NEAR-INFRARED SPECTROSCOPY ANALYSIS TECHNOLOGY BASED ON SINGLE SAMPLE**

Z. Wei, M. Lin*

College of Metrology and Measurement Engineering at China Jiliang University, Hangzhou 310018, China; e-mail: linm@cjlu.edu.cn

Application of near-infrared spectroscopy to the prediction of sample content is strongly limited by signal peak overlap. To analyze the spectral information directly related to the target components and to make the chemometric model more explanatory, an independent characteristic projection algorithm is proposed. The algorithm was applied to the independent spectral analysis of a single sample using corn as a representative example. Moisture, oil, protein, and starch, which are the four main components of corn, were the target components. The pure component spectra were used the projection directions to decompose the near-infrared spectrum of a single corn sample; then four decomposed spectra corresponding to the four pure component spectra were obtained. Their corresponding relationship was determined using their correlation coefficients and by comparing their characteristic peaks, and the molecular absorption patterns corresponding to the characteristic absorption peaks of each decomposed spectrum were analyzed in detail. The theoretical analysis and experimental results indicate that the independent characteristic projection algorithm can be applied to single-sample spectral analysis to extract more complete physicochemical information about the target components and provide a theoretical basis for establishing a robust near-infrared spectral chemometric model with great extrapolation capability and stability.

Keywords: *near-infrared spectral analysis, single sample, independent characteristic projection, pure component spectrum, decomposed spectrum.*

ТЕХНОЛОГИЯ СПЕКТРАЛЬНОГО АНАЛИЗА В БЛИЖНЕЙ ИК-ОБЛАСТИ НА ОСНОВЕ ОДНОГО ОБРАЗЦА

Z. Wei, M. Lin*

УДК 543.42

Колледж метрологии и измерительной техники Китайского университета Цзилиан, Ханчжоу 310018, Китай e-mail: linm@cjlu.edu.cn

(Поступила 21 апреля 2020)

Применение ближней ИК-спектроскопии для прогнозирования состава образца сильно ограничено перекрытием пиков сигнала. Для анализа спектральной информации, непосредственно связанной с изучаемыми компонентами, и для того, чтобы сделать хемометрическую модель более понятной, предлагается алгоритм независимой проекции характеристик. Алгоритм применен к спектральному анализу одного образца кукурузы. Исследованы четыре основных компонента кукурузы – влага, масло, белок и крахмал. Спектры чистых компонентов использованы в направлениях проекций для разложения ближнего ИК-спектра одного образца кукурузы, затем получены четыре разложенных спектра, соответствующих четырем спектрам чистых компонентов. Их соотношения определялись с использованием коэффициентов корреляции и путем сравнения характеристических пиков. Детально проанализированы картины молекулярного поглощения, соответствующие характеристическим пикам поглощения каждого разложенного спектра. Теоретический анализ и экспериментальные результаты показывают, что алгоритм независимой проекции характеристик может быть применен к

** Full text is published in JAS V. 88, No. 3 (http://springer.com/journal/10812) and in electronic version of ZhPS V. 88, No. 3 (http://www.elibrary.ru/title_about.asp?id=7318; sales@elibrary.ru).

спектральному анализу отдельных образцов для извлечения более полной физико-химической информации о компонентах и обеспечения теоретической основы для создания надежной хемометрической модели в ближней ИК-области с отличной экстраполяцией и стабильностью.

Ключевые слова: *спектральный анализ в ближней инфракрасной области, одиночный образец, независимая характеристическая проекция, спектр чистых компонентов, разложенный спектр.*

Introduction. Near-infrared spectroscopy (NIRS) is a mainstream nondestructive technique that provides convenient, high-efficiency, low-cost, and real-time online detection. It is widely used in fields such as agriculture [1, 2], industry [3, 4], commerce [5, 6], biomedicine [7, 8], environmental monitoring [9], and aerospace technology [10]. NIRS provides characteristic information about frequency overtones and combined vibrational absorption frequencies of molecular hydrogen-containing groups (O–H, N–H, C–H, etc.). The more complex the composition of the substance, the more likely the overlap of the characteristic spectral peaks. Therefore, there is no one-to-one correspondence between near-infrared absorption peaks and the chemical components from which they derive. The traditional chemometric model only provides the corresponding relationship between the spectrum and the content or properties of the substance without fully considering the underlying physical and chemical significance in the spectrum, an approach that can yield unconvincing results [11, 12]. To extract more meaningful information about the target components from the spectrum, it is necessary to improve the chemometric model, which is an active and challenging research area in the NIRS field [13].

