

估算有机化合物土壤吸着系数 K_{oc} 的 片段常数法*

朴海善 陶 澍** 胡海瑛 卢晓霞 叶生发

(北京大学城市与环境学系, 北京 100871)

摘 要 根据 592 种有机化合物实测 K_{oc} 数据, 研究了用片段常数法估算有机化合物土壤吸着系数的可能性。将 592 种化合物随机划分为建模组和验证组。前者用于模型建立, 后者用于模型检验。计算结果证明, 包括 74 种结构片段和 24 个结构校正因子的回归模型能够很好地预测有机化合物的吸着系数。对所有 592 种化合物而言, 模型的可决系数高达 0.9696。估算结果的平均绝对误差仅为 0.37 个对数单位。

关键词 K_{oc} , 吸着, 土壤, 有机污染物, 片段常数

中图分类号 S153

土壤和沉积物对化学物质的吸着是影响化学物质迁移与转化等环境行为的主要过程之一。在环境污染研究中, 准确测定或估算有机化合物的吸着系数对评价其在环境中的行为具有很重要的意义。有机化合物在土壤或沉积物上的吸着平衡一般用吸着系数 K_d 描述。它指达到吸附平衡时固相与水相浓度之比。有关研究表明, 多数情况下, 吸着过程中起主导作用的吸着剂是天然有机物。假定土壤或沉积物其他组分的贡献可以忽略不计, 有机物的吸着可以用单一有机组分的吸着系数 K_{oc} 与土壤有机物含量两者表征。可见, K_{oc} 实际上是按照土壤有机质含量标准化了的吸着系数。

任何条件下, 实测总是最可靠的 K_{oc} 获取方法。然而在大多数情况下不可能对所有化合物进行这样的测定。通过各类模型预测有机化合物的 K_{oc} 具有重要的应用前景。目前已建立的估算方法包括利用相关物理化学参数的统计模型和直接根据有关结构参数的统计模型。前者包括 K_{oc} 与溶解度和水-辛醇分配系数等定量关系的对数线性模型^[1,2],⁽¹⁾, 后者主要包括取代基常数法、分子连接性指数法和片段常数法。直接利用分子结构参数的方法在实际应用中的优越性不言而喻。从理论上说, 任何有机分子的吸着特性从根本上取决于其组成和结构特征, 这方面研究实际上也是所谓有机物结构效应定量关系研究 (QSAR) 的一部分。取代基常数法因依赖于结构相似母体物质的相关参数, 其使用受到极大限制。分子连接性指数是目前研究最多的方法, 他根据分子拓扑关系建立估算模型, 然

* 国家自然科学基金重点项目 (49632060) 和国家杰出青年基金 (49525102) 资助

** 通讯联系人

(1) 朴海善, 陶澍, 胡海瑛, 李杭, 卢晓霞. 根据水/辛醇分配系数估算有机化合物的吸着系数. 待发表

收稿日期: 1998-10-20; 收到修改稿日期: 1999-08-10

后直接计算吸着系数^{[3],(2)}。片段常数法实际上可视为一种“彻底的”取代基常数法。Karickhoff指出,“在计算某些热力学参数的时候,可以将一个分子看作多种片段的集合体,其中每一种片段对这些热力学参数都起着相对独立的作用。这种把一个分子拆分成多种片段的方法不仅可以应用于辛醇-水相分配系数等参数的计算,还可以应用于化合物吸着系数的计算。”^[4] Leo曾提出根据片段常数法估算化合物水/辛醇分配系数的模型,根据76种化合物的实测水/辛醇分配系数,计算了200多种片段的片段常数和11类结构校正系数。用这种方法建立的模型的平均估算误差仅0.14个对数单位^[5]。张建平利用从文献中收集到的300种化合物的实测 *K_{oc}* 数据,建立了包括53种片段和15种结构因子的回归模型,初步探讨了用片段常数法估算 *K_{oc}* 的可能性。但由于实测数据较少,因此模型估算误差较大⁽³⁾。

