

Independent Component Analysis (ICA) in the Deconvolution of Overlapping HPLC Aromatic Peaks of Oil

N Pasadakis, V Gaganis, P Smaragdis

TR2003-140 September 2003

Abstract

Chemometric methods like Principal Component Analysis (PCA) and Evolving Factor Analysis (EFA) have been applied to improve peaks separation, especially in HPLC UV-DAD analysis. In this work, the Independent Component Analysis (ICA) was adopted for the separation of overlapping aromatic peaks and the simultaneous determination of the underlying spectra. The application of the method on middle petroleum fractions showed an improved separation between the aromatic groups compared to the previously used methods. The determined by the ICA method UV spectra of the component groups can be further exploited for the compositional characterization of the aromatic fractions.

IMA 2003

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

INDEPENDENT COMPONENT ANALYSIS (ICA) IN THE DECONVOLUTION OF OVERLAPPING HPLC AROMATIC PEAKS OF OIL

N. Pasadakis¹, V. Gaganis¹, P. Smaragdis²

¹ Mineral Resources Engineering Department
Technical University of Crete, 73100 Chania, Greece

² Mitsubishi Electric Research Laboratories
201 Broadway, Cambridge MA 02139, USA

ABSTRACT

Chemometric methods like Principal Component Analysis (PCA) and Evolving Factor Analysis (EFA) have been applied to improve peaks separation, especially in HPLC UV-DAD analysis. In this work, the Independent Component Analysis (ICA) was adopted for the separation of overlapping aromatic peaks and the simultaneous determination of the underlying spectra. The application of the method on middle petroleum fractions showed an improved separation between the aromatic groups compared to the previously used methods. The determined by the ICA method UV spectra of the component groups can be further exploited for the compositional characterization of the aromatic fractions.

INTRODUCTION

The current trend in analytical systems is the fast production of analytical data (i.e. fast GC-MS) thus rendering the employment of chemometric methods for the deconvolution of coupled chromatographic data of high importance. Numerous approaches have been reported in the literature for the determination of composition regions and the deconvolution of coupled chromatographic data, primarily in High-Performance Liquid Chromatography-Diode array detection [1-3].

The most commonly encountered methods for the deconvolution are the Principal Component Analysis (PCA), Evolving Factor Analysis (EFA), deconvolution based on the derivatives and spectral similarity indices [4]. The comparison of these methods attributes several advantages and disadvantages to each. Therefore, the selection of the method to be applied relies on the type of data available and the experience of the analyst.

In this work, we adopt an algorithm obtained from the signal processing field for the deconvolution of coupled HPLC-UV DAD data. Following the well-known Beer's law, the problem formulation can be expressed by $X = C \times S$ (Eq. 1), where $X = [x_{ij}]$ denotes the measured spectra of the eluted mixture at the i^{th} time step and at the j^{th} wavelength. $C = [c_{ij}]$ denotes the elution profile of the i^{th} component group at the j^{th} time step, while $S = [s_{ij}]$ denotes the intensity of the i^{th} components group at the j^{th} wavelength. According to this formulation the component groups elution profiles are stored in the columns of matrix C while their spectra are stored in the rows of the S matrix. It should be noted that both matrices S and W are unknown, while X is the only matrix available.

Let n_f , n_w denote the number of fluids (time steps) and wavelengths recorded, while n_i denotes the number of components groups present in the mixture. The size of the system matrices is $X(n_f \times n_w)$, $C(n_f \times n_i)$ and $S(n_i \times n_w)$.

The ICA as well as the PCA belongs to the statistical decomposition methods. In the case present it is employed as a blind source separation (BSS) application, in a sense that the matrix X is split in two matrices without using any additional information on the kind of the eluting components [5]. For the application of the ICA, the columns of the X and S matrices are assumed to be different samples of two non-Gaussian random vectors. The method aims at finding the matrix S such that the rows of the S matrix are statistically independent, that is the components groups spectra are independent. Recall that the PCA method aims at providing spectra, which are simply orthogonal to each other.

