ISSN 0030-400X, Optics and Spectroscopy, 2014, Vol. 117, No. 5, pp. 703–709. © Pleiades Publishing, Ltd., 2014.

SPECTROSCOPY OF ATOMS AND MOLECULES

Absorption and Quantitative Characteristics of C–H Bond and O–H Bond of NIR[1]

Zhisheng Wu*, Guoqing Ouyang, Xinyuan Shi, Qun Ma, Guang Wan, and Yanjiang Qiao**

Beijing University of Chinese Medicine, Beijing, 100102 China
Key Laboratory of TCM-information Engineering of State Administration of TCM, Beijing, 100102 China
Beijing Key Laboratory for Basic and Development Research on Chinese Medicine, Beijing, 100102 China
*e-mail: *wzs@bucm.edu.cn; **yjqiao@263.net*

Received April 24, 2013; in final form, November 11, 2013

Abstract—The previous study mainly focused on the interpretation of the relationship between absorption characteristics and quantitative contribution in each near-infrared (NIR) frequency range. Furthermore, the absorption characteristics of NIR mainly cover overtones and combinations of molecular vibrations such as CH, OH, SH, and NH bonds. And it has been know that NIR wavelengths of C–H bond and O–H bond are assigned to different radio frequencies. This paper was intended to investigate the absorption characteristics of bond C–H and O–H bonds in NIR spectral range. Water and acetone which correspond to O–H and C–H bonds have been selected as typical solvents, as well as solutes. Calibration models were established using partial least square regression (PLS) and multiple linear regression (MLR). The parameter of the model were optimized by different spectral pretreatment methods. The result showed that the model set by Savitzky–Golay smooth (SG) in the spectral region of 810–1100 nm could successfully make accurate predictions. Short wave-NIR region has been discovered as optimum characteristic absorption of C–H and O–H bonds.

DOI: 10.1134/S0030400X1411023X

1. INTRODUCTION

Near-infrared (NIR) spectroscopy has been widely used as a fast low cost and non-destructive technique. The wavenumber of NIR is from 4000 cm^{-1} to 12500 cm^{-1}, which mainly covers overtones and combinations of molecular vibrations. These signals decrease significantly in absorption cross section, compared with the fundamental vibrational bands from mid-infrared [1–4]. The functional groups, C–H, N–H, O–H, and S–H are almost exclusively involving the hydrogen atom. Based on the molecular vibrations, the NTR frequency range can be divided into four ranges: combination region (CR, 4000–4900 cm^{-1}), first combination-overtone (FCOT, 4900–7100 cm^{-1}), second combination-overtone (SCOT, 7100–10000 cm^{-1}), and third overtone (TO, 10000–12500 cm^{-1}) [5].

Recently, NIR effectiveness for both qualitative and quantitative analysis has proven in different fields such as agriculture, food and petroleum industry [6–8]. However, because of its high detection limit and low sensitivity, the trace analysis of NTR is still a tough challenge. The main difficulty of NTR is overlapping and broad absorption bands due to their overtones and combination tones of hydrogen bonds, thus strong self-absorption of the solvent may seriously interfere with the absorption signal of the solute [9–11].

NIR wavelengths of C–H bond and O–H bond are assigned to different radio frequencies [12, 13]. On the other hand, a great majority of solvents contain C–H bond and O–H bond system. Therefore, we tried to research the absorption and quantitative characteristics of C–H bond and O–H bond of NIR, finding the features of their absorption intension and the diversity of the useful message leading by the concentration of the samples.

In our experiment, water and acetone, which correspond to O–H bond and C–H bond, were selected as typical solvents, as well as solutes (water in acetone and acetone in water). Then calibration models were established by different methods using partial least square regression (PLS) and multiple linear regression (MLR). The model fitting results can generally be evaluated according to the following chemometric indications: low root mean square error in cross-validation (RMSECV), low root mean square prediction error (RMSEP), high determination coefficient (R^2) and low bias, etc. [14–16].

