

DOI: 10.7524/AJE.1673-5897.20180718002

郑玉婷, 乔显亮, 于洋, 等. 有机化学品生物富集因子定量结构-活性关系模型[J]. 生态毒理学报, 2019, 14(2): 214-221

Zheng Y T, Qiao X L, Yu Y, et al. Quantitative structure-activity relationship model for bioconcentration factors of organic chemicals [J]. Asian Journal of Ecotoxicology, 2019, 14(2): 214-221 (in Chinese)

## 有机化学品生物富集因子定量结构-活性关系模型

郑玉婷<sup>1,\*</sup>, 乔显亮<sup>2</sup>, 于洋<sup>1</sup>, 林军<sup>1</sup>, 丁琼<sup>1</sup>

1. 环境保护部固体废物与化学品管理技术中心, 北京 100029

2. 大连理工大学环境学院 工业生态与环境工程教育部重点实验室, 大连 116024

收稿日期: 2018-07-18 录用日期: 2018-09-26

**摘要:** 依据经济合作与发展组织(OECD)关于定量结构-活性关系(QSAR)模型构建和使用导则, 将780个有机化合物, 以4:1的比例随机划分为训练集(624个化合物)和验证集(156个化合物), 通过多元线性回归(MLR)方法构建了一个包含12个描述符的有机化合物鱼类生物富集因子(BCF)的QSAR模型。QSAR模型的调整决定系数  $R_{adj}^2 = 0.809$ , 去一法交叉验证系数  $Q_{LOO}^2 = 0.803$ , 外部验证系数  $Q_{EXT}^2 = 0.732$ , 表明模型具有较好的拟合优度、稳健性和预测能力。采用欧几里德距离方法表征模型应用域, 通过威廉姆斯图分析模型离群点, 并对模型进行机理解释。所构建的模型, 可以用于预测应用域内有机化学品的生物富集因子。

**关键词:** 有机化学品; 生物富集因子; 定量结构-活性关系

文章编号: 1673-5897(2019)2-214-08 中图分类号: X171.5 文献标识码: A

## Quantitative Structure-Activity Relationship Model for Bioconcentration Factors of Organic Chemicals

Zheng Yuting<sup>1,\*</sup>, Qiao Xianliang<sup>2</sup>, Yu Yang<sup>1</sup>, Lin Jun<sup>1</sup>, Ding Qiong<sup>1</sup>

1. Solid Waste and Chemicals Management Center, MEP, Beijing 100029, China

2. Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), Department of Environmental Science and Technology, Dalian University of Technology, Dalian 116024, China

Received 18 July 2018 accepted 26 September 2018

**Abstract:** According to the principles of development and validation for quantitative structure-activity relationship (QSAR) proposed by Organization for Economic Cooperation and Development (OECD), a QSAR model of bioconcentration factors in fish was built. 780 organic compounds were randomly divided into training set (624 compounds) and validation set (156 compounds) at a ratio of 4:1. The method that developed the model with 12 molecular descriptors was multiple linear regressions (MLR). The adjusted coefficient of determination ( $R_{adj}^2$ ), the leave-one-out cross validated coefficient ( $Q_{LOO}^2$ ) and external explained variance ( $Q_{EXT}^2$ ) was 0.809, 0.803 and 0.732, respectively. It is indicated that the model had satisfied goodness-of-fit, robustness and predictive ability. The application domain of the model was characterized by the Euclidean distance, outliers were analyzed by Williams and

mechanism was interpreted subsequently. The obtained model was able to predict the BCF of organic chemicals within the application domain.

**Keywords:** organic chemicals; bioconcentration factor; QSAR

截至2018年7月,美国化学文摘社(CAS)统计出世界上迄今为止的有机和无机物质已超过1.4亿种。为最大限度地避免化学品带来的环境及健康问题,新化学物质在进入市场前应开展风险评估,已进入市场的化学物质,应补充对其开展风险研究并采取相应的管控措施。但开展一种化学品的风险评估,通过传统的实验方法可能需要数年,评估的速度远不及化学品数量增加的速度,且费用高,需依靠动物实验<sup>[1]</sup>。因此有必要发展替代技术,如定量结构-活性/性质关系(QSAR)等预测技术,获取风险评估所需的相关数据。QSAR技术能够弥补实验方法的不足,快速预测化学品的相关属性,得到了各国化学品行政管理部门的重视,并逐步被纳入美国、欧盟等国家和组织的化学品行政管理,在化学物质测试、新化学物质申报登记、现有化学物质风险筛查、全球化学品统一分类和标记系统(GHS)等领域也发挥着越来越大的作用<sup>[2]</sup>。

