Asian Journal of Green Chemistry 2 (2018) 144-159

Contents lists available at Avicenna Publishing Corporation (APC)

Asian Journal of Green Chemistry

Journal homepage: www.ajgreenchem.com

Orginal Research Article

Quantitative structure-retention relationships applied to chromatographic retention of ecotoxicity of anilines and phenols

Mehrdad Shahpar[a,*], Sharmin Esmaeilpoor[b]

[a] Director of Ilam Petrochemical Company, Ilam, Iran

[b] Department of Chemistry, Payame Noor University, P.O. BOX 19395-4697, Tehran, Iran

ARTICLE INFORMATION

Received: 7 October 2017
Received in revised: 10 December 2017
Accepted: 11 December 2017
Available online: 7 February 2018

DOI: 10.22631/ajgc.2018.100313.1023

KEYWORDS

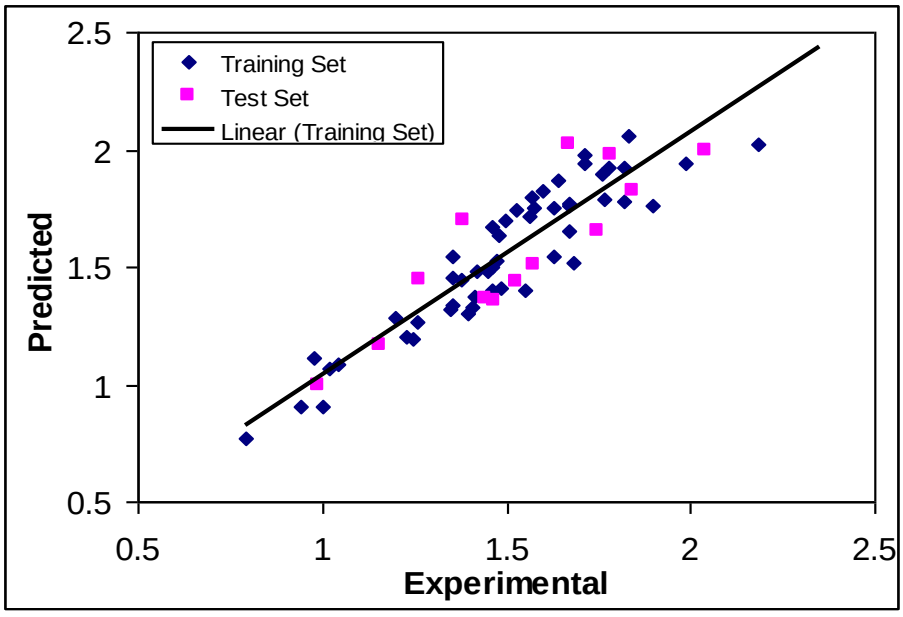
Ecotoxicity
Environmental hazard
Phenols
Anilines
Quantitative stature retention relationship

ABSTRACT

Aniline, phenol, and their derivatives are widely used in industrial chemicals that consequently have a high potential for environmental pollution. Genetic algorithm and partial least square (GA-PLS), kernel partial least square (GA-KPLS) and Levenberg-Marquardt artificial neural network (L-M ANN) techniques were used to investigate the correlation between chromatographic retention (log k) and descriptors for modelling the toxicity to fathead minnows of anilines and phenols. Descriptors of GA-PLS model were selected as inputs in L-M ANN model. The described model does not require experimental parameters and potentially provides useful prediction for log k of new compounds. Finally a model with a low prediction error and a good correlation coefficient was obtained by L-M ANN. The stability and prediction ability of L-M ANN model was validated using external test set techniques.

Corresponding author, e-mail address: shahpar2012@gmail.com (M. Shahpar).
Tel.: +98914 3100801.

Graphical Abstract



Introduction

Aromatic amines and phenols (Anilines and related derivates) are widely used industrial chemicals and are therefore an important class of environmental pollutants. Aniline is the parent molecule of a vast family of aromatic amines. Since its discovery in 1826, it has become one of the hundred most important building blocks in chemistry. Aniline and its derivatives containing chloro-substituents are used as intermediates in many different fields of applications, such as the production of isocyanates, rubber processing chemicals, dyes and pigments, agricultural chemicals and pharmaceuticals. These compounds can be released into the surface water as industrial effluents or as break-down products of pesticides and dyes. A large database on the effects of single chemicals has been developed using the fathead minnow for acute partial and life-cycle tests [1].

