

有机同系物密度递变规律及有关问题的讨论

孙玉坤 孙睿鸣[#]

(海南师范学院化学系 海口 571158 [#]河北师范大学化学系 石家庄 050016)

摘 要 简述了同系线性规律研究的概况,指出了在《有机同系物密度递变规律》等文章中存在的数学原理错误。

关键词 有机同系物密度 递变规律 讨论

Discussion about the Gradation Rule of Density for Organic Homologs

Sun Yukun, Sun Ruiming[#]

(Department of Chemistry, Hainan Teachers College, Haikou 571158)

([#]Department of Chemistry, Hebei Normal University, Shijiazhuang 050016)

Abstract In this paper, study on homologs linear rules in organic chemistry have been discussed, especially some of the mathematical mistakes occurred in the literatures.

Key words Organic homologs Density, Gradation rule, Discussion

同系线性规律在物理有机化学中占有重要地位。长期以来吸引着国内外许多科学家的注意。人们寻求 $P = a + bF(n)$ 关系中, P 对 $F(n)$ 的线性关系。自 1935 年至 1971 年粗略统计, $F(n)$ 函数有 34 种之多^[1], 34 种函数中大多数是经验公式。我国也开展了这方面的研究工作, 由 1977 年至 1987 年粗略统计有 7 种同系公式, 这些公式有理论的, 也有经验的, 比早期公式线性相关性好, 应用范围也较宽。

一个值得注意的问题是: 有些研究一味追求所谓的线性相关性好, 而忽略了基本的数学原理, 例如发表在《中国科学》B 辑上的论文《有机同系分子均衡化学势与电离能》^[2]和在《化学学报》的论文《一个新的有机电离能拟合公式》^[3]就是如此, 笔者曾对其基本数学原理错误进行了讨论^[4]。

《化学通报》上的论文《有机同系物密度递变规律》^[5]、《正烷烃密度与温度的定量关系》^[6], 存在着比文献[2]、[3]更大的原理上的错误, 把一个不存在普遍规律的式子, 处理成线性相关性非常好的回归方程, 文献[5]的标题和计算内容不相符合, 现在讨论如下, 希望起到抛砖引玉的作用。

1 文献[5]、[6]在数学原理上的错误

最基本的直线方程 $y = a + bx$, x 是自变量, y 是函数, y 是纵坐标, x 为横坐标, y 对 x 作

孙玉坤 男, 58 岁, 教授, 主要从事有机化合物同系线性规律研究。

海南省教育厅资助项目(HjKj200015)

图是直线。后来发展出来在数理统计中称作一元非线性回归^[7], 即 $y = a + b F(x)$, 作图时横坐标用 $F(x)$, 得到 y 对 $F(x)$ 的直线。在同系线性规律的研究中就是用的这种一元非线性回归。 $P=a+bF(n)$, P 是函数, 是物理、化学实验测定的有关物理量, 例如: 吸收峰波数, 电离势等等。无论是早期的 34 种同系函数, 还是近期在我国出现的 7 种同系函数都遵守这一原则。但是文献[5]不是这样, 它的基本回归方程为: $M/D = a + bn$, 得到 M/D 与 n 的线性关系。在这个式子中 M 同时也是 n 的函数。例如文献[5]讨论的饱和脂肪烃取代物:

$X - (\text{CH}_2)_n - Y$, $M = M_x + 14n + M_y$, 文献[5]的回归方程就变成 $\frac{M_x + 14n + M_y}{D} = a + bn$, M_x 和 M_y 对于一个具体系列是常数, 例如对文献[5]表 1 的第一个系列 $M_x=1$ 、 $M_y=1$, 那么回归方程为 $\frac{14n+2}{D} = a + bn$ 。左边有 n , 右边也有 n , 谁对谁回归? 文献[6]更明显, 其回归式为 $n/a = a + bn$ 、 $n/B = c + dn$ 。用这种式子回归, 不是纯粹函数对自变量回归, 它包函着相当重要成份的自变量对自变量回归。