开展化学品环境风险评估必不可少的指标参数之一是生物富集因子(BCF)。生物富集因子被定义为达到平衡状态时,污染物在生物体内浓度与环境介质中浓度的比值<sup>[1]</sup>。它能够评价化学品的生物累积性,被认为是筛选具有高风险化学品生物蓄积性不可或缺的重要指标。随着计算机技术的不断更新,预测BCF的QSAR模型也得到了快速发展。Gramatica和Papa<sup>[3]</sup>使用Dragon软件计算了化合物的1150个分子描述符,使用MOPAC-PM3软件计算了3个量子化学描述符,使用EPIWIN package软件包计算了 $\log K_{ow}$ 值。通过GA-VSS进行变量选择,经普通最小二乘法(OLS)回归构建BCF模型,模型进行了外部验证并表征了应用域。Qin等<sup>[4]</sup>根据线性溶解能关系(LSER)理论,以量子化学描述符(Connolly分子表面积、平均分子极化率、分子量、碳原子最负净电荷),采用软件Simca 6.0中偏最小二乘(PLS)回归方法,对122个有机化合物建立BCF模型。该模型进行了应用域表征,满足经济合作与发展组织(OECD)提出的QSAR模型构建导则。量子化学类描述符计算相对复杂,但在模型机理解释方面占一定优势。Piira等<sup>[5]</sup>构建了一个以473个有机化合物作为训练集的BCF模型。该模型在去掉7

个离群点后,训练集与测试集的相关系数分别提高至0.776和0.726。Papa等<sup>[6]</sup>建立了一个包含605种有机化合物的QSAR模型。Dearden等<sup>[7]</sup>选择疏水性描述符、氢键、分子拓扑性指数等描述符,以624个化合物,通过人工神经网络(ANN)构建了BCF模型。但是基于人工神经网络方法构建的模型往往没有明确的表达式,所得描述符也不便于机理解释。以上模型有一个共同的特点是包含了较大数量的化合物或化合物种类繁多,都属于通用模型(universal models)<sup>[8]</sup>。

针对单一类型化合物构建的模型属于局域模型(local models)。局域模型的拟合能力一般较高,但其应用域具有一定局限性。如de Melo<sup>[9]</sup>以57个多氯联苯(PCBs),通过软件Interface Parameter Client计算5种描述符,1个E状态指数,4个拓扑类指数,以偏最小二乘(PLS)方法建立BCF模型。其中变量选择和模型的建立通过OPS算法完成,模型进行了内外部验证。Mansouri等<sup>[10]</sup>通过21个多溴联苯醚(PBDEs)分子构建多元线性回归模型,模型仅包含2个DRAGON描述符,较为简洁。Jackson等<sup>[11]</sup>以93个农药化合物建立BCF模型,模型包含16个分子描述符( $\log P$  and E-state indexes)。郑玉婷等<sup>[12]</sup>针对374种卤代化合物,通过DRAGON描述符构建了满足OECD模型(QSAR)构建导则的BCF模型。综上,目前QSAR模型主要包含两大类,局域模型和通用模型。局域模型的拟合效果和预测能力相对较高,由于其是针对特定类型化合物构建的,因此具有特定的应用价值。为快速获取大量不同种类化合物的BCF,使得BCF-QSAR模型在化学品风险评估与管理的实际工作中发挥作用,本研究构建了有机化学品生物富集因子的通用模型。

## 1 材料与方法 (Materials and methods)

### 1.1 数据集

模型数据集包含780个有机化合物的鱼类 $\log BCF$ 值,数据来源于文献及On-Line EPI Suite<sup>TM</sup> User's Guide (v4.00-v4.10)<sup>[9,11,13-19]</sup>。鱼体种类包括鲤鱼、虹鳟鱼、青鳉鱼、黑头呆鱼、斑马鱼等。化学品种类如卤代类(多氯联苯、多氯联苯醚、多溴联苯醚、全氟磺酸类、全氟羧酸),农药,二恶英类等环境重点关

注化学物质,并包括近20种(类)原环境保护部2017年发布的《优先控制化学品名录(第一批)》中的化学品。

## 1.2 分子结构优化

使用高斯09软件包中的密度泛函理论优化化合物分子结构<sup>[20]</sup>。计算条件:B3LYP方法、6-31+G(d,p)基组水平,使用极化连续介质模型(PCM)考虑溶剂(水)的效应。

