Journal of Visualized Experiments

O-cresol Concentration online measurement based on Near Infrared Spectroscopy via Partial Least Square Regression

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE59077R4
Full Title:	O-cresol Concentration online measurement based on Near Infrared Spectroscopy via Partial Least Square Regression
Keywords:	Near infrared spectroscopy; Polyphenylene ether; O-cresol Concentration; Partial least square regression; Online measurement; Prediction accuracy
Corresponding Author:	Xiaoli Luan Jiangnan University Wuxi, CHINA
Corresponding Author's Institution:	Jiangnan University
Corresponding Author E-Mail:	xlluan@jiangnan.edu.cn
Order of Authors:	Zihao Chen
	Xiaoli Luan
	Niannian Zheng
	Fei Liu
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Wuxi, Jiangnansu,China

TITLE:

O-cresol Concentration online measurement based on Near Infrared Spectroscopy via Partial Least Square Regression

AUTHORS:

Zihao Chen[1], Niannian Zheng[1], Xiaoli Luan[1], Fei Liu[1]
[1]Key Laboratory for Advanced Process Control of Light Industry of the Ministry of Education,
Institute of Automation, Jiangnan University, Wuxi, China

Email Addresses:
Zihao Chen: 1070415628@vip.jiangnan.edu.cn
Niannian Zheng: 6181905058@stu.jiangnan.edu.cn
Xiaoli Luan: xlluan@jiangnan.edu.cn
Fei Liu: fliu@jiangnan.edu.cn

Corresponding Author:
Xiaoli Luan

KEYWORDS:

near infrared spectroscopy, polyphenylene ether, O-cresol concentration, partial least square regression, online measurement, prediction accuracy

SHORT ABSTRACT:

The protocol describes a method of predicting o-cresol concentration during the production of polyphenylene ether using near-infrared spectroscopy and partial least squares regression. To describe the process more clearly and completely, an example of predicting the o-cresol concentration during the production of polyphenylene is used to clarify the steps.

LONG ABSTRACT:

Unlike macroscopic process variables, near-infrared spectroscopy provides process information at the molecular level and can significantly improve the prediction of the components in industrial processes. The ability to record spectra for solid and liquid samples without any pretreatment is advantageous and the method is widely used. However, the disadvantages of analyzing high-dimensional near-infrared spectral data include information redundancy and multicollinearity of the spectral data. Thus, we propose to use partial least squares regression method, which has traditionally been used to reduce the data dimensionality and eliminate the collinearity between the original features. We implement the method for predicting the o-cresol concentration during the production of polyphenylene ether. The proposed approach offers the following advantages over component regression prediction methods: 1) partial least squares regression solves the multicollinearity problem of the independent variables and effectively avoids overfitting, which occurs in a regression analysis due to the high correlation between the independent variables; 2) the use of the near-infrared spectra results in high accuracy because it is a non-destructive and non-polluting method to obtain information at microscopic and molecular scales.

INTRODUCTION:

Near infrared (NIR) spectroscopy (NIRS) has gained wide acceptance as a fast, efficient, non-destructive, and non-polluting modern analytical technology; the method has been used during the past several years for product quality detection and analysis and chemical component measurement in industrial processes. The most essential specialty of the method is its ability to record spectra for solid and liquid samples without any pre-processing, making NIRS especially suitable for the direct and rapid detection and analysis of natural and synthetic products[1,2]. Unlike traditional sensors that measure process variables (e.g., temperature, pressure, liquid level, etc.) at a macroscopic scale and inevitably suffer the external noise and background interference, NIRS detects the structural information of the chemical composition at microscopic and molecular scales. Thus, essential information can be measured more accurately and effectively than with other methods[3,4].

Polyphenyl ether, as one of the engineering plastics, are widely used due to its heat resistance, flame retardant, insulation, electrical properties, dimensional stability, impact resistance, creep resistance, mechanical strength and other properties[5]. More importantly, it is non-toxic and harmless compared to other engineering plastics. At present, 2,6-xylenol is one of the basic raw materials for the synthesis of polyphenylene ether, and it is usually prepared by catalyzed alkylation of phenol with methanol method[6]. There are two main products of this preparation method, o-cresol and 2,6-xylenol. After a series of separation and extraction steps, 2,6 xylenol is used to produce polyphenylene ether. However, trace amounts of o-cresol remain in 2,6-xylenol. O-cresol does not participate in the synthesis of polyphenylene ether and will remain in the polyphenylene ether product, resulting in a decrease in product quality or even the substandard. At present, most companies still analyze the compositions of complex organic mixtures such as liquid phase polyphenyl ether products containing impurities (e.g., o-cresol) by physical or chemical separation analysis such as chromatography[7,8]. The separation principle of chromatography is the use of the mixture of compositions in the fixed phase and the flow phase in the dissolution, analysis, adsorption, desorption or other affinity of the minor differences in the performance. When the two phases move relative to each other, the compositions are separated by the above actions repeatedly in the two phases. Depending on the object, it usually takes a few minutes to a few tens of minutes to complete a complex material separation operation. It can be seen that the measurement efficiency is low.

Nowadays, the measurement of product quality and the advanced control technology based on this analysis for the modern fine process chemical materials industry is the key direction to further improve product quality. In the process industry of polyphenyl ether production, real-time measurement of o-cresol content in polyphenylene ether product is of great development significance. Chromatographic analysis clearly cannot meet the requirements of advanced control technology for real-time measurement of substances and signal feedback. Therefore, we propose the partial least squares regression (PLSR) method to establish a linear model between the NIRS data and the o-cresol concentration, which realize the online measurement of o-cresol content in the liquid polyphenylene ether product of outlet.

The pre-processing for NIRS plays the most important role prior to multivariate statistical modeling. NIRS wavenumbers in the NIR spectrum and the particle sizes of biological samples are comparable, so it is known for unexpected scatter effects that has influence on the recorded sample spectra. By performing appropriate pre-processing methods, these effects are easy to be eliminated largely[9]. The most commonly used pre-processing techniques in NIRS are categorized as scatter correction and spectral derivative methods. First group of methods includes multiplicative scatter correction, detrending, standard normal variate transformations, and normalization. The spectral derivation methods include the use of the first and second derivatives.

Prior to developing a quantitative regression model, it is important to remove the unsystematic scatter variations from the NIRS data because they have a significant influence on the accuracy of the predictive model, its complexity and parsimony. The selection of a suitable pre-processing method should always depend on the subsequent modeling step. Here, if the NIR spectral dataset does not follow the Lambert-Beer law, then other factors tend to compensate for the non-ideal behavior of the prediction for predicted components. The disadvantage of the existence of such needless factors leads to the increase of model complexity, even most likely, a reduction in the robustness. Thus, the application of spectral derivatives and a conventional normalization to the spectral data is an essential part of the method.

