

# 肉豆蔻精油挥发性组分数谱保留指数的预测

陈 艳

(徐州工程学院化学化工学院, 江苏 徐州 221008)

**摘 要:** 计算 42 种肉豆蔻精油挥发性组分的价分子连接性指数  $X_i$ 、电性拓扑状态指数  $E_n$  和电性距离矢量  $M_k$ 。采用最佳子集回归的方法建立这些拓扑指数与肉豆蔻精油挥发性组分数谱保留指数(RI)的定量结构-色谱保留相关性(quantitative structure-retention relationship, QSRR)模型。该模型的相关系数为 0.980, 它的计算值与相应实验值基本吻合, 平均相对误差为 2.71%, 通过 Jackknife 法和交互检验证明该模型具有总体稳健性和良好的预测能力。结果表明, 价分子连接性指数, 电性拓扑状态指数和电性距离矢量能够较好表征分子结构特征, 有效地揭示影响化合物色谱保留指数的本质因素。

**关键词:** 肉豆蔻精油; 挥发性成分; 气相色谱保留指数; 拓扑指数; 定量结构-保留相关性

## Prediction of Chromatographic Retention Index of Volatile Compounds in Nutmeg Essential Oil

CHEN Yan

(School of Chemistry and Chemical Engineering, Xuzhou Institute of Technology, Xuzhou 221008, China)

**Abstract:** The molecular connectivity index ( $X_i$ ), electrotopological state index ( $E_n$ ) and molecular electronegativity distance vector ( $M_k$ ) of 42 volatile compounds in nutmeg essential oil were calculated. A quantitative structure-retention relationship (QSRR) model for retention index (RI) of volatile compounds in nutmeg essential oil was deduced by leaps and bounds regression with the correlation coefficient of 0.980. The calculated values were in good agreement with experimental data with an average error of 2.71%. This model exhibited excellent stability and predictability through the evaluation by Jackknife method and LOO cross-validation procedure. The regression results showed that the  $X_i$ ,  $E_n$  and  $M_k$  could characterize molecular structure and explain the nature of chromatographic retention index.

**Key words:** nutmeg essential oil; volatile components; chromatographic retention index; topological index; quantitative structure-retention relationship (QSRR)

中图分类号: S666

文献标识码: A

文章编号: 1002-6630(2011)12-0274-04

肉豆蔻为肉豆蔻属植物肉豆蔻(*Myristica fragrans* Houtt.)的干燥种仁, 性温、味辛、微苦, 具有温中行气、涩肠止泻的功效, 用于脾胃虚寒、久泻不止、脘腹胀痛、食少呕吐等症<sup>[1-2]</sup>。肉豆蔻具有强烈的甜辛香, 香气浓厚、飘逸, 是东西方烹调普遍采用的辛香料, 广泛用于烘焙食品、巧克力、奶油食品、肉制品、蔬菜等的调味<sup>[3]</sup>。目前, 关于肉豆蔻的研究, 比较多的是其挥发油的提取及其成分的分析<sup>[3-6]</sup>。郑福平等<sup>[3]</sup>用 Discover 聚焦微波合成仪对微波萃取管中的肉豆蔻粉末直接加热、以净化空气吹扫, 挥发性成分吸附于活性炭再洗脱、浓缩的方法, 制得了肉豆蔻精油。对精油进行 GC-MS 分析, 采用计算机质谱库检索结合气相色谱保留指数定性的二维定性方法, 确定了 42 种挥发性成分。

定量结构-保留相关性(quantitative structure-retention relationship, QSRR)的研究与应用已成为化学、环境、生命、食品等学科研究中的一个前沿领域<sup>[7]</sup>, 利用拓扑指数对保留值进行相关性研究, 成为近年来的热门课题。QSRR 方法应用于肉豆蔻精油挥发性组分数谱保留指数的研究未见报道。在以前工作<sup>[8-12]</sup>的基础上, 本研究利用价分子连接性指数( $X_i$ )<sup>[13]</sup>、电性拓扑状态指数( $E_n$ )<sup>[14-16]</sup>和电性距离矢量( $M_k$ )<sup>[17]</sup>关联肉豆蔻挥发油 42 种挥发性化学成分的保留指数, 建立良好的 QSRR 预测模型。

## 1 数据与方法

### 1.1 数据

收稿日期: 2010-07-28

基金项目: 江苏省自然科学基金项目(09KJD150012); 徐州市科技计划研究项目(XM08C015)

