Acta Oceanol. Sin., 2018, Vol. 37, No. 11, P. 116–122
DOI: 10.1007/s13131-018-1254-9
http://www.hyxb.org.cn
E-mail: hyxbe@263.net

Establishment and application of an intelligent treating method for oil spill identification

TAN Liju[1], ZHAO Ruxiang[2], YIN Xiaonan[2], ZHANG Haijiang[2]*, WANG Jiangtao[1]*

[1] College of Chemistry and Chemical Engineering, Ocean University of China, Qingdao 266100, China

[2] Yantai Oil Spill Response Technical Center of Yantai Maritime Safety Administration, Yantai 264000, China

Received 21 December 2017; accepted 2 April 2018

Abstract

In the identifying process of an oil spill accident, manual integral and artificial visual comparison are commonly used at present to determine the oil spill sources, these methods are time-consuming and easily affected by human factors. Therefore, it is difficult to achieve the purpose of rapid identification of an oil spill accident. In this paper, an intelligent method of automatic recognition, integration and calculation of diagnostic ratio of Gas Chromatography-Mass Spectrometer (GC/MS) spectrum are established. Firstly, four hundreds of samples collected around the world were analyzed using a standard method and Retention time locking technology (RTL) was applied to reduce the change of retention time of GC/MS spectrum. Secondly, the automatic identification, integration of *n*-alkanes, biomarker compounds, polycyclic aromatic hydrocarbons and calculation of the diagnostic ratios were realized by MATLAB software. Finally, a database of oil fingerprints were established and applied successfully in a spill oil accident. Based on the new method and database, we could acquire the diagnostic ratios of an oil sample and find out the suspected oil within a few minutes. This method and database can improve the efficiency in spilled oil identification.

Key words: oil fingerprint, automatic identification, diagnostic ratio, standardized database, oil spills sources

Citation: Tan Liju, Zhao Ruxiang, Yin Xiaonan, Zhang Haijiang, Wang Jiangtao. 2018. Establishment and application of an intelligent treating method for oil spill identification. Acta Oceanologica Sinica, 37(11): 116–122, doi: 10.1007/s13131-018-1254-9

1 Introduction

Oil fingerprint identification is not only an important method to determine the source of spilled oil (Wang et al., 1995, 1999; Wang and Fingas, 1997; Wang, 2003; Stout et al., 2001), but also a crucial factor affecting the disposal efficiency in an oil spill accident (Sun et al., 2004, 2006). In the information of an oil fingerprint, the diagnostic ratio is a parameter with a large difference among various oil sources (Wang et al., 2004; Wang and Stout, 2007; Yang et al., 2008). Because this parameter is a ratio of typical characteristic components (oil fingerprint), the analytical error has little effect on the results (Daling et al., 2002; Christensen et al., 2005). Therefore, diagnostic ratios of oil fingerprints are widely applied in spilled oil identification (Wang et al., 2004; Wang and Stout, 2007; Yang et al, 2008; Bayona et al., 2015). For the oil spill accident involved ships, it is required to draw conclusions as soon as possible to release the irrelevant ship because of the huge economic losses of delays in shipment. However, it is difficult to achieve a rapid and accurate conclusion because it is drawn manually in most cases. In the widely used identification process of oil fingerprints, chromatograms are analyzed through visual comparison, chromatographic peaks are integrated artificially and diagnostic ratios are calculated manually, so the process is time-consuming and easily affected by human factors. For example, the European oil spill identification standard (CEN/TR 15522-2: 2012) (CEN/TR, 2012) recommends that more than 120 compounds should be identified for each oil sample, and China's national standard *Specifications for Identification System of Spilled Oils on the Sea* (GB/T 21247–2007) recommends that 150 compounds should be identified for each oil sample. Due to the analyst's habits on selection baseline, the starting and ending time are different, the integral results are also different. Thus, the classical identifying method is time-consuming and lack of accuracy.

With the development of computer technology, various digitized handling methods began to be used to analyze the oil fingerprints so that spilled oil could be identified quickly and objectively (Qin, 2015). The German Hydrographic Bureau built a system to detect an oil spill named computerized oil spill identification (COSI) (Dahlman, 2016). Staniloae et al. (2001) built a pattern recognition system based on infrared spectroscopy and gas chromatography to compare spilled oil with suspected oil. Sun et al. (2012) developed a set of software for oil fingerprint identification which improved the investigation efficiency of the spilled oil. Al-Kaabi et al. (2017) identified some beached oil. Additionally, various measuring methods and oil fingerprint databases were built in some countries (Sun et al., 2006). Because of the different measuring methods of those database, they had no comparability among databases and lacked applicability for every country. Therefore, a standard method and database should be established to analyze the source of spilled oil automatically and accurately.

Based on the above analysis, this paper intends to solve three problems: (1) write a program that can recognize and integrate chromatogram peaks automatically, as well as calculate a large amount of oil diagnostic ratios automatically; (2) establish a database for the diagnostic ratios of known oils; and (3) verify the

---

Foundation item: The National Key Research and Development Program under contract No. 2016YFC1402101; the Project of Maritime Safety Administration.

*Corresponding author, E-mail: yiyouzhongxin@163.com; jtwang@ouc.edu.cn

accuracy of the program and database through an oil spill accident.