## 1.3 计算分子描述符

基于高斯优化的结构,使用Dragon 6.0软件计算描述符<sup>[21]</sup>。Dragon 6.0包括29类4885种分子描述符,是一个全面且高效的分子描述符计算软件。对计算的4885种描述符进行初步筛选,去掉常数和近似常数的描述符,去掉至少有一个缺失值的描述符后,用于建模的描述符集包含描述符3480个。

## 1.4 模型构建

训练集与验证集的划分对于所建模型的泛化能力具有一定影响,因此,以4:1的比例将780个化合物,随机划分为训练集(624个化合物)和验证集(156个化合物)。多元线性回归(MLR)是一种算法简洁且透明的回归方法,本研究使用IBM SPSS Statistics 19.0软件中逐步回归方法,以描述符作为自变量,以logBCF作为因变量,构建多元线性回归(MLR)模型。逐步回归方法可得到一系列模型,综合依据模型的表征参数,描述符的个数(通常不超过样本点个数的1/5),方差膨胀因子(小于10,严格来说小于5),同时参考回归结果(当描述符个数达到12后,继续增加自变量个数, $R_{adj}^2$ 无显著提高),选取最优模型。模型拟合优度、稳健性及预测能力分别通过调整决定系数( $R^2$ ),去一法交叉验证系数( $Q_{LOO}^2$ )和bootstrapping法验证系数( $Q_{BOOT}^2$ ),外部决定系数 $R_{EXT}^2$ 和外部验证系数 $Q_{EXT}^2$ 表征<sup>[22]</sup>。多元线性回归建模应满足任意2个自变量之间不存在相关性,如果线性相关系数很高,就会产生多重相关性问题。因此,本研究使用方差膨胀因子(VIF)及QUICK rule方法诊断多重相关性,模型应用域(D)采用欧几里德距离表征。

## 2 结果(Results)

### 2.1 QSAR模型及参数

$\log\text{BCF}=2.137+0.061*\text{MLOGP2}+0.034*\text{F02}[\text{C-Cl}]-0.312*\text{nROH}-1.282*\text{P}-117+0.323*\text{Mor25m}-0.052*\text{N\%}+0.080*\text{X4v}-0.289*\text{O}-058-1.137*\text{LLS}_-$

$01-1.387*\text{H4v}+0.071*\text{SM12\_AEA}(\text{dm})-0.269*\text{O}-057$   
 $n_{\text{training}}=624, m=12, R^2=0.812, R_{\text{adj}}^2=0.809, Q_{\text{LOO}}^2=0.803, Q_{\text{BOOT}}^2=0.795, \text{RMSE}_{\text{training}}=0.612, P<0.001,$   
 $n_{\text{validation}}=156, R_{\text{EXT}}^2=0.738, Q_{\text{EXT}}^2=0.732, \text{RMSE}_{\text{validation}}=0.736, K_{\text{YX}}=0.397, K_{\text{X}}=0.358$

最优模型包含12个Dragon描述符,每个描述符的变量膨胀因子(VIF)均小于3。其中, $n_{\text{training}}$ 为训练集个数, $m$ 为描述符个数, $R^2$ 为决定系数, $R_{\text{adj}}^2$ 为调整决定系数, $Q_{\text{LOO}}^2$ 为去一法交叉验证系数, $Q_{\text{BOOT}}^2$ 为bootstrapping法验证系数, $\text{RMSE}_{\text{training}}$ 为训练集均方根误差, $P$ 为显著性水平, $n_{\text{validation}}$ 为验证集个数, $R_{\text{EXT}}^2$ 为外部验证的决定系数, $Q_{\text{EXT}}^2$ 为外部验证系数, $\text{RMSE}_{\text{validation}}$ 为外部验证的均方根误差, $K_{\text{YX}}$ 为自变量与因变量组成的矩阵 $M_{\text{YX}}$ 的 $K$ 相关指数, $K_{\text{X}}$ 为自变量矩阵 $M_{\text{X}}$ 的 $K$ 相关指数,且 $K_{\text{YX}}$ 与 $K_{\text{X}}$ 之差大于0.01。

结果显示, $R_{\text{adj}}^2$ 、 $Q_{\text{LOO}}^2$ 、 $\text{RMSE}_{\text{training}}$ 分别为0.809、0.803和0.612,表明模型具有良好的拟合效果及稳健性; $R_{\text{EXT}}^2$ 、 $Q_{\text{EXT}}^2$ 和 $\text{RMSE}_{\text{validation}}$ 分别为0.738、0.732和0.736表明模型具有较好外部预测能力。其预测值和实验值的关系如图1所示。

