第26卷，第8期
2006年8月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 26, No. 8, pp1432-1436
August, 2006

采用盲信号分离算法处理 GC-FTIR 信号

姚志湘[1,3]，黄 洪[2]，刘焕彬[1]

1. 华南理工大学制浆造纸工程国家重点实验室，广东 广州 510640
2. 华南理工大学化工学院，广东 广州 510640
3. 广西工学院生物与化学工程系，广西 柳州 545006

摘 要 针对气相色谱-红外(GC-FTIR)联用的多维数据处理，提出基于盲信号分离的色谱重叠峰分离和红外吸收谱图纯化的定性定量分析方法。优点在于充分利用联用仪器提供的大量信息，解决红外和色谱分析中有机混合物无法完全分离的难题。该方法对二甲苯的同分异构体系进行了研究，验证了理论和算法的合理性，对独立分量分析的数据进行了完整的解释，讨论了针对定量分析产生误差的原因。

主题词 气相色谱-红外联用；盲信号分离；有机混合物
中图分类号：O657.3 **文献标识码**：A **文章编号**：1000-0593(2006)08-1432-05

引 言

许多分析问题的解决都不能单靠某一种分析技术，尤其是对于复杂多组分体系问题的解决。技术的发展对于多组分体系分析的要求日益提高，仪器联用技术是解决此类问题的关键手段。与单一的分析手段相比，联用技术提供了庞大的信息量，信息从单维发展到多维。数据的处理与解释，对数据进行挖掘，得到满足分析要求的更准确、更直观的结果是化学计量学研究的重要任务。

色谱-红外联用技术结合两种分析手段的长处，优势互补，使气相色谱-傅里叶红外光谱联用仪成为一种非常高效的有机混合物结构和含量的分析手段。红外的准确定性是建立在色谱完全分离的基础上，对色谱提出了很高的要求，对难于完全分离的组分色谱红外联用的作用有限。实际上，联用仪器的数据目前并未得到充分的利用，随着化学计量学的发展，研究者在数据的深入挖掘上作了很多有益的探索[1-3]。色谱-红外联用可以提供“时间×浓度×波长吸收”的三维分析数据，多维分析数据提供了比二维数据丰富得多的信息量，从多维数据中进行深度挖掘可以有效地提高仪器的精度和适用范围，降低分析成本。盲信号分离技术是上世纪90年代提出来的信息处理技术，目的在于实现未知独立信号源的分离与识别，电信、多媒体、医疗等多个领域都在积极的研究，旨在突破各自领域内的信息分离难题[4]。仪器联用提供的大量信息可以发挥盲信号分离技术的作用。本文以二甲苯同分异构体系分析为例，对盲信号分离技术在仪器联用技术的作用进行了探讨。

1 盲信号分离与独立元分析的定义

由 n 个未知的源信号 $S_i(t)$，$i=1,\cdots,n$，构成列向量 $\boldsymbol{S}=[S_1(t), S_2(t), \cdots, S_n(t)]^{\mathrm{T}}$，其中 t 是序列值，取值为 0，l，2，…。设 A 为 $p\times n$ 维矩阵，称为混合矩阵。$\boldsymbol{Y}=[y_1(t), y_2(t), \cdots, y_p(t)]^{T}$ 是由 M 个可观察信号 $y_i(t)$，$i=1,\cdots,p$ 构成的列向量，且满足下列方程

$$\boldsymbol{Y}=\boldsymbol{AS},\ p\geqslant n \tag{1}$$

盲信号分离(BSS)的命题是，对任何 t，根据已知的 $y(t)$ 在 A 未知的条件下求未知的 $s(t)$，构成一个无噪声的盲分离问题。设 $N=[n_1(t), n_2(t), \cdots, n_p(t)]^{\mathrm{T}}$ 是由 p 个高斯白噪声信号 $n_i(t)$ 构成的列向量，且 Y 满足下列方程

$$\boldsymbol{Y}=\boldsymbol{AS}+\boldsymbol{N},\ p\geqslant n \tag{2}$$

由已知的 $\boldsymbol{Y}$ 在 $\boldsymbol{A}$ 未知时求 $\boldsymbol{S}$ 是一个有噪声盲分离问题[5,6]。独立分量分析(ICA)是解决盲信号分离的主要方法[4]。

2 色红联用信号的模型描述

气相色谱-红外联用中样品先经过色谱分离，然后进入红外分光光度计进行系列红外谱图测定，输出信号包括色谱谱图和系列的红外光谱图。如果样品里含有 n 个组分，在色谱阶段，将色谱峰离散为 p 个序列值，那么第 i 个组分在色谱