Once the calibration model is developed, favorable predictions could be expected. Then the optimum wavelength will be investigated for the absorption signal of C–H bond and O–H bond of NIR.

[1] The article is published in the original.

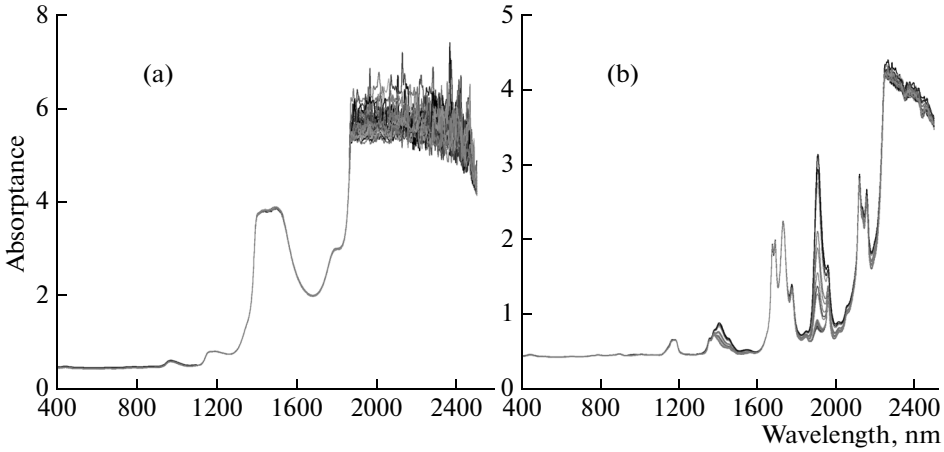

Fig. 1. Raw NIR spectra of acetone (a) and water (b) sample solutions.

2. MATERIALS AND METHODS

2.1. Materials

Acetone was purchased from Concord Technology Co., Ltd. (Tianjin, China). Deionized water was purified by Milli-Q water system (Millipore Corp., Bedford, MA, USA).

2.2. Preparation of Sample Solutions

1.0 mL acetone was measured to a 10 mL volumetric-flask accurately. And then meter the volume to scale with water to dispense the 10% acetone solution in water. 5.0 mL of the solution was transferred to a 250 mL volumetric flask and diluted with water to dispense the 0.2% acetone solution. Measure 0.1, 0.2, 0.3...2.3, 2.4 and 2.5 mL of the 0.2% acetone solution to a 5 mL volumetric flask separately, then meter the volume to scale with water to dispense the 40, 80, 120...920, 960 and 1000 ppm acetone solutions.

Measure 125.0 mL of the 0.2% acetone solution to a 250 mL volumetric flask and then meter the volume to scale with water to dispense the 0.1% acetone solution. Measure 0.1, 0.3, 0.5...4.5, 4.7 and 4.9 mL of the 0.1% acetone solution to a 5 mL volumetric flask separately, then meter the volume to scale with water to dispense the 20, 60, 100...900, 940 and 980 ppm acetone solutions.

The same method was used to prepare water solutions in acetone of the same concentration series. Then all the sample solutions (about 1 mL) were hold in circular sample cuvettes with solid caps separately (8 mm in diameter) for NIR spectra collection.

2.3. NIR Equipment and Software

The NIR spectra were collected in the transmission mode using the XDS rapid liquid analyzer with VISION software (Foss Holographic Grating NTR System, Silver Spring, MD, USA). Each spectrum was in average of 32 scans with a wavelength increment of 0.5 nm. The range of spectra was from 400 ran to 2500 nm. Each sample was analyzed three times and the mean of the three spectra was used in the following analysis.

