Composition analysis of colored dissolved organic matter in Taihu Lake based on three dimension excitation-emission fluorescence matrix and PARAFAC model, and the potential application in water quality monitoring

WANG Zhi-gang1,2, LIU Wen-qing2,∗, ZHAO Nan-jing2, LI Hong-bin2, ZHANG Yu-jun2, SI-MA Wei-cang2, LIU Jian-guo2

1. School of Earth and Space Science, University of Science and Technology of China, Hefei 230026, China. E-mail: wangzg@aiofm.ac.cn
2. Key Laboratory of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China

Received 26 August 2006; revised 30 October 2006; accepted 20 November 2006

Abstract
Taihu Lake is one of the five biggest lakes in China. Surface water samples from 26 sampling sites of Taihu Lake were collected. Furthermore wet chemical analysis (CODCr and BOD5) and measurement of three dimensional excitation-emission matrix (3DEEM) spectra in the laboratory have been conducted. Using parallel factor analysis (PARAFAC) model, three components of colored dissolved organic matter (CDOM) have been identified successfully, based on the analysis of 3DEEM data. The characteristics of the three components also have been described by comparing them to some components of CDOM, identified in earlier researches. Meanwhile, spatial variations of concentration for the three components in Taihu Lake have been analyzed, and the result indicates that the concentration of component 1 depends more on the situation of wastewater pollution and can be used as the indicator of wastewater pollution. The relationship between the concentrations of the three components and results of the wet chemical analysis show that none of the three components can be used as indicators of gross organic matter in water. However, the concentrations of all the three components have obvious linear relationships with the BOD5 value, especially for component 1 (r = 0.72878). Finally, the potential applications of the composition analysis based on 3DEEM and PARAFAC model in water quality monitoring have been illuminated.

Key words: 3DEEM; Taihu Lake; monitoring of water quality; CDOM; PARAFAC

Introduction
Color dissolved organic matter (CDOM) is part of the dissolved organic matter (DOM), which can be mainly divided into two groups— natural organic matter (NOM) and anthropogenic organic matter (AOM, i.e., agricultural waste, and industrial organic pollutants). Rapid development of industry and agriculture has led to large amounts of AOM being released into the water, which results in great differences in the composition and concentration of CDOM. Therefore, changes in concentration and composition of CDOM can not only provide information about organic pollution in a special water region, but also the information of organic pollution sources (Wang et al., 2006). However, it is rather difficult to trace the difference in the composition and concentration of CDOM with traditional chemical analysis, because of the complexity and nonconservative behavior of the composition of CDOM in natural environments. Recently, fluorescence spectroscopy, especially three dimensional excitation-emission fluorescence matrix (3DEEM), which is an autofluorescence of several samples measured at several emission wavelengths for several excitation wavelengths, has been used in the study of CDOM composition and concentration. Coble et al. (1996), has successfully identified terrestrial and marine components of CDOM, based on 3DEEM, showing several excitation-emission peaks, which correspond to the special fluorophore of CDOM. Baker (2001), has detected the presence of treated wastewater effluent from the upstream sewage treatment plants in the river water, according to the 3DEEM of the water sample. Wang et al. (2006), has given the information of CDOM concentration and sources in Chaohu Lake, based on 3DEEM spectra. But traditional 3DEEM analysis identified CDOM components, using simple excitation-emission wavelength pair(s) of the fluorescence peak are not sufficient, maybe they are even misleading in the identification and qualification of CDOM components, because CDOM is a complex mixture of organic molecules and fluorescence from different organic molecules, which may have overlapped. Considering the deficiency of 3DEEM in tracing the differences of CDOM components and concentrations, a new mathematical model named PARAFAC has been applied in the analysis...
of the 3DEEM spectra. Several studies have explored and described the underlying chemical components in 3DEEM spectral data by applying the PARAFAC model. Holbrook et al. (2006), investigated the water samples from natural rivers using fluorescence spectroscopy and PARAFAC, and identified three individual fluorophore moieties. Taihu Lake is one of the five largest fresh water lakes in China, which locates in the southeast of the Jiangsu Province. It has been severely polluted in the last two decades. Especially large amounts of AOM directly result in eutrophication and algae proliferation. In this study, water samples from 26 sites of the Taihu Lake have been investigated and further analyzed using the 3DEEM spectra for individual water sample, with the PARAFAC model. It has been attempted to distinguish the components of CDOM in the Taihu Lake and illustrate the potential application of component analysis in water quality monitoring.