After spectral preprocessing, the NIRS data with a high signal-to-noise ratio and low background interference are obtained. Modern NIRS analysis provides the rapid acquisition of large amounts of absorbance over an appropriate spectral range. The chemical composition of the sample is then predicted by extracting the relevant variables using the information contained in the spectral curve. Generally, NIRS is combined with multivariate analysis techniques for qualitative or quantitative analyses[10]. A multivariate linear regression (MLR) analysis is commonly used for developing and mining the mathematical relationship between the data and the components in industrial processes and has been widely used in NIRS analysis.

However, there are two fundamental problems when implementing an MLR for preprocessed NIRS data. One problem is the variable redundancy. The high dimensionality of the NIRS data often renders the prediction of a dependent variable unreliable because variables are included that have no correlation with the components. These redundant variables reduce the information efficiency of the spectral data and affect the accuracy of the model. In order to eliminate the variable redundancy, it is essential to develop and maximize the correlation between the NIRS data and the predicted components.

Another problem is the issue of multicollinearity in the NIRS data. One of the important assumptions of multiple linear regression models is that there is no linear relationship between any of the explanatory variables of the regression model. If this linear relationship exists, it is proved that there is multicollinearity in the linear regression model and the assumption is violated. In multiple linear regressions, such as an ordinary least squares regression (OLSR), multiple correlations between the variables affect the parameter estimation, increase the model error, and affect the stability of the model. To eliminate the multilinear correlation between the NIR spectral data, we use variable selection methods that maximize the inherent variability of the samples.

Here, we propose to use the PLSR, which is a generalization of multiple linear regression that has been widely used in the field of NIRS[11,12]. The PLSR integrates the basic functions of the MLR, canonical correlation analysis (CCA), and principal component analysis (PCA) and combines the forecasting analysis with a non-model data connotation analysis. The PLSR can be divided into two parts. The first part selects the components of the characteristic variables and the predicted components by partial least squares analysis (PLS). PLS maximizes the inherent variability of principal components by making the covariance of the principal components and predicted components as large as possible when extracting the principal components. Next, the OLSR model of o-cresol concentration is established for the principal components selected. PLSR is suitable for the analysis of noisy data with numerous independent variables that are strongly collinear and highly correlated and for the simultaneous modeling of several response variables. Also, PLSR extracts the effective information of the sample spectra, overcomes the problem of multicollinearity, and has the advantages of strong stability and high prediction accuracy[13,14].

The following protocol describes the process of using the PLSR model for measuring the o-cresol concentration using NIR spectral data. The reliability and accuracy of the model are evaluated quantitatively by using the determination coefficient (R^2), the prediction correlation coefficient (R_p) and the mean square prediction error of cross-validation (MSPECV). Moreover, to intuitively show the advantages of the PLSR, the evaluation indicators are visualized in several plots for a qualitative analysis. Finally, evaluation indicators of an experiment are presented in table format to quantitatively illustrate the reliability and precision of the PLSR model.

PROTOCOL:

1. NIR spectrum data acquisition with Fourier transform (FT)-NIR process spectrometer

1.1 Install the liquid phase optical fiber probe of the near-infrared spectrometer at the outlet of the polyphenyl ether product. And open the **OPUS** software on the upper computer connected to the instrument and start to configure the measurement.

1.2 Connecting to spectrometer

1.2.1 On the **Measure** menu, select the **Optic Setup and Service** command, or click the icon from the toolbar.

1.2.2 On the dialog that opens, click the **Optical Bench** *tab*.

1.2.3 Check whether the spectrometer settings are ok. If yes, close the dialog. If no, continue with step 4.

1.2.4 From the **Configuration** drop-down list, select the particular spectrometer type.

1.2.5 Enter the spectrometer's IP address into the **Optical Bench URL** entry field.

1.2.6 Click the **Connect** button.

1.3 Setting up measurement parameters

1.3.1 On the **Measure** menu, select the **Measurement** command, or click the icon from the toolbar.

1.3.2 On the dialog that opens, define the measurement parameters on the different tabs.

NOTE: Details on the individual measurement parameters are described in the OPUS Reference Manual.

1.3.3 Click the **Accept & Exit** button.

1.4 Storing experiment file

1.4.1 On the **Measure** menu, select the **Advanced Measurement** command. Then, click the **Advanced** tab.

1.4.2 On the dialog that opens, define the resolution as **4 cm^{-1}**.

1.4.3 Define the number of scans as **16** scans in the **Sample/Background Scan Time** entry fields.

1.4.4 Define the path to automatically store the measuring data from **4,000 cm^{-1}-12,500 cm^{-1}**.

1.4.5 Determine the data type for the result spectrum as **Absorbance**.

1.4.6 Click the **Save** button.

1.4.7 On the dialog that opens, define a name for the experiment file and save this name.

1.5 Measuring background spectrum

1.5.1 On the **Measure** menu, select the **Advanced Measurement** command.

1.5.2 Click the **Optic** tab.

1.5.3 On the dialog that opens, click the **Aperture setting** drop-down list and select the same value used to acquire a sample spectrum.

1.5.4 Click the **Basic** tab.

1.5.5 On the dialog that opens, click the **Background Single Channel** button.

1.6 Measuring sample spectrum

1.6.1 Place the sample into the optical path of the spectrometer. The way in which this is done depends on the spectrometer configuration.

1.6.2 On the **Measure** menu, select the **Advanced Measurement** command.

1.6.3 Click the **Basic** tab.

1.6.4 On the dialog that opens, define the sample description and sample form in the particular entry field. This information is stored together with the spectrum.

1.6.5 Click the **Sample Single Channel** button to start online measurement. And save the NIR spectrum of each scan as OPUS file.

1.7 Collect the polyphenylene samples every 6 h and test the o-cresol concentration with liquid chromatography in the laboratory of industry to obtain a chemical reference value.

NOTE: Laboratory staff of industry field take each polyphenyl ether sample from the outlet of the liquid phase polyphenyl ether. The o-cresol content in each sample was measured three times by liquid chromatography. Then, the mean value of the results of the three times analysis was taken as the reference value of the o-cresol content to reduce the accidental error.

1.8 Obtain 600 chemical reference values of o-cresol concentration in the laboratory. The calibration range of o-cresol concentration is from 42.1063 mg/1 g polyphenyl ether product to 51.6763 mg/1 g polyphenyl ether product.

1.9 Combine the NIR spectra at the given test times with the chemical reference values of the o-cresol concentration.

1.10 Use the software OPUS to read the original spectral set as shown in **Figure 1**.

1.10.1 On the **File** menu, click the **Load File** command.

1.10.2 On the dialog that opens, select the particular spectrum file.

1.10.3 Click the **Open** button. The spectrum is displayed in the spectrum window.

2. NIR spectroscopy data pre-processing

2.1 With the spectral preprocessing function in, obtain spectral dataset preprocessed with first-order derivative.

2.1.1 Open **The Unscrambler** which is a multivariate data analysis and experimental design software, select the **Import** command under **File**. Import the OPUS file as original NIR spectral dataset.