In recent years, significant advancements have been made in methods of near-infrared spectroscopic analysis. Yang et al. [14] proposed a spectral peak detection algorithm, CWT–IS, based on the continuous wavelet transform (CWT) and image segmentation (IS). This method can effectively eliminate the adverse effects caused by noise and baseline, clarify the location of characteristic peaks, and facilitate subsequent spectral analysis. Sarraguça et al. [15] verified the feasibility of applying net analyte signal (NAS) theory to the quality control of solid dosage forms in the pharmaceutical industry, providing quantitative and qualitative analysis of active product ingredients in different pharmaceutical formulations. Furthermore, the theory can improve NIRS interpretability and has the advantages of high sensitivity and good selectivity. Lü et al. [16] presented a synchronous two-dimensional correlation spectroscopy combining self-peak and cross-peak information method for selecting the characteristic wavelengths, which introduced interpretability into variable selection and greatly reduced the calculation complexity. Yu et al. [17] proposed a weighted clustering and pruning of wavelength variables partial least squares (WCPV–PLS) method, which can improve the prediction accuracy by reducing the number of wavelength variables. This method has been applied to the content prediction of corn components and has strong robustness. Yang et al. [18] developed a new variable selection method named moving-window partial least-squares coupled with sampling error profile analysis (SEPA-MWPLS). This method shows good stability and reliability in variable selection, model establishment, and physicochemical prediction using spectral datasets obtained from corn and pharmaceutical tablets. Yang et al. [19] used the successive projections algorithm (SPA), information gain, and the Gini index to select the feature bands, combined with a particle swarm optimization–extreme learning machine (PSO–ELM) model, to identify eight tree species at the leaf level with a strong recognition effect. Jiang et al. [20] combined spectral multi-band selection with Savitzky–Golay preprocessing, the correlation coefficient method, and the synergy interval partial least squares (siPLS) algorithm to quantitatively analyze corn components, which significantly simplified the model complexity. Tao et al. [21] introduced an algorithm based on time–frequency domain fractal dimension analysis, which combines wavelet multi-scale observation and measurement of the fractal self-similarity to identify and analyze overlapping spectral peaks.

Despite the above progress, these analytical techniques remain limited to identifying spectral bands associated with the target components, which inevitably leads to some information loss. In addition, the algorithms can be influenced by the choice of initial values and the number of sample components, which need to be determined from the sample characteristics. More importantly, methods for multi-component NIRS analysis of single samples have not evolved. To extract spectral information with practical significance from the substance spectrum as completely as possible, and to establish a robust near-infrared spectral model, this study introduces an independent characteristic projection (ICP) algorithm. This method converts the single sample spectral data into a matrix, with each row containing information about a target component.

Using corn as a test sample, its moisture, oil, protein, and starch were selected for target component analysis. The pure component spectra of moisture, corn oil, corn protein, and corn starch were obtained by independent experiments and regarded as the projection directions in the ICP algorithm. Four decomposed spectra were then obtained from the corn spectrum, and the relationship between the pure component spectra and

decomposed spectra was determined by calculating the correlation coefficient and matching their similar characteristic peaks.

Theory and methods. *Summary of ICP algorithm.* The near-infrared spectrum of a substance comprises mutually independent pure component spectra. Based on this property, the ICP algorithm converts single-sample, one-dimensional spectral data, combined with the pure component spectrum matrix, into a decomposed matrix whose individual rows contain information about each target component. The process is similar to a vector projection, which regards the pure component spectrum as a basis vector, also called the basis spectrum. The ICP algorithm projects the original spectrum of the substance onto each basis spectrum.

The algorithm is an improved method based on independent component analysis (ICA). Traditional ICA decomposes a signal comprising several independent sources into independent components. The disadvantage of ICA is that such a decomposition cannot be conducted using only a single observation channel, but synchronous observations must be made using multiple channels combined by different mixing ratios of each source. Thus, traditional ICA cannot decompose single-sample, one-dimensional spectral data into spectra of each target component. To address this shortcoming, the ICP algorithm improves the input matrix of ICA.

The input matrix of traditional ICA that is applied to near-infrared spectroscopic analysis [22] comprises multiple sample spectral data. Suppose a_i, b_i, and c_i represent the different mixing ratio coefficients of each single component spectrum x_i of the sample, where $i = 1, 2, \ldots n$. As the input for the ICP algorithm, only one sample spectrum is retained. The mixing ratio coefficient of other input channels is set to 0 or 1, the latter representing the pure component spectrum of the target components.