1 Methods

1.1 Collection of water samples

Surface water samples of Taihu Lake were collected during the period October 27–29, in 2005. A total of 26 samples corresponding to an individual sampling site were analyzed by 3DEEM and wet chemistry (COD\textsubscript{Cr} and BOD\textsubscript{5}) in the laboratory. All the samples were transported to the laboratory as soon as possible in polypropylene sampling bottles, which had previously undergone a thorough cleaning process, (Washing in alconox cleaner, running water rinse, alternating 0.5 mol/L HCl baths for a minimum of 30 min, followed by a final, triple distilled water rinse), and stored at 4°C in the laboratory, after by being analyze in 3 d. The accurate positions of all the sampling sites were recorded by the global positioning system (GPS), as shown in Fig.1.

1.2 Measurement of 3DEEM data

All samples were allowed to reach laboratory temperature (approximately 20°C), to minimize temperature influence. Furthermore, all the samples were filtrated through a micro-filtration membrane (0.45 µm), prior to measurement, for removing all the insolvable organic particles. 3DEEM measurement was taken using the Skalar fluorescence imaging system (Skalar M-153, Skalar Company, Netherland). Both the excitation wavelength interval and emission wavelength interval were set at 5 nm. Each sample was determined thrice, the average value was acquired as the 3DEEM spectroscopy data. In the process, it was verified that the Raman scattering intensity of pure water (deionized water), located at 415 nm by the excitation of 360 nm kept stable, and the variance was less than 5%. The Raman scattering intensity above was set at 104, to which fluorescence intensity determined for all the samples was normalized.

1.3 Analysis of wet chemistry in laboratory

Prior to the measurement, all the samples were filtrated through the micro-filtration membrane (0.45 µm). COD\textsubscript{Cr} and BOD\textsubscript{5} were determined for 26 separate samples according to the standard of GB11914-89 and the standard of GB7488-87, released by the Chinese State Environment Protection Agency respectively.

1.4 3DEEM data modeling by PARAFAC

The PARAFAC analysis was performed in MATLAB6.5 (Mathworks), using the “N-way toolbox for MATLAB version 2.0”, with default numerical settings, detailed introduction of the PARAFAC model, and its realization refers to the article of Anderson and Bro (2000). To reduce the influence of Rayleigh scatter and second-order Rayleigh scattering, 3DEEM values for those emission wavelengths between excitation wavelengths −20 nm and excitation wavelengths +20 nm, and also those emissions

![Fig. 1 The sketch map of sampling sites in Taihu Lake.](image-url)
wavelengths between the two excitation wavelengths ~20 nm and 575 nm, were removed, and set as missing, respectively. Meanwhile, 3DEEM of pure water (deionized water), was subtracted from the 3DEEM for all the samples, aiming to reduce the influence of Raman scattering on the modeling of PARAFAC. This gave a data set of size 26 × 64 × 25.

Results of the PARFAC model are highly sensitive to the number of components. Consequently, choosing the correct number of components is a critical step in analyzing 3DEEM for samples with unknown fluorophore composition, when using the PARAFAC model. To choose the correct number of components, several diagnostic and validation techniques have been used. Initially, the core consistency diagnostic of the entire dataset provides a quantitative assessment of how well the raw data can be described by the PARAFAC model, for a given number of components. Table 1 shows the core consistency and explains the variance for PARAFAC models of the fluorescence data with 1–8 components. The explained variance indicates that three or four components are optimal because the increase that is obtained with more than four components is small relative to the decrease in core consistency. Next, a so-called jack-knife technique is used, which includes Identity Match Plots (IMP) and Resample Influence Plots (RIP), to detect the outliers in the entire dataset (Riu and Bro, 2003). The IMP and RIP plots reveal that there are no outliers among these samples for the three and four components of the PARAFAC model. Finally, a split-half analysis provides strong evidence for a three-component model because there is good agreement of the excitation and emission loadings between two randomly divided datasets. However, PARAFAC modeling of four-components cannot be validated by split-half analysis. Consequently, three components is a proper component number for the PARAFAC model.