2.1.2 Select **Transform** command under **Modify**. And select the **Savitzky Golay Derivatives** under **Derivatives**.

2.1.3 Define the Samples and Variables as All Samples and All Variables in Scope. And define the number of **Smoothing points** as **13** and the **Derivative** as 1st derivative in Parameters.

2.1.4 Click **OK** to start the derivative.

CAUTION: The increase of smoothness can reduce the sharp fluctuations of the curve, reduce the noise effect but also weaken the characteristics of the curve and make the curve distorted. Therefore, the appropriate smoothness selected according to the observation of the actual fluctuation intensity of the curve and the effect after processing.

2.2 Perform vector normalization on the sample spectra to normalize the value of the absorbance.

2.2.1 Select the **Normalization** command under **Modify**.

2.2.2 Define the **Samples and Variables** as **All Samples** and **All Variables** in **Scope**.

2.2.3 Select **Vector normalization** in the **Type.**

2.2.4 Click **OK** to perform vector normalization.

3. Establishment of PLSR model

3.1 Creation of the NIR spectral data set

3.1.1 Open **Uncrambler**.**exe**, select **Export** under **File** with the **Matlab** files to export the preprocessed spectral data set into .mat File and to obtain the spectral data set X automatically with 2203 variables.

3.1.2 Obtain a complete NIR spectral dataset X (a matrix of 600 rows and 2203 columns) and the corresponding chemical reference values Y (a vector of 600 rows) in the form of .mat file for subsequent analysis and modeling.

3.2 Selection of the appropriate number of principal components

3.2.1 Open **Matlab** and import the .mat file containing the preprocessed near-infrared spectral data into the workspace by dragging the .mat file to the workspace.

NOTE: The .mat file stores the near-infrared spectral data X as an independent variable and the o-cresol content of the product as a dependent variable in the form of two matrices.

3.2.2 Open the programmed .m file in the **Editor**. Click **Open** under the **Editor** option, select the compiled .m file in the file storage directory, and then click **Confirm**.

3.2.3 Extract *15* principal components according to the optimization objective of **Equation 1** and the OLSR model between the extracted principal components and the predicted values of the o-cresol concentration with the program containing the command plsregress() in **Matlab**.
[XL, YL, XS, YS, BETA, PCTVAR, MSE] = plsregress(X,Y,ncomp,'CV',k);
Consult the MATLAB help document to get the usage details and the return value.

NOTE: $\mathrm{Cov}(\mathrm{u}_i, y) = \sqrt{Var(\mathrm{u}_i)}Cor(\mathrm{u}_i, y) \rightarrow max$ **Equation 1**
$\mathrm{u}_i = Xp_i$, and p_i is the i_{th} principal components of the NIR spectral data;
$Var(\mathrm{u}_i)$ is the projection of the i_{th} principal components of the NIR spectral data;
$Cor(\mathrm{u}_i, y)$ is the Pearson correlation coefficient for the i_{th} principal components and the o-cresol concentration.

3.2.4 Obtain the R^2 value of the NIR spectral data and the predicted values for the different principal components using **Equation 2**.

NOTE: $R^2 = 1 - \frac{SSE}{SST}$ **Equation 2**
SSE is the sum of squares due to error and is defined as $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$;
SST is the total sum of squares and is defined as $SST = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$;
y_i is the reference value of the o-cresol concentration of test dataset;
$\hat{y}_i$ is the predicted value of the o-cresol concentration of test dataset;
$\bar{y}_i$ is the mean value of reference value of the o-cresol concentration of test dataset;
n is the number of samples of test dataset.

3.2.5 Determine the R^2 values and the trend with increasing number of principal components as shown in **Figure 2**. Select 10 as the appropriate number of principal components with the R^2 value of 0.9917.

NOTE: R^2 value is the proportion of the variance in the dependent variable that is predictable by the independent variables. The higher the R^2 value is, the higher the goodness-of-fit is and vice versa.

3.3 Validation of the goodness-of-fit and accuracy of the PLSR model with *10* principal components by using the command plsregress().

3.3.1 Repeat the modeling process with 10 principal components as steps 3.2.1-3.2.5 with 10 principal components.

3.3.2 Evaluate the model based on a 10-fold cross-validation using the plots of the percent variance explained in the NIR spectral data, the residuals, and the MSPECV.

3.3.3 Plot the percent variance explained in NIR spectral data, the residuals, and the MSPECV as **Figures 3**, **4,** and **5**.

3.3.4 Tabulate the evaluation indicators of R^2,R_p, and MSPE of 10-fold cross validation for the PLSR model for a quantitative analysis as shown in **Table 1**.

NOTE: The equations of R_p and MSPE are shown as **Equation 3** and **Equation 4**.

$$R_p = \frac{Cov(y_i,\hat{y}_i)}{\sigma_{y_i}\sigma_{\hat{y}_i}} \qquad \text{Equation 3}$$

$$\text{MSPE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad \text{Equation 4}$$

$Cov(y_i,\hat{y}_i)$ is the covariance of reference value and predicted value of o-cresol concentration;
σ_{y_i} is the standard deviation of reference value of o-cresol concentration;
$\sigma_{\hat{y}_i}$ is the standard deviation of predicted value of o-cresol concentration.

REPRESENTATIVE RESULTS:

The predicted value of o-cresol Impurity in polyphenyl ether products is obtained by PLSR-based near-infrared spectroscopy. **Figure 2** and **Figure 3** respectively show the reliability of the method in the feature selection stage from the curve of the decision coefficient and the error interpretation percentage increasing with the number of principal components.

Specifically, please note that in the selection of the principal components, in order to minimize the complexity of the model, generally, when the two indicators do not increase significantly with the number of principal elements, the subsequent principal elements with less information can be discarded. In this paper, according to the two curves, when R^2 is 0.9917, it is the threshold to discard subsequent principal components.

Visually test the accuracy and stability of the method for predicting the purity of o-cresol products by means in shown **Figure 4** and **Figure 5**. Based on the training set and test set generated by the o-cresol samples obtained from industrial field and their corresponding near-infrared spectroscopy data, we compared the residual and mean square error of the predicted and reference values of the purity of polyphenyl ether products.

The residual here refers to the difference between the o-cresol content reference value and the PLSR model estimate (fitted value). Using the information provided by the residuals can help us examine the rationality of the established PLSR model assumptions and the reliability of collecting near-infrared spectral data. It is shown that PLSR method effectively reduce the residuals to an acceptable range. Moreover, the PLSR has a small fluctuation range from -0.2 to

0.2, while the calibration range is from 42.1063 mg/1 g polyphenyl ether product to 51.6763 mg/1 g polyphenyl ether product. The residual plot data let us conclude qualitatively that the PLSR for the measurement of the o-cresol content based on the NIR spectral data has high accuracy.

