

DOI: 10.7524/j.issn.0254-6108.2017.04.2016072202

徐童, 陈景文, 李超, 等. 气相有机化学品与羟基自由基反应速率常数的 QSAR 模型[J]. 环境化学, 2017, 36(4): 703-709.

XU Tong, CHEN Jingwen, LI Chao, et al. QSAR models for predicting hydroxyl radical reaction rate constants with organic chemicals in the atmosphere[J]. Environmental Chemistry, 2017, 36(4): 703-709.

气相有机化学品与羟基自由基反应速率常数的 QSAR 模型*

徐 童¹ 陈景文¹ 李 超² 李雪花^{1**}

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

2. 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 气相中有机化学品与羟基自由基($\cdot\text{OH}$)反应速率常数(k_{OH})是评价化学品大气持久性的重要参数. 而化学品种类繁多, 现有的 k_{OH} 实验数据不能满足其生态风险评估的需求. 因此, 需要建立一种能够快速预测有机化学品 k_{OH} 的方法, 填补现有数据的缺失. 本研究搜集整理了 917 种有机化合物的 k_{OH} 实验值, 采用逐步多元线性回归法(MLR)和支持向量机法(SVM)分别构建了用于预测多种类化合物 $\lg k_{\text{OH}}$ 的线性和非线性定量结构-活性关系(QSAR)模型. 基于 MLR 的预测模型具有良好的拟合度(经校正的相关系数的平方 $R^2_{\text{adj.tr}} = 0.862$ 、均方根误差 $\text{RMSE}_{\text{tr}} = 0.455$)、稳健性(交叉验证系数 $Q^2_{\text{Loo}} = 0.856$)和预测能力(外部验证系数 $Q^2_{\text{ext}} = 0.850$). 基于 SVM 的预测模型也具有较好的拟合度($R^2_{\text{adj.tr}} = 0.915$, $\text{RMSE}_{\text{tr}} = 0.358$)和预测能力($Q^2_{\text{ext}} = 0.860$). 机理分析表明, 最高占据分子轨道能(E_{HOMO})、卤素原子在分子中所占的百分比($X\%$)和分子中具有 $=\text{CH}-$ 结构的数目(NdsCH)是最重要的 3 个描述符, 解释了数据集 78.3% 的方差. 采用 Williams 法表征模型的应用域. 所建立的模型可用于预测烷烃、烯烃、炔烃、芳香族化合物、醇类、酮类、醚类、醛类、酸类、酯类、卤代化合物、含氮化合物、含硫化合物等室温下的 k_{OH} .

关键词 有机化学品, 羟基自由基反应速率常数, 定量结构-活性关系(QSAR).

QSAR models for predicting hydroxyl radical reaction rate constants with organic chemicals in the atmosphere

XU Tong¹ CHEN Jingwen¹ LI Chao² LI Xuehua^{1**}

(1. Key Laboratory of Industrial Ecology and Environmental Engineering (Ministry of Education), School of Environmental Science and Technology, Dalian University of Technology, Dalian, 116024, China; 2. Key Laboratory for Wetland Ecology and Vegetation Restoration of National Environmental Protection, School of Environment, Northeast Normal University, Changchun, 130024, China)

Abstract: Reaction rate constant of hydroxyl radical with organic chemicals (k_{OH}) is an important parameter for assessing the atmospheric persistence. However, for the large number of chemicals, it is difficult to meet the demand of ecological risk assessment with existing experimental k_{OH} values. Thus, it is significant to develop an effective prediction method for k_{OH} . In this study, we collected k_{OH} values of 917 chemicals, and used Multiple Linear Regression (MLR) and Support Vector

2016 年 7 月 22 日收稿 (Received: July 22, 2016).

* 国家自然科学基金(21477016)和中央高校基本科研业务费专项资金(DUT16LK13)资助.

Supported by the National Natural Science Foundation (21477016) of China and the Fundamental Research Funds for the Central Universities (DUT15TD09).

