Chemometrics and Intelligent Laboratory Systems 173 (2018) 14–20

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Applying Tchebichef image moments to quantitative analysis of the components in complex samples based on raw NIR spectra

Jin Jin Liu, Bao Qiong Li, Hong Lin Zhai *, Xue Wang, Min Li Xu

College of Chemistry & Chemical Engineering, Lanzhou University, Lanzhou 730000, PR China

ARTICLE INFO

Abbreviations:
TMs
Tchebichef image moments
TM-PLS
Tchebichef moment method combined with Partial least squares
SC-3D
Self-construction three-dimension
iPLS
interval partial least square

Keywords:
Near infrared spectroscopy
Quantitative analysis
Tchebichef image moments
Interval partial least square

ABSTRACT

The interferences of irrelevant information, overlapping and shifts of peaks appear mostly in near infrared (NIR) spectroscopy, especially in complex samples, which seriously impede the accurate quantification. In this work, the features of raw NIR spectra represented by Tchebichef image moments (TMs) were employed to partial least square (PLS) modeling. The proposed strategy was applied to quantitative analysis of the components in complex samples based on their raw NIR spectra, and the obtained models were strictly evaluated by their statistical parameters. Our study indicates that the information in raw NIR spectra can be reorganized and represented by TM method owing to its powerful multi-resolution capability and inherent invariance property, which is beneficial to extract the important information of target components. Compared with the PLS and interval partial least square (iPLS) method, the proposed approach could provide accurate and reliable analytical results. Therefore, as an efficient pretreatment method, TMs can be used to improve the analytical precision of PLS based on conventional NIR spectra.

1. Introduction

Near-infrared (NIR) spectroscopy techniques have been widely used in agriculture [1], food [2], chemical industry [3] and pharmaceutical [4] owing to its simplicity, fastness, accuracy, non-destructive, convenience and so on. However, the raw spectra collected from instruments contains not only beneficial information for a calibration but also redundant and irrelevant information. The interferences of irrelevant information, overlapping and shifts of peak in NIR spectra are inevitable, and seriously affect the quantitative analytical results. Although baseline correction and smoothing have been used in experiments for decreasing noise signal, it may lead to the loss of the target signals or unsatisfactory treatment about the noise due to the absence of objective standards. Multivariate calibration methods such as principal component regression (PCR) [5,6], partial least squares (PLS) [7,8], artificial neural network (ANN) [9,10] and support vector regression (SVR) [11,12] have been the key techniques to analysis NIR spectra. Traditional multivariate calibration methods are difficult to find the quantitative relationship from the raw NIR spectra directly. Therefore, to obtain a satisfactory analytical results for complex samples, apart from the improving experimental conditions to obtain better NIR spectra, several pretreatment techniques such as multiplicative scattering correction [13,14], the standard normal variate [15,16], orthogonal signal correction [17,18], continuous wavelet transform [19,20] and iPLS [21,22] have been presented to remove the background and noise in the raw spectra so as to enhance the reliability of NIR spectral analysis. Nevertheless, for quantitative analysis in some complex cases, the performances of many conventional pretreatments methods with the time-consuming and costly are not ideal due to the highly overlapping spectral bands as well as messy signals in the NIR spectra, especially when a great number of complex samples need to be analyzed. A near-infrared (NIR) spectrum usually consists of hundreds or even thousands of measurements or channels, theoretical and experimental evidence indicates that some form of wavelength selection can improve the predictive form a PLS regression model by avoiding the irrelevant information embedded in the response data [23–25], there is still the risk that eliminating all other intervals, as is done iPLS, might result in loss of useful information [26]. Thus, the direct extraction of useful information from raw NIR spectra is still an enormous challenge for building high quality models. It is necessary to develop some simple and effective chemometric methods.

* Corresponding author.
E-mail address: zhaihl@163.com (H.L. Zhai).

https://doi.org/10.1016/j.chemolab.2017.12.011
Received 5 April 2017; Received in revised form 16 December 2017; Accepted 18 December 2017
Available online 22 December 2017

Image is one of the most important ways of information expression. As the shape features extracted from grayscale images, moments have several advantages such as multi-resolution and invariance abilities of translation, rotation and scaling in image descriptions [27]. Several moments have been applied successfully in chemical quantitative analysis based on the three-dimensional (3D) spectra or landscapes of high performance liquid chromatography coupled with photodiode array detection [28], excitation-emission matrix fluorescence [29], ^{1}H nuclear magnetic resonance [30], and liquid chromatography with mass spectrometry [31]. These results indicate that image moment methods can eliminate the adverse effects such as unknown interferences, overlap and drift of peaks in the quantitative analysis of multi-compounds on the basis of 3D spectra or landscapes.

Here, we introduced Tchebichef image moment (TM) method to the analysis of raw NIR spectra for the first time. The NIR 3D spectrum was constructed from traditional NIR spectrum and regarded as grayscale image, then the combination of TM and PLS (TM-PLS) was presented for quantitative analysis the target components of complex samples based on raw spectrum, the obtained models were evaluated strictly by statistical parameters. Moreover, as a comparison, classical PLS and iPLS method was also carried out on the same samples.

2. Data and methodology

2.1. Description of the data sets

The dataset1 was downloaded from http://software.eigenvector.com/Data/Corn/index.html, which consisted of the NIR spectra of 80 corn samples and the reference content values of moisture, oil, protein and starch ranging from 9.38% to 10.99%, 3.09%–3.83%, 7.65%–9.71% and 62.83%–66.47%, respectively. The spectra were measured at Cargill Inc. (Minneapolis, MN, USA) using three different NIR spectrometers (m5, mp5, mp6), and each spectrum was composed of 700 data points recorded in the wavelength range 2498–1100 nm (4003-9091 cm^{-1}) with the digitization interval 2 nm. The spectra of the four ingredients measured on m5 NIR spectrometer were employed in this study.

