

模差渐进投影法解析重叠色谱峰

孙长海, 张舒婷, 任恒鑫, 吴洪斌, 方洪壮

(佳木斯大学 药学院 黑龙江省生物药制剂重点实验室, 黑龙江 佳木斯 154007)

摘要: 提出了一种鉴别重叠色谱峰类型的化学计量学方法——模差渐进投影(MDEP)法。该方法利用二维色谱数据中相邻两个色谱流出点的光谱矢量模的和与该两个矢量和的模之差, 建立模差-时间($dm-t$)曲线判断非纯色谱峰的纯组分区, 同时判定第二组色谱流出点和倒数第二组色谱结束点。分别以纯组分的光谱矢量建立正交阵对该色谱峰进行投影运算, 得到投影后转换色谱。该转换色谱消除了纯组分区所对应化合物的色谱和光谱信息。对转换色谱重复上述操作至整个色谱峰的模差-时间曲线均为 0 时终止, 得到重叠色谱峰各组分的光谱流出点和结束点。利用上述结果, 采用不包含待测组分光谱信息的该重叠峰二维色谱数据建立正交矩阵对该色谱峰进行投影运算, 获得待测组分的相似色谱图。通过比对分析即可得到该组分在重叠色谱峰中的光谱和色谱信息。依次操作, 即可获得重叠色谱峰中各组分的光谱和色谱信息。采用模差渐进投影法对 4 组分模拟重叠二维色谱峰及杜香挥发油 GC-MS 的重叠色谱峰进行了解析, 方法简便、结果准确。

关键词: 模差渐进投影法; 多元分辨; 重叠色谱峰

中图分类号: O657.7; R917 文献标识码: A 文章编号: 1004-4957(2010)09-0878-05

doi: 10.3969/j.issn.1004-4957.2010.09.002

Analysis of Overlapping Peak of Chromatography by Modules Difference Evolving Projection Method

SUN Chang-hai, ZHANG Shu-ting, REN Heng-xin, WU Hong-bin, FANG Hong-zhuang

(Heilongjiang Province Key Laboratory of Biological Medicine Formulation, College of Pharmacy, Jiamusi University, Jiamusi 154007, China)

Abstract The modules difference evolving projection(MDEP) method, a chemometrics method for identification of the types of chromatography was established. The curve of modules difference and time ($dm-t$) was established by computing the difference of the sum of modules and the modules of the sum of adjacent vectors in two dimension data of chromatography. It could give the purity parts and the first chromatographic point of the second component and the end chromatographic point of the second component from end in the overlapping peak. The orthogonal matrix that was structured with the spectra vector of purity parts of the first component and the last component could be projection computed with the matrix of the peak, and the changed chromatogram was gained. The changed chromatogram was continued to deal with above methods until the modules difference became zero over the peak and the first chromatographic point and the end chromatographic point of each component were gained. The orthogonal matrix that was structured without the spectra vector of the analyzed component could be projection computed with the matrix of the peak, and the similar chromatogram of the analyzed component would be gained. The spectral and chromatographic information of the analyzed component could be gained by comparing analysis. In turn, the spectral and chromatographic information of all components could be gained. The impurity peaks of four components with the data of simulated experiment and a real GC-MS data of the volatile oil in *Ledum palustre L. var. Angustum E. Busch* were analyzed. The method was simple and accurate.

Key words MDEP method; multiple analysis; chromatographic overlapping peak

收稿日期: 2010-03-28 修回日期: 2010-06-04

基金项目: 佳木斯大学基金资助项目(S2008-044)

