Vol. 36, No. 4 April 2017

#### DOI: 10.7524/j.issn.0254-6108.2017.04.2016081002

廖立敏, 李建凤, 雷光东.饮用水中挥发性有机物结构与色谱保留时间的关系[J].环境化学,2017,36(4):710-715.

LIAO Limin, LI Jianfeng, LEI Guangdong. Relationship between structure and gas chromatography retention time of volatile organic compounds in drinking water [J]. Environmental Chemistry, 2017, 36(4):710-715.

# 饮用水中挥发性有机物结构与色谱保留时间的关系\*

### 廖立敏 李建凤 雷光东\*\*

(内江师范学院化学化工学院,内江,641100)

摘 要 将有机化合物分子中的非氢原子分为 4 类,将不同非氢原子自身及非氢原子之间的关系作为结构描述符,对部分饮用水中挥发性有机物分子结构进行了参数化表达.采用逐步回归(SMR)与偏最小二乘回归(PLS)相结合的方法构建了化合物结构与色谱保留时间( $t_{\rm R}$ )之间的关系模型,模型的建模相关系数( $R^2$ )为 0.955,"留一法"交互检验的相关系数( $R^2$ )为 0.894,标准偏差(SD)为 0.803.结果表明结构描述符能较好地表征化合物分子结构特征,所建模型稳定性好、预测能力强,对于挥发性有机污染物 QSRR 研究具有一定的参考价值.

关键词 挥发性有机物,色谱保留时间,结构描述符,结构与性质的关系.

# Relationship between structure and gas chromatography retention time of volatile organic compounds in drinking water

LIAO Limin LI Jianfeng LEI Guangdong \*\*

(College of Chemistry and Chemical Engineering, Neijiang Normal University, Neijiang, 641100, China)

**Abstract:** Non-hydrogen atoms in organic molecule were grouped into four categories. Different non-hydrogen atoms and relationship between them were used as descriptors, and the structures of some volatile organic compounds in drinking water were parameterized. The stepwise regression (SMR) combining partial least squares regression (PLS) methods were employed to build the relationship model between the structure and retention time  $(t_R)$ . The modeling correlation coefficient  $(R^2)$  was 0.955, the "leave one out" cross validation correlation coefficient  $(Q^2)$  was 0.894, and the standard deviation (SD) was 0.803. The results showed that the structural descriptors well characterized the molecular structures of the compounds, and the stability and predictive power were good. This study provides certain reference value for QSRR study of volatile organic pollutants.

**Keywords**: volatile organic compounds, retention time, structure descriptor, relationship between structure and properties.

挥发性有机污染物沸点低、种类繁多,是水体中常见和重要的污染物.大多数挥发性有机物在大气中具有反应活性,是光化学烟雾的反应物<sup>[1]</sup>.大气中挥发性有机污染物危害动植物生长、威胁人类健康<sup>[2-3]</sup>,进入水体后造成饮用水污染.对饮用水中有机污染物的种类和性质进行研究具有重要意义,姚祺

<sup>2016</sup>年8月10日收稿(Received: August 10, 2016).

<sup>\*</sup>四川省教育厅青年基金(13ZB0003)和四川省科技厅应用基础项目(2015JY0077)资助.

**Supported by** the Youth Foundation of Education Bureau, Sichuan Province (13ZB0003) and Applied Basic Foundation of Technology Bureau, Sichuan Province (2015JY0077).

<sup>\* \*</sup>通讯联系人, E-mail: leigdnjtc@126.com.

等<sup>[4]</sup>采用气相色谱-质谱联用技术分析了饮用水中的挥发性有机污染物,最后在饮用水中鉴定出 50 多种有机化合物.有机化合物 QSRR 研究对于预测化合物色谱保留时间、解释保留机理、辅助鉴定化合物等具有重要意义,在有机化合物 QSRR/QSAR 研究中,首先需要将分子结构进行参数化表达,然后选用合适的方法构建结构与性质之间的关系模型,研究者们在这方面做过许多有意义的工作<sup>[5-7]</sup>.