The dataset2 was downloaded from the website of http://www.models.life.ku.dk/Marzipan, which applied for a compositional analysis of 32 marzipan samples. Six different NIR and one IR set-up were applied in analysis of the same marzipan samples, traditional moisture and sugar analysis was performed on all samples. The spectra of the sugar measured on NIRS1 spectrometer was employed in this study.

The dataset3 was downloaded from the website of http://www.models.life.ku.dk/NIRsoil, which composed of NIR diffuse reflection spectra of 108 soil sample. The chemical value of soil organic matter was measured as loss on ignition at 550 °C. The spectra were recorded by using an NIR Systems 6500 NIR spectrophotometer (Foss Analytical, Sweden) over a wavelength range from 400 to 2500 nm (1050 variables). The NIR spectra of the soil organic matter (SOM) was studied.

The samples were divided into two groups randomly, training set (75% of the entire data set) and test set (25% of the entire data set). The concentration of target components in sample were listed in Table S1 Table S2 and Table S3 as Exp. columns (Supplement materials). The NIR spectra of the samples for dataset1, dataset2 and dataset3 were shown in Fig. 1, respectively. As can be seen, there exists the seriously overlapping and drift of peaks, which are usually troublesome in analysis. The quantitative models were established by TM-PLS method based on the training set, and their predictive ability were evaluated by the samples in the test set that were "unknown" to the established models since they were excluded from the modeling procedure [32].

2.2. Modeling method

2.2.1. Self-construction of NIR spectra

As one of the simplest ways, the 3D spectrum can be established by means of the replicating 2D spectrum, which named self-construction three-dimension (SC-3D) landscape [33]. Therefore, the NIR SC-3D spectrum was directly constructed through double the corresponding NIR spectrum (the number of SC was set to 2), and the matrix sizes can be arranged as $\boldsymbol{A}$ (data points) × 2 (the number of SC) × $\boldsymbol{B}$ (the number of samples). The information in the SC-3D spectrum and the original NIR spectra is exactly the same, and the inherent correlation among the different signals in the SC-3D spectrum would not be affected. The intent is just to extract the feature information effectively by TM method. As the schematic diagram, the NIR SC-3D spectrum and its grayscale image of one sample were shown in Fig. 2.

2.2.2. The calculation of Tchebichef moments

As the shape descriptors in grayscale image, Tchebichef moments (TMs), also named Chebyshev moments, are a set of orthogonal moment functions based on the discrete Tchebichef polynomials, and the description about TM is provided in the supplement materials, and the more detailed exploration could refer to reference [34]. Here, the brief definition of TM is given as follows:

Let make use of $f(x,y)$ representing the density distribution function of a digital grayscale image with size $N \times M$ in Cartesian coordinates. The TMs with $(n+m)$ order are defined as:

$$T_{nm} = \frac{1}{\tilde{\rho}(n,N)\tilde{\rho}(m,M)} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \tilde{t}_n(x)\tilde{t}_m(y) f(x,y) \quad n = 0,\ 1,\ 2, \ldots,\ N-1,\ m = 0,\ 1,\ 2, \ldots,\ M-1 \tag{1}$$

then Eq. (1) leads to the following inverse moment transform:

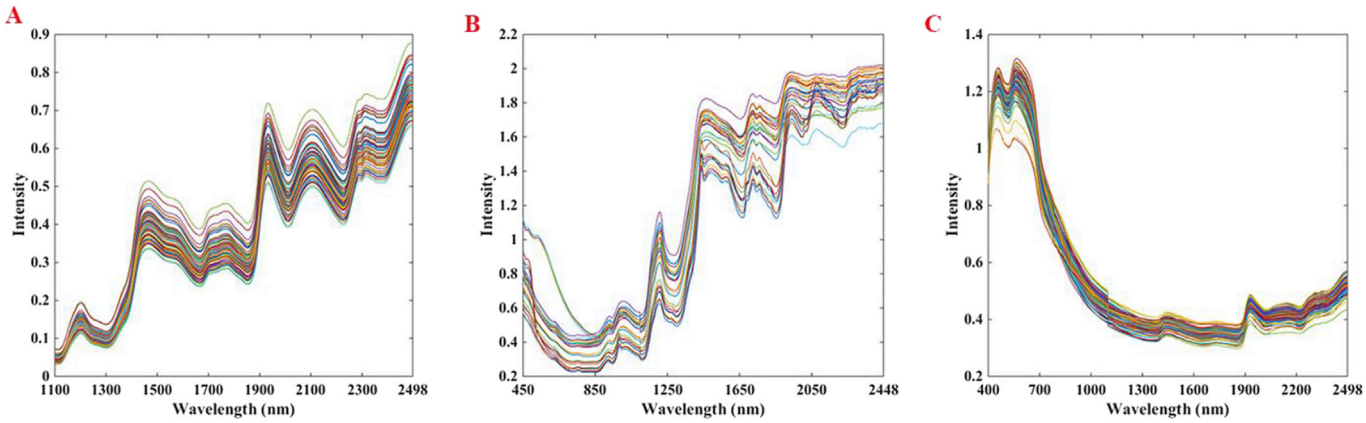

Fig. 1. The NIR spectra of the samples for dataset1(A), dataset2(B) and dataset3(C).

