chemistry proceedings

MDPI

*Proceeding Paper*

# Prediction of n-Octanol/Water Partition Coefficients (Kow) for Pesticides Using a Multiple Linear Regression-Based QSPR Model

**Youssouf Driouche *, Meriem Ferfar, Amina Dridi, Amel Soussa, Ines Bekhouche, Souad Narsis, Rachida Mansouri and Souad Yahi**

Environmental Research Center (CRE), Annaba, Algeria; m.ferfar@cre.dz (M.F.); a.dridi@cre.dz (A.D.); a.soussa@cre.dz (A.S.); i.bekhouche@cre.dz (I.B.); s.narsis@cre.dz (S.N.); r.mansouri@cre.dz (R.M.); s.yahi@cre.dz (S.Y.)

* Correspondence: ydriou@gmail.com or y.driouche@cre.dz; Tel.: +213-(0)663780175

**Abstract:** This study developed a QSPR model to predict the n-octanol/water partition coefficient (log Kow) of 56 pesticides. Molecular descriptors were calculated using Dragon software. A genetic algorithm and variable subset selection identified key descriptors. The model, built by multiple linear regression, showed strong performance ($R^2 = 0.9322$, $Q^2_{LOO} = 0.9089$, $Q^2_{ext} = 0.9277$). The dataset was split using the Kennard-Stone algorithm to ensure representative sampling. Internal and external validations confirmed robustness and predictive power. This model offers a reliable tool for estimating log Kow, supporting environmental risk assessment and the evaluation of pesticide behavior and toxicity.

**Keywords:** QSPR; log kow; pesticides

Academic Editor(s): Name

Published: date

**Citation:** Driouche, Y.; Ferfar, M.; Dridi, A.; Soussa, A.; Bekhouche, I.; Narsis, S.; Mansouri, R.; Yahi, S. Prediction of n-Octanol/Water Partition Coefficients (Kow) for Pesticides Using a Multiple Linear Regression-Based QSPR Model. *Chem. Proc.* **2025**, *volume number*, x. https://doi.org/10.3390/xxxxx



## 1. Introduction

Pesticides are unique among chemicals because they are intentionally introduced into the environment to manage pests and protect crops and industrial goods. However, their effects are not limited to target species; they also impact non-target organisms and ecosystems. Frequent use can reduce biodiversity [1]. Many pesticides are persistent, remaining in soil or leaching into water bodies, which leads to widespread contamination. Due to their chemical nature, they can bioaccumulate in living organisms and affect human health through food chains. The environmental risks of intensive pesticide use are well documented and significant [2].

Quantitative structure–activity relationships (QSARs) offer a way to predict chemical toxicity based on molecular structure, even before synthesis. These models are especially useful when experimental data are scarce [1,2]. QSPR/QSAR methods rely on the idea that molecular properties are linked to specific structural features, called descriptors [3]. These computational tools can forecast chemical behavior, identify key structural elements, and reduce the need for experimental testing, making them cost- and time-effective in areas like drug development [4].

Physicochemical properties—such as vapor pressure, solubility, and partition coefficients-are central to understanding how organic compounds behave in the environment [5]. Among these, the n-octanol/water partition coefficient (Kow) is crucial. A high log Kow suggests a higher potential for bioaccumulation in organisms. This parameter is also used to infer systemic action and environmental fate of compounds [6]. In QSAR models,

log Kow is often employed as a descriptor of toxicity [7–9]. The partition coefficient is defined as the ratio of equilibrium concentrations of a substance in a biphasic system like octanol and water [10], with octanol mimicking biological lipids [11].

This study aimed to develop QSAR models for predicting the acute toxicity of pesticides and to construct a statistical model for estimating their log Kow. Using multiple linear regression (MLR), the model identifies which molecular descriptors significantly influence variation in log Kow across the pesticide dataset.

## 2. Materials and Method

### *2.1. Experimental Data*

This study used a dataset of 56 pesticides obtained from the published work of G.S. Patil [12]. The partition coefficient values were converted to log Kow to minimize data variability. The dataset was split into two subsets: 42 compounds for training and 14 for external validation.