收稿日期：2005-05-28，**修订日期**：2005-08-28
基金项目：国家自然科学基金(20206008)和广西科学基金(桂科基0448010)资助
作者简介：姚志湘，1968年生，华南理工大学制浆造纸工程国家重点实验室博士后

峰中的含量可以用数组 $\boldsymbol{M}_i=[m_{i1},m_{i2},\cdots,m_{ip}]$ 表示，全部 n 个组分在色谱峰中的分布用矩阵 M 表示

$$\boldsymbol{M}=\begin{bmatrix} m_{11} & \cdots & m_{1p} \\ \vdots & m_{ij} & \vdots \\ m_{n1} & \cdots & m_{np} \end{bmatrix} \tag{3}$$

红外中，某个波数 r 下，根据朗伯-比尔定律

$$E=kcl \tag{4}$$

$$\boldsymbol{E}_r=\begin{bmatrix} E_{r1} \\ \vdots \\ E_{rp} \end{bmatrix}^{\mathrm{T}}=\begin{bmatrix} k_{r1}m_{11}+\cdots+k_{rn}m_{n1} \\ \vdots \\ k_{r1}m_{1j}+\cdots+k_{rn}m_{nj} \\ \vdots \\ k_{r1}m_{1p}+\cdots+k_{rn}m_{np} \end{bmatrix}^{\mathrm{T}}=[k_{ri},\cdots,k_{rn}]\begin{bmatrix} m_{11} & \cdots & m_{1p} \\ \vdots & m_{ij} & \vdots \\ m_{n1} & \cdots & m_{np} \end{bmatrix}=[k_{r1},\cdots,k_{rn}]\boldsymbol{M} \tag{6}$$

将整个红外吸收波段离散为从 1 到 v 个序列值，用矩阵 $\boldsymbol{E}$ 表示

$$\boldsymbol{E}=\begin{bmatrix} E_{11} & \cdots & E_{1p} \\ \vdots & E_{rj} & \vdots \\ E_{v1} & \cdots & E_{vp} \end{bmatrix}=\begin{bmatrix} k_{11}, & \cdots & ,k_{1n} \\ & \vdots & \\ k_{v1}, & \cdots & ,k_{vn} \end{bmatrix} \tag{7}$$

$\boldsymbol{K}$ 中的元素 k_{ri} 为组分 i 在波数 r 处的单位组分红外吸收值，数组 $\boldsymbol{k}_i=[k_{1i},\cdots,k_{vi}]$ 是单位含量组分在整个红外吸收中的吸收值，以序列值为横坐标，$\boldsymbol{k}_i$ 为纵坐标作图，即为组分 i 的红外吸收的吸光度谱图。

如果已知组分数目 n，并假设在红外吸收中，各组分是独立的，即各组分的吸收值符合朗伯 - 比尔定律，并能够线形相加，那么式(7)便构成了求取独立红外谱图和样品中组分含量的盲信号分离问题。其中矩阵 $\boldsymbol{K}$ 的每一列是对应每个组分的单位红外吸收谱图，矩阵 $\boldsymbol{M}$ 中的 每行代表每个色谱时刻进入红外分光光度计的各组分的相对含量。式(7)中矩阵的各元素和维数见 Scheme 1。

$$\underset{p\text{行}}{\boxed{\begin{array}{l}\boldsymbol{E}^{\mathrm{T}}\\ \text{按色谱流出序列}\\ \text{排列的红外吸光度}\end{array}}}\overset{p\text{列}}{} = \underset{p\text{行}}{\boxed{\begin{array}{l}\boldsymbol{M}^{\mathrm{T}}\\ \text{各组分含量}\\ \text{单位比值}\end{array}}}\overset{n\text{列}}{} \times \underset{n\text{行}}{\boxed{\begin{array}{l}\boldsymbol{K}^{\mathrm{T}}\\ \text{各组分的单位}\\ \text{红外吸光度谱}\end{array}}}\overset{p\text{列}}{} \tag{8}$$

Scheme 1 The matrix

3　实验试剂和仪器

采用 Perkin-Elmer System 2000 GC-IR 气相色谱-红外联用仪。色谱条件：30 m 柱径 0.32 mm SE54 非极性毛细管柱，柱温：60 ℃，FID 检测器，载气：N_2，载气流量：1 mL·s^{-1}，柱前压：0.5 MPa，毛细管分流进样，分流比 9∶1。用二甲苯(广州化学试剂厂，分析纯，二甲苯含量>80%，乙苯含量<19%)进样，对以上推导加以验证。