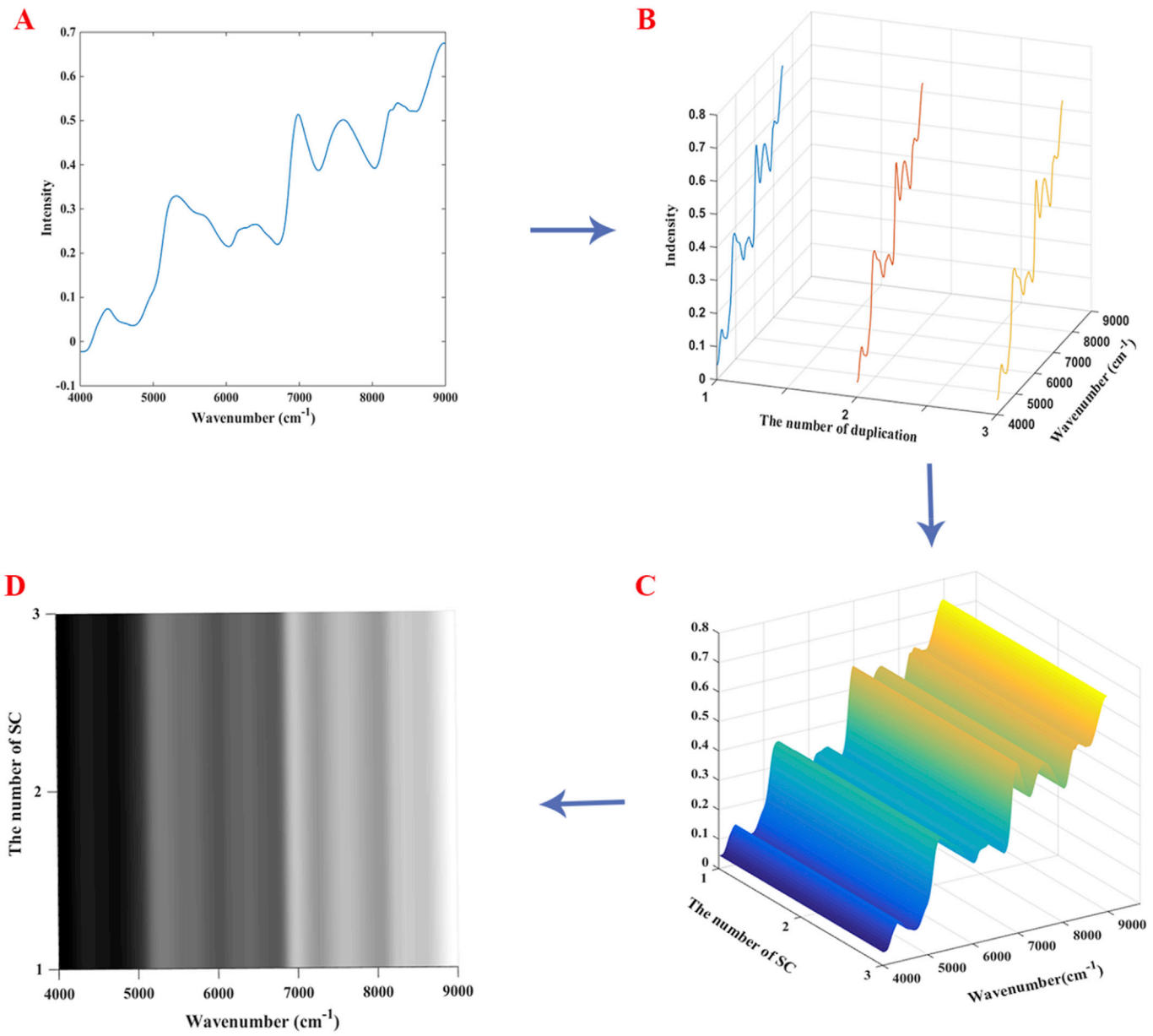

Fig. 2. The schematic diagram of the proposed method. (A) the spectrum of one sample (B) the replicating 2D spectrum (C) the NIR SC-3D spectrum (D) the grayscale image.

$$\widehat{f}(x,y) = \sum_{n=0}^{nM} \sum_{m=0}^{mM} T_{nm} \tilde{t}_n(x) \tilde{t}_m(y) \tag{2}$$

where $\widehat{f}(x,y)$ is the reconstructed image. The nN and mM are the maximum orders of n and m, respectively.

2.2.3. TM-PLS method

As an excellent multivariate data statistical tool, PLS regression method has been broadly and successfully applied in the quantitative analysis of spectral data [35]. Here, the mathematical models for the quantitative analysis of four components in corn samples were established by PLS regression, where Y and X were the concentrations of target compounds (dependent variables) and calculated TMs (independent variables), respectively.

2.2.4. Interval partial least squares algorithm (iPLS)

Nørgaard et al. [21] developed a wavelength selection process driven by a local regression method called interval partial least squares (iPLS) regression to improve the predictive power and to enhance the interpretation of PLS models. The approach used in the interval PLS algorithm splits the response matrix (X for m samples measured at p spectral wavelengths) into n disjoint intervals (X_1, X_2, …,X_n) of equal width (of p/n channels), where each sub-interval may contain different information and variance that is related to the property of interest. $RMSE_{cv}$ values and number of LVs obtained by PLS regression on each sub-interval are employed to focus on the important spectral regions and to eliminate the other regions [36]. The best regression model based on sub-intervals should require the smallest number of PLS components and produce the lowest $RMSE_{cv}$ values.