本文选用部分饮用水中挥发性有机化合物为研究样本,以化合物非氢原子及非氢原子之间的关系构建新的结构描述符,对化合物结构进行参数化表达,采用逐步回归(SMR)与偏最小二乘回归(PLS)相结合的方法构建了化合物结构与气相色谱保留时间( $t_{\rm R}$ )之间的关系模型,对于预测挥发性有机污染物的色谱保留时间( $t_{\rm R}$ )、辅助鉴定挥发性有机污染物具有一定的参考价值.

# 1 材料与方法(Materials and methods)

#### 1.1 实验材料

选取 56 个饮用水中挥发性化合物为研究样本,化合物气相色谱保留时间以  $t_R$ 表示,实验值取自文献[4].化合物及其气相色谱保留时间( $t_R$ )列于表 1.

表 1 化合物结构描述符值及气相色谱保留时间( $t_{\rm R}$ ) **Table 1** Structural descriptors and  $t_{\rm R}$  of the compounds

		10	able 1	onuciura	descript	ors and t	R of the C	ompound				
No.	化合物 Compounds	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	<i>x</i> <sub>7</sub>	$x_8$	$t_{\mathrm{R}}$	Cal.	Err.
1	氯乙烯	4.5838	1.7321	0.0000	0.0000	0.5758	4.4926	0.0000	0.0000	1.300	0.435	-0.865
2	1,1-二氯乙烯	7.7534	0.0000	2.0000	0.0000	1.8462	0.0000	8.4381	0.0000	2.088	2.033	-0.055
3	二氯甲烷	6.3392	1.4142	0.0000	0.0000	0.6945	4.5969	0.0000	0.0000	2.177	2.882	0.705
4	反式-1,2-二氯乙烯	6.3392	3.4641	0.0000	0.0000	0.0636	7.0405	0.0000	0.0000	2.601	3.962	1.361
5	1,1-二氯乙烷	7.7534	0.0000	2.0000	0.0000	1.8462	0.0000	8.4381	0.0000	2.783	2.033	-0.750
6	氯丁二烯	4.5838	5.1962	0.0000	0.0000	0.0018	5.5444	0.0000	0.0000	3.083	3.969	0.886
7	顺式-1,2-二氯乙烯	6.3392	3.4641	0.0000	0.0000	0.0636	7.0405	0.0000	0.0000	3.255	3.963	0.708
8	溴氯甲烷	7.2448	1.4142	0.0000	0.0000	0.7319	4.9691	0.0000	0.0000	3.389	4.086	0.697
9	三氯甲烷	9.5088	0.0000	1.7321	0.0000	2.0835	0.0000	8.4451	0.0000	3.447	4.258	0.811
10	2,2-二氯丙烷	8.3392	0.0000	0.0000	2.0000	2.0712	0.0000	0.0000	8.9271	3.506	3.691	0.185
11	1,2-二氯乙烷	6.3392	2.8284	0.0000	0.0000	0.0417	5.4747	0.0000	0.0000	4.043	4.651	0.608
12	1,1,1-三氯乙烷	10.5088	0.0000	0.0000	2.0000	3.0145	0.0000	0.0000	10.9646	4.129	4.396	0.267
13	1,1-二氯丙烯	7.3392	1.7321	2.0000	0.0000	0.7596	2.4610	6.8490	0.0000	4.327	4.120	-0.207
14	四氯化碳	12.6784	0.0000	0.0000	2.0000	4.1671	0.0000	0.0000	13.0020	4.477	4.753	0.276
