A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, Domain of Application and Prediction
Mabrouk Hamadache, Othmane Benkortbi, Salah Hanini, Abdeltif Amrane, Latifa Khaouane, Cherif Si Moussa

To cite this version:

HAL Id: hal-01220889
https://hal-univ-rennes1.archives-ouvertes.fr/hal-01220889
Submitted on 15 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, Domain of Application and Prediction

Mabrouk Hamadache¹, Othmane Benkortbi¹, Salah Hanini¹, Abdeltif Amrane², Latifa Khaouane¹, Cherif Si Moussa¹

¹ Laboratoire des Biomatiériaux et Phénomènes de Transport (LBMPT), Université de Médéa, Quartier Ain D'heb, 26000, MEDEA, Algérie
² Ecole Nationale Supérieure de Chimie de Rennes, Université de Rennes 1, CNRS, UMR 6226, 11 allée de Beaulieu, CS 50837, 35708 Rennes Cedex 7, France

L. KHAOUANE : latifa_khaouane@yahoo.fr ; O. BENKORTBI : benkortbi_oth@yahoo.fr ;
C. SI MOUSSA : simoussa_cherif@yahoo.fr ; S. HANINI : s_hanini2002@yahoo.fr ;
A. AMRANE : abdeltif.amrane@univ-rennes1.fr

*Corresponding author: Mabrouk HAMADACHE;
mhamdeche@yahoo.fr
Tel: +213 07 78 12 37 50; Fax: +213 25 58 12 53
ABSTRACT

Quantitative Structure Activity Relationship (QSAR) models are expected to play an important role in the risk assessment of chemicals on humans and the environment. In this study, we developed a validated QSAR model to predict acute oral toxicity of 329 pesticides to rats because a few QSAR models have been devoted to predict the Lethal Dose 50 (LD₅₀) of pesticides on rats. This QSAR model is based on 17 molecular descriptors, and is robust, externally predictive and characterized by a good applicability domain. The best results were obtained with a 17/9/1 Artificial Neural Network model trained with the Quasi Newton back propagation (BFGS) algorithm. The prediction accuracy for the external validation set was estimated by the Q²_{ext} and the Root Mean Square error (RMS) which are equal to 0.948 and 0.201, respectively. 98.6% of external validation set is correctly predicted and the present model proved to be superior to models previously published. Accordingly, the model developed in this study provides excellent predictions and can be used to predict the acute oral toxicity of pesticides, particularly for those that have not been tested as well as new pesticides.

Keywords

Acute toxicity, Pesticides, QSAR, Prediction, External validation

Abbreviations: QSAR, quantitative structure-activity relationship; LD₅₀, lethal dose 50; ANN, artificial neural networks; BFGS, Quasi-Newton back propagation algorithm; RMS, root mean square error; REACH, registration, evaluation, authorization and restriction of chemicals; OECD, organization for economic cooperation and development; LOO, leave-one-out; CV, cross-validation; AD, applicability domain; VIF, variation inflation factors; MLP, multi-layer perceptron.

1. Introduction
Pesticides are widely used in agriculture for plant protection and for increasing production yields and quality of agricultural products but also in domestic applications. They are also used to slow the spread of insects, to maintain lawns, recreational areas and highways. Pesticides have also contributed to the control of many human diseases transmitted by insects. The most common pesticides are herbicides, insecticides and fungicides. However, despite these advantages, pesticides have a major drawback such as toxicity [1]. Due to the excessive use of these products, they are found as well as residue in the environment (water, soil, air) than in terrestrial and aquatic food chains [2, 3]. In addition, they also pose a threat to the environment, humans, animals and other organisms [4, 5]. Many studies made internationally highlight the environmental pollution by pesticides. The consequences of this pollution are the widespread presence of residues in air, water, soil and foodstuffs [6-13].

Long-term exposure to pesticides can cause harm to human life and can disrupt the functioning of various organs in the body. This significant relationship between exposure to pesticides and some chronic diseases has been the subject of several scientific publications. Exposure to these persistent pesticides has been associated with health effects including cancer, headache, skin and eye irritation, immune system problems, stomach, kidney, Parkinson and Alzheimer’s disease, reproductive difficulties, birth defects, diabetes, cataracts and anemia [14-17].

As seen, humans and the environment are exposed to thousands of pesticides. This pollution caused by pesticides has become an important issue affecting the survival and development of human being. It is evident that risk assessment for pesticides can provide a precaution against the corresponding pollution. One of the procedures currently used for human and environmental risk assessment is the determination of the acute toxicity of pesticides [18]. Unfortunately, experimental determination of the toxicity takes time, requires a high expense and poses an ethical problem (demands to reduce or abolish the use of animals). Also, there is a very large body of research going on in many countries with the aim of replacing in vivo tests by in silico prediction methods according to the European Directive on the Protection of Laboratory Animals [19] and the Registration, Evaluation, Authorization and restriction of Chemicals (REACH) regulation [20]. Despite being significantly cheaper than in vivo study, in vitro tests are still costly compared with in silico methods [21]. The use of in silico predictive methods, based on computer tools, offers a rapid, cost-effective and ethical alternative to testing toxicity of chemical substances in animals [22]. These methods include the Quantitative Structure–Activity Relationship (QSARs) models. To establish a QSAR model, three elements
are necessary. The first relates to the biological activity (eg toxicity) measured for a set of molecules. The second concerns the descriptors. Finally, the third must be a statistical learning method that is used to connect the first two elements.

The acute toxicity still remains the object of interest in QSAR model building. To date, a large number of QSARs models for predicting the acute toxicity of chemical substances have been developed [23, 24]. Unfortunately, few studies have been devoted to the acute toxicity of pesticides on rats. For example, Enslein et al. [25, 26] developed regression analysis models using two large data sets (425 and 1851 various chemicals, respectively). The R² value for the test set is 0.33, which means that these models are characterized by low power external prediction. A very marked improvement in R² coefficient was obtained following the QSAR models developed with 44, 54, 67, 30 and 62 pesticides by Zakaria et al. [27], Eldred and Jurs [28], Zahouily [29], Guo et al. [30] and Garcia et al. [31] respectively. Recent studies devoted to pesticides [32, 33] have proposed QSAR models with values of 0.93 (27 herbicides) and 0.96 (62 herbicides) for the R² coefficient. The conclusion which can be draw from these studies is that most QSAR models developed are distinguished by two major shortcomings: lack of validation test on the one hand, and a limited field of application because these studies included a relatively small number of pesticides on the other hand.

Since the prediction of potential risks to human health is based on the assumption that test results seen in high-dose animal tests are predictive of effects that will occur in human populations exposed to much lower levels [34], our main goal in this work is to establish a robust QSAR model to predict acute toxicity (log [1/LD₅₀]) of pesticides on rats. The database used consists of 329 pesticides. The QSAR model established by using artificial neural networks and molecular descriptors satisfies the guidelines required by the Organisation for Economic Cooperation and Development (OECD), namely: (1) a defined endpoint; (2) an unambiguous algorithm; (3) a defined domain of applicability; (4) appropriate measures of goodness of fit, robustness, predictability; (5) a mechanistic interpretation, if possible.

2. Materials and method

2.1 Data set

It is well known that high-quality experimental data are essential for the development of high quality QSAR models [35]. If they are unreliable, the model will be unreliable. The rat lethal dose 50 (LD₅₀ - rat, male via oral exposure) values were retrieved from Pesticide Properties Database [36]. The LD₅₀ correspond to the concentration (mg/kg) of pesticide that
lead to the death of 50% of rat. All values of oral acute toxicity were first converted into mmol/kg and then translated to log \[1/ (\text{mmol/kg})\].

The initial database that included 907 pesticides was rigorously reviewed and “cleaned” by removing pesticides whose LD\(_{50}\) was not experimentally determined or whose LD\(_{50}\) was not determined in the same experimental conditions. A total of 329 pesticides with experimental data were selected to form the final database (Table 1). The basis of 329 pesticides was divided into 2 lots. The first with 258 pesticides was dedicated to develop the QSAR model. The second which included 71 pesticides that had not been used for the development of the QSAR model, was left for the external validation.

2.2 Molecular descriptors

One important step in obtaining a QSAR model is the numerical representation of the structural features of molecules, which were named molecular descriptors. Nowadays, there are more than 4000 of molecular descriptors which can be used to solve different problems in Chemistry, Biology and related sciences [1]. In the specific case of this study, for each molecule, 1664 molecular descriptors were calculated, which belong to many classes. All descriptors were obtained through the online program E-Dragon 1.0 (http://www.vcclab.org/lab/edragon).

To avoid the phenomenon of overfitting, the number of descriptors must be reduced. Several methods to simplify a database are used. The method used to select the most significant descriptors was described previously [32]. In the first step, invariant descriptors, namely those with absent values (represented by the code “999”), were manually removed. Next, any descriptor that had identical values for 75% of the samples and any descriptors with a relative standard deviation < 0.05 were removed. Finally, half of the descriptors showing an absolute value of the Pearson correlation coefficient > 0.95 were also removed. The number of descriptors obtained after the selection was 95. For relevant descriptors selection, stepwise regression was then used [37]. Twenty nine descriptors were selected.

2.3 Model development

In this work, all calculations were run on a Sony personal computer with a Core (TM) i3 and windows XP as operating system. The Artificial Neural Networks (ANN) which has extensive applicability in solving non-linear systems was employed to build the QSAR model between the molecular relevant descriptors and the toxicity of pesticides. A three-layer feed-
forward neural network utilizing back-propagation algorithm was employed. The typical back-propagation network consists of an input layer, an output layer and at least one hidden layer. Each layer contains neurons and each neuron is a simple micro-processing unit which receives and combines signals from many neurons.

The use of a neuronal regression goes through the choice of the input parameters but also by optimizing the architecture of the neural network itself. The optimization of both the distribution of the database, the number of hidden layers, the number of neurons per hidden layer, the transfer functions as well as algorithms was carried after extensive testing. The design of the neural model is to evaluate the components of the network according to the desired performance modeling. Model performance is evaluated in terms of root mean square error (RMS) [38] and was calculated with the following equation:

\[ RMS = \sqrt{\frac{\sum_{i=1}^{n} (y_i^{exp} - y_i^{pred})^2}{n}} \]  

where \( n \) is the number of compounds in the dataset, and \( y_i^{pred}, y_i^{exp} \) are the predicted and the experimental values, respectively.

2.4 Model validation

For the validation of the predictive power of a QSAR model, two basic principles (internal validation and external validation) are available. The quality is always judged by the statistical parameters, for instance, the squared R (R\(^2\)) and root mean square error (RMS). These parameters mainly reflect the goodness of fit of the models. However, recent studies [38] have indicated that the internal validation is considered to be necessary for model validation. In the present study, we took the leave-one-out (LOO) cross-validation (CV) for the internal validation to evaluate the internal predictive ability of the developed model, and its result was defined as \( Q^2_{\text{LOO}} \), which could be calculated according to the following equation [38]:

\[ Q^2_{\text{LOO}} = 1 - \frac{\sum_{i=1}^{\text{training}} (y_i^{exp} - y_i^{pred})^2}{\sum_{i=1}^{\text{training}} (y_i^{exp} - \bar{y})^2} \]  

where \( y_i^{exp}, y_i^{pred} \) and \( \bar{y} \) are the experimental, predicted, and average log (1/LD50) values of the samples for the training set, respectively. A value of \( Q^2_{\text{LOO}} > 0.5 \) is considered satisfactory, and a \( Q^2_{\text{LOO}} \) value > 0.9 is excellent [39].