Principle of ICP algorithm. Figure 1 shows the flow diagram of the ICP algorithm. Here, s is the single sample one-dimensional spectral data, and r_i ($i = 1, 2,\ldots N–1$) are the one-dimensional pure component spectral data of known components, each of which have n wavelength numbers. These spectra are combined to form a new matrix x containing N rows. There are N–1 sample target component, which is also the number of decomposed layers in the algorithm. That is, $R = [r_1, r_2, \ldots, r_{N-1}]^T$ constitutes a pure component spectrum matrix. The specific steps of the ICP algorithm are as follows:

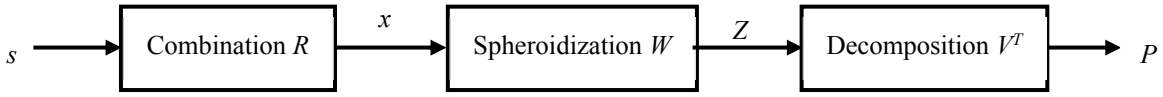

Fig. 1. Flow diagram of the independent characteristic projection algorithm.

1. Construct the input matrix; s and R were combined to form a new matrix $x = [s, r_1, r_2, \ldots, r_{N-1}]^T$.

2. Conduct spheroidization. Spheroidize x to obtain Z that contains orthogonal normalization for each row. That is, $Z = Wx$, W is the spheroidization matrix.

3. Explain the shape of the weight matrix M; $K(Z)$ contains the total fourth-order cumulants of Z, $Q_Z(M)$ is a set of four-dimensional cumulant matrices obtained by weighting $K(Z)$, and M is the weight matrix that generates Q_Z. Construct a weight matrix M from a set of $N \times N$ symmetric basis matrices, whose elements M_{pq} are defined as

$$M_{pq} = \begin{cases} e_p e_q^T, & p = q, \\ \frac{1}{\sqrt{2}}\left[e_p e_q^T + e_q e_p^T \right], & p < q, \\ \frac{1}{\sqrt{2}}\left[e_p e_q^T - e_q e_p^T \right], & p > q, \end{cases} \quad (1)$$

where e_p and e_q are $N \times 1$ unit vector, all of whose elements equal 0 except for elements p and q, which are equal to 1.

4. Choose a fourth-order cumulative matrix. Take M as the weight matrix, then construct the four-dimensional cumulant matrix $Q_Z(M)$ of Z:

$$\left[Q_Z(M) \right]_{ij} = \sum_{k=1}^{N} \sum_{l=1}^{N} K_{ijkl}(Z) m_{pq} , \quad (2)$$

where $i,j = 1, 2, \ldots, N$, $Q_Z(M)$ is $N \times N$ matrix, $K_{ijkl}(Z)$ is the four-dimensional cumulant of the i, j, k, l components of Z, and m_{pq} is the p, q element of matrix M.

5. Conduct diagonalization. According to the definition of the four-dimensional cumulant matrix, $Q_Z(M)$ is a real symmetric matrix, so $Q_Z(M)$ can be diagonalized as follows:

$$Q_Z(M) = V\Lambda(M)V^T, \tag{3}$$

where $\Lambda(M)$ is a diagonal matrix whose diagonal elements contain the N eigenvalues of $Q_Z(M)$. It is required to obtain a matrix V that can jointly diagonalize each $Q_Z(M)$.

6. Obtain the output matrix. Finally, the decomposed matrix can be obtained as

$$P = Dx = V^T Wx, \tag{4}$$

where $P = [p_1, p_2, \ldots, p_{N-1}]^T$ is the decomposed spectral matrix of s, and D is the conversion matrix.

Experimental. *Instrument and reagents.* Pure moisture, corn oil, corn protein, and corn starch reagents were purchased from Shanghai Jinsui Bio-Technology Co., Ltd. A SupNIR-1500 series portable near-infrared analyzer (Focused Photonics Inc., Hangzhou), was used to collect the pure substance spectra. Spectral data for the corn samples were downloaded from http://www.eigenvector.com/data/Corn/index.html, which provides near-infrared data of 80 corn samples measured using three different near-infrared spectrometers (designated "m5," "mp5," and "mp6"). The samples contain approximately 10% moisture, 3.5% oil, 8.5% protein, and 64% starch. To simulate the transformation of spectral characteristic peaks when each single component of corn is retained and removed, pure moisture, corn oil, corn protein, and corn starch reagents were prepared in five mixtures in the proportions shown in Table 1.

TABLE 1. Relative Proportions of Components in Five Mixtures

Mixture	Moisture	Corn oil	Corn protein	Corn starch
1	3	1	3	18
2	0	1	3	18
3	3	0	3	18
4	3	1	0	18
5	3	1	3	0

Experimental method. Spectral data (1100–2500 nm, 2 nm resolution) measured using "instrument m5" were used for analysis. Four pure-substance samples and five prepared mixtures (Table 1) were assembled separately into the ∅15-mm quartz reflection colorimetric dish and placed on the instrument probe successively for spectrum collection. Spectra were acquired at room temperature using an empty ∅15-mm quartz reflection colorimetric dish as a reference and an average of 30 scans. The corn sample spectral data and the spectrum of the mock corn sample mixture 1 is shown in Fig. 2. The high similarity between them indicates that the mixture is representative of a real corn sample.