15	苯	0.0000	10.3923	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	4.534	5.735	1.201
16	二溴甲烷	8.1504	1.4142	0.0000	0.0000	0.7658	5.3413	0.0000	0.0000	5.101	5.295	0.194
17	1,2-二氯丙烷	7.3392	1.4142	1.7321	0.0000	0.3739	2.9287	4.4031	0.0000	5.160	5.639	0.479
18	三氯乙烯	9.5088	1.7321	2.0000	0.0000	0.8218	4.2255	7.3154	0.0000	5.213	5.953	0.740
19	二氯一溴甲烷	13.5840	0.0000	0.0000	2.0000	4.2792	0.0000	0.0000	13.5284	5.257	5.752	0.495
20	环氧氯丙烷	3.1696	4.7807	1.7321	0.0000	0.0000	2.3825	0.5375	0.0000	5.574	5.983	0.409
21	反式-1,3-二氯丙烯	6.3392	4.8783	0.0000	0.0000	0.0016	6.4587	0.0000	0.0000	6.011	5.949	-0.062
22	顺式-1,3-二氯丙烯	6.3392	4.8783	0.0000	0.0000	0.0016	6.4587	0.0000	0.0000	6.532	5.949	-0.583
23	1,1,2-三氯乙烷	9.5088	1.4142	1.7321	0.0000	0.7780	3.1762	6.1676	0.0000	6.666	7.087	0.421
24	甲苯	1.0000	8.6603	2.0000	0.0000	0.0000	0.6366	1.2131	0.0000	6.894	7.378	0.484
25	1,3-二氯丙烷	6.3392	4.2426	0.0000	0.0000	0.0009	5.5363	0.0000	0.0000	6.957	6.075	-0.882
26	一氯二溴甲烷	11.3200	0.0000	1.7321	0.0000	2.2296	0.0000	9.3567	0.0000	7.199	6.413	-0.786
27	1,2-二溴乙烷	8.1504	2.8284	0.0000	0.0000	0.0389	6.2785	0.0000	0.0000	7.472	7.139	-0.333
28	四氯乙烯	12.6784	0.0000	4.0000	0.0000	1.6436	0.0000	16.2594	0.0000	7.727	7.662	-0.065
29	1,1,1,2-四氯乙烷	12.6784	1.7321	0.0000	2.2361	2.2745	4.9307	0.0000	11.8130	8.465	6.893	-1.572
30	氯苯	3.1696	8.6603	2.0000	0.0000	0.0000	1.4604	3.2505	0.0000	8.546	8.801	0.255
31	乙苯	1.0000	10.0745	2.0000	0.0000	0.0000	0.9116	0.2707	0.0000	8.861	9.329	0.468
32	对、间二甲苯	2.0000	6.9282	4.0000	0.0000	0.0007	1.2055	2.5044	0.0000	9.119	9.012	-0.107
33	苯乙烯	1.4142	10.3923	2.0000	0.0000	0.0000	1.6926	0.4921	0.0000	9.518	9.543	0.025
34	邻二甲苯	2.0000	6.9282	4.0000	0.0000	0.0148	0.7062	3.0809	0.0000	9.598	8.926	-0.672
35	1,2,3-三氯丙烷	9.5088	2.8284	1.7321	0.0000	0.0844	5.5363	3.8900	0.0000	9.755	9.572	-0.183
36	甲乙基苯	2.0000	8.3424	4.0000	0.0000	0.0005	1.2600	1.8407	0.0000	10.093	10.917	0.824

