Rapid Determination of the Total Phosphorus and the Nitrate Nitrogen in Denitrifying Phosphorus Removal with iPLS and Near Infrared Spectroscopy

Jian Huang1,2, Xuan Jia1,2, Hua Zhang1,2*, Shanshan Xi1,2, Jun Liu3, Tao Luo1,2, Hao Chen1,2

1Anhui Provincial Key Laboratory of Environmental Pollution Control and Resource Reuse, Hefei, 230601, P.R. China
2School of Environmental and Energy Engineering, Anhui Jianzhu University, Hefei, 230601, P.R. China
3Anhui Zhonghuan Environmental Protection Technology Co., Ltd, Hefei, 230601, P.R. China

Received: 26 August 2020
Accepted: 12 November 2020

Abstract

Denitrifying phosphorus removal was realized in sequencing batch reactor with alternate anaerobic-anoxic procedures. Quantitative analysis models of the total phosphorus and nitrate nitrogen were established with near infrared spectroscopy combined with interval partial least-squares (iPLS). Second derivative, multiple scatter correction and wavelet denoising were respectively used to pretreat the raw near infrared spectral data. The results show that the iPLS model established by wavelet denoising method was the optimal one. As to the interval partial least-squares calibration models (iPLS models) of the total phosphorus and nitrate nitrogen, the results showed that the coefficients ($r_c$) of the correlation between the actual values and corrected values were respectively 0.9449 and 0.8681, with the root mean square errors of cross validation (RMSECV) being 0.0254 and 0.0315. In addition, the testing results of the interval partial least-squares calibration models of the total phosphorus and nitrate nitrogen indicated that the coefficients ($r_p$) of correlation between the actual values and predicted values were 0.9294 and 0.9014, with the root mean square errors of prediction (RMSEP) being 0.0235 and 0.0298. This study suggested that the iPLS models of the total phosphorus and nitrate nitrogen may provide a method for synchronous detection of the total phosphorus and nitrate nitrogen during denitrifying phosphorus removal.

Keywords: iPLS, near infrared spectroscopy, wavelet denoising, denitrifying phosphorus removal

*e-mail: zhanghuapaper@163.com
Introduction

The excess of nitrogen and phosphorus causes water eutrophication. Therefore, the nitrogen and phosphorus removal has become the focus of wastewater treatment. Denitrifying phosphorus removal is a sustainable biological nitrogen and phosphorus removal process [1, 2]. Under anaerobic condition, denitrifying phosphorus removal bacteria convert volatile fatty acids into poly-β-hydroxybutyrate and store them in their cells. Under anoxic condition, denitrifying phosphorus removal bacteria decompose intracellular poly-β-hydroxybutyrate to provide energy for denitrifying nitrogen removal and excessive phosphorus uptake [3, 4]. The content of nitrate nitrogen and phosphorus is the main factor affecting the denitrifying phosphorus removal. Therefore, it is important to quickly determine the concentration of the total phosphorus and the nitrate nitrogen in denitrifying phosphorus removal process [5]. However, the traditional chemical detection methods are rather time-consuming, usually costing considerable materials and causing secondary pollution. In addition, only one single indicator can be measured by a traditional detection instrument. Hence, it is particularly significant to develop a new method for synchronous and rapid determination of nitrate nitrogen and phosphorus, which however is rarely reported in literature [6].

Compared with the traditional chemical methods, the quantitative analysis model established by the combination of chemometrics and near infrared spectroscopy is capable of synchronous and rapid detection of multi-components. Moreover, the method has the advantages of low cost and no secondary pollution. Hence, in recent years, near infrared spectroscopy combined with chemometrics has attracted much attention [7-9]. The near infrared spectrometer is often interfered by various external factors when spectral data are collected. The main interfering factor is the background noise of instrument [10]. Therefore, in order to eliminate the noise and obtain valid information, the raw near infrared spectral data should be pretreated so as to improve the accuracy of the model. At present, the pretreatment methods commonly include second derivative [11, 12], multiple scattering correction [13, 14], and wavelet denoising [15-17]. All the pretreatment methods can reduce the interference of noise with the spectral data. However, different pretreatment methods may have different effects [18, 19].

