A Modified Moving-Window Partial Least-Squares Method by Coupling with Sampling Error Profile Analysis for Variable Selection in Near-Infrared Spectral Analysis

Wuye Yang,* Wenming Wang,* Ruoqiu Zhang,* Feiyu Zhang,* Yinran Xiong,* Ting Wu,* Wanchao Chen,** and Yiping Du*†

*Shanghai Key Laboratory of Functional Materials Chemistry, School of Chemistry & Molecular Engineering, East China University of Science and Technology, Shanghai 200237, P. R. China
**Institute of Edible Fungi, Shanghai Academy of Agriculture Sciences, National Engineering Research Center of Edible Fungi, Key Laboratory of Edible Fungi Resources and Utilization (South), Ministry of Agriculture, Shanghai 201403, P. R. China

In this study, a new variable selection method, named moving-window partial least-squares coupled with sampling error profile analysis (SEPA-MWPLS), is developed. With a moving window, moving-window partial least-squares (MWPLS) is used to find window intervals which show low residual sums of squares (RSS) of a calibration set. Sampling error profile analysis (SEPA) is a useful method based on Monte-Carlo Sampling and profile analysis for cross validation (CV). By combining MWPLS with SEPA, we can obtain more stable and reliable results. Besides, we simplify the plot of the RSS line so that it is easier to determine the informative intervals. In addition, a backward elimination strategy is used to optimize the combination of subintervals. The performance of SEPA-MWPLS was tested with two near-infrared (NIR) spectra datasets and was compared with PLS, MWPLS and Monte Carlo uninformative variable elimination (MC-UVE). The results show that SEPA-MWPLS can improve model performances significantly compared with MWPLS in the number of variables, root-mean-squared errors of CV, calibration and prediction (RMSECVs, RMSECs and RMSEPs). Meanwhile it also exhibits better performances than MC-UVE.

Keywords Moving-window partial least-squares, sampling error profile analysis, variable selection, near infrared spectroscopy

Received July 30, 2019; Accepted September 30, 2019; Advance Publication Released Online by J-STAGE October 11, 2019

Introduction

With the development of high-throughput spectrum instrumental technologies in the past several decades, one can easily obtain spectral data with hundreds or thousands of variables (wavelengths). For example, molecular spectroscopic techniques, like near infrared (NIR) and Raman spectroscopy technology, have been proved to be efficient for fast determination in the field of chemicals, pharmaceuticals, soil, food and agriculture analysis. These high-dimension data have been gaining extensive application in multivariate calibration.

The high-dimension data also bring some problems. The most critical one concerning these problems is that the number of variables is too large for the number of samples to build a good model. The model directly built in such a situation has a high risk of over-fitting. In practice, people usually use partial least-squares (PLS) or principle components regression (PCR) to overcome this problem. However, the uninformative and redundant variables in the whole wavelength have negative influences on the performance of these methods. It is widely accepted that using variable selection methods before a multivariate calibration can improve the prediction performance, interpretability and operation speed of the model.

In recent decades, a number of variable selection methods have been developed, such as moving-window partial least-squares (MWPLS), interval partial least squares (iPLS), uninformative variables elimination (UVE), Monte Carlo based UVE (MC-UVE), competitive adaptive reweighted sampling (CARS), sampling error profile analysis combined with least absolute shrinkage and selection operator (SEPA-LASSO), iteratively retaining informative variables (IRIV), variable combination population analysis (VCPA) and others. Particularly, MWPLS is a classical variable selection method in chemometrics, and has been proven to be an effective tool to find informative intervals in the spectrum. However, it also has several shortcomings. In MWPLS, the calibration set is used to build models during the window moving and collect residual sums of squares (RSS) of the calibration set at different PLS latent variables (refer to the section, Moving-window partial least-squares (MWPLS), for details). But when only directly using the calibration set, it is hard to obtain a stable enough result for the variable selection and a number of RSS lines lead a plot unclear to select reasonable wavelength intervals. Besides, MWPLS is hard to find the combination effect between the wavelength intervals.

Sampling error profile analysis (SEPA) is a new algorithm
based on Monte Carlo Sampling (MCS), and it is useful for outlier detection, cross validation and model evaluation in multivariate calibration. The core idea of SEPA is repetitively establishing sub-models by Monte Carlo Sampling to a dataset and then executing statistics analysis for an error profile of these sub-models.