绘丰1

											<b>绥</b> 老	ξl
No.	化合物 Compounds	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	<i>x</i> <sub>7</sub>	$x_8$	$t_{\rm R}$	Cal.	Err.
37	溴苯	4.0752	8.6603	2.0000	0.0000	0.0000	1.5610	3.7768	0.0000	10.276	9.859	-0.417
38	正丙苯	1.0000	11.4887	2.0000	0.0000	0.0000	1.0509	0.0222	0.0000	10.663	10.826	0.163
39	2-氯甲苯	3.1696	8.6603	2.0000	0.0000	0.0000	1.4604	3.2505	0.0000	10.706	8.801	-1.905
40	4-氯甲苯	3.1696	8.6603	2.0000	0.0000	0.0000	1.4604	3.2505	0.0000	10.818	8.801	-2.017
41	1,3,5-三甲苯	3.0000	5.1962	6.0000	0.0000	0.0022	1.7065	3.8740	0.0000	11.103	10.636	-0.467
42	叔丁苯	3.0000	8.6603	2.0000	2.0000	0.4060	0.1615	0.8120	3.6392	11.404	12.033	0.629
43	1,2,4-三甲苯	3.0000	5.1962	6.0000	0.0000	0.0155	1.2715	4.3764	0.0000	11.581	10.559	-1.022
44	仲丁苯	2.0000	10.0745	3.7321	0.0000	0.0111	1.1047	1.5778	0.0000	11.692	12.543	0.851
45	1,4-二氯苯	6.3392	6.9282	4.0000	0.0000	0.0000	2.9145	6.5083	0.0000	11.782	11.865	0.083
46	异丙基甲苯	3.0000	6.9282	5.7321	0.0000	0.1363	0.4094	4.2678	0.0000	11.986	12.456	0.470
47	1,2-二氯苯	6.3392	6.9282	4.0000	0.0000	0.0574	1.6006	8.0254	0.0000	12.217	11.607	-0.610
48	正丁苯	1.0000	12.9029	2.0000	0.0000	0.0000	1.0649	0.0007	0.0000	12.512	12.238	-0.274
49	1,2-二溴-3-氯丙烷	11.3200	2.8284	1.7321	0.0000	0.0801	5.9674	4.3823	0.0000	12.829	11.965	-0.864
50	1,2,4-三氯苯	9.5088	5.1962	6.0000	0.0000	0.0595	2.9177	11.4415	0.0000	14.672	14.651	-0.021
51	萘	0.0000	13.8564	4.0000	0.0000	0.0000	0.0000	0.0000	0.0000	14.967	15.471	0.504
52	六氯丁二烯	19.0175	0.0000	8.0000	0.0000	1.6538	0.0000	25.6025	0.0000	15.161	17.293	2.132
53	1,2,3-三氯苯	9.5088	5.1962	6.0000	0.0000	0.1169	1.6038	12.9586	0.0000	15.236	14.393	-0.843
54	三溴甲烷	12.2255	0.0000	1.7321	0.0000	2.2973	0.0000	9.8125	0.0000	9.119	7.499	-1.620
55	1,1,2,2-四氯乙烷	12.6784	0.0000	3.4641	0.0000	1.5560	0.0000	13.4101	0.0000	9.598	9.295	-0.303
56	1,3-二氯苯	6.3392	6.9282	4.0000	0.0000	0.0021	2.7837	6.6593	0.0000	11.692	11.846	0.154

注: Cal: 模型预测值, Err: 预测误差.

#### 1.2 实验方法

#### 1.2.1 化合物分子结构表征

有机化合物的色谱保留时间( $t_R$ )除了与测量因素有关外,还与分子的结构相关.构成化合物原子种类、数目、原子之间的连接方式等都会对 $t_R$ 产生影响.在分子结构隐氢图中,不同非氢原子及非氢原子之间的关系对化合物性质具有重要影响,而氢原子的影响可以忽略.首先将非氢原子按文献[7]方法分为4类,与1、2、3、4个其它非氢原子直接相连的非氢原子分别规定为第一、二、三、四类非氢原子,如与2个非氢原子相连的仲碳原子属于第二类非氢原子.然后在参阅文献[8]的基础上按式(1)将非氢原子进行参数化染色.

$$Z_{i} = \left[ m_{i}(n_{i} - 1) \times \frac{x_{C} - 1}{x_{i} - 1} - h_{i} \right]^{1/2}$$
 (1)

式中,i 为非氢原子在分子中的编码, $n_i$ 为非氢原子i 的主量子数, $x_i$ 为电负性, $x_c$ 为碳原子的电负性 (Pauling 标度), $m_i$ 为最外层电子数, $h_i$ 为与其直接连接的氢原子数.