2.3. Evaluation of models

To evaluate the performance of the established models, it is necessary to assess the reliability, accuracy and the predictive ability for its applications. In detail, statistical parameters included the correlation coefficient of calibration square (R_c), adjusted correlation coefficients square (R_{adj}), correlation coefficient of leave-one-out square (R_{loo-cv}), the cor-

relation coefficient of prediction square (R_p), root mean square errors ($RMSE_c$), the root mean square error of prediction ($RMSE_p$), *F-test* and *p-value* were calculated, and *t-test* was used to test the significance of the regression coefficients in models. The optimal number of latent variables (LV_s) was selected according to Leave-one-out cross-validation, and the most appropriate number of LV_s was chosen on basis of the minimum root mean square of leave-one-out cross-validation ($RMSE_{cv}$) [37].

All calculation programs were written in M-file based on MATLAB 7.0 and carried out with a PC (CPU 4.00 GHZ, RAM 32.0 GB).

3. Results and discussion

3.1. Characteristics of Tchebichef moments

The NIR spectrum is made up of the absorbance values of the various functional groups in different ingredients, in which the characteristic absorption peaks of the same functional groups maybe different due to the tiny difference in their chemical environments, thus the phenomenon of overlapping and shift peaks is inevitable. As an important type of discrete orthogonal moments, owing to its powerful multi-resolution property, TMs with different moment orders could separate the information (including the target components and interference components) automatically in the digital grayscale image of NIR SC-3D spectrum during the computational processes. Thus, TMs not only reduces the redundancy information among the independent variants but also corral related information into different categories, which is beneficial to outstand the information of target components for the quantitative modeling. Furthermore, thanks to the inherent invariance property of TMs in the shifting, scaling and rotation of image, the shifts of peaks in NIR spectra are autocorrected during the computational processes without any other pretreatment for raw spectra, which means that calculated TMs can't be affected by the shifts of peaks. Therefore, utilizing the TMs to represent the chemical information is an attractive strategy to overcome some problems such as the interferences of irrelevant information, overlapping and shifts of peaks in NIR spectra so as to build high quality models.

3.2. The determination of maximum orders

The information in the NIR SC-3D spectrum can be represented by the TMs with different ($n + m$) orders. The TMs with lower orders are the significant components of the image, while the TMs with higher orders can provide the details [27]. Because the noise is localized in high frequencies, high order moments are more sensitive to noise than low order moments [38]. To obtained the most important information as well as reduce the unnecessary calculations, the optimum values of maximum orders (*nN* and *mM*) can be determined. Although the *N*-1 and *M*-1 may be regarded as maximum orders according to Eq. (1), it is not necessary. The optimum orders of the Tchebichef moment were determined based on the accuracy in quantification and the performance of the model. When the number of SC is more than 2, the values of obtained TMs are not changed. Therefore, the least number of SC was set to 2, and the maximum order *mM* was directly at 0 (according to Eq. S(1) in supplement materials). There were the $nN = 67$, 128, 98, 93 for the target components of dataset1, the *nN* of dataset2 and dataset3 were 25, 94, respectively.

3.3. Performance of TM-PLS models

3.3.1. Results of dataset1

By means of PLS method based on the TMs, the quantitative TM-PLS models were obtained for the four target components in corn samples, respectively. The calculated concentrations for training set and the predicted results for test set were also listed in Table S1 as Cal. TM-PLS columns, respectively. The plots of experimental vs. calculated concentrations are shown as Fig. 3, and the performance of these linear models are summarized in Table 1. The statistical parameters from obtained models including R_c(0.9980, 0.9458, 0.9829, 0.9782) and R_{adj}(0.9979, 0.9448, 0.9826, 0.9779) for moisture, oil, protein and starch in corn samples support the accuracy of established models. The value of *F-test* and *p-value* demonstrate that the established models have good linear relationship between the contents of the four components and the selected latent variants of PLS obtained from TMs. Moreover, the results of *t*-test for all regression coefficients in these models indicate the credibility of obtained models. In order to confirm whether there is an overfitting [39,40], we have also adopted the leave-one-out cross validation to assess the obtained models, and the results of R_{loo-cv} (0.9969, 0.9043, 0.9723, 0.9556) and $RMSE_{cv}$ (0.0267, 0.0714, 0.1165, 0.2450) show that the established models have good reliability. The R_p (0.9987, 0.9450, 0.9702, 0.9552) and $RMSE_p$ values (0.0139, 0.0370, 0.0699, 0.1336) indicate that the models established by the training set have satisfactory predictive ability for the four target components in test set.

3.3.2. Results of dataset2 and dataset3

To further verify the effect of the proposed method, dataset2 and dataset3 were adopted, and Table 2 lists the statistical parameters of the established models. The established model of the target components with high values of R_c (0.9988, 0.9467), R_{adj} (0.9988, 0.9460), R_p (0.9850, 0.9754) and $R_{loo\text{-}cv}$ (0.9977, 0.9334), as well as low values of $RMSE_c$ (0.6146, 2.5736), $RMSE_p$ (1.4435, 2.2562), $RMSE_{cv}$ (0.8813, 2.8695) for dataset2 and dataset3, respectively, indicating that the established model have good linear fit, high reliability and accuracy. Moreover, *F-test* values and *p-value* of the model indicate that the linear relationships between the concentrations of the target components and the selected $T_{n,m}$ that derived from the SC-3D spectrum are significant. Therefore, the statistical parameters demonstrated that the proposed method may be a good choice to quantitative analysis of components in complex samples based on their raw NIR spectra.