In this paper, a new variable selection method, called moving-window partial least-squares coupled with sampling error profile analysis (SEPA-MWPLS), is proposed which is a combination of MWPLS and SEPA. SEPA is used to repeatedly split the data into a sub calibration set and a sub cross validation (CV) set randomly. For each pair of sub sets, MWPLS is used to build a model in each moving window with the sub calibration set, and to calculate the RSS of cross validation with the sub cross validation set. The repeated splitting will yield a number of RSSs, and their median value is used as the final RSS of CV. Compared with MWPLS, SEPA-MWPLS uses the median of a number of RSSs of CV to replace RSS of the calibration. Thus, the result of variable selection will be more stable and reasonable. MWPLS often produce several intervals that are similar to MWPLS, is to build a series of PLS models with varying PLS latent variable in each position of the window and calculate a series of RSSs by these models. When the window moving is complete, we can acquire a residue matrix with the size of $w \times (p - w + 1)$. Different from MWPLS, the residue matrix in SEPA-MWPLS is calculated by a cross validation set rather than a calibration set. When the loops are done, we can obtain a series of residue matrixes with the number of $N$. Here $N$ is the Monte Carlo Sampling times (loop times) and the other parameters are the same as MWPLS. The maximum number of PLS latent variables was set to the optimal value of the model built on the full spectra range.

In each moving window SEPA-CV is performed, and thus it is possible to select an optimal PLS factor number for the window in the cross-validation process. In SEPA-MWPLS, RSSs at different latent variables (a column of the residue matrix with the size of $v \times (p - w + 1)$) are considered to select the optimal factor number for the window position. If there is a minimal RSS, the number at which the minimum locates is considered to be the PLS factor. Otherwise, the setting maximum number of PLS latent variables, saying $v$ is the factor. This step is equivalent to choose the most suitable PLS latent variables for every window. After the cross-validation step, we can obtain a clean RSS line, not like RSS lines in MWPLS. We then set a series of thresholds. For each threshold, those intervals together with the next $(w - 1)$ channels, whose RSSs are smaller than the threshold, are picked out as informative intervals.

Normally more than one interval can be selected as the informative ones. Then, the backward elimination strategy is used to search for a possible combination of the intervals for further optimization. The best latent variable and relevant Root Mean Square Error of Cross Validation (RMSECV) for the backward elimination strategy can be calculated by SEPA-CV or K-Fold CV.

In the last, the RMSECVs obtained by different thresholds are compared to pick out the best intervals for the model. The flow chart of SEPA-MWPLS is shown in Fig. 1.

### Backward elimination strategy

Backward elimination is a commonly used search strategy in feature subset selection. It will sequentially omit the uninformative features, and use left features building models to reduce the dimensions of the data and retain the informative features. For example, if one chooses 50 features in the data, then each model is based on 49 features, leaving out one feature at a time. The first omitted feature gives the poorest performing model in the cross-validation. This procedure is continued until one feature remains. And the features that give the best
performing model in the process of sequential omitting features are regarded as being the most suitable features.

In this study, the backward elimination strategy was used to search for informative intervals and their combination.

Experimental

Corn dataset

The corn data m5 spectra can be obtained from the website: http://www.eigenvector.com/data/Corn/index.html. It contains 855 samples of corn, and each sample was measured by the m5 NIR spectrometer. The NIR spectra recorded from 1100 to 2498 nm with an interval of 2 nm (700 wavelength points). Table 1 shows the profile of the y value of corn data. For each content, the data set was split into a calibration set (60 samples) and a prediction set (20 samples) by The Kennard and Stone28 in the present study.

Experimental

Pharmaceutical tablets dataset

The pharmaceutical tablets data can be obtained from the website: http://www.eigenvector.com/data/tablets/index.html. It contains 655 samples of pharmaceutical tablets that have been split into 155 calibration samples, 40 validation samples and 460 test samples. The NIR spectra were measured by two spectrometers from 600 to 1898 nm with an interval of 2 nm (650 wavelength points). The data contain the assay values of the active ingredient in the tablets. Referring to the other literature,29 several possible outliers (the calibration set: No. 19, 122, 126 and 127; the test set: No. 11, 145, 267, 294, 295, 313, 341, 342 and 343) were excluded in the study. The remaining data set was used in this study. The profile of the y value of the data is given in Table 1. Also, the calibration set and the validation set of the first spectrometer in the original data were mixed as the calibration set (151 samples) and the test set of the first spectrometer in the original data were the prediction set (451 samples) in this study.