** 通讯联系人, Tel: 0411-84707844, E-mail: lixuehua@dlut.edu.cn

Corresponding author, Tel: 0411-84707844, E-mail: lixuehua@dlut.edu.cn

Machine (SVM) to develop linear and non-linear QSAR model for $\lg k_{\text{OH}}$ prediction. Results of linear QSAR model show that it has satisfactory goodness-of-fit ($R_{\text{adj.tr}}^2 = 0.862$, $\text{RMSE}_{\text{tr}} = 0.455$), robustness ($Q_{\text{Loo}}^2 = 0.856$) and good predictability ($Q_{\text{ext}}^2 = 0.850$). Results of non-linear QSAR model show that it also has good fitness ($R_{\text{adj.tr}}^2 = 0.915$, $\text{RMSE}_{\text{tr}} = 0.358$) and predictability ($Q_{\text{ext}}^2 = 0.860$). Mechanistic analysis reveals that the energy of the highest occupied molecular orbital (E_{HOMO}), the percentage of halogen atoms in a molecule ($X\%$), and the number of $=\text{CH}-$ structure in a molecule (NdsCH) are the most important descriptors in the model, which explain the 78.3% variance of data set. The applicability domain of the models was characterized by the Williams plot. The developed models can be used to predict k_{OH} values (298 K) of alkanes, alkenes, alkynes, ketones, ethers, aromatic compounds, alcohols, acids, esters, halogenated compounds, nitrogen compounds, sulfur compounds.

Keywords: organic chemicals, hydroxyl radical reaction rate constants, quantitative structure-activity relationship.

据美国化学文摘社统计,全世界注册的化学品已达 1.11 亿多种,每天新增约 15000 种^[1].化学品种类繁多,有机化学品的污染已成为重要的环境问题之一^[2].污染物进入大气环境后,会发生一系列的迁移转化过程^[3].因此,有必要评价这些化学品在大气中的持久性,为预防和控制其污染提供依据.

羟基自由基($\cdot\text{OH}$)是大气中最重要的氧化剂,与 $\cdot\text{OH}$ 反应是多种有机化学品在大气环境中的主要去除途径.有机污染物与 $\cdot\text{OH}$ 反应的速率常数(k_{OH})可用于评价其大气持久性,也是对其进行生态风险评估的重要参数.然而,在种类繁多的化学品中仅有少部分(约 1000 种)具有 k_{OH} 数据.近年来,氟代烯烃^[4-5]和不饱和醇^[6-9]在大气中被检出,且浓度呈升高趋势^[10-13],但这两类化合物中仅有少数通过实验和计算方法得到了 k_{OH} ^[14-22].通过实验方法测定化学品 k_{OH} 数据耗时且成本高,为了弥补大量数据的缺失,克服实验测定的瓶颈,有必要发展预测 k_{OH} 的模型.

定量结构-活性关系(QSAR)可以有效地预测化合物的 k_{OH} ^[23-24].多元线性回归(MLR)是简洁并最具透明度的一种建模方法^[25],支持向量机(SVM)方法有较高的泛化能力^[26-27],可有效避免过拟合^[28].目前,已经报道了用于预测室温下多类有机化合物 k_{OH} 的模型^[29-31],但是,现有模型很少覆盖氟代烯烃和不饱和醇这两类化合物,用现有模型对这两类化合物的 k_{OH} 进行预测,结果与实验值相差较大.

本研究搜集了包含氟代烯烃和不饱和醇这两类化合物在内的 917 个化合物室温下的 k_{OH} 实验值,采用 MLR^[32-33]和 SVM^[28,32]方法,根据经济合作与发展组织(OECD)有关 QSAR 的构建和验证导则^[25],构建了用于预测多类化合物的线性和非线性 QSAR 模型,采用 Williams 法表征了模型的应用域.

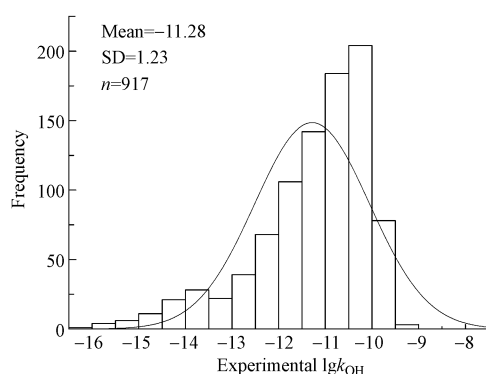
1 材料与方法 (Materials and methods)

1.1 数据收集

基于 Li 等^[29]的数据,补充搜集了氟代烯烃和不饱和醇的 45 个 $\lg k_{\text{OH}}$ 实验值,数据集中共包含 917 个化合物的 $\lg k_{\text{OH}}$ (298 K),图 1 为数据分布图.涵盖了烷烃、烯烃、炔烃、芳香族化合物、醇类、酮类、醚类、醛类、酸类、酯类、卤代化合物、含氮化合物、含硫化合物等多种类化合物.按 4:1 的比例随机划分数据集,734 个化合物用于模型的建立,183 个化合物用于模型的外部验证.