本研究的目的是根据从文献中得到的近600种化合物的实测 *K_{oc}* 资料,以Leo提出的估算水/辛醇分配系数的片段常数模型为参照,用线性多元回归方法建立了估算有机化合物 *K_{oc}* 的统计模型。

1 研究方法

1.1 *K_{oc}* 数据收集与选取

从截止目前的多篇文献中收集了592种有机化合物的 *K_{oc}* 实测数据^{[6~10],(2)}。对于文献中报道了一个以上 *K_{oc}* 测定值的化合物,尽可能选取其中位数。对于以 *K_{om}* 表达的数据,取 $K_{oc} = 1.724K_{om}$ 。收集的化合物包含环境中常见的各类有机污染物,如农药、有机染料、多环芳烃、多氯联苯以及酚类化合物等。从592种化合物中随机抽取430种化合物用于建立模型(建模组),剩余的162种化合物用于模型检验。

1.2 *K_{oc}* 片段常数预测模型建立与验证

假定结构片段和有关校正因子对 $\log K_{oc}$ 的贡献具有简单的线性加和性,可以建立以下多元线性回归模型:

$$\log K_{oc} = \sum_{i=1}^a n_i f_i + \sum_{j=1}^b m_j F_j, \quad (1)$$

式中 a 和 b 分别是结构片段和校正因子种类总数。 f_i 为第 i 种结构片段的片段常数; n_i 表示化合物中第 i 种结构片段的个数。 F_j 代表第 j 种结构因子校正系数,而 m_j 表示第 j 种结构因子的数目。模型建立过程包括以下主要步骤:

(1) 根据结构片段和校正因子的初步划分和确定原则,对建模组430种化合物进行片段和因子计数。在此基础上用多元回归分析方法建立初步模型;

(2) 分析拟合结果,对不合理的片段和校正因子进行修正。同时采用逐步回归方法进一步分析校正因子的合理性。重新拟合建模组实测数据;

(3) 根据验证组162种化合物的实测数据对上述回归模型进行验证,籍以对模型的可推广性进行分

(2) Tao S, Lu X X. Estimation of organic carbon normalized sorption coefficient (K_{oc}) for soils by topological indices and polarity factors. 待发表

(3) 张建平. 用片段常数法估算 *K_{oc}* 值. 北京大学学士论文, 1990

建立初步模型时选用了 86 个结构片段。建模过程中为修正某些片段正负取值的偏差,对片段选取方案修正后得到 74 种片段(见表 1)。

1.4 校正因子确定

对于结构复杂的化合物,仅使用片段不能反映化合物的全部结构特征,诸如分子的不饱和性,卤原子的重叠作用,分子空间构型,极性片段之间的相互影响以及卤原子与极性片段之间的相互影响等。因此,要引进能够反映主要结构特征的结构因子。选择的结构因子如表 1 所列。

初步模型建立时仅用 19 个校正因子(1~19),在作上述修正后增至 24 个。

1.5 数据整理与统计分析

用 Excel®建立了 *Koc* 数据库。各类回归模型计算在 SPSS®下实现。

2 结果与讨论

2.1 *Koc* 预测模型建立

根据建模组中 430 种化合物的片段和结构因子计数以及它们各自的实测 $\log K_{oc}$ 值进行多元回归分析,得到以下包含 86 种片段常数和 24 个校正因子的方程式如(2)。

$$\log K_{oc} = \sum_{i=1}^{86} n_i f_i + \sum_{j=1}^{19} m_j F_j, \quad n = 430, \quad R^2 = 0.9729, \quad s = 0.5633 \quad (2)$$

