Environmental Technology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tent20

Near-infrared spectroscopy as a potential tool with radial basis function for measurement of residual acrylamide in organic polymer
Guocheng Zhu a, Huaili Zheng a, Peng Zhang b, Zhaojie Jiao c & Jun Yin d
a Key Laboratory of the Three Gorges Reservoir Region’s Eco-Environment, Chongqing University, Chongqing, China
b College of Civil Engineering, Hunan University of Science and Technology, Xiangtan, China
c Engineering Research Center for Waste Oil Recovery Technology and Equipment, Chongqing Technology and Business University, Chongqing, China
d Department of Civil and Environmental Engineering, University of Missouri, Columbia, USA
Available online: 12 Apr 2012

To cite this article: Guocheng Zhu, Huaili Zheng, Peng Zhang, Zhaojie Jiao & Jun Yin (2012): Near-infrared spectroscopy as a potential tool with radial basis function for measurement of residual acrylamide in organic polymer, Environmental Technology, DOI:10.1080/09593330.2012.683817
To link to this article: http://dx.doi.org/10.1080/09593330.2012.683817

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Near-infrared spectroscopy as a potential tool with radial basis function for measurement of residual acrylamide in organic polymer

Guocheng Zhua, Huaili Zhenga, Peng Zhangb, Zhaojie Jiaoc and Jun Yind

aKey Laboratory of the Three Gorges Reservoir Region’s Eco-Environment, Chongqing University, Chongqing, China; bCollege of Civil Engineering, Hunan University of Science and Technology, Xiangtan, China; cEngineering Research Center for Waste Oil Recovery Technology and Equipment, Chongqing Technology and Business University, Chongqing, China; dDepartment of Civil and Environmental Engineering, University of Missouri, Columbia, USA

(Received 17 October 2011; final version received 29 March 2012)

Poly(acrylamide-co-diallyldimethylammonium chloride) (PDA), which is usually prepared by free radical polymerization of acrylamide monomer (AM) onto the cationic monomer dimethyl diallyl ammonium chloride (DMDAAC), has been widely applied to wastewater treatment; however, the free-radical polymerization is always incomplete with residual AM remaining in the PDA. The residual AM affects the PDA’s performance while also posing as a potential threat to human health; therefore, during preparation of the PDA, the rapid detection of the residual AM plays an important role in controlling the residual AM while improving the PDA’s performance. The objective of this study was to explore the possibilities for applying near-infrared (NIR) spectroscopy as a potential tool for detecting the residual AM in combination with a statistical tool. In this study, the radial basis function (RBF) network model as the statistical tool was combined with NIR spectroscopy for detection of the residual AM. The experimental results showed that five wavelengths in the NIR spectroscopy were the most important characteristic adsorption peaks, particular at 971.95 and 1077 nm. The simulation of the RBF model presented higher performance with $R^2$-value greater than 0.98, RMSEC and RMSEP less than 7.22 $\times 10^{-5}$ and coefficient of variation (CV) of the predicted residual AM less than 10%, which demonstrated the feasibility of the NIR spectroscopy being a rapid detection tool for prediction of the residual AM using the RBF model. Wavelet de-noising was used for removing the interference/noise in the NIR spectroscopy and improved the generalization ability of the RBF model.

Keywords: near-infrared spectroscopy; radial basis function; acrylamide; organic polymer; environmental monitoring

Introduction

Poly(acrylamide-co-diallyldimethylammonium chloride) (PDA) is a generally used heavy polymer with a broad range of application in many fields, such as water/wastewater treatment and the paper and leather industries. The copolymer can be prepared by copolymerization of cationic monomer di-methyl di-allyl ammonium chloride (DMDAAC) and acrylamide monomer (AM) or another monomer through free radical polymerization [1,2]. During polymerization, if the synthetic reaction of the PDA is incomplete, a certain amount of AM will remain in the PDA. The polymer is considered non-toxic but the residual AM in the PDA may pose a potential threat to human health because of the acute toxicity, neurotoxicity and carcinogenicity involved in food chemistry and biotechnology fields etc. [3–6]. Research on the detection of AM is currently a hot topic [7–11]. The residual AM is not only hazardous to health, but also has a negative effect on the performance of the PDA; therefore, determination of the residual AM assists in controlling the monomer conversion and improving the performance of PDA. There are many available methods including gas chromatography (GC), liquid chromatography (LC), gas chromatography–mass spectroscopy (GC–MS), LC–MS, improved LC–MS/MS etc. [11–14] that have been used to determine the residual AM content; however, the application of these methods may be relatively complex, expensive and also unsuitable for rapid detection of the residual AM. The advantages in using near-infrared (NIR) spectroscopy are the low cost and high accuracy. Previous research has shown that NIR spectroscopy technology is used effectively for rapid detection of organic compound in combination with a numerical method [15–17].