Healthy animals are the most important aspect for a good toxicity test. Emphasis should be placed on determining the quality of the organisms used for producing the test organisms. This report and the video culturing of fathead minnows (Pimephales promelas) were produced by EPA to clarify and expand on culturing methods explained in the acute methods manual. The waters to be used for culturing fathead minnows are any toxicity-free freshwater including natural water, drinking water, or reconstituted water. The water source chosen for culturing may not necessarily be the same type of water used for testing. However, whichever water is chosen for culturing or testing, it must be tested to ensure that good survival and reproduction of the organisms are possible and that consistency is achievable. Before any water is used, it should be tested for possible contamination by

pesticides, heavy metals, major anions and cations, total organic carbon, suspended solids, or any other suspected contaminants. The water quality should ensure adequate survival, growth, and reproduction and it should be from a consistent source to provide constant quality during any given testing period [2]. The fathead has been very commonly used as a baitfish and, more recently, has emerged in the aquarium trade as the rosy-red minnow. This color morph was discovered in several arkansas breeding farms in 1985. Both sexes of this strain have a rosy-golden body and fins and may express dark splotches of wild-type fathead coloration. It is worth mentioning that they are sold in pet shops primarily as feeder fish. They can also be used in home aquariums as pets [3]. This species is also important as a biological model in aquatic toxicology studies. Because of its relative hardiness and large number of offspring produced, EPA guidelines outline its use for the evaluation of acute and chronic toxicity of samples or chemical species in vertebrate animals.

Chemical modelling techniques are based on the premise that the structure of a compound determines all its properties. The study of the type of chemical structure of a foreign substance which will interact with a living system and produce a well-defined biological endpoint is commonly referred to as quantitative structure-retention relationships QSRR [4, 5]. The use of QSRR for toxicity estimation of new chemicals or regulatory toxicological assessment is increasing, especially in aquatic toxicology. Alternatively, quantitative retention relationships QRRR represent other kind of modelling techniques in which chromatographic retention parameters are used as descriptor and/or predictor variables of a given biological response of chemicals. QSRR models which use retention factors (log k) obtain conventional RP-HPLC, micellar liquid chromatography (MLC) and biopartitioning micellar chromatography (BMC) which will be reported [6–10].

The aim of the present study is the estimation of optimal descriptors ability calculated by linear regression (the partial least squares (PLS) and non-linear regressions (the kernel partial least squares (KPLS) and Levenberg- Marquardt artificial neural network (L-M ANN) in QSRR analysis of logarithm of the retention factor in BMC (log k) for toxicity to fathead minnows of anilines and phenols. The stability and predictive power of these models were validated using Leave-Group-Out Cross-Validation (LGO CV) and external test set. This is the first research on the QSAR which uses GA-PLS for the chromatographic retention of ecotoxicity of anilines and phenols**.**

Experimental

Computer hardware and software

A pentium IV personal computer (CPU at 3.06 GHz) with the Windows XP operating system was used. The structures of the compounds were drawn using Hyper Chem version 7.0. All molecules were preoptimized using molecular mechanics AM1 method in the HyperChem program. The output

files were exported from dragon for generating descriptors which were developed by *Todeschini* et al [11]. The GA-PLS, GA-KPLS, L-M ANN, cross validation and other calculations were performed in MATLAB (Version 7.0, Math works, Inc).

Data set

The 65 phenols and anilines for which experimental chromatographic retention (log k) values to fathead minnows were available [12] were used. The name of studied compounds and their experimental log k values for training and test sets are shown in Table 1 and Table 2. These data were obtained by biopartitioning micellar chromatography. An Agilent 1100 chromatograph with a quaternary pump and an UV-vis detector (Variable wavelength detector) was employed. It is equipped with a column thermostat with 9 μL extra-column volume for preheating mobile phase prior to the column and an autosampler with a 20 μL loop. All the assays were carried out at 25 °C. Data acquisition and processing were performed by means of an HP Vectra XM computer (Amsterdam, the netherlands) equipped with HP-Chemstation software (A.07.01 [682] ©HP 1999). Two Kromasil C18 columns (5 μm, 150 mm×4.6 mm i.d.; Scharlab S.L., Barcelona, spain) and (5 μm, 50 mm×4.6 mm i.d.; scharlab) were used. The mobile phase flow rate was 1.0 or 1.5 $mLmin^{-1}$ for the 150 mm and 50 mm column length, respectively. The detection was performed in UV at 254 nm for acetanilide, antipyrine and propiophenone (Reference compounds), and 240 nm for phenols and anilines.