Furthermore, the external validation is a significant and necessary validation method used to determine both the generalizability and the true predictive ability of the QSAR models.
for new chemicals, by splitting the available dataset into a training set and an external prediction set. As mentioned above, the whole dataset in this work has been randomly divided into a training set with 258 compounds for model development, and a prediction set with 71 compounds for model external validation. The external predictive ability of the developed models on the external prediction set was evaluated by $Q^2_{ext}$, which could be calculated as follows [38]:

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{\text{prediction}} (y_i^{exp} - y_i^{pred})^2}{\sum_{i=1}^{\text{prediction}} (y_i^{exp} - \bar{y}_{tr})^2}$$  

(3)

where $y_i^{exp}, y_i^{pred}$ are the experimental and predicted $\log(1/LD_{50})$ values of the samples for the prediction set, and $\bar{y}_{tr}$ is the mean experimental $\log(1/LD_{50})$ values of the samples for the training set.

2.5 Applicability domain

Even the most comprehensive and validated models cannot predict reliably properties for all existing compounds. The QSAR model is not intended to be used outside its domain of applicability, that is to say, outside of the chemical space covered by the training set. Also, the applicability domain (AD) of models must be defined and the predictions of the molecules in this area can be considered admissible. The determination of AD is therefore of great importance [40].

The AD is a theoretical region in the space defined by the descriptors of the model and the modeled response, for which a given QSAR should make reliable predictions. This region is defined by the nature of the compounds in the training set, and can be characterized in various ways. In our work, the AD was verified by the leverage approach. The leverage $h_i$ is defined as follows [41]:

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$  

(4)

Where $x_i$ is the descriptor value of the ith object, and $\bar{x}$ is the average value of the descriptor in the training set, and $n$ is the number of substances in the training set. The warning leverage $h^*$ is, generally, fixed at $3(p + 1)/n$, where $n$ is the total number of samples in the training set and $p$ is the number of descriptors involved in the correlation.

The applicability domain (AD) of QSAR model is defined from the Williams plot. The plot of leverage values versus standardized residuals (Williams plot) was used to give a
graphical detection of both the response outliers (Y outliers) and the structurally influential compounds (X outliers). In this plot, the two horizontal lines indicate the limit of normal values for Y outliers (i.e. samples with standardized residuals greater than 3.0 standard deviation units, ±3.0s); the vertical straight lines indicate the limits of normal values for X outliers (i.e. samples with leverage values greater than the threshold value, h > h*). For a sample in the external test set whose leverage value is greater than h*, its prediction is considered unreliable, because the prediction is the result of a substantial extrapolation of the model. Conversely, when the leverage value of a compound is lower than the critical value, the probability of accordance between predicted and experimental values is as high as that for the compounds in the training set [42].

3. Results and discussion

3.1 Selection of relevant descriptors

To select the most important descriptors and the optimal number, the influences of the number of descriptors on the correlation coefficients \([R^2\text{ and adjusted }R^2(\text{adj})]\) and the RMSE were investigated for 1–29 descriptors. \(R^2\) and \(R^2_{\text{adj}}\) increased with increasing number of descriptors. However, the values of RMSE decreased with increasing number of descriptors. Models with 18–29 descriptors did not significantly improve the statistics of the model. For these reasons, the number of descriptors used to develop the model was 17. Let us note that \(n / k\) is greater than 5 [43] where \(n\) (258) and \(k\) (17) are respectively the number of compounds and the number of descriptors used in the QSAR model.

Multi-collinearity between the 17 descriptors was detected by calculating their variation inflation factors (VIF). If VIF falls into the range of 1–5, the related model is acceptable. All the descriptors have VIF values < 2.873, indicating that the obtained model has statistical significance, and the descriptors were found to be reasonably orthogonal. Order to study the correlation between the selected descriptors, the correlation matrix has been established using the XLSTAT software. The results show that these descriptors are not correlated owing to the fact that the greatest value of the correlation coefficient is 0.512. The list of descriptors used in the development of QSAR model is given in Table 2.

3.2 QSAR modeling

The main objective of this phase of the study is to find the optimal architecture of the neural network to predict the acute oral toxicity of pesticides on rats. A typical multilayer
perceptron (MLP) three-layered network with an input layer, a hidden layer and an output layer is adopted in this work. Increasing the number of the hidden layers decreases the learning accuracy. Theoretical works have shown that a single hidden layer is sufficient for the ANN to approximate to any complex nonlinear function and many experimental results seem to confirm that one hidden layer may be enough for most forecasting problems [44]. The use of a neuronal regression requires the selection of input parameters, but also the optimization of the neural network architecture. Before training the network, the database distribution, the activation functions (for hidden neurons and output neurons), the number of neurons in the hidden layer and the learning algorithms were optimized after many trials. The optimal model performance is evaluated in terms of root mean square error (RMS) [45, 46]. The results of this study and the ANN network optimal adopted are given in Table 3.

The selected parameters (Table 3) were used to develop nonlinear model. The seventeen relevant descriptors were used as inputs to the network. Before training the network, the number of nodes in the hidden layer was optimized, because it is an important parameter influencing the performances of the ANN. Thus, a 17-9-1 network architecture was obtained after trial and error procedure. The main performance parameters of MLP-ANN model are shown in Table 4. The predictive results from the MLP-ANN model for the entire dataset (329 compounds) are obtained and presented in Table 1. Figure 1 and 2 shows the regression line of the model equation, i.e. predicted vs experimental results for the training and validation set highlighted by different symbols.

Fig.1 and Fig.2 indicates that there is a significant correlation between experimental values and predicted values of log (1/ LD50). As can be seen from Table 4, the non-linear MLP-ANN model give good results with higher correlation coefficients (R² and R² extr.), lower RMS, as well as better robustness (Q²) in both training set and validation set, which indicated that the MLP-ANN not only performed well in model development, but also had excellent prediction. This fact suggested a non-linear correlation between the acute toxicity and the relevant descriptors. In addition, the residual of the predicted values of the toxicity data against the experimental values for the present model is shown in Fig. 3. As most of the calculated residuals are distributed on two sides of the zero line, a conclusion may be drawn that there is no systematic error in the development of the present model.

To see the importance of each descriptor for the prediction of LD50 oral toxicity of pesticides towards rats, the relative contributions [47] of the seventeen descriptors to the MLP-ANN model were determined and are plotted in Fig.4. The contribution of the descriptors
decreased in the order: HATS0m (12.81%) > E1u (7.98%) > MATS2p (7.74%) > HATSe (7.63%) > Mor15m (7.14%) > RDF030e (6.48%) > H6m (6.27%) > Mor23u (6.12%) > Du (5.88%) > nS (5.58%) > PJI3 (5.10%) > N-072 (4.68%) > RDF020e (4.29%) > MATS1m (3.47%) ≈ nArX (3.45%) > Mor26u (2.93%) > H-046 (2.45%). The most significant descriptor in the model was therefore HATS0m. It should be noted that for the majority of the descriptors, the difference between two descriptors contribution was not significant, indicating that all selected descriptors were needed in the development of QSAR predictive model.

Generally, QSAR models are functions of a molecule’s structure, electronic properties and hydrophobicity [48]. In the present model, HATS0m, E1u, Mor15m, H6m, Mor23u, Du, nS, PJI3, N-072, MATS1m, nArX, Mor26u and H-046 involve the structure while MATS2p, HATSe, RDF030e and RDF020e represent the electronic properties.

Descriptors used in our model have been used in previous QSAR models in the literature. Hamadache et al. [32] have used MATS2p, HATSe, HATS0m, nS, E1u and N-072 in their MLR and ANN models to predict rat oral acute toxicity of 62 herbicides. In a study by Habibi-Yangjeh and Danandeh-Jenagharad [49], the MATS1m, H-046, Mor23u and PJI3 descriptors were used for global prediction of the toxicity of 250 phenols to Tetrahymena pyriformis in a linear and nonlinear model. In a QSAR model of acute toxicity LD₅₀ for rats caused by aromatic compounds, Bakhtiyor et al. [50] found that the descriptor MATS2p significantly contributes to the toxicity of these compounds. In a study on the penetration of the blood–brain barrier, the human intestinal absorption and the hydrophobicity, Soto et al. [51] proposed linear and nonlinear QSAR/QSPR models that include the descriptor MATS2p. A QSA(P)R model with high internal and external statistical quality for predicting toxicity was developed by Borges [52] with MATS2p for a set of 28 alkyl (1-phenylsulphonyl)-cycloalkane-carboxilates. A QSAR model on rat oral LD₅₀ data of 58 per- and polyfluorinated chemicals developed by Bhattarai and Gramatica [53] employed E1u; the authors concluded that E1u is one of the most important descriptors.

Moreover, some authors [48, 54-57] found that among the descriptors that affect the toxicity of the compounds studied, a substantial number belong to the categories of WHIM descriptors, GETAWAY descriptors, 2D autocorrelations, and Atom-centered fragments. In our study, a large number of descriptors involved in the present model also belong to this category. It is obvious that the descriptors in this category have major significance in the toxicity of pesticides.
3.3 Applicability domain

The applicability domain of the model was analysed using a Williams plot (Fig. 5), where the vertical line is the critical leverage value ($h^*$), and the horizontal lines are $3\sigma$ the cutoff value for Y space. As seen in Fig. 5, one can observe that none of the pesticides compounds in the training set and validation set have a leverage higher than the warning $h^*$ value of 0.16. In the Williams plot, three pesticides can be considered as response outlier (in the Y-response space). In the training set, one pesticide (Pyrazophos: 225) was overestimated, while another pesticide was underestimated (Oxycarboxine: 201). However, in the region of underestimated pesticides, Pyrazophos (329) was from the validation set. These three response outlier (in the Y-response space) could be associated with errors in the experimental values.

It should be noted that 98.6% of the domain was covered by the model when it was applied to predict the acute oral toxicity of the 71 pesticides in the validation set. Thus, these results show that MLP-ANN model complies with the third principle of the OECD. Accordingly, the model developed in this study provides excellent predictions for 329 pesticides. It can be used to predict the acute oral toxicity of pesticides, particularly for those that have not been tested as well as new pesticides.

3.4 Comparison with different models

As indicated in the introduction, there are a limited number of QSAR models available in the literature for predicting the oral acute toxicity of pesticides to rats. The evaluation of their advantages and disadvantages is quite difficult, because each published study used different data sets and a different modeling approach (chemical descriptors, algorithms, etc.). However, it would be worthwhile to evaluate the performance of our model (present work) in light of the few QSAR models published in the literature over the last few years. Our main aim is to compare the predictive power of each model, which gives an estimation of the fitting of the model and its robustness.