NIR spectroscopy technology can be used in qualitative or quantitative analysis [18–24]. The qualitative analysis can be used to determine both the composition and structure of matter, whilst the quantitative analysis determines the material content [24–27]. However, in comparison with the usual chemical analysis methods, the NIR spectroscopy technology is an indirect quantitative analysis method, which is dependent on the establishment of a calibration model between the spectroscopy data (matrix $X$, as an
independent variable) and material content (matrix $Y$, as the dependent variable) [28,29].

There are many available statistical methods for developing the calibration models such as principal component analysis (PCR), partial least squares regression (PLSR), support vector machines (SVM), artificial neural network (ANN) etc. [30–36]. PCR can be used to minimize the problems of co-linearity effectively, but it has limitations with the number of variables and noise filtration. Furthermore, PCR cannot indicate if there is a strong correlation between the principal components of the matrices $X$ and $Y$. PLSR is an orthogonal basis of latent variables constructed in such a way that they are oriented along directions of maximal covariance between the spectral matrix $X$ and the response vector $Y$ [37] and that the latent variables are ordered according to their relevance for predicting the $Y$-variable. Interpretation of the relationship between $X$-data and $Y$-data (regression model) is then simplified because the relationship is based on the smallest possible number of latent variables [37]. Although the PLSR method is suitable for non-linear model development, the computational process is relatively complex. The SVM is a supervised learning method for classification and regression analysis, which plays an important role in modelling non-linear relationships whilst presenting higher efficiency in non-linear analysis using NIR spectroscopy [38,39]. In addition to the PCR, PLSR and SVM model, the ANN algorithm is another simple and effective tool for non-linear fitness between matrix $X$ and matrix $Y$ [40]. In some conditions when the near infrared spectroscopy indicates a severe peak overlapping, weak signal and big noise, the ANN model can effectively solve these problems [41].

The RBF model is a non-linear statistical data model generally used as an adaptive system by changing its structure based on external or internal information flowing through the network during the learning phase [42,43]. The learning process can model complex relationships between inputs and outputs or find patterns in data, and there is no specific formula describing the patterns, thus it acts as a black-box model. Although the RBF model can determine the hidden layer structure automatically, some important parameters affecting the prediction outcome of the model should be pre-defined before simulation. Currently, the response surface method (RSM) would be utilized as an available technology for analysis of the optimization conditions. RSM is an optimization method combining mathematical and statistical techniques. It is useful for the optimization of chemical reactions and/or industrial processes and widely used for experimental design [44,45]. In statistics, RSM can be used to explore the relationships between several explanatory variables and one or more response variables [46]. If the response variable’s value (matrix $Y$) was one or more explanatory variable’s function, when determining optimization values, the response variable’s value as a function of explanatory variables is described graphically; therefore, RSM has attracted more attention as an optimization tool [47–52].

The synthetic technology of the PDA has been widely studied, whereas the detection of the residual AM in the PDA using the NIR spectroscopy has not been well covered in literature. The objective of this study, therefore, was to analyse the feasibility of the radial basis function for the prediction of AM in PDA using NIR spectroscopy. In this study, NIR spectroscopy was collected and pre-processed using wavelet de-noising for removing the interference/noise of NIR spectroscopy. The RBF model was optimized and used to predict the residual AM content in organic polymer PDA.

Materials and method

Materials

The organic heavy polymer PDA was synthesized by free radical polymerization of AM onto the cationic monomer DMDAAC. At room temperature, a set amount of surfactants span 80 and tween 80, and paraffin as liquid-dispersed phase were mixed in fixed proportions in a four-neck flask. The solution was stirred with nitrogen gas until a homogeneous liquid mixture was obtained. The pre-determined proportion of AM and DMDAAC monomer solution was then added to the mixture. After 30 min of rapid mixing at room temperature, a set concentration of sodium bisulphite solution was added and emulsified for 20 min. Lastly, a set amount of potassium persulphate solution was added, followed by emulsification for 30 min in a water bath heated to 40 °C with the presence of nitrogen gas and a milk white copolymer was obtained. The obtained sample was pre-processed by de-emulsifying with acetone, washing with absolute ethyl alcohol and vacuum drying at 65 °C to produce PDA.