4　实验与计算结果

4.1　色谱和红外谱图

图 1 为二甲苯的气相色谱出峰情况。二甲苯存在四种同分异构体：邻、间、对位的二甲苯和乙苯，从色谱图上看出使用的二甲苯样品里至少含有两种同分异构体。将色谱峰均

在确定的测定条件下，吸收厚度 l 保持不变，第 j 个色谱流出时刻，进入比色的物质浓度 c 正比于色谱峰中的含量 m_j，对于多个组分同时进入比色，假设各组分独立，组分的混合不影响吸收强度，r 波数下的吸光度为

$$E_{rj}=k_{r1}m_{1j}+\cdots+k_{rn}m_{nj} \tag{5}$$

对于整个色谱峰流出，则

分成 28 个时刻，记录流出物的红外谱图。为了验证算法的有效性，红外图谱中的 CO_2 和水的背景不扣除，图 2，图 3 和图 4 分别是第 4，13，23 时刻的红外谱图。

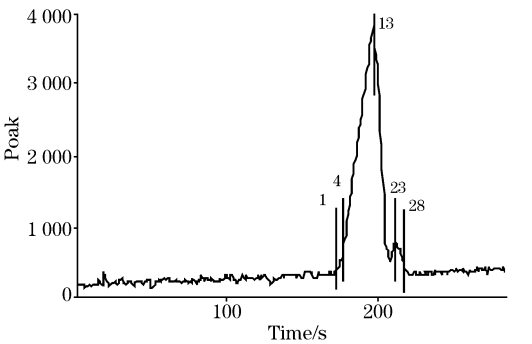

Fig. 1　Gas chromatography flows curve

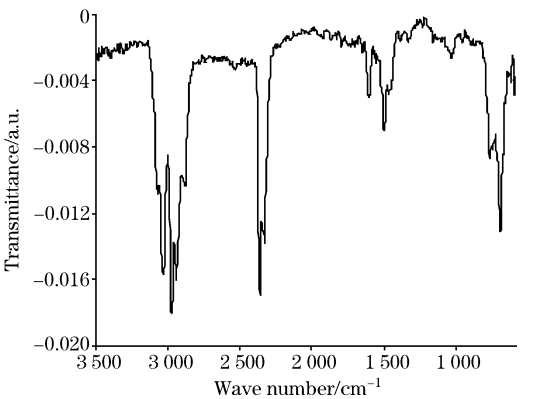

Fig. 2　Infrared spectrogram at time 4

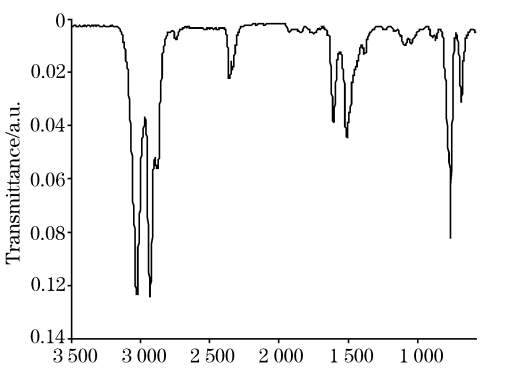

Fig. 3　Infrared spectrograms at time 13

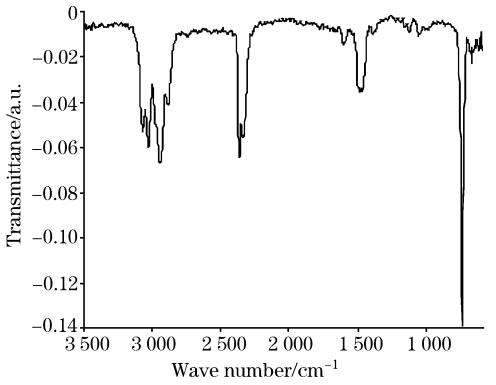

Fig. 4 Infrared spectrogram at time 23

从系列谱图得知该样品含有乙苯、间二甲苯和邻二甲苯，不含对二甲苯。只有图 3 中能大致观察到 2 000～1 600 cm^{-1}处的苯取代泛频吸收，另外两张图上对于弱吸收峰都不能准确表现。从图谱上可以估计出色谱峰中第一个大峰内包含了两种物质，即间二甲苯和乙苯，在该色谱条件下不能分离该两种物质；不完全分离的小峰内含邻二甲苯。