1.2 结构优化及分子描述符计算

模型建立所用的分子结构描述符包括量子化学描述符和 DRAGON 描述符.对分子结构进行优化和频率分析,含有 H、C、N、O、F、P、S、Cl、Se、Br 原子的分子在 B3LYP-6-311+G(d,p)水平下进行优化,含有 I、Si、Hg、Pb 原子的分子在 B3LYP-LANL2DZ 水平下进行优化.计算量子化学描述符:最高占据分子轨道能(E_{HOMO})、最低未占据分子轨道能(E_{LUMO})、分子的极化率(α)、偶极矩(μ)、氢键酸度(Q_{HMAX}).基于优化好的分子结构使用 DRAGON 软件(6.0)计算描述符,去掉常数项和近常数项,共得到 3138 个描述符.

图 1 $\lg k_{\text{OH}}$ 实验值数据分布图Fig.1 Distribution of experimental $\lg k_{\text{OH}}$ in the data set

1.3 模型建立和评价

使用 SPSS 19.0 软件中的逐步 MLR 方法,对变量进行筛选,构建预测模型.SVM 模型的建立使用 MATLAB 2014 软件,由 <http://www.ilovematlab.cn> 获得 SVM 开源程序.将由 MLR 模型筛选出来的描述符引入 SVM 模型,基于遗传算法优化模型参数,设置种群个数为 20(每次优化生成 20 个模型),终止代数为 100(截止条件),最终选取均方差最小的模型.SVM 模型由容量参数(C)、可影响模型预测能力的参数(γ)、不敏感损失函数(ε)等 3 个参数确定^[26-27]. C 是控制最大化边界和最小化误差平衡的规则化参数, C 太小将对训练集产生欠拟合, C 太大则会出现过拟合. ε 选小,回归估计精度要求高,但支持向量机数量增多; ε 选大,回归估计精度降低,支持向量机数量少,支持向量机的稀疏性大.核函数影响 SVM 的泛化能力, γ 为核函数中的参数,通过优化 γ 并找到最优解^[34-35].SVM 模型构建中采用的核函数为径向基函数: $K(u, v) = \exp(-\gamma|u-v|^2)$,其中 u 和 v 是自变量.

模型的建立遵循以下原则:模型自变量即描述符的数量少;模型的统计学参数中,校正的相关系数的平方(R_{adj}^2)、内部验证的交叉验证系数(Q_{LOO}^2)和 bootstrapping 计算的 Q_{BOOT}^2 高,均方根误差(RMSE)较低;模型中各描述符的方差膨胀因子(VIF)小于 10;自变量矩阵 M_X 的 K 相关系数(K_{XX})小于自变量与因变量矩阵 M_{XY} 的 K 相关系数(K_{XY}).模型建立后用去一法和 bootstrapping 法进行内部验证,未参与建模的 183 个化合物用于外部验证.模型评价的统计学参数包括:校正的相关系数的平方(R_{adj}^2)、均方根误差(RMSE)、内部验证的交叉验证系数(Q_{LOO}^2)、bootstrapping 计算的 Q_{BOOT}^2 .

模型应用域的特征采用 Williams 法,计算训练集和验证集中每个化合物的杠杆值(h)和标准残差(δ)作图.标准残差(δ)的绝对值大于 3 认为是离群点.杠杆值(h)和其警戒值(h^*)计算如下:

$$h^i = x_i^T (X^T X)^{-1} x_i \quad (1)$$

$$h^* = 3(k+1)/n \quad (2)$$

其中, x_i 是第 i 个化合物的分子结构描述符的变量; X 是分子结构描述符所构成的矩阵; k 是模型中包含的描述符的个数.

训练集中化合物的 h 值大于 h^* 时,说明在数据集中该分子的子结构出现较少,对模型有影响;验证集中化合物的 h 值大于 h^* 但在标准残差范围内时,说明其预测结果是该模型的外推结果,模型也适用于该化合物,但不确定性较大;化合物 h 值大于 h^* 同时又在标准残差范围外时,该化合物为离群点.

2 结果与讨论 (Results and discussion)

2.1 $\lg k_{\text{OH}}$ 预测模型

基于 MLR 的 QSAR 模型为:

$$\lg k_{\text{OH}} = -8.934 + 13.15 E_{\text{HOMO}} - 0.02400 X\% + 0.2430 \text{ NdsCH} + 0.3310 nR = Cp + 0.3960 \text{ CIC0} + 0.9330 nP - 0.004000 \alpha + 2.918 Q_{\text{HMAX}} + 0.7250 (C-020) + 0.1040 \text{ Mor14i} - 0.03900 \mu$$

$$n_{\text{tr}} = 734, R_{\text{adj}}^2 = 0.862, \text{RMSE} = 0.455, K_{XX} = 0.323, K_{XY} = 0.372, Q_{\text{LOO}}^2 = 0.856, Q_{\text{BOOT}}^2 = 0.797, n_{\text{ext}} = 183,$$