### 2.2 模型应用域表征

为表征模型应用域,本研究采用了欧几里德距离的方法,如图2所示。训练集化合物特征向量到中心点特征向量的欧几里德距离范围为0.192~1.438,因此特征向量欧几里德距离不大于1.438的化合物适用于本模型。

### 2.3 模型离群点分析

标准残差对杠杆值的威廉姆斯图如图3所示,化合物的标准残差落在(-3,+3)以外,被认为是离群点,离群点信息如表1所示。化合物的leverage值大于警戒值 $h^*$ 时( $h^*=3(m+1)/n_{\text{training}}$ ),认为该化合物显著影响模型的回归效果。leverage值大于警戒值( $h^*$ )的化合物其标准残差仍落在(-3,+3)以内,说明本模型具有一定的延展性。训练集化合物的leverage值范围为0.0031~0.4532,说明leverage值于0.0031~0.4532范围内的化合物适用于本模型。

## 3 讨论(Discussion)

### 3.1 模型描述符意义及机理解释

模型中描述符及各描述符的含义、种类与VIF值如表2所示,可知本模型包含7类描述符,分别是基本描述符、2D原子对描述符、3D-MoRSE描述符、



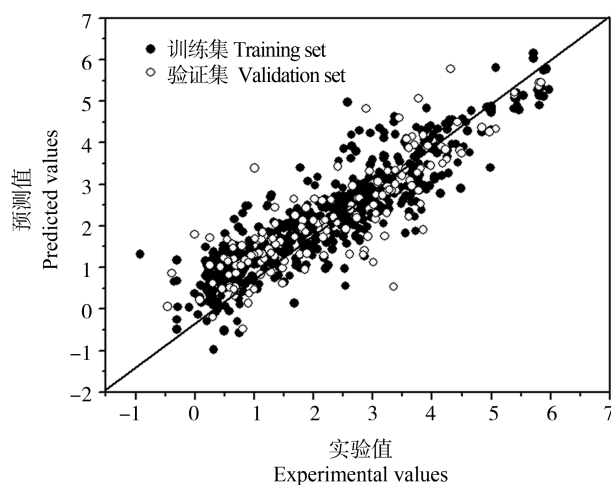


图 1 logBCF 预测值与实测值拟合图

Fig. 1 Fitting figure for predicted and measured values of logBCF

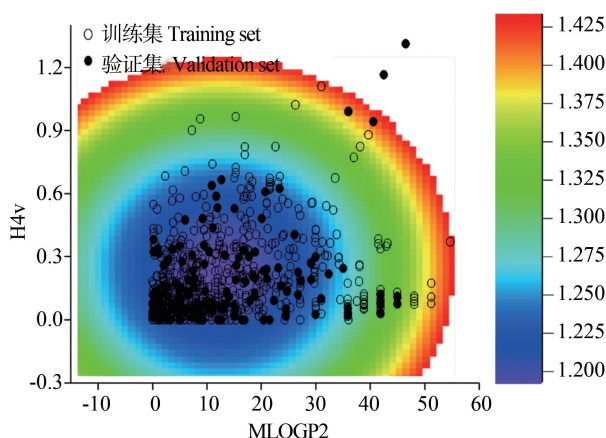


图 2 欧几里德距离方法表征模型应用域

Fig. 2 Model applicability domain based on the Euclidean distance approach

表 1 模型离群点信息

Table 1 The outliers information of the QSAR model

化合物名称 Compound name	标准残差 Standardized residuals
丙烯醛(Acrolein)	3.21
氟甲喹(Flumequine)	3.66
抗氧剂 702(4,4'-Methylenebis(2,6-di-tertbutylphenol))	-3.09
甲基咪草烟(Imazapic)	-3.61
八氯化萘(Octachloronaphthalene)	-3.86