The cross-validation mean square error is a measure of the degree of difference between the reference and the predicted o-cresol content. This can help us evaluate the degree of change in the predictive data. The smaller the value of MSE, the better the accuracy of the predictive model describing o-cresol content. **Figure 5** indicates that the MSPECV for the o-cresol concentration measurement based on the PLSR decrease as the number of principal components increases and reach an acceptable minimum at 10 principal components. Moreover, the error decreases significantly and the descent process is relatively stable. This proves that the PLSR results in high stability for the measurement of the o-cresol concentration using NIRS.

The model evaluation indicators for a 10-fold cross-validation are shown in **Table 1**. The R^2 of 0.98332 is pretty high for the PLSR, indicating that the model based on the PLSR well reflects the linear relationship between the NIR spectral data and the o-cresol concentration (i.e., the model has stronger explanatory power). The Pearson correlation coefficient R_p is a statistic used to reflect the degree of linear correlation between two variables. The larger the absolute value of R_p, the stronger the correlation. This can help to quantitatively observe a linear correlation between the predicted o-cresol content value and the chemical reference value to confirm the reliability of the model. The mean relative prediction error (MRPE) of 0.01106 is very low for the PLSR and the prediction correlation coefficient R_p of 0.99161 is large; therefore, the PLSR model is of great prediction stability and accuracy.

FIGURE LEGENDS:

Figure 1. Original NIR spectrum taken with FT-NIR spectrometer. The figure shows the non-preprocessed spectral dataset collected over a period of time by the FT-NIR spectrometer.

Figure 2. Determination coefficient for the o-cresol concentration determined by PLSR. The figure shows the trend of the determination coefficient for the o-cresol concentration with increasing number of components.

Figure 3. Percent variance explained in the NIR spectral data for the PLSR. The figure shows the percent variance explained in the NIR spectral data for the principal components under the PLSR.

Figure 4. Residual of the PLSR for the test set. The figure shows the residuals of the 200 test set samples for the PLSR.

Figure 5. MSPECV of the PLSR. The precision of the PLSR model was evaluated with a 10-fold cross-validation; the MSPECV of model is shown in Figure 5.

Table 1. Evaluation indices of the model. Table 1 shows the model evaluation indicators for a 10-fold cross-validation.

DISCUSSION:

This protocol describes the process of performing the PLSR on the measurement of the o-cresol concentration remaining in the liquid product of polyphenylene ether with NIRS.

The two critical steps in this process are the pre-processing of the original NIR spectral data and the variables selection of the high-dimensional NIR spectral data.

Generally, the non-systematic background interference leads to the non-systematic scattering deviation or baseline drift of NIR spectrum. Proper NIR spectral pre-processing method (derivative, normalization, multivariate scattering correction, etc.), as the first critical step, eliminates background interference and increases the signal-to-noise ratio of data.

Multiple linear correlations existing in the high-dimensional NIR spectral data introduces unnecessary computation and reduces the operation efficiency of the regression model. Moreover, there are redundant variables that have weak or even no correlation to the o-cresol content in liquid polyphenyl ether products. The variable selection based on PLS, as the second critical step, solves the above two problems, which reduces the complexity of the model and enhances its robustness compared with the direct use of OLSR for modeling.

In addition, the technology has its limitations as following. Firstly, it requires that there is no significant nonlinear correlation between the various dimensional variables of the NIR spectral data. As the PLS algorithm mechanism is based on the expectation of maximizing the variance of NIR spectral data variables and maximizing the Pearson correlation between the selected principal component and the content of o-cresol. Therefore, the operation of the algorithm on data is essentially a linear combination of high-dimensional variables, which does not include the possible nonlinear relationship between data[15]. Secondly, due to the offline modeling, it cannot be self-updated. Our work has not yet covered the effects of interaction between o-cresol and solvent molecules or other impurity molecules on NIRS. The interaction between o-cresol and solvents or other impurity molecules may have an unexpected effect on NIRS, but we still have not found an appropriate method to quantify this effect. Thirdly, the change of environment, working condition and other factors may lead to the concept drift of the model.

However, for the existing methods such as physical or chemical separation analysis for the measurement of chemical substance content in the chemical material industry, NIRS detects composition and structural information about chemical substance content from a microscopic molecular perspective. In general, the qualitative transition at the macroscopic level is caused by the quantitative change at the microscopic level. Therefore, this technology has stronger accuracy, timeliness and sensitivity than other methods. Moreover, the combination with data analysis technology can effectively utilize industrial big data, which is conducive to industrial automation and even intelligent construction.

In the future, this technology can be flexibly applied to online measurement of various liquid or solid phase chemicals content in other fine process industries where the measurement of general sensors and laboratory measurement is difficult or unable to meet the industrial requirements[16]. Moreover, this technology can be widely combined with advanced control technology due to the real-time feedback of analysis variables and thus meet higher industrial requirements[17-19].

ACKNOWLEDGMENTS:

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61722306 and 61473137) and National First-class Discipline Program of Light Industry Technology and Engineering (LITE2018-025).

DISCLOSURE:

The authors have nothing to disclose.

REFERENCE:

1. Nicolai, B.M., et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*. **46** (2), 99-118, (2007).
2. Chang, C.W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Science Society of America Journal*. **65** (2), 480-490, (2001).
3. Chen, Y., et al. Near-infrared spectroscopy for rapid evaluation of different processing products of Sophora japonica L. *Spectroscopy Letters*. **51**(1), 37-44, (2018).
4. Cayuela, J.A., Garcia, J.F. Nondestructive measurement of squalene in olive oil by near infrared spectroscopy. *LWT-FOOD SCIENCE AND TECHNOLOGY*. **88**, 103-108, (2018).
5. Joaquim, M. ''Polyphenyl Ether Lubricants'' Synthetic Lubricants and High-performance Functional Fluids'', R. L. Rudnick and R. L. Shubkin, Eds., NY, Marcel Dekker, Inc., 239, (1999).
6. Grabowska, H., Kaczmarczyk, W., Wrzyszcz, J. Synthesis of 2,6-Xylenol by Alkylation of Phenol with Methanol. *Applied Catalysis*. **47**(2), 351-355, (1989).
7. Jeon, D.B., et al. Determination of volatile organic compounds, catechins, caffeine and theanine in Jukro tea at three growth stages by chromatographic and spectrometric methods. *FOOD CHEMISTRY*. **219**, 443-452, (2016).
8. Davidyuk, E.I., Demchenko, V.F., Klisenko, M.A. Rapid group separation and identification of chlorinated organic compounds by high performance liquid chromatography. *JOURNAL OF ANALYTICAL CHEMISTRY*. **52**(11), 1058-1065, (1997).
9. Rinnan, A., Berg, F., Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*. **28**(10), 1201-1222, (2009).
10. Zou, X.B., Zhao, J.W., Povey, M.J.W., Holmes, M., Mao, H.P. Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*. **667**(1-2), 14-32, (2010).
11. Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S. The potential of near-infrared reflectance spectroscopy for soil analysis - a case study from the Riverine Plain of south-eastern Australia. *Australian Journal of Experimental Agriculture*. **42**(5), 607-614, (2002).
12. Wang, C.K., Zhang, T.L., Pan, X.Z. Potential of visible and near-infrared reflectance spectroscopy for the determination of rare earth elements in soil. *Geoderma*. **306**, 120-126, (2017).
13. Gatius, F., Miralbes, C., David, C., Puy, J. Comparison of CCA and PLS to explore and model

NIR data. *Chemometrics and Intelligent Laboratory Systems*. **164**, 76-82, (2017).