2 Materials and methods

2.1 *Sample treatment*

All oil samples were analyzed with a 7890A Gas Chromatography (GC)-5975C Mass Spectrometer(MS) (Agilent Technologies, USA) equipped with a HP-5MS fused silica capillary column (60 m, 0.25 mm i.d., 0.25 μm film thickness, J&W Scientific, USA) in the pulsed splitless mode. The GC oven temperature was programmed from 60 to 300°C at 6°C/min and maintained at 300°C for 30 min. Helium was used as the carrier gas with 1.0 mL/min constant flow. The temperature of the injector, transfer line and ion source was 290°C, 280°C and 230°C, respectively (Zhang et al., 2016).

Qualitative analysis of the oil samples was performed in selected ion monitoring (SIM) mode. All samples were analyzed twice in order to eliminate the effect of analytical variability and sample heterogeneity.

Approximately 0.15 g of pure oil sample was extracted with 10.0 mL of high purity *n*-hexane (Merck, HPLC, Germany)/dichloromethane (Merck, HPLC, Germany) (1:1, v/v), and approximately 1.0 g of anhydrous sodium sulfate was added to remove water. The mixture was vortexed for 30 s and centrifuged at 3 000 r/min. for 5 minutes. After that, 1.0 mL of the supernatant was transferred to a vial containing 1.0 g of silica gel which was activated at 180°C for 2 hours. Then, 5.0 mL of high purity n-hexane/dichloromethane (1:1, v/v) was added to the vial. The mixture was vortexed and centrifuged at 3 000 r/min. An aliquot of 1.0 mL of each extract was subjected to analysis (Zhang et al., 2016).

2.2 *Retention time locking (RTL) technology*

Retention time is a very important parameter for chromatographic instruments which affects the accuracy of separation and identification of interesting analytes from complex mixtures. However, the retention time is difficult to repeat and consistent due to the replacement, aging and cutting the chromatographic column, and the fluctuation of temperature and air pressure.

Retention time locking (RTL) technology is the feature of Agilent Chem Station software, which is developed to lock the retention time of the target analyte by adjusting the head pressure. In this work, RTL was applied to reduce the change of retention time at different time and on different GC/MS instruments.

2.3 *Automatic identification and integration of GC/MS chromatogram peaks*

Before automatic identification and integration by computer, manual identification and integration was needed to obtain relevant parameters of characteristic peaks, such as the ionic species, peak name and type (categorized into single peak, double peak and multi peak), retention time, and start time and end time of the peak. Among these parameters, the retention time could provide the positional information of the peak height, and the start time and end time could provide the integral range of peak area. The MATLAB software was used to establish a method of automatic identification and integration of chromatogram peaks on the basis that the retention time of the target compounds was basically unchanged by RTL technology. The specific steps to establish the method are as follows:

(1) Smoothing the chromatographic peaks

The chromatographic peaks need to be smoothed before identification because of the unavoidable noise, which may affect the identification of chromatographic peaks and cause misjudgment. It should be pointed out that the smoothed data were just used in the process of identification (found the appeared time of the peaks), and the raw data were still used in the integral operation for the precision.

(2) Automatic recognition of chromatographic peaks

According to the retention time, start time and end time obtained by manual integration, the peaks and valleys of chromatographic peaks were founded, and then chromatographic peaks were drew through our own program language.

(3) Automatic calculation of peaks' areas

After the chromatographic peaks were identified automatically, the areas of the corresponding peaks were calculated by the trapezoid area formula with the determined starting position and end position and the corresponding horizontal baseline.

2.4 *Establishment the database of oil diagnostic ratios*

A total of 440 crude oil samples and fuel oil samples were collected worldwide, mainly collected from China, Kingdom of Saudi Arabia, Iraq, Kuwait, Venezuela, Korea, Singapore, etc, and then these samples were analyzed by GC/MS locked by RTL technology. Due to the use of RTL technique, the retention time of the target compounds in the oil samples varied little, which made it possible to automatically identify and integrate the spectrogram. Automatic recognition and integration of chromatographic peaks of oil samples were realized by programming language and MATLAB software. After that, the corresponding diagnostic ratios of oil samples were calculated, and then the oil fingerprint database was established based on these diagnostic ratios.

2.5 *Automatic retrieval of diagnostic ratios for suspected oil samples*

The correlation coefficient method was used to study the correlation between unknown spilled oil and oil samples in the database. A set of relative content ratios (x_1, x_2, ..., x_n) of hydrocarbon was used to represent unknown spilled oil samples and another set of ratios (y_1 , y_2, ..., y_n) represented oil samples in the database. The Pearson correlation coefficient R between two samples was calculated by the formula below (Xu, 2012):

$$R=\frac{N\sum x_i y_i-\sum x_i\sum y_i}{\sqrt{N\sum x_i^2-\left(\sum x_i\right)^2}\sqrt{N\sum y_i^2-\left(\sum y_i\right)^2}}. \tag{1}$$

According to the results of R, we can distinguish the relationship between unknown spilled oil sample X and suspected oil Y in the database. The closer the R value approaches to 1, the greater the degree of correlation is between the two oils, while the closer the R value approaches to 0, the lower the correlation is between the two oils.