It should be noted that the most of these QSAR models were obtained using small databases [33] and generally with structurally similar chemicals such as amide herbicides [27, 58], benzimidazoles herbicides [59] or phenylurea herbicides [60]. Also, the number of statistical parameters used for validation of this QSAR models is limited, especially in old publications. Devillers [61] developed a QSAR model for acute oral toxicity in rodents (rats). He used artificial neural networks (ANN) to predict the LD$_{50}$ values of organophosphate
pesticide. The 51 chemicals of the training set and the nine compounds of the external testing set were described by a set of descriptors. The acute toxicities (1/log LD$_{50}$) were converted to mmol/kg and a series of 8 descriptors has been used. The best results were obtained with an 8/4/1 ANN model. The root mean square error (RMS) values for the training set and the external testing set equaled 0.29 and 0.26, respectively. This study demonstrated the usefulness of descriptors such as lipophilicity and molar refractivity.

Structure-toxicity relationships were studied for a set of 47 insecticides with three-layer perceptron and use of a backpropagation algorithm [29]. A model with three descriptors showed good statistics in the artificial neural network model with a configuration of 3/5/1 ($r = 0.966$, RMS = 0.200 and $Q^2 = 0.647$). The statistics for the prediction on toxicity [log LD$_{50}$, oral, rat] in the test set of 20 organophosphorus insecticides derivatives was $r = 0.748$, RMS = 0.576. The model descriptors indicate the importance of molar refraction toward toxicity of organophosphorus insecticides derivatives used in this study. Otherwise, different topological descriptors were used by Garcia-Domenech et al. [31] in the prediction of the oral acute toxicity (LD$_{50}$) of 62 organophosphorus pesticides on rats. The LD$_{50}$ values were expressed in mmol/kg with a logarithmic transformation before use. A model with eight variables ($r = 0.906$, $Q^2 = 0.701$) was generated. Zhu et al. [62] have developed a number of QSAR models for acute oral toxicity in rats using large datasets (7385 compounds). Several sets of descriptors and different modeling methods were used. It should be noted an improvement of the prediction compared to other works. However, the complexity of the modeling approach, while being interesting and promising, renders these models little useful in practice.

The statistical parameters of the results obtained from the present study and studies published in the literature are shown in Table 5. It is possible to observe that all of those models could give high prediction ability (correlation coefficient $R^2$, $Q^2$). However, our model exceeds the previously published models in all statistical indices available for comparison. Indeed, it gives the higher correlation coefficient and the lower RSM error if compared to the other models. It can be seen that the database for this study (training set and validation set) was wider than that of previous models with the exception of the base used by Zhu et al. [62]. According to these results, the present model can be promisingly used for predicting the toxicity of new chemicals, thus contributing to the risk assessment, saving substantial amounts of money and time.

4. Conclusion
The aim of the present work was to develop a QSAR study and to predict the oral acute toxicity of pesticides to rats. This study involved 258 pesticides with an additional external set of 71 pesticides modelled for their oral acute toxicity on rat based on the artificial neural network (multi-layer perceptron: MLP-ANN) with descriptors calculated by Dragon software and selected by a stepwise MLR method. The seventeen selected descriptors showed that the electronic properties and the structure of the molecule play a main role in the toxicity activity of the pesticides. The built MLP-ANN model was assessed comprehensively (internal and external validations). It showed good values of $R^2 = 0.963$ and $Q^2_{LOO} = 0.962$ for the training set and high predictive $R^2_{ext}$ and $Q^2_{ext}$ values (0.950 and 0.948) for the validation set. All the validations indicate that the built QSAR model was robust and satisfactory. Based on the comparison with models previously published, the proposed QSAR model achieved good results and provided 98.6% predictions that belong to the applicability domain. In conclusion, the model developed in this study meets all of the OECD principles for QSAR validation and can be used to predict the acute oral toxicity of pesticides, particularly for those that have not been tested as well as new pesticides and thus help reduce the number of animals used for experimental purposes.

References


PPDB (Pesticide Properties DataBase), http://sitem.herts.ac.uk/aeru/footprint/ (accessed 14/05/2014).


Table 1.

Observed (experimental) log (1/LD$_{50}$), predicted log (1/LD$_{50}$) and leverage of pesticide compounds.