Software and calculation

All calculations including PLS, MWPLS, SEPA-MWPLS and MC-UVE were carried out in Matlab (Ver. 2016b, The MathWorks, Inc) on a personal computer with an Intel Core CPU (i7- 8700k) and 32 GB RAM. All the data including NIR spectra X and properties y were mean centered for modeling. And the Monte Carlo sampling number for MC-UVE and SEPA-CV was 500. Because of the repeated Monte Carlo sampling the calculation of SEPA-MWPLS needs more time than PLS, MWPLS, and MC-UVE. In this study the calculation time of the proposed method is about 20 - 50 min.

Results and Discussion

Corn dataset

The results of variable selection methods including MCUVE, MWPLS and SEPA-MWPLS on corn datasets are summarized in Table 2, and the variables selected by these methods are shown in Fig. 2. We can easily find from the Table 2 that all of these variable selection methods can improve the performance of the model compared with the PLS method without variable selection. Also compared to the other variable selection methods, SEPA-MWPLS exhibit the minimum RMSEC, RMSECV and RMSEP on almost all of the four observed properties. It shows a ranking of prediction ability for these methods, which is as follows: SEPA-MWPLS > MC-UVE > MWPLS > PLS. Here, we take the property, content of protein as an example to discuss the results in detail.

Figure 3 shows the RSS lines obtained by MWPLS and SEPA-MWPLS. Figure 3(a) shows the trend of all lines curves.
obtained by different numbers of latent variables varying 1 to 9 (here 9 is the optimal number of latent variables of the PLS model built on the full spectral range). We need to observe the trend of all the RSS lines obtained from the calibration set in MWPLS to search for possible informative intervals. Unfortunately, we can find from the Fig. 3(a) that the plots are complex and have a lot of small peaks that will enhance the difficulty of finding informative intervals. However, compared to Figs. 3(a) and 3(b) obtained with SEPA-MWPLS is much more clear and easy to observe because SEPA-MWPLS only contain one RSS line. It should be mentioned that the RSS line in Fig. 3(b) is not one line in Fig. 3(a), because the former RSSs are yielded with SEPA-CV due to $N_{sub}$ models with sub calibration set and cross-validation set, which can reduce the contingency and instability of the results while the later one is calculated with only the calibration set. Therefore, the RSS line obtained with SEPA-MWPLS is not only cleaner, but also more robust and reliable.

There are two incisive peaks in the spectral range around 1750 and 2150 nm in Fig. 3(b), which are picked by SEPA-MWPLS, while MWPLS did not suggest the two intervals because they did not show low RSS in Fig. 3(a). These two intervals have been reported as important regions for modeling in some papers, they are assigned to C–H groups and N–H groups, respectively. Although MWPLS did not find the spectral intervals at around 1750 and 2150 nm with very small errors,

| Response | Method  | $nV_{AR}$ | RMSECV, % | $nLV$ | RMSEC, % | RMSEP, % |
|----------|---------|-----------|-----------|-------|-----------|-----------|
| Moisture | PLS     | 700       | 0.0206    | 10    | 0.0127    | 0.0156    |
|          | MC-UVE  | 60        | 0.0038    | 10    | 0.0022    | 0.0030    |
|          | MWPLS   | 207       | 0.0065    | 10    | 0.0045    | 0.0063    |
|          | SEPA-MWPLS | 71    | 0.0037    | 10    | 0.0024    | 0.0029    |
| Oil      | PLS     | 700       | 0.0716    | 9     | 0.0501    | 0.0747    |
|          | MC-UVE  | 81        | 0.0435    | 10    | 0.0236    | 0.0389    |
|          | MWPLS   | 84        | 0.0660    | 8     | 0.0396    | 0.0504    |
|          | SEPA-MWPLS | 76 | 0.0216    | 9     | 0.0139    | 0.0216    |
| Protein  | PLS     | 700       | 0.1484    | 9     | 0.0986    | 0.1209    |
|          | MC-UVE  | 81        | 0.0779    | 9     | 0.0664    | 0.0640    |
|          | MWPLS   | 652       | 0.1314    | 9     | 0.0920    | 0.1160    |
|          | SEPA-MWPLS | 82 | 0.0200    | 9     | 0.0110    | 0.0167    |
| Starch   | PLS     | 700       | 0.3034    | 10    | 0.2006    | 0.2381    |
|          | MC-UVE  | 150       | 0.1954    | 10    | 0.1395    | 0.1343    |
|          | MWPLS   | 133       | 0.2590    | 10    | 0.1772    | 0.2680    |
|          | SEPA-MWPLS | 73 | 0.0820    | 10    | 0.0511    | 0.1073    |

$nV_{AR}$: number of selected variables; RMSECV: root-mean-square error of cross validation; $nLV$: number of latent variables; RMSEC: root-mean-square error of calibration set; RMSEP: root-mean-square error of prediction.