14. Wold, S., Sjostrom, M., Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics & Intelligent Laboratory*. **58**(2), 109-130, (2001).

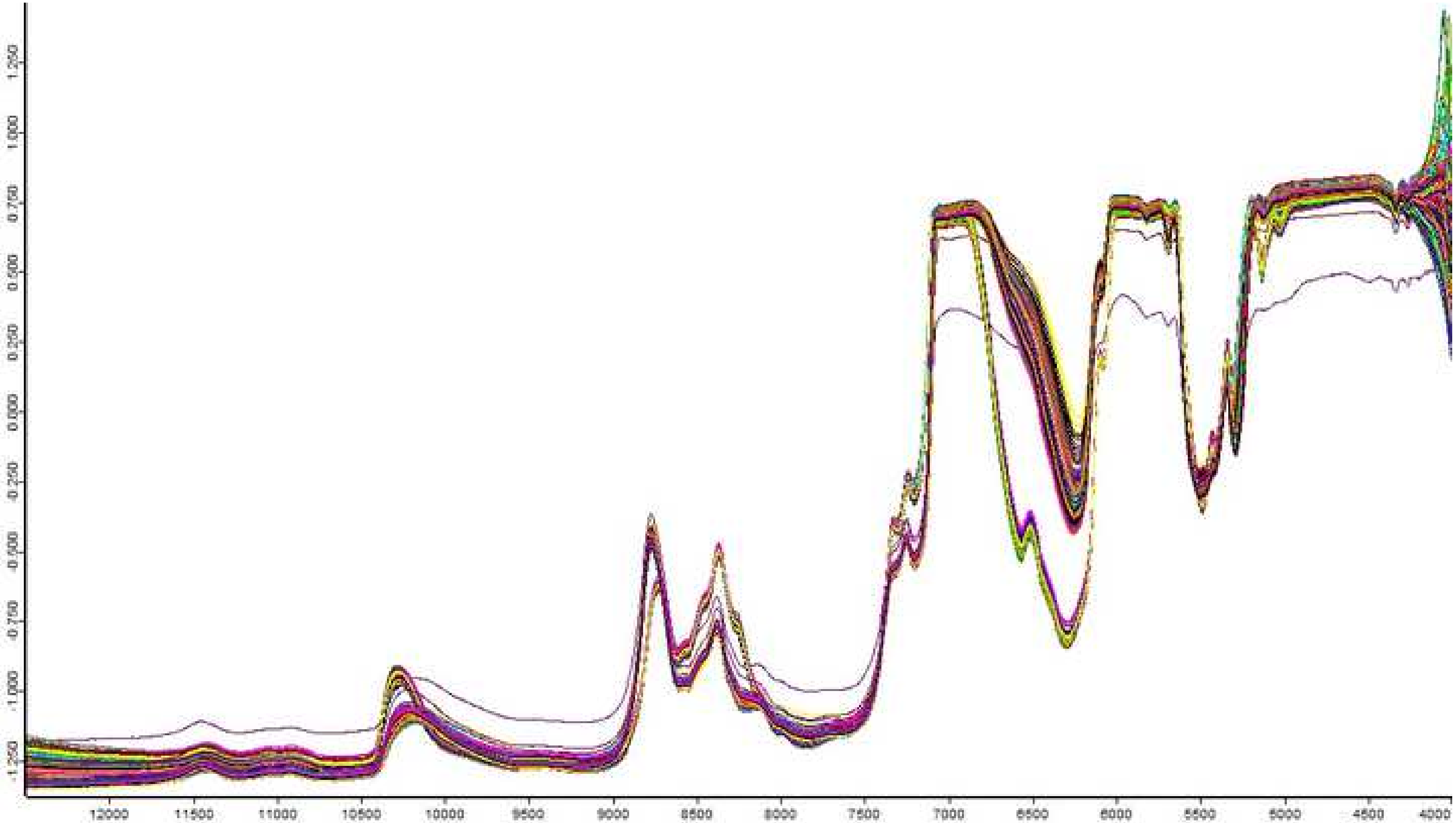
15. Douglas, R.K., Nawar, S., Alamar, M.C., Mouazen, A.M., Coulon, F. Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques. *SCIENCE OF THE TOTAL ENVIRONMENT*. **616**, 147-155, (2017).

16. Grassi, S., Alamprese, C. Advances in NIR spectroscopy applied to process analytical technology in food industries. *CURRENT OPINION IN FOOD SCIENCE*. **22**(SI), 17-21, (2018).

17. Trung, T., Downes, G., Meder, R., Allison, B. Pulp mill and chemical recovery control with advanced analysers - from trees to final product. *APPITA*. **68**(1), 39-46, (2015).

18. Vann, L., Sheppard, J. Use of near-infrared spectroscopy (NIRs) in the biopharmaceutical industry for real-time determination of critical process parameters and integration of advanced feedback control strategies using MIDUS control. *Journal of Industrial Microbiology& Biotechnology*. **44**(12), 1589-1603, (2017).

19. Modrono, S., Soldado, A., Martinez-Fernandez, A., de la Roza-Delgado, B. Handheld NIRS sensors for routine compound feed quality control: Real time analysis and field monitoring. *TALANTA*. **162**, 597-603, (2017).


1.250
1.000
0.750
0.500
0.250
0.000
-0.250
-0.500
-0.750
-1.000
-1.250
12000
11500
11000
10500
10000
9500
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000