2.4. Calibration Model

Spectral data were selected under the concentrations of 20, 40, 80, 100, 140, 160...860, 880, 920, 940, 980, 1000 ppm (about two thirds of all samples) to set calibration model in pertinent method and the rest for prediction. In this work, two calibration models were set: PLS model and MLR model. The PLS model was evaluated by parameters such as determination coefficient of calibration (R_c^2), determination coefficient of prediction (R_p^2), prediction residual error sum of squares (PRESS), root mean square error of calibration (RMSEC), RMSECV, RMSEP, residual predictive deviation (RPD). The MLR model was evaluated by the parameters like R_c^2, R_p^2, RMSEC, RMSEP and RPD.

Data analysis was performed with home-made routines programmed in MATLAB code (MATLAB v 7.0, MathWorks) and VISION software.

3. RESULTS AND DISCUSSION

3.1. NIR Spectral Characteristics

Figure 1 shows the raw NIR spectra of sample solutions. As shown in Fig. 1, there was too much noise in the CR region, which was interfered with the quantitative analysis of the solutes. On the other hand, there were large signal fluctuations in the spectral region of 800–1900 nm, which might mean that this spectral region contained the main information which could reflect the concentration message of different samples.

Table 1. Comparison of PLS models using different spectral pretreatment methods

Solvents	Pretreated method	Factor	R_c^2	RMSEC	RMSECV	PRESS	R_p^2	RMSEP	RPD
Acetone	Raw	6	0.9244	85.01	89.73	700524.94	0.9274	103.72	2.81
	1D	7	0.8669	113.49	120.53	1263860.6	0.8680	138.95	2.10
	2D	3	0.5481	204.02	209.57	3820944.5	0.7311	192.55	1.51
	SG	12	0.9663	46.37	57.38	121539.57	0.9577	75.13	3.88
	NPS	7	0.9048	95.97	103.86	938462.94	0.9101	115.24	2.53
	BC&NPS	8	0.9121	92.82	100.73	882677.69	0.9194	109.82	2.65
	BC&NPS&SG	4	0.8928	99.97	107.20	999703.13	0.8911	127.53	2.29
Water	Raw	2	0.2037	259.33	264.18	6490401.5	0.3125	300.50	0.97
	1D	2	0.1724	264.38	273.59	6961432.0	0.0975	309.51	0.94
	2D	6	0.7383	152.09	159.38	2362369.2	0.8009	173.90	1.68
	SG	4	0.9687	58.03	72.95	368520.19	0.9754	72.40	4.03
	NPS	2	0.2041	259.27	263.94	6478759.0	0.3032	304.36	0.96
	BC&NPS	2	0.1925	261.14	265.89	6574750.5	0.3199	269.74	1.08
	BC&NPS$SG	10	0.9616	60.78	100.06	715815.31	0.9505	91.35	3.19

As a consequence, the spectral region of 800–1900 nm was selected for model setting, in order to set a reasonable model which could make accurate quantitative prediction.

3.2. Comparison of Different Spectral Pretreatment Methods

The type of data pre-processing technique prior to the development of calibration model may greatly influence model performance. Table 1 shows the comparison of different spectral pretreatment, i.e., 1st order derivative (1D), 2nd order derivative (2D), Savitzky–Golay smoothing (SG), N-Point Smoothing (NPS), Baseline Correction + N-Point Smoothing (BC and NPS), Baseline Correction + N-Point Smoothing + Savitzky–Golay smoothing (BC and NPS and SG). To our basic knowledge, the nearer the value of R_c^2 and R_p^2 come to 1, the better the line of calibration and prediction result could be; the lower the value of RMSEC, RMSECV and RMSECP is, the more accurate the calibration model and prediction could be. As to the RPD, it is equal to the value of SD (mean square deviation) being dividing by RMSEP, with the use of evaluating the model's ability of predictions, the bigger the value of RPD is, the higher prediction ability the model has, and when it is bigger than three, it means that the model can predict accurately [17].