4.2 分辨重叠色谱峰和初步提纯红外谱图

将得到的 29 组红外吸收数据构成矩阵 ***E***，3 500～580 cm^{-1}每个波数取一个测量值，***E*** 是一个 29×2 920 的矩阵。从对谱图的初步辨识，可知进入红外光度计并引起吸收的组分有五种，分别是二甲苯的三个同分异构体、CO_2 和水，那么按五个独立分量对 ***E*** 进行 ICA 计算。

图 5 是计算出来的 5 个独立分量，对照特征吸收，可以分别辨认出依次对应的是 CO_2、邻二甲苯、乙苯、间二甲苯、水引起的不规则噪声谱带，从分量中可以看出谱图中 CO_2 引起的吸收得以去除，独立为单独的 CO_2 吸收分量。将所得分量按照波数对应重新绘于图 6、图 7 和图 8，乙苯和邻二甲苯的致特征吸收高于基线，出峰位置未发生改变，这可能是由于间二甲苯的覆盖造成的，因此还需要选择间二甲苯含量相对较低的数据对谱图进一步纯化，将在讨论中加以分析。

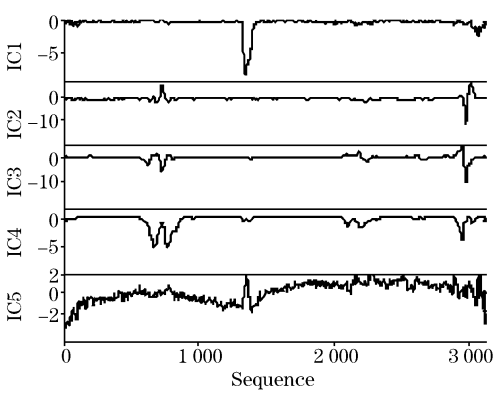

Fig. 5 ICA component computing for 5 components

计算出的 ***M*** 是一个 5×28 的矩阵，前 4 行的各元素值与前四个独立分量对应，也就是在每个色谱流出时刻各组分含量的相对值，测定中 CO_2 的绝对含量保持不变，***M*** 矩阵中各列与 CO_2 对比可计算出在整个色谱过程中各组分含量变化的情况，对比后的比值绘于图 9，在图中可以看出通过盲分离算法实现了对重叠和被覆盖色谱峰的分峰。对各列比值求总和，可计算出三种同分异构体在样品中的相对含量，其中乙苯的含量为 3.7%。