Determination of molecular descriptors

Molecular descriptors are defined as numerical characteristics associated with chemical structures. The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number applied to correlate physical properties. The Dragon software was used to calculate the descriptors in this research and a total of molecular descriptors, from 18 different types of theoretical descriptors, was calculated for each molecule. Since the values of many descriptors are related to the bonds length and bonds angles etc., the chemical structure of every molecule must be optimized before calculating its molecular descriptors. For this reason, the chemical structure of the 65 studied molecules was drawn using hyperchem software and saved with the HIN extension. To optimize the geometry of these molecules, the AM1 geometrical optimization was applied. After optimizing the chemical structures of all compounds, the molecular descriptors were calculated using dragon. A wide variety of descriptors have been reported in the literature, and used in QSRR analysis.

Table 1. The compounds and log retention factor for calibration and prediction sets

Emtry	Compounds calibration set	log k
1	2,6-Dinitrophenol	0.793
2	2,4-Dinitrophenol	0.943
3	4,6-Dinitro-2-methylphenol	1.004
4	2,5-Dinitrophenol	1.017
5	3-Hydroxyphenol	1.044
6	2-Nitrophenol	1.2
7	Phenol	1.245
8	4-Nitroaniline	1.257
9	2,3,6-Trichlorophenol	1.349
10	2,3,5,6-Tetrachlorophenol	1.352
11	Pentabromophenol	1.354
12	4-Nitrophenol	1.378
13	2,3,4,6-Tetrachlorophenol	1.352
14	4-Mehtylphenol	1.354
15	2,4,6-Tribromophenol	1.378
16	3-Nitrophenol	1.394
17	2,4,6-Triiodophenol	1.411
18	2,6-Dichlorophenol	1.417
19	2,4-Dinitroaniline	1.448
20	2,4,6-Trichlorophenol	1.459
21	2-Chloro-4-nitroaniline	1.46
22	4-Chlorophenol	1.476
23	4-Ethylphenol	1.477
24	2,4-Dimethylphenol	1.496
25	2-Chloro-4-methylaniline	1.529
26	4-Chloro-3-methylphenol	1.552
27	2,3,6-Trimethylphenol	1.567
28	N,N-Dimethylaniline	1.576
29	Pentafluoroaniline	1.6
30	2,3,4-Trichloroaniline	1.626
31	N,N-Dimethylaniline	1.642
32	Pentafluoroaniline	1.668

33	2,3,4-Trichloroaniline	1.67
34	4-Phenoxiphenol	1.683
35	2-Phenylphenol	1.709
36	2,3,5-Trichlorophenol	1.757
37	3,5-Dichlorophenol	1.765
38	3,4,5-Trichlorophenol	1.817
39	2,3,4,5-Tetrachlorophenol	1.822
40	2,6-Diisopropylaniline	1.896
41	2,6-Diisopropylphenol	1.986
42	2,6-Di(tert)butil-4-methylphenol	2.351
43	2,6-Dimethoxiphenol	0.979
44	4-Methylaniline	1.227
45	N-Methylaniline	1.351
46	4-Chloroaniline	1.408
47	2-Chloroaniline	1.459
48	2-Chlorophenol	1.485
49	3,4-Dichloroaniline	1.563
50	2,4,6-Trimethylphenol	1.629
51	2,4-Dichlorophenol	1.673
52	4-Butylaniline	1.713
53	2,4,5-Trichlorophenol	1.778
54	2,3,5,6-Tetrachloroaniline	1.832
55	Nonylphenol	2.186

Genetic algorithm for descriptor selection

In QSRR studies, after calculating the molecular descriptors from optimized chemical structures of all the components available in the data set, the problem is to find an equation that can predict the desired property with the least number of variables as well as highest accuracy. In other words, the problem is to find a subset of variables (Most statistically effective molecular descriptors for the log k) from all the available variables (All molecular descriptors) that can predict log k with the minimum error in comparison to the experimental data. A generally accepted method for this problem is the genetic algorithm based linear and non linear regressions (GA-PLS and GA-KPLS). In these methods,

the genetic algorithm is applied for the selection of the best subset of variables with respect to an objective function.

