Applying near Infrared Spectroscopy and iPLS to Quantitative Analysis of PHB, Poly-P, and GLY in Denitrifying Phosphorus Removal

Hua Zhang1,2, Changcheng Li1, Jian Huang1*, Xianhuai Huang1, Wenchao Zhang1, Xiong Zhang1, Zhaoliang Wu1, Xiaokun Yu1

1Key Laboratory of Anhui Province of Water Pollution Control and Wastewater Reuse, School of Environment and Energy Engineering, Anhui Jianzhu University, Hefei, China
2Civil Engineering, National University of Ireland, Galway, Ireland

Received: 18 September 2017
Accepted: 3 October 2017

Abstract

Near infrared spectroscopy and interval partial least-squares (iPLS) were applied to rapid quantitative analysis of the poly-β-hydroxybutyrate (PHB), polyphosphate (Poly-P), and glycogen (Gly) during denitrifying phosphorus removal. Wavelet denoising was used to pretreat the raw near infrared spectroscopy, and the quantitative analysis models (iPLS models) of PHB, Poly-P, and GLY were established with interval partial least-squares (iPLS). The iPLS was used to select the optimal spectral interval for modeling. The total phosphorus decreased from 7.9 mg/L to 0.67 mg/L during denitrifying phosphorus removal. The region from 4,320 to 4,640 cm⁻¹ was selected to establish the iPLS model of intracellular PHB. The region from 4,000 to 4,320 cm⁻¹ was selected to establish the iPLS model of intracellular Poly-P. Finally, the region from 5,103 to 5,379 cm⁻¹ was chosen to establish the iPLS model of intracellular GLY. Statistical tests of these iPLS models of PHB, Poly-P, and GLY show that the correlation coefficients ($r_c$) between the correction values and the chemical values are 0.9637, 0.9582, and 0.9437, with the root mean square error of cross validation (RMSECV) being 0.0069, 0.0039, and 0.0025. Test results of iPLS models show that the correlation coefficients ($r_p$) between the prediction value (by iPLS model) and the chemical value were 0.9430, 0.9389, and 0.9133, with the root mean square error of prediction (RMSEP) being 0.0523, 0.0040, and 0.0058. These research results show that the proposed models may provide a rapid and effective quantitatively analysis of intracellular PHB, Poly-P, and GLY, and that the effect of the denitrifying phosphorus removal process can be quickly judged from the cell metabolism perspective.

Keywords: near infrared spectroscopy, iPLS, denitrifying phosphorus removal, PHB, Poly-P, GLY
Introduction

Phosphorus is the main factor causing water eutrophication [1], and denitrifying phosphorus removal is a sustainable biological nitrogen and phosphorus removal process [2-3]. The metabolism of microorganisms plays an important role in denitrifying phosphorus removal [4]. Under anaerobic conditions, denitrifying phosphorus removal bacteria absorb volatile fatty acids and synthesize intracellular poly-β-hydroxybutyrate (PHB) using the energy of the degradation of intracellular polyphosphate (Poly-P) and glycogen (GLY), and then release PO₄³⁻ into the outside [5-7]. However, under anoxic conditions, denitrifying phosphorus removal bacteria decompose stored PHB as an energy source to excessively absorb PO₄³⁻ and cell-growth [8-11]. Therefore, the metabolism of PHB, Poly-P, and Gly are considered to determine the efficiency of denitrifying phosphorus removal [12]. So the rapid measurements of the content of intracellular PHB, Poly-P, and GLY are important for regulating the denitrifying phosphorus removal process. The traditional methods to analyze intracellular PHB, Poly-P, and GLY are complicated, costing much time and resources, and often causing secondary pollution [13]. Therefore, it is of much significance to develop a rapid and pollution-free method to analyze intracellular PHB, Poly-P, and GLY.

Near infrared spectroscopy is a fast and green analytical technique. With the characteristics of simple operation, low cost, rapid analysis, less consumption of chemicals, and simultaneous measurements of multiple components [14-19], it has drawn much attention in recent years and has been widely used in food, medicine, the environment, etc. [20-28]. However, there is little research on its application for measuring intracellular PHB, Poly-P, and GLY in the denitrifying phosphorus removal process. It is often inevitable that near infrared spectra will be affected by the uneven particles distribution of samples and background noise of near infrared spectroscopy instruments. Therefore, raw near infrared spectra data should be preprocessed to eliminate such interference [29]. Wavelet denoising is an effective preprocessing method that can effectively eliminate scattering and noise [30-31]. Interval partial least squares (iPLS) is a new chemometric method developed on the basis of PLS. Compared with PLS, iPLS has better performance in quantitative analysis [32]. With the iPLS method, the full-spectra are divided into several regions and then the optimal intervals are selected to establish the quantitative analysis model (iPLS model) [33-34]. The near infrared spectroscopy combined with iPLS algorithm has achieved good results in rapid quantitative analysis of polyphenols in green tea and inorganic nitrogen in water and in the rapid diagnosis of some diseases [19].