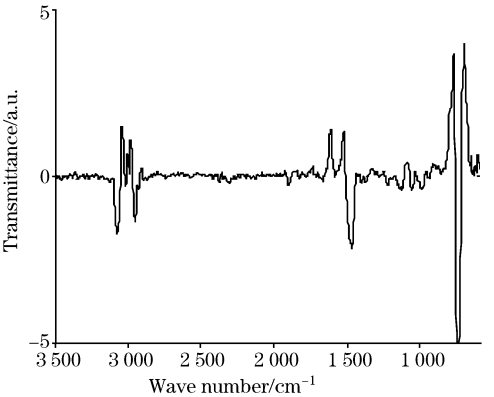

Fig. 6 ICA component corresponding to ethylbenzene

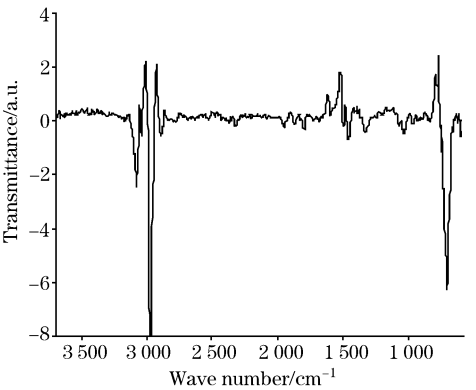

Fig. 7 ICA component corresponding to *m*-xylene

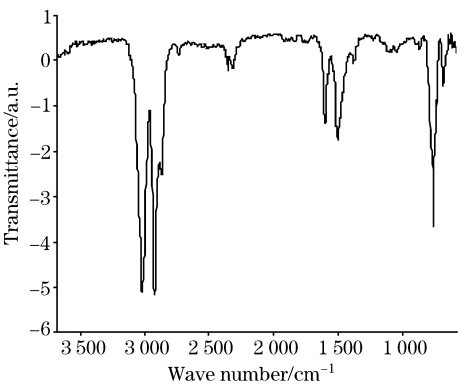

Fig. 8 ICA component corresponding to *o*-xylene

4.3 红外谱图的提纯

按照所有样本直接计算出来的分量，由于间二甲苯的含量大，同时，间二甲苯在各特征波段上的吸收都占有优势，导致其他两种物质的谱图出现吸收峰上移；从图 9 中看出前五个色谱时刻中不含邻二甲苯，乙苯，间二甲苯含量相当高，可从前五个时刻中提取“纯”乙苯的红外谱图，避免间二甲苯的掩盖；同样的 23～28 时刻邻二甲苯占优，图样的可以

提出“纯”邻二甲苯的谱图，12～17 中间二甲苯含量最大，可以最大限度的避免其他两种物质的干扰。

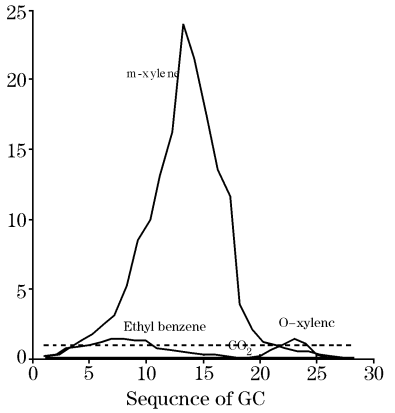

Fig. 9　Reconstruction of chromatograph flow following M matrix

图 10，图 11 和图 12 分别是从 1～5，12～17，23～28 色谱时刻中提出来与各个物质对应的独立分量，图中对比的虚线是经方差标准化以后的标准 Sadtler 图谱，标准 Sadlter 谱图取自仪器自身所带数据库。经过提纯后的谱图与标准谱图重合，基线平稳，各特征吸收峰没有出现移动和变形，其中 2 000～1 600 cm^{-1}处的苯环取代泛频可以被清晰辨识，重叠的背景吸收和 CO_2 吸收得到较好扣除。

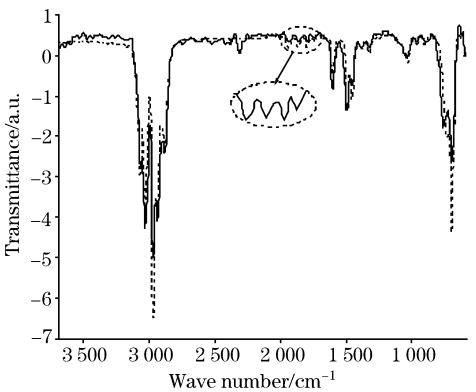

Fig. 10　Purified ethylbenzene infrared spectrograms

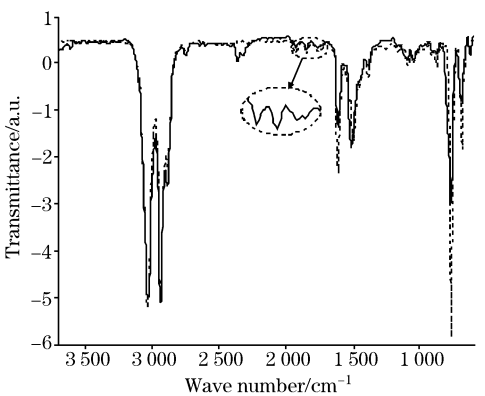

Fig. 11　Purified *m*-xylene infrared spectrograms

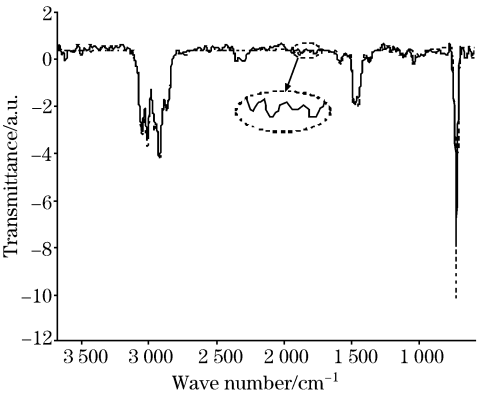

Fig. 12　Purified *o*-xylene infrared spectrograms

5　讨　论

在前面的模型分析中有三个假设，一是各组分的吸收完全符合朗伯-比尔定律的线性关系；二是各组分吸收值是各自独立的简单线性加和，没有考虑组分间的分子间力影响；三是认为各物质的红外特征吸收值彼此独立，符合 ICA 定义的统计独立要求。同时对于气相的红外吸收图谱，仪器随机误差较大。文中所用的是线性的 ICA 算法，对于不满足以上假设所带来的误差，直接采用目前的 ICA 算法无法消除，在 2 400～2 300 cm^{-1}处的 CO_2 吸收仍然有很少量的残留，但误差在可接受范围内，算法误差主要存在于定量分析和吸收较强的图谱对吸收较弱的图谱的干扰上。