The quantitative analysis model can be established via the pretreated spectral data. Partial least squares (PLS) is a chemometrics method which can establish quantitative analysis model in the full spectrum range [20-23]. Interval partial least squares (iPLS) is a new chemometric method developed on the basis of PLS [24, 25], dividing the full spectrum range into several intervals with equal width. Then PLS regression model is established in the full spectrum range and each interval, and each corrected root mean square errors (RMSECV) are calculated respectively [26-29]. The optimal interval which has the minimum RMSECV is selected for modeling. IPILS model is featured by simple operation and easy determination of optimal interval [30-32]. Therefore, IPILS model is believed to be more effective has better prediction effect than PLS model [33-36]. However, there have been few studies on rapid quantitative analysis of the total phosphorus and the nitrate nitrogen in wastewater treatment process. This motivated this study in which water samples from denitrifying phosphorus removal process are collected and scanned by near infrared spectroscopy to obtain the raw near infrared spectral data. The second derivative, multiple scattering correction and wavelet denoising are then applied to pretreat the spectral data. After that, an examination of the results of different pretreatment methods determined the optimal pretreatment method, which combined with iPLS, was used to establish the quantitative analysis models of the total phosphorus and the nitrate nitrogen in denitrification phosphorus removal process. It is hoped that in this way, simultaneous, rapid and pollution-free detection of the total phosphorus and nitrate nitrogen was realized in denitrifying phosphorus removal process. Our research may have technological significance for the rapid determination of the total phosphorus and the nitrate nitrogen in wastewater treatment.

Material And Methods

Experimental Reactor and Water Quality

The experimental reactor was made by our own. The reactor is cylindrical, with the volume of 15 L. The operation of the reactor at room temperature was controlled by micro-computer, under alternating anaerobic-anoxic conditions. The experimental raw water is taken from the influent water of wastewater treatment plant. The water quality is as follows: the total phosphorus is 7.5 mg/L, NH$_4^+$-N is 32 mg/L, COD is 180 mg/L and pH is 7.5. Potassium nitrate solution was added with a concentration of 360 mg/L.$^+$

Experimental Controls and Sample Collection

The operation cycle is as follows: 30 min for influent, 105 min for anaerobic stage, 150 min for anoxic stage, 25 min for precipitation, 5 min for drainage, 45 min for stabilization. The reactor was fed with wastewater by peristaltic pump. The volume of water inflow and the volume of drainage in each cycle are both 5.5 L. During the stable operation of the reactor, water sample was collected every 15 minutes. A total of 140 water samples were collected in 7 cycles. The water samples were centrifuged for 5 min at speed of 1200 r/min, and the supernatant was collected. Then the supernatant was filtered with 0.45 μm filter membrane. Samples were
Chemical Analysis and Spectral Data Collection

The total phosphorus was determined by ammonium molybdate spectrophotometry. Nitrate nitrogen was determined by ultraviolet spectrophotometry. Near infrared spectral data were collected with a transform near infrared spectrometer (Bruker, Germany). The spectral range is 4000-12500 cm\(^{-1}\), the scanning times are 32, and the resolution is 8 cm\(^{-1}\).

Establishment and Evaluation Index of the iPLS Model

The full spectrum ranges from 4000 to 12500 cm\(^{-1}\) with 2203 raw near infrared spectral data. The full spectrum is divided into 20 intervals. The PLS model was built for each interval. The interval with the minimum RMSECV is considered as the optimal interval for iPLS model [37]. The iPLS model is evaluated with the calibration correlation coefficient \(r_c\), predictive correlation coefficient \(r_p\), corrected root mean square error (RMSECV), and predicted root mean square error (RMSEP) [38]. The calculation formulas are as follows:

\[
\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{m} (x_i - x'_i)^2}{m-1}}
\]

\[
r_c = \sqrt{1 - \frac{\sum_{i=1}^{m} (x_i - x'_i)^2}{\sum_{i=1}^{m} (x_i - \bar{x})^2}}
\]

\[
\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - y'_i)^2}{n-1}}
\]

\[
r_p = \sqrt{\frac{\sum_{i=1}^{n} (y_i - y'_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]

In the formulas, \(m\) is the number of the calibration set samples; \(x_i\) are the truth values of the calibration set samples; \(x'_i\) are corrected values by iPLS model; \(\bar{x}\) is the average of the truth values of the calibration set samples; \(n\) is the number of the test set samples; \(y_i\) is the truth values of the test set samples; \(y'_i\) is predicted values by iPLS model; \(\bar{y}\) is the average of the truth values of the test set samples. The correlation coefficient is an index evaluating the correlation between the two sets of data. The root mean square error is an index evaluating the degree of deviation of the corrected values (or the predicted value) from the truth values.

Test and Evaluation Index of the Ipls Model

External validation method was applied to test the iPLS model. 40 water samples from 2 cycles that are not used in modeling were centrifuged and filtered with 0.45 um filter. The spectral data of these 40 samples are collected to test the iPLS model whose effectiveness is evaluated by the predictive correlation coefficient \(r_p\) and the predictive root mean square error (RMSEP).