<table>
<thead>
<tr>
<th>No.</th>
<th>Compound</th>
<th>Type</th>
<th>log [1/LD$_{50}$] (mmol/kg)$^{-1}$</th>
<th>Leverage ($h_i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Observed</td>
<td>Predicted</td>
</tr>
<tr>
<td>1</td>
<td>1,2-Dichloropropane</td>
<td>Insecticide</td>
<td>−1.24</td>
<td>−1.26</td>
</tr>
<tr>
<td>2</td>
<td>2,4,5-Trichlorophenol</td>
<td>Herbicide</td>
<td>−0.62</td>
<td>−0.30</td>
</tr>
<tr>
<td>3</td>
<td>2,4-DB</td>
<td>Herbicide</td>
<td>−0.55</td>
<td>−0.53</td>
</tr>
<tr>
<td>4</td>
<td>2,4-Dimethylphenol</td>
<td>Fongicide</td>
<td>−0.30</td>
<td>−0.28</td>
</tr>
<tr>
<td>5</td>
<td>2-Amino butane</td>
<td>Fongicide</td>
<td>−0.68</td>
<td>−0.59</td>
</tr>
<tr>
<td>6</td>
<td>Acephate</td>
<td>Insecticide</td>
<td>−0.71</td>
<td>−0.73</td>
</tr>
<tr>
<td>7</td>
<td>Acetamiprid</td>
<td>Insecticide</td>
<td>0.02</td>
<td>−0.08</td>
</tr>
<tr>
<td>8</td>
<td>Acetochlor</td>
<td>Herbicide</td>
<td>−0.85</td>
<td>−0.84</td>
</tr>
<tr>
<td>9</td>
<td>4-CPA</td>
<td>Herbicide</td>
<td>−0.66</td>
<td>−0.63</td>
</tr>
<tr>
<td>10</td>
<td>Acrolein</td>
<td>Herbicide</td>
<td>0.29</td>
<td>0.32</td>
</tr>
<tr>
<td>11</td>
<td>Alachlor</td>
<td>Herbicide</td>
<td>−0.54</td>
<td>−0.71</td>
</tr>
<tr>
<td>12</td>
<td>Alanycarb</td>
<td>Insecticide</td>
<td>0.08</td>
<td>0.12</td>
</tr>
<tr>
<td>13</td>
<td>Aldicarb</td>
<td>Insecticide</td>
<td>2.31</td>
<td>2.41</td>
</tr>
<tr>
<td>14</td>
<td>Aldrin</td>
<td>Insecticide</td>
<td>0.97</td>
<td>0.94</td>
</tr>
<tr>
<td>15</td>
<td>Allyxycarb</td>
<td>Insecticide</td>
<td>0.49</td>
<td>0.27</td>
</tr>
<tr>
<td>16</td>
<td>Alpha-endosulfan</td>
<td>Insecticide</td>
<td>1.03</td>
<td>0.90</td>
</tr>
<tr>
<td>17</td>
<td>Amicarbazone</td>
<td>Herbicide</td>
<td>−0.62</td>
<td>−0.75</td>
</tr>
<tr>
<td>18</td>
<td>Amidithion</td>
<td>Insecticide</td>
<td>−0.34</td>
<td>−0.08</td>
</tr>
<tr>
<td>19</td>
<td>Aminocarb</td>
<td>Insecticide</td>
<td>0.84</td>
<td>0.82</td>
</tr>
<tr>
<td>20</td>
<td>Amiprofos-methyl</td>
<td>Herbicide</td>
<td>−0.01</td>
<td>−0.04</td>
</tr>
<tr>
<td>21</td>
<td>Amitraz</td>
<td>Insecticide</td>
<td>−0.44</td>
<td>−0.37</td>
</tr>
<tr>
<td>22</td>
<td>Ancymidol</td>
<td>Herbicide</td>
<td>−0.83</td>
<td>−0.77</td>
</tr>
<tr>
<td>23</td>
<td>Anilazine</td>
<td>Fongicide</td>
<td>−1.22</td>
<td>−1.41</td>
</tr>
<tr>
<td>24</td>
<td>Anilofos</td>
<td>Herbicide</td>
<td>−0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>25</td>
<td>Asomate</td>
<td>Fongicide</td>
<td>0.11</td>
<td>0.21</td>
</tr>
<tr>
<td>26</td>
<td>Azaconazole</td>
<td>Fongicide</td>
<td>−0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>27</td>
<td>Azametiphos</td>
<td>Insecticide</td>
<td>−0.56</td>
<td>−0.59</td>
</tr>
<tr>
<td>28</td>
<td>Azinphos-methyl</td>
<td>Insecticide</td>
<td>1.55</td>
<td>1.48</td>
</tr>
<tr>
<td>29</td>
<td>Benalaxil</td>
<td>Fongicide</td>
<td>−0.32</td>
<td>−0.65</td>
</tr>
<tr>
<td>30</td>
<td>Bendiocarb</td>
<td>Insecticide</td>
<td>0.82</td>
<td>0.83</td>
</tr>
<tr>
<td>31</td>
<td>Benfuracarb</td>
<td>Insecticide</td>
<td>0.30</td>
<td>0.00</td>
</tr>
<tr>
<td>32</td>
<td>Benquinox</td>
<td>Fongicide</td>
<td>0.38</td>
<td>0.36</td>
</tr>
<tr>
<td>33</td>
<td>Bentazone</td>
<td>Herbicide</td>
<td>−0.32</td>
<td>−0.27</td>
</tr>
<tr>
<td>No.</td>
<td>Compound</td>
<td>Type</td>
<td>log $[1/LD_{50}]$ (mmol/kg)$^{-1}$</td>
<td>Leverage ($h_i$)</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>---------------</td>
<td>-----------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Observed</td>
<td>Predicted</td>
</tr>
<tr>
<td>34</td>
<td>Benzthiazuron</td>
<td>Herbicide</td>
<td>−0.79</td>
<td>−0.48</td>
</tr>
<tr>
<td>35</td>
<td>Binapacryl</td>
<td>Fongicide</td>
<td>0.75</td>
<td>0.45</td>
</tr>
<tr>
<td>36</td>
<td>Brodifacoum</td>
<td>Rodonticide</td>
<td>3.12</td>
<td>3.16</td>
</tr>
<tr>
<td>37</td>
<td>Bromacil</td>
<td>Herbicide</td>
<td>−0.70</td>
<td>−0.60</td>
</tr>
<tr>
<td>38</td>
<td>Bromocyclen</td>
<td>Insecticide</td>
<td>−1.50</td>
<td>−1.43</td>
</tr>
<tr>
<td>39</td>
<td>Bromophos</td>
<td>Insecticide</td>
<td>−0.64</td>
<td>−0.73</td>
</tr>
<tr>
<td>40</td>
<td>Bromophos-ethyl</td>
<td>Insecticide</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td>41</td>
<td>Bromoxynil</td>
<td>Herbicide</td>
<td>0.53</td>
<td>0.86</td>
</tr>
<tr>
<td>42</td>
<td>Bromoxynil heptanoate</td>
<td>Herbicide</td>
<td>0.13</td>
<td>0.09</td>
</tr>
<tr>
<td>43</td>
<td>Bromoxynil octanoate</td>
<td>Herbicide</td>
<td>0.23</td>
<td>0.17</td>
</tr>
<tr>
<td>44</td>
<td>Bromuconazole</td>
<td>Fongicide</td>
<td>0.06</td>
<td>0.25</td>
</tr>
<tr>
<td>45</td>
<td>Bronopol</td>
<td>Fongicide</td>
<td>−0.10</td>
<td>−0.04</td>
</tr>
<tr>
<td>46</td>
<td>Bupirimate</td>
<td>Fongicide</td>
<td>−1.10</td>
<td>−0.96</td>
</tr>
<tr>
<td>47</td>
<td>Butachlor</td>
<td>Herbicide</td>
<td>−0.81</td>
<td>−0.97</td>
</tr>
<tr>
<td>48</td>
<td>Butamifos</td>
<td>Herbicide</td>
<td>−0.28</td>
<td>−0.07</td>
</tr>
<tr>
<td>49</td>
<td>Butylate</td>
<td>Herbicide</td>
<td>−1.21</td>
<td>−1.29</td>
</tr>
<tr>
<td>50</td>
<td>Butocarboxim</td>
<td>Insecticide</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>51</td>
<td>Butonate</td>
<td>Insecticide</td>
<td>−0.53</td>
<td>−0.48</td>
</tr>
<tr>
<td>52</td>
<td>Butoxycarboxim</td>
<td>Insecticide</td>
<td>−0.31</td>
<td>−0.10</td>
</tr>
<tr>
<td>53</td>
<td>Butralin</td>
<td>Herbicide</td>
<td>−0.55</td>
<td>−0.66</td>
</tr>
<tr>
<td>54</td>
<td>Cadusafos</td>
<td>Insecticide</td>
<td>0.95</td>
<td>1.05</td>
</tr>
<tr>
<td>55</td>
<td>Camphechlor</td>
<td>Insecticide</td>
<td>0.92</td>
<td>0.41</td>
</tr>
<tr>
<td>56</td>
<td>Carbanolate</td>
<td>Insecticide</td>
<td>0.85</td>
<td>0.82</td>
</tr>
<tr>
<td>57</td>
<td>Carbaryl</td>
<td>Insecticide</td>
<td>−0.48</td>
<td>−0.39</td>
</tr>
<tr>
<td>58</td>
<td>Carbetamide</td>
<td>Herbicide</td>
<td>−0.86</td>
<td>−1.01</td>
</tr>
<tr>
<td>59</td>
<td>Carbofuran</td>
<td>Insecticide</td>
<td>1.50</td>
<td>1.38</td>
</tr>
<tr>
<td>60</td>
<td>Carbophenothion</td>
<td>Insecticide</td>
<td>1.54</td>
<td>1.44</td>
</tr>
<tr>
<td>61</td>
<td>Carbosulfan</td>
<td>Insecticide</td>
<td>0.58</td>
<td>0.82</td>
</tr>
<tr>
<td>62</td>
<td>Carboxin</td>
<td>Fongicide</td>
<td>−1.04</td>
<td>−0.85</td>
</tr>
<tr>
<td>63</td>
<td>Chlordane</td>
<td>Insecticide</td>
<td>−0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>64</td>
<td>Chlordecone</td>
<td>Insecticide</td>
<td>0.73</td>
<td>0.77</td>
</tr>
<tr>
<td>65</td>
<td>Chlorethoxyfos</td>
<td>Insecticide</td>
<td>2.27</td>
<td>2.28</td>
</tr>
<tr>
<td>66</td>
<td>Chlorfenac</td>
<td>Herbicide</td>
<td>−0.87</td>
<td>−0.72</td>
</tr>
<tr>
<td>67</td>
<td>Chlorfenethol</td>
<td>Insecticide</td>
<td>−0.27</td>
<td>−0.53</td>
</tr>
<tr>
<td>68</td>
<td>Chloridazon</td>
<td>Herbicide</td>
<td>−0.98</td>
<td>−1.01</td>
</tr>
<tr>
<td>69</td>
<td>Chlorobenzilate</td>
<td>Insecticide</td>
<td>−0.93</td>
<td>−0.95</td>
</tr>
<tr>
<td>70</td>
<td>Chloromethiluron</td>
<td>Insecticide</td>
<td>−2.04</td>
<td>−2.10</td>
</tr>
<tr>
<td>71</td>
<td>Chlorophacinone</td>
<td>Rodonticide</td>
<td>2.08</td>
<td>1.98</td>
</tr>
<tr>
<td>No.</td>
<td>Compound</td>
<td>Type</td>
<td>( \log [1/\text{LD}_{50}] ) (mmol/kg)(^{-1} )</td>
<td>Leverage ((h_i))</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>72</td>
<td>Chloropicrin</td>
<td>Insecticide</td>
<td>-0.18</td>
<td>0.003</td>
</tr>
<tr>
<td>73</td>
<td>Chlorpyrifos</td>
<td>Insecticide</td>
<td>0.74</td>
<td>0.005</td>
</tr>
<tr>
<td>74</td>
<td>Chlorpyrifos-methyl</td>
<td>Insecticide</td>
<td>-0.94</td>
<td>0.007</td>
</tr>
<tr>
<td>75</td>
<td>Chlorthiamid</td>
<td>Herbicide</td>
<td>-0.56</td>
<td>0.005</td>
</tr>
<tr>
<td>76</td>
<td>Chlorthion</td>
<td>Insecticide</td>
<td>-0.47</td>
<td>0.004</td>
</tr>
<tr>
<td>77</td>
<td>Clethodim</td>
<td>Herbicide</td>
<td>-0.50</td>
<td>0.004</td>
</tr>
<tr>
<td>78</td>
<td>Clodinafop-propargyl</td>
<td>Herbicide</td>
<td>-0.60</td>
<td>0.005</td>
</tr>
<tr>
<td>79</td>
<td>Cloethocarb</td>
<td>Insecticide</td>
<td>0.86</td>
<td>0.006</td>
</tr>
<tr>
<td>80</td>
<td>Clomazone</td>
<td>Herbicide</td>
<td>-0.76</td>
<td>0.006</td>
</tr>
<tr>
<td>81</td>
<td>Coumachlor</td>
<td>Rodonticide</td>
<td>1.33</td>
<td>0.010</td>
</tr>
<tr>
<td>82</td>
<td>Crotroxphos</td>
<td>Insecticide</td>
<td>0.68</td>
<td>0.005</td>
</tr>
<tr>
<td>83</td>
<td>Cyanazine</td>
<td>Herbicide</td>
<td>-0.08</td>
<td>0.003</td>
</tr>
<tr>
<td>84</td>
<td>Cyanophos</td>
<td>Insecticide</td>
<td>-0.40</td>
<td>0.004</td>
</tr>
<tr>
<td>85</td>
<td>Cycloxydim</td>
<td>Herbicide</td>
<td>-1.08</td>