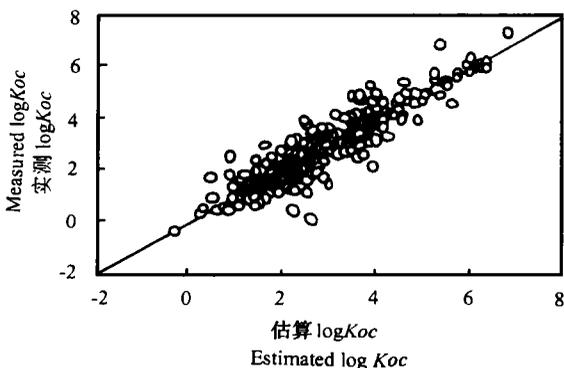


图1 建模组430种化合物实测与估算*Koc*关系

Fig.1 Relationship between measured *Koc* and estimated *Koc* of 430 chemicals for modeling

用式 2 求得建模组中所有 430 种化合物的计算 *Koc*。图 1 为计算值与实测值的关系。

图 1 表明,片段常数模型可以很好地表述化合物结构与其吸着系数的定量关系。然而,拟合得到的回归参数中个别片段常数的正负关系与理论不符。有机化合物在土壤/沉积物有机质上的吸着可近似地视为相似相溶,因此在母体化合物上加合一个疏水官能团的结果可以使新化合物的吸着系数增加,这样的官能团的片段常数应当为正。反之,亲水官能团的片段常数则应为负,

但包括在内的 9 种官能团的正负取值与预想的相反。虽然其中一些出现频率很低的片段(如 $-\text{SP}(\text{S}) (\text{O}-)\text{S}-$)的此类偏差可能与个别化合物的实测误差有关。它们对模型的总体精度影响不大。但其他一些在数据库中出现频率很高的片段则不然。如亲水的 $-\text{NH}_2$ 的片段常数取 $+0.1722$, 而憎水的 $-\text{C}$ 片段常数却为 -0.2851 。类似的片段还包括 $-\text{NH}$ 和 $-\text{CH}$ 等。其原因可能是方案中有关片段拆分不合理。因此,对上述方案进行修正,将 $-\text{N}-$, $-\text{NH}-$ 和 $-\text{NH}_2$ 当作等价片段处理,同时针对 $-\text{NH}-$ 和 $-\text{NH}_2$ 分别引进新的校正因子 F_{NH} 和 F_{NH_2} 。同样,将 $-\text{C}$, $-\text{CH}$, $-\text{CH}_2$ 和 $-\text{CH}_3$ 看作相同的片段,再对 $-\text{C}$, $-\text{CH}$ 和 $-\text{CH}_2$ 引进校正因子 F_{C} , F_{CH} 和 F_{CH_2} 。对 430 种化合物实测数据,以重新划分的片段(74

种)和校正因子(24个)为基础进行多元回归,得到新的拟合方程。虽然拟合结果与修正前几无差别($R^2 = 0.9725$, $s = 0.5618$),但原模型中的主要不合理片段常数不复存在。对建模组中的430种化合物,估算 K_{oc} 值与实测值的平均绝对误差为0.36个对数单位。

2.2 校正因子确认

为进一步判断初步选定的24个结构校正因子在模型中的有效性,用逐步回归方法对它们各自的贡献进行检验。以对模型贡献大小为序,将24个因子逐一引入回归模型中,图2即为因子引入过程中模型可决系数的变化趋势。

从图中结果可见,所有因子的引入均可在某种程度上提高模型的拟合程度。引入24个因子后,模型可决系数从无校正的0.963增至0.973。此外,与其余因子相比,前14个因子对模型的贡献相对较大。事实上,拟合系数的绝对值的大小同样反映了不同因子的相对贡献(参见表3中列举的最终模型的校正因子系数)。