As shown in Table 1, the values of the RMSEC, RMSECV, RMSEP and PRESS were the lowest, while the value of R_c^2, R_p^2 and RPD was the highest when the spectra was pretreated and the PLS model was established with the method of SG smooth, it means that the model set by the method of SG smooth was the best one, so it was chosen to set the PLS model.

Table 2 shows the comparison of different MLR models set in different spectral pretreatment methods. According to the above selective rules, it is not difficult to make a conclusion that the method of SG smooth was the best one to make pretreatment and set MLR model.

3.3. Waveband Selection and Absorption Characteristics

Based on the results given above, we can get that the use of each NIR frequency region involves the development of calibration models to reflact the absorption characteristics.

Table 3 shows each PLS model under SG smooth in different spectral regions. Other regions in which an accurate model could not be set were not shown (the result of MLR model is show in Table 4). As seen in Tables 3 and 4 we can find that when models set by the method of SG smooth in the spectral region of 810–1100 nm, its values of R_c^2 and R_p^2 turned to be nearest to 1, and its values of RMSEC and RMSEP were the lowest and the value of RPD was the highest. Therefore, 810–1100 nm (short wave-NTR region) was selected as the best spectral region for PLS model and MLR model.

3.4. Determination of the Optimum LV Numbers

In PLS model, it is generally known that the number of latent factors is critical parameters. The optimum number of latent factors is determined by the

Table 2. Comparison of MLR models using different spectral pretreatment methods (acetone solution)