Table 2. The data set and log k for test set

Entry	Compounds	log k
1	Aniline	0.988
2	4-Methoxyphenol	1.158
3	3-Methoxyphenol	1.266
4	Pentachlorophenol	1.384
5	4-Ethylaniline	1.446
6	2-Methylphenol	1.465
7	4-Ethoxy-2-nitroaniline	1.526
8	2,6-Dichloro-4-aniline	1.576
9	4-Propylphenol	1.669
10	4-Tert-butylphenol	1.748
11	4-Hexyloxyaniline	1.785
12	4-Tert-pentylphenol	1.841
13	4-Octylaniline	2.043

GA is a stochastic optimization method that has been inspired by evolutionary principles. The distinctive aspect of GA is that it investigates many possible solutions simultaneously, each of which explores different regions in parameter space. GA has been applied as an optimization technique in several scientific fields [13, 14]. In GA for variable selection, the chromosome and its fitness in the species represent a set of variables and predictivity of the derived QSRR model, respectively. GA consists of three basic steps: (I) an initial population of chromosomes is created. The number of the population is dependent on the dimensions of application problems. A binary bit string represents each chromosome. Bit “1” denotes a selection of the corresponding variable, and bit “0” denotes a non selection. The values of a binary bit are determined in a random way (Probability of initial variable selection). (II) A fitness of each chromosome in the population is evaluated by predictivity of the QSRR model derived from the binary bit string. (III) The population of chromosomes in the next generation is reproduced. The third step can be divided into three operations: selection, crossover, and mutation. The application probability of these operators was varied linearly with a generation renewal. For a typical run, the evolution of the generation was stopped when 90% of the generations had taken the same fitness. In this paper, size of the population is 30 chromosomes, the probability of initial variable selection is 5:V (V is the number of independent variables), crossover

is multi point, the probability of crossover is 0.5, mutation is multi point, the probability of mutation is 0.01 and the number of evolution generations is 1000. For GA-PLS and GA-KPLS programs, 3000 runs were performed.

Data pre-processing

Each set of the calculated descriptors was collected in a separate data matrix D_i with a dimension of (m×n) where m and n are the number of molecules and the number of descriptors, respectively. Grouping of descriptors was based on the classification achieved by Dragon software. In each group, the calculated descriptors were searched for constant or near constant values for all molecules and those detected were removed. Before applying the analysis methods and due to the quality of data, a previous treatment of the data is required. Scaling and centering can be considered as the pre-processing methods which are needed before performing the regression methods as combined with FE. The results of projection methods depend on the normalization of the data. Descriptors with small absolute values have a small contribution to overall variances; this biases towards other descriptors with higher values. With appropriate scaling, equal weights are assigned to each descriptor so that the important variables in the model can be focused. In order to give all variables the same importance, they are standardized to unit variance and zero mean (Autoscaling).

Nonlinear model

Artificial neural network

A three-layer back propagation artificial neural network ANN with a sigmoid transfer function was used in the investigation of feature sets. The descriptors from the calibration set were used for the model generation whereas the descriptors from the prediction set were used to stop the overtraining of network. Moreover, the descriptors from the test set were used to verify the predictivity of the model. Before training the networks, the input and output values were normalized with auto-scaling of all data [15, 16]. The goal of training the network is to minimize the output errors by changing the weights between the layers.

$$\Delta W_{ij,n} = F_n + \alpha \Delta W_{ij,n-1} \quad (1)$$

In this, ΔW_{ij} is the change in the weight factor for each network node, α is the momentum factor, and F is a weight update function, which indicates how weights are changed during the learning process. The weights of hidden layer were optimized using the Levenberg-Marquardt algorithm, a second derivative optimization method [17].

Levenberg-Marquardt Algorithm

In Levenberg-Marquardt algorithm, the update function, F_n, is calculated using the following equations.

$$F_0 = -g_0 \tag{2}$$

$$g = J^T e \tag{3}$$

$$F_n = -[J^T \times J + \mu I]^{-1} \times J^T \times e \tag{4}$$

Where g is gradient and J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights, and e is a vector of network errors. The parameter μ is multiplied by some factor (λ) whenever a step would result in an increased e and when a step reduces e, μ is divided by λ [18].