However, as to near infrared spectroscopy combined with iPLS algorithm, there is little study on the analysis of intracellular PHB, Poly-P, and GLY in the denitrifying phosphorus removal process. Therefore, in this study wavelet denoising was used to preprocess raw near infrared spectra of samples in the denitrifying phosphorus removal process. After pretreatment, iPLS was used to select the optimum spectral interval and establish quantitative analysis models of intracellular PHB, Poly-P, and GLY. It is hoped that this study may provide a rapid and non-polluting method for the analysis of intracellular PHB, Poly-P, and GLY in the denitrifying phosphorus removal process.

Material and Methods

Experimental Equipment and Water Quality

A sequencing batch reactor was used in this experiment (Fig. 1). Denitrifying phosphorus removal was achieved by alternate anaerobic and anoxic condition. The volume of the reactor was 15 L. The reaction temperature was maintained as room temperature. The operational cycle was 360 min. The influent time was 30 min, anaerobic stage was 105 min, anoxic stage was 150 min, precipitation time was 25 min, drainage was 5 min, and stabilization time was 45 min. The raw water quality in this experiment was as follows: 180 mg/L for COD, 7.5 mg/L for TP, 32 mg/L for NH₄⁺-N, and 7.5 for pH.

Sample Collection and Analysis Method

When the denitrifying phosphorus removal process was in stable operation, sludge samples were collected every 15 min, and 20 samples were collected in a cycle. Finally, 140 samples were collected in 7 cycles. The 100 samples were used to establish iPLS models, and the remaining 40 samples were used to test iPLS models. All the samples were centrifuged for 5 min, and then the sludge samples were dried for 20 h. Every sample was divided into two sub-samples. One was measured with chemical method to get the chemical value of intracellular PHB, Poly-P, and GLY; the other was measured with near infrared spectrometer to collect spectral data. The iPLS method was employed to establish the iPLS models. Gas chromatography was used for measuring PHB. Molybdenum antimony spectrophotometric method was used for measuring Poly-P. The anthrone method was used for measuring GLY. Near infrared spectrometer was used for measuring spectral data.

Establishing and Evaluating the iPLS Model

The establishment of iPLS models includes the following steps. First, raw near infrared spectra were pretreated with wavelet denoising. Second, the full spectrum was divided into several intervals. Then independent PLS regression analysis was conducted in each subinterval, and the interval with the minimum RMSECV was the optimal modeling interval. After that, the iPLS model established in the optimal interval was...
Applying near Infrared Spectroscopy...

the optimal model, and the accuracy of the iPLS model was evaluated with the root mean square error of cross validation (RMSECV) and the correlation coefficient ($r_c$). Finally, the established models were tested, and the performance of the models was evaluated with the root mean square error of prediction (RMSEP) and the correlation coefficient ($r_p$). These evaluation parameters are calculated with the following formula [35-36]:

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

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

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - y'_i)^2}{n-1}}$$  \hspace{1cm} (3)

$$r_p = \sqrt{1 - \frac{\sum_{i=1}^{n} (y_i - y'_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$  \hspace{1cm} (4)

where $m$ and $n$ are the numbers of the modeling sample and test sample, $x_i$ and $y_i$ are the chemical values of the modeling sample and the test sample, $x'_i$ is corrected value by iPLS model, $y'_i$ is predicted value by iPLS model, and $\bar{x}$ and $\bar{y}$ are the average of chemical values of the modeling sample and the test sample.

**Results and Discussion**

**Change of Intracellular Substances of Denitrifying Phosphorus Removal Process**

Fig. 2 shows the changes of intracellular substances during the denitrifying phosphorus removal process. As we can see in Fig. 2, intracellular Poly-P and GLY were decomposed to produce energy to synthesize PHB in the anaerobic stage. Therefore, the content of Poly-P...
decreased from 5.31% to 1.19%, and the content of GLY decreased from 7.81% to 6.31%. However, stored PHB is decomposed during the anoxic stage, producing energy to absorb phosphate from sewage and to synthesize Poly-P and GLY. The content of PHB decreased from 12.41% to 3.11%, the content of Poly-P increased from 1.19% to 6.01%, and the content of GLY increased from 6.31% to 8.81%. These data show that the denitrifying phosphorus removal process has been stable.

