

估算有机化合物在鱼体中生物富集因子的片段常数法

胡海瑛 陶澍 卢晓霞 习晓环

(北京大学城市与环境学系, 北京 100871)

摘要: 根据脂肪烃、多卤联苯、卤代苯和稠环芳烃等 80 种非极性有机化合物在鱼体中生物富集因子(BCF)的实测数据, 建立了有机物在鱼体中生物富集因子的片段常数法估算模型。模型包含 9 个片段常数和 16 个结构校正因子, 模型可决系数为 0.995, 平均绝对误差为 0.1836 个对数单位, 学生化残差呈随机分布。以调整可决系数和平均绝对误差为描述统计量, 进行了类别剔除的 jackknife 检验, 在此基础上讨论了模型的稳健性。

关键词: 片段常数; 生物富集因子; 非极性有机化合物

中图分类号: X174 **文献标识码:** A **文章编号:** 1000-3207(2002)01-0025-006

水体中有机污染物对水生生物的危害是当前倍受关注的环境问题之一。评价有机物在生物体中的富集, 通常用生物富集因子(BCF)表示, 其定义指达到平衡时化合物在生物体中的浓度与在环境(水体)中浓度的比值。由于环境中化合物种类繁多, 所需 BCF 数据不可能全部通过实测获得。就大多数化合物而言, 直接根据相关参数或结构特征估算 BCF 具有重要价值。

依据化合物性质与结构之间的特定关系, 对 BCF 进行估算的方法目前已报道的主要为分子连接性指数法^[1,2]。另一种用于环境参数估算的方法是 \log 和 Hansch 建立的片段常数法, 该方法已成功用于对辛醇-水分配系数(K_{ow})和土壤吸着系数(K_a)的估算^[3-5], 但对片段常数法在估算 BCF 中的应用, 目前还未见文献报道。本研究根据 BCF 实测数据建立非极性化合物 BCF 片段常数估算模型, 并讨论所建模型的精度和稳健性。

1 研究方法

1.1 实测值选取 从已发表文献中收集到 80 种非极性有机物在虹鳟(*Salmo gairdneri*)、胖头鱼(*Pimephales promelas*)、蓝鳃太阳鱼(*Lepomis macrochirus*)和红鲷(*Poecilia reticulatas*)、金鱼(*Carassius auratus*)等鱼种中的 BCF 实测数据(流动实验)^[2]。涉及的化合物包括卤代苯、卤代联苯、氯化萘、氯化脂肪烃、多环芳烃以及 DDT、DDE 等常见微量有机污染物。80 种化合物的实测 $\log BCF$ 值在 0.30 到 5.36 之间。对收集到两个以上实测数据的化合物取 BCF 中位数。

收稿日期: 1999-11-01; 修订日期: 2001-04-20

基金项目: 自然科学基金重点项目(49632060); 国家基金创新群体(40024101)资助

作者简介: 胡海瑛(1974—), 女, 湖南省岳阳市人; 现主要从事环境化学方面的教学科研

通讯作者: 陶澍 E-mail: taos@urban.pku.edu.cn

1.2 模型建立 根据线性自由能原理,可将化合物分子看做多种片段的集合,各片段对其物理化学性质的贡献应具有加和性。因此选择以下线性估算模型:

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

式中 f_i 为第 i 种片段的片段常数, n_i 为化合物中该种片段的个数; F_j 为第 j 种结构因子校正系数, m_j 为该因子的个数; a 、 b 分别代表化合物中片段和结构因子种类数。化合物片段划分和结构因子确定以 Leo 方法为准则^[3],参考其它研究结果^[5]对部分片段进行调整,并根据回归分析结果对结构因子进行取舍。回归在 SPSS7.5 统计软件中进行。

1.3 模型误差分析及稳健性检验 利用残差、学生化残差、可决系数(R^2)、调整的可决系数(R_{adj}^2)和标准差(s)等统计参数对回归模型的误差进行分析。并通过抽取一类化合物的 jackknife 检验研究模型的稳健性^[6]。

