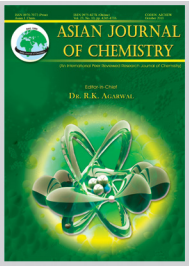
Asian Journal of Chemistry; Vol. 24, No. 4 (2012), 1449-1452

ASIAN JOURNAL OF CHEMISTRY

www.asianjournalofchemistry.co.in



Application of Wavelet Transform-Principal Component Regression Method for Simultaneous Determination of Colourants in Ternary Mixtures Using Spectrophotometry

SHENGZHU SI[1,*], LIPING WANG[1], WA SI[2] and XIONGZI DONG[1]

[1]Department of Chemistry and Chemical Engineering, Hefei Normal University, Hefei 230061, Anhui Province, P.R. China
[2]Graduate School of Information, Production and Systems, Waseda University, Fukuoka 44081536, Japan

*Corresponding author: E-mail: sishengzhu@163.com

(*Received*: 5 January 2011; *Accepted*: 17 November 2011) AJC-10688

A new multivariate method combined wavelet transform and principal component regression techniques for the simultaneous determination of safranine, phloxine B and eosinY using total absorbance measurements is proposed. In this method, the visible absorption spectra of ternary mixtures was pretreated by wavelet transform at first,and then the principal component regression procedure was performed on the wavelet coefficients of the original spectra. The absorption spectra and related wavelet coefficients (both approximation and detail) at different scales were used to perform the optimization of the calibration matrices by the principal component regression method for a comparative study, it showed that the model based on wavelet approach coefficients at lower scale is better than that based on the full visible absorption spectra. A principal component regression multivariate calibration model on the first approach coefficients of the original visible absorption spectra extracted by Bior1.1 wavelet transform, using cross-validation method leaving one out calibration sample at a time to select the optimum number of components, produced a satisfactory result with good prediction accuracies.

Key Words: Wavelet transform, Principal component regression, Colourants, Spectrophotometry, Multivariate calibration.

INTRODUCTION

The application of multivariate calibration methods in chemometrics, particularly principal component regression (PCR) and partial least squares (PLS), to multicomponent spectrophotometric data has been increasing in recent years[1-5]. Each method needs a calibration step where the relationship between the spectra and the component concentration is deduced from a set of reference samples, followed by a prediction step in which the results of the calibration are used to determine the component concentrations from the sample spectrum. Among the multivariate calibration methodologies, there are two categories including direct calibration methods and indirect calibration methods. Principal component regression and partial least squares are factor analysis-based indirect calibration methods that have many of the full-spectrum advantages of the direct calibration methods without suffering the disadvantages of this more classical statistical tool[6-8]. Principal component regression was chosen because experience shows that, if applied correctly, it generally performs as well as the partial least squares methods but the mathematical background is easier to understand.

Wavelet transform (WT) is a mathematical method developed on the basis of Fourier transform (FT)[9,10]. Compared with the Fourier transform, wavelet transform has good local features in both time domain and frequency domain and can effectively extract more information from the signals of interest. Wavelet transform is also called mathematical microscope because it can solve many problems which are impossible for Fourier analysis by multi-scale analyzing the signal through operations such as scaling and translation. Wavelet analysis already has many application aspects in chemistry[11-15]. Because of its ability of multi-resolution analysis, wavelet analysis has the capacity of decomposing the signal at different scales, thus can get the discrete approximation and discrete details of the original signal in different frequencies of wavelet domain. By using the discrete approximation absorbance signal instead of the original absorbance data for multivariate calibration model, the data can be compressed and the impact of noise on the calibration results can be reduced.

In this paper, we combine discrete wavelet transform with principal component regression method to analyze the visible spectrum data of three colourants which are safranine, phloxine B and eosinY. After the data processing, we achieve the simultaneous determination of three components and obtain satisfactory prediction results. Further research shows that models using wavelet coefficients of specific scale is superior to the traditional full spectrum model.

Theoretical background: Wavelets are a new family of basis functions, which can be used to describe instrumental

signals. Projection of the signal onto wavelet basis functions is called wavelet transform. As any transform, the wavelet transform aims to transform the signal from the original to another domain in which some operations on the signal (*i.e.* denoising, compression) can be caried out in an easier way. Applying wavelet transform on a signal decomposes it into different frequency sub-bands. We now briefly review wavelet-based multi-resolution decomposition. More details can be found in Mallat's paper[16]. To have multi-resolution representation of signals we can use discrete wavelet transform (DWT). According to the discrete wavelet transform and Mallat algorithm[16], after j-scale (level) decomposition, the original signal C_0 can be expressed as:

$$C_0 = C_1 + D_1 = C_2 + D_2 + D_1 = C_j + \sum_{k=1}^{j} D_k \quad (1)$$

$$\text{where } C_j = HC_{j-1},\ D_j = GC_{j-1}, \quad (2)$$

Operators H and G represent low pass filter and high pass filter respectively, while C_j and D_j represent the discrete approximation (low frequency factor) and the discrete details (high frequency coefficients) of C_0 (original signal) under 2_j resolutions, the data points of both C_j and D_j of C_0 under the j scale decomposition is $1/2^j$ of the original data. Because the wavelet transform is a linear transformation, and C_j and D_j are the linear projection of C_0 in wavelet space, we can use less wavelet coefficients (C_j or D_j) instead of raw data C_0 for analyzing.