Solvents	Method	X value ($K = 8$, 780–1900 nm)	R_c^2	RMSEC	F value	R_p^2	RMSEP	RPD
Acetone	Raw	1446.0, 1411.0, 1485.5, 1858.0, 1863.0, 1451.5, 1506.0, 1501.5	0.2681	267.84	3.57	0.1278	349.51	0.83
	1D	1665.0, 1164.0, 1655.5, 1199.0, 1693.0, 1631.0, 16410, 1174.0	0.9513	69.09	190.47	0.9541	83.88	3.48
	2D	1704.0, 1685.0, 1671.5, 814.5, 1712.5, 848.5, 1733.5, 1419.5	0.9447	73.59	166.71	0.9515	88.73	3.29
	SG	810.0, 814.5, 993.0, 793.0, 819.0, 850.0, 882.5, 857.0	0.9889	33.02	866.78	0.9907	37.71	7.73
	SNV	1897.0, 1886.5, 1857.5, 1892.0, 1485.5, 1491.5, 1557.5, 1730.0	0.3966	243.19	6.40	0.0548	361.23	0.81
	MSC	1897.0, 1886.5, 1857.5, 1892.0, 1485.5, 1491.5, 1557.5, 1729.5	0.4048	241.54	6.62	0.0337	356.82	0.82
	NPS	1483.5, 1455.5, 1446.5, 1533.5, 1509.0, 1499.5, 1452.5, 1863.0	0.1581	287.27	1.83	0.0242	300.87	0.97
	1D&SG	820.5, 813.5, 1054.0, 810.5, 802.5, 860.5, 890.5, 1115.0	0.9583	63.97	223.80	0.9437	94.91	3.07
	2D&SG	815.0, 829.5, 917.5, 855.0, 1015.5, 850.5, 1044.0, 1143.5	0.9586	63.71	225.69	0.9693	68.26	4.27
	1D&NPS	1665.0, 1164.5, 1678.5, 1259.5, 1654.5, 1253.0, 1633.0, 1716.5	0.9439	74.17	163.96	0.9548	84.80	3.44
	2D&NPS	1704.5, 1728.5, 815.0, 849.0, 1684.0, 1671.0, 1712.5, 1419.0	0.9374	78.32	146.05	0.9565	84.70	3.44
	MSC&SG	810.0, 814.5, 933.0, 819.0, 793.0, 826.5, 1029.0, 850.0	0.9882	33.99	817.36	0.9874	43.87	6.65
	BC&NPS	1774.5, 1557.5, 1814.5, 1865.0, 1623.0, 17485, 1831.0, 1604.0	0.7125	167.86	24.26	0.6819	208.40	1.40
	BC&NPS&SG	814.5, 827.0, 803.5, 830.5, 1356.0, 819.0, 1112.0, 854.0	0.9456	73.07	169.40	0.9049	119.49	2.44
Water	Raw	1140.5, 956.0, 1639.5, 980.0, 1176.5, 1724.5, 1671.5, 1662.0	0.4040	232.23	7.11	0.2688	314.22	0.93
	1D	904.5, 1108.0, 802.5, 841.5, 1828.0, 1505.0, 819.0, 790.0	0.7519	149.84	31.82	0.5708	263.39	1.11
	2D	951.5, 813.5, 845.5, 935.5, 1352.5, 1609.5, 833.5, 1034.0	0.8952	97.43	89.74	0.8869	131.68	2.21
	SG	818.5, 8125, 809.5, 853.5, 859.0, 1052.0, 841.0, 844.5	0.9724	50.00	369.30	0.9778	59.98	4.86
	SNV	1380.0, 1786.5, 1773.0, 1547.5, 1490.5, 1519.5, 1767.5, 1832.0	0.4382	225.47	8.18	0.0213	919.46	0.32
	MSC	1379.0, 17600, 1484.0, 1532.5, 1536.0, 17855, 1845.0, 1848.5	0.6016	189.88	15.95	0.5004	290.76	1.00
	NPS	1140.5, 959.5, 1639.0, 976.5, 1178.0, 1672.0, 785.0, 1183.0	0.5288	206.49	11.78	0.1596	326.26	0.89
	1D&SG	808.5, 839.5, 1046.5, 798.5, 1052.5, 894.0, 813.0, 816.0	0.9101	90.21	106.23	0.9156	116.85	2.50
	2D&SG	814.5, 1041.0, 831.5, 828.5, 798.0, 1581.5, 1274.5, 811.5	0.9148	87.81	112.83	0.8705	140.13	2.08
	1D&NPS	904.0, 801.0, 1828.0, 1505.0, 1355.5, 1349.0, 1157.0, 1685.0	0.7813	140.67	37.51	0.4790	424.91	0.69
	2D&NPS	951.5, 1116.0, 812.0, 846.5, 1038.5, 1351.0, 1595.5, 936.5	0.8722	107.53	71.67	0.8320	159.92	1.82
	MSC&SG	818.5, 812.5, 793.0, 1691.0, 1253.0, 1068.5, 832.5, 845.0	0.9741	43.37	395.67	0.9678	74.58	3.91
	BC&NPS	1386.0, 1852.5, 1761.5, 1799.5, 815.0, 1734.5, 1858.0, 1406.0	0.6075	188.46	16.25	0.1845	917.36	0.32
	BC&NPS&SG	814.5, 1040.5, 831.5, 797.5, 811.5, 818.5, 1285.0, 1683.5	0.9477	68.79	190.30	0.9061	119.84	2.43

Table 3. PLS models set using different spectral regions

Solvents	Region, nm	R_c^2	RMSEC	R_p^2	RMSEP	PRD
Acetone	810–850	0.9933	26.37	0.9877	45.13	6.46
	810–1100	0.9976	15.25	0.9909	37.55	7.76
	810–1300	0.9963	18.89	0.9693	70.76	4.12
Water	810–850	0.9687	52.02	0.9754	65.50	4.45
	810–1100	0.9894	30.79	0.9786	51.86	5.62
	810–1300	0.8924	46.46	0.8797	140.41	2.08
	810–1500	0.9561	39.10	0.7669	191.50	1.52

