

DOI:10.7524/j.issn.0254-6108.2013.07.011

有机污染物土壤吸附预测模型及其影响因素^{*}

刘 娟 闻 洋^{**} 赵元慧

(国家环境保护湿地生态与植被恢复重点实验室, 东北师范大学环境学院, 长春, 130024)

摘 要 以 701 种有机化合物的土壤吸附系数作为数据集, 选取 594 种有机化合物作为训练集, 剩余 107 种作为测试集. 根据训练集化合物建立土壤吸附系数 K_{oc} 与辛醇/水分配系数 K_{ow} 的线性 and 非线性模型, 应用平均残差 (AE), 平均绝对残差 (AAE) 和均方根误差 (RMSE) 来检验模型的预测能力, 模型具有良好的预测能力. 同时, 比较不同类型的化合物的实测值与预测值, 发现部分同系物的预测值与实测值存在系统的偏差, 这些偏差主要是由吸附机理, 溶解度, 水解作用, 挥发作用, 实验误差等原因造成, 这些因素均会对土壤吸附系数的预测产生影响.

关键词 土壤吸附, 分配系数, 模型检验, 疏水性.

土壤沉积物对有机化合物的吸附作用是影响其在土壤中迁移、转化等环境行为的重要过程之一, 是污染物从非饱和溶液转入固相的主要途径, 能够强烈影响化合物在水环境与土壤环境中的含量、生物活性以及生物毒性. 因此, 研究化学物质在环境中的迁移转化过程, 讨论它们在环境中的归趋, 对评价有毒有害化学物质的环境行为及其对环境的潜在危险都具有重要意义.

土壤/水分配系数 (K_d) 是评价有机化合物环境行为的重要参数. 当土壤中有有机碳的质量分数 $f_{oc} > 0.001$, 溶解度较低时, 土壤/水分配系数 (K_d) 可以被标化为土壤吸附系数 (K_{oc}), $K_{oc} = K_d/f_{oc}^{[1-2]}$. 土壤的亲脂性可由正辛醇来替代, 所以正辛醇/水分配系数 (K_{ow}) 可以用来表达有机化合物在土壤/水体系中的分配行为. 前人通过建立 $\lg K_{oc}$ 与 $\lg K_{ow}$ 的定量结构-活性关系 (QSAR) 模型来预测土壤吸附系数^[3-4], 但这些模型都是根据非离子型化合物建立的, 并且数据量较少, 没有外部检验. 此外, 这些模型对于疏水性化合物 ($1 < \lg K_{ow} < 4$) 的预测结果较好^[5], 但对高疏水性化合物 ($\lg K_{ow} > 5$) 的预测能力较差^[6]. 目前文献中对 $\lg K_{oc}$ 与 $\lg K_{ow}$ 的非线性模型还没有报道.

本文以 701 种有机化合物的土壤吸附系数为研究对象, 计算了 19 种分子结构描述符, 建立了 $\lg K_{oc}$ 与 $\lg K_{ow}$ 的线性 and 非线性模型, 并检验了模型的预测能力. 同时分析了土壤吸附的影响因子.

1 材料与方法

1.1 K_{oc} 实测数据

本文所用的 K_{oc} 实测数据来自两篇文献^[7-8], Schrüüman 等^[7] 发表了 571 种 K_{oc} 数据, Gramatica 等^[8] 发表了 643 种 K_{oc} 数据, 两组数据中有 513 种有机化合物重复出现. 比较 K_{oc} 的重复值发现二者具有很好的可比性, 因此本研究选用 513 种重复出现 K_{oc} 的数据为 Schrüüman 文献值. 两篇文献中共有 701 种有机化合物的 K_{oc} 实测数据, 实测 $\lg K_{oc}$ 值范围为 $-0.30 \sim 6.50$, 这些数据通过电子邮件向作者索取. 本研究化合物类型包括醇、卤代烷烃和烯烃、酯、卤代环烷烃、烷基苯、卤代苯、硝基苯、苯酚、苯胺、有机酸、苯甲酸酯、联苯、多环芳烃、乙酰苯胺、芳香族化合物、有机磷和有机硫化合物等. 在建立模型过程中, 选择数据集的 85% 作为训练集 (594 种), 其余作为测试集 (107 种).

1.2 分子描述符

本研究计算了 19 种分子结构描述符来表征分子的大小、疏水性、溶解性、极性、可离子化率、柔韧

2013 年 1 月 30 日收稿.