AM would influence the performance of the PDA, and thus an approximate amount was required during the preparation for PDA. In addition, increasing the DMDAAC content would contribute to increasing the cationic degree and charge neutralization ability of the PDA. This would, however, increase the cost of materials and an appropriate DMDAAC/AM mass ratio was therefore required for reaching better PDA performance. The drawback for the preparation was that the synthesis reaction was incomplete, resulting in residual AM in the PDA. Due to the influence of residual AM, the PDA sample used in the treatment of wastewater did not perform effectively during coagulation–floculation.

Method

Spectroscopy data collection

The residual AM content in the PDA samples was detected by a UV-visible spectrophotometer (TU-1901, Purkinje Gener Instrument Co. Ltd., China). The samples’ NIR spectrosopes were scanned by a near-infrared (NIR) analyser (NIRQuest 512, Ocean Optics Co. Ltd., USA), giving access to the spectral interval 900–1700 nm wavelength;
Environmental Technology

3

however, there were many small peaks and troughs leading to a spectral interference. The interference would influence the fitness performance of the calibration model, and thus spectral interference filters were necessary.

Pre-processing of spectroscopic data

The NIR spectroscopy technology mainly depends on the hydrogen vibration frequency and carbon monoxide (CO) absorbance frequency [15], and different groups have different absorbances at the same wavelength. The NIR spectroscopy is affected by factors such as weak absorption intensity in the near-infrared spectral region, sample light scattering, spurious and instrumental response etc., resulting in interference/noise. The interference/noise would influence the prediction accuracy of a calibration model as the NIR spectroscopy analysis was aimed at extracting the spectral information for further statistical analysis [53]. In controlling interference/noise data, there are many spectrum pre-process methods such as wavelet transform, smoothing, baseline correction, derivation, normalization etc. [54].

Wavelet processing has been used in many fields. It is faster and simpler in a sparse representation with few coefficients and thus of interest in this study. In addition, it can reduce high frequency noise and direct wavelet transforms can be computed from the original spectrum. In the present work, the effect of the wavelet de-noising on the elimination of interference/noise data for NIR spectroscopy was investigated. The wavelet de-noising was performed on the MatLab7.5 platform according to three steps involving wavelet decomposition of spectrum, threshold of wavelet decomposition with high frequency wavelet coefficients and wavelet reconstruction of spectrum.

RBF model

RBF neural network architecture is a three-layer network consisting of an input layer, a hidden layer and an output layer with a transfer function for a radial basis neuron [55]. The input layer is simply a fan-out layer, which was defined as \( x = (x_1, x_2, x_3, ..., x_n) \), consisting of \( n \)-nodes. Each node represents a variable. The \( n \)-value is the number of variables, which was wavelength in this research. Better simulation effect may be obtained by reducing the number of input layers [56]; therefore, in this study, characteristic peak was chosen as the node variable of input layer. The hidden layer was composed of many nodes, which performed a non-linear mapping from the input space into a higher dimensional space with the patterns becoming linearly separable [57]. The number of hidden nodes was optimized by the system. The output layer performed a simple weighted sum with a linear output, which was defined as \( y = (y_1, y_2, ..., y_m, ..., y_p) \).

The non-linear activation function allowed for non-linear mapping of the input space to the network output. The mapping for input layer was \( x \rightarrow R_i(x) \) and the output layer \( R_i(x) \rightarrow y_k \). The output layer value for neuron \( k \) was given by Equation (1):

\[
\hat{y}_k = \sum_{i=1}^{m} w_{ik} R_i(X), \quad k = 1, 2, ..., p \tag{1}
\]

where \( x = \) input vector; \( c_i = \) center vector for neuron \( i \) and determines the network performance at some degree; \( \sigma_i = \) \( i \)th perceptual variable and determines the width of focal point that the basis functions around, specifically the size of perceptual vision; \( m = \) number of hidden layers; \( \|x - c_i\| = \) vector \( x - c_i \) norm, which was usually denoted as the distance between \( x \) and \( c_i \); \( R_i(x) \) has only one maximum value at \( c_i \), and with the increase of \( x - c_i \) the value of \( R_i(x) \) rapidly reduced to zero.