3.4. Comparison with PLS and iPLS method

3.4.1. Comparison with traditional PLS method

Traditional PLS method was carried out based on the raw conventional NIR spectra of the same samples, the performance of these linear models are also summarized in Table 1 for dataset1. As can be observed from the statistical parameters, the related parameters (R_c, R_{adj}, R_p, $RMSE_c$, $RMSE_p$) of the models obtained from TM-PLS method are preponderant than that of PLS regression method, especially, the values of R_p for oil, protein, starch. This phenomenon shows that TM method as an efficient pretreatment method, the established models by TM-PLS method have higher robustness, accuracy and reliability even in the presence of different degree of overlapping signals. The calculated concentrations for training set and the predicted results for test set were also listed in Table S1 as Cal. PLS columns, and the plots of experimental vs. calculated concentrations are also shown as Fig. 3, respectively. For dataset2 and dataset3, we can obtain the similar results, the calculated concentrations for training set and the predicted results for test set were also listed in Table S2 and Table S3 as Cal. PLS columns, and the plots of experimental vs. calculated concentrations were shown as Fig. S2 and Fig. S3.

3.4.2. Comparison with classical iPLS method

For comparison purpose, advanced iPLS method was carried out to choose the best spectrum interval or even several of the better spectrum intervals. Fig. 4 shows the best spectrum interval for dataset1, the statistical parameters of these models are summarized in Table 1 as iPLS columns, respectively. The best spectrum interval for dataset2 and dataset3 were shown in Supplement materials as Fig. S1 and the statistical parameters of these models are summarized in Table 2 as iPLS columns. As for the results by TM method, although the some difference

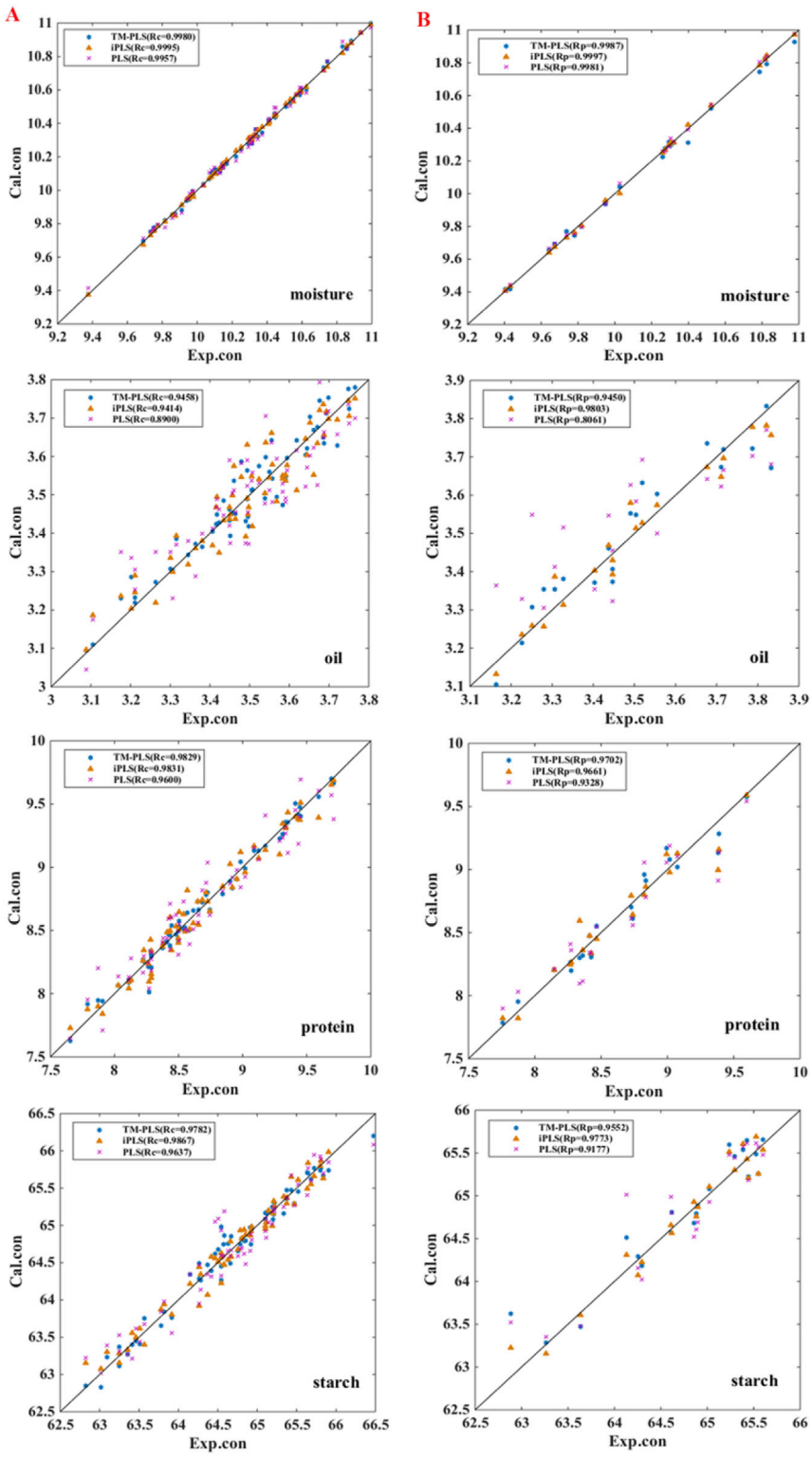

Fig. 3. The plots of experimental vs. calculated concentration for dataset1. (A) training set and (B) test set.

can be seen, the satisfactory results can be obtained by using established models.

