Journal of the Serbian Chemical Society
JSCS@tmf.bg.ac.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. *90* (9) 1055–1067 (2025)
JSCS–5439
Original scientific paper
Published 30 September, 2025

Environmental behavioral controls of polychlorinated biphenyls: Prediction of the soil sorption coefficient (K_{oc}) using multiple linear regression

AMINA CHAOUI[1,2], AMEL BOUAKKADIA[2,3], NOUREDDINE KERTIOU[2,3]* and HAMZA HADDAG[3]

[1]*Laboratory of Engineering and sciences of Advanced Materials (ISMA), ABBES Laghrour-University, Khenchela P.O 1252, 40004, Algeria,* [2]*Matter Sciences Department, Abbes Laghrour-University, Khenchela P.O 1252, 40004, Algeria and* [3]*Synthesis and Bio Catalysis Organic Laboratory, University of Badji Mokhtar-Annaba, Algeria*

(Received 19 February, revised 24 April, accepted 12 May 2025)

Abstract: The application of the quantitative structure property relationship (QSPR) approach helps to predict physicochemical properties of chemicals from their molecular structures. They are based on the translating and encoding of input information on theoretical molecular descriptors running by genetic algorithm techniques (GAs). In this research, the focus is mainly on the controlling and understanding the environmental fate and transport mechanisms of polychlorobiphenyls (PCBs) in soils. To achieve this goal, multiple linear regression (MLR) models were handed-down on a series of 188 PCBs to predict their soil sorption coefficient (log K_{oc}) which is critical to measuring their affinity. All statistical parameters obtained have indicated that the QSPR-MLR model has a very high predictive ability with a higher coefficient of determination (R^2 = = 0.9984) and lower root mean squared errors (*RMSE* = 0.0610). Moreover, the dominant molecular descriptors mentioned that the sorption process of these compounds on the surface of soils have been influenced by the size of molecule, polarizability and density.

Keywords: QSAR/QSPR study; polychlorinated biphenyls; molecular descriptors; partition coefficient; environmental chemistry.

INTRODUCTION

The study on the quantitative structure property relationship (QSPR) approaches, since their first formulation in the 1960s, is steadily growing following advances in environmental topics, medicine, pharmacology, biomedical and industrial research. The application of the QSPR method helps to predict physicochemical properties of compounds from their molecular structures. They are based

* Corresponding author. E-mail: kertiou.noureddine@univ-khenchela.dz
https://doi.org/10.2298/JSC250219035C

on the establishment of a mathematical model between theoretical molecular descriptors (MDs) and responses using linear and non-linear learning machines. These MDs encode and represent all input information running by genetic algorithm techniques (GAs). Besides, using these models helps researchers make predictions for new or untested compounds, thereby assessing their potential environmental impact and aiding in the development of safer and more sustainable chemicals. All the published papers, that have been done on the development of QSAR and QSPR models, revolve around their search to obtain the optimal model in order to predict the properties/activities or even though toxicities of chemicals under the guidelines of the five OECD[1] roles including here below: 1) a defined endpoint, 2) unambiguous algorithm, 3) a defined domain of applicability, 4) appropriate measures of goodness-of-fit, robustness and predictivity and 5) mechanistic interpretation of the molecular descriptors, if possible.[2]

Thus, in this research, the focus is mainly on the controlling and understanding the environmental fate and transport mechanisms of polychlorobiphenyls (PCBs) in soils. To achieve this goal, these chemometrics methods were applied to predict their soil sorption partition coefficient (K_{oc}) which is crucial for effective environmental management and pollution control (*e.g.*, some selected examples around modeling of PCBs using linear and non-linear QSAR/QSPR approaches.[3–10] log K_{oc} is a key parameter used to describe and to determine how organic compounds distribute themselves between aqueous and non-aqueous phases in the environment. It provides essential information for predicting the movement, bioavailability and potential environmental impact of organic pollutants.

PCBs, due to their resistance, lipophilicity, accumulation, their volatility and persistence for a long time, are frequently considered as hazardous compounds that can expose gravity impacts on human health, animals and on the environment. Several studies mentioned that PCB concentrations are highest in sediments and soil, particularly in industrial and urban area.[11,12] Over 90 % of PCB emissions originated from transformers and capacitor leaks and from fragmentizing operations.[13] That being the case, the movement of these organic compounds in soil is the primary and the principal cause of their toxic effects in the ecosystems. They migrate via the soil to vegetables, fruits, trees and eventually to humans through eating food, drinking water and breathing polluted air, causing disruption and contamination on biological systems in the form of diseases as cancer reproductive difficulties, immune system problems, kidney, stomach and Parkinson's diseases.[14–17] The number and the position of chlorines determine both the physical and the biological properties of each congener.[15]

In this study, the models were developed on a series of 188 polychlorinated biphenyls (PCBs) by leveraging the power of multiple linear regression techniques resulting from their simplicity, reproducibility and their higher predictive ability. The basis of these methods is the generation and selection of only influenced

molecular descriptors by using genetic algorithms. The following sections describe the methodologies and materials employed to build and validate QSPR models.