Results and discussion

Linear model

Results of the GA-PLS model

The best model is selected on the basis of the highest square correlation coefficient leave-group-out cross validation (R^2), the least root mean squares error (*RMSE*) and relative error (*RE*). These parameters are probably the most popular measures of how well a model fits the data. The best GA-PLS model contains 13 selected descriptors in 5 latent variables space. These descriptors were obtained constitutional descriptors [sum of conventional bond orders (H-depleted) (SCBO)], topological descriptors (Balaban-type index from polarizability weighted distance matrix (Jhetp) and eccentricity (ECC)), 2D autocorrelations (Broto-Moreau autocorrelation of a topological structure-lag 6 / weighted by atomic Sanderson electronegativities (ATS6e), Burden eigenvalues (lowest eigenvalue n.1 of Burden matrix / weighted by atomic Sanderson electronegativities (BELe1), RDF descriptors (Radial Distribution Function-4.5 / unweighted (RDF045u), Radial Distribution Function-11.5 / unweighted (RDF115u), Radial Distribution Function-6.5 / weighted by atomic masses (RDF065m) and Radial Distribution Function-12.5 / weighted by atomic masses (RDF125m), WHIM descriptors (1st component symmetry directional WHIM index / weighted by atomic van der Waals volumes (G1v)), functional group counts (number of total tertiary C(sp^3) (nCt) and number of aromatic C(sp^2) (nCar)) and quantum descriptors [lowest unoccupied molecular orbital (LUMO)]. The R^2 and mean *RE* for training and test sets were (0.864, 0.751) and (8.04, 18.84), respectively. The predicted values of log k are plotted against the experimental values for training and test sets in

Figure 1. Generally, the number of components (Latent variables) is less than the number of independent variables in PLS analysis. The PLS model uses higher number of descriptors that allow the model to extract better structural information from descriptors in order to result in a lower prediction error.

Nonlinear model

Results of the GA-KPLS model

In this paper a radial basis kernel function, $k(x,y)= \exp(||x-y||^2/c)$, was selected as the kernel function with $c = rm\sigma^2$ where r is a constant that can be determined by considering the process to be predicted (Here r was set to be 1), m is the dimension of the input space and σ^2 is the variance of the data [19, 20]. It means that the value of c depends on the system under the study. The 10 descriptors in 5 latent variables space chosen by GA-KPLS feature selection methods were contained. These descriptors were obtained geometrical descriptors (gravitational index G2 (bond-restricted) (G2), spherosity (SPH) and HOMA total (HOMT)), RDF descriptors (Radial Distribution Function - 3.0 / weighted by atomic masses (RDF030m)), 3D-MoRSE descriptors (3D-MoRSE-signal 10 / unweighted (Mor10u) and 3D-MoRSE-signal 18 / weighted by atomic masses (Mor18m)), GETAWAY descriptors (Leverage-weighted autocorrelation of lag 2 / unweighted (HATS2u) and H autocorrelation of lag 7 / weighted by atomic masses (H7m)), charge descriptors (relative positive charge (RPCG)) and quantum descriptors (Dipole moment (μ)). The R^2 and mean *RE* for training and test sets were (0.827, 0.709) and (9.43, 20.82), respectively. It can be seen from these results that statistical results for GA-PLS model are superior to GA-KPLS method. Figure 2 shows the plot of the GA-KPLS predicted versus experimental values for log k of all of the molecules in the data set.

Results of the L-M ANN model

With the aim of improving the predictive performance of nonlinear QSRR model, L-M ANN modeling was performed. The networks were generated using the thirteen descriptors appearing in the GA-PLS models as their inputs and log k as their output. For ANN generation, data set was separated into three groups: calibration and prediction (Training) and test sets. All molecules were randomly placed