<td>0.008</td>
</tr>
<tr>
<td>86</td>
<td>Cyhexatin</td>
<td>Insecticide</td>
<td>0.16</td>
<td>0.003</td>
</tr>
<tr>
<td>87</td>
<td>Cymoxanil</td>
<td>Fongicide</td>
<td>-0.58</td>
<td>0.005</td>
</tr>
<tr>
<td>88</td>
<td>Cypermethrin</td>
<td>Insecticide</td>
<td>0.16</td>
<td>0.003</td>
</tr>
<tr>
<td>89</td>
<td>Cyphenothrin</td>
<td>Insecticide</td>
<td>0.07</td>
<td>0.003</td>
</tr>
<tr>
<td>90</td>
<td>Cyprofuram</td>
<td>Fongicide</td>
<td>0.21</td>
<td>0.003</td>
</tr>
<tr>
<td>91</td>
<td>Cyromazine</td>
<td>Insecticide</td>
<td>-1.31</td>
<td>0.011</td>
</tr>
<tr>
<td>92</td>
<td>Dalapon</td>
<td>Herbicide</td>
<td>-1.81</td>
<td>0.018</td>
</tr>
<tr>
<td>93</td>
<td>Dazomet</td>
<td>Insecticide</td>
<td>-0.41</td>
<td>0.004</td>
</tr>
<tr>
<td>94</td>
<td>Deltamethrin</td>
<td>Insecticide</td>
<td>0.76</td>
<td>0.005</td>
</tr>
<tr>
<td>95</td>
<td>Demeton-S-methyl sulfone</td>
<td>Insecticide</td>
<td>0.91</td>
<td>0.006</td>
</tr>
<tr>
<td>96</td>
<td>Desmetryn</td>
<td>Herbicide</td>
<td>-0.81</td>
<td>0.006</td>
</tr>
<tr>
<td>97</td>
<td>Diafenthiuron</td>
<td>Insecticide</td>
<td>-0.73</td>
<td>0.006</td>
</tr>
<tr>
<td>98</td>
<td>Di-allate</td>
<td>Herbicide</td>
<td>-0.16</td>
<td>0.003</td>
</tr>
<tr>
<td>99</td>
<td>Dibromochloropropane</td>
<td>Insecticide</td>
<td>0.14</td>
<td>0.003</td>
</tr>
<tr>
<td>100</td>
<td>Dichloroacetic</td>
<td>Fongicide</td>
<td>0.15</td>
<td>0.003</td>
</tr>
<tr>
<td>101</td>
<td>Dichloroprop</td>
<td>Herbicide</td>
<td>-0.55</td>
<td>0.005</td>
</tr>
<tr>
<td>102</td>
<td>Dichlorvos</td>
<td>Insecticide</td>
<td>0.44</td>
<td>0.004</td>
</tr>
<tr>
<td>103</td>
<td>Dicofane</td>
<td>Insecticide</td>
<td>0.50</td>
<td>0.004</td>
</tr>
<tr>
<td>104</td>
<td>Dicofol</td>
<td>Insecticide</td>
<td>-0.19</td>
<td>0.003</td>
</tr>
<tr>
<td>105</td>
<td>Dicrotophos</td>
<td>Insecticide</td>
<td>1.14</td>
<td>0.008</td>
</tr>
<tr>
<td>106</td>
<td>Dieneochlor</td>
<td>Insecticide</td>
<td>-0.82</td>
<td>0.006</td>
</tr>
<tr>
<td>107</td>
<td>Diethatyl ethyl</td>
<td>Herbicide</td>
<td>-0.87</td>
<td>0.007</td>
</tr>
<tr>
<td>108</td>
<td>Difenamidine</td>
<td>Herbicide</td>
<td>-0.61</td>
<td>0.005</td>
</tr>
<tr>
<td>109</td>
<td>Diflovidazin</td>
<td>Insecticide</td>
<td>-0.29</td>
<td>0.004</td>
</tr>
<tr>
<td>No.</td>
<td>Compound</td>
<td>Type</td>
<td>log $[1/\text{LD}_{50}]$ (mmol/kg)$^{-1}$</td>
<td>Leverage ($h_i$)</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------</td>
<td>-----------------</td>
<td>------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Observed</td>
<td>Predicted</td>
</tr>
<tr>
<td>110</td>
<td>Diflumetorim</td>
<td>Fongicide</td>
<td>−0.14</td>
<td>−0.33</td>
</tr>
<tr>
<td>111</td>
<td>Dimethachlor</td>
<td>Herbicide</td>
<td>−0.80</td>
<td>−0.68</td>
</tr>
<tr>
<td>112</td>
<td>Dimethenamid</td>
<td>Herbicide</td>
<td>−0.16</td>
<td>−0.22</td>
</tr>
<tr>
<td>113</td>
<td>Dimethenamid-(P)</td>
<td>Herbicide</td>
<td>−0.19</td>
<td>−0.31</td>
</tr>
<tr>
<td>114</td>
<td>Dimethomorph</td>
<td>Fongicide</td>
<td>−1.00</td>
<td>−0.99</td>
</tr>
<tr>
<td>115</td>
<td>Dimethylvinphos</td>
<td>Insecticide</td>
<td>0.53</td>
<td>0.71</td>
</tr>
<tr>
<td>116</td>
<td>Dimexano</td>
<td>Herbicide</td>
<td>−0.05</td>
<td>−0.09</td>
</tr>
<tr>
<td>117</td>
<td>Dinobuton</td>
<td>Fongicide</td>
<td>0.37</td>
<td>0.27</td>
</tr>
<tr>
<td>118</td>
<td>Dinoseb</td>
<td>Herbicide</td>
<td>0.98</td>
<td>1.11</td>
</tr>
<tr>
<td>119</td>
<td>Dinoterb</td>
<td>Insecticide</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>120</td>
<td>Dioxathion</td>
<td>Insecticide</td>
<td>1.30</td>
<td>1.09</td>
</tr>
<tr>
<td>121</td>
<td>Diphenacine</td>
<td>Rodonticide</td>
<td>2.17</td>
<td>2.18</td>
</tr>
<tr>
<td>122</td>
<td>Diquat</td>
<td>Herbicide</td>
<td>−0.06</td>
<td>−0.11</td>
</tr>
<tr>
<td>123</td>
<td>Dithianon</td>
<td>Fongicide</td>
<td>−0.01</td>
<td>−0.07</td>
</tr>
<tr>
<td>124</td>
<td>Diuron</td>
<td>Herbicide</td>
<td>−0.27</td>
<td>−0.47</td>
</tr>
<tr>
<td>125</td>
<td>Edifenphos</td>
<td>Fongicide</td>
<td>0.32</td>
<td>0.17</td>
</tr>
<tr>
<td>126</td>
<td>Endothal</td>
<td>Herbicide</td>
<td>0.56</td>
<td>0.35</td>
</tr>
<tr>
<td>127</td>
<td>EPN</td>
<td>Insecticide</td>
<td>1.36</td>
<td>1.26</td>
</tr>
<tr>
<td>128</td>
<td>EPTC</td>
<td>Herbicide</td>
<td>−0.68</td>
<td>−0.86</td>
</tr>
<tr>
<td>129</td>
<td>Ethanedial</td>
<td>Herbicide</td>
<td>−0.31</td>
<td>−0.24</td>
</tr>
<tr>
<td>130</td>
<td>Ethoate-methyle</td>
<td>Insecticide</td>
<td>−0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>131</td>
<td>Ethoxysulfuron</td>
<td>Herbicide</td>
<td>−0.91</td>
<td>−0.94</td>
</tr>
<tr>
<td>132</td>
<td>Fenamidone</td>
<td>Fongicide</td>
<td>−0.81</td>
<td>−0.77</td>
</tr>
<tr>
<td>133</td>
<td>Fenchlorphos</td>
<td>Insecticide</td>
<td>−0.19</td>
<td>−0.17</td>
</tr>
<tr>
<td>134</td>
<td>Fenobucarb</td>
<td>Insecticide</td>
<td>−0.48</td>
<td>−0.22</td>
</tr>
<tr>
<td>135</td>
<td>Fenoprop</td>
<td>Herbicide</td>
<td>−0.38</td>
<td>−0.63</td>
</tr>
<tr>
<td>136</td>
<td>Fenpropathrin</td>
<td>Insecticide</td>
<td>−0.40</td>
<td>−0.38</td>
</tr>
<tr>
<td>137</td>
<td>Fenpropidin</td>
<td>Fongicide</td>
<td>−0.73</td>
<td>−0.71</td>
</tr>
<tr>
<td>138</td>
<td>Fenpropimorph</td>
<td>Fongicide</td>
<td>−0.74</td>
<td>−0.53</td>
</tr>
<tr>
<td>139</td>
<td>Fensulfothion</td>
<td>Insecticide</td>
<td>2.15</td>
<td>2.18</td>
</tr>
<tr>
<td>140</td>
<td>Fentin acetate</td>
<td>Fongicide</td>
<td>0.47</td>
<td>0.62</td>
</tr>
<tr>
<td>141</td>
<td>Fenvalerate</td>
<td>Insecticide</td>
<td>−0.03</td>
<td>−0.12</td>
</tr>
<tr>
<td>142</td>
<td>Fipronil</td>
<td>Insecticide</td>
<td>0.68</td>
<td>0.85</td>
</tr>
<tr>
<td>143</td>
<td>Florasulam</td>
<td>Herbicide</td>
<td>−1.14</td>
<td>−1.17</td>
</tr>
<tr>
<td>144</td>
<td>Fluazifop-butyl</td>
<td>Herbicide</td>
<td>−0.90</td>
<td>−1.11</td>
</tr>
<tr>
<td>145</td>
<td>Fluchloralin</td>
<td>Herbicide</td>
<td>−0.64</td>
<td>−0.59</td>
</tr>
<tr>
<td>146</td>
<td>Flucythrinat</td>
<td>Insecticide</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td>147</td>
<td>Flufenacet</td>
<td>Herbicide</td>
<td>−0.22</td>
<td>−0.21</td>
</tr>
<tr>
<td>No.</td>
<td>Compound</td>
<td>Type</td>
<td>log $[1/\text{LD}_{50}]$ (mmol/kg)$^{-1}$</td>
<td>Leverage ($h_i$)</td>
</tr>
<tr>
<td>-----</td>
<td>------------------</td>
<td>------------</td>
<td>------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Observed</td>
<td>Predicted</td>
</tr>
<tr>
<td>148</td>
<td>Flumorph</td>
<td>Fongicide</td>
<td>−0.86</td>
<td>−1.01</td>
</tr>
<tr>
<td>149</td>
<td>Fluoroacetamide</td>
<td>Insecticide</td>
<td>0.77</td>
<td>0.66</td>
</tr>
<tr>
<td>150</td>
<td>Fluquinconazole</td>
<td>Fongicide</td>
<td>0.53</td>
<td>0.41</td>
</tr>
<tr>
<td>151</td>
<td>Flusilazole</td>
<td>Fongicide</td>
<td>−0.33</td>
<td>−0.27</td>
</tr>
<tr>
<td>152</td>
<td>Fluvalinate</td>
<td>Insecticide</td>
<td>0.28</td>
<td>0.18</td>
</tr>
<tr>
<td>153</td>
<td>Fomesafen</td>
<td>Herbicide</td>
<td>−0.45</td>
<td>−0.38</td>
</tr>
<tr>
<td>154</td>
<td>Fonofos</td>
<td>Insecticide</td>
<td>1.56</td>
<td>1.55</td>
</tr>
<tr>
<td>155</td>
<td>Formetanate</td>
<td>Insecticide</td>
<td>1.17</td>
<td>1.21</td>
</tr>
<tr>
<td>156</td>
<td>Formothion</td>
<td>Insecticide</td>
<td>−0.15</td>
<td>−0.25</td>
</tr>
<tr>
<td>157</td>
<td>Fospirate</td>
<td>Insecticide</td>
<td>−0.45</td>
<td>−0.34</td>
</tr>
<tr>
<td>158</td>
<td>Fosthiazate</td>
<td>Insecticide</td>
<td>0.70</td>
<td>0.90</td>
</tr>
<tr>
<td>159</td>
<td>Furathio carb</td>
<td>Insecticide</td>
<td>0.86</td>
<td>0.65</td>
</tr>
<tr>
<td>160</td>
<td>Furfural</td>
<td>Fongicide</td>
<td>0.17</td>
<td>−0.06</td>
</tr>
<tr>
<td>161</td>
<td>Gamma-cyhalothrin</td>
<td>Insecticide</td>
<td>0.91</td>
<td>0.76</td>
</tr>
<tr>
<td>162</td>
<td>Halfenprox</td>
<td>Insecticide</td>
<td>0.56</td>
<td>0.61</td>
</tr>
<tr>
<td>163</td>
<td>Halosulfuron-methyl</td>
<td>Herbicide</td>
<td>−1.25</td>
<td>−1.11</td>
</tr>
<tr>
<td>164</td>
<td>Heptenophos</td>
<td>Insecticide</td>
<td>0.42</td>
<td>0.39</td>
</tr>
<tr>
<td>165</td>
<td>Hexaconazole</td>
<td>Fongicide</td>
<td>−0.84</td>
<td>−0.88</td>
</tr>
<tr>
<td>166</td>
<td>Hexazinone</td>
<td>Herbicide</td>
<td>−0.83</td>
<td>−0.81</td>
</tr>
<tr>
<td>167</td>
<td>Hymexazol</td>
<td>Fongicide</td>
<td>−1.21</td>
<td>−1.43</td>
</tr>
<tr>
<td>168</td>
<td>Icaridin</td>
<td>Insecticide</td>
<td>−0.99</td>
<td>−1.07</td>
</tr>
<tr>
<td>169</td>
<td>Imiprothrin</td>
<td>Insecticide</td>
<td>−0.45</td>
<td>−0.29</td>
</tr>
<tr>
<td>170</td>
<td>Ioxynil</td>
<td>Herbicide</td>
<td>0.46</td>
<td>0.62</td>
</tr>