For understanding the relation of equation (1) to the ICA method one should ignore the continuity of time and consider the fraction eluted at each time segment as a different fluid. Therefore, the number of these fluids equals that of the time steps recorded. Each fluid may contain one or more of the mixture components groups, depending on the groups overlapping.

Using the UV detector each fluid is scanned at a number of wavelengths. The larger the number of wavelengths obtained by the detector, the more information is collected regarding the fluid. Therefore the spectra of each fluid sample can be considered as a different signal which, following the notation of equation (1), is a linear combination of the signals contained in the rows of the S matrix e.g. the spectra of the components groups present in the mixture.

The basic observation on which the applicability of the ICA method is based on is that even if the spectra of any subset of the groups were known, there would be still no information for the remaining unknown spectra since they would contain entirely different components. This argument is expressed in a mathematical way by the notion of statistical independence.

It should be noted that statistical independence is in general a much stronger requirement than uncorrelatedness provided by the PCA method, which simply implies orthogonality of the groups spectra. However, in the latter case, information of a group's spectra is still hidden in the spectra of other samples. Independence requires that statistics of order higher than two are also equal to zero. From a mathematical point of view, the ICA method employs information on the distribution of the required spectra, which is not contained in the covariance matrix.

As an example consider the case of two people speaking together and been recorded by two microphones. The employment of a PCA method for splitting the recorded mixed signals in two separate signals S weighted by a constant matrix C according to equation (1), would lead to two orthogonal signals each one containing both men's voices. On the other hand, the ICA method will provide two statistically independent signals which separate clearly the mixed voices.

EXPERIMENTAL RESULTS AND DISCUSSION

The applicability of the ICA method for the deconvolution of overlapping 3-D chromatographic data was initially tested on experimental data derived from the HPLC-DAD analysis of a synthetic mixture containing three aromatic components, namely o-xylene, naphthalene and dibenzothiophene (90:9:1 % w.). The analysis was carried out using an HPLC system consisting of a Waters pump model 600 and a Waters Diode Array Ultraviolet Detector (UV-DAD) model 996. Two analytical columns, Versapack NH_2 4.1x300 mm from Alltech, connected in series were used with the column oven at 35°C and a flow rate of 1ml/min. As mobile phase, a mixture of n-

hexane and isopropyl ether (95:5 v:v) was used. Detailed description of the experimental conditions is given in [7]. The chromatogram at 254nm and the UV spectra acquired at the peak maximum for each one of the mixture components are presented in Figures 1&2. In order to acquire UV-DAD data exhibiting strong overlapping between the peaks, the mobile phase composition was changed to n-hexane and isopropyl ether (25:75 v:v). The selection was based on the relationship that exists between the resolution and the mobile phase composition presented in [1]. The derived chromatogram under the new conditions for the mixture of o-xylene, naphthalene and dibenzothiophene (90:9:1 % w.) at 254nm is presented in Figure 3, where a strong overlapping between the three components can be observed.