Table 4. MLR model set using different spectral regions

Solvents	Region, nm	X value ($K = 8$, 780–1900 nm)	R_c^2	RMSEC	F value	R_p^2	RMSEP	PRD
Acetone	810–850	814.5, 819.0, 826.0, 846.0, 831.0, 811.5, 822.0, 837.0	0.9677	56.23	292.44	0.9691	69.22	4.21
	810–1100	933.0, 814.5, 858.0, 986.5, 1044.0, 922.5, 853.5, 926.5	0.9815	43.05	489.55	0.9838	53.55	5.44
	810–1300	933.0, 814.5, 858.0, 986.5, 1044.0, 922.5, 853.5,	0.9810	43.15	503.47	0.9810	58.61	4.97
	810–1500	1917.0933.0, 814.5, 8580, 986.5, 1044.0, 922.5, 853.5, 1917.0	0.9810	43.15	503.47	0.9810	58.61	4.97
	810–1700	933.0, 814.5, 858.0, 986.5, 1044.0, 922.5, 853.5, 1917.0	0.9810	43.15	503.47	0.9810	58.61	4.97
Water	810–850	818.5, 812.5, 844.5, 828.5, 823.5, 835.5, 835.5, 815.5	0.9641	57.02	281.75	0.9636	78.09	3.73
	810–1100	818.5, 812.5, 844.5, 1010.0, 853.5, 859.5, 828.5, 835.0	0.9720	50.33	364.53	0.9707	68.52	4.25
	810–1300	818.5, 812.5, 844.5, 1010.0, 1282.0, 1252.0, 1068.0, 861.0	0.9714	50.84	357.14	0.9504	89.35	3.26
	810–1500	818.5, 812.5, 844.5, 1010.0, 1282.0, 1252.0, 1068.0, 861.0	0.9714	50.84	357.14	0.9504	89.35	3.26
	810–1700	818.5, 812.5, 844.5, 1010.0, 1282.0, 1068.0, 861.0, 857.0	0.9715	59.00	385.28	0.9599	79.72	3.66

lowest PRESS value. Usually, the first minimum value on the PRESS plot is used to determine the optimum number of factors with the best prediction for the cross validation samples. Figure 2 shows the effect of latent factors on PRESS values for the method of SG smooth in the spectra region of 810–1100 nm (see supporting information). According to the result presented in Fig. 2 the best factor number of acetone solution in the PLS model under SG smoothing 810–1100 nm was 8, and that of the water solution was 7.

3.5. Prediction Results Using Optimum Wavelength Range

Figure 3 shows the prediction results of the PLS and MLR models set by the method of SG smoothing within the spectral region of 810–1100 nm. The results show that this calibration model set by the selected method and spectral region has a favorable ability for making predictions, for its values of R_{cal}^2, R_{val}^2 were quite near to 1, and its values of SEC, SEP were very low.

There is some distinct regularity in the absorption characteristics of C–H bond and O–H bond of NIR. The intension of their NIR spectral message is related to the concentration of the samples, and the relationship may be directly correlated even linear in short wave-NIR. Therefore, complete differentiation of C–H bond and O–H bond for NTR absorption can be presented in short wave-NIR.