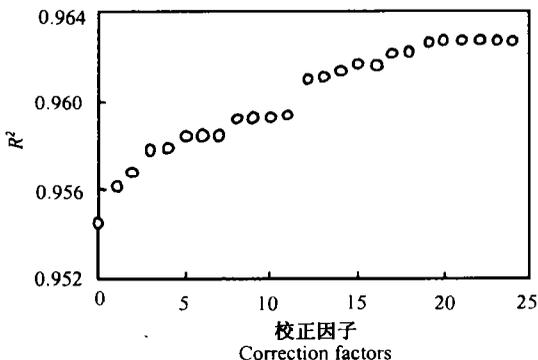


图2 对结构因子逐步回归过程中回归可决系数随因子逐个引入的变化

Fig.2 Variation of coefficient of determination during a stepwise regression (By turns correction factors are $F_{\text{a}}, F_{\text{NH}_2}, F_{\text{C}}, F_{\text{CH}}, F_{\text{P}_3}, F_{\text{NH}}, F_{\text{P}_0}, F_{\text{P}_2}, F_{\text{P}_1}, F_{\text{CH}_2}, F_{\text{P}_4}, F_{\text{HP}_1}, E_{\text{b}}, F_{\text{P}_5}, F_{\text{mhG}_2}, F_{\text{HP}_2}, F_{\text{mhG}_3}, E_{\text{mhv}}, F_{\text{e}}, F_{\text{CB}}, F_{\text{a}}, F_{\text{P}_6}, F_{\text{mhG}_1}, F_{\text{b}}$)

2.3 Koc 预测模型验证

利用由随机选取的162种化合物构成的验证组检验了据建模组数据得到的估算模型的可靠性。根据这些化合物的结构片段和结构特征,计算了各自的片段数及校正因子个数,用上述估算模型(式2)计算了它们的 K_{oc} 值。计算 $\log K_{oc}$ 值与实测 $\log K_{oc}$ 值的比较如图3所示。

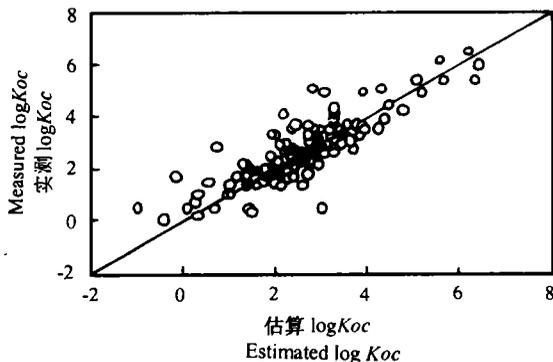


图3 验证组实测Koc和计算(据上述估算模型)Koc比较

Fig.3 Comparison of measured K_{oc} with calculated K_{oc} for testing set

图中数据点基本围绕1:1直线分布,说明估算模型基本可靠。验证组的计算 $\log K_{oc}$ 的平均绝对误差为0.468个对数单位,略高于建模组(0.358个对数单位)。这一结果当在预料之中。尽管如此,验证组中60%化合物的绝对误差仍小于0.5个对数单位。

2.4 基于全部数据的预测模型

在分别用建模组和验证组数据建立初步模型并对其可靠性进行验证后,为提高模型的预测精度,利用全部数据(592种)对模型参数进行修正,得到最终预测模型。模型形式与初步模型(式2)相似,仅参数个数略有不同,且拟合参数也得以修正。据此得到的片段常数和结构校正因子列于表2和表3中。表中参数和式2即构成片段常数法估算有机化