$R^2_{\text{adj,ext}}=0.840$, $\text{RMSE}_{\text{ext}}=0.468$, $Q^2_{\text{ext}}=0.850$

MLR 模型具有良好的拟合度和稳健性.模型进行外部验证,有良好的外部预测能力. K_{XX} 小于 K_{XY} ,各变量之间相关系数均小于 0.8,11 个描述符的 VIF 值均小于 10,说明了模型并不存在多重共线性.Y-scrambling 检验得出的 $R^2_{\text{YS}}(0.0143)$ 和 $Q^2_{\text{YS}}(-0.0190)$ 说明模型没有偶然相关性.描述符的含义和 VIF 值、 t 检验值见表 1.

最终选取的 SVM 模型中 $C=50.09$ 、 $\gamma=0.2318$ 、 $\varepsilon=0.03106$.统计学参数分别为 $n_{\text{tr}}=734$, $R^2_{\text{adj,tr}}=0.915$, $\text{RMSE}_{\text{tr}}=0.358$; $n_{\text{ext}}=183$, $R^2_{\text{adj,ext}}=0.860$, $\text{RMSE}_{\text{ext}}=0.454$, $Q^2_{\text{ext}}=0.860$.SVM 模型同样具有良好的拟合能力和预测能力.图 2 为 MLR 模型和 SVM 模型对化合物的 $\lg k_{\text{OH}}$ 的预测值和实验值的拟合图.

表 1 QSAR 模型中描述符含义、VIF 值、 t 值

Table 1 Meanings, VIF, and t values of descriptors in the QSAR model

描述符 Descriptor	含义 Meaning	VIF	$t(p)$
E_{HOMO}	最高占据分子轨道能量	2.165	21.24 (<0.001)
$X\%$	卤素原子在分子中所占的百分比	2.876	-14.96 (<0.001)
NdsCH	分子中具有=CH—结构的数目	1.154	12.52 (<0.001)
$nR=Cp$	末端 sp^2 杂化 C(sp^2) 主碳的个数	1.298	8.185 (<0.001)
CIC0	补偿信息论指数	4.662	9.364 (<0.001)
nP	分子中磷原子的个数	1.063	7.796 (<0.001)
α	分子的极化率	2.822	-5.623 (<0.001)
Q_{HMAX}	氢原子作为氢键给体的能力(氢键酸度)	1.198	7.484 (<0.001)
C-020	分子中=CX2 结构信息	1.400	6.362 (<0.001)
Mor14i	离子化势加权的 3D 分子结构描述符	2.168	3.250 (<0.001)
μ	分子偶极矩	1.275	-2.771 (<0.001)

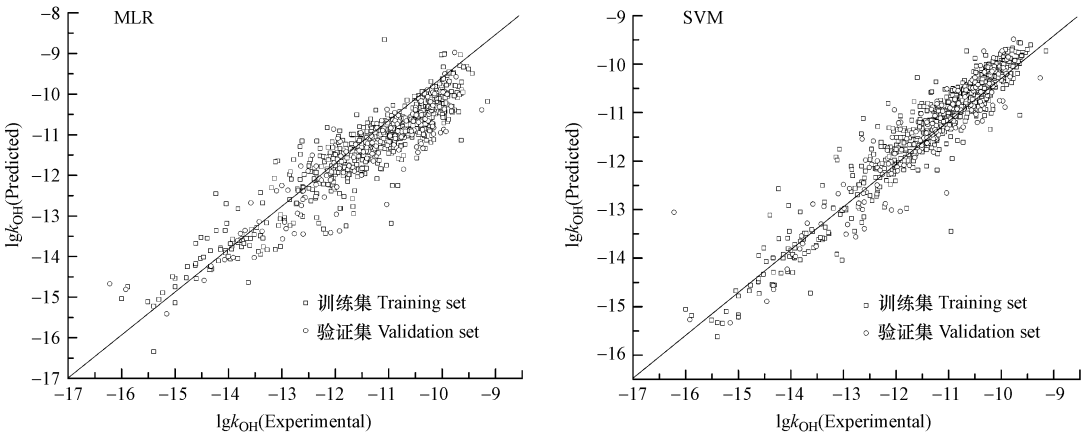


图 2 基于 MLR 和 SVM 的 QSAR 模型的 $\lg k_{\text{OH}}$ 实验值与预测值的拟合图

Fig.2 Plots of the predicted versus experimental $\lg k_{\text{OH}}$ values for the QSAR models based on MLR and SVM