对于不同类型非氢原子自身对化合物性质的影响,按式(2)进行分类累加.

$$x_k = \sum_{i=k} Z_i$$
 (k=1, 2, 3, 4) (2)

式中,k 表示非氢原子i 的原子类型, $Z_i$ 按式(1)计算.根据非氢原子的分类,对于一个有机化合物分子中最多含有4类非氢原子,因此最终可得到4个非氢原子自身对化合物性质贡献项,用 $x_1,x_2,x_3$ 和 $x_4$ 表示.

对于非氢原子之间的关系对分子性质的影响,采用 Gaussian 形距离关系函数式(3)计算,这种关系不是原子之间的具体作用方式,而是要反映出非氢原子之间的相关程度随距离增减呈反向变化以及随原子性质改变呈正向变化.

$$x_r = m_{nl} = \sum_{i \in n, i \in l} Z_i \times Z_j \times \exp(-\alpha \times d_{ij}^2) \qquad (n = 1, 2, 3, 4; n \le l \le 4)$$
(3)

Z 按式(1) 计算; $d_{ij}$ 是非氢原子 i j 之间的相对距离(即键长之和与碳碳单键键长的比值,如果 i j 之间有多条路径,则以最短的为准);n 和 l 为原子所属类型; $\alpha$  取 0.5. 化合物分子中 4 类非氢原子可以组合出以下 10 种关系项:  $m_{11}$ 、 $m_{12}$ 、 $\dots$ 、 $m_{44}$ ,简写为  $x_5$  、 $x_6$  、 $\dots$  、 $x_{14}$  .这样最多将有 14 个变量(结构描述符)来描述化合物结构信息.

#### 1.2.2 QSRR 建模与检验

对于一个样本集,并非每一个结构描述子都含与化合物保留时间相关的信息,对于那些相关性不大

的描述符即噪声,若将它们留在模型中不仅会影响回归效果,而且还会大大降低预测能力.因此,建模前对变量进行筛选就显得尤为必要,而逐步回归(Stepwise regression, SMR)是筛选变量的常用方法.偏最小二乘回归(Partial least squares, PLS)是目前较流行的一种建模方法,该方法可有效解决变量间的多重共线性问题,详细原理请参见文献[9].本文采用逐步回归(SMR)与偏最小二乘回归(PLS)相结合的方法进行建模分析,运用"留一法"对模型预测能力进行检验.一般认为,建模相关系数( $R^2$ )在0.64—1.0之间,表明模型高度相关;标准偏差(SD)与数值范围的比例小于10%时,表明模型预测准确性良好;交互检验相关系数 $Q^2 \ge 0.5$ ,表明模型具有良好的稳健性和预测能力[10].

## 2 结果与讨论(Results and discussion)

将化合物结构进行参数化表征得到结构描述符(篇幅所限,在此没有全部列出,如读者需要可以直接向作者索取),首先采用逐步回归(SMR)对变量进行筛选,然后运用偏最小二乘回归(PLS)建立模型,最终 SMR 共挑选出 12 个变量组合用于建模,变量筛选及相应的 PLS 结果见表 2.