作者简介: 陈艳(1968—), 女, 教授, 硕士, 主要从事污染物、药物和食品构效关系研究。E-mail: chenyan681110@126.com

本研究所采用的数据来自于文献[3], 肉豆蔻挥发油中共计 42 个结构多样的化合物, 其中, 单萜类化合物 11 种, 氧化单萜类化合物 13 种, 倍半萜类化合物 7 种, 芳香类化合物 4 种, 酯类化合物 4 种。色谱保留指数是文献作者以  $C_6 \sim C_{25}$  系列正烷烃在弱极性 VF5 MS 柱上测定, 见表 1。

用 ChemDraw Ultra 9.0 软件分别构建 42 个化合物的结构, 存为 .mol 文件, 然后在 Matlab 环境下, 调用上述分子结构, 应用文献[18-19]的方法编制程序, 计算得

到 10 种价分子连接性指数, 46 种电性拓扑状态指数及 91 种电性距离矢量, 共有 147 个描述子。根据统计学原理, 变量值的个数少于 5% 的自变量, 其对因变量的贡献可以忽略不计。因此首先对自变量集中自变量值少于  $2(42 \times 5\%)$  的自变量予以剔除, 剩余 42 个描述子来表征分子的结构。

## 1.2 方法

将每种化合物的 42 种描述子作为自变量  $X$ 、相应的色谱保留指数为因变量  $Y$  构建数据集, 应用最佳子集回

表 1 肉豆蔻精油挥发性组分的拓扑指数和色谱保留指数  
Table 1 Retention index (RI) and topologic index of volatile compounds in nutmeg essential oil

序号	化合物	$X_1$	$E_1$	$E_2$	$M_1$	$M_2$	$M_{32}$	RI		
								实验值	计算值	相对误差 /%
1	$\alpha$ -侧柏烯	7.146	7.029	2.841	3.011	6.312	0.000	924	997	7.90
2	$\alpha$ -蒎烯	7.146	7.131	2.810	3.716	6.327	0.000	934	978	4.66
3	蒎烯	6.983	4.721	4.315	6.566	5.342	0.000	948	953	0.52
4	香桉烯	6.983	4.720	4.181	3.753	6.057	0.000	976	1012	3.68
5	$\beta$ -蒎烯	6.983	4.798	4.146	4.913	5.984	0.000	981	985	0.38
6	月桂烯	6.983	4.216	2.127	3.811	15.090	0.000	988	1008	2.00
7	$\alpha$ -水芹烯	7.094	6.722	1.236	2.586	8.001	0.000	1005	1046	4.05
8	$\Delta^3$ -蒎烯	7.146	7.091	2.723	3.273	6.311	0.000	1008	992	1.56
9	$\alpha$ -松油烯	7.146	6.730	2.546	2.541	8.171	0.000	1016	1011	0.45
10	对伞花烃	6.887	6.541	0.000	2.275	9.225	0.000	1025	1047	2.16
11	柠檬烯	6.983	4.355	3.816	3.926	7.162	0.000	1030	1026	0.42
12	1,8-按叶油素	7.814	6.757	5.315	3.900	7.204	0.000	1032	1034	0.23
13	$\gamma$ -松油烯	7.146	6.730	2.358	2.436	7.043	0.000	1057	1032	2.37
14	顺式松烯水合物	7.814	6.757	5.315	3.900	7.204	0.000	1076	1034	3.87
15	$\gamma$ -松油醇	7.723	5.985	3.357	2.623	6.544	0.432	1082	1162	7.41
16	反式松烯水合物	7.723	6.595	3.552	3.187	5.216	0.740	1091	1136	4.13
17	芳樟醇	7.723	5.893	1.670	3.826	17.377	0.112	1102	1075	2.46
18	顺-对薄荷烷-2-烯-1-醇	7.671	6.338	2.035	2.629	8.994	-0.104	1108	1141	2.98
19	对乙烯基茴香醚	6.002	1.656	0.000	0.156	10.365	0.000	1132	1127	0.43
20	反-对薄荷烷-2-烯-1-醇	7.671	6.338	2.035	2.629	8.994	-0.104	1140	1141	0.09
21	樟脑	7.607	6.674	3.252	4.804	4.531	2.989	1147	1142	0.40
22	4-松油醇	7.723	6.322	2.845	2.622	6.361	-0.129	1187	1153	2.90
23	$\alpha$ -松油醇	7.723	5.985	3.357	2.623	6.544	0.432	1212	1162	4.11
24	薄荷酮	7.762	6.478	3.190	2.599	3.722	1.291	1219	1203	1.32
25	2-甲基-5-(1-甲基乙基)双环[3.1.0]己-2-醇	7.723	6.595	3.552	3.187	5.216	0.740	1243	1136	8.60
26	$\gamma$ -松油醇	7.723	5.985	3.357	2.623	6.544	0.432	1246	1162	6.73
27	乙酸龙脑酯	9.593	8.396	3.512	5.929	6.020	-0.913	1281	1329	3.74
28	黄樟油素	6.747	0.000	1.212	0.000	10.338	0.000	1292	1319	2.11
29	乙酸香茅酯	9.800	7.878	3.831	2.206	8.286	-5.635	1342	1333	0.65
30	乙酸香叶酯	9.593	7.671	2.493	2.073	8.364	-5.818	1370	1343	1.96
31	$\alpha$ -古巴烯	10.585	9.806	4.374	4.207	9.248	0.000	1374	1463	6.48
32	甲基丁香酚	8.040	3.264	0.848	0.619	13.125	0.000	1396	1402	0.41
33	$\beta$ -石竹烯	10.475	7.144	6.482	5.137	13.231	0.000	1419	1431	0.83
34	双表柏木烯	10.475	7.540	7.093	6.634	12.405	0.000	1428	1368	4.23
35	$\beta$ -金合欢烯	10.475	6.496	4.452	4.202	20.286	0.000	1455	1475	1.38
36	异石竹烯	10.475	7.144	6.482	5.137	13.231	0.000	1480	1431	3.33
37	$\gamma$ -依兰油烯	10.422	7.062	5.290	4.320	12.085	0.000	1489	1494	0.32
38	甲基异丁香酚	8.203	5.251	0.000	0.404	8.182	0.000	1494	1428	4.40
39	杜松烯	10.585	9.409	4.004	3.503	11.874	0.000	1516	1482	2.26
40	肉豆蔻醚	8.078	1.620	1.066	0.238	11.443	0.000	1526	1504	1.41
41	榄香素	9.371	4.803	0.775	1.246	13.270	0.000	1545	1590	2.90
42	肉豆蔻酸乙酯	12.509	4.120	15.605	0.028	6.513	-1.916	1791	1820	1.64

