Journal of Chromatography A, 1198–1199 (2008) 181–187

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Identification of sources of diesel oil spills using parallel factor analysis: A bridge between American society for testing and materials and Nordtest methods

Diako Ebrahimi, D. Brynn Hibbert*

School of Chemistry, The University of New South Wales, Sydney, NSW 2052, Australia

ARTICLE INFO

Article history:
Received 20 March 2008
Received in revised form 6 May 2008
Accepted 8 May 2008
Available online 14 May 2008

Keywords:
Diesel oil
Oil spills
PARAFAC2
Multi-way data analysis
Multi-way preprocessing
ASTM for oil spill matching
Nordtest for oil spill matching
Photo-oxidation of diesel oil

ABSTRACT

American Standards for Testing and Materials method (ASTM5739-00) and Nordtest methodology, as the two major approaches for identifying the source of spilled oils using gas chromatography–mass spectrometry (GC–MS) data, are critically compared and a new method based on multi-way parallel factor analysis (PARAFAC2) is proposed. The new approach exploits both ASTM and Nordtest methodologies by using the entire extracted ion chromatogram (EIC) and taking into account the concentration diversities of different compound classes, respectively. A multi-way data preprocessing is proposed to preserve the diagnostic properties of the original GC–MS data, which are destroyed in the ASTM method by normalizing the EICs individually. Petroleum oils, in particular diesel oils, that are difficult to classify using current methods are shown to be excellent candidates for PARAFAC2 in which EIC matrices of different sizes can be analyzed simultaneously. A diesel oil sample from an oil spill and seven very similar suspect diesel source oils, which had undergone controlled weathering for 2–15 days, were compared by this method. 79% of pairwise group comparisons were separated, in contrast to the method in which EICs were each normalized to 100, which gave 32% separation of the comparisons.

1. Introduction

Regulatory agencies such as Environmental Protection Agencies (EPAs) institute different chemical analysis methods [1–3] to identify the responsible sources of pollution among which gas chromatography–mass spectrometry (GC–MS) [2,4,5] is the method of choice. Identification of oil spills are through the examination of ‘source specific’ and ‘weathering stable’ extracted ion chromatograms (EICs) while the others may be utilized in determination of the extent of weathering and the age of spill oils [6]. The two major approaches for identification of the source of oil spills by GC–MS are ASTM5739-00 [2] and a methodology from Nordtest [5]. In the ASTM method the assessments are based on overlaying and visually comparing the EICs for the oil spill and each suspect source. Inspecting the entire EIC (all structural isomer peaks) is advantageous, but there might be a risk of subjective error associated with visual comparisons of spectra in this method, particularly in cases with highly similar EIC patterns such as diesel oils [7,8]. The concentration diversity of chemical compound classes which emerges as differences in the EIC peak heights is neglected in ASTM as a result of normalizing the EICs to a maximum of 100 individually [2]. Thus only the pattern of isomers in an EIC (related to within EIC concentration variability) are examined, not the relative concentrations of compounds from different EICs (between EIC concentration variability). The method therefore might be prone to missing matches between samples if the patterns of EICs change during biodegradation and photo-oxidation as a result of a preference of some isomers over others [5,9–14]. The bacterial degradation issue is tackled in Nordtest methodology using concentration ratios (called Diagnostic Ratios, DRs) of isomers known to be bio-resistant instead of the entire EIC peaks [5]. When the dominant weathering process is biodegradation (and evaporation and dissolution) use of the proposed DRs instead of the entire EIC is beneficial and has been successfully exploited in the assessment of biodegraded petroleum spills [6]. Photo-oxidation, which has been shown to impact hydrocarbon families in an opposite trend to biodegradation [10,13,14] is not taken into account in assigning DRs [7]. In real oil spills weathering may be a combination of evaporation, dissolution, photo-oxidation and biodegradation; As a result there would be many discrepancies between the EIC patterns of isomers, for which assigning stable DRs would be a demanding, if not impossible, task. This is particularly an issue in lighter petroleum product spills in which almost

---

* Corresponding author. Tel.: +61 2 9385 4713; fax: +61 2 9385 6141.
E-mail address: b.hibbert@unsw.edu.au (D.B. Hibbert).

0021-9673/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2008.05.016

all the hydrocarbons are degradable. Tenacious and heavy biomarkers are removed from these products (diesels for example) during refining [15,16]. Thus weathering stable DRs pertaining to biomarkers do not exist in diesel oils. For this reason, examination of the entire EIC, instead of a few bio-stable DRs, could be helpful in these cases.

While the majority of the proposed DRs in Nordtest method represent the ratios of isomers within a particular EIC, few DRs such as C2-DBT (two carbon substituted dibenzothiophene)/C2-Ph (two carbon substituted phenanthrene), C3-DBT/C3-Ph, C3-DBT/C3-Chrysene and Retene/C4-Ph [5] express the between EIC concentration diversities. This is one of the interesting aspects of Nordtest method which is not considered in the ASTM approach.