第一作者: 孙长海(1974-), 男, 黑龙江佳木斯人, 副教授, 硕士

通讯作者: 方洪壮, Tel: 0454-8611265, E-mail: fhz-dj@163.com

高效液相色谱法是分离分析复杂物质的有力工具, 其强大的分离功能使复杂体系简单化, 但实际色谱图中仅有 20% 的色谱峰为纯组分峰, 其他绝大多数仍为重叠峰(即非纯峰)^[1]。非纯色谱峰的组成及类型的正确判断是复杂物质研究中的常见问题。目前二维色谱峰常用的分辨方法有自模式曲线分辨法、迭代目标转换因子分析法、窗口因子分析法、直观推导式演进特征投影法、正交投影分辨法、子窗口因子分析法及二维色谱的一阶微分矩阵顺序秩分析方法等^[2-15]。本文利用两个矢量模(即其 Frobenius 范数或 Euclidean 范数)的和不小于其和的模, 当且仅当这两个矢量的方向相同时等号成立的原理^[16], 建立模差-时间曲线判断色谱峰的纯组分区。以纯组分区的光谱矢量建立正交阵对待分析色谱峰进行投影运算, 再利用模差-时间曲线判断投影后的纯组分区, 重复上述操作至整个区间内的模差-时间曲线均为 0, 建立了模差渐进投影(Modules difference evolving projection, MDEP)法。利用模拟的色谱-光谱数据及杜香挥发油 GC-MS 二维重叠色谱峰数据对方法进行验证, 结果表明该方法简便、直观、准确, 可用于二维重叠色谱峰的多元分辨分析。

1 方法原理

1.1 三角不等式

根据范数的定义, 令 V 是复向量空间。函数 $\|x\|: V \rightarrow \mathbf{R}$ 称为向量 x 的范数, 若对所有 $x, y \in V$, 下面的范数公式成立:

$$\|x+y\| \leq \|x\| + \|y\| \quad \text{或} \quad \|x\| + \|y\| - \|x+y\| \geq 0 \quad (1)$$

式(1)即为三角不等式, 当且仅当向量 x 与 y 方向相同时等式成立。

1.2 模差-时间曲线

设 X 为采用联用色谱测得的二维色谱指纹图谱的 $n \times m$ 阶数据矩阵, 其行方向表征色谱的流出信息(m 个流出时间点), 列的方向表征光谱信息(n 个吸收波长); 分别对矩阵 X 的每一列光谱矢量 x_i 及 $x_i + x_{i+1}$ 进行求模运算。并按下式计算其模差:

$$dn = \|x_i\| + \|x_{i+1}\| - \|x_i + x_{i+1}\| \quad (2)$$

其中 x_i 为色谱第 i 个时间点的光谱信息(即该时间点所对应的光谱矢量)。

式(2)中, 仅当 $x_i = kx_{i+1}$, 其中 k 为常数时, 亦即第 i 点与第 $i+1$ 点流出的是同一组分时, dn 为 0。以 dn 对时间作图, 当 $dn = 0$ 时, 为纯组分区, 否则为多组分区。

1.3 投影运算

对于两个光谱列矢量 x_1 和 x_2 , 及其和矢量 $x = x_1 + x_2$, 按下式建立矩阵 M : $M = I - x_1 x_1^+$, 其中: x_1^+ 为 x_1 的广义逆, I 为单位矩阵。如果有 $M x_2 = (I - x_1 x_1^+) x_2 = p$, 则有 $M x = (I - x_1 x_1^+) x = p$, 因为 $M x = (I - x_1 x_1^+) x = (I - x_1 x_1^+) (x_1 + x_2) = (I - x_1 x_1^+) x_1 + (I - x_1 x_1^+) x_2$

根据广义逆的性质 $x_1 x_1^+ x_1 = x_1$ 及 $I x_1 = x_1$, 所以 $M x = (I - x_1 x_1^+) x = p$ 。

重叠色谱峰上的色谱重叠区的每一时间都可以看成是由光谱矢量组成的广义混合光谱, 在模差-时间曲线纯组分区任选一点所对应的光谱列矢量 x_j 按下式建立矩阵 M : $M = I - x_j x_j^+$ (3)

其中: x_j 为第 j 个组分的纯光谱矢量, x_j^+ 为 x_j 的广义逆, I 为单位矩阵。

以矩阵 X 对 M 进行投影运算得矩阵 P : $P = MX$ (4)