2.2 应用域表征

图 3 为基于 MLR 和 SVM 模型的化合物的杠杆值对标准残差作出的 Williams 图.MLR 模型中离群点有 11 个,SVM 模型的离群点为 15 个.MLR 模型的离群点主要含有多氯代、全氟代、 $\text{C}=\text{O}$ 、 $\text{N}=\text{O}$ 等碎片结构,SVM 模型中离群点多含 $\text{C}=\text{O}$ 键和全卤代化合物.模型对全氟代正丁胺预测误差较大,该化合物含 27 个氟原子,此类化合物在数据集中仅 1 个,对该化合物预测误差较大的可能原因是因为数据集中该类化合物较少.数据集含 $\text{C}=\text{O}$ 结构的化合物中,仅有约 5%是离群点,预测误差较大,其他含 $\text{C}=\text{O}$ 结构化合物的 $|\delta|<3$,说明所构建的模型可以预测大多数此类化合物.另外,环丙烷、全氟代正丁胺、二硫二甲基在前人模型^[29]中也未被准确预测,为离群点.基于 MLR 的 QSAR 模型,剔除离群点后,模型的 R^2_{adj} 为 0.885,模型的拟合度和稳健性都有所提高.

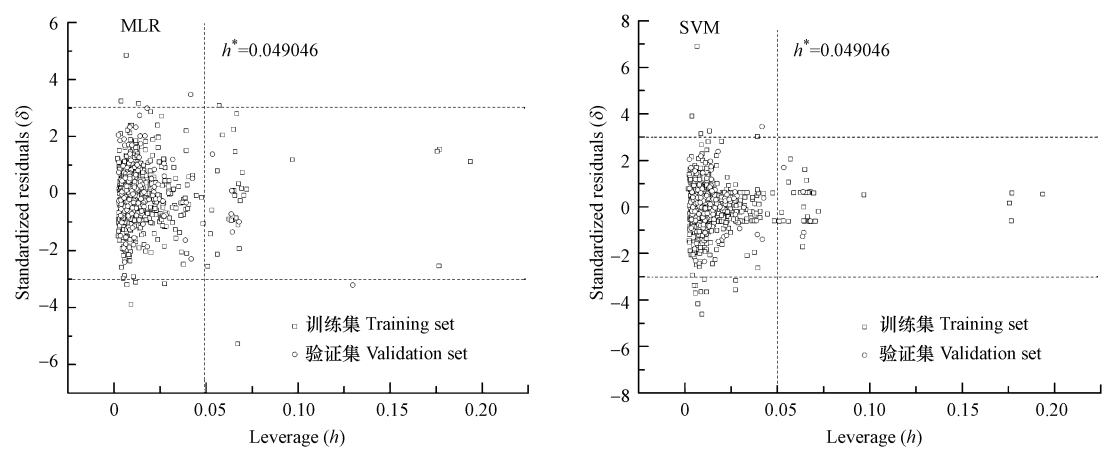


图 3 基于 MLR 和 SVM 的 QSAR 模型的 Williams 图

Fig.3 Williams plots of the QSAR models based on MLR and SVM

剔除离群点后基于 MLR 的 QSAR 模型为:

$$\lg k_{\text{OH}} = -8.632 + 14.18 E_{\text{HOMO}} - 0.02300 X\% + 0.2320 \text{ NdsCH} + 0.3250 nR = \text{Cp} + 0.4110 \text{ CICO} + 0.9220 nP - 0.005000 \alpha + 2.932 Q_{\text{HMAX}} + 0.7180 (\text{C-020}) + 0.1220 \text{ Mor14i} - 0.02800 \mu$$

$n_{\text{tr}} = 725, R^2_{\text{adj}} = 0.885, \text{RMSE} = 0.415, K_{\text{XX}} = 0.327, K_{\text{XY}} = 0.377, Q^2_{\text{LOO}} = 0.881, Q^2_{\text{BOOT}} = 0.795, n_{\text{ext}} = 181, R^2_{\text{adj,ext}} = 0.851, \text{RMSE}_{\text{ext}} = 0.434, Q^2_{\text{ext}} = 0.860$