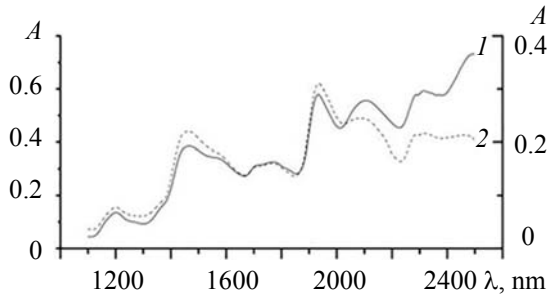

Fig. 2. Near-infrared spectra of (1) corn sample and (2) mixture 1.

Results and discussion. *Spectral analysis of mixtures.* Figure 3 compares the spectra of mixtures 1–5 and shows the transformation of characteristic peaks when individual components of corn are removed. The left ordinate axis corresponds to the mixture 1 spectrum, and the right axis corresponds to the spectra of mixtures 2–5. As shown in Table 1, mixture 1 is a mock sample prepared using all the corn components, while mixtures 2, 3, 4, and 5 contain all components except moisture, oil, protein, and starch, respectively.

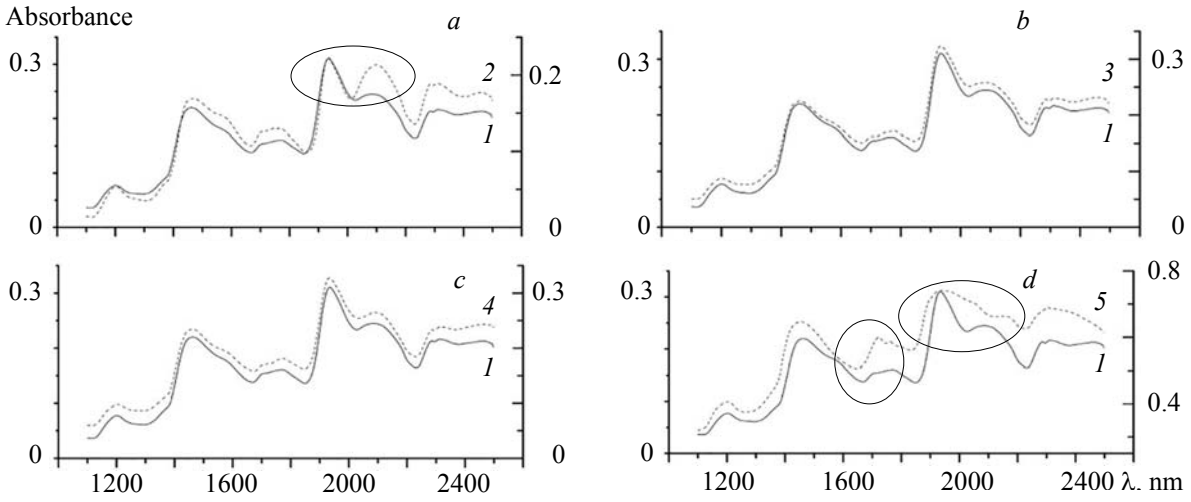

Fig. 3. Spectral comparison of mixture 1 with (a) mixture 2, (b) mixture 3, (c) mixture 4, and (d) mixture 5.

In Fig. 3a the absorbance difference between the hydroxyl absorption band at 1940 nm (assigned to moisture) and the adjacent peak is visibly smaller in the spectrum of mixture 2 compared with that of mixture 1, consistent with the omission of the moisture component from mixture 2. In Fig. 3d, the methylene absorption bands assigned to oil at 1725 and 1762 nm are visible in the spectrum of mixture 5 because of the absence of the starch component; other absorption bands assigned to starch at 1930 and 2100 nm in the spectrum of mixture 1 are not seen in mixture 5, consistent with the removal of this component. After removing individual corn components, the characteristic peaks of the remaining sample components are not seen in the spectrum except in three obvious regions in Fig. 3. This is because the spectral peaks strongly overlap, and, in particular, the low characteristic absorption corresponding to low-abundance components is difficult to observe in the spectrum. As shown in Fig. 3b,c, no visible spectral changes are observed when the low-abundance oil and protein components are individually removed from the mock corn sample. However, the physicochemical information about each sample component is contained in the spectrum.

Figure 4 shows the difference spectra of mixture 1 and mixtures 2–5 and compares them with the corresponding spectra of pure corn components. As shown in Fig. 4, the difference between the spectra of mixture 1 (the mock corn sample) and mixture 2 (the mock corn sample without moisture) is similar to the moisture spectrum; the difference spectrum therefore provides physicochemical information about the sample moisture. By analogy, the difference between mixtures 1 and 3, 1 and 4, and 1 and 5, can provide information about the oil, protein, and starch components, respectively.