2 结果与讨论

2.1 片段常数划分和结构校正因子确定

按照 Leo 提出的原则略加修改,进行片段划分和确定结构因子^[3]。按照 Leo 的原则,对数据库中所有化合物可划分出 20 种片段和校正因子。以此为起点,逐步尝试修改片段划分方法,最终确定以下修正原则:

(1) 对“独立碳原子”计总数,以 F_C 、 F_{CH} 和 F_{CH_2} 结构因子对片段 C、CH 和 CH_2 进行校正;增加 $F_{CH_3}^\Phi$ (Φ 表示与苯环相连),对苯环上甲基取代进行校正,计入模型时系数为取代甲基的个数;

(2) 不区分直链上双键和脂肪环上双键,统一以 $F_{=}$ 代表双键结构校正因子,计算方法与 Leo 相同;

(3) 整体考虑连接在一个“独立碳原子”上多个卤原子的综合作用,认为碳原子上的氢均被卤素取代时 4 个卤原子的综合影响为零,即只引入二个和三个卤原子校正因子, F_{Cl_2} 和 F_{Cl_3} , 计算方法与 Leo 相同;

(4) 对苯和联苯苯环上多卤素取代进行位置校正,考虑对称效应不计算对位的影响,仅对邻、间位进行校正,校正因子分别为 F_o 、 F_m , 计入模型时按邻、间位的个数加和;

(5) 由于联苯 α 位上的取代基对分子空间构象的作用,造成分子横截面积的变化,化合物生物富集过程受到影响^[7,8],故对其进行校正,结构因子为 F_α , 计入模型时按取代基个数线性加和;

(6) 对稠环芳烃电子云相对密集的 α 、 γ 位上卤素取代进行校正,结构因子表示为 $F_{\alpha\gamma}$, 计入模型时以取代基数目加和。

经上述修正,模型的变量数增至表 2 和表 3 中列举的 25 种(9 种片段和 16 种校正因子),调整可决系数从 0.985 增至 0.992,平均绝对误差从 0.2812 降至 0.1836。

2.2 模型建立

由于数据量偏低,预留检验组的办法很难保证模型的可靠性,故采用替代办法,即首先利用全部数据进行回归分析,建立模型,最后采用依次抽取各类化合物的方法进行 jackknife 检验,据此判断模型的稳健性。这样的做法实质上与依次将各类化合物作为检验组,而将其它各类化合物用作建模组的方法等价。根据上述的原则对所有 80 种化合物

计算片段数和校正因子数, 作为变量对实测 $\log BCF$ 数据按方程(1)形式进行强制过原点的多元线性回归。模型可决系数(R^2) 0.994, 标准差(s) 0.3214。得到的回归系数, 即片段常数 f_i 和因子校正常数 F_j , 如表 1 和表 2 所列:

表 1 片段常数回归结果

Tab. 1 Regression results of fragment constants

片段	F	片段	f	f^Φ	f^{Φ^2}	片段	f^{AR}
- H	- 0.032	- Cl	0.423	0.816	0.430	- C(H)=	0.286
- C(-)	- 0.615	- Br		0.797		- C=	0.097
- C ₆ H ₅	1.241						