^{*} 国家自然科学基金项目 (20977015); 东北师范大学中央高校基本科研业务费专项资金 (11SSXT138) 资助.

^{**} 通讯联系人, Tel: 0431-85099550; E-mail: weny501@nenu.edu.cn

性、氢键酸度和碱度. 包括分子量 (MW)、辛醇/水分配系数 (K_{ow})、表观分配系数 (D)、酸性电离常数 (pK_a -A)、碱性电离常数 (pK_a -B)、非离子态的百分率 (F_0)、正离子价态的百分率 (F_+)、负离子价态的百分率 (F_-)、耦合离子价态的百分率 ($F_{\pm}$) 等. 上述分子描述符由 ACD/Labs (Advanced Chemistry Development, Inc., <http://www.acdlabs.com>) 软件计算, 其中 $\lg K_{ow}$ 是碎片法计算值.

1.3 模型建立与评价

线性分析采用 Minitab 14.0 软件中逐步回归分析选择变量参数, 多元线性回归分析 (MLR) 建立模型, 非线性分析采用 TcWin 软件建立模型. 模型的评价指标包括: 化合物的个数 (n)、决定系数 (R^2)、标准偏差 (S)、Fisher 检验值 (F). 模型的预测能力通过外部检验来实现, 由平均残差 (AE)、平均绝对残差 (AAE)、均方根误差 (RMSE) 来表征. 相关参数的定义为:

$$AE = \sum_{i=1}^n (y_i - \hat{y}_i) / n \quad (1)$$

$$AAE = \sum_{i=1}^n |y_i - \hat{y}_i| / n \quad (2)$$

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n} \quad (3)$$

其中, n 为模型中化合物的个数, y_i 和 $\hat{y}_i$ 分别为第 i 个化合物 $\lg K_{oc}$ 的实测值和预测值.

2 结果与讨论

2.1 $\lg K_{oc}$ 与 $\lg K_{ow}$ 线性和非线性模型的建立

首先对训练集中有机化合物的土壤吸附系数和计算的 19 种分子结构描述符进行了逐步回归分析, 发现 $\lg K_{oc}$ 与 $\lg K_{ow}$ 具有显著的相关性, 引入其它分子描述符并不能显著地增加回归方程的相关性, 对化合物分类研究进一步发现对于所有化合物, 在广泛的范围内 $\lg K_{oc}$ 与 $\lg K_{ow}$ 并不完全为线性关系, 而是一种非线性关系 (图 1), 对于疏水性化合物 $0.5 < \lg K_{ow} < 7.5$ 的范围内 (见图 1 中 2 区), $\lg K_{oc}$ 与 $\lg K_{ow}$ 存在线性关系, 线性相关模型见方程 (4). 应用 TcWin 分析软件对训练集中所有化合物 $\lg K_{oc}$ 与 $\lg K_{ow}$ 进行非线性回归分析, 发现模型 (5) 能很好地描述 $\lg K_{oc}$ 与 $\lg K_{ow}$ 的非线性关系.

$$\lg K_{oc} = 0.661 + 0.672 \lg K_{ow} \quad (n=558 \quad R^2=0.76 \quad S=0.55 \quad F=1793) \quad (4)$$

$$\lg K_{oc} = 1.15 + 0.334 \lg K_{ow} + 0.0396 (\lg K_{ow})^2 + 0.00290 (\lg K_{ow})^3 - 0.000868 e^{\lg K_{ow}} \quad (n=594 \quad R^2=0.79 \quad S=0.55 \quad F=538) \quad (5)$$

线性和非线性模型的决定系数 (R^2) 分别为 0.76 和 0.79, 表明模型的拟合能力较好, $\lg K_{oc}$ 实测值与 $\lg K_{ow}$ 的拟合结果如图 1 所示.

图 1 中直线是模型 (4) 的拟合线, 曲线是模型 (5) 的拟合线. 模型 (5) 可以表示训练集中有机化合物土壤吸附数据 79% 的变化, 虽然结果相对较好, 但是观察图 1 发现一些化合物离群, 如图中空心方框所表示的有机化合物. 根据模型 (5) 对 $\lg K_{oc}$ 的实测值与预测值之间残差进行分析, 结果发现一些有机化合物存在系统偏差. 例如醇类、卤代烷烃、烯烃类和酯类的实测值均低于模型 (5) 的预测值. 对于可离子化合物, 有机酸类实测值低于模型 (5) 的预测值. 芳香族化合物的实测值高于模型 (5) 的预测值, 尤其是多环芳烃和聚芳香族杂环化合物. 表 1 列出了根据这些化合物实测值与模型 (5) 预测值的平均残差 (AE) 计算出的校正因子.