MATERIALS AND METHODS

Dataset collection and preparation

The collection of the experimental data is the significant and the unique source of input information accessible to QSPR modelers.[2] Experimental log K_{oc} values of 188 PCBs in this work were obtained from the literature[18] ranging from 4.2 to 8.03. The PCB codes, the formula, experimental and predicted log K_{oc} values of these PCBs as well as their Chemical Abstracts Service (CAS) numbers are listed in Table S-I of the Supplementary material to this paper. After this step, all molecular structures were drawn using the software ChemDraw[19] and saved as MDL.Molfile. A suggested format for the Hyperchem package[20] to optimize their geometries by the molecular mechanics force field method (MM+) on selecting the lowest energy. Besides, for more precise configuration a semi-empirical PM3 method was done to find the optimal conformation by computation Polak–Ribiere algorithm.

Descriptors calculation

Several molecular descriptors were calculated using Dragon software.[21] We sort these descriptors keeping only significant ones. Finally, 1143 molecular descriptors were retained for the QSPR model construction.

Dataset splitting

For the subsamples selection, a Kennard and Stone algorithm (CADEX) in MATLAB was applied. This non-random technique is usable just for only one splitting. So, it's considered as a deterministic method for the partitioning of the data into training and prediction sets in a way that ensures they are as uniformly spread out as possible. The algorithm, first determines which two samples are the furthest apart and have the largest Euclidean distance in order to determine which samples are the most representative. The remaining samples that are most distant from the previously chosen samples are then picked in each subsequent phase and appended to the bottom of the prior rank list. This process is repeated until a certain number of samples have been selected and ranked. The training set, is selected to build the model and learn the relationship between descriptors and property. It contains 132 compounds of PCBs (70 % of the total). While prediction set is used to evaluate the goodness-of-fit, robustness and the external predictive ability of the model. It contains 56 compounds of PCBs (30 % of the total).

MLR-model development and validation

The aim of multiple linear regression (MLR) is to model in the simplest way, capturing only the most relevant aspects of the chemical structure related to the end point.[2] QSARINS (version 2.2.4)[22,23] software served to achieve this step by running genetic algorithms. Different complemented steps were being verified in order to guarantee the goodness of the optimal MLR model selected. On the one hand, internal and external validation must be done with checking model predictivity of each training and predicting sub-sets, respectively. The internal validation was verified through: Determination Coefficient (R^2), adjusted squared regression coefficient (R^2adj), leave-one-out cross validation (Q^2_{LOO}), leave-many-out cross validation (Q^2_{LMO}), concordance correlation coefficient (*CCC*), root mean squared errors (*RMSE*), mean absolute error (*MAE*), *Y*-scrambling, standard error (*s*). And the external validation by CCC_{ext}, $RMSE_{ext}$, MAE_{ext}, standard error (s_{ext}).[2] Model performance was checked by QUIK rule, which referred to K_{XX} (the inter correlation of the independent variables) < K_{XY} (correlation We're not gonna

do it remove the external between independent variable matrix and dependent variable matrix) that indicate stability and reasonable prediction.[24] On the other hand, the definition of the applicability domain (AD) must be visualized by William's diagram aids in the knowledge and the definition of stable chemicals and the others which appeared as outlier. The outlier (the point with absolute value of standard deviation > 3) and high leverage points (the point where leverage values (h_i) were greater than the warning leverage (h^*)) could be simply and directly identified by Williams plot.[25] The leverage h_i is the i-th diagonal element of the Hat matrix, which links the target values to their predictions:[26]

$$hi = x_i^{\mathrm{T}} (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} x_i \quad (1)$$

Here, $\mathbf{X}$ is a 2D-descriptor matrix with n rows (compounds) and P columns (descriptors), $\mathbf{X}^{\mathrm{T}}$ is its transpose, $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ is the inverse of the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, and x_i is the row vector representing the descriptors of the query compound. A leverage value hi greater than the warning threshold h^* indicates that the compound i lies far from the center of the descriptor space and has a significant influence on the regression model. The critical value of leverage (h^*) is fixed at $3(p+1)/n$, where p is the number of predictor variables, and n is the number of observations. In this work, the ADs of the QSPR-MLR model were represented by the method of leverage value and Euclidean distance.[27]

RESULTS AND DISCUSSION

An excellent MLR-QSPR developed model was obtained based on two theoretical molecular descriptors (*EPS1* and *RDF030v*), the model is given as:

$$\log K_{\mathrm{oc}} = -1.8617 - 0.9159\,EPS1 - 0.0327\,RDF030v \quad (2)$$

For further assessing the predictive power of any model in QSPR field analysis internal and external validations are a necessary part. All statistical parameters of the QSPR-MLR model represented in Eq. (2) are presented in Table I.