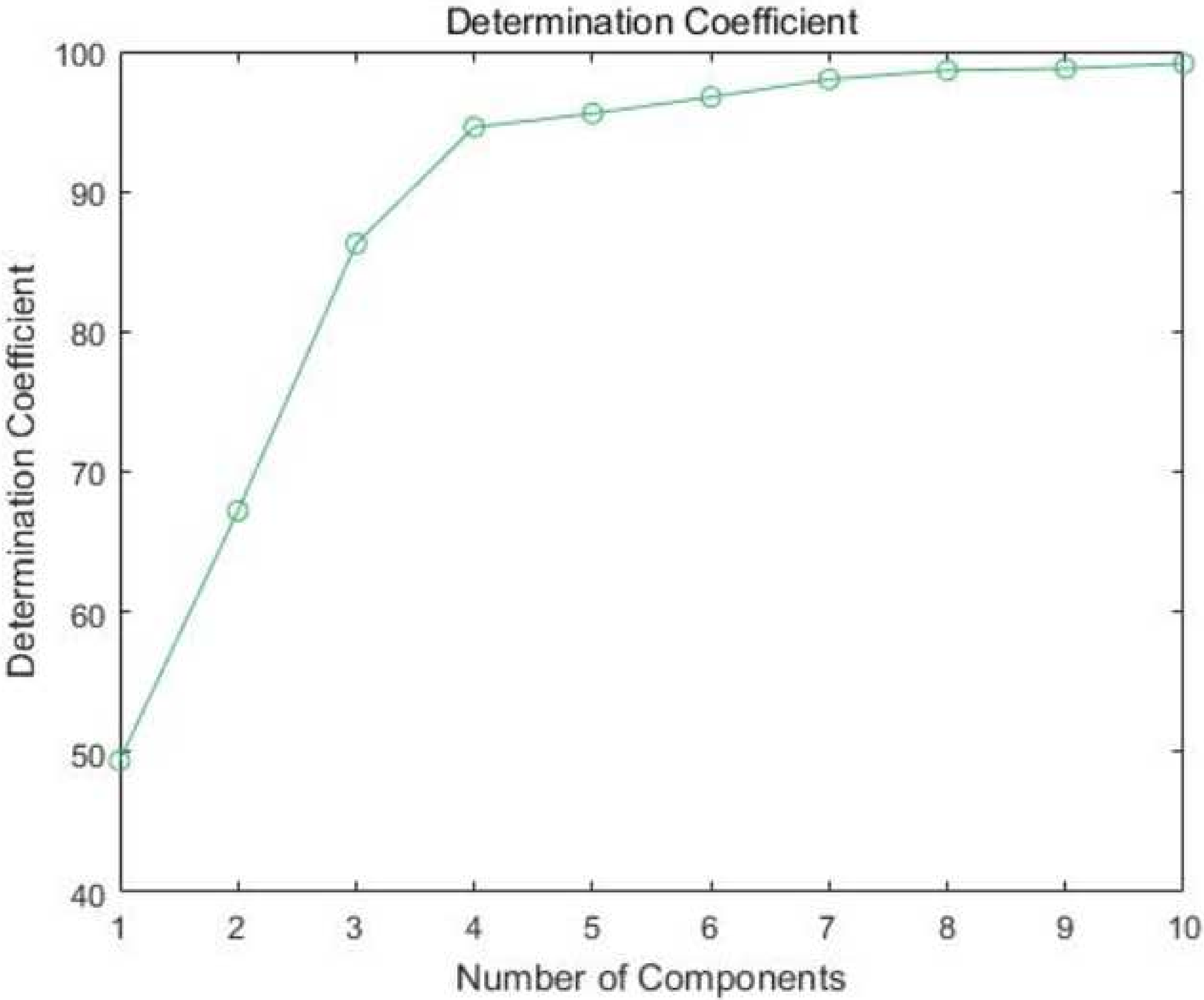
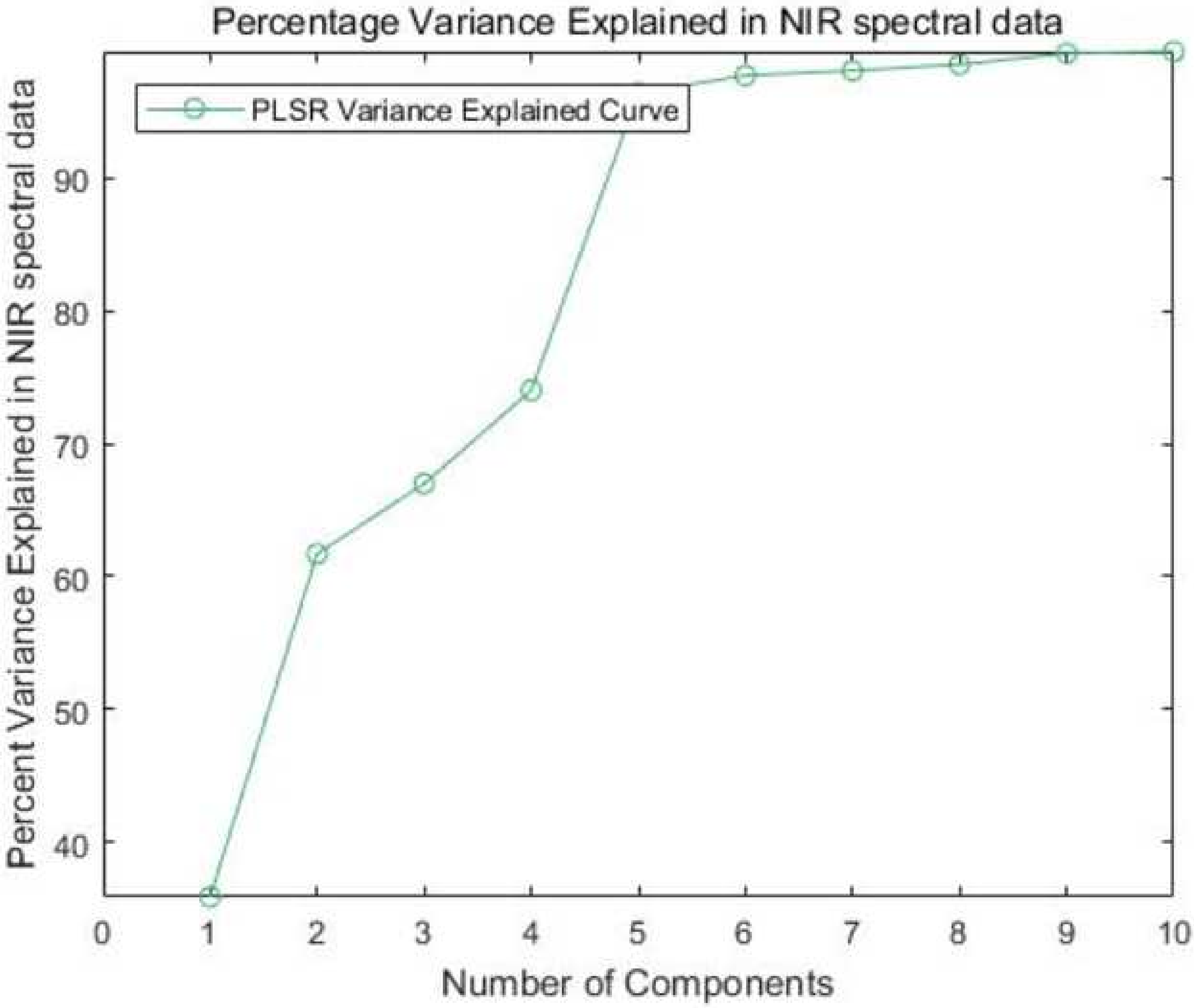