将校正因子加入模型 (5) 建立一个新模型, 如模型 (6) 所示.

$$\lg K_{oc} = \text{方程 (5)} + \text{校正因子} \quad (6)$$

模型 (6) 能够很好地预测所有化合物的土壤吸附系数. 加入校正因子后标准偏差明显降低 ($S = 0.49$). 这些化合物存在系统偏差可能是由于吸附机理、溶解度、水解作用、挥发作用、实验误差等原因, 这在下面将做详细讨论.

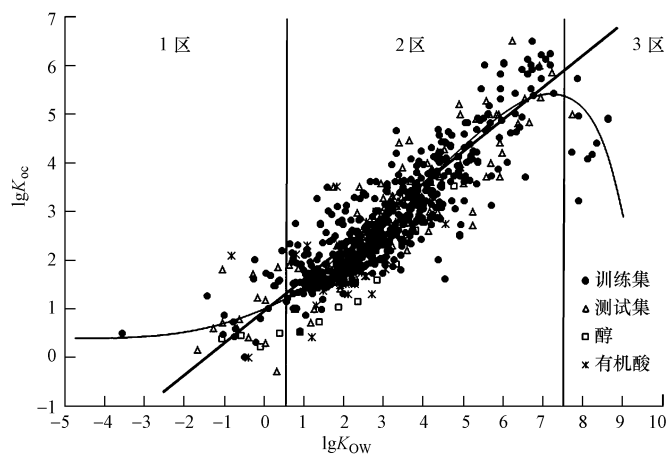


图 1 lgK_{oc}实测值与 lgK_{ow}的相关性图

Fig. 1 Plot of measured lgK_{oc} against lgK_{ow}

表 1 特殊组化合物的校正因子

Table 1 Correction factors for some class-specific compounds	
化合物类型	校正因子
醇	-0.78
卤代烷烃和烯烃	-0.31
酯	-0.37
有机酸	-0.40
多环芳烃	0.44
聚芳香族杂环化合物	0.97

2.2 模型检验

应用测试集 107 种有机化合物对模型的预测能力进行外部检验. 同时与 KOCWIN 软件中(EPI Suite 4.0)两种模型进行比较(以 1-MCI 和 lgK_{ow}为参数的计算模型), 所计算的平均残差(AE), 平均绝对残差(AAE)和均方根误差(RMSE)值见表 2.

表 2 模型(4)、(5)、(6)和 KOCWIN 模型的 lgK_{oc} 预测值分析

Table 2 lgK _{oc} values predicted from models (4), (5), (6) and KOCWIN models									
模型	训练集(n = 594)			测试集(n = 107)			数据集(n = 701)		
	AE	AAE	RESE	AE	AAE	RESE	AE	AAE	RESE
(4)	0.00	0.45	0.60	0.16	0.52	0.65	0.03	0.46	0.61
(5)	0.00	0.41	0.55	0.08	0.49	0.63	0.01	0.43	0.56
(6)	0.01	0.36	0.49	0.01	0.44	0.57	0.01	0.37	0.50
lgK _{ow} -based	-0.01	0.38	0.53	0.14	0.52	0.67	0.01	0.40	0.56
MCI-based	-0.05	0.39	0.54	-0.12	0.54	0.73	-0.06	0.41	0.57

表 2 结果表明,对于所有化合物,非线性模型(5)的预测能力要好于线性模型(4). 然而,对于疏水性化合物,线性模型和非线性模型的预测能力相近. 对所有数据集,最优的模型是模型(6),平均绝对残差(AAE)为 0.37,平均残差(AE)为 0.01. 为进一步评价模型质量,将本研究模型与 KOCWIN 软件的两种模型(lgK_{ow}-based 和 MCI-based)进行了比较(表 2). 与其相比,非线性模型(5)的精确度稍差,但是模型(5)简单,并且只用了一个参数. 非线性模型(6)的预测能力要好于其它模型,其预测值的平均绝对残差小于 0.4 个 lg 单位,见图 2(图中虚线表示 3 倍标准偏差).