TABLE I. Statistical parameters of QSPR model (internal validation); training set: n(tra) = 132

R^2	R^2_{adj}	*CCC*	Q^2_{LOO}	Q^2_{LMO}	R^2_{Yscr}	Q^2_{Yscr}	*s*	*F*	*RMSE*	*MAE*
0.9925	0.9924	0.9962	0.9922	0.9920	0.0148	–0.0318	0.0617	8525.0396	0.0610	0.0352

$R^2 = 0.9925 > 0.7$, $Q^2 = 0.9922 > 0.6$, $CCC = 0.9962 > 0.85$.[28,29] The difference between R^2 and Q^2 is low (no more than 10%).[2] R^2_{Yscr} and Q^2_{Yscr} values implied that the model had no accidental correlation ($R^2_{\mathrm{Yscr}} = 0.0148 < 0.3$, $Q^2_{\mathrm{Yscr}} = -0.0318 < 0.05$).[30] The great value of Fisher's parameter ($F = 8525.0396$) shows a highly significant model. Additionally, the lowest value of both *RMSE* (0.0610) and *MAE* (0.0352) factors indicate the less errors in predictions. Furthermore, a QUIK rule ($K_{XX} = 0.6727 < K_{XY} = 0.775$) showing a strength estimation. Despite the excellent internal results, this is not enough to propose it as a good model. So, an external validation (listed in Table II) must be undertaken to verify the predictive capabilities of our model, we resorted to its validation on a set designed for this purpose and chosen from the start.

TABLE II. Statistical parameters of QSPR model (external validation); prediction set: n(test) = 56

R^2_{ext}	CCC_{ext}	$RMSE_{ext}$	MAE_{ext}	Q^2-F1	Q^2-F2	Q^2-F3	r^2m
0.9757	0.9865	0.0880	0.0380	0.9746	0.9738	0.9844	0.9427

All values of Q^2-Fn > 0.70 and r^2m > 0.65,[2] with a noticeable decrease in the R^2_{ext} value by approximately 2 %. The regression line is shown in (Fig. 1), representing the relation between the measured values and the model's predictions, providing insight into the model's accuracy and predictive capabilities. In the training set, the observed log K_{oc} values were plotted against the corresponding experimental log K_{oc} values. A linear regression line was fitted to these points, ideally indicating a strong correlation and minimal deviation from the line of best fit. The closer the points are on this line, the more accurate the model is in predicting log K_{oc} values based on experimental data. The training set, used to develop the model, typically shows a high degree of fit, reflecting relationships within the data it was trained on. For the prediction set, which includes data not used during the model training, the observed log K_{oc} values were also plotted against the experimental log K_{oc} values. The fit of the linear regression line in this context demonstrates the model's generalizability and robustness in predicting set should align closely with the line, though some variability in environmental data. Upon examining the scatter plots and the linear regression lines, four outlier points were identified: PCB-000, PCB-095, PCB-155 and PCB-209. These points significantly deviated from the line adjustment, including a substantial discrepancy between the observed and the experimental log K_{oc} values. Such outliers suggest that the model's assumptions or inputs may not adequately account for specific characteristics or behaviors of these particular PCB compounds. PCB-000, PCB-209, PCB-155 and PCB-095 could represent unique cases where the chemical properties or environmental interactions differ significantly from the majority of the data set. A final point to highlight the MLR-QSPR model reliability is the definition of applicability domain (Fig. 1). It represents log K_{oc} experimental values *vs.* predictive values.

As shown (Fig. 2), the values of log K_{oc}(std) are between the limits ±3. One chemical compound PCB-209 (decachlorobiphenyl) from the training set, which appear clearly outside the limit set by the critical value h^* = 0.068 symbolized by the right parallel to the axis of log K_{oc}(std), is considered as influential compound. One other compound corresponding to PCB-000 (biphenyl) from training set appeared a little bit outside this interval with two others PCB-095 (2,2',3,5',6-pentachlorobiphenyl) from test set and PCB-155 (2,2',4,4',6,6'-hexachlorobiphenyl) are wrongly predicted. They are considered as outliers' compounds and cleaning of the data set is required. These outliers highlight the importance of further investigation to understand the underlying reasons for their deviation,

which could include experimental errors, unique chemical behaviors, or limitations in the model's scope. Addressing these outliers might involve refining the model or incorporating additional parameters to improve its overall accuracy and applicability. They were removed to observe their impact on the model accuracy and robustness. With removal of the outlier, the model was recalibrated and re-evaluated.