It is clear that the sample spectrum contains physicochemical information about each component, but this information cannot be clearly observed because of spectral overlap. The ICP algorithm proposed herein addresses this issue by extracting the pure component physicochemical information in the form of decomposed spectra.

Component information corresponding to each decomposed spectrum. The spectra of pure moisture, corn oil, corn protein, and corn starch components were acquired and used as the projection directions, and the spectra of the corn samples were then decomposed to obtain physicochemical information. Because there are four projection directions, the number of decomposed layers in the algorithm was set to 4. Using the ICP algorithm, four decomposed spectra representing the individual moisture, corn oil, corn protein, and corn starch components were obtained and numbered by their row order in the decomposed matrix.

After the decomposed matrix was obtained, the relationship between the four decomposed spectra and the spectra of the pure components was determined by calculating the correlation coefficient across the whole spectrum and matching similar characteristic peaks. The comparison between the decomposed spectrum and the pure component spectrum is shown in Fig. 5. The vertical coordinate of the decomposed spectrum is the absorption coefficient.

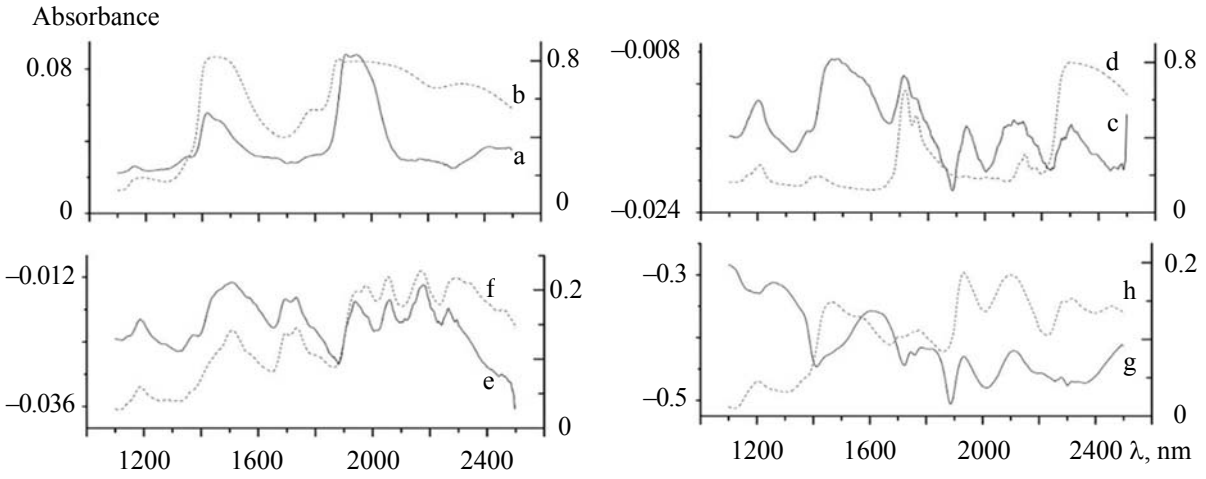

Fig. 4. Comparison of the difference spectra obtained from the mixtures shown in Table 1 and the corresponding pure component spectra: a) difference spectrum of mixtures 1 and 2; b) moisture spectrum; c) difference spectrum of mixtures 1 and 3; d) oil component spectrum; e) difference spectrum of mixtures 1 and 4; f) protein component spectrum; g) difference spectrum of mixtures 1 and 5; h) starch component spectrum.