2.3 土壤吸附的影响因子

有机化合物在土壤中的吸附是一个复杂的过程,可能包括多种作用模式和机制,许多因子可以影响有机化合物在水相和固相的浓度,如吸附机理、溶解度、水解作用、挥发作用、实验误差等,导致土壤吸附

系数的预测值高于或低于实测值,影响土壤吸附的预测。

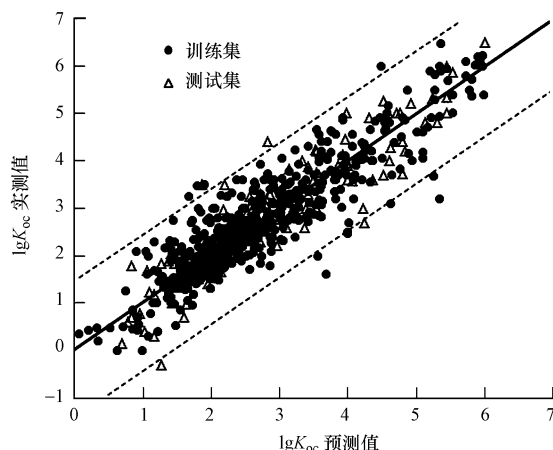


图2 训练集和测试集化合物 $\lg K_{oc}$ 的实测值与预测值的相关性图

Fig.2 Plot of observed $\lg K_{oc}$ against predicted $\lg K_{oc}$ for training and test sets

2.3.1 吸附机理

对于疏水性化合物($0.5 < \lg K_{ow} < 7.5$),吸附过程中土壤有机碳含量可以作为一个重要的决定因素,土壤有机质对非极性化合物的吸附起主要作用,所以疏水性化合物的 $\lg K_{oc}$ 与 $\lg K_{ow}$ 呈线性相关(图1中2区).对于亲水性化合物($\lg K_{ow} < 0.5$),疏水性吸附减弱,土壤中其它物质可以成为重要的吸附剂,因此根据 $\lg K_{ow}$ 预测的 $\lg K_{oc}$ 值并不准确.土壤有机质的亲水性吸附变得显著^[2,9],导致 $\lg K_{oc}$ 不随 $\lg K_{ow}$ 的增加而变化,而是接近于常数(图1中1区).

2.3.2 溶解度

土壤吸附系数是有机化合物吸附在土壤/沉积物中的浓度与溶解在水中浓度的比值.只有完全溶解在水中的有机化合物才能够被土壤或沉积物吸附^[10-11].低溶解度导致可用性下降,同时水中溶解的有机质对化合物也具有吸附作用,导致高疏水性化合物($\lg K_{ow} > 7.5$)的吸附作用下降, $\lg K_{oc}$ 随着 $\lg K_{ow}$ 的增加而下降(图1中3区).

2.3.3 离子化

可离子化化合物在水中以多种形态存在,有机化合物的离子态和非离子态对土壤吸附系数贡献不同,已为人们所公知^[12-13].为了研究离子化率对土壤吸附系数的影响,分别计算了有机物中性态的辛醇/水分配系数(非离子态在辛醇/水中浓度的比值) K_{ow} 和可离子化有机物的表观分配系数 D ,即非离子态在辛醇中浓度和水中总浓度(非离子态和离子态)的比值, D 和 K_{ow} 的关系如公式(7、8)所示.

$$D(\text{一元酸}) = K_{ow} / (1 + 10^{pH - pK_a}) = K_{ow} F_0 \quad (7)$$

$$D(\text{一元碱}) = K_{ow} / (1 + 10^{pK_a - pH}) = K_{ow} F_0 \quad (8)$$

其中, F_0 是有机酸或碱的非离子态在水中总浓度的百分比.

如果离子态和非离子态对土壤吸附系数有相同的贡献,用模型(4、5)根据 $\lg K_{ow}$ 计算的 $\lg K_{oc}$ 预测值应接近实测值.如果只有非离子态对土壤吸附系数有贡献,用模型(4、5)根据 $\lg D$ 计算的 $\lg K_{oc}$ 预测值应接近实测值.考察有机酸类化合物的 $\lg K_{oc}$ 实测值与预测值发现,用 $\lg K_{ow}$ 预测的 $\lg K_{oc}$ 要偏高于实测值(表1),而用 $\lg D$ 预测的 $\lg K_{oc}$ 要偏低低于实测值,这表明虽然非离子态在土壤吸附过程中起主要作用,但是阴离子对土壤吸附过程也有贡献,带阴离子的羧基可以与土壤中铝离子相互作用,对吸附系数产生贡献,但是与非离子态的疏水性作用相比,这种作用相对较弱.