模型	人选变量	主成分数(A)	$R^2$	$Q^2$
1	<i>x</i> <sub>3</sub>	_	_	_
2	$x_3 x_2$	1	0.706	0.682
3	$x_3 x_2 x_1$	2	0.760	0.710
4	$x_3 x_2 x_1 x_7$	3	0.899	0.757
5	$x_3 x_2 x_1 x_7 x_8$	4	0.889	0.820
6	$x_3 x_2 x_1 x_7 x_8 x_4$	5	0.925	0.881
7	$x_3 x_2 x_1 x_7 x_8 x_4 x_6$	3	0.837	0.716
8	$x_3 x_2 x_1 x_7 x_8 x_4 x_6 x_5$	7	0.955	0.894
9	$x_3$ $x_2$ $x_1$ $x_7$ $x_8$ $x_4$ $x_6$ $x_5$ $x_{12}$	7	0.959	0.889
10	$x_3$ $x_2$ $x_1$ $x_7$ $x_8$ $x_4$ $x_6$ $x_5$ $x_{12}$ $x_{11}$	5	0.884	0.807
11	$x_3$ $x_2$ $x_1$ $x_7$ $x_8$ $x_4$ $x_6$ $x_5$ $x_{12}$ $x_{11}$ $x_{10}$	5	0.893	0.834
12	$x_3 x_2 x_1 x_7 x_8 x_4 x_6 x_5 x_{12} x_{11} x_{10} x_{13}$	5	0.891	0.802

表 2 逐步回归变量筛选及 PLS 结果

Table 2 The results of stenwise variable selection and partial least squares regression

一个好的预测模型不但对内部样本具有较好的拟合能力,而且还应该对外部样本具有较强的预测能力.因此在选择模型时,在保证对内部样本具有良好拟合效果的情况下,尽量选择交互检验相关系数  $(Q^2)$  较大的模型,以确保模型具有较强的预测能力.表 2 中可以看出,应该选择由逐步回归(SMR)第八步挑选的变量组合(所选结构描述符列于表 1)建模所得模型,此时建模相关系数 $(R^2)$ 达到 0.955(接近最大值 0.959),而交互检验相关系数 $(Q^2)$ 达到最大值 0.894. $R^2$ 处于 0.64—1.0 之间、 $Q^2 \ge 0.5$ ,说明此模型高度相关、稳健性好、预测能力强.建模标准偏差(SD)为 0.803,样本的数值范围(最大值 15.236—最小值 1.300)为 13.936,而标准偏差(0.803)与数值范围(13.936)的比例为 5.76%,远小于 10%的标准,说明预测准确性良好.

图 1 为 56 个样本在 PLS 前两个主成分得分空间散点分布图,90%以上的样本点都落在 95%置信度 Hotelling T<sup>2</sup>椭圆置信圈内,说明结构描述符能够恰当表现挥发性有机化合物分子结构特征,并在统计模型中得到正确反映.为了考察样本在 X 空间的拟合情况,对样本在 X 空间的规格化模型距离作了分析(图 2),可以看到,95%以上的样本规格化模型距离都在 95%的置信区间内(只有 1 个样本超出此范围),小于临界值 2.414,同样说明模型质量良好.超出此范围的是 42 号样本,42 号样本中含有一个叔丁基,而其它化合物不含此基团,具有一定的特殊性.

图 3 为 PLS 中变量重要性(VIP)投影图,通常认为 VIP 大于 1 的变量与 Y之间具有较大的相关性,对 Y影响显著.图 3 中可以发现  $x_3$ , $x_2$ 的 VIP 值大于 1,说明这两个变量对化合物的气相色谱保留时间  $(t_R)$ 影响显著,而变量  $x_3$ 对应于第三类非氢原子的自身贡献项,变量  $x_2$ 对应于第二类非氢原子的自身贡献项,因而第三类原子、第二类原子越多,化合物可能具有较大的保留时间 $(t_R)$ 值;其次是  $x_7$ 对化合物保

留时间 $(t_R)$ 亦有一定的影响, $x_7$ 对应于第一类原子与第三类原子的关系,以上都说明取代基种类、数量及分布情况对化合物的色谱保留时间 $(t_R)$ 有重要的影响.