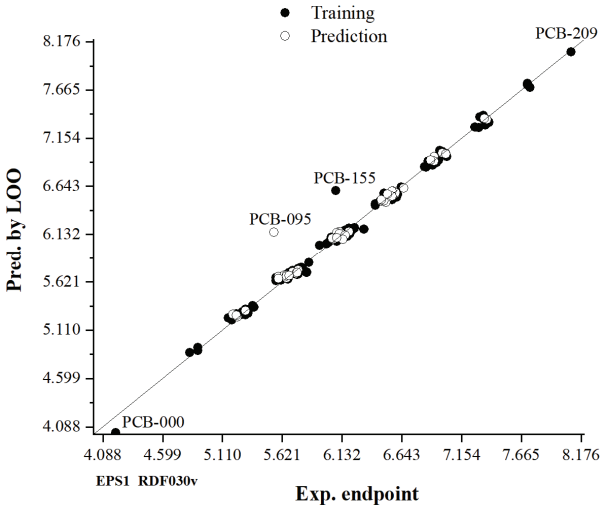


Fig. 1. Predicted *vs.* experimental log K_{oc} values for the training, validation sets.

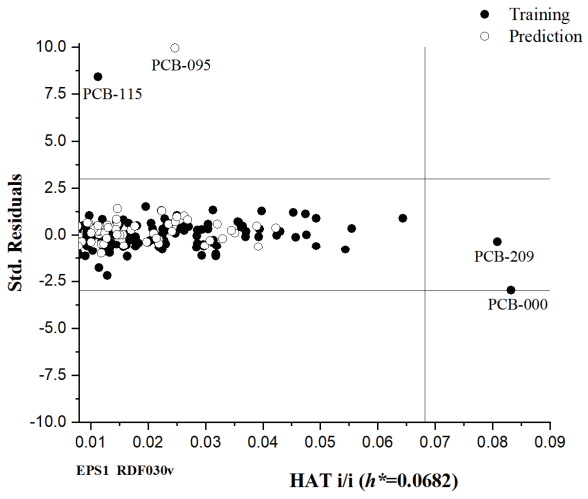
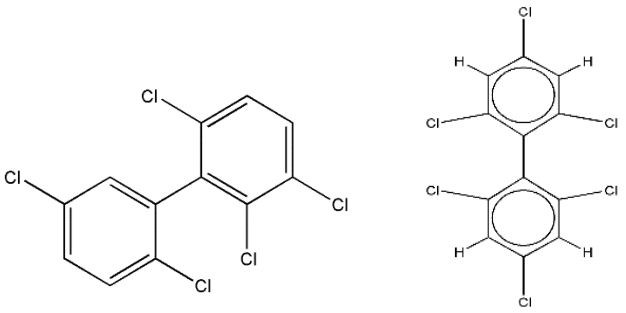


Fig. 2. Williams plot of the developed model.

Model refinement after outlier removal

PCBs, polychlorinated biphenyls are chemical compounds that consist of several isomers, differing in the position and number of chlorine atoms attached to the

biphenyl structure. In particular, PCB-095 and PCB-155 stand out from other PCBs due to their specific number of chlorine atoms and their configuration (Fig. 3).



A B

Fig. 3. 2D structure of PCB-095 (A) and PCB-155 (B).

The arrangement of chlorine atoms on the biphenyl rings is unique to PCB--095 and PCB-155. For example, in PCB-095, the chlorine atoms can be located in ortho, meta, or para positions, which influences the reactivity and stability of the compound. In the case of PCB-155, some chlorine atoms occupy the ortho and meta positions, while other PCBs may have different configurations.

PCBs are mixture of several isomers based on the arrangement of chlorine atoms. PCB-095 has characteristics that distinguish it from other isomers, which can impact its environmental and toxicological properties.

The structural difference of PCB-095 compared to other PCBs lies in the number and position of chlorine atoms, which influences its chemical and environmental properties. Its polarity is intermediate, affected by its chlorinated structure. When compared to other PCBs, variations in the number and position of chlorine atoms lead to differences in polarity, thus impacting the solubility, toxicity, and environmental behavior of these compounds.

In summary, the structural distinction between PCB-095 and PCB-155, as well as other PCBs, is based on the number and position of chlorine atoms in their structure.

The QSPR-MLR models re-developed after the elimination of PCB-095 and PCB-155 is as follows in Tables III and IV.