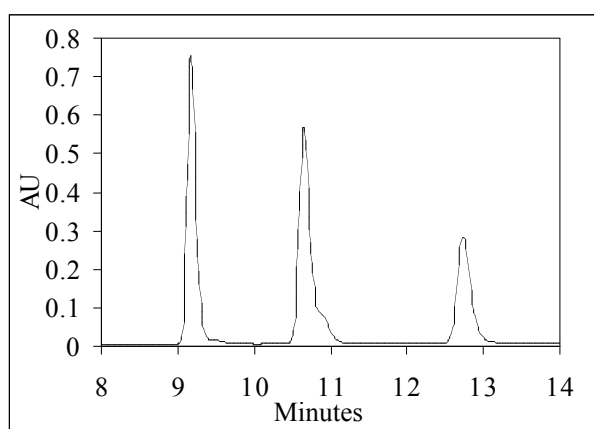


Fig1. HPLC chromatogram of the mixture

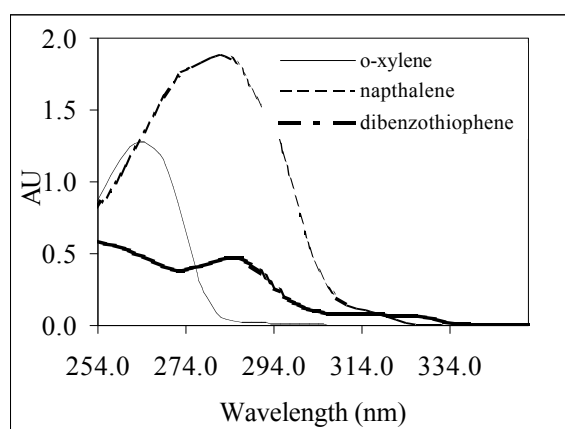


Fig2. UV spectra of the mixture components

The data matrix for the ICA analysis (X matrix) was constructed using the Millennium software package. The chromatographic signals derived for all the odd wavelengths between 254 and 352nm were integrated and the areas under the signal for time slices of 0.05min were calculated. The 21x50 matrix X contains in its rows the “spectra” over the range 254-352nm that correspond to the fraction of the mixture eluted within each time slice. The columns of X respectively represent the elution profiles of the whole mixture at each wavelength. The ICA analysis was performed using the JADE algorithm [8] in the MATLAB environment. The eigenvalues of X indicated the presense of three main features (spectra) as it was expected. Therefore three Independent Components (matrix S) were determined which are illustrated in Figure 4 together with the experimentally derived spectra of the pure components. Note that the independent components have been brought to the same scale to the experimental spectra for comparison purposes.

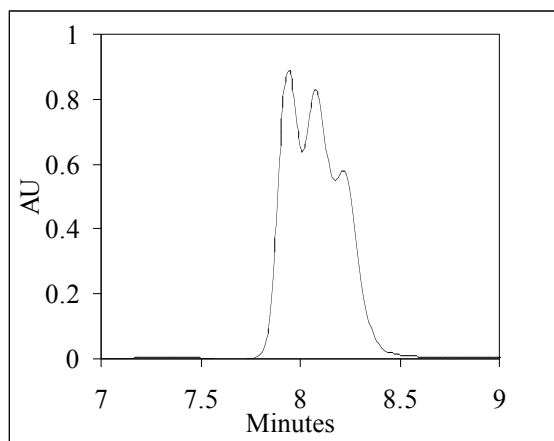


Fig3. Mixture analysis with overlapping

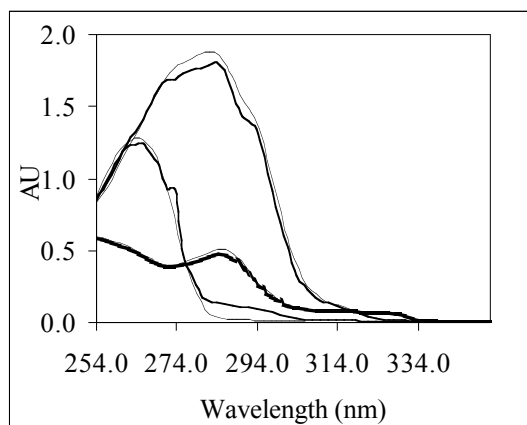


Fig4. Calculated and experimental spectra

Visual comparison of the two sets shows that there is an excellent match between the the spectra calculated by the ICA and the experimental ones. The calculated correlation coefficients for each pair of spectra were 0.996, 0.999 and 0.999 respectively. The elution profiles (matrix C) are calculated from equation (1), resulting in an excellent deconvolution of the components peaks.