2 Results and discussion

2.1 CDOM component characteristics

Three components have been identified, from the data set of 26 samples from the Taihu Lake, by the PARAFAC model, and the excitation and emission loadings including contour plots of each component are shown in Fig.2. Component 1 presents a fluorescence peak at an excitation/emission wavelength pair of 322 nm/407 nm, component 2 and component 3 correspond to a fluorescence peak at an excitation/emission wavelength pair of 360 nm/425 nm and 294 nm/320 nm, respectively. All the three components identified in this study resemble some fluorophores described in earlier literature. Component 1 in this study is similar to component 4 in the study of Stedmon et al. (2003), and the combination of component 5 and component 6 in the study of Stedmon et al. (2005). Moreover, component 1 also resembles peak M, which is thought to represent marine humic matter (Coble et al., 1998), and peak B, which is considered as one of

![Fig. 2 Excitation (solid lines), emission loadings (dotted lines), and contour plots of component 1 (top), component 2 (middle), and component 3 (bottom).](image)

<table>
<thead>
<tr>
<th>Number of components</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explained variance (%)</td>
<td>85.7</td>
<td>95.6</td>
<td>97.5</td>
<td>98.4</td>
<td>98.7</td>
<td>99.1</td>
<td>99.3</td>
<td>99.5</td>
</tr>
<tr>
<td>Core consistency (%)</td>
<td>100</td>
<td>99.7</td>
<td>60.4</td>
<td>46.8</td>
<td>17.9</td>
<td>4.2</td>
<td>3.0</td>
<td>0.8</td>
</tr>
</tbody>
</table>
the characteristics of wastewater (Baker et al., 2003). Component 2 in this study resembles the component 3 in the study of Stedmon et al. (2003) and the component 4 in the study of Stedmon et al. (2005), both of them are attributed to fulvic acid fluorophore group and are rarely present in wastewater (Baker et al., 2003). Component 3 can be considered as a combination of component 7 and component 8, which corresponds to the tryptophan-like fluorophore group and the tyrosine-like fluorophore group, respectively (Stedmon et al., 2005). Detailed relationships of components in this study and those in other earlier studies have been shown in Table 2.

2.2 Spatial variations of CDOM components in Taihu Lake

Different regions of a large lake usually show different concentrations of CDOM components, which can give information about the pollution situation. The PARAFAC model can quantify the scores of a specific fluorophore component, which is directly proportional to the fluorophore’s concentration (Holbrook et al., 2006). Spatial variations of scores for three CDOM components in Taihu Lake have been shown in Fig. 3. Obviously, all the three components of sample 4 show the highest scores, which means the highest component concentrations are included in 4# sample. Sampling site 4# is located at the outlet of Meiliang Bay—this is generally regarded as one of the most seriously polluted water regions in Taihu Lake (Zhang et al., 2004). As a whole, the three components show similar distribution variation to some extent. For example, 1#–7# samples show higher scores of component 1 and component 2 compared to the other samples, component 3 also keeps a relatively higher scores in 1#–4# samples, possibly because the sampling sites above are near the downtown of Wuxi City and affected more seriously by wastewater. For samples 8#–22#, which are far away from the downtown, scores of component 1, component 2, and component 3 keep relatively lower values, indicating less pollution by wastewater. Moreover, an increase in the trend of the three component scores has occurred in the last several samples with the decrease of distance from the Meiliang Bay. However, an obvious difference has been observed among the three components, the greatest distribution variation occurred in component 1 when compared with the other two components. In conclusion, the scores of the three components identified by the PARAFAC model changes with wastewater pollution, and the scores of component 1 have the greatest distribution variation among the three components.

2.3 Relationship between wet chemical analysis results and CDOM component scores

COD, value is one of the most common index used in water quality analysis, which directly represents the amount of oxygen consumed in the oxidation of organic matter with high oxidative chemical reagents (such as K2Cr2O7 and K2MnO4), and hence can indicate the amount of gross organic matters in water. Here, an analysis of the relationship between CODCr value and CDOM component scores has been conducted and the result has been shown in Table 3. Obviously, no linear relativity has been observed between CDOM component scores and CODCr, the linear relative coefficient for component 1, 2, and component 3 are 0.03611, 0.0326, and 0.15828, respectively. It has been concluded that none of the three CDOM components can indicate the amount of gross organic matter in water. Not all the organic matter in water can emit fluorescence, and many of the organic components identified with PARAFAC model.