表 2 模型的结构校正因子

Table 2 Structural correction factors (SCF) of the model

编号 No.	符号 Symbol	结构校正因子 SCF	计数原则 Rule of counting
1	F _b	脂链键因子	片段间不与H相连的直链键数减1
2	F _β	脂环键因子	脂环中不与H相连的直链键数减1
3	F _{CBr}	支链因子	支链数
4	F ₋	直链中双键因子	非(类)芳环中片段间双键数
5	F ₋	脂肪环内双键因子	(类)芳环中片段间双键数
6	F ₌	三键因子	片段间叁键数
7	F _{mbG1}	双卤素因子	有2个卤原子取代的碳数
8	F _{mbG2}	三卤素因子	有3个卤原子取代的碳数
9	F _{mbG3}	四卤素因子	有4个卤原子取代的碳数
10	F _{mbv}	相邻两碳多卤素因子	单键相连两碳上卤原子取代数减1
11	F _{P0}	极性片段相互作用因子	直接相连的极性片段对数
12	F _{P1}	极性片段相互作用因子	两性极性片段取代的脂链碳原子数
13	F _{P2}	极性片段相互作用因子	两性极性片段取代的相邻脂链碳原子对数
14	F _{P3}	极性片段相互作用因子	两性极性片段取代的脂环碳原子数
15	F _{P4}	极性片段相互作用因子	两性极性片段取代的相邻脂环碳原子对数
16	F _{P5}	极性片段相互作用因子	两性极性片段取代的芳环碳原子数
17	F _{P6}	极性片段相互作用因子	两性极性片段取代的相邻芳环碳原子对数
18	F _{HP1}	卤原子与极性片段作用因子	卤原子与极性片段取代的脂碳数
19	F _{HP2}	卤原子与极性片段作用因子	卤原子与极性片段取代的芳碳数
20	F _C	四级碳因子	四级碳数
21	F _{CH}	三级碳因子	三级碳数
22	F _{CH2}	二级碳因子	二级碳数
23	F _{NH}	二级氮因子	二级氮数
24	F _{NH2}	一级氮因子	一级氮数

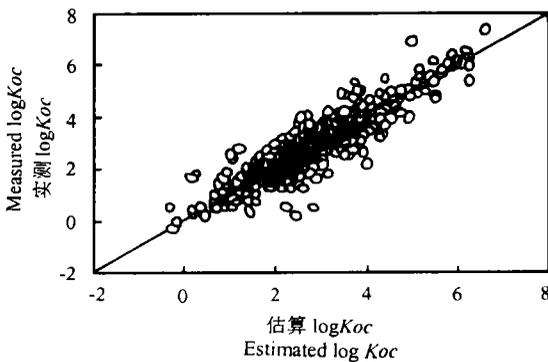


图 4 基于全部592种化合物的模型的估算与实测Koc比较

Fig.4 Comparison of estimated K_{oc} with measured K_{oc} of the model based on whole data set (592)

合物土壤/沉积物吸附系数(K_{oc})的模型。

对本研究收集的所有实测数据而言,模型拟合精度($R^2 = 0.9696$, $s = 0.5617$, $n = 592$)与基于430种化合物的初步模型没有显著差异。图4为根据592种化合物数据建立的模型的估算结果与实测 $\log K_{oc}$ 值关系。从图中可以看出,基于592种化合物建立的简单线性模型估算的结果与实测值相当一致。事实上,用于建模的592种化合物实测

表 3 基于全部数据(592种)的估算模型的结构因子

Table 3 Structural factors of the model based on whole data set (592)

结构因子	系数	结构因子	系数	结构因子	系数	结构因子	系数
Factor	Coefficient	Factor	Coefficient	Factor	Coefficient	Factor	Coefficient
F_b	0.0884	F_{mbG1}	0.0446	F_{P2}	0.4291	F_{HP2}	-0.1074
\underline{F}_b	0.2560	F_{mbG2}	0.1314	F_{P3}	0.6357	F_c	-1.0231
F_{CBr}	-0.1019	F_{mbG3}	-0.0354	F_{P4}	-0.0854	F_{CH}	-0.7309
F_{-}	0.1143	F_{mbv}	0.0085	F_{P5}	0.1658	F_{CH2}	-0.3764
\underline{F}_{-}	0.1264	F_{P0}	0.5604	F_{P6}	0.0748	F_{NH}	0.4302
$F_{=}$	-1.0277	F_{P1}	0.3731	F_{HP1}	-0.1332	F_{NH2}	0.7345

K_{oc} 的变异的约 98% 可由这一回归模型描述。

图 5 为估算 $\log K_{oc}$ 与实测值相对误差的分布。所有 592 种化合物估算和实测 K_{oc} 的平均绝对误差仅 0.37 个对数单位, 其中 74% 的化合物的绝对误差小于 0.5 个对数单位。

与根据相同数据库、用不同方法进行的估算结果相比, 片段常数法建立的模型的精度最高。譬如, 根据水/辛醇分配系数法估算 K_{oc} 的经平均相对误差达 0.48 个对数单位。