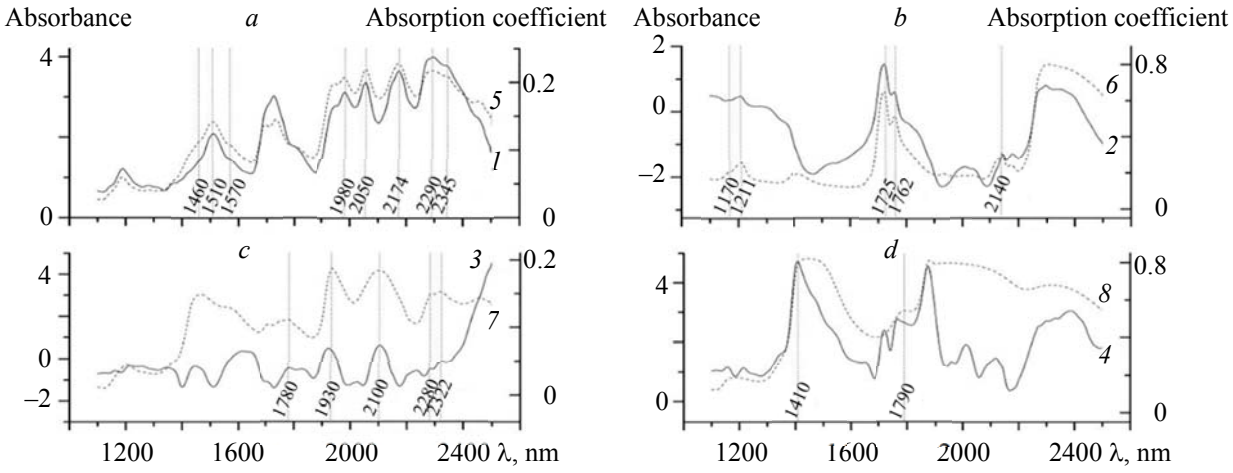

Fig. 5. Comparison of (a) decomposed spectrum 1 and corn protein spectrum (5); (b) decomposed spectrum 2 and corn oil spectrum (6); (c) decomposed spectrum 3 and corn starch spectrum (7); (d) decomposed spectrum 4 and moisture spectrum (8).

The highest correlation coefficient (r) was obtained for the decomposed spectrum 1 and the corn protein spectrum, reaching 0.95. Figure 5a confirms that these spectra contain several identical characteristic peaks and share similar spectral morphology, which shows that the decomposed spectrum 1 provides information about the protein component of the corn sample. Similarly, the decomposed spectra 2, 3, and 4 corresponded to the spectra of corn oil (r = 0.5867), corn starch (r = 0.2231), and moisture (r = 0.4722), respectively.

Protein is a polypeptide chain comprising amino acids and forms a spatial structure by winding and folding. In Fig. 5a, decomposed spectrum 1 and the corn protein spectrum have similar characteristic absorption peaks in the 1400–1600 and 1900–2400 nm regions. Consistent with the structure of protein, 1460, 1510, and 1570 nm are the first-order frequency-doubling bands of amide N–H fundamental symmetric stretching vibration, 1980 nm is the combined band of the amide N–H antisymmetric stretching and in-plane bending vibrations, 2050 nm is the combined band of N–H and C=O stretching vibrations, 2174 nm is the combined band of N–H bending and C–N stretching vibrations, 2290 nm is the C=O fundamental stretching vibrational absorption band, and 2345 nm is the combined spectral band of symmetric stretching and bending vibrations of protein side chain methylene C–H groups [23]. These spectral bands belong to the similar characteristic absorption of decomposed spectrum 1 and the corn protein spectrum.

Vegetable oil is rich in many fatty acids. In Fig. 5b, decomposed spectrum 2 and corn oil spectrum have similar characteristic absorption peaks in the 1100–1300, 1700–1800, and 2100–2200 nm regions. According to the vegetable oil component characteristic, 1170 nm is the second-order frequency-doubling band of olefin C–H fundamental stretching vibration, 1211 nm is the second-order frequency-doubling band of methylene C–H fundamental stretching vibration, 1725 nm is the first-order frequency-doubling band of methylene C–H fundamental stretching vibration, 1762 nm is the first-order frequency-doubling band of methylene C–H fundamental symmetric stretching vibration, and 2140 nm is the combined spectral band of C–H and C=O stretching vibrations [23]. These spectral bands belong to the similar characteristic absorption of decomposed spectrum 2 and the corn oil spectrum.

Starch is polymerized from glucose molecules; its chemical formula is $(C_6H_{10}O_5)_n$. In Fig. 5c, decomposed spectrum 3 and the corn starch spectrum have similar characteristic absorption peaks in the 1700–1800, 1900–2000, 2050–2150, and 2200–2400 nm regions. Combined with starch structure, 1780 nm is the first-order frequency-doubling band of methylene C–H fundamental stretching vibration, 1930 nm is the O–H fundamental stretching vibrational absorption band, and 2100 nm is the third-order frequency-doubling band of the polymer C=O–O fundamental stretching vibration. Features at 2280 and 2322 nm correspond to the combined spectral bands of C–H stretching and deformation vibrations [23]. These spectral bands belong to the similar characteristic absorption of decomposed spectrum 3 and the corn starch spectrum.

The main functional group of moisture is the O–H bond. In Fig. 5d, the decomposed spectrum 4 and the moisture spectrum have similar characteristic absorption peaks in the 1400–1500 and 1700–1800 nm regions. Combined with the molecular structure of moisture, 1410 nm is the first-order frequency-doubling band of O–H fundamental stretching vibration, and 1790 nm is the combined spectral band of O–H stretching and bending vibrations [23]. These spectral bands belong to the similar characteristic absorption of decomposed spectrum 4 and moisture spectrum.