### *2.2. Descriptors Generation*

The molecular structures of all compounds were built using HyperChem [13] and initially optimized via the MM+ force field with the Polak–Ribiere algorithm. Final geometries corresponding to the lowest energy conformers were obtained using the semi-empirical PM3 method at the restricted Hartree–Fock level, without configuration interaction, applying a gradient norm threshold of 0.001 kcal·Å$^{-1}$·mol$^{-1}$. These optimized structures were used to calculate 1664 molecular descriptors with Dragon software (version 5.4) [14]. Additionally, quantum chemical descriptors—including HOMO, LUMO, the HOMO–LUMO gap (ΔHL), and ionization potential-were computed using the PM3 method in HyperChem and considered during descriptor selection for model development.

## 3. Results and Discussion

The dataset of 56 compounds was partitioned into two subsets using the Kennard–Stone algorithm implemented in CADEX: a calibration set of 42 compounds and a validation set of 14, as shown in Table 1. The objective was to select a reduced set of descriptors that best account for the variation in the dependent variable (log Kow). Descriptor selection was performed using genetic algorithms provided in the Mobydigs software [15].

The application of the GA–VSS led to several effective models for predicting the Kow of pesticide chemicals based on various sets of molecular descriptors. The best model, derived from the 56 pesticide compounds, demonstrated high predictive accuracy and was established using the following regression equation:

$$\textbf{log Kow} = 1.63 + 0.301 \times \textbf{Polarizability} - 0.798 \times \textbf{O-058} - 0.230 \times \textbf{nHAcc} - 4.55 \times \textbf{E1u} \quad (1)$$

Equation (1) incorporates four distinct categories of molecular descriptors, summarized in Table 2. The model's performance was assessed using predictive metrics such as $Q^2_{LOO}$ and $Q^2_{LMO}$, along with the coefficient of determination ($R^2$) to evaluate the goodness of fit. Additionally, the standard deviation of prediction error (SDEP) and the standard deviation of calculation error (SDEC) within the applicability domain are reported.

As shown in Table 3, the fitting and validation metrics are consistently high, confirming the model's strong predictive performance. The selected descriptors effectively capture the variation in the partition coefficient. The $R^2$ value reflects a well-fitted model, and the small difference between $R^2$ and $Q^2_{LOO}$ indicates strong internal robustness, further supported by a high Fisher statistic. The close values of SDEC and SDEP suggest that the model's predictive ability is consistent with its fitting accuracy. External validation results,

including $Q^2_{ext}$ and $SDEP_{ext}$, demonstrate the model's reliability in predicting data not used during its training phase.

**Table 1.** The data set and the corresponding observed and predicted values of log (Kow) by MLR for the training and test sets.