表 1 本文分离变量后回归结果与文献[5]回归结果对比

X—(CH ₂) _n —Y			回归 点数	本文分离变量后结果			文献[5]结果		
X	Y	n		a	b	r	a	b	r
H	H	5~17,19~24,26~32,35	27	0.8583	-25.4814	-0.9943	31.7192	16.3072	1.0000
H	CHCH ₂	2,4~6,8~11,13~16	12	0.8699	-24.9004	-0.9987	60.2204	16.1650	1.0000
H	CHCl ₂	1~3,6	4	0.8329	29.2105	0.9947	66.3695	17.0264	0.9982
H	NH ₂	1~18	18	0.8564	-11.0729	-0.9603	35.4567	16.0174	0.9987
i-Pr	i-Pr	0~4	5	0.8480	-25.0265	-0.9873	131.1029	16.4196	0.9994
H	CCET	2~6,8	6	0.8458	-13.3289	-0.9568	79.7732	16.5183	0.9994
Br	Br	1~5	5	0.8535	114.0107	0.9990	52.7043	16.6013	0.9994
H	CO ₂ Me	1~5,8,10	7	0.8285	7.5051	0.9548	63.0557	16.7717	0.9994
H	CO ₂ H	1~9,17	10	0.8270	12.3120	0.9940	38.0933	17.4798	0.9997
H	CHCL Me	1~3,5	4	0.8761	-0.9245	-0.3697	74.4206	16.0807	0.9998
Cl	Cl	1~6	6	0.8773	28.7916	0.9972	47.4401	16.1586	0.9998
H	i-Pr	2~6	5	0.8790	-29.8225	-0.9989	83.8473	15.9652	0.9999
线性相关系 数 r 平均值						-0.1105			0.9995

文献[5]、[6]的错误在于违背了基本数学原理, 不分离变量就进行线性回归, 即函数和自变量没有分开就回归。看一看分离变量后会有什么结果, 分离变量的方法参照文献[4]。分离变量后, 由文献[5]表 1 中选了 12 个 r 较差的数据系列进行回归分析(r 较差便于比较, 随机抽取), 计算结果列于本文表 1 中。对于文献[6]的处理列于本文表 2 中。

2 结果讨论

2.1 文献[5]、[6]数学错误的进一步讨论

线性回归方程 $y=a+bx$, 实际是 y 的变化是否与 x 变化同步的问题(以 b 为系数)。 y 以 bx 变化, 线性相关系数 r 就是 1, 否则偏离 1。现在文献[5]和[6]不分离变量就线性回归, 等于用 $y_x = a + bx$ 进行线性回归。左边实验值 y 乘以自变量 x , y_x 对 x 进行线性回归, 当然比 y 对 x 线性回归更容易同步了。为了证实这个问题, 本文作者选用了文献[4]表 1 第一个和第二个数据系列, 线性回归方

程用

表 2 本文分离变量后回归结果与文献[6]的对比

本文分离变量后回归结果		文献[6]的结果	
对文献[6]中式(6)分离变量	对文献[6]式(7)分离变量后	式(6)	式(7)
a 0.87009	c 6.042×10^{-4}	a 1.830560	c -0.550207
b -1.64267	d 2.335×10^{-4}	b 1.151970	d 0.172787
r -0.9986	r 0.9405	r 1.000	r 0.9995
回归点数 33	回归点数 33	回归点数 33	回归点数 33

文献[8]的公式 $E = a + b \frac{2k-1}{N+\frac{1}{2}}$ $N = n + t$, 在这里 $F(n) = \frac{2k-1}{n+t+\frac{1}{2}}$, 线性相关系数分别为:

$r_1 = 0.930$ $r_2 = 0.925$, 属于 Hammett 方程标准的中。用 $E \times \frac{2k-1}{N+\frac{1}{2}} = a + b \frac{2k-1}{N+\frac{1}{2}}$ 进行线性回归,

线性相关系数: $r_1 = 1.000$ $r_2 = 1.000$ 属于 Hammett 方程标准的优了(用其它方程结果相似)。从这里可以充分看出不分离变量而回归所引起的严重错误。