The network determined the number of hidden layers, using a linear combination of radial basis functions as activation functions. In this study, the Gaussian basis function, which is the most commonly used radial-basis function, was used as indicated in Equation (2):

\[
R_i(x) = \exp \left( -\|x - c_i\|^2/2\sigma_i^2 \right), \quad i = 1, 2, ..., m \tag{2}
\]

where \( n = \) number of the input units; \( m = \) number of the hidden units; \( p = \) number of the output units; \( w_{ik} = \) weight between the \( i \)th neuron of hidden layer and the \( k \)th neuron of output layer; \( R_i(x) = \) activation function for the \( i \)th neuron of the hidden layer.

The simulation platform Matlab7.5 was used for the program design of the RBF neural network. A previous study observed that the spread of radial basis functions and the maximum number of network neurons affected the neural network [58]. In the case of large or small spread, many neurons would be required to fit a fast-changing function and a smooth function; however, the model may not generalize well [59,60] and in order to achieve the best forecasting capability, the appropriate spread of radial basis functions and the maximum number of network neurons were required.

Training and testing set

In this research, input data was the absorbance value at a specific wavelength, matrix \( X \) and output data as the residual AM content, matrix \( Y \). Input data was divided into two separate datasets: training set (training sample) and testing set (testing sample). The training set was used to fit the linear or non-linear model and the testing set was used to test the feasibility of the model.

Calibration models were composed of input and output variables. The input \( (X) \) and output variables \( (Y) \) were absorbance at a certain wavelength and the residual AM in the PDA, respectively. In addition, the absorbance at certain wavelength range could be selected as a variable; however, if the number of the selected wavelength was too large, the number of the variable was increased, causing the model to become more complex. The characteristic peak is often the
primary means used to identify substances; therefore, the absorbance at the characteristic peak was generally defined as the independent variable’s value. In this research, seven characteristic peaks were selected. The effect of absorbance at certain wavelength on the residual AM in the PDA was obtained using the regression coefficient ($A$) of standard equation. The $A$-value of standard equation was lower than 0.05 at 1506.94 nm and 1638.32 nm. At these points, the absorbance of the residual AM content was small. In contrast, the other five wavelengths were the maximum impact factors, particular at 971.95 and 1077 nm. After eliminating spectral interference, the characteristic peak of the PDA in the NIR range was easy to identify.

**Optimization of calibration model**

The response surface method (RSM) is generally employed to find the relationship between one or more response variables and a set of quantitative, experimental variables or factors [30–33]. In this research, the relationship between the parameters (independent variable), such as the spread of radial basis functions and the maximum number of network neurons, was investigated to find the most suitable combination of these variables that gave maximum turbidity efficiency. The effects of these parameters, with different levels of simulated performance, were assessed and three levels of each factor were selected for RSM analysis using central composite experimental design. The low, middle and high levels of each variable were designated as $-1.41$, 0 and $+1.41$, respectively, and the corresponding actual values (ranges and coded levels) for each variable are presented in Table 1. The experimental runs (1–14) were performed and presented in Table 2. The correlation coefficient ($R$) between observed and simulation value was selected as the response value (dependent variable).

$F$-test and $P$-values (probability), with 95% confidence level included in the variance analysis, were used to evaluate the statistical significance of the model. The quality of the model was expressed by the coefficient of determination $R^2$.

**Results and discussion**

**Pre-processing result of spectroscopic data**

The spectrum was pre-processed though the wavelet de-nosing and the first deviation to remove the interference information, thereby improving the discriminatory performance. After data pre-processing, the spectrum became smoother, which was improved simulation accuracy. The experimental results are shown in Figure 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Neurons number</th>
<th>Spread</th>
<th>R-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurons number</td>
<td>8.59</td>
<td>9</td>
<td>0.9883</td>
</tr>
<tr>
<td>Spread</td>
<td>1.79</td>
<td>2</td>
<td>0.9883</td>
</tr>
</tbody>
</table>

**Optimization result of calibration model**

In order to investigate the effect of interaction on the spread of radial basis functions and the maximum number of network neurons to obtain the optimum value, RSM was applied to determine optimization conditions based upon the NIR spectrums with wavelet de-nosing. The relationship between the spread of radial basis functions, the maximum number of neurons and $R$-value (calibration set) was predefined according to the following Equation (3):

$$R = \beta_0 + \beta_1 N + \beta_2 S + \beta_{12} N \cdot S + \beta_{11} N^2 + \beta_{22} S^2$$

where $\beta_0 =$ constant; $\beta_1$ and $\beta_2 =$ linear coefficients; $\beta_{11}$ and $\beta_{22}$ = quadratic coefficients.