Fig. 3  (a) RSS lines obtained by MWPLSR under different numbers of PLS latent variables varying from 1 to 9. (b) RSS line obtained by SEPA-MWPLS.

Fig. 2  Variables selected by MC-UVE, MWPLS and SEPA-MWPLS on corn dataset.
it found other informative intervals of around 1930, 2000, 2050 nm and so on. If combining these intervals including those at 1750 and 2150 nm, the model was also better than that built with the PLS method using whole wavelengths. But many variables (high to 650 spectral channels) were selected (see in Table 2).

It is shown in Fig. 2 that SEPA-MWPLS picked two spectra intervals. One is around 1750 nm and the other is around 2150 nm. Except for two informative intervals close to 1750 and 2150 nm regions, MC-UVE also picked several other intervals, but they do not make the model have better performance than the two intervals found by SEPA-MWPLS. RMSECV, RMSEC and RMSEP of the SEPA-MWPLS method were quite smaller than those of MC-UVE with values of 0.0200, 0.0110 and 0.0167 to 0.0779, 0.0664 and 0.0640, respectively.

As for the other properties, the wavelength channels selected by SEPA-MWPLS are also reasonable and effective. Concerning corn moisture, SEPA-MWPLS select spectra intervals at around 1916 and 2112 nm, which correspond to the water absorption and the combination of O–H bonds according to the literature.31,32 On oil dataset, the intervals at around 1726 and 2302 nm which respectively correspond to the second and first overtones of the C–H stretching mode and the combination of C–H vibrations are selected by SEPA-MWPLS.12 Concerning starch dataset, the intervals at around 1754 and 1982 nm, respectively, correspond to the second overtone of C–H bonds33 and the combination of O–H bonds. On the all of these four properties, the numbers of variables selected by SEPA-MWPLS are less than MWPLS (71, 76, 82 and 73 to 207, 84, 652 and 133). And the RMSECV, RMSEC and RMSEP obtained by SEPA-MWPLS are lower than those of MC-UVE and MWPLS except for the property of moisture. The performance of SEPA-MWPLS is similar to MC-UVE on the moisture dataset. SEPA-MWPLS shows better performance concerning most properties. However, in the property of moisture, SEPA-MWPLS does not have advantages. Because the moisture data only need 7 variables to build an excellent model,17 MC-UVE may not still select such a small number of variables. Besides, the strongly combined effect between two information intervals also influences work of MWPLS in SEPA-MWPLS.

Pharmaceutical tablets dataset

The results of variable selection methods, including MC-UVE, MWPLS and SEPA-MWPLS, on the pharmaceutical tablets dataset are summarized in Table 3, while the variables selected by these methods are shown in Fig. 4. Table 3 shows that SEPA-MWPLS has the best prediction performance among these selection methods in view of the errors of RMSECV, RMSEC and RMSEP. The values of RMSECV, RMSEC and RMSEP are, respectively, 2.647, 2.597 and 2.300, while these properties obtained by the method of the second-best performance (MC-UVE) are respectively 2.894, 2.757 and 2.870. Similar to the corn data, Fig. 5 shows that the RSS lines drawn by MWPLS and SEPA-MWPLS have several differences. In Fig. 5(b), SEPA-MWPLS suggests four subintervals which have relative low RSSs. They are, respectively, around 1131, 1200, 1350 and 1628 nm (respectively called a, b, c and d subinterval in the following). However, the subinterval d is not suggested by MWPLS as an informative interval because the RSS in the subinterval is not very low (shown in Fig. 5(a)). That once again reflects that the results directly given by the whole calibration set are different from the results given by a number pairs of calibration and cross validation sets with MCS, and SEPA-MWPLS may find out more stable informative intervals. SEPA-MWPLS preliminarily find out four subintervals.