2.3.4 转化

土壤吸附过程中存在代谢转化过程, $\lg K_{oc}$ 的实测值是根据母体化合物的浓度得到的,而不是母体化合物和代谢产物的总浓度,所以转化可以降低母体化合物在水相中的浓度,导致预测值偏高.一些有机化合物在溶液中能够迅速转化,如酯类化合物在溶液中可以水解为羧酸类和醇类化合物,羟基苯酚类在水中易被氧化成苯醌类化合物,这也解释了为什么酯类化合物的预测值要高于实测值(表1).

2.3.5 挥发作用

挥发作用可以影响土壤吸附系数的实验测定,减少有机化合物在溶液中的浓度^[14].通过HENRYWIN 软件(EPI Suite 4.0)计算 701 种有机化合物的亨利常数来研究挥发作用,结果表明对于大部分高挥发性化合物(卤代烷烃和烯烃),lgK_{oc}的实测值低于模型(5)的预测值(表 1),这种差异是由吸附过程中有机化合物挥发作用导致的.

2.3.6 土壤/溶液比例

在平衡状态下,土壤/溶液比例是影响吸附实验准确度的一个重要参数.当吸附量在 20%—80% 之间时,吸附实验的准确度较高,其最优值在 50% 左右^[15].土壤吸附量按公式(9)计算.

土壤吸附量(%) = $\frac{m_s}{m_s + m_w} = \frac{K_{oc} \times f_{oc} \times m_w \times W_s/V_w}{K_{oc} \times f_{oc} \times m_w \times W_s/V_w + m_w} = \frac{K_{oc} \times f_{oc} \times W_s/V_w}{K_{oc} \times f_{oc} \times W_s/V_w + 1}$ (9)

其中,m_s为土壤吸附有机化合物的质量(mg);W_s为土壤的重量(kg);m_w为有机化合物溶解在水中的质量(mg);V_w为溶液的体积(L);f_{oc}为土壤中有有机碳的质量分数.

根据公式(9)可以计算不同土壤/溶液比例时的土壤吸附量,结果见表 3.

表 3 不同吸附系数的化合物在不同土壤/溶液比例时的土壤吸附量

Table 3 Percentage of sorption at different soil/solution ratios							
(土壤/溶液比例)/ (kg·L ⁻¹)	lgK _{oc}						
	0	1	2	3	4	5	6
1/1	1	9	50	91	99	100	100
1/50	0.02	0.2	2	17	67	95	100
1/100	0.01	0.1	1	9	50	91	99

表 3 结果表明对于 lgK_{oc} < 1 和 lgK_{oc} > 5 化合物,土壤吸附量不在 20%—80% 的范围内.所以较低或较高吸附系数的化合物实验准确度较差,这解释了多环芳烃和聚芳香族杂环化合物的实测值与预测值存在较大误差的原因(表 1).

2.3.7 实验误差

实验数据来自两篇文献,两组数据集集中有 513 种重复的有机化合物,其中 384 种具有相同或相近的实测值,它们可能来自同一数据源.其余 129 种有机化合物 lgK_{oc} 的平均残差(AE)和平均绝对残差(AAE)分别为 0.01 和 0.29,有些化合物的土壤吸附差值高于一个数量级,因此实验误差是影响土壤吸附系数的一个重要的原因.另外土壤类型、暴露浓度、实验条件和测定时间都可能导致 lgK_{oc} 实测值存在一定的差异.有些有机化合物也可能与土壤存在特殊的相互作用,例如多环芳烃和聚芳香族杂环化合物,导致模型(5)的预测值要低于实测值(表 1).

3 结论

对于疏水性化合物(0.5 < lgK_{ow} < 7.5),疏水性是其在土壤吸附过程中的主要驱动力,lgK_{oc} 与 lgK_{ow} 呈线性相关.然而对于所有的化合物,lgK_{oc} 与 lgK_{ow} 呈非线性相关.根据训练集 594 种有机化合物建立了土壤吸附系数的非线性模型,决定系数(R²)为 0.79.应用测试集 107 种有机化合物对模型的预测能力进行检验,检验结果表明,模型具有良好的预测能力.此外,残差分析发现一些化合物存在系统偏差,导致这些偏差的原因包括,亲水性与疏水性化合物的吸附机理不同,高疏水性化合物的溶解度较低,酯类化合物的水解作用,挥发性化合物的挥发作用,土壤吸附系数较高或较低时实验误差较大等.