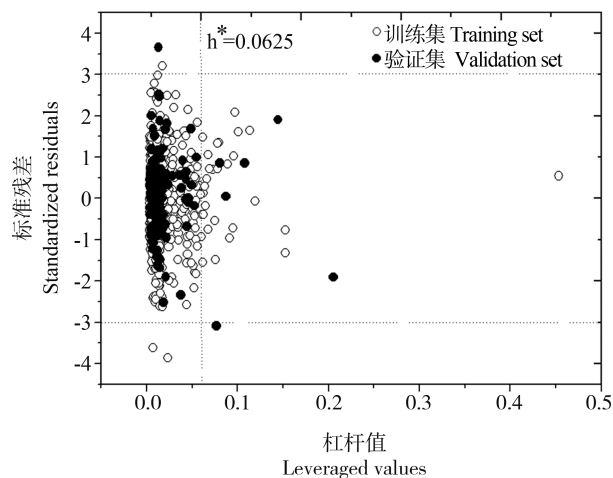


图 3 标准残差对杠杆值的威廉姆斯图

Fig. 3 Standard residuals versus leverage values of Williams

分子连接性指数、基本指数、GETAWAY 描述符、边界邻接指数。其中 MLOGP2 代表摩尔辛醇-水分配系数(Moriguchi- $\log K_{ow}$ )的平方, Moriguchi 模型是描述  $\log K_{ow}$  的方程<sup>[22]</sup>。F02[C-CI]代表原子对“C-CI”拓扑距离为 2 时出现的频率, nROH 代表羟基的个数, P-117 代表原子中心片段“X3-P=X”, Mor25m 代表质量加权的散射函数, N% 代表 N 原子的百分比, X4v 是四价连接性指数, O-058 代表原子中心片段“=O”, LLS\_01 是一个类药指数, H4v 代表氢加权范德华体积滞后 4 的自相关函数, SM12\_AEA (dm)代表增广边缘邻接矩阵的 12 阶谱距(加权偶极矩), O-057 代表苯酚/烯醇/羧基中“OH”结构。仅选择 MLOGP2 构建模型,  $R_{adj}^2$  可达到 0.676, 说明辛醇水分配系数在构建生物富集因子模型时的重要性, 这与其他作者观点一致<sup>[23-24]</sup>。化学物质的生物富集受化学物质本身的亲/疏水性、生物体类脂含量及代谢作用等影响, 仅通过辛醇水分配系数往往不能充分表征生物富集的全过程, 所以本模型中还包括其他类型的描述符, 更加全面地描述大量化合物的相关结构。如: P-117 代表分子中“X3-P=X”结构, “X3-P=X”结构是细胞膜磷脂结构中的疏水部分。nROH、O-057 用于描述羟基、苯酚、烯醇、羧基中“OH”的结构信息, “OH”结构往往决定了化学物质本身的亲/疏水性。另外, 模型中所有描述符的方差膨胀因子(VIF)均小于 3, 自变量与因变量组成的矩阵  $M_{YX}$  的  $K$  相关指数  $K_{YX}$  大于自变量矩阵  $M_X$  的  $K$  相关指数  $K_X$ , 都表明描述符之间不存在严重的多重共线性。

表2 分子描述符含义与方差膨胀因子  
Table 2 The meaning of molecule descriptors and values of variance inflation factor

描述符 Descriptor	含义 Meaning	类型 Block	方差膨胀因子 VIF
MLOGP2	摩尔辛醇-水分配系数的平方 Squared Moriguchi octanol-water partition coefficient	基本描述符 Basic descriptor	2.5
F02[C-Cl]	原子对 C-Cl 拓扑距离为 2 时出现的频率 Frequency of C-Cl at topological distance 2	2D 原子对描述符 2D atom pairs	1.8
nROH	羟基的个数 Number of hydroxyl groups	基本描述符 Basic descriptor	1.3
P-117	原子中心片段 X3-P=X Atom-centred fragments X3-P=X	基本描述符 Basic descriptor	1.2
Mor25m	信号 25/质量加权的散射函数 Signal 25/weighted by mass	3D-MoRSE 描述符 3D-MoRSE descriptor	1.2
N%	N 原子的百分比 Percentage of N atoms	基本描述符 Basic descriptor	1.5
X4v	四价连接性指数 Valence connectivity index of order 4	连接性指数 Connectivity indices	2.0
O-058	原子中心片段 =O Atom-centred fragments =O	基本描述符 Basic descriptor	1.8
LLS_01	类药指数 Drug-like indices	基本指数 Basic index	2.6
H4v	氢加权范德华体积滞后 4 的自相关函数 H autocorrelation of lag 4/weighted by van der Waals volume	GETAWAY 描述符 GETAWAY descriptor	2.8
SM12_AEA(dm)	增广边缘邻接矩阵的 12 阶谱距(加权偶极矩) Spectral moment of order 12 from augmented edge adjacency matrix weighted by dipole moment	边界邻接指数 Boundary adjacency index	3.0
O-057	苯酚/烯醇/羧基中的 OH Phenol/enol/carboxyl OH	基本描述符 Basic descriptor	1.3