Fig. 3 Spatial variations of three CDOM components in the Taihu Lake, identified with PARAFAC model.

<table>
<thead>
<tr>
<th>Component in this study</th>
<th>Excitation and emission maxima</th>
<th>Corresponding component in earlier study</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component 1</td>
<td>Ex = 322 nm, Em = 407 nm</td>
<td>Combination of component 5 and component 6</td>
<td>Humic fluorophore, dominant in wastewater</td>
</tr>
<tr>
<td>Component 2</td>
<td>Ex = 360 nm, Em = 425 nm</td>
<td>Component 3 (Stedmon et al., 2003)</td>
<td>Fulvic acid fluorophore group, present in all environment</td>
</tr>
<tr>
<td>Component 3</td>
<td>Ex = 294 nm, Em = 320 nm</td>
<td>Component 5 (Stedmon et al., 2003)</td>
<td>Tryptophan-like fluorescence and tyrosine-like fluorescence</td>
</tr>
</tbody>
</table>

Table 3 Linear relationships between scores of three CDOM components and COD and BOD

<table>
<thead>
<tr>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>COD</td>
<td>r = 0.03611</td>
<td>n = 26</td>
</tr>
<tr>
<td>BOD</td>
<td>r = 0.72878</td>
<td>n = 25</td>
</tr>
</tbody>
</table>

BOD value of 7# sample is missing.
matters such as sugar and lipid, do not emit fluorescence. Therefore, COD\textsubscript{Cr} values of the samples from Taihu Lake depend on the amount of nonfluorescence organic matter to a great extent, and it is not reasonable to evaluate the amount of gross organic matter merely according to the scores of the CDOM components.

Nevertheless, BOD\textsubscript{5} can reflect the amount of labile organic matters in water, and showed an obvious linear relationship with the scores of three components. The relative coefficients for component 1, component 2, and component 3 are 0.72878, 0.6471 and 0.53541, respectively (Table 3). This fact indicates that all the three components of CDOM make a great contribution to BOD\textsubscript{5} value. Especially component 1, which ascribes to humic fluorophore and is dominant in wastewater, it has the strongest linear relationship with BOD\textsubscript{5}. The direct reason may be that component 1 is the largest proportion of CDOM in Taihu Lake and is the largest contributor to BOD\textsubscript{5}. Thus, the scores of component 1 can be used to indicate BOD\textsubscript{5} and wastewater pollution situation. In conclusion, all the three components of CDOM identified by the PARAFAC model make a contribution to the BOD\textsubscript{5} value; especially the scores of component 1 can be used as an indicator of BOD\textsubscript{5} and situation of wastewater pollution.

3 Conclusions

On the basis of the analysis of 3DEEM data of 26 sampling sites in the Taihu Lake, applying the PARAFAC model, three components of CDOM have been identified. The scores of all the three components changed with the distance of the sampling site from downtown, especially component 1, which showed the highest variation. The relationships between the scores of the three CDOM components and traditional wet chemical indexes (COD\textsubscript{Cr} value, BOD\textsubscript{5} value), indicated that the scores of the three components had an obvious linear relativity with the BOD\textsubscript{5} value, especially for component 1. However, no linear relationship had been found between the scores of the three components and COD\textsubscript{Cr}. Considering the results of the analysis above, several potential applications of component analysis, based on 3DEEM and the PARAFAC model in water quality monitoring, can be recommended as follows:

1. On the basis of the analysis of 3DEEM spectra, applying the PARAFAC model, a better understanding of CDOM components and characteristics can be achieved.

2. Of the three components identified by the PARAFAC model, the distribution variations of component 1 is the most suitable to be used as the indicator of wastewater pollution situation, which is very valuable in investigating wastewater pollution of a large water region.

3. None of the three component scores can be used to indicate the gross amount of organic matter in water, however, the scores of component 1 can be regarded as an indicator of labile organic matter.

References