2.3 模型机理分析

由表 1 可得 E_{HOMO} 、 $X\%$ 、 NdsCH 等 3 个描述符的 t 值明显高于其它描述符,可以解释数据集 78.3% 的方差.因此, E_{HOMO} 、 $X\%$ 、 NdsCH 是模型中最重要的 3 个描述符. E_{HOMO} 是指最高占据分子轨道,是与能量相关的量子化学描述符,可以表示分子的亲核性. $\cdot\text{OH}$ 是亲电试剂,高 E_{HOMO} 值的化合物更容易与 $\cdot\text{OH}$ 发生反应. $X\%$ 代表卤素原子在分子中所占的百分比,模型表示化合物中含有越多的卤素原子则与 $\cdot\text{OH}$ 的反应活性越低. NdsCH 表示分子中具有 $=\text{CH}-$ 结构的数目,化合物中含 $=\text{CH}-$ 结构越多,越容易与 $\cdot\text{OH}$ 反应.模型中描述符 $nR = \text{Cp}$ 、 CICO 、 nP 、 Q_{HMAX} 、 C-020 分别表示末端 sp^2 杂化 C (sp^2) 主碳的个数、补偿信息论指数、分子中磷原子的个数、氢原子作为氢键给体的能力(氢键酸度)、分子中 $=\text{CX}_2$ 结构信息.这些描述符的值与 $\lg k_{\text{OH}}$ 呈正相关.描述符 Mor14i 是 3D-Morse 描述符,是基于电子衍射建立的表征分子结构特征的描述符,与 $\lg k_{\text{OH}}$ 呈正相关.描述符 α 和 μ 分别为分子的极化率和偶极矩,为量子化学描述符,分子的极化率和偶极矩越大,其 $\lg k_{\text{OH}}$ 越小. α 主要表示分子的体积性质和分子电荷变形性,其数值越大,分子的电荷变形性就大,不利于与 $\cdot\text{OH}$ 发生反应. μ 与 α 呈正比,分子的极性越大,偶极矩也越大.

2.4 模型比较

表 2 列出了前人模型^[29-31]和本研究模型的评价参数.与 Li 等^[29]的模型相比,对氟代烯烃和不饱和醇两类化合物预测效果更好(Li 等^[29]所建模型的 $\text{RMSE} = 0.419$.本研究建立的基于 MLR 模型的 $\text{RMSE} = 0.455$,基于 SVM 模型的 $\text{RMSE} = 0.358$),其中,非线性模型比线性模型预测效果更好.本研究建立的模型包含的描述符中量子化学描述符物理意义明确,易于机理分析.

表 2 模型的统计学参数对比

Table 2 Comparison of statistical parameters for different k_{OH} prediction models									
模型 Models	算法 Methods	A	训练集 Training set			验证集 Validation set			
			n_{tr}	R^2	RMSE	n_{ext}	R^2_{ext}	RMSE_{ext}	Q^2_{ext}
Li 等 ^[29]	MLR	12	696	0.883	0.419	176	0.858	0.489	0.851
Roy 等 ^[30]	MLR	4	460	0.824	0.430	—	—	—	—
Roy 等 ^[30]	MLR	4	460	0.806	0.450	—	—	—	—
Wang 等 ^[31]	PLS	22	576	0.878	0.391	146	—	0.430	0.872
本研究	MLR	11	734	0.862	0.455	183	0.852	0.468	0.850
本研究	SVM	11	734	0.915	0.358	183	0.860	0.454	0.860

注: A 表示模型中描述符的个数. The number of descriptors in the model.

3 结论 (Conclusion)

本研究采用 MLR 和 SVM 两种方法,分别建立了用于预测气相中有机化学品的 $\lg k_{\text{OH}}$ 的线性和非线性 QSAR 模型,所构建模型均具有良好的预测能力和稳健性,对评价有机化学品的大气持久性和生态风险评价有重要意义.机理分析表明,最高占据分子轨道能 (E_{HOMO})、卤素原子在分子中所占的百分比 ($X\%$) 和分子中具有 $=\text{CH}-$ 结构的数目 (NdsCH) 等 3 个描述符对化合物的 $\lg k_{\text{OH}}$ 影响较大.所建立的模型可用于预测烷烃、烯烃、炔烃、芳香族化合物、醇类、酮类、醚类、醛类、酸类、酯类、卤代化合物、含氮化合物、含硫化合物等室温下的 k_{OH} 值.

