," *Biologicals* **34**, 55–59 (2006).
4. T. Burnouf, "Modern plasma fractionation," *Transf. Med. Rev.* **21**, 101–117 (2007).
5. E. J. Cohn, J. L. Oncley, L. E. Strong, W. L. Hughes, S. H. Armstrong, "Chemical, clinical, and immunological studies on the products of human plasma fractionation. I. The characterization of the protein fractions of human plasma," *J. Clin. Invest.* **23**, 417–432 (1944).
6. P. Kistler, H. Nitschmann, "Large scale production of human plasma fractions," *Vox Sang.* **7**, 414–424 (1962).
7. H. P. Zhang, H. Du, S. H. Wang, "Influence study of Ethanol on the pH of the solution," *Guangzhou Chem. Ind.* **44**(20), 76–78 (2016).
8. G. Reich, "Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications," *Adv. Drug Deliv. Rev.* **57**, 1109–1143 (2005).
9. M. Jamrogiewicz, "Application of the near-infrared spectroscopy in the pharmaceutical technology," *J. Pharm. Biomed. Anal.* **66**, 1–10 (2012).
10. T. Hu, T. T. Li, L. Nie, L. X. Zang, H. C. Zang, Y. Z. Zeng, "Rapid monitoring the water extraction process of Radix Paeoniae Alba using near infrared spectroscopy," *J. Innov. Opt. Health Sci.* **10**(3) (2017), doi: 1142/S179354581750002x.
11. G. B. da Costa, D. D. S. Fernandes, A. A. Gomes, V. E. de Almeida, G. Veras, "Using near infrared spectroscopy to classify soybean oil according to expiration date," *Food Chem.* **196**, 539–543 (2016).
12. P. Siriphollakul, S. Kanlayanarat, R. Rittiron, J. Wanitchang, T. Suwonsichon, P. Boonyariththongchai, K. Nakano, "Pasting properties by near-infrared reflectance analysis of whole grain paddy rice samples," *J. Innov. Opt. Health Sci.*, **8**(6) (2015).
13. M. B. S. Scholz, N. F. Pagiatto, C. S. G. Kitzberger, L. F. P. Pereira, F. Davrieux, P. Charmetant, T. Leroy, "Validation of near-infrared spectroscopy for the quantification of cafestol and kahweol in green coffee," *Food Res. Int.* **61**, 176–182 (2014).
14. K. Phuangsombut, N. Suttiwijitpukdee, A. Terdwongworakul, "Nondestructive classification of mung bean seeds by single kernel near-infrared spectroscopy," *J. Innov. Opt. Health Sci.* **10**(3) (2017).
15. A. N. Sen, S. P. Gopinath, C. S. Robertson, "Clinical application of near-infrared spectroscopy in patients with traumatic brain injury: A review of the progress of the field," *Neurophotonics* **3** (2016), doi: 10.1117/1.Nph.3.3.031409.
16. L. Li, B. Y. Ding, Q. Yang, S. Chen, H. Y. Ren, J. F. Wang, H. C. Zang, F. S. Wang, L. X. Zang, "The relevance study of effective information between near infrared spectroscopy and chondroitin sulfate in ethanol precipitation process," *J. Innov. Opt. Health Sci.* **7** (2014).
17. P. Wang, H. Zhang, H. L. Yang, L. Nie, H. C. Zang, "Rapid determination of major bioactive isoflavonoid compounds during the extraction process of kudzu (*Pueraria lobata*) by near-infrared transmission spectroscopy," *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **137**, 1403–1408 (2015).

18. X. B. Zhang, Y. C. Feng, C. Q. Hu, "Feasibility and extension of universal quantitative models for moisture content determination in beta-lactam powder injections by near-infrared spectroscopy," *Anal. Chim. Acta* **630**, 131–140 (2008).
19. N. Prieto, R. Roehe, P. Lavin, G. Batten, S. Andres, "Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review," *Meat Sci.* **83**, 175–186 (2009).
20. F. Wang, W. Jiang, C. Li, H. Zhang, L. Nie, L. Li, P. Wang, H. C. Zang, "Application of near infrared spectroscopy in monitoring the moisture content in freeze-drying process of human coagulation factor VIII," *J. Innov. Opt. Health Sci.* **8** (2015).
21. H. Zhang, A. H. Liu, H. C. Zang, H. Li, W. Jiang, L. Li, J. F. Wang, "Rapid determination of immunoglobulin G concentration in cold ethanol precipitation process of raw plasma with near-infrared spectroscopy," *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **116**, 370–373 (2013).
22. C. Li, F. Wang, L. Zang, H. Zang, M. Alcala, L. Nie, M. Wang, L. Li, "Near infrared spectroscopy combined with multivariate analysis for monitoring the ethanol precipitation process of fraction I+II+III supernatant in human albumin separation," *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **175**, 17–23 (2017), doi: 10.1016/j.saa. 2016.12.016.
23. G. ElMasry, D. W. Sun, P. Allen, "Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef," *J. Food Eng.* **110**, 127–140 (2012).
24. H. B. Qu, D. L. Ou, Y. Y. Cheng, "Background correction in near-infrared spectra of plant extracts by orthogonal signal correction," *J. Zhejiang Univ. Sci. B* **6**, 838–843 (2005), doi: 10.1631/jzus.2005.B0838.
25. S. Wold, H. Antti, F. Lindgren, J. Ohman, "Orthogonal signal correction of near-infrared spectra," *Chemometr. Intell. Lab. Syst.* **44**(1–2), 175–185 (1998), doi: 10.1016/S0169-7439(98)00109-9.
26. H. D. Li, Y. Z. Liang, Q. S. Xu, D. S. Cao, "Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration," *Anal. Chim. Acta* **648**, 77–84 (2009).
27. X. W. Liu, X. Y. Cui, X. M. Yu, W. S. Cai, X. G. Shao, "Understanding the thermal stability of human serum proteins with the related near-infrared spectral variables selected by Monte Carlo-uninformative variable elimination," *Chin. Chem. Lett.* **28**, 1447–1452 (2017).
28. Z. Y. Mao, W. S. Cai, X. G. Shao, "Selecting significant genes by randomization test for cancer classification using gene expression data," *J. Biomed. Inf.* **46**, 594–601 (2013).
29. L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, S. B. Engelsen, "Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy," *Appl. Spectrosc.* **54**, 413–419 (2000).
30. R. Taylor, "Interpretation of the Correlation Coefficient: A Basic Review," *J. Diagn. Med. Sonogr.* **6**(1), 35–39 (1990).
31. P. Geladi, B. R. Kowalski, "Partial Least-Squares Regression - a Tutorial," *Anal. Chim. Acta* **185**, 1–17 (1986), doi 10.1016/0003-2670(86)80028-9.
32. W. L. Li, Z. W. Cheng, Y. F. Wang, H. B. Qu, "A study on the use of near-infrared spectroscopy for the rapid quantification of major compounds in Tanreqing injection," *Spectrochim. Acta A* **101**, 1–7 (2013).