Multiple regression analysis was used to determine the relationship between independent variables and dependent variables. The multiple regression analysis was showed that $\beta_0$, $\beta_1$, $\beta_2$, $\beta_{12}$, $\beta_{11}$ and $\beta_{22}$ are 0.970, $-0.000500$, 0.0135, 0, 0.000025 and $-0.00210$, respectively. The results of analysis of variance and regression analysis are as shown in Table 3.

Table 3 shows that $R^2 = 0.999$ and the adjusted $R^2 = 0.998$. After performing an $F$-statistic test, the $F$-value = 1151.47 and the $P$-value = 0. The $F$-statistic test at $P < 0.0001$ was considered as significant; therefore, the multiple regression was feasible.

The interaction effect of the spread of radial basis functions and the maximum number of network neurons on $R$-value using the regression function is illustrated in Figure 2. Figure 2 indicates that the spread of radial basis functions and the maximum number of network neurons has a significant interaction on $R$-value. In addition, it shows that the $R$-value achieves better results when the spread of radial basis functions is more than 3.0. In order to obtain the optimized conditions using the regression equation, the partial derivatives of the spread of radial basis functions and the maximum number of network neurons were set to zero. The calculated extreme values were adopted as the
optimum value. In this study, the calculation result shows that the spread of radial basis functions and the maximum number of network neurons were 3.21 and 10, respectively.

Simulation analysis of calibration model
In order to investigate the performance of the RBF model on the prediction of the residual AM in the PDA, the simulation experiment was conducted at the spread of radial basis functions of 3.21 and the maximum number of network neurons i.e., 10. In addition, a simple comparison was made between PLSR and RBF. The simulation performance of the RBF model was assessed using the squared correlation coefficient ($R^2$) for the observed and modelled residual AM value, the root mean squared error of the modelled residual AM for the calibration samples (RMSEC) and the root mean squared error of the modelled residual AM for the validation samples (RMSEP) between the observed and modelled value. Coefficient of variation (CV), in statistics, is a normalized measure of dispersion of a probability distribution.

Table 3. Regression analysis and analysis of variance.

<table>
<thead>
<tr>
<th>Item</th>
<th>$R^2$</th>
<th>Adj. $R^2$</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance $F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression analysis</td>
<td>0.999</td>
<td>0.998</td>
<td>5</td>
<td>2.04887E-05</td>
<td>4.09773E-06</td>
<td>1151.47</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual error</td>
<td></td>
<td></td>
<td>8</td>
<td>2.84697E-08</td>
<td>3.55871E-09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td>13</td>
<td>2.05171E-05</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. NIR spectrum after pre-treatment with (a) calibration set with wavelet de-nosing; (b) validation set wavelet de-nosing; (c) calibration set with first deviation; and (d) validation set with first deviation.
The CV as the ratio of the standard error of estimate to the mean value of the response for a model aims to describe the model fit in terms of the relative sizes of the squared residuals and outcome values, and defines reproducibility of the model [61,62]. As a general rule, a model can be considered reasonably reproducible if its CV is not greater than 10% [62,63]. In this research, CV\textsubscript{train} and CV\textsubscript{test} of the modelled residual AM fitting training sample and testing sample, respectively, were examined. The statistical results are presented in Table 4.

Table 4 shows the parameters’ values for evaluation of RBF and PLSR simulation performance. Using raw NIR spectroscopy and the treated spectroscopy with wavelet de-nosing, RBF simulation presented higher $R^2$-values of 0.979–0.997 for both training and testing samples in comparison to NIR spectroscopy after pre-treatment of the first deviation. In general, RMSEC is used to describe the training accuracy of the calibration model fit of the training set and the RMSEP is used to estimate the generalization ability of the calibration model [64]. It was observed that the RBF simulation with the NIR spectroscopy pre-processed by the wavelet de-nosing and raw spectroscopy both indicated better RMSEC and RMSEP values of $<7.22 \times 10^{-5}$; however, the RBF simulation with spectroscopy pre-treated with wavelet de-nosing had a lower RMSEP/RMSEC value (close to zero) and a lower CV-value than the raw spectroscopy. Although the simulation with NIR spectroscopy after pre-treatment of the first deviation indicated a lower level of RMSEC ($<3.06 \times 10^{-4}$), the RMSEP value of the predicted residual AM was $>8 \times 10^{-4}$ with a lower $R^2$-value. With the raw NIR spectroscopy and the spectroscopy pre-processed by the wavelet de-nosing, the RBF simulation showed that the CV of the prediction value of the residual AM for both training and testing samples was $<10\%$. This indicated better precision and reliability of the experiments; however, using the NIR spectroscopy