Generally, when the correlation coefficient R is greater than 0.95, the unknown oil can be regarded as the same kind of oil sample in the oil fingerprint database. In contrast, when the maximum correlation coefficient R is less than 0.8, it indicates that the same oil spill was not found in the oil fingerprint database. When the parameter R is between 0.8–0.95, there is a certain degree of similarity between unknown spilled oil sample and suspected oil sample.

3 Results

3.1 *Effect after using the RTL*

Automatic recognition and automatic integration of chromatographic peaks by computer required highly accurate and repetitive retention time, therefore, RTL method was applied to reduce the change of retention time at different time and on different GC/MS instruments.

Before using RTL, the change of retention time of C_{17} on two different GC/MS instruments was 0.013 min (Fig. 1a), however, the corresponding time reduced to 0.003 min when using RTL (Fig. 1b). Thus, RTL technology could effectively reduce the change of retention time of the target compound.

3.2 *Verification of the accuracy of automatic recognition and automatic integration of chromatographic peaks*

To verify the accuracy of automatic recognition of chromatographic peaks, the peaks redrawn by computer and the original GC/MS spectrograms were compared. The verification of the accuracy of the automatic integral program was performed by comparing the diagnostic ratios calculated by computer and manual methods. Three kinds of typical oil fingerprints were analyzed and are listed below.

3.2.1 *Automatic recognition of n-alkanes*

The original GC/MS chromatogram of alkanes, pristane and phytane (*m/z* 85) and the chromatogram redrawn by computer according to the retention time, start time and other information are shown in Figs 2a and b.

It can be seen from Fig. 2 that the trend of *n*-alkanes redrawn by computer was completely consistent with the original chromatogram. This shows that the accuracy of automatic recognition of alkanes by computer was high and satisfactory.

3.2.2 *Automatic integral of n-alkanes*

The characteristic peaks of C_{17}, pristane (Pr), C_{18}, phytane (Ph) and their diagnostic ratios of C_{17}/Pr, C_{18}/Ph and Pr/Ph were commonly used in the semi-quantitative analysis of a spilled oil accident. Three kinds of crude oils and three kinds of fuel oils were selected to verify the accuracy of automatic integration. To ensure the integration accuracy of the featured peak, it is necessary to integrate the parallel sample data of oil samples, obtain their average, and then calculate the diagnostic ratios.

The comparison between the data of diagnostic ratios obtained by manual and automatic integration is shown in Table 1.

In Table 1, the data from different means were very close and the data of relative standard deviation (RSD%) were between 0.02 and 1.10 with an average of 0.47. All of the RSD% values were below 5%, so the automatic integration of a characteristic peak and the calculation of diagnostic ratios by computer can obtain accurate results.

3.2.3 *Automatic recognition of biomarkers*

Biomarkers can reflect the source properties and generate geological conditions of oil. Therefore, biomarker compounds such as terpane (*m/z* 191) and sterane (*m/z* 217 and *m/z* 218),

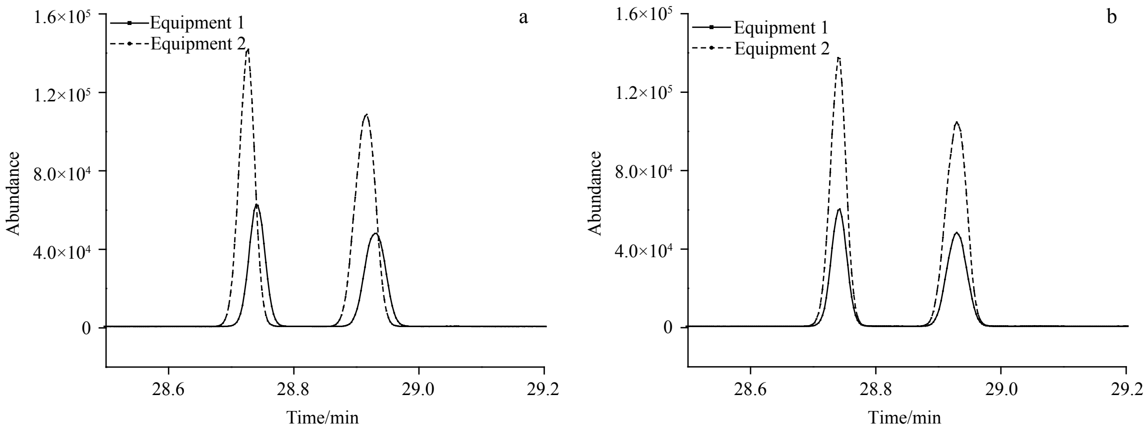

Fig. 1. The retention time of Pristane in two sets of GC/MS before (a) and after (b) RTL (*m/z* 85).