注: f^Φ : 片段与芳香环相连; f^{Φ^2} : 片段与双键相连; f^{AR} : 芳香环上片段

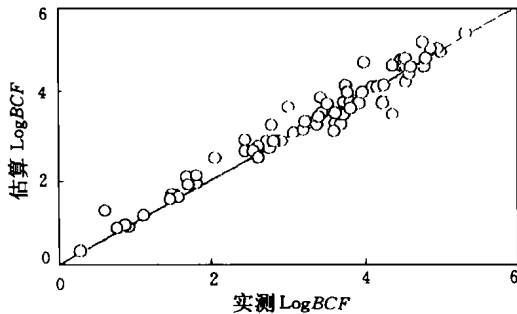
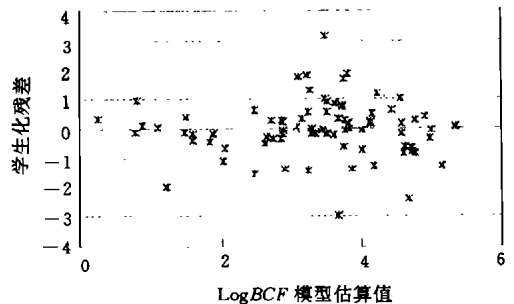
表 2 结构因子回归结果

Tab. 2 Regression results of structure factors

结构因子	F	结构因子	F	结构因子	F	结构因子	F
F_C	1.524	F_b	- 0.369	$F^{\Phi=}$	0.781	F_O	- 0.264
F_{CH}	0.508	F_b	0.197	$F^{\Phi^2=}$	1.875	F_m	- 0.153
F_{CH_2}	0.688	F_{rCl}	- 4.886	F_{mhG1}	0.221	F_α	- 0.435
$F^{\Phi}_{CH_3}$	1.072	$F_=$	0.922	F_{mhG2}	0.123	$F_{\alpha\gamma}$	- 0.127

注: 除文中特别提及, 符号意义与 Leo 相同^[3]; F^Φ : 校正对象与芳环相连; F^{Φ^2} : 校正对象与双键相连

图 1 为 $\log BCF$ 模型估算值与实测值的对应关系。所有数据点都围绕 1:1 直线两侧, 模型可决系数高达 0.995。虽然没有采用有重复因变量的回归方式, 在选定的变量条件下, 模型的线性特征一目了然。此外, 对所有 25 个自变量的总体参数是否为零的 t-检验结果表明, 在 0.10 水平, 7 个变量不显著。但考虑到这些片段的物理意义, 模型仍保留所有选定变量。

图 1 $\log BCF$ 实测值与模型估算值对应关系Fig. 1 Relationship between calculated and measured $\log BCF$ 图 2 学生化残差与 $\log BCF$ 模型估算值Fig. 2 Plot of studentized residual against calculated $\log BCF$

2.3 误差分析

对所研究的 80 种化合物, 拟合模型的平均绝对误差仅为 0.1836 个对数单位, 其中 78.8% 的绝对误差小于 0.3 个对数单位, 92.5% 的绝对误差小于 0.5 个对数单位。为进一步考察模型对不同化合物的误差和模型效果, 计算了学生化(Studentized)残差, 结果如图 2 所示。

如图所示,学生化残差围绕“0”值线上下波动,多数随机分布在[-2,2]区间内。落在该区间以外的化合物包括苯(残差-0.5740)、六溴苯(-0.6431)、2,2,4,4-四氯联苯(-0.6768)和2,3-二氯联苯(0.9315),他们分属单环芳烃、溴苯和多氯联苯。其中前三种化合物的较大偏差很可能与它们特别长的半衰期有关。据认为,鱼体内半衰期越长的化合物达到鱼体-水体分配平衡所需要的时间也越长^[9],对这样的化合物,惟动态实验才能得到可靠的实测BCF值,静态实验的结果则可能显著偏低。由于所收集的实测资料中部分未给出实验方法,很可能包括了部分静态实验数据,从而导致实测数据偏低和较大的负残差。尽管如此,残差偏大的化合物所占比例很小,加上表征其对模型本身影响的Cook's距离都不大(最大值仅0.42),它们的存在不应对模型的可靠性产生显著影响。

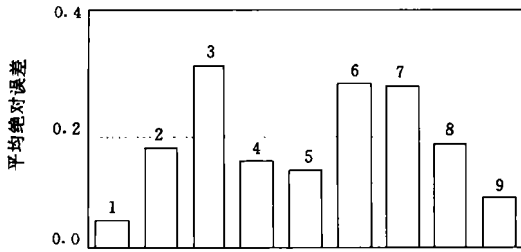


图3 9类化合物模型平均绝对误差(对数单位)