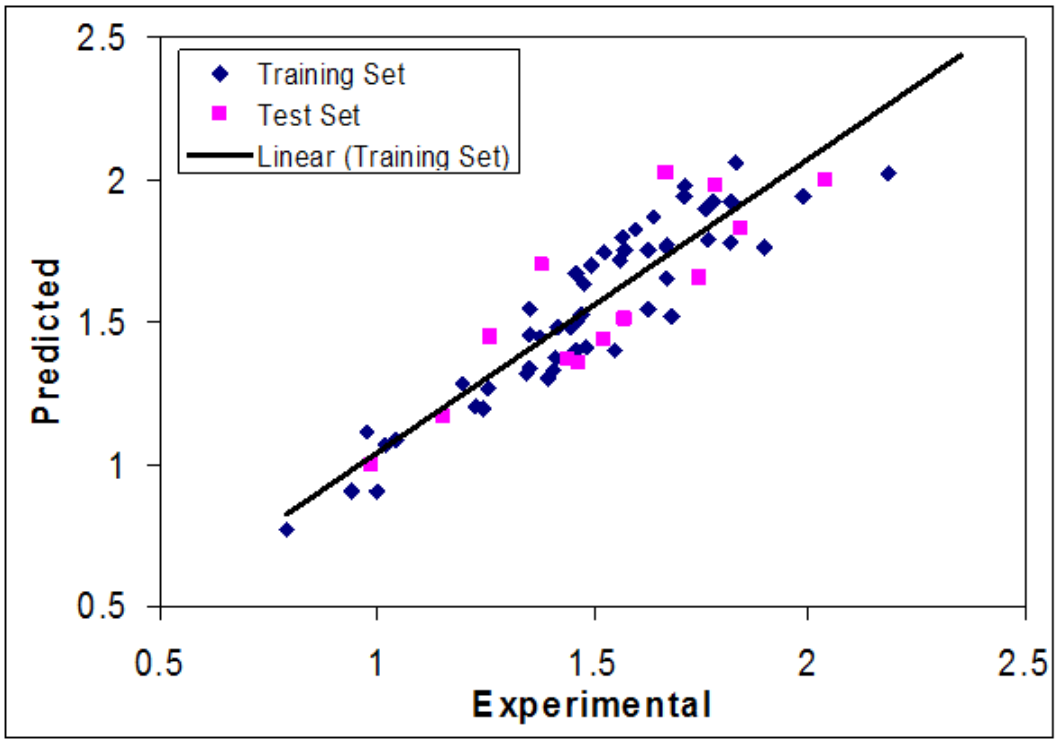


Figure 1. Plots of predicted retention time against the experimental values by GA-PLS model

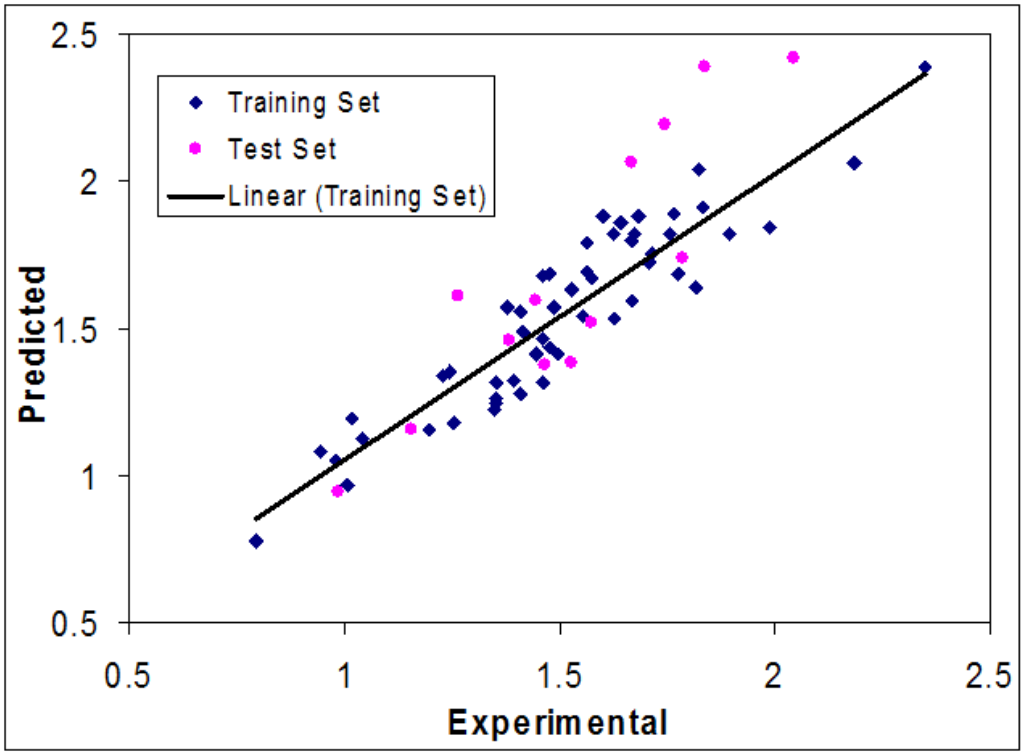


Figure 2. Plots of predicted log K versus the experimental values by GA-KPLS model

