Oil source recognition technology using concentration-synchronous-matrix-fluorescence spectroscopy combined with 2D wavelet packet and probabilistic neural network

Xiao-Dong Huang [a,b], Chun-Yan Wang [a,c,d,*], Xin-Min Fan [a,b], Jin-Liang Zhang [c], Chun Yang [d], Zhen-Di Wang [d]

[a] Department of Physics and Electronic Science, Weifang University, Weifang 261061, China
[b] Institute of New Electromagnetic Materials, Weifang University, Weifang 261061, China
[c] College of Resources Science and Technology, Beijing Normal University, Beijing 100875, China
[d] Emergencies Science and Technology Section (ESTS), Science and Technology Branch, Environment Canada, 335 River Rd., Ottawa, Ontario K1A 0H3, Canada

HIGHLIGHTS

- Unconventional Oil source recognition released into the environment is a more complexand difficult task.
- Oil source recognition technology using concentration-synchronous-matrix-fluorescence spectroscopy is proposed.
- Optimal recognition of the samples, the accuracy and speed of four classification algorithms were compared.

GRAPHICAL ABSTRACT
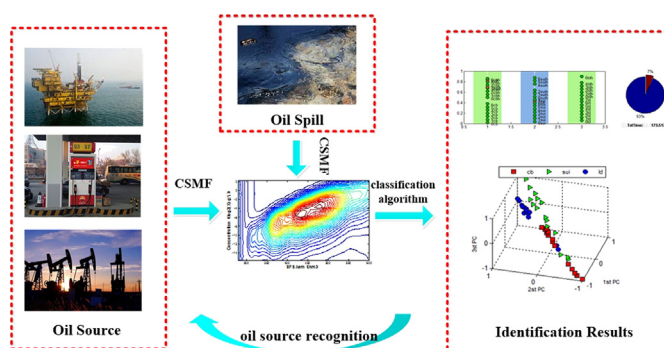
ARTICLE INFO

ABSTRACT

Developing an accurate, rapid and economic oil source recognition method is essential for water recourses protection. Concentration-synchronous-matrix-fluorescence (CSMF) spectroscopy combined with 2D wavelet packet and probabilistic neural network (PNN) was proposed for source recognition of crude oil and petroleum products samples in this study. 2D wavelet packet was used to extract wavelet packet coefficients as the feature vectors from CSMF contour image and four algorithms, Back-propagation (BP) neural network, Radial based function neural network (RBFNN), Support vector Machine (SVM) and probabilistic neural network (PNN) were carried out for pattern recognition. With the introduction of interference factors such as weathering and sea water adulteration to the three samples from Bohai bay territory of China, the comparison about accuracy and recognition time of the four methods was discussed and the results showed that PNN network maintain the highest recognition accuracy and speed. These findings may offer potential application for oil spill recognition for unconventional oil.

© 2017 Published by Elsevier B.V.

1. Introduction

Due to the depletion of conventional oil resources, the production and transportation of unconventional oil increase dramatically.

* Corresponding author at: Department of Physics and Electronic Science, Weifang University, Weifang 261061, China.
E-mail addresses: ssamxd@163.com (X.-D. Huang), wangcy@bnu.edu.cn (C.-Y. Wang).

Unconventional Oil source recognition released into the environment is a more complex and difficult task. The traditional methods such as GC-FID and GC–MS mainly for the detection and identification of the conventional crude oil meet the new challenge (Wang et al., 2002). The reality has put forward new requirements of technology for the identification of oil spill samples, the evaluation of weathering and the tracing of the source (Wang et al., 2009; Yang et al., 2009; Shi et al., 2012; Shi et al., 2013; Wang et al., 2013; Xiao et al., 2014).

The abundance and distribution of polycyclic aromatic hydrocarbons (PAHs) and their derivatives could vary significantly in crude oils from different regions and petroleum products of different types (Li et al., 2014; Yang et al., 2014; Han et al., 2015; Lilla and S., 2015). PAHs are high molecular mass, weathering-resistant compounds that can be source specified and the fluorescence spectroscopy is quite applicable to the PAHs identification and qualification (Radović et al., 2014; Weng et al., 2015).

However, one major problem with the analysis of oil-related samples of conventional fluorescence spectroscopy is the severe overlapping of emission and excitation bands (Kumar and Mishra, 2013; Cho et al., 2014). In such case, synchronous fluorescence (SF) could prove beneficial as both the excitation and emission monochromators are scanned simultaneously in such a manner that a constant wavelength interval is kept between emission and excitation wavelength $\Delta\lambda$, which can increase the spectra selectivity(Deepa et al., 2006; C.Y. Wang et al., 2010).

