Chemosphere 113 (2014) 79–86

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Identifying changes in dissolved organic matter content and characteristics by fluorescence spectroscopy coupled with self-organizing map and classification and regression tree analysis during wastewater treatment

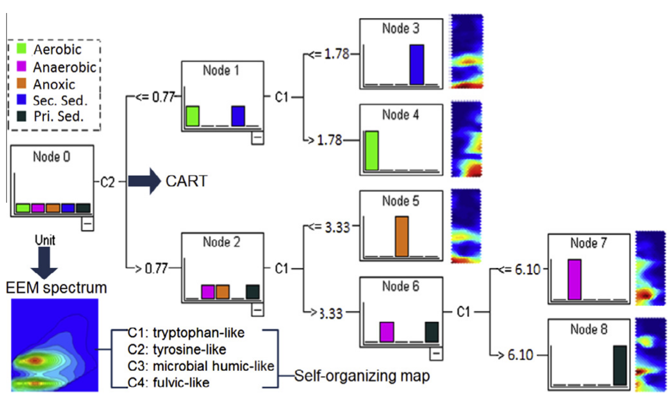
Huibin Yu [a,b], Yonghui Song [a,b,*], Ruixia Liu [a,b], Hongwei Pan [a], Liancheng Xiang [a,b], Feng Qian [a,b]

[a] State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China
[b] Department of Urban Water Environmental Research, Chinese Research Academy of Environmental Science, Beijing 100012, China

HIGHLIGHTS

- Wastewater samples from different units in WWTP were investigated.
- DOM fractions of wastewater were identified by EEM with self-organizing map.
- Latent tracers were sought to monitor DOM removal using CART.
- Protein-like material was dominant component of DOM and deeply removed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 January 2014
Received in revised form 9 April 2014
Accepted 11 April 2014
Available online 14 May 2014

Handling Editor: O. Hao

Keywords:
Dissolved organic matter
Fluorescence spectroscopy
Self-organizing map
Classification and regression tree analysis
Wastewater

ABSTRACT

The stabilization of latent tracers of dissolved organic matter (DOM) of wastewater was analyzed by three-dimensional excitation–emission matrix (EEM) fluorescence spectroscopy coupled with self-organizing map and classification and regression tree analysis (CART) in wastewater treatment performance. DOM of water samples collected from primary sedimentation, anaerobic, anoxic, oxic and secondary sedimentation tanks in a large-scale wastewater treatment plant contained four fluorescence components: tryptophan-like (C1), tyrosine-like (C2), microbial humic-like (C3) and fulvic-like (C4) materials extracted by self-organizing map. These components showed good positive linear correlations with dissolved organic carbon of DOM. C1 and C2 were representative components in the wastewater, and they were removed to a higher extent than those of C3 and C4 in the treatment process. C2 was a latent parameter determined by CART to differentiate water samples of oxic and secondary sedimentation tanks from the successive treatment units, indirectly proving that most of tyrosine-like material was degraded by anaerobic microorganisms. C1 was an accurate parameter to comprehensively separate the samples of the five treatment units from each other, indirectly indicating that tryptophan-like material was decomposed by anaerobic and aerobic bacteria. EEM fluorescence spectroscopy in combination with self-organizing map and CART analysis can be a nondestructive effective method for characterizing structural component of DOM fractions and monitoring organic matter removal in wastewater treatment process.

* Corresponding author at: Chinese Research Academy of Environmental Sciences, Beijing 100012, China. Tel./fax: +86 10 84928380.
E-mail address: songyh@craes.org.cn (Y. Song).

http://dx.doi.org/10.1016/j.chemosphere.2014.04.020
0045-6535/

1. Introduction

Dissolved organic matter (DOM) consists of a heterogeneous mixture of aliphatic and aromatic polymers containing oxygen, nitrogen and sulfur functional groups, which is widespread in engineered systems (Stedmon et al., 2003; Downing et al., 2009; Yu et al., 2012). Dissolved organic carbon (DOC), a substantial part of DOM, is frequently indicative of DOM (Herzsprung et al., 2012). DOM faces a challenge in engineered systems, as it deeply affects all wastewater treatment performance. The content and composition of DOM can dominate coagulation process, disinfection by-product formation, membrane fouling, oxidant demand, microbial activity and pollutant transformation (Swietlik et al., 2004; Zularisam et al., 2006; Shin et al., 2008). Hence effective tools are required to monitor wastewater treatment facilities and evaluate the effect and removal of DOM in wastewater treatment plant (WWTP). Such methodologies can correct treatment parameters (dissolved oxygen concentration, hydraulic retention time, sludge recycling ratio and so on) in response to DOM alterations in raw water and ensure efficiency and reliability of wastewater treatment performance (Ishii and Boyer, 2012; Yu et al., 2013).