Results and Analysis

Change of Water Quality During Stable Period

Fig. 1 shows the water quality during the stable cycles of the reactor. As can be seen in Fig. 1, the total phosphorus is 7.9 mg /L and COD is 180 mg /L in raw water. After the anaerobic stage, the total phosphorus increased to 24.2 mg /L, and COD decreased to 22.3 mg /L. After anoxic stage, the total phosphorus decreased to 0.67 mg /L and nitrate nitrogen decreased from 18.1 mg/L to 0.41 mg/L. These results indicate that, in the anaerobic stage, denitrifying phosphorus removal bacteria decompose intracellular polyphosphate and release phosphate, and that in anoxic stage, denitrifying phosphorus removal bacteria use nitrate nitrogen as electron acceptor to excessively absorb phosphate, thus reducing the concentration of phosphate and nitrate nitrogen [39].

Collection of Raw Near Infrared Spectral Data

140 water samples were collected for modeling, and then were filtered with 0.45 \(\mu\)m filter membrane. These water samples were divided into two groups. 100 water samples were used for iPLS modeling, and 40 water samples was used to test the iPLS models. The spectral data are recorded at intervals of 3.66 cm\(^{-1}\). Each full spectrum has 2203 data. The raw near infrared spectra are shown in Fig. 2, which shows that the shapes of
near infrared spectra of different samples are similar, indicating similar water quality composition.

Pretreatment of Raw Near Infrared Spectral Data

The iPLS models established by different pretreatment methods have different degrees of accuracy. To find out the most accurate iPLS model established through the optimal pretreatment method, we compared the results of iPLS models established by three pretreatment methods, which are shown in Table 1 and Table 2.

From Table 1 and Table 2, it can be seen that the wavelet denoising, which has been widely used in signal processing, image processing and many nonlinear science fields, is the optimal pretreatment method for establishing iPLS models. The results indicates that the $r_c$ and RMSECV of the iPLS model of the total phosphorus are respectively 0.9449 and 2.5448, and the $r_c$ and RMSECV of the iPLS model of the nitrate nitrogen are respectively 0.8681 and 3.1475.

### Establishment of iPLS Calibration Model of the Total Phosphorus

The establishment of the iPLS model for the total phosphorus is shown in Figs 3-6. Fig. 3 gives the regression analysis of PLS in each interval. The RMSECV is taken as evaluation index, and the interval with the minimum RMSECV is selected as modeling interval. As can be seen in Figure 3, the minimum RMSECV is obtained when the principal component number is 10.

Fig. 4 shows the variation of the RMSECV in different intervals. In Fig. 4, the number in each bar represents the optimal number of principal components in each interval. The Height of the bar represents the corresponding RMSECV. The dotted line represents the minimum RMSECV in the full spectrum. The curve represents the raw spectra. As can be seen from Fig. 4, the RMSECV of the 17th interval is the minimum, which means that this interval contains the optimal spectral information. Therefore, the 17th interval (5272~5693 cm$^{-1}$) is selected as the optimal interval to establish iPLS model (Fig. 5). As shown in Fig. 6, the iPLS model of the total phosphorus is established in the optimal interval. The $r_c$ of the iPLS model for the total phosphorus is 0.9449 and the RMSECV is 0.0254.

### Establishment of iPLS Calibration Model of Nitrate Nitrogen

The establishment of the iPLS model of nitrate nitrogen is shown in Figs 7-10. Fig. 7 gives the regression analysis of PLS in each interval. The RMSECV is taken as evaluation index, and the interval with the minimum RMSECV is selected as modeling interval. As can be seen from Fig. 3, the minimum RMSECV is obtained when the principal component number is 11.

Fig. 8 shows the variation of the RMSECV in different intervals. In Fig. 8, the number in each bar

---

**Table 1.** Three pretreatment methods of iPLS model for the total phosphorus.

<table>
<thead>
<tr>
<th>Pretreatment method</th>
<th>Number of principal component</th>
<th>Optimal interval</th>
<th>$r_c$</th>
<th>RMSECV (%)</th>
<th>$r_p$</th>
<th>RMSEP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second derivatives</td>
<td>8</td>
<td>17</td>
<td>0.8459</td>
<td>2.7315</td>
<td>0.8192</td>
<td>2.9633</td>
</tr>
<tr>
<td>Multiple scatter correction</td>
<td>8</td>
<td>18</td>
<td>0.9011</td>
<td>2.6031</td>
<td>0.8837</td>
<td>2.5753</td>
</tr>
<tr>
<td>Wavelet denoising</td>
<td>10</td>
<td>17</td>
<td>0.9449</td>
<td>2.5448</td>
<td>0.9294</td>
<td>2.3459</td>
</tr>
</tbody>
</table>