<table>
<thead>
<tr>
<th>Analytical method</th>
<th>Correction method of NIR spectroscopy</th>
<th>Calibration set</th>
<th>Validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$R^2_{\text{train}}$</td>
<td>RMSEC</td>
</tr>
<tr>
<td>RBF</td>
<td>None</td>
<td>0.993</td>
<td>2.80 $\times 10^{-5}$</td>
</tr>
<tr>
<td>RBF</td>
<td>Wavelet de-nosing</td>
<td>0.979</td>
<td>5.98 $\times 10^{-5}$</td>
</tr>
<tr>
<td>RBF</td>
<td>The first deviation</td>
<td>0.294</td>
<td>3.06 $\times 10^{-4}$</td>
</tr>
<tr>
<td>PLSR</td>
<td>None</td>
<td>0.995</td>
<td>6.03 $\times 10^{-5}$</td>
</tr>
<tr>
<td>PLSR</td>
<td>Wavelet de-nosing</td>
<td>0.986</td>
<td>4.11 $\times 10^{-5}$</td>
</tr>
<tr>
<td>PLSR</td>
<td>The first deviation</td>
<td>0.718</td>
<td>1.96 $\times 10^{-4}$</td>
</tr>
</tbody>
</table>
pre-processed by the first deviation, a higher CV\text{test} value of 10.587% was achieved, suggesting that the wavelet de-nosing contributed to improving the generalization ability of the RBF model. Table 4 also illustrates that the PLSR simulation with the NIR spectroscopy pre-processed by the first deviation showed lower simulation performance with lower $R^2$-values, although the CV values were <10%. With the pre-processed spectroscopy by the wavelet de-nosing, the simulation results of the PLSR and RBF had the similar levels of RMSE and CV; however, RBF indicated better predication performance with higher $R^2$ values using the validation set than with the spectroscopy pre-processed by wavelet de-nosing. The above discussion shows that the RBF, as well as the conventional tool PLSR model, could be used to predict the residual AM in the PDA sample using the NIR spectroscopy.

Conclusions
This paper presents an available method to detect the residual AM in the PDA with NIR spectroscopy and the RBF model statistical tool. The experimental results show that the simulation for both testing dataset and training dataset revealed a high simulation performance with $R^2$-value of >0.97, RMSEC and RMSEP of $5.98 \times 10^{-5}$ and $7.22 \times 10^{-5}$, respectively, and a CV of the predicted residual AM <10%. The statistical analysis suggests that using NIR spectroscopy as a rapid detection tool for predicting the residual AM in the PDA in combination of the RBF model is feasible. Due to large amounts of interference/noise in the collected NIR spectrum, pre-treatment of the NIR spectroscopy was required. For removing the interference/noise information, the NIR spectroscopy was pre-processed using wavelet de-nosing and first deviation. After the wavelet de-nosing of the NIR spectroscopy, the spectrum became smoother and the generalization ability of the RBF model in comparison to the raw spectroscopy was increased. With NIR spectroscopy pre-processed by wavelet de-nosing, it would present better effect on fitness. In contrast, the simulation performance using NIR spectroscopy pre-processed by the first deviation was investigated and showed a lower prediction level and did not achieve the desirable accuracy compared to wavelet de-nosing. The wavelet de-nosing was therefore an approximate and favourable method to pre-treat the NIR spectroscopy and it improved the simulation performance in this study.

Acknowledgements
The authors are grateful for the financial support provided by the Scholarship Award for Excellent Doctoral Student granted by Ministry of Education (Project NO.0903005109081-7), China; National Natural Science Foundation of China (Project NO. NSFC, 51078366, 21177164); and Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (Project NO.708071)

References