Fig. 3 Mean absolute residual of each chemical group

1. 氯代脂肪烃(10)
2. 单环芳烃(5)
3. 溴苯(7)
4. 氯苯(9)
5. 多溴联苯(3)
6. 多氯联苯(14)
7. 氯萘(8)
8. 稠环芳烃(11)
9. 其它(多氯环己烷、氯丹、DDT及DDE等)(13) (括号内为化合物个数)

对于不同类别的化合物,模型的误差有较大差别。按照本研究的分类方法,8类化合物的平均误差在图3中给出。所研究的8类化合物中,溴苯、多氯联苯和氯萘的平均绝对误差超过模型平均水平(图中虚线),分别为0.31,0.28,0.27,而氯代脂肪烃最小,仅为0.05。

2.4 化合物类别与模型稳健性

模型的稳健性可以用jackknife法检验。考虑到本研究的数据较少,涉及的化合物种类相对较少,故运用jackknife法针对不同类别化合物进行稳健性检验,藉以判断模型在各种类别化合物之间外推的可靠性。

检验时依次剔除每类化合物,并对余下化合物进行回归,最后通过比较9次回归的调整可决系数(R_{adj}^2)和平均绝对误差进行稳健性检验。检验结果如图4所示:

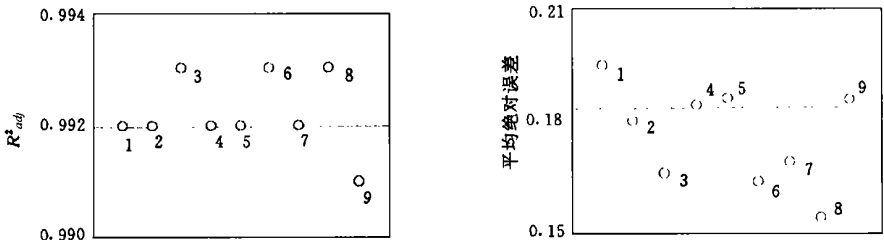


图4 剔除一类化合物 jackknife 检验结果(标号为被剔除类别)

Fig. 4 Adjusted coefficient of determination and mean absolute residual from jackknife test with deletion of each chemical group

1. 氯代脂肪烃
2. 单环芳烃
3. 溴苯
4. 氯苯
5. 多溴联苯
6. 多氯联苯
7. 氯萘
8. 稠环芳烃
9. 其它

如图所示,无论剔除哪类化合物模型的上述两个参数波动都不大,可见模型对不同化合物类别具有相当好的稳健性。相对而言,剔除溴苯、多氯联苯和稠环芳烃三类化合物后模型的 R_{adj}^2 增高而平均残差降低,这三类化合物,特别是前两类恰好是平均绝对误差较高

者(图 3)。模型对它们的估算精度相对较差。氯萘与多氯联苯不仅原子个数多, 结构也趋复杂。多氯取代时氯原子对化合物性质的影响已不是简单加和关系, 且有多种效应共同发挥作用, 但对每种效应的作用程度还不能区分和定量。模型虽然引入校正因子 F_o 、 F_m 和 $F_{\alpha\gamma}$ 代表这种结构效应, 使拟合效果有所改善, 但简单的线性校正因子很难将这种复杂关系完全描述清楚。此外, 许多文献提到许多此类分子较大化合物在实验期限内未必达到平衡, 虽然用动态方法代替静态方法能够消除这种偏差, 但并非所有实验者都能提供动态方法测定的数据, 故数据库中此类化合物的误差可能要大一些。