in these sets. A three-layer network with a sigmoid transfer function was designed for each ANN. Before training the networks, the input and output values were normalized between -1 and 1. The network was then trained using the training set by the back propagation strategy for optimization of the weights and bias values. The proper number of nodes in the hidden layer was determined by training the network with different number of nodes in the hidden layer. The root-mean-square error (RMSE) value measures how good the outputs are in comparison with the target values. It should be noted that for evaluating the overfitting, the training of the network for the prediction of log k must stop when the RMSE of the prediction set begins to increase while RMSE of calibration set continues to decrease. Therefore, training of the network was stopped when overtraining began. All of the above mentioned steps were carried out using basic back propagation, conjugate gradient and Levenberge-Marquardt weight update functions. It was realized that the RMSE for the training and test sets are minimum when three neurons were selected in the hidden layer. Finally, the number of iterations was optimized with the optimum values for the variables. It was realized that after 16 iterations, the *RMSE* for prediction set were minimum. The mean relative error and R^2 for calibration, prediction and test sets were (0.959, 0.942, 0.903) and (4.49, 5.34, 7.12), respectively. Comparison between these values and other statistical parameter reveals the superiority of the L-M ANN model over other model. The key strength of neural networks, unlike regression analysis, is their ability to flexible mapping of the selected features by manipulating their functional dependence implicitly. The statistical parameters reveal the high predictive ability of L-M ANN model. The whole of these data clearly displays a significant improvement of the QSRR model consequent to nonlinear statistical treatment. Plot of predicted log k versus experimental log k values by L-M ANN for training and test sets are shown in Figure **3a** and Figure **3b**. Obviously, there is a close agreement between the experimental and predicted log k and the data represent a very low scattering around a straight line with respective slope and intercept close to one and zero. As can be seen in this section, the L-M ANN is more reproducible than other models for modeling the log k of compounds.

Model validation and statistical parameters

The accuracy of proposed models was illustrated using the evaluation techniques such as leave group out cross-validation (LGO-CV) procedure and validation through an external test set. In addition, chance correlation procedure is a useful method for investigating the accuracy of the resulted model by which one can make sure if the results were obtained by chance or not.

Cross validation is a popular technique used to explore the reliability of statistical models. Based on this technique, a number of modified data sets are created by deleting in each case one or a small