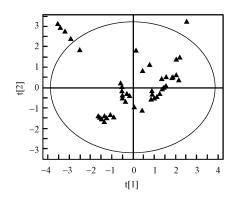


图 1 样本在前 2 个主成分得分分布

**Fig.1** Compounds' score distribution in the front two principal components

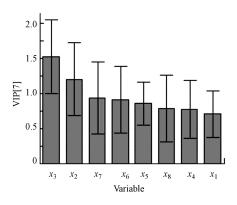


图 3 变量重要性投影

Fig.3 Importance of variables

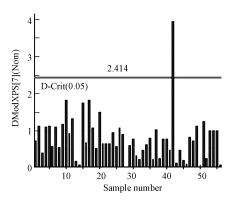


图 2 偏最小二乘 X 空间规格化模型距离

Fig.2 Normalized distance to PLS model in X space

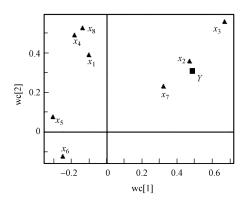


图 4 样本在 PLS 中的载荷

Fig.4 Plot of PLS loadings for compounds

图 4 为样本在前两个主成分的载荷图.图 4 可以发现  $x_3$ 、 $x_2$ 、 $x_7$ 处在图的右上方区域,即在两个主成分中都有较大的值,并且与 Y 呈正相关,同样说明  $x_3$ 、 $x_2$ 、 $x_7$ 对 Y 影响显著,与上述分析结果一致. $x_1$ 、 $x_4$ 、 $x_5$ 、 $x_8$ 在第一主成分中与 Y 呈负相关,在第二主成分中与 Y 呈正相关; $x_6$ 处在图的左下方区域,即在两个主成分中与 Y 均呈负相关. $x_1$ 、 $x_2$ 、 $x_3$  、 $x_4$  、 $x_5$  、 $x_8$  在图中所处位置之间有一定的距离,说明所选变量没有多大的相似性,变量之间共线性小.

模型对样本的气相色谱保留时间( $t_R$ )进行了预测,预测值列于表 1 的 Cal.列,误差列于 Err.列.图 5 为模型预测值与实验值之间的相关图,图 6 为预测误差分布图.

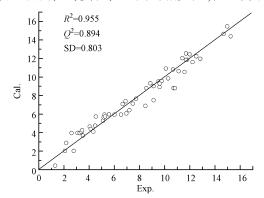


图 5 预测值与实验值相关图

Fig.5 Plot of predicted values vs. experimental values

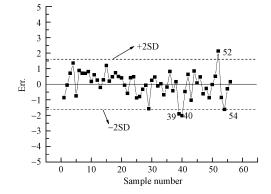


图 6 预测误差分布图

Fig.6 Plot of predicted residuals scattered

从图 5 可以看出,所有样本点都分布在 45°对角线上或者紧靠对角线,说明预测值与实验值非常接近,总体预测效果好.同样,从图 6 可以看出模型的预测准确性高、误差小,绝大部分样本的预测误差都处在±2SD 范围内,仅 4 个样本的误差略超出此范围,同样说明总体误差可以接受.其中 52 号样本误差最大,可能与其含有最多的氯原子(6 个)有关,当然也有可能实验数据本身存在较大误差从而影响计算结果.

与文献[5-6]相比,本文构建的结构描述符无需进行分子结构优化,简单易懂、计算量小;由于其基于分子二位平面结构计算得到,因而也有所不足,如不能反映分子的立体结构特征、不能区分顺反异构体、对于一些特殊的立体结构特征还难以表达,这些将在今后的研究中予以克服.

## 3 结论(Conclusion)

将分子中的非氢原子进行分类并参数化染色,将非氢原子自身及不同非氢原子之间的关系作为分子结构描述符,对部分饮用水中挥发性有机化合物结构进行了参数化表征.采用逐步回归与偏最小二乘相结合的方法(SMR-PLS)构建了化合物结构与气相色谱保留时间( $t_{\rm R}$ )的关系模型,模型相关系数( $R^2$ )及交互检验的相关系数( $Q^2$ )均较为理想,一定程度上揭示了影响化合物气相色谱保留时间( $t_{\rm R}$ )的结构因素.模型可以较准确地预测饮用水中挥发性有机化合物的气相色谱保留时间( $t_{\rm R}$ ),对于有机污染物的QSPR/QSAR 研究具有一定的参考价值.