表3 BCF 的 QSAR 局域模型  
Table 3 Local QSAR models of BCF

化合物数 <sup>1)</sup> Number of compound <sup>1)</sup>	描述符数 Number	描述符 Descriptor	相关系数 $R^{2\ 2)}$ Correlation coefficient $R^{2\ 2)}$	建模方法 Method	应用域 Application domain	参考文献 Reference
非离子型有机化合物 Nonionic 239	13	分子连接性指数、基团校正因子 Molecular connectivity index and group correction factor	0.810	非线性回归 Nonlinear regression	无 No	(2000) <sup>[13]</sup>
持久性有机污染物 Persistent organic pollutants 200/36	7	电负性距离矢量 Electronegativity distance vector	0.808/0.771	预测变量选择 Prediction variable selection	无 No	(2007) <sup>[30]</sup>
非离子型有机化合物 Nonionic 122/70	6	量子化学描述符 Quantum chemical descriptors	0.868/0.755	偏最小二乘 Partial least squares	有 Yes	(2009) <sup>[31]</sup>
农药类有机化合物 Pesticide 93/16	16	E-状态指数、 $\log P$ E-state index, $\log P$	0.80/0.88	多元线性回归 Multiple linear regression	无 No	(2009) <sup>[11]</sup>
非极性有机化合物 Nonpolar 85/29	5	电负性距离矢量 Electronegativity distance vector	0.937/0.902	预测变量选择 Prediction variable selection	无 No	(2010) <sup>[29]</sup>

续表3

化合物数 <sup>1)</sup> Number of compound <sup>1)</sup>	描述符数 Number	描述符 Descriptor	相关系数 $R^2$ <sup>2)</sup> Correlation coefficient $R^2$ <sup>2)</sup>	建模方法 Method	应用域 Application domain	参考文献 Reference
多氯联苯类有机物 Polychlorinated biphenyls 47/10	5	E-状态指数、拓扑指数 E-state index, topological index	0.923/0.968	偏最小二乘 Partial least squares	无 No	(2012) <sup>[9]</sup>
卤代有机化合物 Halogenated 299/75	9	DRAGON	0.877/0.776	多元线性回归 Multiple linear regression	有 Yes	(2013) <sup>[12]</sup>

注:1) 化合物数;训练集/验证集;2) 相关系数  $R^2$ ;训练集  $R^2$ /验证集  $R^2$ 。

Note: 1) Number of compounds: training set/validation set; 2) Correlation coefficient  $R^2$ : training set  $R^2$ /validation set  $R^2$ .

表 4 BCF 的 QSAR 通用模型

Table 4 The universal QSAR models of BCF

化合物数 <sup>1)</sup> Number of compound <sup>1)</sup>	描述符数 Number	描述符 Descriptor	相关系数 $R^2$ <sup>2)</sup> Correlation coefficient $R^2$ <sup>2)</sup>	建模方法 Method	应用域 Application domain	参考文献 Reference
290/315	5	DRAGON	0.73/0.72	遗传算法-变量子集选择策略 Genetic algorithm variable subset selection strategy	有 Yes	(2007) <sup>[6]</sup>
376/124/142	7	$\log P$ , max(Ea), Topology	0.765/0.774/0.739	人工神经网络 Artificial neural network	有 Yes	(2010) <sup>[7]</sup>
310/156/161	5	VC-logS etc.	0.751/0.616/ 0.464	最佳多元线性回归 Optimal multiple linear regression	有 Yes	(2013) <sup>[8]</sup>
624/156	12	DRAGON	0.809/0.738	多元线性回归 Multiple linear regression	有 Yes	本研究 This study

注:1) 化合物个数;训练集/验证集/测试集;2) 相关系数  $R^2$ ;训练集  $R^2$ /验证集  $R^2$ /测试集  $R^2$ 。

Note: 1) Number of compounds: training set/validation set/test set; 2) Correlation coefficient  $R^2$ : training set  $R^2$ /validation set  $R^2$ /test set  $R^2$ .

### 3.2 模型比较

国内外关于鱼类 BCF 的 QSAR 发展趋势是从早期的通过  $\log K_{ow}$  估算<sup>[25-26]</sup>逐步发展为基于片段常数法<sup>[27]</sup>、分子连接指数<sup>[13]</sup>、电性拓扑状态指数<sup>[28]</sup>、电负性距离矢量<sup>[29-30]</sup>、量子化学描述符<sup>[31]</sup>等参数的预测方法。根据模型训练集化合物种类不同,鱼类 BCF 的 QSAR 模型可分为两类,局域模型和通用模型。