除此之外, 剔除后造成 R_{adj}^2 或平均残差变化相对较大的化合物还包括其它类和氯代脂肪烃。氯代脂肪烃结构简单, 常规片段和结构因子能够充分和恰当表达结构效应与化合物物化性质之间的关系, 这些化合物 $\log BCF$ 也较小(0.3–3.76), 在此范围内 BCF 与热力学参数之间线性关系一般较好^[10, 11]。本研究建立的模型能很好地描述这类化合物, 故剔除后导致平均绝对误差上升。归入其它类的 13 种化合物剔除后虽然对平均残差影响不大, 却导致 R_{adj}^2 下降。由于这些化合物结构差异较大, 很难直接判断其根本原因。

根据 80 种非极性化合物 BCF 实测数据建立片段常数法估算模型, 模型可决系数高达 0.995, 平均绝对误差为 0.1836 个对数单位。说明 80 种化合物片段常数和结构因子与 BCF 之间存在很好的定量关系。个别可疑值对模型影响不大。jackknife 检验结果证明模型在化合物类别外推方面具有较高的稳健性。

参考文献:

- [1] Koch R. Molecular connectivity index for assessing ecotoxicological behaviour of organic compounds [J]. *Toxicol. Environ. Chem.*, 1983, **6**: 87–96
- [2] 卢晓霞, 陶澍, 胡海瑛, 等. 根据分子连接性指数估算非极性有机化合物在鱼体中的生物富集因子[J]. *环境科学*, 1999, **20**: 9–12
- [3] Leo A J. Symposium on structure-activity correlations in studies of toxicity and bio-concentration with aquatic organisms [M]. Great Lakes Research Advisory Board, Burlington, Ontario, 1975: 151
- [4] Hammett L. Physical organic chemistry: reaction rates, equilibria and mechanisms (2nd ed.) [M]. New York: McGraw-Hill, 1970
- [5] 朴海善, 陶澍, 胡海瑛. 估算有机化合物土壤吸着系数 K_{oc} 的片段常数法[J]. *土壤学报*, 2000, **37**: 209–216
- [6] Nirmalakhandan N N, Speece R E. Prediction of aqueous solubility of organic chemicals based on molecular structure [J]. *Environ. Sci. Technol.*, 1988, **22**: 328–338
- [7] Sparling J et al. The effects of degree of ortho chloro substitution on the pharmacokinetics of five hexachlorobiphenyls in the rat [J]. *Chemosphere*, 1980, **9**: 29–137
- [8] Shaw G R, Connell D W. Relationships between steric factors and bioconcentration of polychlorinated biphenyls (PCB's) by the sea mullet (*Mugil cephalus* Linnaeus) [J]. *Chemosphere*, 1980, **9**: 731–743
- [9] Oliver B G, Niimi A J. Bioconcentration factors of some halogenated organics for rainbow trout: limitations in their use for prediction of environmental residues [J]. *Environ. Sci. Technol.*, 1985, **19**: 842–849
- [10] Muir D C G. Bioconcentration of PCDD's by fish: effects of molecular structure and water chemistry [J]. *Chemosphere*, 1985, **14**: 829–833
- [11] Karin Fox et al., Kinetics of bioconcentration and clearance of 28 polychlorinated biphenyl congeners in Zebrafish (*Brachydanio rerio*) [J]. *Ecotox. Environ. Safety*, 1994, **28**: 9–109

PREDICTION OF BIOCONCENTRATION FACTOR OF NON-POLAR ORGANIC COMPOUNDS BY FRAGMENT CONSTANT REGRESSION

HU Hai-ying, TAO Shu, LU Xiao-xia and XI Xiao-huan

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

Abstract: Relationship between bioconcentration factor and fragment constants was investigated. A prediction model was developed on the basis of the observed BCF values of 80 non-polar organic compounds. The precision of the model was evaluated using adjusted coefficient of determination and studentized residuals. Jackknife tests was applied to examine the robustness of the prediction model. The coefficient of determination and the mean residual of the model was 0.995 and 0.1836, respectively.

Key words: Robustness; Bioconcentration factor; Fragment constant