This phenomenon shows that TM method as an efficient pretreatment method, the established models by TM-PLS method have higher robustness, accuracy and reliability even in the presence of different degree of overlapping signals. The calculated concentrations for training set and the predicted results for test set by iPLS method were also listed in Table S1 Table S2 and Table S3 as Cal. iPLS columns, and the plots of experimental vs. calculated concentrations are also shown as Fig. 3, Fig. S2 and Fig. S3 for dataset1, dataset2 and dataset3, respectively. Owing to their the powerful multi-resolution capability and inherent invariance property, TMs could improve the accuracy and reliability of PLS method and obtain the more satisfactory results. As a simple and effective pretreatment method for raw NIR spectra, TMs were employed to extract and represent the important information of chemical images of NIR spectra.

Table 1
The performance of models by TM-PLS iPLS and PLS method for the four components of dataset1.

Model		*LVs*	Training set					Cross-validation		Test set	
			R_c	R_{adj}	*F-test*	*p-value*	$RMSE_c$	$R_{loo\text{-}cv}$	$RMSE_{cv}$	R_p	$RMSE_p$
TM-PLS	Moisture	5	0.9980	0.9979	1.43×10^{4}	3.91×10^{-71}	0.0217	0.9969	0.0267	0.9987	0.0139
	Oil	6	0.9458	0.9448	4.91×10^{2}	5.18×10^{-30}	0.0542	0.9043	0.0714	0.9450	0.0370
	Protein	6	0.9829	0.9826	1.65×10^{3}	2.63×10^{-44}	0.0918	0.9723	0.1165	0.9702	0.0699
	Starch	8	0.9782	0.9779	1.29×10^{3}	2.60×10^{-41}	0.1723	0.9556	0.2450	0.9552	0.1336
iPLS	Moisture	6	0.9995	0.9995	6.80×10^{4}	1.01×10^{-90}	0.0100	0.9993	0.0122	0.9997	0.0076
	Oil	6	0.9414	0.9404	4.52×10^{2}	4.45×10^{-29}	0.0562	0.9068	0.0704	0.9803	0.0234
	Protein	6	0.9831	0.9828	1.67×10^{3}	1.72×10^{-44}	0.0911	0.9755	0.1097	0.9661	0.0739
	Starch	7	0.9867	0.9864	2.13×10^{3}	1.85×10^{-47}	0.1350	0.9769	0.1781	0.9773	0.0940
PLS	Moisture	6	0.9957	0.9957	6.79×10^{3}	8.44×10^{-62}	0.0314	0.9935	0.0386	0.9981	0.0164
	Oil	5	0.8900	0.8880	2.20×10^{2}	1.93×10^{-21}	0.0761	0.8373	0.0914	0.8061	0.0728
	Protein	6	0.9600	0.9594	6.82×10^{2}	9.00×10^{-34}	0.1394	0.9358	0.1762	0.9328	0.1029
	Starch	8	0.9637	0.9630	7.56×10^{2}	5.80×10^{-35}	0.2217	0.9369	0.2912	0.9177	0.1801

TM-PLS: Tchebichef moment method combined with partial least square.
LVs: Number of Latent Variables.

Table 2
The performance of models by TM-PLS iPLS and PLS method for the dataset2 and dataset3.

Model		Dataset2			Dataset3		
		TM-PLS	iPLS	PLS	TM-PLS	iPLS	PLS
Number of Latent Variables		6	5	6	4	4	4
Training set	R_c	0.9988	0.9981	0.9981	0.9467	0.8884	0.9421
	R_{adj}	0.9988	0.9980	0.9980	0.9460	0.8595	0.9414
	F-test	9.12×10^{3}	5.81×10^{3}	5.83×10^{3}	6.82×10^{2}	2.95×10^{2}	6.24×10^{2}
	p-value	1.31×10^{-30}	3.68×10^{-28}	3.55×10^{-28}	1.24×10^{-40}	1.95×10^{-28}	2.99×10^{-39}
	$RMSE_c$	0.6146	0.7940	0.7927	2.5736	3.6692	2.6792
Cross-validation	$R_{loo\text{-}cv}$	0.9977	0.9963	0.9956	0.9334	0.8595	0.9297
	$RMSE_{cv}$	0.8813	1.1147	1.2079	2.8695	4.0914	2.9447
Test set	R_p	0.9850	0.9642	0.9811	0.9754	0.9599	0.9768
	$RMSE_p$	1.4435	2.1003	1.5721	2.2562	2.8990	2.2399

TM-PLS: Tchebichef moment method combined with partial least square.