此时, P 矩阵是每一列均为不含 j 组分光谱信息的转换二维色谱矩阵。对矩阵 P 重复“1.2”和“1.3”的操作至全区间的 $dn-t$ 曲线均为 0, 即可判断该二维色谱峰的组分数及各组分的色谱流出点和结束点。

2 实验部分

2.1 仪器与药品

6890/5973 气相色谱-质谱仪, G1701DA 质谱工作站, NIST02 版质谱数据库(美国安捷伦公司提供)。标准挥发油提取器。

细叶杜香药材采自大兴安岭, 经佳木斯大学生药教研室宗希明高级实验师鉴定为杜鹃花科(Ericaceae)杜香属(Ledum)植物细叶杜香(*Ledum palustre* L. var. *angustum* E. Busch)的嫩枝和叶。正己烷、无水硫酸钠均为分析纯。

2.2 挥发油的提取及测定条件

称取干燥细叶杜香嫩枝和叶约 50 g 粉碎, 置于 2 000 mL 圆底烧瓶中, 加蒸馏水 400 mL 浸泡 6 h 用水蒸气蒸馏法蒸馏 5 h, 将提取的挥发油用无水硫酸钠干燥, 放入冰箱中备用。

色谱条件: SE54 石英毛细管柱 (30 m × 0.25 mm × 0.25 μm)。载气为 He, 流速 1 mL · min⁻¹。程序升温为: 起始温度 60 °C, 以 4 °C · min⁻¹ 升至 280 °C。进样口温度: 260 °C, 溶剂切断时间为 3 min。

质谱条件: 标准 EI 源, 离子源温度 230 °C, 接口温度 280 °C; 电子能量 70 eV; 扫描范围: 20~500 u, 扫描速率 3.12 次 · s⁻¹, 分流比为 20:1。

2.3 模拟数据

2.3.1 光谱数据 用计算机模拟的紫外光谱共有 4 条, 光谱数据 S 中, 每条光谱均由两个 100 个波长点的高斯曲线构成。高斯曲线参数: 均值位置 (μ)、标准差数值 (σ) 及高斯峰的高度 (h) 见表 1。

2.3.2 色谱数据 色谱模拟数据 C 由 100 个保留时间点的 4 个高斯曲线构成, 模拟色谱峰所设定的峰位所在时间点分别为 30、40、50、60 min, 标准差分别为 8、5、7、8, 色谱峰峰高分别为 0.5、0.5、0.6、0.6。模拟色谱结果如图 1A。

2.3.3 色谱-光谱数据 由以上模拟所得的色谱数据 C 和光谱数据 S 按式 $D = CS + E$ 构成模拟色谱-光谱矩阵数据。其中 C 的行数为保留时间点, 列数为色谱峰的个数; S 的行数为背景和化学成分数之和, 列数为波长点数; E 为服从正态分布的随机误差项。

2.4 GC-MS 数据

取细叶杜香挥发油 1 μL 注入气相色谱仪进行分析, 化学成分 GC-MS 检测的总离子流图见图 2A。在 9.00~9.60 min (即 1 100~1 250 色谱点范围) 的色谱峰 s 峰, 其总离子图见图 2B。通过质谱库检索, 前半部分为水芹烯, 匹配度为 65, 后半部分为松柏烯, 匹配度为 62, 色谱峰中的组分定性无法实现。因此选择此峰簇的色谱-波谱数据矩阵作为实验数据。

2.5 计算分析

分别将模拟数据及实验数据代入模差渐进投影法 MATLAB 程序, 绘制 $dm-t$ 曲线, 考察色谱峰的类型。通过直观推导式演进特征投影 (HELP) 法对实验数据进行解析, 获取纯色谱峰曲线, 与 MDEP 法的结果进行对照。