Principal component regression (PCR) is a two-step process. The first step consists of the principal component analysis (PCA) of the original data, *i.e.* absorption spectra, to obtain a reduced number of variables, the factor scores. Then, multiple linear regression is used to relate these scores to the concentration values. Suppose a calibration matrix of absorbance X consists of 'm' calibration samples at 'n' wavelengths and Y of the concentration matrix of l components in 'm' mixture samples. According to the principal component analysis (PCA), we decompose the X matrix:

$$X_{m\times n} = T_{m\times h}P_{m\times n} + E_{m\times n} \quad (3)$$

where, h is the number of principal components and E is the residual matrix. The aim of principal component analysis is to represent X by a set of new orthogonal variables called principal components (PCs). The principal components are linear combinations of explanatory variables that maximize the data variance. The data matrix, X, is decomposed to a score matrix, T and loadings matrix, P. The elements of the loadings matrix give information about the contribution of the original variables to each principal component. Since the columns of T matrix is orthogonal, h is the optimum number of principal components to reproduce the original data matrix X by T and P within experimental error. After performing principal component analysis on X, the second step in principal component regression consists of the linear regression of the scores and the Y of concentration. The linear model between Y and T is of the form:

$$Y_{m\times l} = T_{m\times h}B_{h\times l} \quad (4)$$

with the least square solution:

$$B = (T^tT)^{-1}T^tY = \Lambda^{-1}T^tY \quad (5)$$

where, the superscript t denotes the transpose of the matrix. Since the columns of T matrix is orthogonal, $T^tT = \Lambda$. Λ is a matrix constructed by the h larger eigenvalues of X^tX, thus Λ^{-1} is the form-only inverse. Actually, Λ^{-1} is a diagonal matrix composed of inverses of eigenvalues, which can avoid errors caused by other methods [for example, the error caused by multiple linear regression (MLR) method in the inverse calculate process]. Besides, the orthogonality of T and P eliminates the possible collinearity when calculating the coefficient matrix by using absorbance matrix X and concentration matrix Y directly and thus improves the accuracy of prediction.

As for the unknown samples, Y_u can be obtained by the following formula:

$$Y_u = T_uB = X_uP^tB \quad (6)$$

We replace the original absorbance matrix by wavelet coefficients for principal component regression modeling and predicting in this paper. For the evaluation of the predictive ability of a multivariate calibration model, the root mean square difference (RMSD) and the relative error of prediction (REP) can be used:

$$RMSD = [\frac{1}{N}\sum_{i=1}^{N}(\hat{X}_i - X_i)^2]^{1/2},\ REP(\%) = \frac{100}{\bar{x}}[\frac{1}{N}\sum_{i=1}^{N}(\hat{X}_i - X_i)^2]^{1/2} \quad (7)$$

where, x_i is the true concentration of the analyte in the sample i, $\hat{x}_i$ represents the estimated concentration of the analyte in the sample i and N is the total number of samples used in the prediction set, were calculated. The values of the root mean square difference (RMSD) is an indication of the average error in the analysis for each component. The relative error of prediction (REP), which is the square root of the mean square of the error in prediction expressed as a percentage of the mean of the true concentrations, can also be used to evaluate the predictive ability of each method and for each component.

EXPERIMENTAL

A Shimadzu UV-260 ultraviolet-visible spectrophotometer was used for spectral acquisition and storage of the spectrophotometric data. An in-home program set in MATLAB Version 7.0 for principal component regression and wavelet transform processing was implemented on a personal computer.

Safranine, phloxineB and eosinY stock solutions of 500 μg/mL were prepared with redistilled water.

A full set of calibration and test solutions for three-component system were formed by orthogonal design. Calibration and test set solutions were prepared serial dilution of the stock solutions. Spectrophotometric measurements were carried out with a Shimadzu UV-260 ultraviolet-visible spectrophotometer, employing a 10 mm quartz cell. Absorbances of the mixtures of three colourants between 380-600 nm wavelengths by 0.1 nm intervals against a blank of solvent were scanned and recorded for subsequent treatment.

RESULTS AND DISCUSSION

Absorption spectrum: Fig. 1 shows the absorption spectrum of safranine, phloxine B and eosinY which overlapped seriously. Using classical spectrophotometric method for quantitative analysis may cause serious interference, while combining chemometrics methods with spectrophotometric analysis enable us to quantitate each component simultaneously and accurately without separation.