TABLE III. Statistical parameters of re-developed QSPR model (internal validation);); training set: n(tra) = 131

R^2	R^2_{adj}	CCC	Q^2_{LOO}	Q^2_{LMO}	R^2_{Yscr}	Q^2_{Yscr}	s	F	$RMSE$	MAE
0.9966	0.9966	0.9983	0.9964	0.9964	0.0151	–0.0318	0.0416	18857.1091	0.0411	0.0313

The model is given as:

$$\log K_{oc} = -1.8940 + 0.9206EPS1 - 0.0348RDF030v \quad (3)$$

Based on the results obtained, it is observed that the elimination of PCB-095 and PCB-155 did not affect the robustness of the model and its internal and external predictive capacities.

TABLE IV. Statistical parameters of re-developed QSPR model (external validation); test set: n(ext) = 55

R^2_{ext}	CCC_{ext}	$RMSE_{ext}$	MAE_{ext}	Q^2-F1	Q^2-F2	Q^2-F3	r^2m
0.9968	0.9979	0.0352	0.0283	0.9959	0.9958	0.9975	0.9783

According to Fig. 4, it is clear that the predicted log K_{oc} values were very similar to the experimental ones, *i.e.*, all points lying excellently along the regression line except PCB-000 which refers to the biphenyl compound. The compound PCB 209 is located on the line, but it deviates significantly due to its distinct structure. This is confirmed by the Williams diagram, which places it outside the application domain.

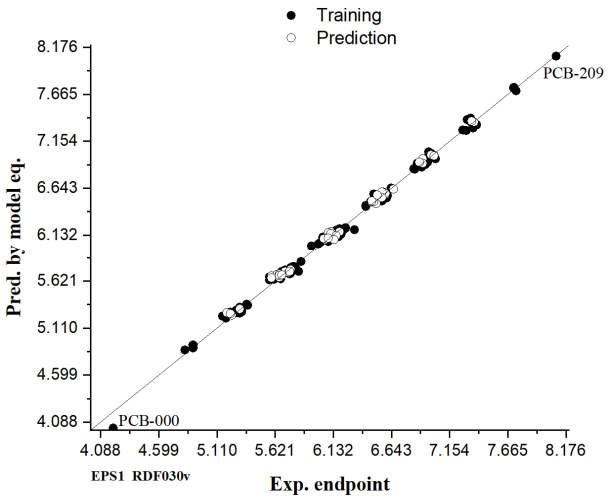


Fig. 4. Plot of predicted *vs.* experimental log K_{oc} partition coefficient.

The Fig. 5 shows that most data points from both training and prediction sets have leverage values below the critical threshold excluding PCB-000 and PCB--209 included in the training set, they have very unique structure, are considered influential chemical, and appear in the Williams graph to leverage higher than the warning h^* value of 0.069. In addition to being an influential point, PCB-000 is an outlier on the y-axis where it presents a standardized error lower than –3, along with PCB-120, which is also slightly aberrant. These points do not disproportionately influence the regression model. Consequently, the model predictions remain robust and unaffected by a few data points, maintaining overall stability.

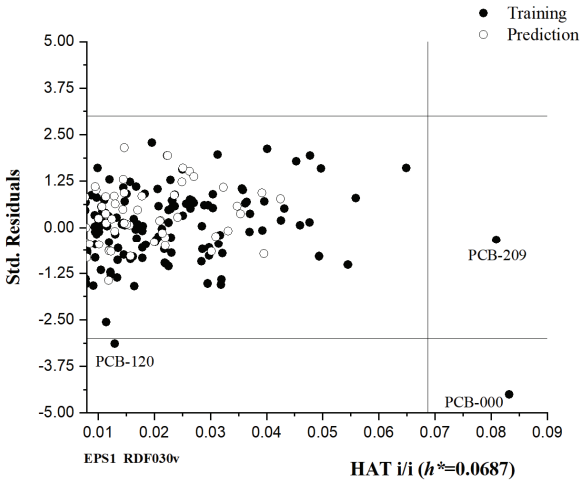


Fig. 5. Williams plot of the re-developed model.

Mechanistic interpretation

The *EPS1* descriptor is a measure used in molecular chemistry to describe the polarity and surface area of a molecule. It accounts for the surface area occupied by polar atoms, which is significant for understanding molecule solubility, permeability, and interactions with other molecules. The second descriptor *RDF030v* typically refers to a statistical measure that describes the spatial distribution of atoms within a molecule. Their negative coefficient implies that there is contradictory relationship between both *RDF030v* and log K_{oc} (the *RDF030v* value increases, log K_{oc} decreases). This could indicate that a higher number of atom pairs at the specified distance negatively impact the binding affinity, potentially due to steric interference or unfavorable spatial configuration. Specially, this descriptor might represent a particular distance range, capturing how atoms are distributed at a distance of 3.0 Å from a reference point within the molecule. This descriptor can provide insights into the three-dimensional arrangement and overall shape of the molecule.