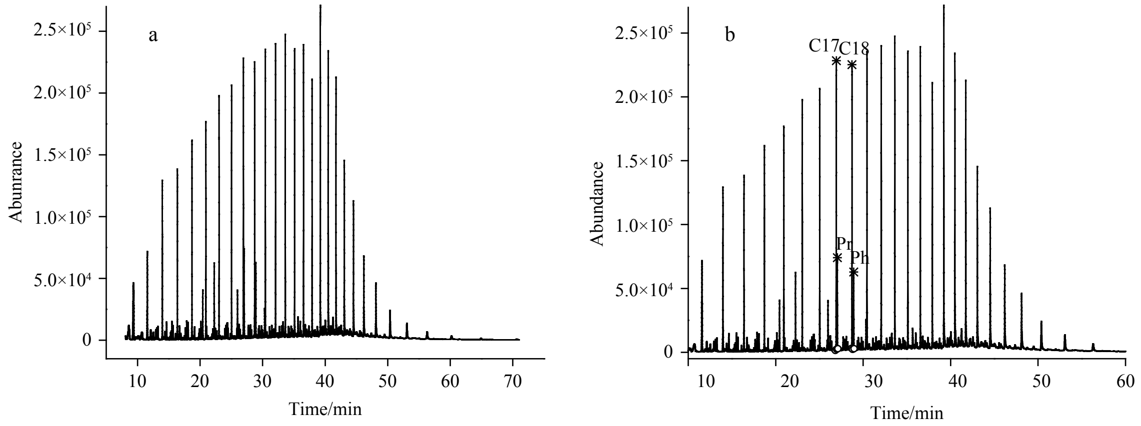

Fig. 2. GC/MS chromatograms of *n*-alkanes, Pristane and Phytane (*m/z* 85) (a) the original chromatogram and (b) the redrawn one by computer.

Table 1. Comparison of diagnostic ratios of *n*-alkanes between manual and automatic integration

Samples	Diagnostic ratio	Manual results	Automatic results	RSD/%
YY036	$n\mathrm{C}_{17}$/Pr	2.59	2.60	0.36
	$n\mathrm{C}_{18}$/Ph	2.79	2.78	0.19
	Pr/Ph	1.11	1.10	0.61
YY088	$n\mathrm{C}_{17}$/Pr	1.29	1.29	0.46
	$n\mathrm{C}_{18}$/Ph	0.94	0.94	0.43
	Pr/Ph	0.78	0.77	0.23
YY167	$n\mathrm{C}_{17}$/Pr	1.06	1.06	0.48
	$n\mathrm{C}_{18}$/Ph	1.80	1.77	1.10
	Pr/Ph	1.76	1.73	1.05
RL022	$n\mathrm{C}_{17}$/Pr	2.23	0.02	0.60
	$n\mathrm{C}_{18}$/Ph	2.25	0.00	0.08
	Pr/Ph	0.90	0.01	0.44
RL046	$n\mathrm{C}_{17}$/Pr	2.39	2.38	0.32
	$n\mathrm{C}_{18}$/Ph	2.22	2.22	0.10
	Pr/Ph	0.88	0.88	0.19
RL121	$n\mathrm{C}_{17}$/Pr	5.44	5.44	0.02
	$n\mathrm{C}_{18}$/Ph	6.63	6.55	0.93
	Pr/Ph	1.18	1.16	0.95

and their corresponding diagnostic ratios, play a very important role on identifying spilled oil, especially for the analysis of weathered oil.

The original and redrawn GC/MS chromatograms of terpane were shown in Figs 3a and b.

It can be seen that the chromatogram redrawn by computer were completely consistent with the original one. This shows that the accuracy of the automatic identification of biomarkers by computer was high and satisfactory.

3.2.4 *Automatic integration of biomarkers*

The results of the diagnostic ratios of biomarkers obtained by manual and automatic integration are shown in Table 2.

Table 2 shows that the data obtained by different means are very close and most RSD% are below 5.00% and only one datum is larger than 5%. This demonstrates that the automatic integration for diagnostic ratios of biomarkers by computer are reliable.

3.2.5 *Automatic recognition of polycyclic aromatic hydrocarbons (PAHs) and their homologues*

PAHs and their homologues in oil include naphthalene, phenanthrene, dibenzothiophene, fluorine, chrysene and alkylated homologues. The composition of PAHs and their homologues in crude oil and petroleum products obtained by different refining processes are different. The relative molecular mass of PAHs and their homologues is large, their stability is high, and their weathering resistance is stronger than saturated hydrocarbons. Therefore, they are important indicators for the identification of spilled oil.

The diagnostic ratios in PAHs and their homologues such as 2-MP/1-MP (2- methyl phenanthrene/1- methyl phenanthrene, 192 ion), 4-MD/1-MD (4-methyl dibenzothiophene/1- methyl dibenzothiophene, 198 ion), C2-dbt/C2-phe (C2- dibenzothiophene/C2- phenanthrene, 212/206), C3-dbt/C3-phe (C3-dibenzothiophene/C3-phenanthrene, 226/220), and C3-dbt/C3-chr) (C3- dibenzothiophene/C3-chrysene, 226/270) were used commonly in oil spill identification. The original GC/MS spectrogram and the redrawn one by computer of the PAHs and their corresponding alkylated homologues are shown in Fig. 4.

3.2.6 *Automatic integration of PAHs and their homologues*

The comparison results of the diagnostic ratios for PAHs and their homologues obtained by manual and automatic integration are shown in Tables 3 and 4.

From the tables, we can see that in the crude oil and fuel oil samples, some diagnostic ratios such as 2-MP/1-MP, 4-MD/1-MD, C2-dbt/C2-phe, and C3-dbt/C3-chr have little differences between manual and automatic integration, and the RSD% values are almost lower than 5%. While the C3-dbt/C3-phe has slightly larger differences, the overall results are satisfactory.