2.2 表 1 对比的结果

由表 1 的对比可以看到, 有机同系物密度与重复度之间的回归方程, 分离变量后, 线性相关系数 r 有正有负, 本文共计算 12 个系列, 5 个系列 r 为正值, 7 个系列 r 为负值。由表 1 中选取了两个系列, 将其密度 D 随结构单元 n 变化列于表 3 中, 其中 $H-(CH_2)_n-NH_2$ 系列线性相关系数 $r = -0.9603$, 代表 r 为负值的系列。另一系列 $H-(CH_2)_n-COOH$ 线性相关系数 $r = 0.9940$, 代表 r 为正值系列的系列。数据来自文献[9], 和文献[5]的密度数据出处相同。由表 3 的数据可以看出, 对于 $H-(CH_2)_n-NH_2$ 系列随着 n 的增大密度 D 在增大, 而 $H-(CH_2)_n-COOH$ 系列, 随着 n 的增大密度 D 在减小, 两类系列密度变化趋势正好相反。用任何一个同系规律来进行线性回归都可以得到两种不同正负号的线性相关系数。密度 D 随着 n 的变化而发生的不同倾向的变化是客观存在的。 r 为正值系列, 随着 n 的增大, 密度在减小; 而为负值系列, 随着 n 的增大, 密度在增大。 r 由正值变为负值, 一定要经过零点, 即有些系列是不相关的。实际上本文表 1 的数据中就有 $r = -0.3697$ 的系列。并且 $\bar{r} = -0.1105$ 。由数理统计理论来讲, 一个规律要么是正相关, r 是正值; 要么是负相关, r 为负值, 不可能在一个规律当中既有正相关又有负相关。考察以前出现的 34 种和近期出现的 7 种同系线性规律, 没有一种规律既有正相关又有负相关。所以普遍的密度 D 和结构单元 n 之间的线性规律可能不成立。

2.3 有公式的推导和数据的正确度

文献[5]的基本公式存在原理上的错误, 没有考虑化学因素和分子之间作用力, 将各部分体积机械相加。例如文献[5]的(1)、(2)、(3)、(4)分别为 $D = M/V$, $V = V_x + V_y + nV_A$, $V = a + bn$, $D = M/(a + bn)$, 用后三个公式解释不了表 3 所示的两类密度变化截然相反的现象。文献[5]、[6]数据处理存在不少问题, 所取密度的测定温度不一致, 有效数字计算不规范, 文献[6]表 1、表 2 截距和斜率有效数字普遍和密度有效数字不相符。

表 3 密度 D 和 $\frac{M}{D}$ 随结构单元 n 的变化($\frac{M}{D}$ 取整数)

n	1	2	3	4	5	6
$\text{H}-(\text{CH}_2)_n-\text{NH}_2$ D	0.6628 ²⁰	0.6829 ^{20/4}	0.7173 ^{20/4}	0.7414 ^{20/4}	0.7547 ^{20/4}	0.7660 ²⁰
$\text{H}-(\text{CH}_2)_n-\text{NH}_2$ $\frac{M}{D}$	47	66	82	98	115	132
$\text{H}-(\text{CH}_2)_n-\text{COOH}$ D	1.0492 ^{20/4}	0.9930 ²⁰	0.9577 ^{20/4}	0.9391 ^{20/4}	0.9274 ^{20/4}	0.9200 ^{20/4}
$\text{H}-(\text{CH}_2)_n-\text{COOH}$ $\frac{M}{D}$	57	74	92	109	125	141
n	7	8	9	10	11	12
$\text{H}-(\text{CH}_2)_n-\text{NH}_2$ D	0.7754 ^{20/4}	0.7826 ^{20/4}	0.7886 ^{20/4}	0.7936 ^{20/4}	0.7979 ^{20/4}	0.8015 ^{20/4}
$\text{H}-(\text{CH}_2)_n-\text{NH}_2$ $\frac{M}{D}$	148	165	181	198	214	231
$\text{H}-(\text{CH}_2)_n-\text{COOH}$ D	0.9088 ²⁰	0.9057 ^{20/4}	0.8858 ^{40/4}			
$\text{H}-(\text{CH}_2)_n-\text{COOH}$ $\frac{M}{D}$	158	174	194			
n	13	14	15	16	17	18
$\text{H}-(\text{CH}_2)_n-\text{NH}_2$ D	0.8049 ^{20/4}	0.8079 ^{20/4}	0.8104 ^{20/4}	0.8129 ^{20/4}	0.8510 ^{20/4}	0.8618 ^{20/4}
$\text{H}-(\text{CH}_2)_n-\text{NH}_2$ $\frac{M}{D}$	247	264	280	296	300	312
$\text{H}-(\text{CH}_2)_n-\text{COOH}$ D					0.8532 ⁶⁰	
$\text{H}-(\text{CH}_2)_n-\text{COOH}$ $\frac{M}{D}$					333	