<tr>
<td>171</td>
<td>Iprobenfos</td>
<td>Fongicide</td>
<td>−0.37</td>
<td>−0.59</td>
</tr>
<tr>
<td>172</td>
<td>Isocarbophos</td>
<td>Insecticide</td>
<td>0.76</td>
<td>0.58</td>
</tr>
<tr>
<td>173</td>
<td>Isoprocarb</td>
<td>Insecticide</td>
<td>−0.32</td>
<td>−0.41</td>
</tr>
<tr>
<td>174</td>
<td>Isoprotiolan</td>
<td>Fongicide</td>
<td>−0.61</td>
<td>−1.11</td>
</tr>
<tr>
<td>175</td>
<td>Isoproturon</td>
<td>Herbicide</td>
<td>−0.95</td>
<td>−0.73</td>
</tr>
<tr>
<td>176</td>
<td>Isoxathion</td>
<td>Insecticide</td>
<td>0.45</td>
<td>0.56</td>
</tr>
<tr>
<td>177</td>
<td>Kelevan</td>
<td>Insecticide</td>
<td>0.42</td>
<td>0.40</td>
</tr>
<tr>
<td>178</td>
<td>Lambda-cyhalothrin</td>
<td>Insecticide</td>
<td>0.91</td>
<td>1.03</td>
</tr>
<tr>
<td>179</td>
<td>Lindane</td>
<td>Insecticide</td>
<td>0.25</td>
<td>0.14</td>
</tr>
<tr>
<td>180</td>
<td>Linuron</td>
<td>Herbicide</td>
<td>−0.66</td>
<td>−0.79</td>
</tr>
<tr>
<td>181</td>
<td>Malathion</td>
<td>Insecticide</td>
<td>−0.73</td>
<td>−0.44</td>
</tr>
<tr>
<td>182</td>
<td>MCPA-thioethyl</td>
<td>Herbicide</td>
<td>−0.26</td>
<td>−0.33</td>
</tr>
<tr>
<td>183</td>
<td>MCPB</td>
<td>Herbicide</td>
<td>−1.27</td>
<td>−1.31</td>
</tr>
<tr>
<td>184</td>
<td>Mecarbam</td>
<td>Insecticide</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>185</td>
<td>Mepiquat</td>
<td>Herbicide</td>
<td>−1.12</td>
<td>−1.23</td>
</tr>
<tr>
<td>No.</td>
<td>Compound</td>
<td>Type</td>
<td>log [1/LD$_{50}$ (mmol/kg)$^{-1}$]</td>
<td>Leverage ($h_i$)</td>
</tr>
<tr>
<td>-----</td>
<td>---------------</td>
<td>--------------</td>
<td>-----------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>186</td>
<td>Metalaxyl</td>
<td>Fongicide</td>
<td>−0.36</td>
<td>0.004</td>
</tr>
<tr>
<td>187</td>
<td>Metamitron</td>
<td>Herbicide</td>
<td>−0.77</td>
<td>0.006</td>
</tr>
<tr>
<td>188</td>
<td>Methomyl</td>
<td>Insecticide</td>
<td>0.73</td>
<td>0.005</td>
</tr>
<tr>
<td>189</td>
<td>Metominostrobin</td>
<td>Fongicide</td>
<td>−0.40</td>
<td>0.004</td>
</tr>
<tr>
<td>190</td>
<td>Metsulfovax</td>
<td>Fongicide</td>
<td>−1.23</td>
<td>0.010</td>
</tr>
<tr>
<td>191</td>
<td>Mevinphos</td>
<td>Insecticide</td>
<td>1.81</td>
<td>0.016</td>
</tr>
<tr>
<td>192</td>
<td>Monocrotophos</td>
<td>Insecticide</td>
<td>1.20</td>
<td>0.008</td>
</tr>
<tr>
<td>193</td>
<td>Morphothion</td>
<td>Insecticide</td>
<td>0.18</td>
<td>0.003</td>
</tr>
<tr>
<td>194</td>
<td>Naled</td>
<td>Insecticide</td>
<td>0.66</td>
<td>0.004</td>
</tr>
<tr>
<td>195</td>
<td>Naptalam</td>
<td>Herbicide</td>
<td>−0.78</td>
<td>0.006</td>
</tr>
<tr>
<td>196</td>
<td>Nithiazine</td>
<td>Insecticide</td>
<td>−0.15</td>
<td>0.003</td>
</tr>
<tr>
<td>197</td>
<td>Nitrapyrin</td>
<td>Bactéricide</td>
<td>−0.49</td>
<td>0.004</td>
</tr>
<tr>
<td>198</td>
<td>Nitrofen</td>
<td>Herbicide</td>
<td>−0.97</td>
<td>0.007</td>
</tr>
<tr>
<td>199</td>
<td>Octhilinone</td>
<td>Fongicide</td>
<td>−0.41</td>
<td>0.004</td>
</tr>
<tr>
<td>200</td>
<td>Ofurace</td>
<td>Fongicide</td>
<td>−0.97</td>
<td>0.007</td>
</tr>
<tr>
<td>201</td>
<td>Oxycarboxin</td>
<td>Fongicide</td>
<td>−0.79</td>
<td>0.006</td>
</tr>
<tr>
<td>202</td>
<td>Oxydemeton-methyl</td>
<td>Insecticide</td>
<td>0.71</td>
<td>0.005</td>
</tr>
<tr>
<td>203</td>
<td>Paraquat</td>
<td>Herbicide</td>
<td>0.23</td>
<td>0.003</td>
</tr>
<tr>
<td>204</td>
<td>Parathion</td>
<td>Insecticide</td>
<td>2.16</td>
<td>0.021</td>
</tr>
<tr>
<td>205</td>
<td>Parathion methyl</td>
<td>Insecticide</td>
<td>1.94</td>
<td>0.018</td>
</tr>
<tr>
<td>206</td>
<td>Pebulate</td>
<td>Herbicide</td>
<td>−0.74</td>
<td>0.006</td>
</tr>
<tr>
<td>207</td>
<td>Pethoxamid</td>
<td>Herbicide</td>
<td>−0.52</td>
<td>0.004</td>
</tr>
<tr>
<td>208</td>
<td>Phenkapton</td>
<td>Insecticide</td>
<td>0.93</td>
<td>0.006</td>
</tr>
<tr>
<td>209</td>
<td>Phenthoate</td>
<td>Insecticide</td>
<td>0.11</td>
<td>0.003</td>
</tr>
<tr>
<td>210</td>
<td>Phosalone</td>
<td>Insecticide</td>
<td>0.49</td>
<td>0.004</td>
</tr>
<tr>
<td>211</td>
<td>Picloram</td>
<td>Herbicide</td>
<td>−1.22</td>
<td>0.010</td>
</tr>
<tr>
<td>212</td>
<td>Piperophos</td>
<td>Herbicide</td>
<td>0.04</td>
<td>0.003</td>
</tr>
<tr>
<td>213</td>
<td>Pirimicarb</td>
<td>Insecticide</td>
<td>0.22</td>
<td>0.003</td>
</tr>
<tr>
<td>214</td>
<td>Plifenate</td>
<td>Insecticide</td>
<td>−1.47</td>
<td>0.013</td>
</tr>
<tr>
<td>215</td>
<td>Prallethrin</td>
<td>Insecticide</td>
<td>−0.18</td>
<td>0.003</td>
</tr>
<tr>
<td>216</td>
<td>Pretilachlor</td>
<td>Herbicide</td>
<td>−1.29</td>
<td>0.011</td>
</tr>
<tr>
<td>217</td>
<td>Prometon</td>
<td>Herbicide</td>
<td>−0.83</td>
<td>0.006</td>
</tr>
<tr>
<td>218</td>
<td>Propanil</td>
<td>Herbicide</td>
<td>−0.64</td>
<td>0.005</td>
</tr>
<tr>
<td>219</td>
<td>Propargite</td>
<td>Insecticide</td>
<td>−0.88</td>
<td>0.007</td>
</tr>
<tr>
<td>220</td>
<td>Propiconazole</td>
<td>Fongicide</td>
<td>−0.45</td>
<td>0.004</td>
</tr>
<tr>
<td>221</td>
<td>Propoxur</td>
<td>Insecticide</td>
<td>0.62</td>
<td>0.004</td>
</tr>
<tr>
<td>222</td>
<td>Prosulfuron</td>
<td>Herbicide</td>
<td>−0.11</td>
<td>0.003</td>
</tr>
<tr>
<td>223</td>
<td>Prothiofos</td>
<td>Insecticide</td>
<td>−0.43</td>
<td>0.004</td>
</tr>
<tr>
<td>No.</td>
<td>Compound</td>
<td>Type</td>
<td>( \log [1/\text{LD}_{50}] ) (mmol/kg)(^{-1} )</td>
<td>Leverage (( h_i ))</td>
</tr>
<tr>
<td>-----</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Observed</td>
<td>Predicted</td>
</tr>
<tr>
<td>224</td>
<td>Pymetrozine</td>
<td>Insecticide</td>
<td>−1.43</td>
<td>−1.46</td>
</tr>
<tr>
<td>225</td>
<td>Pyrazophos</td>
<td>Fongicide</td>
<td>0.39</td>
<td>−0.18</td>
</tr>
<tr>
<td>226</td>
<td>Pyrazoxyfen</td>
<td>Herbicide</td>
<td>−0.61</td>
<td>−0.77</td>
</tr>
<tr>
<td>227</td>
<td>Pyridaben</td>
<td>Insecticide</td>
<td>0.36</td>
<td>0.37</td>
</tr>
<tr>
<td>228</td>
<td>Pyridafenthion</td>
<td>Insecticide</td>
<td>−0.35</td>
<td>−0.34</td>
</tr>
<tr>
<td>229</td>
<td>Pyrifox</td>
<td>Fongicide</td>
<td>−0.99</td>
<td>−1.07</td>
</tr>
<tr>
<td>230</td>
<td>Pyrimethanil</td>
<td>Fongicide</td>
<td>−1.32</td>
<td>−1.34</td>
</tr>
<tr>
<td>231</td>
<td>Pyroquiline</td>
<td>Fongicide</td>
<td>−0.27</td>
<td>−0.52</td>
</tr>
<tr>
<td>232</td>
<td>Quinalphos</td>
<td>Insecticide</td>
<td>0.62</td>
<td>0.41</td>
</tr>
<tr>
<td>233</td>
<td>Quinclorac</td>
<td>Herbicide</td>
<td>−1.04</td>
<td>−1.01</td>
</tr>
<tr>
<td>234</td>
<td>Sethoxydim</td>
<td>Herbicide</td>
<td>−1.06</td>
<td>−1.08</td>
</tr>
<tr>
<td>235</td>
<td>Simetryn</td>
<td>Herbicide</td>
<td>−0.38</td>
<td>−0.58</td>
</tr>
<tr>
<td>236</td>
<td>Sulfoxaflor</td>
<td>Insecticide</td>
<td>1.74</td>
<td>1.71</td>
</tr>
<tr>
<td>237</td>
<td>Sulprofos</td>
<td>Insecticide</td>
<td>0.24</td>
<td>0.32</td>
</tr>
<tr>
<td>239</td>
<td>Tebuconazole</td>
<td>Fongicide</td>
<td>−0.68</td>
<td>−0.83</td>
</tr>
<tr>
<td>240</td>
<td>Tecloftalam</td>
<td>Fongicide</td>
<td>−0.95</td>
<td>−1.02</td>
</tr>
<tr>
<td>241</td>
<td>Tecnazene</td>
<td>Fongicide</td>
<td>−0.52</td>
<td>−0.47</td>
</tr>
<tr>
<td>242</td>
<td>Tefluthrin</td>
<td>Insecticide</td>
<td>1.31</td>
<td>1.27</td>
</tr>
<tr>
<td>243</td>
<td>Thiocarboxime</td>
<td>Insecticide</td>
<td>1.16</td>
<td>0.90</td>
</tr>
<tr>
<td>244</td>
<td>Thiopicarb</td>
<td>Insecticide</td>
<td>0.64</td>
<td>0.77</td>
</tr>
<tr>
<td>245</td>
<td>Thiofanox</td>
<td>Insecticide</td>
<td>1.46</td>
<td>1.38</td>
</tr>
<tr>
<td>246</td>
<td>Thimeton</td>
<td>Insecticide</td>
<td>0.79</td>
<td>0.73</td>
</tr>
<tr>
<td>247</td>
<td>Tolfenpyrad</td>
<td>Insecticide</td>
<td>−0.05</td>
<td>−0.12</td>
</tr>
<tr>
<td>248</td>
<td>Teloxoxydim</td>
<td>Herbicide</td>
<td>−0.15</td>
<td>−0.11</td>
</tr>
<tr>
<td>249</td>
<td>Tri-allate</td>
<td>Herbicide</td>
<td>−0.44</td>
<td>−0.34</td>
</tr>
<tr>
<td>250</td>
<td>Tribufos</td>
<td>Herbicide</td>
<td>0.04</td>
<td>−0.15</td>
</tr>
<tr>
<td>251</td>
<td>Trichlorfon</td>
<td>Insecticide</td>
<td>0.20</td>
<td>0.13</td>
</tr>
<tr>
<td>252</td>
<td>Trichloronate</td>
<td>Insecticide</td>
<td>1.03</td>
<td>0.89</td>
</tr>
<tr>
<td>253</td>
<td>Tricyclazole</td>
<td>Fongicide</td>
<td>0.01</td>
<td>−0.14</td>
</tr>
<tr>
<td>254</td>
<td>Tridiphane</td>
<td>Herbicide</td>
<td>−0.88</td>
<td>−0.91</td>
</tr>
<tr>
<td>255</td>
<td>Trietazine</td>
<td>Herbicide</td>
<td>−0.08</td>
<td>−0.18</td>
</tr>
<tr>
<td>256</td>
<td>Triflumizole</td>
<td>Fongicide</td>
<td>−0.47</td>
<td>−0.68</td>
</tr>
<tr>
<td>257</td>
<td>Trimethacarb</td>
<td>Insecticide</td>
<td>−0.25</td>
<td>−0.23</td>
</tr>
<tr>
<td>258</td>
<td>Vamidothion</td>
<td>Insecticide</td>
<td>0.65</td>
<td>1.02</td>
</tr>
</tbody>
</table>