归(Leaps-and-Bounds regression)选择最佳变量组合,建立相应 QSRR 模型。以逐一剔除法(leave-one-out, LOO)交互检验判断其稳定性与预测能力,同时引入 Kubinyi 函数(Kubinyi function, FIT)<sup>[20-21]</sup>,其计算公式为:

$$\text{FIT} = \frac{R^2(y - b - 1)}{(y + b^2)(1 - R^2)} \quad (1)$$

式中:  $y$  为化合物数;  $b$  为变量数。FIT 值越大,所建的模型越稳定,预测能力越高。

## 2 结果与分析

### 2.1 QSRR 模型的构建及 LOO 交互验证

表 2 RI 与  $X_i$ 、 $E_n$  及  $M_j$  的最佳变量子集回归结果

Table 2 Results of  $X_i$ ,  $E_n$ ,  $M_j$  and RI with leaps and bounds regression

序号	$R$	$R^2$	$Q$	$Q^2$	$S$	$F$	FIT	变量
1	0.831	0.691	0.820	0.672	119.719	89.357	2.080	$X_1$
2	0.923	0.852	0.910	0.829	83.925	112.114	4.881	$X_1, M_1$
3	0.946	0.894	0.927	0.859	71.901	106.875	6.284	$X_1, M_1, E_1$
4	0.964	0.928	0.952	0.905	59.866	120.078	8.222	$X_1, M_1, E_1, E_2$
5	0.972	0.945	0.958	0.917	53.227	123.682	9.096	$X_1, M_1, E_1, E_2, M_{32}$
6	0.980	0.960	0.967	0.935	45.893	140.880	10.769	$X_1, M_1, E_1, E_2, M_{32}, M_2$
7	0.981	0.962	0.964	0.929	45.772	121.564	9.459	$X_1, M_1, E_1, E_2, M_{32}, M_2, E_4$

将肉豆蔻精油 42 个挥发性组分色谱保留指数的实验值及其 3 种拓扑指数 42 个描述子引入 SPSS 13.0 统计软件,在 95% 置信区间内,就上述化合物数据集进行多元线性统计分析,经最佳子集回归构建 QSRR 模型,见表 2。对所得的不同参数组合模型进行 LOO 交叉验证,即每次抽取 42 个化合物作为样本,剔除 1 个化合物作为被预测对象,重复 42 次,得到一组预测值,再由预测值与实验值进行线性回归,即得交叉验证系数( $Q$  及  $Q^2$ )。由表 2 可见,随着模型中变量数增多,其  $Q^2$  先是增大至 0.935,而后下降, FIT 值在同样位置出现转折点,说明六元数学模型具有最好的稳定性及预测能力。