**Table 2.** Three pretreatment methods of iPLS model for nitrate nitrogen.

<table>
<thead>
<tr>
<th>Pretreatment method</th>
<th>Number of principal component</th>
<th>Optimal interval</th>
<th>$r_c$</th>
<th>RMSECV (%)</th>
<th>$r_p$</th>
<th>RMSEP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second derivatives</td>
<td>17</td>
<td>19</td>
<td>0.8233</td>
<td>3.5529</td>
<td>0.8012</td>
<td>3.7617</td>
</tr>
<tr>
<td>Multiple scatter correction</td>
<td>19</td>
<td>18</td>
<td>0.8431</td>
<td>3.3084</td>
<td>0.8301</td>
<td>3.4157</td>
</tr>
<tr>
<td>Wavelet denoising</td>
<td>21</td>
<td>18</td>
<td>0.8681</td>
<td>3.1475</td>
<td>0.9014</td>
<td>2.9794</td>
</tr>
</tbody>
</table>
represents the optimal number of principal components in each interval. The height of the bar represents the corresponding RMSECV. The dotted line represents the minimum RMSECV in the full spectrum. The curve represents the raw spectra. As can be seen from Fig. 8, the RMSECV of the 1st interval is the minimum interval, which means that this interval contains the optimal spectral information. Therefore, the 1st interval (4000-4423 cm⁻¹) is selected as the optimal interval to establish iPLS model (Fig. 9). As shown in Fig. 10, the iPLS model of the total phosphorus is established in the optimal interval. The $r_c$ of the iPLS model for the total phosphorus is 0.8681 and the RMSECV is 0.0315.

Test of iPLS Models of the Total Phosphorus and Nitrate Nitrogen

The iPLS models were tested with the spectral data of the 40 samples that were not used in modeling. The actual values of the total phosphorus and the nitrate...
nitrogen are determined through traditional chemical methods. The results of the test for the iPLS model of the total phosphorus are shown in Fig. 11, which indicates that the correlation between the predicted value and the actual value is high: the $r_p$ is 0.9294 and the RMSEP is 0.0235. These results show that the iPLS model of the total phosphorus is accurate.

The results of the test for the iPLS model of nitrate nitrogen are shown in Fig. 12 which indicates that the correlation between actual values and the predicted values is high: the $r_p$ is 0.9014 and the RMSEP is 0.0298. These results show that the iPLS model of nitrate nitrogen is accurate.

**Conclusions**

Sequencing batch reactor was used in our study to realize the denitrifying phosphorus removal. Combined...
with iPLS, near infrared spectroscopy was employed to analyze simultaneously the total phosphorus and the nitrate nitrogen. The second derivative, multiple scattering correction and wavelet denoising were used to pretreat the raw near infrared spectral data. The results show that the wavelet denoising is the best pretreatment method. Near infrared spectroscopy combined with iPLS method was used to establish iPLS models of the total phosphorus and the nitrate nitrogen. These models can realize the rapid analysis of the total phosphorus and the nitrate nitrogen. The $r_1$ and $r_2$ of the iPLS model of the total phosphorus were respectively 0.9449 and 0.9294, with the RMSECV and RMSEP being 0.0254 and 0.0235. The $r_1$ and $r_2$ of the iPLS model of nitrate nitrogen were respectively 0.9014 and 0.9014, with the RMSECV and RMSEP being 0.0254 and 0.0298. It is thus concluded that the near infrared spectroscopy combined with iPLS can provide a new approach to the detection of water quality indicators. In future research, it is suggested that this method should be used in the rapid detection of other indicators in water.

Acknowledgements

Acknowledgements: This research was supported by Anhui Province Natural Science Foundation (1908085ME142, 1808085MB34), the Natural Science Project for Colleges of Anhui Province (KJ2019ZD52, KJ2020A0466), the Project of National key research and development program (2019YFC0408504).

Conflict of Interest

The authors declare no conflict of interest.

Reference

20. HOSEINPOUR S., AGHBASHLO M., TABATABAEI M., KHALIFE E. Exact estimation of biodiesel cetane number (CN) from its fatty acid methyl esters (FAMEs) profile using partial least square (PLS) adapted by artificial neural network (ANN). Energy Conversion and Management. 124, 389, 2016.