Pretreatment of Raw Near Infrared Spectra

The raw near infrared spectra of the samples are shown in Fig. 3. The X-axis is the wave number, which ranges from 4,000 to 12,000 cm$^{-1}$, and the Y-axis is the absorbance of near infrared spectra. Fig. 3 shows that there are similar shape and change in the raw near infrared spectra of all the samples. However, the raw near infrared spectra are also affected by the uneven distribution of particles and instrument noise. Therefore, there are scattering and noise signals in raw near infrared spectra.

Fig. 3. Raw near infrared spectra by near infrared spectrometer.

Fig. 4. Pretreated spectra of sludge samples; wavelet denoising was used to pretreat the spectra.

Fig. 5. Establishing the iPLS model of PHB: a) change of RMSECV in full spectrum, b) change of RMSECV in each interval, c) selected wave number, and d) iPLS model of Poly-P; linear correlation between the corrected value and the truth value of PHB.
Applying near Infrared Spectroscopy...

Table 1. Selected wave band, optimal internal, principle component number, and evaluation parameters of PHB, Poly-P, and GLY.

<table>
<thead>
<tr>
<th>Intracellular polymer</th>
<th>Selected wave band (cm(^{-1}))</th>
<th>Optimal internal</th>
<th>Principle component number</th>
<th>(r_c)</th>
<th>RMSECV</th>
<th>(r_p)</th>
<th>RMSEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHB</td>
<td>4320-4640</td>
<td>24</td>
<td>10</td>
<td>0.9637</td>
<td>0.0069</td>
<td>0.9430</td>
<td>0.0523</td>
</tr>
<tr>
<td>Poly-P</td>
<td>4000-4320</td>
<td>25</td>
<td>2</td>
<td>0.9582</td>
<td>0.0039</td>
<td>0.9389</td>
<td>0.0040</td>
</tr>
<tr>
<td>GLY</td>
<td>5103-5379</td>
<td>25</td>
<td>9</td>
<td>0.9437</td>
<td>0.0025</td>
<td>0.9133</td>
<td>0.0058</td>
</tr>
</tbody>
</table>

spectra, especially in the wave number of 4,000 to 5,000 cm\(^{-1}\), where a large number of burr peaks appear. These interferences will affect the true characteristic signal of the substances. So effective pretreatment is necessary to obtain desired spectral information and improve the accuracy of the quantitative analysis model. Fig. 4 shows near infrared spectra pretreated with wavelet denoising. After pretreatment, the interferences in the raw near infrared spectra were effectively reduced, and the near infrared spectra are smoother.

Establishing the iPLS Model

After pretreatment by wavelet denoising, the near infrared spectra (4,000-12,000 cm\(^{-1}\), 1,037 spectral data) were divided into several intervals by iPLS. Then the interval with the minimum RMSECV was selected as the optimal interval, and the iPLS model in the optimal interval was the best model.

Establishing the iPLS Model of Intracellular PHB

Establishing the iPLS model of intracellular PHB is shown in Figs 5(a-d) and Table 1. In the full-spectrum, the RMSECV changed with the principal component number (Fig. 5a), and the minimum RMSECV is 0.69% when the principal component number is 10. Then the full spectrum is divided into 25 intervals. For each interval, the optimal principal component number and the corresponding RMSECV are shown in Fig. 5b). The italic number in the bar column of each interval represents the optimum principal component number of the interval, and the height of each column presents the RMSECV of the corresponding interval. In addition, the dotted line represents the minimum RMSECV in full spectrum. As shown in Figs 5(b-c), the RMSECV of the 24th interval (4,320-4,640 cm\(^{-1}\)) is the minimum compared with full-spectrum, with RMSECV being 0.69%. Therefore, the 24th interval is the optimal interval for modeling, which has the characteristics of higher information content. In this interval, the iPLS model of intracellular PHB was considered the best model (Fig. 5d). In Fig. 5d), the X axis is the chemical value of intracellular PHB, and the Y axis is the corrected value of the iPLS model. The \(r_c\) between the chemical value and the correction value, is 0.9637, and the corresponding RMSECV is 0.0069. The result shows that the iPLS model of intracellular PHB has a good correction effect.