After all statistical and graphical tests of the MLR-model created, the model represented in Eq. (3) is the most excellent. This is after removing the response outliers. It is challenging to determine the correct explanation for why the model failed to predict them, especially since all the chemicals have similar structures to each other. In addition to this, it is crucial to remember that variations in experimental conditions and measurement errors can also cause deviations from the expected results. Natural variations in soil composition and properties can also lead to differences in adsorption behavior. This means the model is reliable and produces consistent results. Also, we can see that the R^2 and Q^2 values for the permuted models (shown as red and yellow points, Fig. 6) are clustered near zero,

while the R^2 and Q^2 values of the actual model are significantly higher. This indicates that the original model is not due to chance correlations and that has identified a real relationship between the variables (the model is robust).

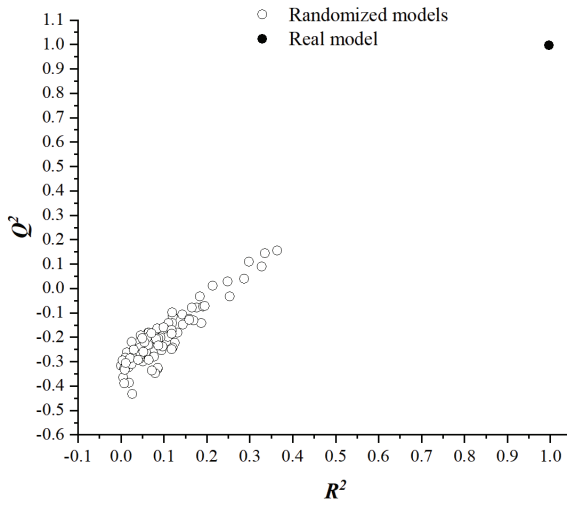


Fig. 6. *Y*-Scrambling plot.

Furthermore, variance inflation factor (*VIF*),[31] *T* and *p* values has several benefits (Table V) for improving the interpretation of the regression model results and enhancing its reliability. The larger *T* and the lower *VIF* and *p* values indicates that the independent variable (descriptor) has a significant effect on the dependent variable (log K_{oc}). *VIF* values are ensuring that there is no significant linear relationship between the independent variables.

TABLE V. Descriptions involved in the best developed MLR-models and corresponding *T* (test), *p* (statistically significant) and *VIF* (variance inflation factor) values

Descriptor	*T*	*P*	*VIF*
EPS1	196.45	0.000	1.601
RDF030v	–11.67	0.000	1.601

Comparison with previous work

To highlight the strengths of the MLR-QSPR model, a comparison of results with similar studies in the field was done and summarized in Table VI.

Our model's statistics are better than the two other models summarized in Table VI, except for the external validation (Q^2_{ext}). Yu *et al.*[3] model exhibits a $Q^2_{ext} = 0.9980$ greater than our model. This can be explained by the percentage of test compounds. We used 30 % of the dataset to externally validate our model

where Yu *et al.* used only ~19 % of their 138 PCBs. The external predictive ability of our model could be explained by relevant selection of descriptors.

TABLE VI. Comparison between our present work and previously published work for organic carbon/water partition coefficient (K_{oc}); KS – Kennard–Stone; SPM – SPM derived

No.	Ref.	Compd.	Dataset	Splitting method	Training	Test	R^2	Q^2_{LOO}	Q^2_{ext}	*RMSE*
1	Curr. study	PCBs	185	KS	130	55	0.9970	0.9968	0.9968	0,0378
2	Yu *et al.*[3]	PCBs	138	SPM	112	26	0.9960	0.9960	0.9980	0.0380
3	Yuan *et al.*[7]	PCBs	73	KS	55	18	0.9170	0.9020	0.8570	0.2530

CONCLUSION

To estimate the soil sorption partition coefficient and to determine which physicochemical molecular features have an influence on the sorption of a set of 188 polychlorinated biphenyls, a multiple linear regression QSPR model has been established in compliance with the OECD[1] principles. The entire database was divided into training and test sets by Kennard–Stone algorithm. Using QSARINS software, MLR model was developed using genetic algorithm techniques to select the main molecular descriptors. Two molecular descriptors were identified, namely *EPS1* and *RDF030v*, showing that the distribution behavior of the PCBs compounds is influenced by molecular size, polarizability and density. The final optimal model, obtained after removing outliers which had high R^2 and Q^2 values, had the distinct advantages of satisfactory stability and prediction performance as well as wide applicability domains. The proposed model can be used in the future to make reliable predictions for new compounds or compounds which have not yet been tested.