本研究选用最佳六元相关模型作为肉豆蔻精油挥发组分色谱保留指数的预测模型:

$$\text{RI} = 88.941 + 204.476X_1 - 46.285E_1 - 33.460E_2 - 22.062M_1 - 10.520M_2 + 23.212M_{32} (n=42, R=0.980, R^2=0.960, Q=0.967, Q^2=0.935, S=45.893, F=140.880) \quad (2)$$

式中,  $n$ 、 $R$ 、 $R^2$ 、 $Q$ 、 $Q^2$ 、 $s$ 、 $F$  依次为样本数、相关系数、判定系数、交叉验证系数、估计标准误差及 Fisher 检验值。

用模型(2)计算肉豆蔻挥发性成分的色谱保留指数,列于表 1,和文献值基本吻合,平均相对误差为 2.71%。

### 2.2 模型的稳健性和预测能力的再检验

#### 2.2.1 模型中变量的相关性检验

评价模型的稳定性及是否存在自相关性,可用变异膨胀因子(variance inflation factors, VIF)<sup>[22]</sup>予以判定。如  $VIF=1$ ,表明各自变量间完全不相关;当  $VIF < 5$  时,说明变量间没有明显的自相关性,所建模型是稳定的;当  $VIF > 5$  时,说明变量间存在明显的共线性,所建模型不能用于估算与预测。

$$\text{VIF} = \frac{1}{1 - R^2} \quad (3)$$

式中:  $R^2$  为自变量  $X$  中某一变量与余下变量的判定系数。模型(2)中各自变量的自相关矩阵见表 3。

由表 3 可知,所有变量的 VIF 值均小于 5,说明变量和变量之间存在较低的自相关性,模型的稳定性较高。

表 3 模型(2)中各描述符的自相关矩阵与 VIF 值

Table 3 VIF values and self-correlation matrix described in Model 2

变量	$X_1$	$E_1$	$E_2$	$M_1$	$M_2$	$M_{32}$	$R^2$	VIF
$X_1$	1.000						0.793	4.831
$E_1$	0.413	1.000					0.623	2.653
$E_2$	0.662	0.208	1.000				0.694	3.268
$M_1$	0.162	0.555	0.285	1.000			0.526	2.110
$M_2$	0.363	-0.114	-0.080	0.007	1.000		0.518	2.075
$M_{32}$	-0.351	-0.139	-0.140	0.193	-0.064	1.000	0.215	1.274

#### 2.2.2 模型的 Jackknife<sup>[23]</sup>检验

用 Jackknife 法对模型进行稳健性再检验。每次从 42 个化合物中去掉一个化合物,用剩余化合物为建模组,根据模型(2)的形式进行回归,重复 42 次,得到 42 个相关系数,这 42 个数在 0.977~0.983 之间波动,但波动不大,从图 1 可以看出,呈良好的正态分布,平均值为 0.980,和模型(2)的相关系数一致,说明回归模型具有可接受的总体稳健性。

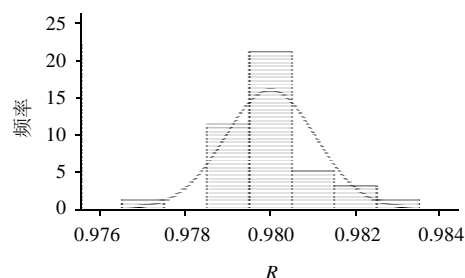


图 1 Jackknifed 相关系数直方图

Fig.1 Histogram of Jackknifed correlation coefficients

## 3 结论

影响气相色谱保留指数很多,但主要是待测组分和

固定相之间的分子间作用力。本研究用分子连接性指数, 电子拓扑状态指数和电性距离矢量 3 种拓扑指数全面地表征了肉豆蔻精油挥发性化合物的分子结构, 揭示了影响其色谱保留指数的本质因素, 所建 QSRR 模型的相关系数达到 0.980, 而且经 Jackknife 法、交互检验法验证该模型具有总体稳健性和良好的预测能力。本研究建模的思路, 可以用于其他天然产物中挥发性组分色谱保留指数的预测, 在 QSRR 研究中值得推广。