ICA 算法是新发展起来的多变量统计方法，在独立性判据等方面还需要进一步的改进，在样本中如果组分间的含量存在较大差异，含量低的会受到较严重的干扰。该分析中间二甲苯的在各个特征吸收谱带上比其他物质具有更大的吸收，将分析中几种物质的 Sadtler 图谱按照相近比例模拟系列吸收图谱进行 ICA 计算，仍然会出现图 6 和图 8 中吸收基线下移的现象。从式(7)可知，$\boldsymbol{M}$ 矩阵的值和 $\boldsymbol{K}$ 矩阵相关，计算得到的红外吸收基线下移，导致 $\boldsymbol{M}$ 中间二甲苯比例较实际值高，而其他组分含量相应下降。

要解决误差问题，一方面可以改变样本的选取范围，避免与含量和吸收大的组分同时计算；另一方面，通过计算发现 ICA 计算导致固定的相对误差，通过基线校准等技术调整，由调整前后的红外谱图积分求取比例，可以用于校准 $\boldsymbol{M}$ 矩阵的定量结果。校正后间二甲苯和邻二甲苯的总和为 88%，与试剂标称一致，具体的调整校正步骤和定量分析将在另文中进行阐述。

以上分析表明，采用盲信号处理方法可以对红外谱图进行提纯，帮助实现清晰辨识，联用型色谱提供的三维色谱数据可以很好地对色谱重叠峰进行分离，改善色谱效能，盲分离算法处理多维分析数据的前景是非常令人高兴和期待的。

6　结　论

色谱-红外联用得到的多维分析数据可以被归纳为一个盲信号分离问题，通过对盲信号分离问题的求解，可以同时

实现色谱重叠峰的分离和红外谱图的“提纯”。通过对二甲苯同分异构体混合物的分析证实了这一方法的可行性，效果令人满意。

参 考 文 献

[1] HU Yun, LIANG Yi-zeng, LI Bo-yan(胡 芸, 梁逸曾, 李博岩). Acta Chimica Sinica(化学学报), 2003, 61(9): 1466.

[2] Praisler M J, Bocxlaer Van, De Leenheer A, et al. Journal of Chromatography A, 2002, 962: 161.

[3] LI Yan, WANG Jun-de, CHEN Zuo-ru, et al(李 燕, 王俊德, 陈作如, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2002, 22(5): 758.

[4] Hyvarinen A, Karhunen J, Oja E. Independent Component Analysis. Wiley and Sons Press. 2001.

[5] YANG Xing-jun, ZHENG Jun-li(杨行峻, 郑君里). Artificial Neural Network and Blind Siginal Processing(人工神经网络与盲信号处理). Beijing: Tsinghua University Press(北京: 清华大学出版社), 2003.

[6] YANG Zhu-qing, LI Yong, HU De-wen(杨竹青, 李 勇, 胡德文). Acta Automatica Sinica(自动化学报), 2002, 28: 762.

Processing GC-FTIR by the Blind Source Separation

YAO Zhi-xiang[1,3], HUANG Hong[2], LIU Huan-bin[1]

1. National Key Lab of Pulp & Paper-Making Engineering, South China University of Technology, Guangzhou 510640, China

2. College of Chemical Engineering, South China University of Technology, Guangzhou 510640, China

3. Department of Biological Chemical Engineering, Guangxi University of Technology, Liuzhou 545006, China

Abstract An analysis method for separating chromatographic overlapped peaks and purifying infrared spectra is put forward, based on the blind source separation technique and the multi-dimensional data of GC-FTIR, Using various information from hyphenated instruments, this method was used to separate completely a organic mixture, the xylene isomerism system, a problem unable to solve usually. The method can confirm the rationality of theory and algorithm and give integral explanations of the independent component analysis data. The reason for the error in quantitative analysis is discussed.

Keywords GC-FTIR; Blind source separation; Organic mixture

(Received May 28, 2005; accepted Aug. 28, 2005)