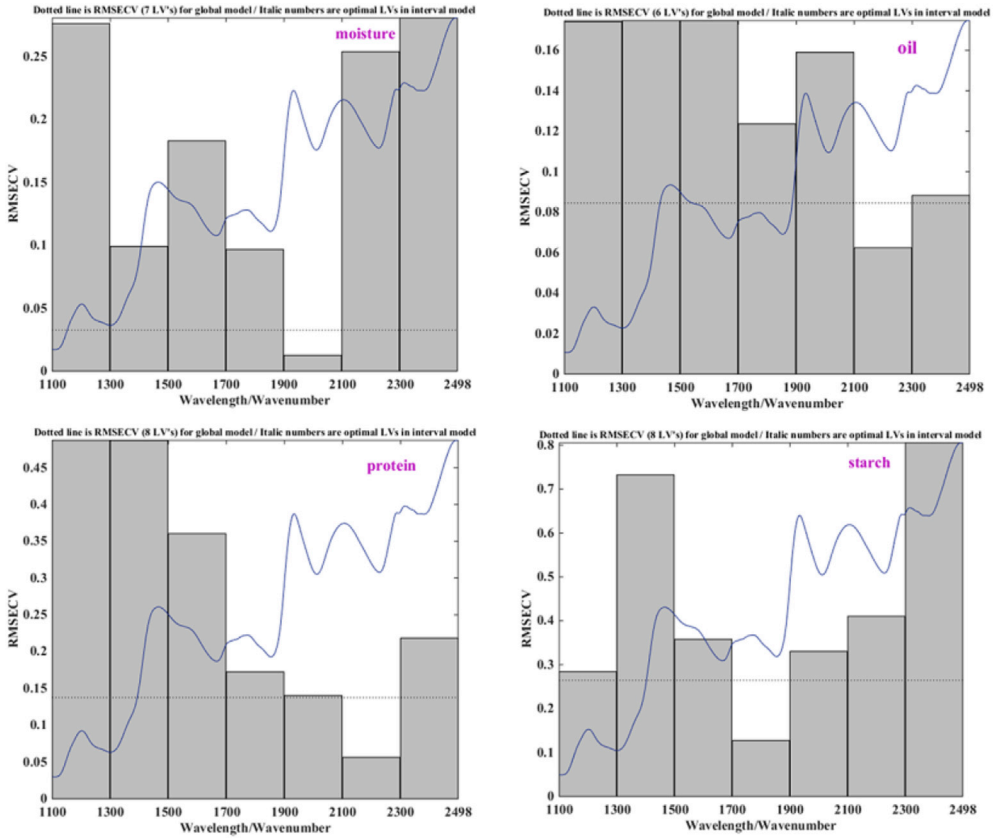

Fig. 4. The best spectrum interval for four target components of dataset1 by iPLS method.

4. Conclusion

We introduced Tchebichef image moments to PLS modeling for quantitative analysis of components in complex samples based on their raw NIR spectra. TM-PLS approach can effectively overcome the challenges that NIR spectra typically consist of broad, weak, overlapping peaks and interference of the background in the quantitative analysis as an efficient pretreatment method based on whole range spectra. Compared with traditional PLS and classical iPLS method, the TM-PLS method can accurately and robustly quantify the contents in complex samples on all range spectrum, which indicates that TMs is an effective pretreatment method owing to their inherent invariance and powerful multi-resolution properties. Therefore, the proposed approach could be an attractive strategy to the quantitative analysis of components in complex samples based on conventional NIR spectra, which will extend the application of NIR spectroscopy.

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgement

The author gratefully acknowledge the financial support from the National Science Foundation Committee of P. R. China (grant no. 21275067)

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.chemolab.2017.12.011.