Compared with manual integral and artificial visual comparison method, the automatic recognition and integration method significantly shortened the working time. However, the difficulty of the new method lied in how to accurately determine the starting and ending position of the chromatographic peaks, especially for low concentration chromatographic peaks.

3.3 *Application of the diagnostic ratios database*

In 2015, an unidentified oil spill was found in the sea area of Shandong Province, China. Samples of spilled oil were collected immediately and brought to the laboratory for analysis and identification.

3.3.1 *Automatic calculation of the diagnostic ratios for unknown spilled oil*

After the analysis was completed, the spectrogram of the un-

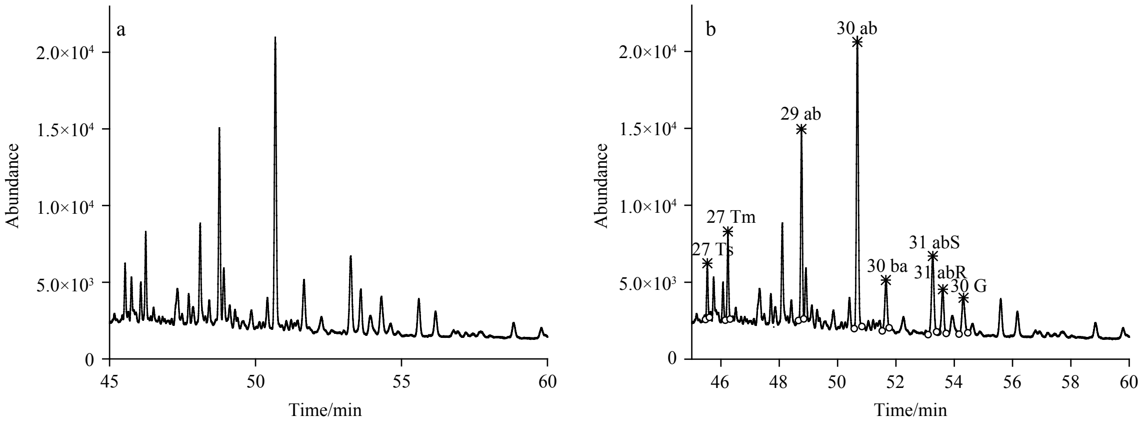

Fig. 3. GC/MS spectrum of terpane (m/z 191) the original chromatogram (a) and the redrawn one by computer (b).

Table 2. Comparison of the diagnostic ratios of biomarkers between manual and automatic integration

Samples	Diagnostic ratio	Manual results	Automatic results	RSD/%
YY036	27Ts/27Tm	0.61	0.59	1.86
	29αβ/30αβ	0.54	0.55	0.75
	29Ts/30αβ	0.16	0.15	2.60
	29βα/30αβ	0.08	0.08	0.99
	30O/30αβ	0.11	0.12	6.49
	30βα/30αβ	0.21	0.22	0.55
	30G/30αβ	0.19	0.18	3.13
	C31αβ(*S*/(*S*+*R*))	0.64	0.64	0.39
	C32αβ(*S*/(*S*+*R*)	0.57	0.58	0.52
	C33αβ(*S*/(*S*+*R*)	0.56	0.59	3.01
	27ββ(*S*+*R*)/[28ββ(*S*+*R*)+ 29ββ(*S*+*R*)]	0.39	0.39	1.25
	28ββ(*S*+*R*)/[27ββ(*S*+*R*)+ 29ββ(*S*+*R*)]	0.57	0.59	1.62
	29ββ(*S*+*R*)/[27ββ(*S*+*R*)+ 28ββ(*S*+*R*)]	0.55	0.54	0.55
	29αα*S*/29αα*R*	0.91	0.92	0.83
	29ββ(*S*+*R*)/29αα(*S*+*R*)	0.97	0.91	4.73

known spilled oil was recognized and integrated automatically using the self-written program language. The diagnostic ratios were calculated according to the obtained peak area, and the data are shown in Table 5.

3.3.2 *Automatic retrieval of the diagnostic ratios for unknown spilled oil*

The diagnostic ratios of the unknown spilled oil were compared with the known samples in the oil fingerprint database using the self-written program language. The diagnostic ratios of the samples in the database were screened and sorted and 10 samples closest to the unknown one are shown in Table 6. According to the results of the automatic retrieval by computer, the fuel sample with the highest similarity to the oil fingerprint database was a sample collected in a certain engine room in 2007.

3.3.3 *Comparison of spectral characteristics of unknown spilled oil and suspected oil*

According to the automatic retrieval results from the oil fingerprint database, a known sample that had the highest similarity with the suspected oil was found in the database. This known sample was a fuel sample that was collected from a certain engine room in 2007. The spectrograms of *n*-alkanes, PAHs, biomarkers were extracted and analyzed further based on the experience of the analyst, and the detailed features of the spilled oil sample and the known one were achieved.

From the spectra, we can see that the distribution characteristics— spectrogram of alkanes, biomarker and PAHs of the two samples were very similar; therefore, the sample in our database and the unknown sample were likely from the same source. Their relationship was calculated and analyzed further.