The ICA method, was subsequently applied to the HPLC UV-DAD analysis of a diesel sample for the determination of the existing component groups. The chromatographic analysis was carried out using the optimized procedure described in [7]. The chromatogram at 254nm is presented in Figure 5. The 3-D UV-DAD signal was now represented by a 25x50 data matrix with the time step of the integration set at 0.4min. The eigenvalues plot of this data matrix indicated 3 significant patterns within the data and the ICA algorithm determined the corresponding spectra features (Figure 6), which are attributed to the mono-, di- and tri-aromatic components of the diesel. These profiles are in good agreement with the experimentally derived in our previous work [9] UV spectra of aromatic fractions in middle petroleum distillates. As expected, the calculated by the ICA elution profiles that correspond to the aromatic groups exhibit significant overlapping.

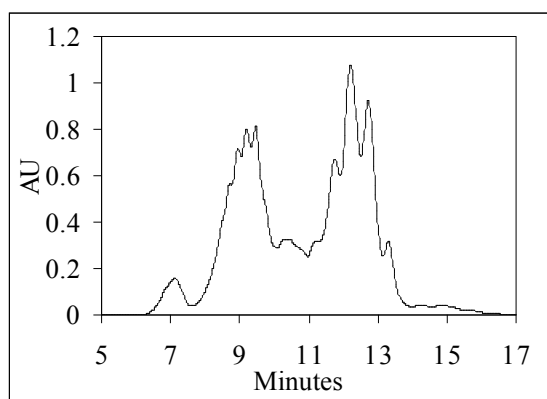


Fig5. Diesel HLPC analysis at 254nm

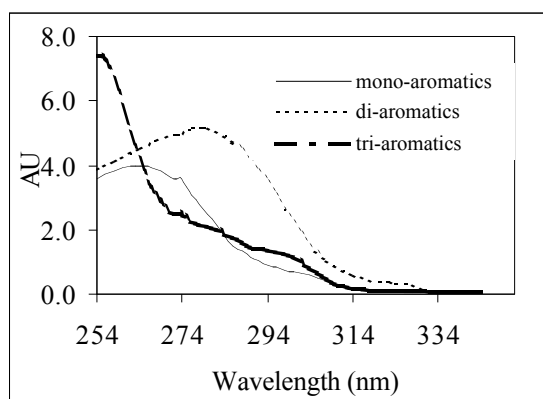


Fig6. UV spectra of the components groups

CONCLUSIONS

A new multivariate analysis method, the ICA, was applied for the interpretation of the HPLC-UV-DAD data of hydrocarbon mixtures. It was shown that the method improves significantly the separation in the analysis of synthetic mixtures, by providing very accurate estimation of the underlying spectra. The method can be also applied for the identification of component groups in complex hydrocarbon mixtures.

REFERENCES

- [1] R.G. Brereton, A.K. Elbergali, J. Chemom., (1994) 8, 423-437.
- [2] H.R. Keller, D.L. Massart, Anal. Chem. Acta, (1992) 263, 21.
- [3] R. Brereton, S. Gurden, J. Groves, Chemom. Intell. Lab. Syst. (1995), 27, 73-87.
- [4] S. Dunkerley, R. Brereton, J. Crosby, Chemom. Intell. Lab. Syst. (1999), 48, 99-119.
- [5] A. Hyvärinen, Neural Computing Surveys, (1999), Vol. 2, pp. 94-128.
- [6] P. Comon, Signal Proccesing, (1994), 36, 287-314.
- [7] N. Pasadakis, V. Gaganis, N. Varotsis, Proceedings of International Conference Instrumental Methods of Analysis IMA 2001, Vol II, 472-476.
- [8] J.F. Cardoso, A. Souloumiac. IEE Proceedings F, (1993), 140(6), 362-370
- [9] N. Pasadakis, V, Gaganis, N. Varotsis. Fuel, 2001, 80, 147-153.