### Table 3 Results of the pharmaceutical tablets dataset

| Methods or intervals | nVAR | RMSECV/ mg | nLV | RMSEC/ mg | RMSEP/ mg |
|----------------------|------|------------|-----|-----------|-----------|
| PLS                  | 650  | 3.787      | 6   | 3.177     | 3.947     |
| MC-UVE               | 102  | 2.894      | 6   | 2.757     | 2.870     |
| MWPLS                | 187  | 2.884      | 6   | 2.777     | 3.219     |
| SEPA-MWPLS (b&d)     | 52   | 2.647      | 6   | 2.597     | 2.300     |
| b                    | 29   | 3.341      | 6   | 3.115     | 3.760     |
| b&c                  | 56   | 3.248      | 6   | 3.078     | 3.641     |
| a&b&c&d              | 95   | 3.284      | 6   | 3.067     | 3.510     |

nVAR: number of variables; RMSECV: root-mean-square error of cross validation set; nLV: number of latent variables; RMSEC: root-mean-square error of calibration set; RMSEP: root-mean-square error of prediction. Spectral intervals: a, 1116 – 1146 nm; b, 1172 – 1228 nm; c, 1324 – 1376 nm; d, 1606 – 1650 nm.

Fig. 4 Variables selected by MC-UVE, MWPLS and SEPA-MWPLS on the pharmaceutical tablets dataset.
(a, b, c and d) in Fig. 5(b). However, in Table 3, we find their direct combination does not lead to good result (the RMSEP obtained by these four subintervals is 3.510, larger than 2.870), and simply decrease the number of subintervals by RSS also do not lead good performance (the RMSEP obtained by combined interval b and c or individual interval of b is 3.641 and 3.760 respectively, still too large). But finally SEPA-MWPLS gives a good result with the subintervals of b and d (the RMSEP obtained by combined interval b and d is 2.300). This should be attributed to the strategy of backward elimination. The process of how backward elimination simplifies the intervals is shown in Fig. 6. In Stage 1, it will use all of these four subintervals to build the model and obtain the RMSECV. Then, in Stage 2, it will successively omit one of the subintervals and use the left three to build models and obtain four RMSECVs. Then the intervals with the lowest RMSECV (subintervals a, b and d) are selected and preserved. In Stages 3 and 4, it will do a similar operation to Stage 2. And only one interval is preserved. Then, in Stage 5, it will compare the RMSECVs obtained by anterior stages, and the interval with lowest RMSECV is regarded as the final result. Though the subintervals b and c have the lowest RSS when they are used alone, their combination did not give the best result. The combination of subintervals b and d has the lowest RMSECV and RMSEP. The best two subintervals may have repetitive information, so we only need one subinterval of these two. Thus, only using several optimal subintervals to build the model may miss important information. Because of the combination effect of the variates, suboptimal intervals should also be in view. That is the reason why we use the backward elimination strategy.

Conclusions

In this study, a new variable selection method based on the WMPLS and SEPA, called moving window partial least-squares coupled with sampling error profile analysis (SEPA-MWPLS), is proposed. SEPA-MWPLS was applied to NIR spectroscopic analysis of corn and pharmaceutical tablets datasets. The results show that SEPA-MWPLS is better than MWPLS, giving more stable and reliable estimations of the number of PLS latent variables and CV error. And it is more likely to retain the intervals which have combination effects than WMPLS. Besides, it simplifies the figure of the RSS lines and makes it easier to determine the informative intervals. Compared to MC-UVE, SEPA-MWPLS shows good performance including smaller number of variables, trend to establish a model with lower RMSECV, RMSEC and RMSEP, indicating it is a promising variable selection method.
Acknowledgements

This work was supported by Technology Center, Shanghai Tobacco Group Corporation, Key Laboratory of Cigarette Smoke Research of CNTC (K2018-1-056P).