参考文献 (References)

- [1] Chemical Abstracts Service[EB/OL]. [2016-5-10]. <https://www.cas.org/content/chemical-substances>.
- [2] 秦红. 有机污染物生物富集因子定量结构-活性关系的研究[D]. 大连:大连理工大学学位论文,2008.
QIN H. Study of quantitative structure-activity relationship models for bioconcentration factors of organic pollutants[D]. Dalian: Dalian University of Technology, 2008(in Chinese).
- [3] MACLEOD M, MCKONE T E, FOSTER K L, et al. Applications of contaminant fate and bioaccumulation models in assessing ecological risks of chemicals: A case study for gasoline hydrocarbons [J]. Environmental Science & Technology, 2004, 38(23):6225-6233.
- [4] BROWN J S. Introduction to hydrofluoro-olefin alternatives for high global warming potential hydrofluorocarbon refrigerants [J]. Hvac & Research, 2013, 19(6):693-704.
- [5] WUEBBLES D J, WANG D, PATTEN K O, et al. Analyses of new short-lived replacements for HFCs with large GWPs [J]. Geophysical Research Letters, 2013, 40(17):4767-4771.
- [6] WILDT J, KOBEL K, SCHUH-THOMAS G, et al. Emissions of oxygenated volatile organic compounds from plants part I: Emissions from lipoxygenase activity [J]. Journal of Atmospheric Chemistry, 2003, 45(2):143-172.
- [7] FALL R, KARL T, HANSEL A, et al. Volatile organic compounds emitted after leaf wounding: On-line analysis by proton-transfer-reaction mass spectrometry [J]. Journal of Geophysical Research Atmospheres, 1999, 104(D13):15963-15974.
- [8] ATKINSON R, AREY J. Atmospheric degradation of volatile organic compounds [J]. Cheminform, 2003, 35(103):4605-4638.
- [9] DALAI A, SCHOENAU G, DAS D, et al. Volatile organic compounds emitted during high-temperature alfalfa drying [J]. Biosystems Engineering, 2006, 94(1):57-66.
- [10] BAASANDORJ M, RAVISHANKARA A R, BURKHOLDER J B. Atmospheric chemistry of (*Z*)- $\text{CF}_3\text{CH}=\text{CHCF}_3$: OH radical reaction rate coefficient and global warming potential [J]. The Journal of Physical Chemistry A, 2011, 115(38):10539-10549.
- [11] GIERCZAK T, BAASANDORJ M, BURKHOLDER J B. OH + (*E*)- and (*Z*)-1-chloro-3,3,3-trifluoropropene-1 ($\text{CF}_3\text{CH}=\text{CHCl}$) reaction rate coefficients: Stereoisomer-dependent reactivity [J]. The Journal of Physical Chemistry A, 2014, 118(46):11015-11025.
- [12] VOLLMER M K, REIMANN S, HILL M, et al. First observations of the fourth generation synthetic halocarbons HFC-1234yf, HFC-1234ze (*E*), and HCFC-1233zd(*E*) in the atmosphere [J]. Environmental Science & Technology, 2015, 49(5):2703-2708.
- [13] PAPASAVVA S, LUECKEN D J, WATERLAND R L, et al. Estimated 2017 refrigerant emissions of 2,3,3,3-tetrafluoropropene (HFC-1234yf) in the United States resulting from automobile air conditioning [J]. Environmental Science & Technology, 2009, 43(24):9252-9259.
- [14] HENNE S, SHALLCROSS D E, REIMANN S, et al. Future emissions and atmospheric fate of HFC-1234yf from mobile air conditioners in Europe [J]. Environmental Science & Technology, 2012, 46(3):1650-1658.
- [15] RUSSELL M H, HOOGEWEG G, WEBSTER E M, et al. TFA from HFO-1234yf: Accumulation and aquatic risk in terminal water bodies [J]. Environmental Toxicology and Chemistry, 2012, 31(9):1957-1965.
- [16] TOVAR C M, BLANCO M B, BARNES I. Gas-phase reactivity study of a series of hydrofluoroolefins (HFOs) toward OH radicals and Cl atoms at atmospheric pressure and 298 K [J]. Atmospheric Environment, 2014, 88:107-114.
- [17] PERKINS R A, HUBER M L. Measurement and correlation of the thermal conductivity of 2,3,3,3-tetrafluoroprop-1-ene (R1234yf) and trans-1,3,3,3-tetrafluoropropene (R1234ze(*E*)) [J]. Journal of Chemical & Engineering Data, 2011, 56(12):4868-4874.
- [18] ORKIN V L, MARTYNOVA L E, ILICHEV A N. High-accuracy measurements of OH reaction rate constants and IR absorption spectra: $\text{CH}_2=\text{CFCF}_3$ and trans- $\text{CHF}=\text{CHCF}_3$ [J]. The Journal of Physical Chemistry A, 2010, 114(19):5967-5979.
- [19] RAMANJANEYULU C, RAJAKUMAR B. Kinetic parameters for the reaction of OH radical with cis- $\text{CHF}=\text{CHCHF}_2$, trans- $\text{CHF}=\text{CHCHF}_2$, $\text{CF}_2=\text{CHCHF}_2$ and $\text{CF}_2=\text{C}=\text{CHF}$: Hybrid meta DFT and CVT/SCT/ISPE calculations [J]. Journal of Fluorine Chemistry, 2015, 178:266-278.
- [20] HURLEY M D, BALL J C, WALLINGTON T J, et al. Atmospheric chemistry of the *Z* and *E* isomers of $\text{CF}_3\text{CF}=\text{CHF}$: Kinetics, mechanisms, and products of gas-phase reactions with Cl atoms, OH radicals, and O_3 [J]. Journal of Physical Chemistry A, 2007, 111