Establishing the iPLS Model of Intracellular Poly-P

Figs 6(a-d) show the establishment of an iPLS model of intracellular Poly-P, and Fig. 6a) shows the RMSECV changes with the number of principal components in the full-spectrum. In the full-spectrum, the minimum RMSECV is 0.39%, with the principal component number being 2. Fig. 6b) showed the optimal principal component number and the RMSECVs of 25 intervals. As can be seen from Figs 6(b-c), the RMSECV is the minimum in the 25th interval (4,000-4,320 cm\(^{-1}\)), with the RMSECV being 0.0039. Therefore, the 25th interval is considered the optimal interval, and the iPLS model in this interval was considered a good model for intracellular Poly-P (Fig. 6 d). In Fig. 6d), the X axis and Y axis respectively represent the chemical value and the corrected value. Between the chemical value and the correction value, \(r_c\) is 0.9582 and the corresponding RMSECV is 0.0039. The result shows that the iPLS model of intracellular Poly-P has a satisfying effect of correction.

Establishing the iPLS Model of Intracellular GLY

The establishment of the iPLS model of intracellular GLY are presented in Figs 7(a-d) and Table 1. Fig. 7a) shows how RMSECV changes with different numbers of principal components. The RMSECV is the minimum (0.25%) when the principal component number is 9. The full spectrum is divided into 29 subintervals (Fig. 7b). As can be seen from Figs 7(b-c), the RMSECV of the 25th interval (5,103-5,379 cm\(^{-1}\)) is the minimum 0.25%. Therefore, the 25th interval is the optimal interval. The established model is shown in Fig. 7d. The \(r_c\) between the chemical value and the correction value is 0.9437, with the RMSECV being 0.0025. These results indicate that the iPLS model of intracellular GLY has a perfect correction effect.

Tests of iPLS Models of Intracellular PHB, Poly-P, and GLY

The established iPLS models of intracellular PHB, Poly-P, and GLY were tested on the basis of the data of the 40 samples that were not used in modeling. The test results are shown in Fig. 8, where the X axis is the chemical value and the Y axis is the prediction value by iPLS models. The test results of iPLS model of PHB
Fig. 6. Establishing the iPLS model of Poly-P: a) change of RMSECV in full spectrum, b) change of RMSECV in each interval, c) selected wave number, and d) iPLS model of Poly-P; a linear correlation between the corrected value and the truth value of Poly-P.

Fig. 7. Establishing the iPLS model of GLY: a) change of RMSECV in full spectrum, b) change of RMSECV in each interval, c) selected wave number, and d) iPLS model of GLY; a linear correlation between the corrected value and the truth value of GLY.
Applying near Infrared Spectroscopy...

shows that, between the prediction value and the chemical value, \( r_p \) is 0.9430, and the root mean square error of prediction (RMSEP) is 0.0052. The test result of the iPLS model of Poly-P shows that, between the prediction value and the chemical value, \( r_p \) is 0.9389 and the root mean square error of prediction is 0.0040. The test results of iPLS model of GLY show that between the prediction value and the chemical value, \( r_p \) is 0.9133 and the root mean square error of prediction (RMSEP) is 0.0058. These results show that the iPLS models are capable of satisfactorily and promptly predicting the content of intracellular PHB, Poly-P, and GLY.

Conclusions

A fast and effective method for proving quantitative analysis for intracellular PHB, Poly-P, and GLY was developed based on near infrared spectroscopy. The noise of raw near infrared spectra may be reduced by wavelet denoising, and the effective spectral information was extracted. The established quantitative analysis models that make use of the iPLS algorithm are powerful analytical tools for intracellular PHB, Poly-P, and GLY. The feasibility of the iPLS models based on near-infrared spectroscopy was evaluated by root mean square error of cross validation and correlation coefficient. In addition, the performance of iPLS models was validated by test samples, thereby demonstrating their capability as simple, fast analytical tools for intracellular PHB, Poly-P, and GLY. These results demonstrated that the iPLS models of intracellular PHB, Poly-P, and GLY are rapid and reliable for detecting and quantifying the content of intracellular PHB, Poly-P, and GLY in denitrification phosphorus removal. In the future, the influence of different pretreatment methods or different modeling methods on the rapid analysis of intracellular substances should be researched further.

Acknowledgements

This research was supported by Anhui Province Natural Science Foundation (1808085MB34), the Research Projects of Visiting Scholars for Excellent Young and Backbone Talents both at Home and Abroad for Colleges of Anhui Province (gxfx2017054), Key Projects to Support the Talent for Colleges of Anhui Province (gxyqZD2017059), and the Major Science and Technology Program for Water Pollution Control and Treatment of China (2014ZX07405-003-03).

References