**Validation set**

<table>
<thead>
<tr>
<th></th>
<th>Compound</th>
<th>Type</th>
<th>Observed</th>
<th>Predicted</th>
<th>Leverage (( h_i ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>259</td>
<td>2,4-D</td>
<td>Herbicide</td>
<td>−0.33</td>
<td>−0.39</td>
<td>0.004</td>
</tr>
<tr>
<td>No.</td>
<td>Compound</td>
<td>Type</td>
<td>log [1/LD50] (mmol/kg)$^{-1}$</td>
<td>Leverage ($h_i$)</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---------------------------</td>
<td>---------------</td>
<td>-------------------------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Observed</td>
<td>Predicted</td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>Aldoxycarb</td>
<td>Insecticide</td>
<td>0.92</td>
<td>0.87</td>
<td>0.006</td>
</tr>
<tr>
<td>261</td>
<td>Allethrin</td>
<td>Insecticide</td>
<td>v0.35</td>
<td>−0.14</td>
<td>0.004</td>
</tr>
<tr>
<td>262</td>
<td>Alpha-cypermethrin</td>
<td>Insecticide</td>
<td>0.86</td>
<td>0.86</td>
<td>0.006</td>
</tr>
<tr>
<td>263</td>
<td>Azinphos-ethyl</td>
<td>Insecticide</td>
<td>1.46</td>
<td>1.42</td>
<td>0.011</td>
</tr>
<tr>
<td>264</td>
<td>Barban</td>
<td>Herbicide</td>
<td>−0.31</td>
<td>−0.31</td>
<td>0.004</td>
</tr>
<tr>
<td>265</td>
<td>Bensulide</td>
<td>Herbicide</td>
<td>0.17</td>
<td>0.21</td>
<td>0.003</td>
</tr>
<tr>
<td>266</td>
<td>Bensultap</td>
<td>Insecticide</td>
<td>−0.41</td>
<td>−0.16</td>
<td>0.004</td>
</tr>
<tr>
<td>267</td>
<td>Beta-cypermethrin</td>
<td>Insecticide</td>
<td>0.65</td>
<td>0.52</td>
<td>0.004</td>
</tr>
<tr>
<td>268</td>
<td>Chlorbromuron</td>
<td>Herbicide</td>
<td>−0.86</td>
<td>−0.85</td>
<td>0.007</td>
</tr>
<tr>
<td>269</td>
<td>Chlorbufam</td>
<td>Herbicide</td>
<td>−1.03</td>
<td>−0.88</td>
<td>0.008</td>
</tr>
<tr>
<td>270</td>
<td>Chlorpropham</td>
<td>Herbicide</td>
<td>−1.29</td>
<td>−1.58</td>
<td>0.011</td>
</tr>
<tr>
<td>271</td>
<td>Closantel</td>
<td>Insecticide</td>
<td>0.40</td>
<td>0.12</td>
<td>0.003</td>
</tr>
<tr>
<td>272</td>
<td>Crimidine</td>
<td>Rodonticide</td>
<td>2.14</td>
<td>2.18</td>
<td>0.021</td>
</tr>
<tr>
<td>273</td>
<td>Demeton-S-methyl</td>
<td>Insecticide</td>
<td>0.76</td>
<td>1.05</td>
<td>0.005</td>
</tr>
<tr>
<td>274</td>
<td>Dichlorprop-P</td>
<td>Herbicide</td>
<td>−0.38</td>
<td>−0.76</td>
<td>0.004</td>
</tr>
<tr>
<td>275</td>
<td>Dimethoate</td>
<td>Insecticide</td>
<td>−0.03</td>
<td>−0.03</td>
<td>0.003</td>
</tr>
<tr>
<td>276</td>
<td>Dinocap</td>
<td>Fongicide</td>
<td>−0.52</td>
<td>−0.54</td>
<td>0.004</td>
</tr>
<tr>
<td>277</td>
<td>Dioxabenzophos</td>
<td>Insecticide</td>
<td>0.24</td>
<td>0.26</td>
<td>0.003</td>
</tr>
<tr>
<td>278</td>
<td>Ditalimfos</td>
<td>Fongicide</td>
<td>−1.22</td>
<td>−1.15</td>
<td>0.010</td>
</tr>
<tr>
<td>279</td>
<td>DNOC</td>
<td>Herbicide</td>
<td>0.90</td>
<td>0.87</td>
<td>0.006</td>
</tr>
<tr>
<td>280</td>
<td>Endosulfan</td>
<td>Insecticide</td>
<td>1.03</td>
<td>0.71</td>
<td>0.007</td>
</tr>
<tr>
<td>281</td>
<td>Etaconazole</td>
<td>Fongicide</td>
<td>−0.61</td>
<td>−0.52</td>
<td>0.005</td>
</tr>
<tr>
<td>282</td>
<td>Ethiofencarb</td>
<td>Insecticide</td>
<td>0.05</td>
<td>−0.05</td>
<td>0.003</td>
</tr>
<tr>
<td>283</td>
<td>Ethiprole</td>
<td>Insecticide</td>
<td>−1.25</td>
<td>−1.25</td>
<td>0.010</td>
</tr>
<tr>
<td>284</td>
<td>Fenarimol</td>
<td>Fongicide</td>
<td>−0.88</td>
<td>−1.11</td>
<td>0.007</td>
</tr>
<tr>
<td>285</td>
<td>Fenazaquin</td>
<td>Acaricide</td>
<td>0.36</td>
<td>0.38</td>
<td>0.003</td>
</tr>
<tr>
<td>286</td>
<td>Fenitrothion</td>
<td>Insecticide</td>
<td>−0.08</td>
<td>−0.08</td>
<td>0.003</td>
</tr>
<tr>
<td>287</td>
<td>Flonicamid</td>
<td>Insecticide</td>
<td>−0.59</td>
<td>−0.47</td>
<td>0.005</td>
</tr>
<tr>
<td>288</td>
<td>Fluazifop-P-butyl</td>
<td>Herbicide</td>
<td>−0.81</td>
<td>−0.51</td>
<td>0.006</td>
</tr>
<tr>
<td>289</td>
<td>Fluoroglycofen</td>
<td>Herbicide</td>
<td>−0.55</td>
<td>−0.43</td>
<td>0.005</td>
</tr>
<tr>
<td>290</td>
<td>Furalaxyl</td>
<td>Fongicide</td>
<td>−0.50</td>
<td>−0.75</td>
<td>0.004</td>
</tr>
<tr>
<td>291</td>
<td>Furmecyclox</td>
<td>Fongicide</td>
<td>−1.18</td>
<td>−1.36</td>
<td>0.009</td>
</tr>
<tr>
<td>292</td>
<td>Glufosinate</td>
<td>Herbicide</td>
<td>−0.95</td>
<td>−0.91</td>
<td>0.007</td>
</tr>
<tr>
<td>293</td>
<td>Glutaraldehyde</td>
<td>Fongicide</td>
<td>−0.13</td>
<td>0.05</td>
<td>0.003</td>
</tr>
<tr>
<td>294</td>
<td>Halofenozide</td>
<td>Insecticide</td>
<td>−0.94</td>
<td>−1.08</td>
<td>0.007</td>
</tr>
<tr>
<td>295</td>
<td>Imazalil</td>
<td>Fongicide</td>
<td>0.12</td>
<td>0.35</td>
<td>0.003</td>
</tr>
<tr>
<td>296</td>
<td>Indoxacarb</td>
<td>Insecticide</td>
<td>0.29</td>
<td>0.19</td>
<td>0.003</td>
</tr>
<tr>
<td>297</td>
<td>Isofenphos-methyl</td>
<td>Insecticide</td>
<td>1.19</td>
<td>1.37</td>
<td>0.008</td>
</tr>
<tr>
<td>No.</td>
<td>Compound</td>
<td>Type</td>
<td>log $[1/LD_{50}]$ (mmol/kg)$^{-1}$</td>
<td></td>
<td>Leverage ($h_i$)</td>
</tr>
<tr>
<td>-----</td>
<td>---------------</td>
<td>--------------</td>
<td>-----------------------------------</td>
<td>---</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Observed</td>
<td>Predicted</td>
<td></td>
</tr>
<tr>
<td>298</td>
<td>Leptophos</td>
<td>Insecticide</td>
<td>0.98</td>
<td>0.76</td>
<td>0.007</td>
</tr>
<tr>
<td>299</td>
<td>MCPA</td>
<td>Herbicide</td>
<td>−0.68</td>
<td>−0.79</td>
<td>0.005</td>
</tr>
<tr>
<td>300</td>
<td>Mecoprop</td>
<td>Herbicide</td>
<td>−0.73</td>
<td>−1.09</td>
<td>0.006</td>
</tr>
<tr>
<td>301</td>
<td>Metazachlor</td>
<td>Herbicide</td>
<td>−1.10</td>
<td>−0.95</td>
<td>0.009</td>
</tr>
<tr>
<td>302</td>
<td>Metconazole</td>
<td>Herbicide</td>
<td>−0.27</td>
<td>−0.06</td>
<td>0.003</td>
</tr>
<tr>
<td>303</td>
<td>Methazole</td>
<td>Herbicide</td>
<td>−0.47</td>
<td>−0.61</td>
<td>0.004</td>
</tr>
<tr>
<td>304</td>
<td>Methidathion</td>
<td>Insecticide</td>
<td>1.08</td>
<td>1.03</td>
<td>0.007</td>
</tr>
<tr>
<td>305</td>
<td>Metolachlor</td>
<td>Herbicide</td>
<td>−0.63</td>
<td>−0.63</td>
<td>0.005</td>
</tr>
<tr>
<td>306</td>
<td>Metribuzin</td>
<td>Herbicide</td>
<td>0.83</td>
<td>0.52</td>
<td>0.005</td>
</tr>
<tr>
<td>307</td>
<td>Molinate</td>
<td>Herbicide</td>
<td>−0.41</td>
<td>−0.18</td>
<td>0.004</td>
</tr>
<tr>
<td>308</td>
<td>Monolinuron</td>
<td>Herbicide</td>
<td>−0.99</td>
<td>−0.63</td>
<td>0.008</td>
</tr>
<tr>
<td>309</td>
<td>Nitenpyram</td>
<td>Herbicide</td>
<td>−0.76</td>
<td>−0.85</td>
<td>0.006</td>
</tr>
<tr>
<td>310</td>
<td>Oxadixyl</td>
<td>Fongicide</td>
<td>−0.82</td>
<td>−1.12</td>
<td>0.006</td>
</tr>
<tr>
<td>311</td>
<td>Oxamyl</td>
<td>Herbicide</td>
<td>1.94</td>
<td>2.04</td>
<td>0.018</td>
</tr>
<tr>
<td>312</td>
<td>Pendimethalin</td>
<td>Herbicide</td>
<td>−1.05</td>
<td>−1.27</td>
<td>0.008</td>
</tr>
<tr>
<td>313</td>
<td>Phosmet</td>
<td>Insecticide</td>
<td>0.45</td>
<td>0.04</td>
<td>0.004</td>
</tr>
<tr>
<td>314</td>
<td>Profenofos</td>
<td>Insecticide</td>
<td>0.02</td>
<td>−0.01</td>
<td>0.003</td>
</tr>
<tr>
<td>315</td>
<td>Promecarb</td>
<td>Herbicide</td>
<td>0.77</td>
<td>0.58</td>
<td>0.005</td>
</tr>
<tr>
<td>316</td>
<td>Propazine</td>
<td>Herbicide</td>
<td>−1.22</td>
<td>−1.49</td>
<td>0.010</td>
</tr>
<tr>
<td>317</td>
<td>Prosulfocarb</td>
<td>Herbicide</td>
<td>−0.86</td>
<td>−0.80</td>
<td>0.007</td>
</tr>
<tr>
<td>318</td>
<td>Prothoate</td>
<td>Herbicide</td>
<td>1.55</td>
<td>1.14</td>
<td>0.012</td>
</tr>
<tr>
<td>319</td>
<td>Tebutam</td>
<td>Herbicide</td>
<td>−1.43</td>
<td>−1.51</td>
<td>0.012</td>
</tr>
<tr>
<td>320</td>
<td>Tebutiuron</td>
<td>Herbicide</td>
<td>−0.16</td>
<td>−0.07</td>
<td>0.003</td>
</tr>
<tr>
<td>321</td>
<td>Tepraloxydim</td>
<td>Herbicide</td>
<td>−1.36</td>
<td>−1.48</td>
<td>0.011</td>
</tr>
<tr>
<td>322</td>
<td>Terbufos</td>
<td>Herbicide</td>
<td>2.25</td>
<td>2.14</td>
<td>0.023</td>
</tr>
<tr>
<td>323</td>
<td>Tetraconazole</td>
<td>Herbicide</td>
<td>−0.46</td>
<td>−0.79</td>
<td>0.004</td>
</tr>
<tr>
<td>324</td>
<td>Thiacloprid</td>
<td>Herbicide</td>
<td>−0.18</td>
<td>−0.24</td>
<td>0.003</td>
</tr>
<tr>
<td>325</td>
<td>Thiobencarb</td>
<td>Herbicide</td>
<td>−0.44</td>
<td>−0.66</td>
<td>0.004</td>
</tr>
<tr>
<td>326</td>
<td>Tralomethrin</td>
<td>Herbicide</td>
<td>0.57</td>
<td>0.65</td>
<td>0.004</td>
</tr>
<tr>
<td>327</td>
<td>Triazamate</td>
<td>Herbicide</td>
<td>0.71</td>
<td>0.49</td>
<td>0.005</td>
</tr>
<tr>
<td>328</td>
<td>Tridemorph</td>
<td>Fongicide</td>
<td>−0.19</td>
<td>−0.31</td>
<td>0.003</td>
</tr>
<tr>
<td>329</td>
<td>Vernolate</td>
<td>Herbicide</td>
<td>−0.72</td>
<td>−0.12</td>
<td>0.006</td>
</tr>
</tbody>
</table>
Table 2.