- (39):9789-9795.
- [21] GONZALEZ S, JIMENEZ E, BALLESTEROS B, et al. Hydroxyl radical reaction rate coefficients as a function of temperature and IR absorption cross sections for $\text{CF}_3\text{CH}=\text{CH}_2$ (HFO-1243zf), potential replacement of $\text{CF}_3\text{CH}_2\text{F}$ (HFC-134a) [J]. Environmental Science and Pollution Research, 2015, 22(7):4793-4805.
- [22] BALAGANESH M, RAJAKUMAR B. Rate coefficients and reaction mechanism for the reaction of OH radicals with (*E*)- $\text{CF}_3\text{CH}=\text{CHF}$, (*Z*)- $\text{CF}_3\text{CH}=\text{CHF}$, (*E*)- $\text{CF}_3\text{CF}=\text{CHF}$, and (*Z*)- $\text{CF}_3\text{CF}=\text{CHF}$ between 200 and 400 K; Hybrid density functional theory and canonical variational transition state theory calculations [J]. Journal of Physical Chemistry A, 2012, 116(40):9832-9842.
- [23] OBERG T. A QSAR for the hydroxyl radical reaction rate constant; validation, domain of application, and prediction [J]. Atmospheric Environment, 2005, 39(12):2189-2200.
- [24] KWOK E S C, ATKINSON R. Estimation of hydroxyl radical reaction rate constants for gas-phase organic compounds using a structure-reactivity relationship; An update [J]. Atmospheric Environment, 1995, 29(14):1685-1695.
- [25] OECD. Guidance document on the validation of (Quantitative) structure-activity relationship [(Q)SAR] models[M]. 2007. Available: <http://www.oecd-ilibrary.org>.
- [26] LIU H X, XUE C X, ZHANG R S, et al. Quantitative prediction of $\log k$ of peptides in high-performance liquid chromatography based on molecular descriptors by using the heuristic method and support vector machine [J]. Journal of Chemical Information and Computer Science, 2004, 44(6):1979-1986.
- [27] LIU H X, YAO X J, ZHANG R S, et al. Accurate quantitative structure-property relationship model to predict the solubility of C_{60} in various solvents based on a novel approach using a least-squares support vector machine [J]. The Journal of Physical Chemistry B, 2005, 109(43):20565-20571.
- [28] LI F, LIU J L, CAO L L. A comparative QSAR study on the estrogenic activities of persistent organic pollutants by PLS and SVM [J]. Emerging Contaminants, 2015, 1:8-13.
- [29] LI C, YANG X H, LI X H, et al. Development of a model for predicting hydroxyl radical reaction rate constants of organic chemicals at different temperatures [J]. Chemosphere, 2014, 95:613-618.
- [30] ROY P P, KOVARICH S, GRAMATICA P. QSAR model reproducibility and applicability: A case study of rate constants of hydroxyl radical reaction models applied to polybrominated diphenyl ethers and (benzo-) triazoles [J]. Journal of Computational Chemistry, 2011, 32(11):2386-2396.
- [31] WANG Y N, CHEN J W, LI X H, et al. Predicting rate constants of hydroxyl radical reactions with organic pollutants: Algorithm, validation, applicability domain, and mechanistic interpretation [J]. Atmospheric Environment, 2009, 43:1131-1135.
- [32] VAPNIK V N. Statistical Learning Theory[M]. New York: John Wiley & Sons, 1998.
- [33] LI X H, ZHAO W, LI J, et al. Development of a model for predicting reaction rate constants of organic chemicals with ozone at different temperatures [J]. Chemosphere, 2013, 92:1029-1034.
- [34] REN Y R, LIU H X, YAO X J, et al. Prediction of ozone tropospheric degradation rate constants by projection pursuit regression [J]. Analytica Chimica Acta, 2007, 589:150-158.
- [35] YAO X J, PANAYE A, DOUCET J P, et al. Comparative study of QSAR/QSPR correlations using support vector machines, radical basis function neural networks, and multiple linear regression [J]. Journal of Chemical Information and Computer Sciences, 2004, 44(4):1257-1266.