| ID | Object | Status | Log $Kow_{Exp}$ | Log $Kow_{Calc}$ | Log $Kow_{Pred}$ |
|---|---|---|---|---|---|
| 1 | aa Aldicarb | Training | 1.1300 | 1.6633 | 1.7055 |
| 2 | ac Azinphos-ethyl | Training | 3.400 | 3.8846 | 3.9365 |
| 3 | ad Azinphos-methyl | Training | 2.6900 | 3.1298 | 3.1679 |
| 4 | af Bromophos-ethyl | Training | 5.6800 | 4.6503 | 4.5145 |
| 5 | ah Carbofuran | Training | 1.6300 | 2.4009 | 2.4373 |
| 6 | ak Chlordimeform | Training | 2.8900 | 3.1089 | 3.1360 |
| 7 | al Chlorfenvinphos | Training | 3.8100 | 3.7036 | 3.6872 |
| 8 | am Chlorpyrifos | Training | 4.9600 | 4.2289 | 4.1797 |
| 9 | aq Dichlorvos | Training | 1.4700 | 0.7627 | 0.6021 |
| 10 | as Disulfoton | Training | 4.0200 | 4.0222 | 4.0225 |
| 11 | at Disulfoton-sulfone | Training | 1.8700 | 1.8826 | 1.8843 |
| 12 | au Disulfoton-sulfoxide | Training | 1.7300 | 2.7337 | 2.8112 |
| 13 | av Ethion | Training | 5.0700 | 5.0293 | 5.0176 |
| 14 | ax Fenitrothion | Training | 3.400 | 3.0789 | 3.0405 |
| 15 | ay Fensulfothion | Training | 2.2300 | 2.9470 | 2.9820 |
| 16 | az Fensulfothion-sulfide | Training | 4.1600 | 4.3343 | 4.3540 |
| 17 | bb Fensulfothion-sulfoxide | Training | 2.5900 | 1.0347 | 0.8113 |
| 18 | bd Fenofos | Training | 3.8900 | 3.9289 | 3.9362 |
| 19 | bf Isofenphos | Training | 4.1200 | 4.3999 | 4.4441 |
| 20 | bg Leptophos | Training | 5.8800 | 5.4460 | 5.3194 |
| 21 | bh Malathion | Training | 2.8400 | 2.4827 | 2.3990 |
| 22 | bi Methidathion | Training | 2.4200 | 2.4721 | 2.4771 |
| 23 | bk Paraoxon | Training | 1.9800 | 1.9711 | 1.9700 |
| 24 | bl Parathion | Training | 3.7600 | 3.4904 | 3.4617 |
| 25 | bm Parathion-amino | Training | 2.6000 | 3.5076 | 3.5685 |
| 26 | bn Parathion-methyl | Training | 2.9400 | 2.7758 | 2.7507 |
| 27 | bo Phorate | Training | 3.8300 | 3.7061 | 3.6934 |
| 28 | bp Phorate sulfoxide | Training | 1.7700 | 2.3529 | 2.3898 |
| 29 | bq Phorate sulfone | Training | 1.9800 | 1.5262 | 1.4635 |
| 30 | br Phosalone | Training | 4.3800 | 4.0743 | 4.0489 |
| 31 | bs Phosmet | Training | 2.7800 | 2.4588 | 2.4096 |
| 32 | bt Phoxim | Training | 4.3900 | 3.9275 | 3.8851 |
| 33 | bu Pirimiphos ethyl | Training | 4.8500 | 4.7015 | 4.6813 |
| 34 | bv Pirimiphos methyl | Training | 4.2000 | 3.9870 | 3.9580 |
| 35 | bw Propoxur | Training | 1.5500 | 2.1930 | 2.2259 |
| 36 | bx Ronnel | Training | 4.8100 | 3.7066 | 3.6269 |
| 37 | by Temephos | Training | 5.9500 | 6.4691 | 6.6715 |
| 38 | bz Terbufos | Training | 4.4800 | 4.2510 | 4.1733 |
| 39 | cb Terbufos-sulfoxide | Training | 2.4800 | 2.9511 | 3.0095 |
| 40 | cc Terbufos sulfone | Training | 2.2100 | 2.1428 | 2.1249 |
| 41 | cd Triazophos | Training | 3.5500 | 4.2176 | 4.2735 |
| 42 | ce Trichlorfon | Training | 0.4300 | 1.0645 | 1.1735 |
| 43 | ab Aminocarb | Test | 1.7300 | - | 2.3518 |
| 44 | ae Bromophos | Test | 4.8800 | - | 3.9279 |
| 45 | ag Carbaryl | Test | 2.3100 | - | 2.5399 |
| 46 | ai Carbophenothion | Test | 5.1200 | - | 5.1524 |

| | | | | | |
|---|---|---|---|---|---|
| 47 | aj Carbophenothion-methyl | Test | 4.8200 | - | 4.4169 |
| 48 | an Chlorpyrifos-methyl | Test | 4.3000 | - | 3.4384 |
| 49 | ao Diazinon | Test | 3.8100 | - | 4.1168 |
| 50 | ap Dicapthon | Test | 3.6200 | - | 3.1335 |
| 51 | ar Dimethoate | Test | 0.7700 | - | 1.6071 |
| 52 | aw Fenamiphos | Test | 3.2300 | - | 3.6289 |
| 53 | bc Fenthion | Test | 4.0900 | - | 3.9071 |
| 54 | be Iodofos | Test | 5.1600 | - | 4.3957 |
| 55 | bj Methomyl | Test | 0.1300 | - | 0.8981 |
| 56 | cf Trichloronat | Test | 5.2200 | - | 4.4896 |