TABLE-1
RELATIVE ERROR OF PREDICTION (RMSD) AND RELATIVE ERROR OF PREDICTION (REP) VALUES OF PREDICTING THE 3 COMPONENTS IN 12 SAMPLES BY WETLET-PRINCIPAL COMPONENT REGRESSION METHOD AT DIFFERENT SCALES

	Safranine		Phloxine B		Eosin Y		Scale j
	RMSD	REP (%)	RMSD	REP (%)	RMSD	REP (%)	
Discrete Approximation	0.0907	3.887	0.0894	4.472	0.0894	3.250	0
	0.0902	3.866	0.0892	4.459	0.0890	3.237	1
	0.0903	3.869	0.0893	4.464	0.0891	3.238	2
	0.0905	3.881	0.0894	4.472	0.0892	3.242	3
	0.0903	3.872	0.0894	4.4168	0.0890	3.237	4
	0.0902	3.865	0.0895	4.473	0.0889	3.234	5
	0.0959	4.109	0.0933	4.666	0.0927	3.372	6
Discrete detail	/	/	/	/	/	/	0
	0.9764	41.84	0.3677	18.39	0.4192	15.24	1
	0.6078	26.05	0.2331	11.65	0.1971	7.165	2
	0.2613	11.20	0.0977	4.886	0.1138	4.136	3
	0.1969	8.437	0.0934	4.668	0.1086	3.948	4
	0.1378	5.907	0.0787	3.934	0.0916	3.329	5
	0.1162	4.978	0.0721	3.604	0.0762	2.769	6

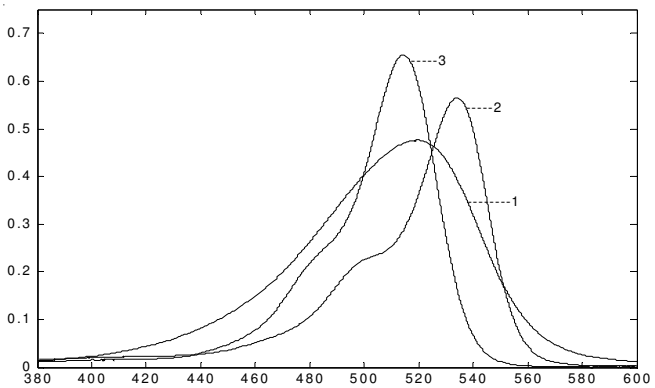


Fig. 1. Absorption spectrum of three colourants; 1 safranine, 2 phloxine B 3 eosinY

Selection of optimum number of factors: It is very important to choose the proper number of principal components when using principal component regression method to build calibration model. In this paper, we used cross-validation leave-one-out method to evaluate the optimum number of principal components for calibration. As for the absorption spectrum of the calibration set system of 12 samples (wavelength range 380-600 nm, with 0.1 nm intervals), we use the wavelet coefficients obtained by bior1.1 wavelet decomposition for modeling. The optimum number of factors (latent variables) to be included in the calibration model was determined by computing the prediction error sum of squares (PRESS) for cross validated models. The prediction error sum of squares values provide a measure of how well the training set is predicting the concentration for each number of factors. A common choice for the optimum number of factors would be that number which yielded the minimum prediction error sum of squares. To avoid overfitting for unknown samples that were not included in the model, the significance of prediction error sum of squares values greater than the minimum can be determined by F statistical test. The proper number of principal components are 4, 5 for the three colourants.

Predict results of wavelet coefficients at different scales: In order to investigate the influence on prediction results by wavelet coefficients of different scales, we use 12 samples for calibration and 12 samples for prediction and process principal component regression regression by the data matrix of C_0 (where C_0 is direct absorbance signal actually), C_1-C_6 and D_1-D_6. Statistical parameters of root mean square difference and relative error of prediction values were summarized in Table-1.

From Table 1, it is observed that since the prediction errors of C_1-C_5 are comparatively small than that of C_0, this means that the prediction results of C_1-C_5 are better than C_0. The fact that the errors of C_6 are a little larger may come of that some information details of the original signal are lost and the influence of edge effects is increasing as the scale increasing. Among the prediction errors based on D_1-D_6, the prediction errors of D_3-D_6 are smaller while the errors of D_1 and D_2 is rather large which tells us that in wavelet analysis, discrete details of small-scale have random noise (high frequency), while large-scale of high frequency coefficients represent the frequency characteristics of the signal. As for the simultaneous visible-spectrometric determination of three colourants in this paper, it is better to use the low frequency wavelet coefficients at proper small scale of the original absorbance data to build the model for principal component regression regression and the prediction results is better than using full spectrum of direct absorption spectra.