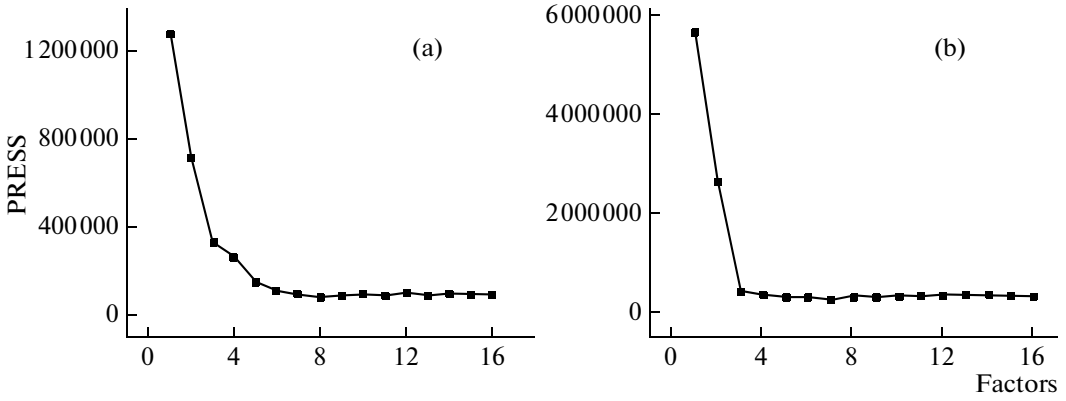

Fig. 2. Effect of latent factors on PRESS values for the method of SG smooth in 810–1100 nm: acetone solution (a), water solution (b).

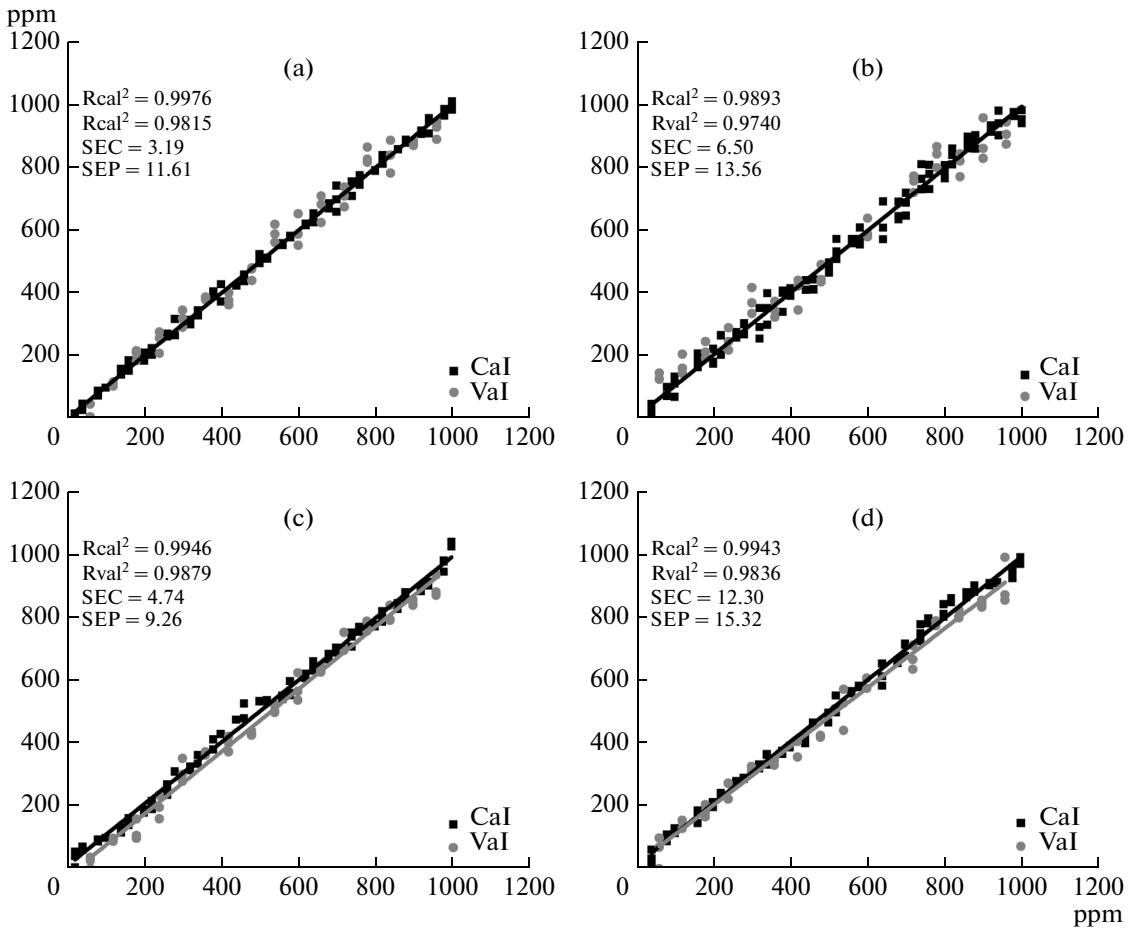

Fig. 3. Prediction results of PLS (a, b) and MLR (c, d) models set by the method of Savitzky–Golay smoothing within the spectral region of 810–1100 nm, acetone (a, c) and water (b, d) solutions.