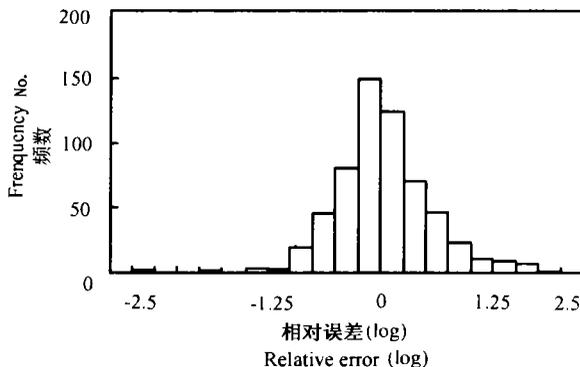


图 5 模型估算相对误差分布

Fig.5 Distribution of relative error of model estimation

3 结 论

基于 592 种有机化合物的实测 K_{oc} 值建立的片段常数估算模型估算的 K_{oc} 值平均相对误差仅 0.37 个对数单位。考虑到实测数据本身的误差, 这一模型具有相当高的估算精度。

参 考 文 献

1. Gawlik B M, Sotiriou N, Feicht E A, Schulte-Hostede S, Kettrup A. Alternatives for the determination of the soil adsorption coefficient, K_{oc} , of non-ionic organic compounds—a review. *Chemosphere*, 1997, 34: 2525~2551
2. A Sabljic. QSAR modelling of soil sorption, improvements and systematics of $\log K_{oc}$ vs. $\log K_{ow}$ correlation. *Chemosphere*, 1995, (31):4489~4514
3. Meylan W, Howard, P H. Molecular topology / fragment contribution method for predicting soil sorption coefficients. *Environ Sci & Technol*, 1992, 26:1560~1567
4. W Karrickhoff. Organic pollutants sorption in aquatic systems. *J. Hydraulic Engineering*, 1984, 110:707~735

5. Leo A J. Octanol/water partition coefficient. In Lyman W J, Rosenblatt D H. eds. Handbook of Chemical Property Estimation Method. McGraw-Hill, NY, 1982. 1~54
6. Sabljic H, Gusten H, Verhaar H, Hermens J. QSAR modeling of soil sorption, improvements and systematics of log *K_{oc}* correlations. Chemosphere, 1995, 31:4489~4514
7. Kenaga E E, Goring C A. Relationship between water solubility, soil sorption, octanol-water partitioning, and concentration of chemicals in biota. In: Eaton J G, Parish P P, Hendricks A C. eds. Aquatic Toxicology, Proceedings of the Third Annual Symposium on Aquatic Toxicology. ASTM, STP 707, Philadelphia, 1980. 78~115
8. Briggs G G. A simple relationship between soil sorption of organic chemicals and their Octanol-water partitioning behavior. Proc 7th British Insecticide and Fungicide Conf, 1973. 11:475~478
9. Briggs G G. Theoretical and experimental relationships between soil adsorption, octanol-water partition coefficients, water solubilities, bioconcentration factors, and the parachor. J Agric Food Chem, 1981, 29: 1050~1059
10. Karickhoff S W, Brown D S, Scott T A. Sorption of hydrophobic pollutants on natural sediments. Water Res, 1979, 13:241~248

FRAGMENT CONSTANT MODEL FOR *K_{oc}* ESTIMATION

Piao Hai-shan Tao Shu Hu Hai-ying Lu Xiao-xia Ye Sheng-fa

(Department of Urban and Environmental Sciences, Peking University, Beijing 100871)

Summary

Measured *K_{oc}* values of 592 chemicals collected from literature were used to develop a sorption coefficient estimation model based on fragment constants plus structural correction factors. The results showed that *K_{oc}* of organic chemicals can be well estimated based on the fragment constant modeling. For the 592 chemicals, the coefficient of determination of the regression model was as high as 0.9696 with a mean absolute error of 0.37 log-unit.

Key words *K_{oc}*, Soil, Sediment, Sorption, Fragment constant