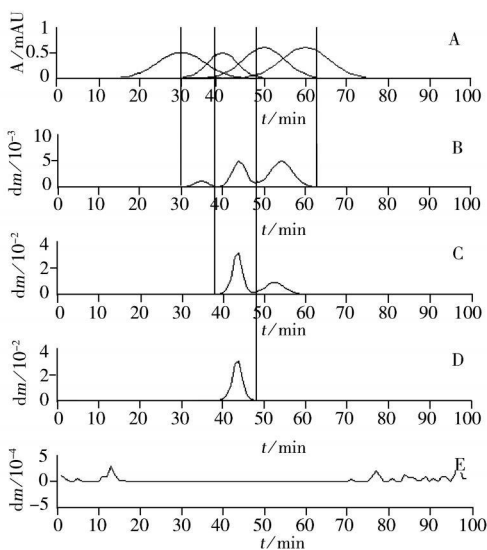
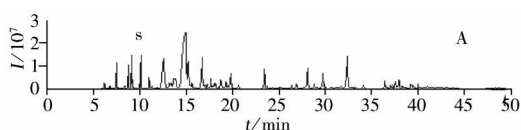


图 1 四组分的模拟色谱图 (A) 及其 $dm-t$ 曲线 (B~E)

Fig 1 Simulated chromatographic profiles (A) of four components and the curves of $dm-t$ (B~E)

A. chromatogram profiles of four simulated components B. the $dm-t$ curves of simulated chromatography C. the $dm-t$ curves of simulated chromatography after projecting with the first simulated component D. the $dm-t$ curves of simulated chromatography after projecting with the first and fourth simulated components E. the $dm-t$ curves of simulated chromatography after projecting with the first, second and fourth simulated components

图 2 细叶杜香挥发油 (A) 及峰 s (B) 的总离子流图

Fig 2 TIC chromatograms of volatile oil of *Ledum palustre* L. var. *Angustum* E. (A) and peak s (B)

3 结果与讨论

3.1 模拟数据判定

3.1.1 模拟光谱参数及结果 图 3 为模拟的光谱图。由图 3 可见, 按表 1 中 4 个模拟光谱的 6 个参数峰位 1 和 2 (μ_1, μ_2), 峰宽 1 和 2 (σ_1, σ_2), 峰高 1 和 2 (h_1, h_2) 获得 4 个各不相同的模拟光谱图。可用于构建二维模拟色谱数据。

3.1.2 模拟色谱参数及判定结果 四组分模拟数据按照模差渐进投影法绘制的 $dn-t$ 曲线结果如图 1。图中 A 为模拟色谱图, B、C、D、E 分别为投影前和依次利用第一、第二和第四组分投影后的 $dn-t$ 曲线。 $dn-t$ 曲线中 B 部分中曲线的起始点为第二个组分的色谱流出点, 结束点对应色谱峰中倒数第二个组分峰结束点。下一级的 $dn-t$ 曲线峰位起始点与结束点所对应的色谱信息判别与上一级 $dn-t$ 曲线的判别方法相同, 仅是组分峰序号加一。由此, 可得到全部组分的色谱流出点和结束点。同法反向运算, 也可得到全部组分的色谱流出点和结束点。由各 $dn-t$ 曲线可直观地判断二维重叠色谱峰的纯组分区及各组分的色谱流出点及终止点, 如图中竖线所示。

3.2 细叶杜香挥发油 GC-MS 数据判定

采用 HELP 方法对细叶杜香色谱峰 s 进行峰类型解析, HELP 方法解析结果表明, 色谱峰 s 由 4 种组分组成, 其结果见图 4A。对色谱峰 s 二维色谱数据进行模差分析, 得到 0 次投影后的 $dn-t$ 曲线, 如图 4B 所示, 可以获得色谱峰 s 第一组分 (9.00~9.07 min) 及最后一组分 (9.40~9.50 min) 的纯组分区, 第二组分色谱流出点 (9.07 min) 以及倒数第二组分色谱结束点 (9.40 min)。利用 0 次投影后的 $dn-t$ 曲线判断得到的第一组分纯组分区的光谱矢量建立正交矩阵对色谱峰 s 的二维色谱数据进行投影运算, 得到不包含第一组分信息的投影转换色谱。对该转换色谱进行模差分析, 得到 1 次投影后的 $dn-t$ 曲线, 如图 4C 所示。可以获得色谱峰 s 第一组分色谱结束点 (9.18 min) 和第三组分色谱流出点 (9.18 min)。利用 0 次投影后的 $dn-t$ 曲线判断得到的第一组分和最后一组分纯组分区的光谱矢量建立正交矩阵对色谱峰 s 的二维色谱数据进行投影运算, 得到不包含第一组分和最后一组分信息的投影转换色谱。对该转换色谱进行模差分析, 得到 2 次投影后的 $dn-t$ 曲线, 如图 4D 所示。可以获得色谱峰 s 第二组分色谱结束点 (9.27 min) 和第四组分流出点 (9.27 min)。因此, 采用 MDEP 法对实验数据判定也为 4 种组分, 其结果见图 4B~E, 各组分色谱开始点及结束点如图中竖线所示。