**Table 2.** Description of the selected descriptors by GA.

| Descriptors | Class | Signification |
|---|---|---|
| Polarizability | Hyperchem descriptor | Polarizability defined as the dipole moment of a molecule induced by an electric field of unit intensity. |
| O-058 | Atom-centered fragments | Defined hydrophobicity. |
| nHAcc | Functional group counts | Total number of Ns, Os and Fs in the molecule, excluding N with a formal positive charge, higher oxidation states and the pyrrolyl form of N. |
| E1u | WHIM descriptors | 1st component accessibility directional WHIM index/unweighted. |

**Table 3.** Results and statistical parameters of GA-MLR.

| $R^2$ | $Q^2_{LOO}$ | $Q^2_{EXT}$ | SDEC | SDEP | $SDEP_{EXT}$ | F | s |
|---|---|---|---|---|---|---|---|
| 93.22 | 90.89 | 92.77 | 0.450 | 0.520 | 0.546 | 92.4052 | 0.511 |

The $R^2$ value reflects the model's fitting quality, while the small difference between $R^2$ and $Q^2_{LOO}$ indicates strong robustness, reinforced by a high Fisher statistic. The similarity between SDEC and SDEP suggests that the model's predictive accuracy is consistent with its calibration performance. External validation metrics, including $Q^2_{ext}$ and $SDEP_{ext}$, confirm the model's reliability in predicting the behavior of compounds outside the training set.

The symmetrical distribution of errors around the zero line indicates that the model does not exhibit systematic bias. Figure 1 presents the $Q^2$ and $R^2$ coefficients, comparing the real model (black dot) with the randomized models (red circles). The $Q^2$ values of the randomized models are consistently below 20, and in many cases negative, confirming that the developed model is based on real structure–property relationships rather than random chance.

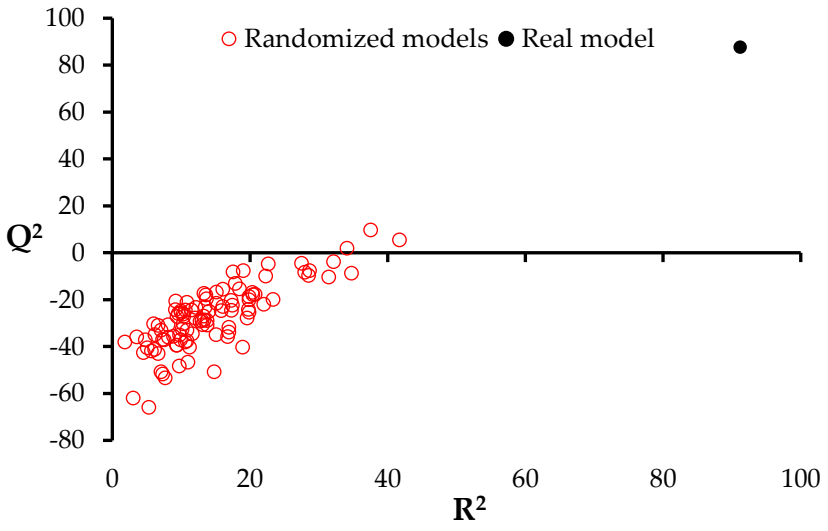


**Figure 1.** Randomization test.

## 4. Conclusions

In this study, the QSPR approach was employed to correlate the log Kow values of 56 pesticides with theoretical molecular descriptors selected using a genetic algorithm. Multiple linear regression was applied to identify linear relationships between the descriptors (independent variables) and log Kow (dependent variable). The resulting model demonstrated optimal performance in terms of goodness of fit, internal and external validation, and predictive accuracy.

**Author Contributions:** Conceptualization, Y.D.; methodology, Y.D. and M.F.; software, Y.D.; validation, Y.D., M.F. and A.D.; formal analysis, M.F.; investigation Y.D., M.F. and A.D.; resources, Y.D.; data curation, Y.D., M.F. and A.D.; writing—original draft preparation, I.B. and S.Y.; writing—review and editing, R.M., S.N. and A.S.; visualization, Y.D.; supervision, Y.D.; project administration, Y.D. All authors have read and agreed to the published version of the manuscript.