Application to synthesis mixtures: We use the low-frequency wavelet coefficient C_1 of original absorption spectra of 12 synthesis samples as calibration set, with 5 factors model, to predict the ternary mixtures of safranine, phloxine B and eosin Y by principal component regression procedure. The results are listed in Table-2.

As we can see that the satisfactory results with recoveries ranging from 94.73 % to 114.09 % of safranine, 89.00 % to 108.60 % of phloxine B and 92.48 % to105.33 % of eosin Y. The relative errors of prediction (REP) for tree components in 12 synthesis mixtures are 3.87, 4.46 and 3.24 % respectively, were obtained by the proposed method. Furthermore, using wavelet coefficients for principal component regression model can not only improve the prediction accuracy by eliminating some noises of the original signal, but can also improve the computing efficiency by large amount of data compression. It

TABLE-2
PREDICTION RESULTS OF MIXTURE SAMPLES BY WAVELET- PRINCIPAL COMPONENT REGRESSION METHOD (PRINCIPAL COMPONENT NUMBER IS 5)

Mixture	Safranine		Phloxine B		Eosin Y	
	Added (μg/mL)	Recovery	Added (μg/mL)	Recovery	Added (μg/mL)	Recovery
1	0.50	114.09	3.50	100.70	1.50	96.92
2	1.00	103.76	2.50	99.34	3.50	101.06
3	1.00	94.73	1.00	103.80	2.00	102.25
4	1.50	94.40	0.50	93.22	4.00	100.99
5	2.00	102.50	3.50	95.24	1.50	92.48
6	2.00	107.79	2.50	98.25	3.50	96.91
7	2.50	97.99	1.50	108.60	2.00	105.33
8	3.00	98.80	0.50	106.99	4.00	100.18
9	3.00	99.01	3.50	101.16	1.50	102.21
10	3.50	103.14	2.50	92.39	3.50	96.18
11	4.00	95.04	1.50	103.10	2.00	100.81
12	4.00	98.23	0.50	89.68	4.00	95.45
RMSD	0.0902		0.0892		0.0890	
REP	3.866		4.459		3.237	

has been shown that combining wavelet transform-principal component regression method with visible spectrophotometric analysis provides us a new simple and reliable way for simultaneous analysis of mixture colourant mixtures without separation.

Conclusion

Multivariate calibration method wavelet transform-principal component regression allows the simultaneous determination of safranine, phloxine B and eosin Y in ternary mixtures based on the wavelet coefficients of absorption spectra at different scales. The present study shows that the wavelet transform can be a good method conceiving ability of signal denoized and data compressed for calibration modelling. A satisfactory result with good prediction accuracies in synthetic samples demonstrates the utility of this procedure for the simultaneous determination of safranine, phloxine B and eosin Y, without tedious pretreatment.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support from the Anhui Province Funded Project (2008jyxm467).

REFERENCES

1. M.G. Trevisan and R.J. Poppi, *Talanta*, **75**, 1021 (2008).
2. X. Dong, L. Wang, Y. Tian and S. Si, *Asian J. Chem.*, **22**, 5759 (2010).
3. G. Absanlan and M. Nekoeinia, *Anal. Chim. Acta*, **531**, 293 (2005).
4. B. Lavine and J. Workman, *Anal. Chem.*, **82**, 4699 (2010).
5. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch and M. Redon, *Appl. Spectrosc.*, **48**, 37 (1994).
6. H. Martens and T. Naes, Multivariate Calibration, John Wiley & Sons, New York, (1989).
7. D.M. Haaland and E.V. Thomas, *Anal. Chem.*, **60**, 1193 (1988).
8. M.M. Galera, J.L.M. Vidal, A.G. Frenich and P. Parrilla, *Analyst*, **119**, 1189 (1994).
9. D.B. Percival and A.T. Walden, Wavelet Methods for Time Series Analysis, Cambridge University Press, 2000 (reprinted in China in2004).
10. G. Strang, *SIAM Rev.*, **31**, 614 (1989).
11. M. Bos and J.A.M. Vrielink, *Chemom. Intell. Lab. Syst.*, **23**, 115 (1994).
12. D. Jouan-Rimbaud, B. Walczak, R.J. Poppi, O.E. de Noord and D.L. Massart, *Anal. Chem.*, **69**, 4317 (1997).
13. F.T. Chau, T.M. Shih, J.B. Gao and C.K. Chan, *Appl. Spectrosc.*, **50**, 339 (1996).
14. J.B. Gao, F.T. Chau and T.M. Shih, *SEA Bull. Math,* **20**, 85 (1996).
15. F.T. Chau, J.B. Gao, T.M. Shih ands J. Wang, *Appl. Spectrosc.*, **51**, 649 (1997).
16. S.G. Mallat, *IEEE Pattern Anal. Machine Intell.*, **11**, 674 (1989).