SUPPLEMENTARY MATERIAL

Additional data and information are available electronically at the pages of journal website: https://www.shd-pub.org.rs/index.php/JSCS/article/view/13257, or from the corresponding author on request.

ИЗВОД

ЕКОЛОШКА КОНТРОЛА ПОНАШАЊА ПОЛИХЛОРОВАНИХ БИФЕНИЛА: ПРЕДВИЂАЊЕ КОЕФИЦИЈЕНТА СОРПЦИЈЕ ЗЕМЉИШТА (K_{OC}) КОРИШЋЕЊЕМ МЕТОДЕ ВИШЕСТРУКЕ ЛИНЕАРНЕ РЕГРЕСИЈЕ

AMINA CHAOUI[1,2], AMEL BOUAKKADIA[2,3], NOUREDDINE KERTIOU[2,3] и HAMZA HADDAG[3]

[1]Laboratory of Engineering and sciences of Advanced Materials (ISMA), ABBES Laghrour-University, Khenchela P.O 1252, 40004, Algeria, [2]Matter Sciences Department, Abbes Laghrour-University, Khenchela P.O 1252, 40004, Algeria и [3]Synthesis and Bio Catalysis Organic Laboratory, University of Badji Mokhtar-Annaba, Algeria

Примена квантитативног односа структуре и особина (QSPR) помаже у предвиђању физичко–хемијских својстава на основу молекулских структура. Овај приступ заснован је на превођењу и кодирању улазних података о теоријским молекулским дескрипторима коришћењем техника генетских алгоритама. У овом истраживању, фокус је првенствено

на контроли и разумевању утицаја у животној средини и механизама транспорта полихлорованих бифенила (PCB) у земљишту. Да би се постигао овај циљ, модели вишеструке линеарне регресије (MLR) су примењени на серију од 188 PCB да би се предвидео њихов коефицијент сорпције земљишта (log K_{OC}) што је кључно за мерење њиховог афинитета. Сви добијени статистички параметри су показали да QSPR-MLR модел има веома високу предиктивну способност (R^2 = 0,9984 и *RMSE* = 0,0610). Поред тога, доминантни молекулски дескриптори указују на то да величина молекула, као и њихова поларизабилност и густина утичу на сорпциони процес ових једињења на површини земљишта.

(Примљено 19. фебруара, ревидирано 24. априла, прихваћено 12. маја 2025)

REFERENCES

1. S. Kherouf, N. Bouarra, A. Bouakkadia, D. Messadi, *J. Serb. Chem. Soc.* **84** (2019) 575 (https://doi.org/10.2298/JSC180820016K)
2. P. Gramatica, *Int. J. Quant. Struct. Prop. Relat.* **5** (2020) 63 (https://doi.org/10.4018/IJQSPR.20200701.oa1)
3. S. Yu, S. Gao, Y. Gan, Y. Zhang, X. Ruan, Y. Wang, L. Yang, J. Shi, *SAR QSAR Environ. Res.* **27** (2016) 2 (http://dx.doi.org/10.1080/1062936X.2016.1158734)
4. W. J. Doucette, *Environ. Toxicol. Chem.* **22** (2003) 1771 (http://doi.org/10.1897/01-362)
5. P. Gramatica, E. Giani, E. Papa, *J. Mol. Graph. Modell.* **25** (2007) 755 (https://doi.org/10.1016/j.jmgm.2006.06.005)
6. R. S. Kookana, R. Ahmad, A. Farenhorst, in *Non-First Order Degradation and Time-Dependent Sorption of Organic Chemicals in Soil,* W. Chen, A. Sabljic, S. A. Cryer, R. S. Kookana, Eds., ACS Symposium Series, Vol. 1174, American Chemical Society, Washington DC, 2014, pp. 222–236 (http://doi.org/10.1021/bk-2014-1174.ch012)
7. J. Yuan, S. Yu, T. Zhang, X. Yuan, Y. Cao, X. Yu, X. Yang, W. Yao*, Ecotox. Environ. Safety* **128** (2016) 171 (http://dx.doi.org/10.1016/j.ecoenv.2016.02.022)
8. Y. Wang, J. Chen, X. Yang, F. Lyakurwa, X. Li, X. Qiao, *Chemosphere* **119** (2015) 438 (http://dx.doi.org/10.1016/j.chemosphere.2014.07.007)
9. M. H. Fatemi, Z. G. Chahi, *SAR QSAR Environ. Res.* **23** (2012) 155 (http://dx.doi.org/10.1080/1062936X.2011.645876)
10. P. Singh, S. Pandit, R. Parthasarathi, in *Safety Evaluation and Risk Assessment*, H. Hong Ed., Elsevier Science, Amsterdam, 2023, pp. 77–87 (http://doi.org/10.1016/B978-0-443-15339-6.00026-6)
11. S. Mao, S. Liu, Y. Zhou, Q. An, X. Zhou, Z. Mao, Y. Wu, W. Liu, *Environ. Poll.* **271** (2021) 116171 (https://doi.org/10.1016/j.envpol.2020.116171)
12. Y. Dumanoglu, E. O. Gaga, E. Gungormus, S. C. Sofuoglu, M. Odabasi, *Sci. Tot. Environ. J.* **580** (2017) 920 (http://doi.org/10.1016/j.scitotenv.2016.12.040)
13. S. Kakareka, T. Kukharchyk, in *EMEP/CORINAIR Guidebook*, K. Breivik, Ed., European Environment Agency, Copenhagen, 2005 (https://www.eea.europa.eu/publications/EMEPCORINAIR5/Sources_of_PCB_emissions.pdf)
14. K. B. Moysich, R. J. Menzes, J. A. Baker, K. L. Falkner, *Rev. Environ. Health* **17** (2002) 263 (http://doi.org/10.1515/REVEH.2002.17.4.263)
15. D. O. Carpenter, *Rev. Environ. Health* **21** (2006) 1 (https://doi.org/10.1515/REVEH.2006.21.1.1)
16. L. M. Frazier, *J. Agromed.* **12** (2007) 27 (https://doi.org/10.1300/J096v12n01_04)
17. N. Othman, Z. Ismail, M. I. Selamat, S. H. S. Abdul Kadir, N. A. Shibraumalisi, *Int. J. Environ. Res. Public Health* **19** (2022) 13923 (https://doi.org/10.3390/ijerph192113923)