3.3.4 *Comparison of the diagnostic ratios of spilled oil and suspected oil with the repeatability limit*

The comparison method of the repeatability limit was recommended by the European Oil Identification System (PD CEN/TR) and had been verified by researchers (Sun et al., 2011).

(1) Brief introduction to the repeatability limit

The basic principle and comparison method of the repeatability limit were described in the literature (Sun et al., 2011; Qin, 2015). The repeatability limit means that the probability of the difference between two methods for the same sample should be less than the repeatability limit *r*, which is 95% (General Administration of quality supervision, inspection and Quarantine of the People's Republic of China, 2007). The formula of repeatability limit $r_{95\%}$ is shown as below:

$$r_{95\%} = 2\sqrt{2}r_s = 2.8r_s, \qquad (2)$$

where r_s is the relative standard deviation.

If the absolute deviation of two diagnostic ratios does not exceed the repeatability limit, the two diagnostic ratios are consistent. Generally, RSD% of diagnostic ratios used for spilled oil identification should not be more than 5%; therefore, we used this value for subsequent calculation. The sample mean was used in the formula to replace the population mean, and formula 3 was changed into

$$r_{95\%} = 2.8 \times \bar{x} \times 5\% = \bar{x} \times 14\%, \qquad (3)$$

where $\bar{x}$ is the sample mean.

For spilled oil identification, if the absolute deviation of two diagnostic ratios is less than $r_{95\%}$, then the two samples can be

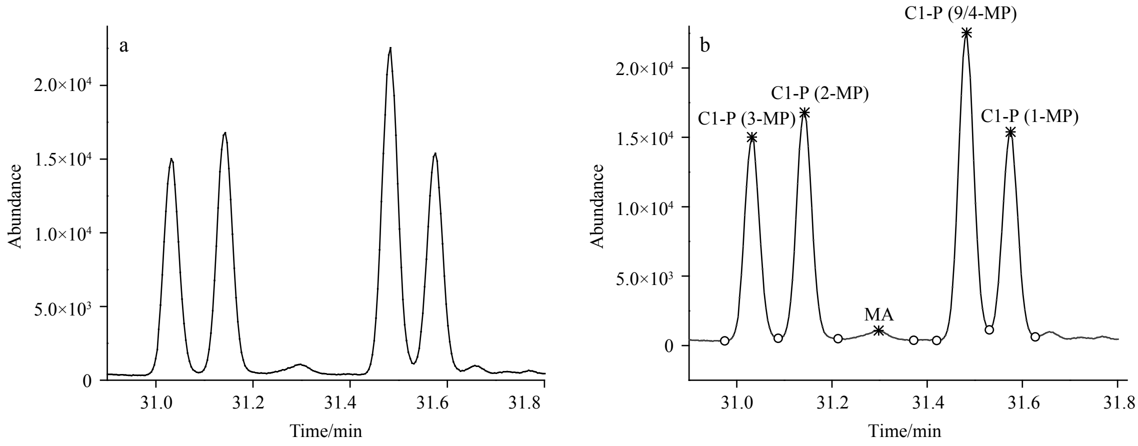

Fig. 4. GC/MS spectrum of methyl phenanthrene (*m*/*z* 192). a. The original chromatogram; and b. the redrawn one by computer. As shown in this figure, the spectrogram automatically recognized by the computer agrees well with the original one obtained by GC/MS.

Table 3. Comparison of the diagnostic ratios of PAHs and their homologues between manual and automatic integration of crude oil

Samples	Diagnostic ratio	Manual results	Automatic results	RSD/%
YY036	2-MP/1-MP	1.12	1.11	0.14
	4-MD/1-MD	3.11	3.11	0.05
	C2-dbt/C2-phe	0.28	0.29	1.72
	C3-dbt/C3-phe	0.33	0.35	5.19
	C3-dbt/C3-chr	1.20	1.14	3.43

Table 4. Comparison of the diagnostic ratios of PAHs and their homologues between manual and automatic integration of fuel oil

Samples	Diagnostic ratio	Manual results	Automatic results	RSD%
RL022	2-MP/1-MP	1.23	1.22	0.56
	4-MD/1-MD	2.07	2.06	0.39
	C2-dbt/C2-phe	1.49	1.54	2.52
	C3-dbt/C3-phe	2.15	1.98	5.64
	C3-dbt/C3-chr	10.75	11.37	3.93

Table 5. Diagnostic ratios of unknown spilled oil

Diagnostic ratios	Results
C_{17}/Pr	3.95
C_{18}/Ph	2.31
Pr/Ph	0.50
27Ts/27Tm	0.75
29αβ/30αβ	1.19
29Ts/30αβ	0.17
29βα/30αβ	0.10
30O/30αβ	0.10
30βα/30αβ	0.09
30G/30αβ	0.16
C31αβ(*S*/(*S*+*R*))	0.58
C32αβ(*S*/(*S*+*R*))	0.57
C33αβ(*S*/(*S*+*R*))	0.62
27ββ(*S*+*R*)/[28ββ(*S*+*R*)+29ββ(*S*+*R*)]	0.48
28ββ(*S*+*R*)/[27ββ(*S*+*R*)+29ββ(*S*+*R*)]	0.41
29ββ(*S*+*R*)/[27ββ(*S*+*R*)+28ββ(*S*+*R*)]	0.62
29αα*S*/29αα*R*	0.93
29ββ(*S*+*R*)/29αα(*S*+*R*)	1.17
2-MP/1-MP	1.17
4-MD/1-MD	2.10
C2-dbt/C2-phe	0.52
C3-dbt/C3-phe	0.53
C3-dbt/C3-chr	2.16