List of descriptors used in the development of QSAR model.

<table>
<thead>
<tr>
<th>Category</th>
<th>Descriptor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2D Autocorrelations indices</strong></td>
<td>MATS2p</td>
<td>Moran autocorrelation of lag 2 weighted by polarizability</td>
</tr>
<tr>
<td></td>
<td>MATS1m</td>
<td>Moran autocorrelation of lag 1 weighted by mass</td>
</tr>
<tr>
<td>Atom-centred fragments</td>
<td>N-072</td>
<td>RCO-N&lt;/&gt;N − X = X</td>
</tr>
<tr>
<td></td>
<td>H-046</td>
<td>H attached to C0(sp3) no X attached to next C</td>
</tr>
<tr>
<td>Geometrical descriptors</td>
<td>PJI3</td>
<td>3D Petitjean shape index</td>
</tr>
<tr>
<td></td>
<td>H6m</td>
<td>H autocorrelation of lag 6/weighted by mass</td>
</tr>
<tr>
<td>Getaway descriptors</td>
<td>HATSe</td>
<td>Leverage-weighted total index/weighted by Sanderson electronegativity</td>
</tr>
<tr>
<td></td>
<td>HATS0m</td>
<td>Leverage-weighted autocorrelation of lag 0/weighted by mass</td>
</tr>
<tr>
<td>RDF descriptor</td>
<td>RDF020e</td>
<td>Radial distribution function—020/weighted by Sanderson electronegativity</td>
</tr>
<tr>
<td></td>
<td>RDF030e</td>
<td>Radial distribution function—030/weighted by Sanderson electronegativity</td>
</tr>
<tr>
<td></td>
<td>Mor15m</td>
<td>Signal 15/weighted by mass</td>
</tr>
<tr>
<td>3D-Morse descriptor</td>
<td>Mor23u</td>
<td>Signal 23/unweighted</td>
</tr>
<tr>
<td></td>
<td>Mor26u</td>
<td>Signal 26/unweighted</td>
</tr>
<tr>
<td></td>
<td>Du</td>
<td>D total accessibility index/unweighted</td>
</tr>
<tr>
<td>Whim descriptors</td>
<td>E1u</td>
<td>1st component accessibility directional WHIM index/unweighted</td>
</tr>
<tr>
<td>Functional group counts</td>
<td>nArX</td>
<td>Number of X on aromatic ring</td>
</tr>
<tr>
<td>Constitutional indices</td>
<td>nS</td>
<td>Number of sulfur atoms</td>
</tr>
</tbody>
</table>
Table 3.
Selected parameters of the optimal multi-layer perceptron.

<table>
<thead>
<tr>
<th>Parameters studied</th>
<th>MSE (minimum value)</th>
<th>Selected parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The database distribution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training (80%) and validation (20%)</td>
<td>0.0311</td>
<td></td>
</tr>
<tr>
<td>Training (79%) and validation (21%)</td>
<td>0.0317</td>
<td></td>
</tr>
<tr>
<td>Training (78.5%) and validation (21.5%)</td>
<td><strong>0.0295</strong></td>
<td>Training (78.5%) and validation (21.5%)</td>
</tr>
<tr>
<td>Training (78%) and validation (22%)</td>
<td>0.0345</td>
<td></td>
</tr>
<tr>
<td>Training (77%) and validation (23%)</td>
<td>0.0382</td>
<td></td>
</tr>
<tr>
<td><strong>Activation functions (hidden neurons/output neurons)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sigmoid–sigmoid</td>
<td>0.0291</td>
<td></td>
</tr>
<tr>
<td>Sigmoid–linear</td>
<td>0.0293</td>
<td></td>
</tr>
<tr>
<td>Sigmoid–tangent hyperbolic</td>
<td>0.1054</td>
<td></td>
</tr>
<tr>
<td>Tangent hyperbolic–sigmoid</td>
<td>0.1719</td>
<td></td>
</tr>
<tr>
<td>Tangent hyperbolic–linear</td>
<td><strong>0.0290</strong></td>
<td>Tangent hyperbolic–linear</td>
</tr>
<tr>
<td>Tangent hyperbolic–tangent hyperbolic</td>
<td>0.0293</td>
<td></td>
</tr>
<tr>
<td>Linear–sigmoid</td>
<td>0.1563</td>
<td></td>
</tr>
<tr>
<td>Linear–tangent hyperbolic</td>
<td>0.0306</td>
<td></td>
</tr>
<tr>
<td>Linear–linear</td>
<td>0.0299</td>
<td></td>
</tr>
<tr>
<td><strong>Number of neurons in the hidden layer</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–16</td>
<td>0.0290</td>
<td>9 Neurons</td>
</tr>
<tr>
<td><strong>Learning algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quasi–Newton back propagation (BFGS)</td>
<td><strong>0.0290</strong></td>
<td></td>
</tr>
<tr>
<td>Levenberg–Marquardt (LM)</td>
<td>0.0293</td>
<td>Quasi–Newton back propagation (BFGS)</td>
</tr>
<tr>
<td>Scaled conjugate gradient (SCG)</td>
<td>0.0395</td>
<td></td>
</tr>
<tr>
<td>Conjugate gradient descent (CGP)</td>
<td>0.0346</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.
Performance of MLP-ANN model for pesticides.

\begin{align*}
R^2 & \quad 0.963 \\
\text{Training set (} n = 258 \text{)} & \quad Q^2_{\text{LOO}} \quad 0.962 \\
& \quad \text{RMS} \quad 0.164 \\
& \quad R^2_{\text{ext}} \quad 0.95 \\
\text{Validation set (} n = 71 \text{)} & \quad Q^2_{\text{ext}} \quad 0.948 \\
& \quad \text{RMS} \quad 0.201
\end{align*}