Analysis using the ICP algorithm showed that the characteristic peaks in the decomposed spectrum can be attributed to the different absorption patterns of each component's functional groups, which indicates that the method for NIRS analysis of single sample is more explanatory than the traditional chemometric model. In addition to the above-mentioned characteristic absorption bands, the absorption peaks at other wavelengths are considered to result from the overlap of various substance molecule absorption patterns.

Comparison of the corn sample spectrum and each decomposed spectrum. Comparing the near-infrared spectrum of the corn sample with the four decomposed spectra obtained using the ICP algorithm (Fig. 6), it can be seen that the characteristic absorption of some sample components is directly reflected in the corn sample spectrum. Some characteristic absorption bands overlap with or are close to those of other components, which increases the absorbance in this region but does not reflect the physicochemical information of this component in the spectrum. Other bands are obscured by more intense nearby characteristic absorption bands associated with higher-abundance sample components.

The characteristic absorption bands belonging to protein at 1460 and 1570 nm, oil at 1211 and 1725 nm, and starch at 1930, 2100, and 2322 nm, are clearly observed in the spectrum without interference from other overlapping or adjacent characteristic absorption bands (case 1).

The adjacent characteristic absorption bands belonging to protein at 2290 nm and to starch at 2280 nm are close to absorption peak at 2288 nm observed in the spectrum of the corn sample, which suggests that the characteristic absorption at 2288 nm is caused by two components, protein and starch (case 2). A similar situation also occurs for the 1762, 1780, and 1790 nm peaks of oil, starch, and moisture, respectively, yielding the characteristic absorption at 1774 nm in the corn sample spectrum.

The characteristic absorption of starch component mainly appears in the 1700–2400 nm region. The high abundance of starch in the sample obscures characteristic peaks of lower-abundance components in this region (case 3). For example, the characteristic absorption band at 1980 nm belonging to protein is obscured by the nearby starch absorption band at 1930 nm. Similarly, the characteristic absorption peaks of protein at 2050, 2174, and 2345 nm, and of oil at 2140 nm, are also obscured. Decomposition of the spectrum of the corn sample shows that the near-infrared spectrum of a substance is a superposition of the individual absorption patterns of pure components. The decomposed spectra of the pure components within the corn sample can be obtained using the ICP algorithm, and physicochemical analysis of each underlying component in the substance spectrum can be made, which provides a theoretical basis for establishing a more interpretive model of near-infrared spectral analysis.

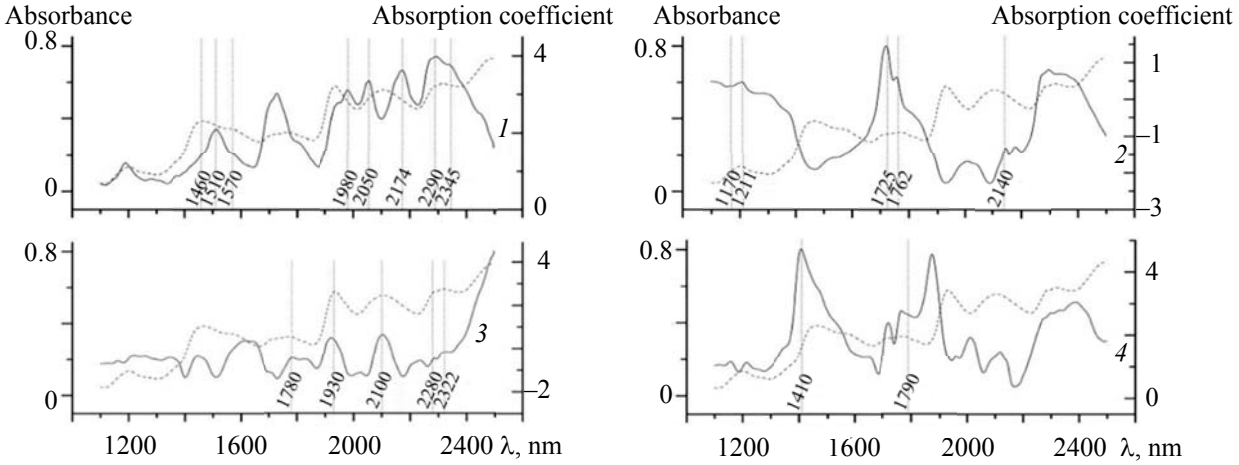

Fig. 6. Comparison of (dotted line) corn spectrum with (solid line) decomposed spectrum 1, decomposed spectrum 2, decomposed spectrum 3, and decomposed spectrum 4.