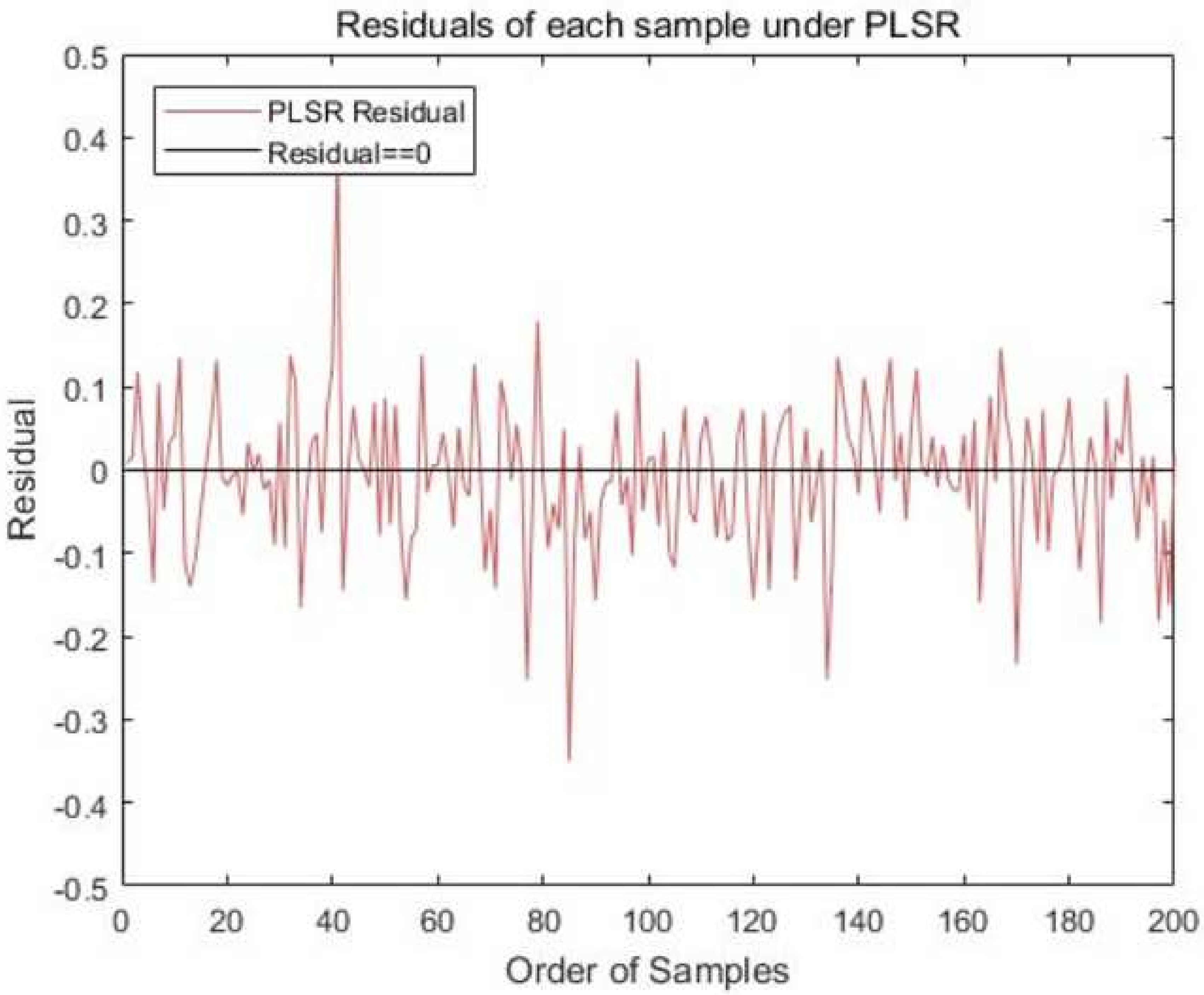