**Funding:**

**Institutional Review Board Statement:**

**Informed Consent Statement:**

**Data Availability Statement:**

**Acknowledgments:** The authors gratefully acknowledge the Environmental Research Center in Annaba, Algeria, for their support and valuable contributions to this work.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Bradbury, S.P. Predicting modes of toxic action from chemical structure: An overview. *SAR QSAR Environ Res.* **1994**, *2*, 89–104. https://doi.org/10.1080/10629369408028842.
2. Bearden, A. P.; Schultz, T. W. Structure-Activity Relationships for Pimephales and Tetrahymena: A Mechanism of Action Approach. *Environ. Toxicol. Chem.* **1997**, *16*, 1311–1317. https://doi.org/10.1002/etc.5620160629.
3. Liu, H. X.; Hu, R. J.; Zhang, R. S.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; & Fan, B. T. The prediction of human oral absorption for diffusion rate-limited drugs based on heuristic method and support vector machine. *J. Comput. Aided Mol. Des.* **2005**, *19*, 33–46. https://doi.org/10.1007/s10822-005-0095-8.
4. Si, H.; Yuan, S.; Zhang, K.; Fu, A.; Duan, Y.-B.; ,Hu, Z. Quantitative structure–activity relationship study on $EC_{50}$ of anti-HIV drugs. *Chemometr. Intell. Lab. Syst.* **2008**, *90*, 15–24. https://doi.org/10.1016/j.chemolab.2007.06.011.

5. Xu, H. Y.; Zhang, J. Y.; Zou, J. W.; Chen, X. S. QSPR models for the physicochemical properties of halogenated methyl-phenyl ethers. *J. Mol. Graph. Model.* **2008**, *26*, 1076–1081. https://doi.org/10.1016/j.jmgm.2007.09.004.
6. Wania, F.; Mackay, D. The evolution of mass balance models of persistent organic pollutant fate in the environment. *Environ. Pollut.* **1999**, *100*, 223–240. https://doi.org/10.1016/s0269-7491(99)00093-7.
7. Fisher, S. W.; Lydy, M. J.; Barger, J.; Landrum, P. F. Quantitative structure–activity relationships for predicting sediment-sorbed chlorobenzene bioavailability to Chironomus decorus larvae. *Environ. Toxicol. Chem.* **1993**, *12*, 1307–1318. https://doi.org/10.1002/etc.5620120721.
8. Van Leeuwen, C. J.; Vanderzandt, P. T.; Aldenberg, T.; Verhaar, H. J. M.; Hermens, J. L. M. The application of QSARs, extrapolation and equilibrium partitioning in aquatic effects assessment for narcotic pollutants. *Sci. Total Environ.* **1991**, *109*, 681–690. https://doi.org/10.1016/0048-9697(91)90222-z.
9. Niculescu, S. P.; Kaiser, K. L. E.; Schüürmann, G. Influence of Data Preprocessing and Kernel Selection on Probabilistic Neural Network Modeling of the Acute Toxicity of Chemicals to the Fathead Minnow and Vibrio fischeri Bacteria. *Water Qual. Res. J. Can.* **1998**, *33*, 153–166. https://doi.org/10.2166/wqrj.1998.009.
10. Wong, S.L. Algal assay evaluation of trace contaminants in surface water using the nonionic surfactant Triton X-100. *Aquat. Toxicol.* **1985**, *6*, 115–131. https://doi.org/10.1016/0166-445X(85)90011-6.
11. Chessells, M.; Hawker, D. W.; Connell, D. W. Critical evaluation of the measurement of the 1-octanol/water partition coefficient of hydrophobic compounds. *Chemosphere* **1991**, *22*, 1175–1190. https://doi.org/10.1016/0045-6535(91)90213-W.
12. Patil, G.S. Prediction of aqueous solubility and octanol–water partition coefficient for pesticides based on their molecular structure. *J. Hazard Mater.* **1994**, *36*, 34–43. https://doi.org/10.1016/0304-3894(93)E0049-8.
13. HyperchemTM. Release 6.02 for Windows. Molecular Modeling System. 2000. Available online: http://www.hyper.com/ (accessed on).
14. RTodeschini, V. Consonni, A.; Mauri, M. Pavan, Dragon Software, version 5.4, Copyright TALETE srl, 2005.
15. RTodeschini, D.; Ballabio, V.; Consonni, A.; Mauri, M. *Paven*, *MobyDigs*, version 1.1, Copyright TALETE srl, 2009. Available online: http://www.talete.mi.it/ (accessed on).