Parallel factor analysis (PARAFAC) [17–19] has attracted great attention in chemistry in recent years due to its unique properties and wide variety of applications [18–20]. With regard to the analysis and characterization of petroleum oil hydrocarbons, few applications of PARAFAC using GC–MS [21], gel permeation chromatography with UV detection [22] and fluorescence spectroscopy [23–25] have been reported.

In a recent paper [7] we demonstrated the application of a modification of PARAFAC known as PARAFAC2 [26,27] to the entire extracted ion chromatogram, as a means of discriminating among different kinds of petroleum oils (diesel, crude, transformer, lubricating oils and their mixtures). Similar diesel oils were not distinguished by exploratory data analysis but supervised classification using a variable selection by analysis of variance-least significant difference (ANOVA-LSD) [7] improved the performance of the method. In this paper a new aspect of PARAFAC2 is proposed to group, without supervision, very similar diesel oils which were not separated by the previous method. The proposed method brings together the advantages of ASTM and Nordtest methods by utilizing the entire data set (instead of only DRs) and taking into account the concentration diversity of oil compounds through a multi-way preprocessing procedure. The performance of the proposed method is compared with those of ASTM and Nordtest using the data from a real diesel oil spill and seven different artificially weathered diesel oils.

The main purpose of this study is to exploit and discover new features of the PARAFAC2 model and multi-way preprocessing that has direct application for identification of spill oils using routine instrumentation. Compared to our previous study no chromatographic improvement is achieved or aimed, instead the importance of multi-way data analysis for field data with fairly low quality is emphasized and discussed.

2. Methodology

GC–MS analysis of several samples provides mass spectrometer detector currents in a three-way tensor of size (sample × elution time × m/z channel). Using PARAFAC the GC–MS data tensor is decomposed into three loadings matrices representing sample (concentration), chromatogram and m/z profiles, respectively. To be analyzed by the PARAFAC algorithm [17], the data set has to be of a complete tensor form, i.e. the matrices (slabs) forming the data array should be of the same dimension. This is a constraint on the data and if not fulfilled the data set cannot comply with PARAFAC algorithm. Data matrices with different row or column size can be modeled simultaneously with a modification of PARAFAC known as PARAFAC2 [19,26,27]. This aspect of the PARAFAC2 method is exploited in this study to enable the analysis of EICs with different dimensions of the elution time mode which results from removing undesirable and non-diagnostic peaks from the EICs. The theory and algorithms for fitting PARAFAC and PARAFAC2 models have already been discussed by many authors [19,26,27] and will not be reiterated here, however a graphical presentation of the methods is given in Fig. 1.

Where $\mathbf{X}$ is a tensor data set, and $\mathbf{E}$ the PARAFAC (or PARAFAC2) model error. $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are the loading matrices representing the first, second and third dimension profiles, respectively. N is the number of factors explaining the significant variation of the data set and I, J and K are the size of different modes of the data. The difference between PARAFAC and PARAFAC2 is in the second mode profiles. In contrast to PARAFAC with one $\mathbf{B}$ matrix, there are K matrices of $\mathbf{B}$ ($\mathbf{B}_1$ to $\mathbf{B}_K$) with different J dimension (J_1 to J_K) in PARAFAC2. The presence of individual loading matrices for each frontal slab of K in PARAFAC2 removes the restriction induced in PARAFAC by which all the frontal slabs are bounded by one common $\mathbf{B}$ matrix. As a consequence of this flexibility, matrices with different sizes in one of the modes can be modeled simultaneously.

To take advantage of the PARAFAC2 structure, GC–MS data of diesel oils are arranged in a way that the first and third dimensions of tensor $\mathbf{X}$ are fixed and only the second dimension changes. An array in which each frontal slab holds the information of one EIC in all the samples fulfills this requirement.

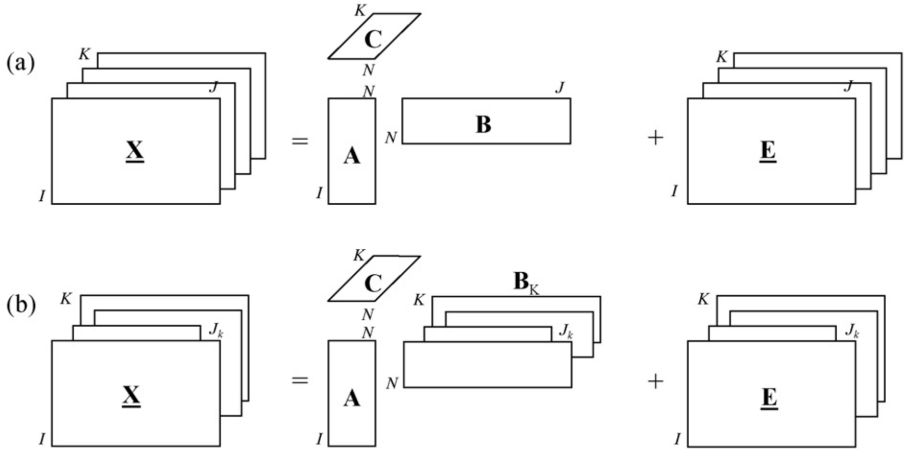

Fig. 1. Graphical presentation of (a) PARAFAC and (b) PARAFAC2.