The other problem is that the fluorescence spectra of oil-related samples are quite sensitive to the solution concentration. The conventional fluorescence spectroscopy technique was used to dilute the oil samples solution to so-called linear concentration range, and thought the main peak intensity of fluorescence spectra has a linear relationship with concentration in that range (Zhou et al., 2012; Zhou et al., 2013). However, due to different fluorescence linear range of different PAHs, the lower ring PAHs would be quenched at high concentration only remaining the information of high ring PAHs, and the dilution to the linear concentration range, would loss the fluorescence information of high ring PAHs in this process. Then single concentration value cannot obtain the aromatics composition ratio of spilled oils (Patra and Mishra, 2001; Patra, 2008; C. Wang et al., 2010).

To address the above two problems, Synchronous scanning fluorescence spectroscopy and Concentration dimension are introduced to form the multi-dimensional concentration-synchronous-matrix-fluorescence (CSMF) spectra, which could reflect much more fluorescence information of different aromatic components under different excitation wavelength and concentration values (Wang et al., 2016).

For better extraction of spectral information, 2D wavelet packet was used to decompose and reconstruct the CSMF contour images. And for optimal recognition of the samples, the accuracy and speed of four classification algorithms, BP neural network, Radial based function neural network (RBFNN), Support vector Machine (SVM) and PNN network were compared.

2. Materials and methods

2.1. Samples preparation

10 sample sets were chosen to test the recognition accuracy and speed of the four methods, which includes 3 similar crude oil samples and 7 samples of different oil types (Table 1).

In order to verify the accuracy of the method in case of the seawater weathering and other external disturbances, we designed a weathering experiment, and different Interference factors such as weathering and seawater have been taken into consideration in the experiment. The 3 similar crude oil samples were dispensed into pre-combusted glass bottles containing seawater and deionized water respectively. All the bottles were placed near the south windows inside the laboratory to

Table 1
Samples used for test.

Sample sets	Name	Abbrev	Oil source
Similar crude oil samples	Chengbei 306#	CB	Bohai Sea Platform in China
	Lvda A-12#	LD	
	Suizhong 36-1#	Sui	
Samples of different oil types	zhongyuan PJ3	ZY	Zhongyuan Oil field
	Tuo 167#	Tuo	Shengli Oil Field
	Huang 16#	Huang	Jiangsu Oil Field
	4# Fuel	4#F	
	−10#Diesel	10#D	
	0#Diesel	0#D	
	93#Gasoline	93#G	

receive natural solar radiation. During the 60-d experimental period, the average temperature was 30 °C.

For each sample, 19 solutions from the stock solution 5 mg/mL to 10^{-4} mg/mL with a serial of reduced concentration (5.00, 3.30, 2.50, 1.67, 1.25, 0.830, 0.625, 0.417, 0.313, 0.208, 0.15625, 0.104, 0.078, 0.052, 0.039, 0.026, 0.020, 0.013, 0.01 g/L respectively) were acquired by the step-by-step method for dilution. No dehydration, purification and nitrogen concentration were applied before the Fluorescence measurement to test the robustness of this technique with a view to the purpose of practical application.

2.2. Measurements of CSMF

The shimadzu RF5301 synchronous fluorescence spectrophotometer was used to measure the synchronous fluorescence of prepared samples, the slit width of excitation wavelength and emission wavelength were set as 5.0 nm and 1.5 nm, respectively. The excitation wavelengths ranging from 220 nm to 600 nm were used in 1 nm increments at a scan rate of 1500 nm/min with a constant offset $\Delta\lambda$, and 40 nm was chosen as the optimum wavelength difference on the basis of the peak number and peak intensity of different concentrations and samples (C. Wang et al., 2010; Wang et al., 2016).

19×340 CSMF spectra could be formed, which was based on the synchronous spectra by scanning the 19 series of sample concentrations.

2.3. Spectral preprocessing

The derivative smoothing function of the Shimadzu fluorescence spectrophotometer was used to reduce the noise from the system and fluorescence signal itself in the process of data collection. All the spectra were normalized to their maximum intensity before modeling, and a non-negativity constrain was imposed. 10-Order chebyshev filter and cubic interpolation method were also used to optimum all the spectra matrixes to 128×128 normalized spectral images.

2.4. Discrete wavelet packet

DWT methods have been successfully applied for the resolution of overlapped spectra for the quantitative measurement of multi-component mixtures (Barclay et al., 1996; Ren and Gao, 2008; Nicolis et al., 2011; And and Griffiths, 2007). The DWT is the sum of the signal multiplied by the scaled and shifted mother function, which can be expressed as:

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi\left(2^{-j}t - k\right)$$

where indices j, k are the scale and the shift parameter respectively.

Discrete wavelet packet transformation is an extension of DWT, whereby all nodes in the tree structure are allowed to split further at each level of decomposition (Jay Liu et al., 2007; Zhang et al., 2015).