局域模型是指训练集只包含相同基团或某一特定种类的化合物,这类模型针对性强,但应用域较为局限,如表 3 所示。

通用模型则是指训练集包含的化合物种类多,应用域较宽的 QSAR 模型。表 4 列出了通用模型,如 Papa 等<sup>[6]</sup>以 5 个 DRAGON 描述符,605 个化合物数据,通过遗传算法-变量子集选择(GA-VSS)建立了 QSAR 模型。Dearden 等<sup>[7]</sup>以拓扑指数和正辛醇/水分配系数共 7 个描述符,通过人工神经网络(ANN)建立了包含 642 个化合物的 QSAR 模型。Pi-

ira 等<sup>[8]</sup>以 VCC-logS 等 5 个描述符通过最佳多元线性回归(BMLR)方法,建立了包含 627 个化合物的 QSAR 模型。但目前已有通用模型的拟合能力还有待提高。

由比较可知,本研究所构建的模型涵盖化合物种类与个数多,且算法简洁透明,满足 OECD 关于 QSAR 模型的构建要求。从拟合度、稳健性、预测能力和应用域来看,能够作为预测有机化学品生物富集因子的 QSAR 模型,可用于化学品环境风险评估与管理。

通讯作者简介:郑玉婷(1989-),女,工程师,主要从事化学品环境风险评估等领域研究。

#### 参考文献 (References):

- [1] Petoumenou M I, Pizzo F, Cester J. Comparison between bioconcentration factor (BCF) data provided by industry to the European Chemicals Agency (ECHA) and data de-

- rived from QSAR models [J]. *Environmental Research*, 2015, 142: 529-534
- [2] Nendza M, Kühne R, Lombardo A. PBT assessment under REACH: Screening for low aquatic bioaccumulation with QSAR classifications based on physicochemical properties to replace BCF *in vivo* testing on fish [J]. *Science of the Total Environment*, 2018, 616-617: 97-106
- [3] Gramatica P, Papa E. QSAR modeling of bioconcentration factor by theoretical molecular descriptors [J]. *QSAR and Combinatorial Science*, 2003, 22: 374-385
- [4] Qin H, Chen J W, Wang Y, et al. Development and assessment of quantitative structure-activity relationship models for bioconcentration factors of organic pollutants [J]. *Chinese Science Bulletin*, 2009, 54(4): 628-634
- [5] Piira G, Silda S, Roncaglioni A, et al. QSAR model for the prediction of bioconcentration factor using aqueous solubility and descriptors considering various electronic effects [J]. *SAR and QSAR in Environmental Research*, 2010, 21(7-8): 711-729
- [6] Papa E, Dearden J C, Gramatica P. Linear QSAR regression models for prediction of bioconcentration factors by physicochemical properties and structural theoretical molecular descriptor [J]. *Chemosphere*, 2007, 67(2): 351-358
- [7] Dearden J C. QSAR modeling of bioconcentration factor using hydrophobicity, hydrogen bonding and topological descriptors [J]. *SAR and QSAR in Environmental Research*, 2010, 7-8(21): 671-680
- [8] Piir G, Sild S, Maran U. Comparative analysis of local and consensus quantitative structure-activity relationship approaches for the prediction of bioconcentration factor [J]. *SAR and QSAR in Environmental Research*, 2013, 24(3): 175-199
- [9] de Melo E B. A new quantitative structure-property relationship model to predict bioconcentration factors of polychlorinated biphenyls (PCBs) in fishes using E-state index and topological descriptors [J]. *Ecotoxicology and Environmental Safety*, 2012, 75: 213-222
- [10] Mansouri K, Consonni V, Durjava M K, et al. Assessing bioaccumulation of polybrominated diphenyl ethers for aquatic species by QSAR modeling [J]. *Chemosphere*, 2012, 89(4): 433-444
- [11] Jackson S H, Cowan-Ellsberry C E, Thomas G. Use of quantitative structural analysis to predict fish bioconcentration factors for pesticides [J]. *Journal of Agricultural and Food Chemistry*, 2009, 57(3): 958-967
- [12] 郑玉婷, 乔显亮, 杨先海, 等. 卤代有机化合物生物富集因子的定量结构-活性关系模型[J]. *生态毒理学报*, 2013, 8(5): 772-777
- Zheng Y T, Qiao X L, Yang X H, et al. Quantitative structure-activity relationship model for bioconcentration factors of halogenated organic compounds [J]. *Asian Journal of Ecotoxicology*, 2013, 8(5): 772-777 (in Chinese)
- [13] Lu X X, Tao S, Hu H Y, et al. Estimation of bioconcentration factors of nonionic organic compounds in fish by molecular connectivity indices and polarity correction factors [J]. *Chemosphere*, 2000, 41(10): 1675-1688
- [14] Devillers J, Bintein S, Domine D. Comparison of BCF models based on log P [J]. *Chemosphere*, 1996, 33(6): 1047-1065
- [15] Adolfsson-Erici M, Åkerman G, McLachlan M S. Measuring bioconcentration factors in fish using exposure to multiple chemicals and internal benchmarking to correct for growth dilution [J]. *Environmental Toxicology and Chemistry*, 2012, 31(8): 1853-1860
- [16] Inoue Y, Hashizume N, Yoshida T, et al. Comparison of bioconcentration and biomagnification factors for poorly water-soluble chemicals using common carp (*Cyprinus carpio* L.) [J]. *Archives of Environmental Contamination and Toxicology*, 2012, 63(2): 241-248
- [17] Martin J W, Mabury S A, Solomon K R, et al. Bioconcentration and tissue distribution of perfluorinated acids in rainbow trout (*Oncorhynchus mykiss*) [J]. *Environmental Toxicology and Chemistry*, 2003, 22(1): 196-204
- [18] Wei D B, Zhang A Q, Wu C D, et al. Progressive study and robustness test of QSAR model based on quantum chemical parameters for predicting BCF of selected polychlorinated organic compounds (PCOCs) [J]. *Chemosphere*, 2001, 44(6): 1421-1428
- [19] Park J H, Lee H J. Estimation of bioconcentration factor in fish, adsorption coefficient for soil and sediments and interfacial tension with water for organic nonelectrolytes based on the linear salvation energy relationships [J]. *Chemosphere*, 1993, 26(10): 1905-1906
- [20] Frisch M J, Trucks G W, Schlegel H B, et al. Gaussian 09, revision a.02 [CP]. Wallingford, CT: Gaussian, Inc., 2009
- [21] Talete S R L. DRAGON (Software for Molecular Descriptor Calculation) Version 6.0 [CP/OL]. [2018-07-18]. <http://www.talete.mi.it/>
- [22] Mohammad H F, Zahra G. Estimation of the volume of distribution of some pharmacologically important compounds from their structural descriptors [J]. *Journal of the Serbian Chemical Society*, 2011, 76(7): 1003-1014
- [23] Zarei K, Salehabadi Z. The shuffling multivariate adaptive regression splines and adaptive neuro-fuzzy inference system as tools for QSPR study bioconcentration factors of