参 考 文 献

[1] Karickhoff S W. Organic pollutant sorption in aquatic systems[J]. Journal of Hydraulic Engineering, 1984,110(6): 707-735

[2] Baker J R, Mihelcic J R, Luehrs D C, et al. Evaluation of estimation methods of organic carbon normalized sorption coefficients[J]. Water Environment Research, 1997, 69(2): 136-145

[3] Gwalik B M, Sotirious N, Feicht N, et al. Alternative for the determination of soil adsorption coefficient, K_{oc} of non-ionicorganic

- compounds — A review[J]. *Chemosphere*, 1997, 34(12): 2525-2551
- [4] Doucette W J. Quantitative structure- activity relationships for predicting soil- sediment sorption coefficients for organic chemicals[J]. *Environmental Toxicology Chemistry*, 2003, 22(8): 1771-1788
- [5] Sabljic A, Güsten H, Verhaar H, et al. QSAR modeling of soil sorption. Improvements and systematics of $\lg K_{oc}$ vs. $\lg K_{ow}$ correlations[J]. *Chemosphere*, 1995, 31(12): 4489-4514
- [6] Baker J R, Mihelcic J R, Shea E. Estimating K_{oc} for persistent organic pollutants: Limitations of correlations with K_{ow} [J]. *Chemosphere*, 2000, 41(6): 813-817
- [7] Schüttmann G, Bert R U, Kühne R. Prediction of the sorption of organic compounds into soil organic matter from molecular structure[J]. *Environmental Science Technology*, 2006, 40(22): 7005-7011
- [8] Gramatica P, Giani, Papa E. Statistical external validation and consensus modeling: A QSPR case study for K_{oc} prediction[J]. *Journal of Molecular Graphics Modeling*, 2007, 25(6): 755-766
- [9] Franco A, Fu W J, Trapp S. Influence of soil pH on the sorption of ionizable chemicals: Modeling advance[J]. *Environmental Toxicology Chemistry*, 2009, 28(3): 458-464
- [10] Cheng H H. Organic residues in soils: Mechanisms of retention and extractability[J]. *International Journal of Environmental Analytical Chemistry*, 1990, 39(2): 165-171
- [11] Mackay D, Fraser A, Bioaccumulation of persistent organic chemicals: mechanisms and models[J]. *Environmental Pollution*, 2000, 110(3): 375-391
- [12] Bintein S, Devillers J. QSAR for organic chemical sorption in soils and sediments[J]. *Chemosphere*, 1994, 28(6): 1171-1188
- [13] Franco A, Trapp S. Estimation of the soil-water partition coefficient normalized to organic carbon for ionizable organic chemicals[J]. *Environmental Toxicology Chemistry*, 2008, 27(10): 1995-2004
- [14] Cousins I T, Beck A J, Jones K C. A review of the processes involved in the exchange of semi-volatile organic compounds (SVOC) across the air-soil interface[J]. *Science of the Total Environment*, 1999, 228(1): 5-24
- [15] Delle S A. Factors affecting sorption of organic compounds in natural sorbent/water systems and sorption coefficients for selected pollutants. A review[J]. *Journal of Physical Chemical Reference Data*, 2001, 30: 187-439

Predictive model for soil sorption of organic pollutants and influencing factors

LIU Xian WEN Yang* ZHAO Yuanhui

(Key Laboratory of Wetland Ecology and Vegetation Restoration of National Environmental Protection,
School of Environmental Sciences, Northeast Normal University, Changchun, 130024, China)

ABSTRACT

The hydrophobic parameter represented by the octanol/water partition coefficient (K_{ow}) is commonly used to predict the soil sorption coefficient (K_{oc}). In the present paper, soil sorption data for 701 compounds were analyzed. The results show that $\lg K_{oc}$ is linearly related to $\lg K_{ow}$ for the compounds with $\lg K_{ow}$ in the range of 0.5—7.5 and non-linearly related to $\lg K_{ow}$ for the compounds in a wide range of $\lg K_{ow}$. A non-linear model was developed between $\lg K_{oc}$ and $\lg K_{ow}$ for a wide range of compounds in the training set. The models were validated terms of average error (AE), average absolute error (AAE) and root-mean squared error (RMSE) by using an external test set with 107 compounds. Systemic predictive deviations were been observed for some class-specific compounds. The reasons for systemic deviations may be attributed to the difference of sorption mechanism for hydrophilic compounds, low solubility for highly hydrophobic compounds, hydrolysis of esters in solution, volatilization for volatile compounds and highly experimental errors for compounds with extremely high or low sorption coefficient.

Keywords: soil sorption, partition coefficient, model validation, hydrophobicity.