3.3 模拟数据及实验数据解析

根据模拟数据及细叶杜香 GC-MS 数据判断第一组分的色谱流出点和结束点, 选取从第一组分结束点到最后一组分结束点所包含的色谱二维数据, 建立正交矩阵对待分析色谱进行正交投影运算, 得到第一组分的相似色谱图。利用该相似色谱图与第一组分的纯组分区色谱数相比较得到两色谱图的比例系数, 以相似色谱图乘以比例系数, 即可得到原色谱图。以原色谱图乘以纯组分区的光谱即可得到每

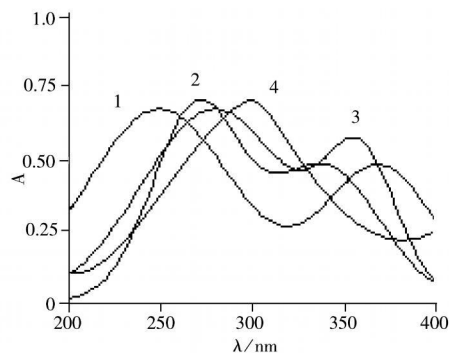


图 3 模拟组分 1~4 的紫外光谱图

Fig 3 Simulated ultraviolet spectra of the different simulated components

1. the first simulated component 2. the second simulated component 3. the third simulated component 4. the fourth simulated component

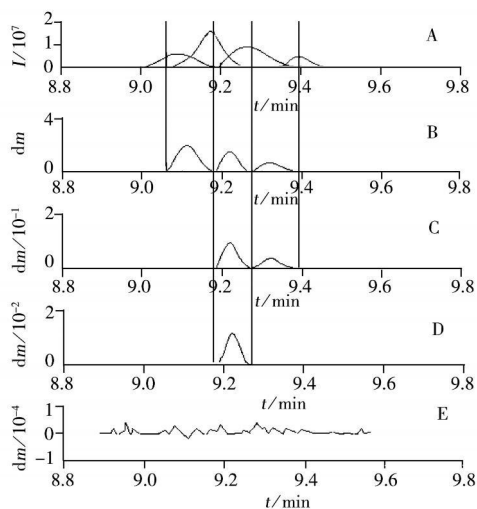


图 4 色谱峰 s 的 HELP 方法解析结果 (A) 及其 $dn-t$ 曲线 (B~E)

Fig 4 The result (A) of HELP method and the curves (B~E) of $dn-t$ for peak s

A. chromatographic profiles of four components in peak s B. the $dn-t$ curves of peak s C. the $dn-t$ curves of peak s after projecting with the first component D. the $dn-t$ curves of peak s after projecting with the first and fourth components E. the $dn-t$ curves of peak s after projecting with the first, second and fourth components