Conclusions. An independent characteristic projection algorithm for near-infrared spectral analysis was proposed. The analytical method was used to project the original sample spectrum onto the different pure component spectra, and the obtained decomposed spectra were compared with the target component information to extract characteristic information about the vibrational absorptions of different molecular hydrogen-containing groups in the relevant wavelength range. The feasibility of the method was verified by investigating the physicochemical information of moisture, corn oil, corn protein, and corn starch in corn samples. The experimental results suggest that the algorithm has no limit on the number of input sample spectra and can be applied to the near-infrared spectral analysis of a single sample. The decomposition of each sample spectrum is independent, in that the spectral peaks of each target component can be effectively separated and analyzed. Moreover, it is not necessary to determine the algorithm parameters (the number of decomposed layers) based on the sample characteristic, and the parameters are only related to the preset projection directions. In addition, compared with the traditional spectral analysis method to find the most relevant wavelength region of the sample spectrum, the ICP algorithm ensures the integrity of the component information to the greatest extent, which could provide a theoretical basis for establishing a robust chemometric model. The algorithm may also be of value for spectral analysis research in other fields.

Acknowledgments. This project was performed with financial support from the National Major Scientific Instruments and Equipment Development Projects (No. 2014YQ470377).

REFERENCES

1. D. Girolamo, C. Holst, M. Cortese, S. Cervellieri, M. Pascale, F. Longobardi, L. Catucci, A. C. R. Porricelli, V. Lippolis, *Food Chem*., **282**, 95–100 (2019).
2. K. H. S. Peiris, S. R. Bean, A. Chiluwal, R. Perumal, S. V. Krishna Jagadish, *Cereal Chem*., **96**, 678–688 (2019).
3. J. C. L. Alves, R. J. Poppi, *Talanta*, **104**, 155–161 (2013).
4. R. M. Balabin, R. Z. Safieva, *Anal. Chim. Acta*, **689**, 190–197 (2011).
5. R. J. Yang, G. M. Dong, X. S. Sun, Y. P. Yu, H. X. Liu, Y. R. Yang, W. Y. Zhang, *Anal. Methods*, **7**, 4302–4307 (2015).
6. Z. D. Liu, W. X. Li, Z. H. Wei, *Text. Res. J*., **90**, 1057–1066 (2020).
7. L. Miao, Y. Liu, H. Li, Y. P. Qi, F. Lu, *Drug Test. Anal.,* **9**, 221–229 (2017).
8. J. Luypaert, D. L. Massart, Y. Vander Heyden, *Talanta*, **72**, 865–883 (2007).
9. X. Zhang, P. P. Liu, B. L. Pan, M. J. Wei, Z. Y. Zhang, *J. Near Infrared Spectrosc*., **27**, 370–378 (2019).
10. M. A. Carvalho, M. Talhavini, M. F. Pimentel, J. M. Amigo, C. Pasquini, S. A. Junior, I. T. Weber, *Anal. Methods*, **10**, 4711–4717 (2018).
11. X. H. Bian, P. Y. Diwu, C. X. Zhang, L. G. Lin, G. H. Chen, X. Y. Tan, Y. G. Guo, B. W. Cheng, *Anal. Chim. Acta*, **1009**, 20–26 (2018).
12. P. B. Harrington, *Anal. Chim. Acta*, **1010**, 20–28 (2018).

13. Y. L. Pei, Z. S. Wu, X. Y. Shi, L. W. Zhou, Y. J. Qiao, *Spectrosc. Spectral Anal*., **34**, 2391–2396 (2014).
14. G. F. Yang, J. C. Dai, X. J. Liu, M. Chen, X. L. Wu, *Anal. Methods*, **12**, 169–178 (2020).
15. M. C. Sarraguça, J. A. Lopes, *Anal. Chim. Acta*, **642**, 179–185 (2009).
16. C. X. Lü, L. J. Chen, Z. L. Yang, X. Liu, L. J. Han, *Appl. Spectrosc*., **68**, 844–851 (2014).
17. S. H. Yu, J. Liu, *Anal. Methods*, **11**, 4593–4599 (2019).
18. W. Y. Yang, W. M. Wang, R. Q. Zhang, F. Y. Zhang, Y. R. Xiong, T. Wu, W. C. Chen, Y. P. Du, *Anal. Sci.*, **36**, 303–309 (2020).
19. R. Yang, J. Kan, *J. Appl. Spectrosc.*, **87**, 184–193 (2020).
20. W. W. Jiang, C. H. Lu, Y. J. Zhang, W. Ju, J. Z. Wang, M. X. Xiao, *Anal. Methods*, **11**, 3108–3116 (2019).
21. W. L. Tao, Y. Liu, X. P. Wang, Q. S. Wu, *Spectrosc. Spectral Anal*., **37**, 3664–3669 (2017).
22. L. M. Fang, M. Lin, *Chinese J. Anal. Chem*., **36**, 815–818 (2008).
23. J. Workman. Jr., L. Weyer, *Practical Guide to Interpretive Near-Infrared Spectroscopy*. CRC Press, Inc., Boca Raton, Florida, 240–263 (2007).