Table 6. Research results of the unknown spilled oil

No.	Information of samples	*R*
1	goal	0.98
2	RL056	0.92
3	RL049	0.92
4	RL057	0.90
5	RL058	0.90
6	RL053	0.90
7	RL060	0.89
8	YY043	0.88
9	RL093	0.88
10	RL055	0.88

considered to be the same.

(2) Screening of diagnostic ratios

Parallel samples were analyzed for all oil samples using GC/MS, then all spectrograms were integrated and the diagnostic ratios were calculated. The absolute deviations of all pairs of parallel diagnostic ratios were compared with the repeatability limit $r_{95\%}$. Those diagnostic ratios whose absolute deviations were larger than repeatability limit $r_{95\%}$ were rejected, while those diagnostic ratios whose absolute deviations were smaller than $r_{95\%}$ were used to identify the unknown spilled oil.

(3) Comparison of the diagnostic ratios

The screened diagnostic ratios were compared. If the absolute deviation between samples was larger than repeatability limit $r_{95\%}$, the oil fingerprints of two samples were different. If all absolute deviations of the diagnostic ratios were smaller than $r_{95\%}$, the oil fingerprints of two samples were the same.

According to the rule of repeatability limit, the diagnostic ratios of the unknown spilled oil and the fuel sample in our database were analyzed and are shown in Table 7.

By comparing the absolute deviation with $r_{95\%}$ of each diagnostic ratio, we can intuitively judge the consistency of the unknown spilled oil with the fuel oil sample in our database. It can be seen from Table 7 that all of the absolute deviations of the diagnostic ratios after screening were less than the corresponding $r_{95\%}$. Therefore, the oil fingerprint characteristics of the unknown spilled oil were consistent with that of the suspected oil samples. Using this comparison method, we can obtain more precise results that are consistent with the literature (Sun et al., 2011).

Through this database of diagnostic ratios of different kinds of oil, we can determine the unknown oil rapidly and accurately. It

Table 7. Absolute deviation and $r_{95\%}$ of the diagnostic ratios of unknown spilled oil and the oil in the database

Diagnostic ratios	Unknown oil	Oil in database	Absolute deviation	$r_{95\%}$
nC_{17}/Pr	3.95	4.17	0.22	0.57
nC_{18}/Ph	2.31	2.42	0.11	0.33
Pr/Ph	0.50	0.51	0.01	0.07
2-MP/1-MP	2.09	2.06	0.03	0.29
4-MD/1-MD	2.10	2.08	0.02	0.29
C2-dbt/C2-phe	0.52	0.53	0.01	0.07
C3-dbt/C3-phe	0.53	0.53	0.00	0.07
C3-dbt/C3-chr	2.16	2.25	0.09	0.31
27Ts/27Tm	0.75	0.67	0.08	0.10
29αβ/30αβ	1.19	1.23	0.04	0.17
29Ts/30αβ	0.17	0.18	0.01	0.02
30d/30αβ	0.03	0.03	0.00	0.00
29βα/30αβ	0.10	0.10	0.00	0.01
30O/30αβ	0.10	0.08	0.02	0.01
30βα/30αβ	0.09	0.10	0.01	0.01
30G/30αβ	0.16	0.15	0.01	0.02
C31αβ(*S*/(*S*+*R*))	0.58	0.58	0.00	0.08
C32αβ(*S*/(*S*+*R*))	0.57	0.57	0.00	0.08
C33αβ(*S*/(*S*+*R*))	0.62	0.63	0.01	0.09
27ββ(*S*+*R*)/[28ββ(*S*+*R*)+29ββ(*S*+*R*)]	0.48	0.51	0.03	0.07
28ββ(*S*+*R*)/[27ββ(*S*+*R*)+29ββ(*S*+*R*)]	0.41	0.40	0.01	0.06
29ββ(*S*+*R*)/[27ββ(*S*+*R*)+28ββ(*S*+*R*)]	0.62	0.61	0.01	0.09
29αα*S*/29αα*R*	0.93	0.95	0.02	0.13
29ββ(*S*+*R*)/29αα(*S*+*R*)	1.17	1.28	0.11	0.17

will be a very important application in the identification of spilled oil sources.

4 Conclusions

Based on the results and discussion, the main conclusions are listed as follows:

(1) The program is written with MATLAB software to realize the automatic recognition and integral of chromatographic peaks and calculate the diagnostic ratios automatically. The characteristic chromatographic peaks can be *n*-alkanes, biomarker compounds, PAHs, etc.

(2) Hundreds of oil samples collected worldwide were analyzed by GC/MS and diagnostic ratios were obtained. Using these diagnostic ratios, a database of oil fingerprints was established and could be used to determine the watercrafts causing the oil spill accident.