一组分纯光谱及纯色谱二维数据。以原重叠色谱峰的二维数据减去该计算得到的第一组分数数据,即可将第一组分从重叠色谱峰数据中消除,依次类推,即可获得该重叠色谱峰全部组分的纯光谱、色谱二维数据。至此,二维重叠色谱峰解析完毕。细叶杜香挥发油 GC-MS 色谱峰 *s* 的二维数据模差渐进投影 (MDEP)法和直观推导式演进特征投影 (HELP)法解析结果见图 5。图 5 表明,两种解析方法所得结果基本一致。但在判定色谱峰纯组分区时,HELP法较 MDEP法更加直观简便。MDEP法对于大峰包小峰的重叠色谱峰类型仅能给出色谱峰的类型,由于没有小峰的纯组分区,因此不能得到小峰的纯光谱及大峰的纯色谱数据,对于这种情况需进一步探讨。

4 结 论

模差渐进投影法用于二维色谱重叠峰类型的鉴别,方法简便,直观,以色谱峰整体数据作为分析对象,避免了窗口设置及迭代转换等复杂的技术步骤。程序简单,计算快速,可用于二维重叠色谱的多元分辨分析。

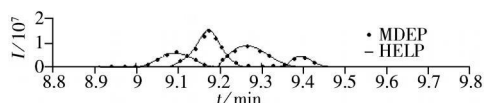


图 5 细叶杜香色谱峰 *s* 的 HELP法和 MDEP法解析结果
Fig 5 The results for peak *s* of volatile oil of *Ledum palustre* L. var. *Angustum* E. by HELP and MDEP methods

参考文献:

- [1] DAVIS JM, GIDDINGS JC. Statistical theory of component overlap in multicomponent chromatogram [J]. Anal Chem, 1983, 55(3): 418-424.
- [2] ABODOLLAH IH, SOROURADDIN MH, NASER IA H. Local resolution of two way data from multicomponent equilibria [J]. Anal Chim Acta, 2006, 562(1): 94-102.
- [3] LIH, HOU JR, WANG K. Resolution of multicomponent overlapped peaks: A comparison of several curve resolution methods [J]. Talanta, 2006, 70(2): 336-343.
- [4] MAEDER M, ZUBBUHLER A D. The resolution of overlapping chromatographic peaks by evolving factor analysis [J]. Anal Chim Acta, 1986, 181(2): 287-291.
- [5] MAEDER M. Evolving factor analysis for the resolution of overlapping chromatographic peaks [J]. Anal Chem, 1989, 61(3): 527-530.
- [6] LIANG Y Z, MANNE R, KVALHEM O M. Resolution of embedded chromatographic peaks by sequential rank analysis of the first order differentiated elution profile in the time domain [J]. Chemom Intell Lab Sys, 1994, 22(2): 229-240.
- [7] KELLER H R, MASSART D L. Evolving factor analysis [J]. Chemom Intell Lab Sys, 1991, 12(3): 209-224.
- [8] HOPKE P K. The evolution of chemometrics [J]. Anal Chim Acta, 2003, 500(2/3): 365-377.
- [9] LIANG Y Z, KVALHEM O M. The heuristic evolving latent projection: Resolving hyphenated chromatographic profiles by component stripping [J]. Chemom Intell Lab Sys, 1993, 20(2): 115-125.
- [10] 许禄, 邵学广. 化学计量学方法 [M]. 北京: 科学出版社, 2004: 200.
- [11] MANNE R, SHEN H L, LIANG Y Z. Subwindow factor analysis [J]. Chemom Intell Lab Sys, 1999, 45(1/2): 171-178.
- [12] MANNE R. On the resolution problem in hyphenated chromatography [J]. Chemom Intell Lab Sys, 1995, 27(1): 89-94.
- [13] DE JUAN A, TAULER R. Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution [J]. Anal Chim Acta, 2003, 500(1/2): 195-210.
- [14] WANG Z G, JIANG J H, DING Y J. Trilinear evolving factor analysis for the resolution of three way multicomponent chromatogram [J]. Anal Chim Acta, 2006, 558(1/2): 137-143.
- [15] DE JUAN A, TAULER R. Factor analysis of hyphenated chromatographic data: Explanation, resolution and quantification of multicomponent systems [J]. J Chromatogr A, 2007, 1158(1/2): 184-195.
- [16] 张贤达. 矩阵分析与应用 [M]. 北京: 清华大学出版社, 2004: 35.