#### 参考文献:

- [1] 国家药典委员会. 中华人民共和国药典: 一部[M]. 北京: 化学工业出版社, 2000: 102.
- [2] 李秀芳, 吴立军, 贾天柱, 等. 肉豆蔻的化学成分[J]. 沈阳药科大学学报, 2006, 23(11): 698-673.
- [3] 郑福平, 孙宝国, 谢建春, 等. 无溶剂聚焦微波提取 / 气 - 质联机分析肉豆蔻挥发油[J]. 食品科学, 2007, 28(9): 484-487.
- [4] 王莹, 杨秀伟, 陶海燕, 等. 商品肉豆蔻挥发油成分的 GC-MS 分析[J]. 中国中药杂志, 2004, 29(4): 339-342.
- [5] 袁子民, 贾天柱, 王静. 肉豆蔻挥发油的研究进展[J]. 时珍国医国药, 2005, 16(12): 1201-1202.
- [6] POLITEO O, JUKIC M, MILOS M. Chemical composition and anti-oxidant activity of essential oils of twelvespiceplants[J]. Croatica Chemica Acta, 2006, 79(4): 545-552.
- [7] 余训民, 杭义萍. 多氯代二苯并呋喃的结构信息连接性指数与其在不同色谱柱上的气相色谱保留行为的关系[J]. 色谱, 2002, 20(1): 6-11.
- [8] 陈艳. 多氯代二苯并呋喃(PCDFs)在不同色谱柱上气相色谱保留行为的 QSRR 研究[J]. 计算机与应用化学, 2007, 24(7): 968-972.
- [9] 陈艳, 冯长君. 部分有机污染物气相色谱保留指数的 QSRR 研究[J]. 计算机与应用化学, 2007, 24(10): 1404-1408.
- [10] 陈艳. 单硫醚气相色谱保留指数的分子拓扑研究[J]. 武汉理工大学学报, 2007, 29(3): 82-85.
- [11] 陈艳, 堵锡华. 天然山楂香料挥发性组分定量结构-色谱保留研究[J]. 食品科学, 2009, 30(18): 270-273.
- [12] 陈艳, 堵锡华. 红梅发酵酒香气成分定量结构-色谱保留值构效关系研究[J]. 食品科学, 2009, 30(21): 39-42.
- [13] KIER L B, HALL L H. Molecular connectivity in structure-activity analysis[M]. New York: Wiley, John & Sons, 1986: 1074-1080.
- [14] HALL L H, KIER L B. Molecular similarity based on novel atom-type electrotopological state indices[J]. J Chem Inf Comput Sci, 1995, 35(6): 1074-1080.
- [15] HALL L H, KIER L B. Electrotopological state indices for atom types. A novel combination of electronic, topological, and valence state information[J]. J Chem Inf Comput Sci, 1995, 35(6): 1039-1045.
- [16] HALL L H, MOHNEY B, KIER L B. The electrotopological state: structure information at the atomic level for molecular graphs[J]. J Chem Inf Comput Sci, 1991, 31(1): 76-82.
- [17] LIU Shushen, YIN Chunsheng, LI Zhiliang. QSAR study of steroid benchmark and dipeptides based on medv-13[J]. J Chem Inf Comput Sci, 2001, 41(2): 321-329.
- [18] 胡黔楠, 梁逸曾, 王亚丽, 等. 直观队列命名法的基本原理及其在矩阵与拓扑指数计算中的应用[J]. 计算机与应用化学, 2003, 20(4): 386-390.
- [19] 张婷, 梁逸曾, 赵晨曦, 等. 基于分子结构预测气相色谱程序升温保留指数[J]. 分析化学, 2006, 34(11): 1607-1610.
- [20] SAIZ-URRA L, GONZAEZ M P, TEIJEIRA M. 2D-autocorrelation descriptors for predicting cytotoxicity of naphthoquinone ester derivatives against oral human epidermoid carcinoma[J]. Bioorganic and Medicinal Chemistry, 2007, 15(10): 3565-3571.
- [21] SAIZ-URRA L, GONZAEZ M P, TEIJEIRA M. QSAR studies about cytotoxicity of benzophenazines with dual inhibition toward both topoisomerases I and II: 3D-MoRSE descriptors and statistical considerations about variable selection[J]. Bioorganic and Medicinal Chemistry, 2006, 14(21): 7347-7358.
- [22] 冯长君, 沐来龙, 杨伟华, 等. 有机污染物的生物富集因子与拓扑指数的数学模型[J]. 物理化学学报, 2008, 24(6): 1053-1057.
- [23] DIETRICH W S, DREYER N D, HANSCH C. Confidence interval estimators for parameters associated with quantitative structure/activity relationship[J]. J Med Chem, 1980, 23(11): 1201-1205.