Determination Coefficient
Determination Coefficient
Number of Components
100
90
80
70
60
50
40
1
2
3
4
5
6
7
8
9
10

Figure 3


Percentage Variance Explained in NIR spectral data
PLSR Variance Explained Curve
Percent Variance Explained in NIR spectral data
40
50
60
70
80
90
0
1
2
3
4
5
6
7
8
9
10
Number of Components


Residuals of each sample under PLSR
PLSR Residual
Residual==0
Residual
Order of Samples
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
0
20
40
60
80
100
120
140
160
180
200

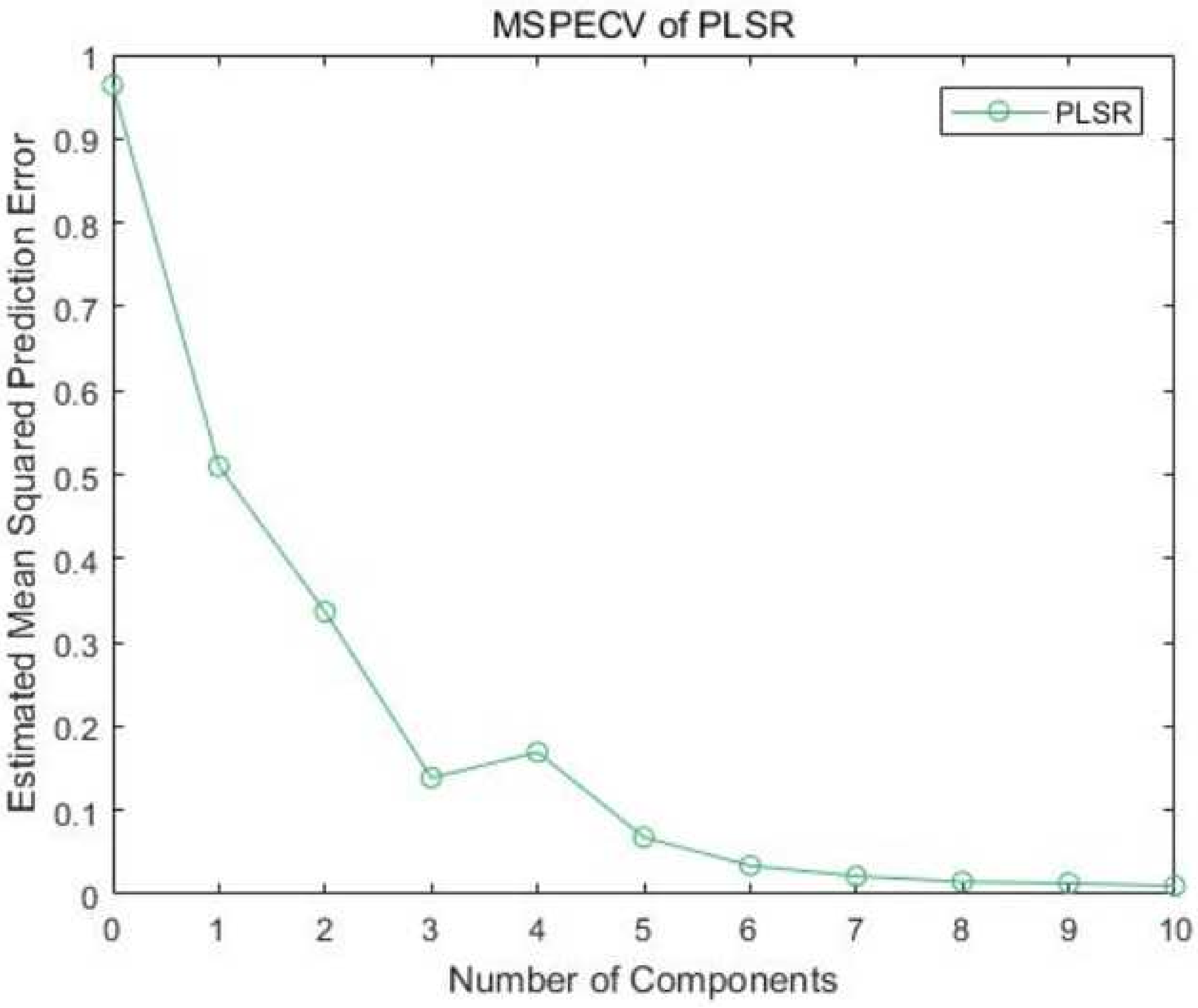

MSPECV of PLSR
PLSR
Estimated Mean Squared Prediction Error
Number of Components
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
1
2
3
4
5
6
7
8
9
10

Figure 2 of SVG

Figure 3 of SVG

Click here to access/download

Video or Animated Figure

Percentage Variance Explained in NIR spectral data.svg

Figure 4 of SVG

Click here to access/download

Video or Animated Figure

Residuals of each sample under PLSR.svg

Figure 5 of SVG

Click here to access/download

Video or Animated Figure

MSPECV of PLSR.svg

Table 1

	Interpretability	Prediction accuracy	
	R^2	R_p	MRPECV
PLSR	0.98332	0.99161	0.01106

Name	Company	Number
MPA II Multi Purpose FT-NIR Analyzer	Bruker	1
Fiber Optic Probes(Liquid phase)	Bruker	1
Liquid chromatography analyzer	/	1
Laboratory Equipment and Supplies(e.g. test tube, etc.)	/	
MATLAB	MathWork	1
OPUS	Bruker	1
Principal computer	DELL	1
The Unscrambler	CAMO	1

JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: Prediction of O-cresol Concentration based on Near Infrared Spectroscopy via Partial Least Square Regression

Author(s): Zihao Chen, Niannian Zheng, Xiaoli Luan*, Fei Liu

Item 1 (check one box): The Author elects to have the Materials be made available (as described at http://www.jove.com/author) via: ☑ Standard Access ☐ Open Access

Item 2 (check one box):

☑ The Author is NOT a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments;* "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Protection of the Work. The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/ or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. Indemnification. The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in *JoVE* or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. Fees. To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. Transfer, Governing Law. This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name:	Xiaoli Luan
Department:	Institute of Automation
Institution:	Jiangnan University
Article Title:	Prediction of O-cresol Concentration based on Near Infrared Spectroscopy via Partial Least Square Regression

Signature: Xiaoli Luan Date: 2018.09.09

Please submit a signed and dated copy of this license by one of the following three methods:

1) Upload a scanned copy of the document as a pfd on the JoVE submission site;
2) Fax the document to +1.866.381.2236;
3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

Authors' response to the reviews of "JoVE59077R2"

Zihao Chen, Niannian Zheng, Xiaoli Luan*, Fei Liu

First of all, we would like to express our sincere appreciation to the editor and reviewers for their insightful comments and suggestions again, many of which have been incorporated into the revised manuscript such that the revised version of our paper is more interesting and informative. Our response to the reviewers' comments (in the manner of one by one and 1:1 correspondence) is given as follows:

Editorial comments:

Changes to be made by the author(s) regarding the manuscript:

1. Step 1.9: Please ensure that all text is written in the imperative tense.

   Response 1:

   Thanks. We have modified this step into the following imperative tense:

   1.9 Combine the NIR spectra at the given test times with the chemical reference values of the o-cresol concentration.

2. 3.2.2, 3.2.4, 3.3.4: Is a software used in this step. If yes, please mention what button is clicked on in the software, or which menu items need to be selected to perform the step. If no, please add more details to your protocol steps. Please ensure you answer the "how" question, i.e., how is the step performed?

   Response 1:

   Thanks. Starting from 3.2, our work has entered the part of building a regression model based on pre-processed near-infrared spectral data using the Matlab program. Matlab, as a matrix computing software we use, focuses on the simulation programs we actually use. Its software operation simply involves the import of .mat files with near-infrared spectral data and the opening of the simulation program, which we have given in the protocol section below:

3.2.1 Open Matlab and import the .mat file containing the preprocessed near-infrared spectral data into the workspace by dragging the .mat file to the workspace.

Note：The .mat file stores the near-infrared spectral data X as an independent variable and the o-cresol content of the product as a dependent variable in the form of two matrices.

3.2.2 Open the programmed .m file in the Editor. Click Open under the Editor option, select the compiled .m file in the file storage directory, and then click Confirm.

3.2.3 Extract 15 principal components according to the optimization objective of Equation 1 and the OLSR model between the extracted principal components and the predicted values of the o-cresol concentration with the program containing the command plsregress() in Matlab.

Note:[XL,YL,XS,YS ,BETA,PCTVAR,MSE] = plsregress(X,Y,ncomp)
This command returns a 2-by-ncomp matrix PCTVAR containing the percentage of variance

explained by the model. The first row of PCTVAR contains the percentage of variance explained in X by each PLS component, and the second row contains the percentage of variance explained in Y. Also, it returns a 2-by-(ncomp+1) matrix MSE containing estimated mean-squared errors for PLS models with 0:ncomp components. The first row of MSE contains mean-squared errors for the predictor variables in X, and the second row contains mean-squared errors for the response variable(s) in Y.

In order not to destroy the regularity of the contents of the protocol, the focused simulation program is given in the form of a note of its core command plsregression(). Some of the more lengthy visualization programs are not appearing in the protocol.

3. 3.2.3: Please add more details to your protocol steps. Please ensure you answer the “how” question, i.e., how is the step performed? Alternatively, add references to published material specifying how to perform the protocol action.

   Response 1:

   Thanks. As mentioned above, the specific algorithm implementation involves Matlab code and compilation. We have provided the core command plsregress() for 3.2.3 and subsequent steps and the key data that can be fed back after execution. We believe that the presentation of a large number of Matlab simulation code in the protocol section is not suitable for the overall format. In particular, most of the code is used to visualize data.

4. Lines 363-410: These are figure and table legends. For representative results, please include at least one paragraph of text to explain the Representative Results in the context of the technique you have described, e.g., how do these results show the technique, suggestions about how to analyze the outcome, etc.

   Response 1:

   Thanks. We have modified the Representative Results and add some notes in the Protocol of the revision to explain the Representative Results in the context of the technique you have described, e.g., how do these results show the technique, suggestions about how to analyze the outcome, etc. Please read the 361-454 lines of revision.