- polychlorinated biphenyls (PCBs) [J]. *Structural Chemistry*, 2012, 23(6): 1801-1807
- [24] Fatemi M H, Ghorbannezhad Z. Estimation of the volume of distribution of some pharmacologically important compounds from their structural descriptors [J]. *Journal of the Serbian Chemical Society*, 2011, 76(7): 1003-1014
- [25] Neely W B, Branson D R, Blau G E. Partition coefficient to measure bioconcentration potential of organic chemicals in fish [J]. *Environmental Science & Technology*, 1974, 8(13): 1113-1115
- [26] Veith G D, DeFoe D L, Bergstedt B V. Measuring and estimating the bioconcentration factor of chemicals on fish [J]. *Journal of Fisheries Research Board of Canada*, 1979, 36: 1040-1048
- [27] 胡海瑛, 陶澍, 卢晓霞. 用片段常数法估算有机化合物在鱼体内的生物富集因子[J]. *环境科学学报*, 2001, 21(3): 271-276
- Hu H Y, Tao S, Lu X X. Predicting of BCFs in fish for organic chemicals using fragment constant method [J]. *Acta Scientiae Circumstantiae*, 2001, 21(3): 271-276 (in Chinese)
- [28] 冯长君, 唐自强, 杨伟华, 等. 有机物在鱼体中生物富集因子的预测[J]. *石油化工高等学校学报*, 2009, 22(1): 31-40
- Feng C J, Tang Z Q, Yang W H, et al. Prediction of bioconcentration factors of nonionic organic compounds in fish [J]. *Journal of Petrochemical Universities*, 2009, 22(1): 31-40 (in Chinese)
- [29] Qin L T, Liu S S, Liu H L. QSPR model for bioconcentration factors of nonpolar organic compounds using molecular electronegativity distance vector descriptors [J]. *Molecular Diversity*, 2010, 14(1): 67-80
- [30] Cui S H, Yang J, Liu S S, et al. Predicting bioconcentration factor values of organic pollutants based on MEDV descriptors derived QSARs [J]. *Science in China Series B (Chemistry)*, 2007, 50(5): 587-592
- [31] Qin H, Chen J W, Wang Y, et al. Development and assessment of quantitative structure-activity relationship models for bioconcentration factors of organic pollutants [J]. *Chinese Science Bulletin*, 2009, 54(4): 628-634 ◆