(3) The methods of automatic recognition and integration and the database were tested in a sudden marine oil spillage. The suspected sample in the oil fingerprint database was retrieved rapidly and accurately; therefore, this intelligent treating method could improve the efficiency of oil spill identification in routine enforcement.

References

Al-Kaabi N S, Kristensen M, Zouari N, et al. 2017. Source identification of beached oil at Al Zubarah, Northwestern Qatar. Journal of Petroleum Science and Engineering, 149: 107–113, doi: 10.1016/j.petrol.2016.10.034

Bayona J, Domínguez C, Albaigés J. 2015. Analytical developments for oil spill fingerprinting. Trends in Environmental Analytical Chemistry, 5: 26–34, doi: 10.1016/j.teac.2015.01.004

Christensen J H, Tomasl G, Hansen A B. 2005. Chemical fingerprinting of petroleum biomarkers using time warping and PCA. Environmental Science Technology, 39(1): 255–260, doi: 10.1021/es049832d

Dahlman G, Kienhuis P G M. 2016. Development and application of online computerized oil spill identification–COSIWeb. Standard Handbook Oil Spill Environmental Forensics (Second Edition), 729–745

Daling P S, Faksness L G, Hansen A B, et al. 2002. Improved and standardized methodology for oil spill fingerprinting. Environmental Forensics, 3(3–4): 263–278

European Committee for Standardization. 2012. CEN/TR 15522-2 Oil spill identification-Waterborne petroleum and petroleum products-Part 2: Analytical methodology and interpretation of results based on GC-FID and GC-MS low resolution analyses. Brussels, Belgium: European Committee for Standardization

General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. 2007. GB/T 21247-2007 Specifications for identification system of spilled oils on the sea (in Chinese). Beijing: China Standard Press

Qin Cuihong. 2015. Oil fingerprint digital identification methods of spilled oil. Energy Conservation & Environmental Protection in Transportation (in Chinese), (2): 51–55

Staniloae D, Petrescu B, Patroescu C. 2001. Pattern recognition based software for oil spills identification by gas-chromatography and IR spectrophotometry. Environmental Forensics, 2(4): 363–366, doi: 10.1006/enfo.2001.0060

Stouts S A, Uhler A D, McCarthy K J. 2001. A strategy and methodology for defensibly correlating spilled oil to source candidates. Environmental Forensics, 2(1): 87–98, doi: 10.1006/enfo.2001.0027

Sun Peiyan, Bao Mutai, Gao Zhenhui, et al. 2004. Identification of crude oil in Bohai Sea by gas chromatogram fingerprint. Periodical of Ocean University of China (in Chinese), 34(Sup.): 23–26

Sun Peiyan, Bao Mutai, Wang Xinping, et al. 2006. Existing state of spilled oil identification and oil fingerprint database construction at home and abroad. Journal of Xi'an Shiyou University (Natural Science Edition) (in Chinese), 21(5): 72–75, 78

Sun Peiyan, Wang Xinping, Zhou Qing, et al. 2012. Management system of fast analysis and identification of information visualization of oil fingerprint. Marine Environmental Science (in Chinese), 31(5): 729–732

Sun Peiyan, Zhao Yuhui, Cao Lixin, et al. 2011. Repeatability limit comparison method of diagnostic ratios in oil fingerprint identification. Marine Environmental Science (in Chinese), 30(1): 110–113, 117

Wang Zhendi. 2003. Fate and identification of spilled oils and petroleum products in the environment by GC-MS and GC–FID. Energy Source, 25(6): 491–508, doi: 10.1080/00908310390195570

Wang Zhendi, Fingas M. 1997. Developments in the analysis of petroleum hydrocarbons in oils, petroleum products and oil-spill-related environmental samples by gas chromatography. Journal of Chromatography A, 774(1–2): 51–78

Wang Zhendi, Fingas M, Lambert P, et al. 2004. Characterization and identification of the Detroit River mystery oil spill (2002). Chromatography, 1038(1–2): 201–214

Wang Zhendi, Fingas M, Lanadriault M, et al. 1995. Identification of alkylbenzenes and direct determination of BETX and BETX+C3-Benzenes in oils by GC/MS. Analytical Chemistry, 67(19): 3491–3500, doi: 10.1021/ac00115a018

Wang Zhendi, Fingas M, Lanadriault M, et al. 1999. Source identification of an unknown spilled oil from Quebec (1998) by unique biomarkers and diagnostic ratios of "Source-Specific Marker" Compounds. Environmental Technology, 20(8): 851–862

Wang Zhendi, Stout S. 2007. Oil Spill Environmental Forensics: Fingerprinting and Source Identification. Boston, MA: Academic Press, 344

Xu Weichao. 2012. A review on correlation coefficients. Journal of Guangdong University of Technology (in Chinese), 29(3): 12–17

Yang Baijuan, Xu Xiaoqin, Li Qingling, et al. 2008. Application of oil spill identification by GCMS-A case study. Marine Environmental Science (in Chinese), 27(6): 661–665, 670

Zhang Haijiang, Wang Chuanyuan, Zhao Ruxiang, et al. 2016. New diagnostic ratios based on phenanthrenes and anthracenes for effective distinguishing heavy fuel oils from crude oils. Marine Pollution Bulletin, 106(1–2): 58–61