#### 参考文献(References)

- [1] 解鑫,邵敏,刘莹,等. 大气挥发性有机物的日变化特征及在臭氧生成中的作用——以广州夏季为例[J]. 环境科学学报,2009,29 (1):54-62.
  - XIE X, SHAO M, LIU Y, et al. The diurnal variation of ambient VOCs and their role in ozone formation: Case study in summer in Guangzhou[J]. Acta Scientiae Circumstantiae, 2009, 29(1):54-62 (in Chinese).
- [2] KIMATA H. Exposure to road trafic enhances allergic skinwheal responses and increases plasma neuropetides and neurophins in patients with atopic eczema/dermatitissyndrome[J]. International Journal of Hygiene and Environmental Health, 2004, 207(1): 45-49.
- [3] HUSS MARP J, EBERLEIN KOENIG B, DARSOW U, et al. Short term exposure to volatile organic compounds enhances atopy patch test reaction[J]. Journal of Allergy and Clinical Immunology, 2004, 113(s2): 56-57.
- [4] 姚祺,沈培明. 水中 57 中 VOCs 的快速 GCMS 分析[J]. 净水技术,2016,35(s1): 83-88,112.

  YAO Q, CHEN P M. Rapid determination of 57 volatile organic compounds in drinking water by GC/MS with purge and trap sampling [J].

  Water Purifiction Technology, 2016, 35(s1): 83-88, 112 (in Chinese).
- [5] 范德玲,刘济宁,王蕾,等. 羟基自由基反应常数定量预测模型[J]. 环境化学,2015,34(10): 1924-1931. FAN D L, LIU J N, WANG L, et al. QSAR model for predicting hydroxyl radical reaction constant of organic chemicals[J]. Environmental Chemistry, 2015, 34(10): 1924-1931 (in Chinese).
- [6] SHU M, ZHANG Y R, TIAN F F, et al. Molecular docking and 3D-QSAR research of biphenyl carboxylic acid MMP-3 inhibitors[J]. Chinese J Struct Chem, 2012, 31(3): 443-451.
- [7] 李建凤,谢永红,雷光东. 部分聚合物结构与热容变化关系研究[J]. 计算机与应用化学,2016,33(7):833-837. LI J F, XIE Y H, LEI G D. Study on relationship of structure and change in heat capacity for some polymers [J]. Chemical Research and Application, 2016, 33(7):833-837 (in Chinese).
- [8] 岳玮,堵锡华. 卤代烷烃 lgS 和 lgKow 的拓扑学研究[J]. 南京工业大学学报,2006,28(4): 27-31.

  YUE W, DU X H. Topological research on the lgS<sub>w</sub> and lgK<sub>ow</sub> of halogenated hydrocarbons [J]. Journal of Nanjing University of Technology, 2006, 28(4): 27-31 (in Chinese).
- [9] 梅虎,周原,孙立力,等. 氨基酸结构描述子矢量 VHSE 及其在肽 QSAR 中的应用[J]. 化学通报,2005,68(7):534-540.

  MEI H, ZHOU Y, SUN L L, et al. A new set of descriptors of amino acids and its application in peptide QSAR[J]. Chemistry, 2005, 68

  (7):534-540 (in Chinese).
- [10] 顾云兰,陈鑫,简美玲. 苯胺类化合物结构与毒性的密度泛函理论研究[J]. 化学研究与应用,2015,27(8): 1139-1144. GU Y L, CHENG X, JIAN M L. Study on the structure-toxicity relationship of aniline compounds by density functional theory[J]. Chemical Research and Application, 2015, 27(8): 1139-1144 (in Chinese).