Three-dimensional excitation–emission matrix (EEM) fluorescence, a rapid, inexpensive and reagentless tool, has shown great promise for indicating water treatment efficiency and finished water quality (Murphy et al., 2011; Ishii and Boyer, 2012). EEM can commonly capture protein-like material, dissolved microbial byproducts, fulvic-like material and humic-like material in the same sample by a peak-picking method (McKnight et al., 2001; Hudson et al., 2007); whereas multivariate data analysis techniques have been increasingly applied to quantitatively interpret EEM. The currently most widely used technique, parallel factor analysis model with a non-negativity constraint, discriminates independent fluorescent components from complex EEM, whose maximum intensities are used to monitor DOM alterations in wastewater treatment process (Bieroza et al., 2009). Moreover Bro and Vidal (2011) have developed an EEMizer with parallel factor analysis, which can implement on-line monitoring WWTP performance.

Self-organizing map, an unsupervised neural network algorithm, is utilized to deconvolute complex EEM and to represent distribution and relationship between organic matter fractions and fluorescence groups (Basheer and Hajmeer, 2000; Rhee et al., 2005). This algorithm has been applied to EEM data clustering, where input feature vectors can be explored to distinguish any reasonable correlations among the data, without prior knowledge or hypotheses concerning the given data set (Lee et al., 2005). Self-organizing map, as parallel factor analysis, may extract key features of the dataset in form of self-organizing map normalized weights, which can quantitatively characterize EEM clusters (Bieroza et al. 2009, 2011). Hence the weights can indicate relative abundance of fluorescence components. The algorithm has been commonly used to recognize olive oil fluorescence spectroscopy, monitor various fermentation processes, and estimate DOM removal in drinking water plant (Scott et al., 2003; Rhee et al., 2005; Bieroza et al., 2011). However, few systematic studies have been implemented to evaluate removal efficiency of DOM fractions from wastewater in WWTP by self-organizing map.

The aims of this study were 2 folds: (i) to characterize EEM fluorescence spectroscopy coupled with self-organizing map for identifying changes in DOM fractions and content of wastewater; and (ii) to seek latent tracers to monitor organic matter removal in wastewater treatment performance using classification and regression tree (CART) analysis.

2. Materials and methods

2.1. Sample collection

The water samples were collected in a large-scale WWTP in Beijing, China. The maximum capacity of the WWTP is approximately 1 000 000 $m^3\,d^{-1}$, and its serving population is about 2 400 000. A traditional anaerobic/anoxic/oxic (A^2O) process is adopted in the WWTP for simultaneous removal of nitrogen, phosphorus and carbon (Yu et al., 2013). The treated wastewater moves into the disinfection unit and is subsequently discharged into the river. The influent wastewater in the WWTP is domestic sewage only mixed with a small amount of industrial wastewater. The average influent concentrations of chemical oxygen demand, biochemical oxygen demand, suspended solids and ammonium are approximately 500, 200, 250 and 30 $mg\,L^{-1}$ respectively, while the effluent from the WWTP meets further decreases in their concentrations to less than 60, 20, 30 and 3 $mg\,L^{-1}$ respectively.

Water samples were collected from corresponding effluents of the units, i.e. samples #1 and #2 from effluent of the primary sedimentation tank, #3 and #4 from the anaerobic tank, #5 and #6 from the anoxic tank, #7 and #8 from the oxic tank, and #9 and #10 from the secondary sedimentation tank. Duplicate samples at each sample site were collected with a Wildco Kemmerer 1.2 L sampler, completely mixed and transferred into an EE BOD bottle. Samples were filtered through glass fiber filters (Whatman GF/F, 0.7 μm, pre-combusted at 450 °C for 4 h). The filtrate was collected into precombusted glass amber bottles and stored in the dark at 4 °C until analyzed. All samples were analyzed within 2 d of collection.

2.2. EEM fluorescence spectroscopy

DOC concentrations of all filtrate samples were determined by a TOC analyzer (analytic jena multi N/C 3100 TOC, Germany). EEM spectroscopy was recorded for each sample on a Hitachi Fluorescence Spectrophotometer (F-7000) equipped with the fluorescence solutions 2.1 software (Hitachi high-Technologies Corporation 1998, 2008) for data processing. Scans were conducted with excitation wavelengths from 200 to 450 nm at 5 nm steps, emission wavelengths from 280 to 550 nm at 5 nm steps, 5 nm bandwith, and 0.5 s integration time. Instrument excitation and emission were adjusted, before EEM spectra of Mili-Q water were subtracted from all sample EEM. Inner filtering effect was corrected by absorbance spectroscopy (McKnight et al., 2001). The fluorescence intensities were normalized the area under the water Raman peak (382–412 nm emission ranges at 350 nm excitation), and then converted to quinine sulfate units. (Murphy et al., 2011; Dahm et al., 2013).

2.3. Self-organizing map model

Self-organizing map can carry out a transformation of initial high-dimensional matrix of input data into a two-dimensional map, i.e. it projects input vectors of high-dimension onto a specific set of single processing elements (neurons), whereas keeping the topological and metric correlations of the input data sets (Garcia et al., 2007). The algorithm consists of two-layered artificial neural networks, namely an input layer and an output layer. Each vector in the input layer is adequately connected to each neuron of the output layer which is associated with the reference vector that has self-organizing map weights. Similarity between the input vector and the reference vector is determined by the Euclidean distance metric. Subsequently, the output neuron which is closest