References

[1] J. Moros, I. Llorca, M.L. Cervera, A. Pastor, S. Garrigues, M. de la Guardia, Chemometric determination of arsenic and lead in untreated powdered red paprika by diffuse reflectance near-infrared spectroscopy, Anal. Chim. Acta 613 (2008) 196–206.
[2] J.C. Tewari, V. Dixit, B.K. Cho, K.A. Malik, Determination of origin and sugars of citrus fruits using genetic algorithm, correspondence analysis and partial least square combined with fiber optic NIR spectroscopy, Spectroc. Acta Pt. A-Molec. Biomolec. Spectr 71 (2008) 1119–1127.
[3] S.K. Tomlinson, O.R. Ghita, R.M. Hooper, K.E. Evans, The use of near-infrared spectroscopy for the cure monitoring of an ethyl cyanoacrylate adhesive, Vib. Spectrosc. 40 (2006) 133–141.
[4] M. Blanco, J. Cruz, M. Bautista, Development of a univariate calibration model for pharmaceutical analysis based on NIR spectra, Anal. Bioanal. Chem. 392 (2008) 1367–1372.
[5] E.V. Thomas, D.M. Haaland, Comparison of multivariate calibration methods for quantitative spectral-analysis, Anal. Chem. 62 (1990) 1091–1099.
[6] R. Dunia, S.J. Qin, T.F. Edgar, T.J. McAvoy, Identification of faulty sensors using principal component analysis, AIChE J. 42 (1996) 2797–2812.
[7] P. Geladi, B.R. Kowalski, Partial least-squares regression - a tutorial, Anal. Chim. Acta 185 (1986) 1–17.
[8] D. Chen, B. Hu, X.G. Shao, Q.D. Su, A new hybrid strategy for constructing a robust calibration model for near-infrared spectral analysis, Anal. Bioanal. Chem. 381 (2005) 795–805.
[9] C. Borggaard, H.H. Thodberg, Optimal minimal neural interpretation of spectra, Anal. Chem. 64 (1992) 545–551.
[10] B.J. Wythoff, Backpropagation neural networks - a tutorial, Chemom. Intell. Lab. Syst 18 (1993) 115–155.
[11] A.I. Belousov, S.A. Verzakov, J. von Frese, Applicational aspects of support vector machines, J. Chemometr. 16 (2002) 482–489.
[12] Y.K. Li, X.G. Shao, W.S. Cai, A consensus least squares support vector regression (LS-SVR) for analysis of near-infrared spectra of plant samples, Talanta 72 (2007) 217–222.
[13] P. Geladi, D. Macdougall, H. Martens, Linearization and scatter-correction for near-infrared reflectance spectra of meat, Appl. Spectrosc. 39 (1985) 491–500.
[14] C. Pizarro, I. Esteban-Diez, A.J. Nistal, J.M. Gonzalez-Saiz, Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy, Anal. Chim. Acta 509 (2004) 217–227.
[15] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and deternding of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777.
[16] Q. Guo, W. Wu, D.L. Massart, The robust normal variate transform for pattern recognition with near-infrared data, Anal. Chim. Acta 382 (1999) 87–103.
[17] S. Wold, H. Antti, F. Lindgren, J. Ohman, Orthogonal signal correction of near-infrared spectra, Chemom. Intell. Lab. Syst 44 (1998) 175–185.
[18] N.A. Woody, R.N. Feudale, A.J. Myles, S.D. Brown, Transfer of multivariate calibrations between four near-infrared spectrometers using orthogonal signal correction, Anal. Chem. 76 (2004) 2595–2600.
[19] B.K. Alsberg, A.M. Woodward, D.B. Kell, An introduction to wavelet transforms for chemometricians: a time-frequency approach, Chemom. Intell. Lab. Syst 37 (1997) 215–239.
[20] D. Chen, W.S. Cai, X.G. Shao, An adaptive strategy for selecting representative calibration samples in the continuous wavelet domain for near-infrared spectral analysis, Anal. Bioanal. Chem. 387 (2007) 1041–1048.
[21] 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 (2000) 413–419.
[22] R. Leardi, L. Norgaard, Sequential application of backward interval partial least squares and genetic of relevant spectral regions, J. Chemometr. 18 (2004) 486–497.
[23] L. Xu, I. Schechter, Wavelength selection for simultaneous spectroscopic analysis. Experimental and theoretical study, Anal. Chem. 68 (1996) 2392–2400.
[24] A. Hoskuldsson, Variable and subset selection in PLS regression, Chemometr. Intell. Lab. Syst 55 (2001) 23–38.
[25] S. Karimi, B. Hemmateenejad, Identification of discriminatory variables in proteomics data analysis by clustering of variables, Anal. Chim. Acta 767 (2013) 35–43.
[26] H.W. Tan, S.D. Brown, Multivariate calibration of spectral data using dual-domain regression analysis, Anal. Chim. Acta 490 (2003) 291–301.
[27] C. Deng, X.B. Gao, X.L. Li, D.C. Tao, A local Tchebichef moments-based robust image watermarking, Signal Process. 89 (2009) 1531–1539.
[28] B.Q. Li, J. Chen, J.J. Li, X. Wang, H.L. Zhai, The application of a Tchebichef moment method to the quantitative analysis of multiple compounds based on three-dimensional HPLC fingerprint spectra, Analyst 140 (2015) 630–636.
[29] J. Chen, B.Q. Li, M.L. Xu, X. Wang, Y.H. Jing, H.L. Zhai, Krawtchouk image moment method for the simultaneous determination of three drugs in human plasma based on fluorescence three-dimensional spectra, Talanta 161 (2016) 99–104.
[30] B.Q. Li, J. Chen, M.L. Xu, X. Wang, H.L. Zhai, The determination of multi-components utilizing 1H NMR three-dimensional spectra combined Tchebichef moments, Chemom. Intell. Lab. Syst 156 (2016) 128–136.
[31] M.L. Xu, B.Q. Li, X. Wang, J. Chen, H.L. Zhai, Quantitative analysis of multiple components based on liquid chromatography with mass spectrometry in full scan mode, J. Sep. Sci. 39 (2016) 3054–3061.
[32] T.M. Martin, P. Harten, D.M. Young, E.N. Muratov, A. Golbraikh, H. Zhu, A. Tropsha, Does rational selection of training and test sets improve the outcome of QSAR modeling? J. Chem. Inf. Model. 52 (2012) 2570–2578.
[33] B.Q. Li, M.L. Xu, X. Wang, H.L. Zhai, J. Chen, J.J. Liu, An approach to the simultaneous quantitative analysis of metabolites in table wines by 1H NMR self-constructed three-dimensional spectra, Food Chem. 216 (2017) 52–59.
[34] R. Mukundan, S.H. Ong, P.A. Lee, Image analysis by Tchebichef moments, IEEE Trans. Image Process. 10 (2001) 1357–1364.
[35] B. Hemmateenejad, M. Akhond, F. Samari, A comparative study between PCR and PLS in simultaneous spectrophotometric determination of diphenylamine, aniline, and phenol: effect of wavelength selection, Spectroc. Acta Pt. A-Molec. Biomolec. Spectr 67 (2007) 958–965.
[36] W.D. Ni, S.D. Brown, R.L. Man, Stacked partial least squares regression analysis for spectral calibration and prediction, J. Chemometr. 23 (2009) 505–517.
[37] Q.S. Chen, J.W. Zhao, H.D. Zhang, X.Y. Wang, Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration, Anal. Chim. Acta 572 (2006) 77–84.
[38] X.B. Dai, H.Z. Shu, L.M. Luo, G.N. Han, J.L. Coatrieux, Reconstruction of tomographic images from limited range projections using discrete Radon transform and Tchebichef moments, Pattern Recogn. 43 (2010) 1152–1164.
[39] M.W. Browne, Cross-validation methods, J. Math. Psychol. 44 (2000) 108–132.
[40] J. Shao, Linear-model selection by cross-validation, J. Am. Stat. Assoc. 88 (1993) 486–494.