References

1. R. M. Balabin and S. V. Smirnov, *Anal. Chim. Acta*, 2011, 692, 63.
2. A. Lopez, S. Arazuri, I. Garcia, J. Mangado, and C. Jaren, *J. Agric. Food Chem.*, 2013, 61, 5413.
3. T. De Beer, A. Burggraeve, M. Fonteyne, L. Saerens, J. P. Remon, and C. Vervaet, *Int. J. Pharm.*, 2011, 417, 32.
4. A. Giraudo, S. Grassi, F. Savorani, G. Gavoci, E. Casiraghi, and F. Geobaldo, *Food Control*, 2019, 99, 137.
5. M. Meenu and B. Xu, *Food Chem.*, 2019, 289, 545.
6. E. Stella, R. Moscetti, R. P. Haff, D. Monarca, M. Cecchini, M. Contini, and R. Massantini, *J. Near Infrared Spectrosc.*, 2015, 23, 197.
7. S. E. Holroyd, *J. Near Infrared Spectrosc.*, 2013, 21, 311.
8. I. M. Johnstone and D. M. Titterington, *Philos. Trans.* 2009, 367, 109.
9. S. Wold, M. Sjostrom, and L. Eriksson, *Chemom. Intell. Lab. Syst.*, 2001, 58, 109.
10. M. Hartnett, G. Lightbody, and G. W. Irwin, *Chemom. Intell. Lab. Syst.*, 1998, 40, 215.
11. I. Guyon, *J. Mach. Learn. Res.*, 2003, 3, 1157.
12. J. Jian-Hui, B. R. James, H. W. Siesler, and O. Yukihiro, *Anal. Chem.*, 2002, 74, 3555.
13. L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, and S. B. Engelsen, *Appl. Spectrosc.*, 2000, 54, 413.
14. V. Centner, D. L. Massart, O. E. de Noord, J. S. De, B. M. Vandeginste, and C. Sterna, *Anal. Chem.*, 1996, 68, 3851.
15. W. S. Cai, Y. K. Li, and X. G. Shao, *Chemom. Intell. Lab. Syst.*, 2008, 90, 188.
16. H. Li, Y. Liang, Q. Xu, and D. Cao, *Anal. Chim. Acta*, 2009, 648, 77.
17. R. Q. Zhang, F. Y. Zhang, W. C. Chen, H. M. Yao, J. Ge, S. C. Wu, T. Wu, and Y. P. Du, *Chemom. Intell. Lab. Syst.*, 2018, 175, 47.
18. Y. H. Yun, W. T. Wang, M. L. Tan, Y. Z. Liang, H. D. Li, D. S. Cao, H. M. Lu, and Q. S. Xu, *Anal. Chim. Acta*, 2014, 807, 36.
19. Y. H. Yun, W. T. Wang, B. C. Deng, G. B. Lai, X. B. Liu, D. B. Ren, Y. Z. Liang, W. Fan, and Q. S. Xu, *Anal. Chem. Acta*, 2015, 862, 14.
20. K. Zheng, Q. Li, J. Wang, J. Geng, C. Peng, S. Tao, W. Xuan, and Y. Du, *Chemom. Intell. Lab. Syst.*, 2012, 112, 48.
21. R. Q. Zhang, F. Y. Zhang, W. C. Chen, Q. Xiong, Z. K. Chen, H. M. Yao, J. Ge, Y. Hu, and Y. P. Du, *Chemom. Intell. Lab. Syst.*, 2019, 184, 132.
22. X. B. Zou, J. W. Zhao, M. J. W. Povey, M. Holmes, and H. P. Mao, *Chemom. Intell. Acta*, 2010, 667, 14.
23. R. Leardi and L. Norgaard, *J. Chemom.*, 2004, 18, 486.
24. W. C. Chen, Y. P. Du, F. Y. Zhang, R. Q. Zhang, B. Y. Ding, Z. K. Chen, and Q. Xiong, *J. Chemom.*, 2018, 32, 11.
25. J. M. Sutter and J. H. Kalivas, *Microchem. J.*, 1993, 47, 60.
26. Q. S. Xu and Y. Z. Liang, *Chemom. Intell. Lab. Syst.*, 2001, 56, 1.
27. K. Z. Mao, *IEEE Trans. Syst. Man Cybern. Part B*, 2004, 34, 629.
28. R. W. Kennard and L. A. Stone, *Technometrics*, 1969, 11, 137.
29. W. Du, Z. P. Chen, L. J. Zhong, S. X. Wang, R. Q. Yu, A. Nordon, D. Littlejohn, and M. Holden, *Anal. Chim. Acta*, 2011, 699, 64.
30. B. C. Deng, Y. H. Yun, D. S. Cao, Y. L. Yin, W. T. Wang, H. M. Lu, Q. Y. Luo, and Y. Z. Liang, *Anal. Chim. Acta*, 2016, 908, 63.
31. H. D. Li, Y. Z. Liang, Q. S. Xu, and D. S. Cao, *Anal. Chim. Acta*, 2009, 648, 77.
32. B. C. Deng, Y. H. Yun, P. Ma, C. C. Lin, D. B. Ren, and Y. Z. Liang, *Analyst*, 2015, 140, 1876.
33. Y. H. Yun, W. T. Wang, B. C. Deng, G. B. Lai, X. B. Liu, D. B. Ren, Y. Z. Liang, W. Fan, and Q. S. Xu, *Anal. Chim. Acta*, 2015, 862, 14.