4. CONCLUSIONS

Our work has studied the absorption characteristics of C–H bond and O–H bond of NIR. We found that there is some special regularity in the absorption characteristics of C–H bond and O–H bond of NTR in short-wave NIR region, and good calibration model could be set by pertinent chemometrics method, and the so-called "good model" could successfully make

accurate predictions for the sample. Short-wave NIR region has truly been found as characteristic absorption of C–H bond and O–H bond of NIR, but further research should be required of the features of C–H, O–H, N–H, S–H and their deep relationship between each other.

ACKNOWLEDGMENTS

This work was financially supported by the Ministry of Science and Technology of China Major Special Project "Significant Creation of New Drugs" (no. 2010ZX09502-002), National Natural Science Foundation of China (no. 81303218) and Innovation Team Foundation of Beijing University of Chinese Medicine, Beijing (2011-CXTD-11, Research Centre of TCM-information Engineering).

REFERENCES

1. G. Reich, Adv. Drug Deliv. Rev. **57**, 1109–1143 (2005).
2. W. F. McClure, J. Near Infrared Spectrosc. **11**, 487–518 (2003).
3. J. Luypaert, D. L. Massart, and Y. V. Heyden, Talanta **72**, 865–883 (2007).
4. Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, and N. Jent, J. Pharm. Biomed. Anal. **44**, 683–700 (2007).
5. Z. S. Wu, M. Du, B. Xu, Z. Z. Lin, X. Y. Shi, and Y. J. Qiao, J. Mol. Struct. **1019**, 97–102 (2012).
6. M. T. J. Benito, C. B. Ojeda, and F. S. Rojas, Appl. Spectrosc. Rev. **43**, 452–484 (2008).
7. B. M. Nicolai, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K. I. Theron, and J. Lammertyn, Postharvest Biol. Tec. **46**, 99–118 (2007).
8. H. Chung, Appl. Spectrosc. Rev. **42**, 251–285 (2007).
9. Z. S. Wu, M. Du, C. L. Sui, X. Y. Shi, and Y. J. Qiao, Anal. Methods **4**, 1084–1088 (2012).
10. B. Xu, Z. S. Wu, Z. Z. Lin, C. L. Sui, X. Y. Shi, and Y. J. Qiao, Anal. Chim. Acta **720**, 22–28 (2012).
11. Z. S. Wu, B. Xu, M. Du, C. L. Sui, X. Y. Shi, and Y. J. Qiao, J. Pharm. Biomed. Anal. **62** 1–6 (2012).
12. M. Alcala, J. Leon, J. Ropero, M. Blanco, and R. J. Romanach, J. Pharm. Sci. **97**, 5318–5327 (2008).
13. M. Blanco, M. Castillo, A. Peinado, and R. Beneyto, Anal. Chim. Acta **581**, 318–323 (2007).
14. M. J. Lee, D. Y. Seo, H. E. Lee, I. C. Wang, W. S. Kim, M. Y. Jeong, and G. J. Choi, Int. J. Pharm. **403**, 66–72 (2011).
15. W. L. Li and H. B. Qu, J. Pharm. Biomed. Anal. **52**, 425–431 (2010).
16. W. L. Li, L. H. Xing, L. M. Fang, J. Wang, and H. B. Qu, J. Pharm. Biomed. Anal. **53**, 350–358 (2010).
17. S. Kuriakose and H. Joe, Food Chem. **135**, 213–218 (2012).