18. K. Jagiello, A. Sosnowska, S. Walker, M. Haranczyk, A. Gajewicz, T. Kawai, N. Suzuki, J. Leszczynski, T. Puzyn, *Struct. Chem.* **25** (2014) 997 (https://doi.org/10.1007/s11224-014-0419-1)
19. *ChemDraw*, CambridgeSoft Corporation, Cambridge, MA (http://camsoft.com/)
20. *Hyperchem^{TM}, release 6.03 for windows*, Hypercube, Inc., Gainesville, FL, 2000
21. R. Todeschini, V. Consonni, M. Pavan, *DRAGON Software for the Calculation of Molecular Descriptors, release 5.3 for windows*, Talete srl, Milano, 2006
22. P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, *J. Comp. Chem.* **34** (2013) 2121 (https://doi.org/10.1002/jcc.23361)
23. P. Gramatica, S. Cassani, N. Chirico, *J. Comp. Chem.* **35** (2014) 1036 (https://doi.org/10.1002/jcc.23576)
24. L. Gu, T. Zhu, M. Chen, *J. Environ. Chem. Eng*. **9** (2021) 105615 (https://doi.org/10.1016/j.jece.2021.105615)
25. X. Zhao, Y. Zhao, Z. Ren, Y. Li, *Comp. Biol. Chem*. **79** (2019) 177 (http://doi.org/10.1016/j.compbiolchem.2019.02.008)
26. H. Yu, D. Wondrousch, F. Li, J. Chen, H. Lin, L. Ji, *Chem. Res. Toxicol.* **28** (2015) 1541 (http://doi.org/10.1021/acs.chemrestox.5b00127)
27. H. Liu, M. Wei, X. Yang, C. Yin, X. He, Sci. *Total Environ.* **574** (2017) 1371 (https://doi.org/10.1016/j.scitotenv.2016.08.051)
28. N. Chirico, P. Gramatica, *J. Chem. Inf. Model.* **52** (2012) 2044 (https://doi.org/10.1021/ci300084j)
29. E. Carnescchi, A. A. Toropov, A. P. Toropova, N. Kramer, C. Svendsen, J. L. Dorne, E. Benfenati, *Sci. Total Environ.* **704** (2020) 135302 (http://doi.org/10.1016/j.scitotenv.2019.135302)
30. T. Zhu, H. Yan, R. P. Singh, Y. Wang, H. Cheng, *Env. Sci. Poll. Res. J.* **27** (2020) 17550 (http://doi.org/10.1007/s11356-019-06389-z)
31. N. Bouarra, S. Kherouf, N. Nadji, L. Nouri, A. Boudjemaa, S. Djerad, K. Bachari, *Chem. Prod. Process Model*. **19** (2024) 251 (http://doi.org/10.1515/cppm-2023-0024).