2.4 粗浅看法

由以上讨论可以看出密度 D 与结构单元 n 之间普遍线性规律不成立, 为什么文献[5]又得到非常好的线性相关系数呢? 这个问题可以由两方面来讨论: (1) 由数学角度来分析。文献[5]、[6]没有分离变量, 进行自变量和函数混合回归, 例如文献[5]表 1 的第一个系列回归方程为 $\frac{14n+2}{D} = a + bn$, 两边都有自变量 n , 用这种式子对密度 D 进行回归会使线性相关系数提高非常大(文献[4]已论述, 2.1 有例证)。数学上到目前为止还没有给出这类式子的回归方法, 用这类式子计算出的线性相关系数不具有通常定义的线性相关系数意义; (2) 由物理意义来分析。文献[5]标题是有机同系物密度递变规律, 它实际上计算的是受到密度轻微干扰的分子量 M 与结构单元 n 之间的线性相关系数, 它的回归通式: $M/D = a + bn$, D 与 n 之间线性无关, 但 M 与 n 之间严格线性相关, 例如文献[5]讨论的饱和脂肪烃取代物 $x-(\text{CH}_2)_2-Y$, $M = M_x + 14n + M_y$ 。笔者在表 3 中给出了不同 n 的 $\frac{M}{D}$ 数据。对于 r 为负值的系列, 例如表 3 中的 $\text{H}-(\text{CH}_2)_n-\text{NH}_2$, 随着结构单元 n 的增加, 密度 D 在增加, 增加的幅度不大, 而分子量 M 随着 n 的增加而增加, 增加的幅度大, 这就使 $\frac{M}{D}$ 随着 n 的增加照样增加。对于 r 为正值系列, 例如表 3 中的 $\text{H}-(\text{CH}_2)_n-\text{COOH}$, 随着 n 的增加, D 在减小, 减小的幅度不大, M 随着 n 的增加大幅度的增加, $\frac{M}{D}$ 随 n 的增加也同样增加。也就是说不管密度 D 随着 n 的增加而增加, 还是随着 n 的增加而减小, 由于 D 变化

幅度小, M 增加的幅度大, 改变不了 $\frac{M}{D}$ 的增加趋势, 于是 $\frac{M}{D}$ 随 n 增加而增加, 这正是文献[5]中所有线性回归系数 r 都是正值的原因。可以看出, 文献[5]中决定 r 都是正值的原因是 M 而不是 D 。所以说文献[5]计算的是受到密度轻微干扰的分子量 M 和结构单元 n 之间的线性相关系数。这种论点还可以由文献[5]表 1 数据加以说明, 在文献[5]表 1 中 b 值变化不大, 大多在 16.5 左右。这是由于分子量与结构单元 n 之间关系为: $M = M_x + 14n + M_y$, 用这个式子回归, 线性相关系数 r 为正值, b 应在 14.0 左右, 但是文献[5]回归式 $\frac{M}{D} = a + bn$, 因为对大多数 $X-(CH_2)_n-Y$ 类化合物, D 为小数, 但接近 1。由于除于一个接近 1 的小数, 使 $\frac{M}{D} > M$, 因此 b 由 14.0 变到 16.5 左右。分子量 M 与结构单元 n 之间严格线性关系是无人怀疑的, 根本不需论证。

致谢中国科学院院士北京大学徐光宪教授在百忙之中冒酷暑给笔者指导和帮助, 深表衷心感谢

参考文献

- [1] 蒋明谦. 有机化合物的同系线性规律. 北京: 科学出版社, 1980:123~127, 336.
- [2] 曹晨忠. 中国科学 B 辑, 1995, 25(8): 793~798.
- [3] 刘亚军, 郑世钧. 化学学报, 1998, 56(3): 223~232.
- [4] 孙玉坤. 科学通报, 2000, 45(10): 1109~1113.
- [5] 曹晨忠. 化学通报, 1995, (7): 50~53.
- [6] 曹晨忠. 化学通报, 1994, (10): 44~48.
- [7] 刘汉生, 张宝山. 应用数理统计基础. 太原: 山西科学教育出版社, 1987: 182~186.
- [8] 孙玉坤. 物理化学学报, 1987, 3(5): 539~543.
- [9] C W Robert et al. CRC Handbook of Chemistry and Physics, 63 rd ed, Ohio, CRC Press, 1982.