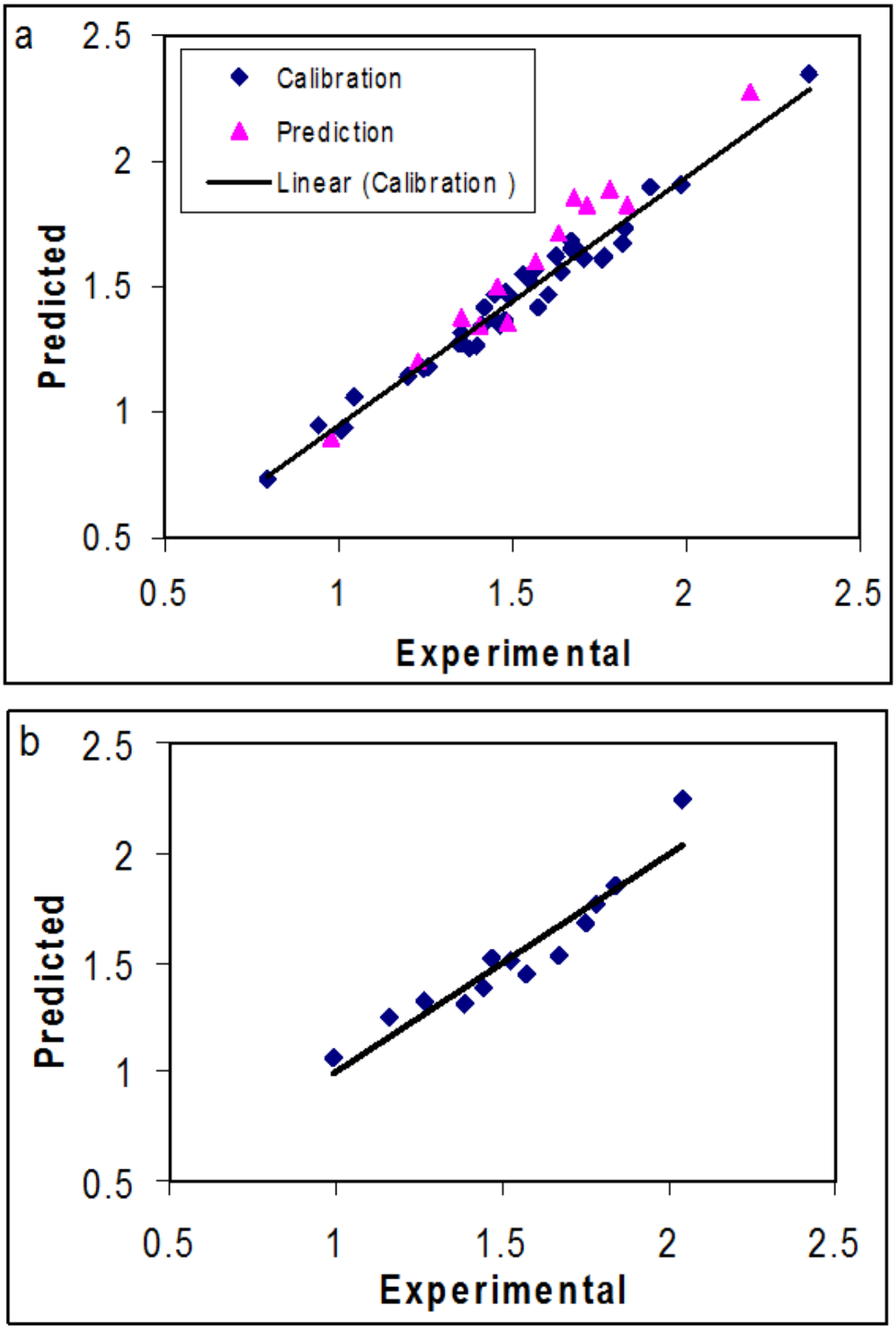


Figure 3. Plot of predicted log k obtained by L-M ANN against the experimental values a) for training set and b) test set

group (Leave-some-out) of objects. For each data set, an input–output model is developed, based on the utilized modeling technique. Each model is evaluated, by measuring its accuracy in predicting the responses of the remaining data (The ones or group data that have not been utilized in the development of the model). In particular, the LGO-CV procedure was utilized in this study. A QSRR model was then constructed on the basis of this reduced data set and subsequently used to predict the removed data. This procedure was repeated until a complete set of predicted was obtained. The statistical significance of the screened model was judged by the correlation coefficient (R^2). The predictive ability was evaluated by the cross validation coefficient (R^2). The accuracy of cross validation results is extensively accepted in the literature considering the R^2 value. In this sense, a high value of the statistical characteristic ($R^2 > 0.5$) is considered as proof of the high predictive ability of the model.

The data set should be divided into three new sub-data sets, one for calibration and prediction (Training), and the other one for testing. The calibration set was used for model generation. The prediction set was applied deal with overfitting of the network, whereas test set which its molecules have no role in model building was used for the evaluation of the predictive ability of the models for external set [21].

In the other hand by means of training set, the best model is found and then, the prediction power of it is checked by test set, as an external data set. In this work, 60% of the database was used for calibration set, 20% for prediction set and 20% for test set [22], randomly (In each running program, from all 65 components, 39 components are in calibration set, 13 components are in prediction set and 13 components are in test set).

The result clearly displays a significant improvement of the QSRR model consequent to non-linear statistical treatment and a substantial independence of model prediction from the structure of the test molecule. In the above analysis, the descriptive power of a given model has been measured by its ability to predict log k of unknown compounds.

For the constructed models, two general statistical parameters were selected to evaluate the prediction ability of the model for log k values. For this case, the predicted log k of each sample in the prediction step was compared with the experimental log k. The root mean square error of prediction (*RMSE*) is a measurement of the average difference between predicted and experimental values, at the prediction stage. The *RMSE* can be interpreted as the average prediction error, expressed in the same units as the original response values. The *RMSEP* was obtained using the following formula:

$$RMSE = [\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2]^{\frac{1}{2}} \quad (5)$$

The second statistical parameter was the relative error of prediction (*RE*) that shows the predictive ability of each component, and is calculated as:

$$RE(\%) = 100 \times \left[\frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{y_i} - y_i)}{y_i}\right] \quad (6)$$

Where y_i is the experimental log k value of the anilines and phenols in the sample i, $\hat{y}_i$ represents the predicted log k value in the sample i, $\bar{y}$ is the mean of experimental log k values in the prediction set and n is the total number of samples used in the test set [23, 24].

Conclusion

The GA-PLS, GA-KPLS and L-M ANN models was applied for the prediction of the log k values of ecotoxicity of anilines and phenols. High correlation coefficients and low prediction errors confirmed the good predictability of models. All methods seemed to be useful, although a comparison between these methods revealed the slight superiority of the L-M ANN over other models. Application of the developed model to a testing set of 13 compounds demonstrates that the new model is reliable with good predictive accuracy and simple formulation. The QSRR procedure allowed us to achieve a precise and relatively fast method for determination of log k of different series of these compounds to predict with sufficient accuracy the log k of new substituted compounds.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1]. Sowers A.D., Gaworecki K.M., Mills M.A., Roberts A.P., Klaine S.J. *Aquat. Toxicol.*, 2009, **95**:173

[2]. Aruoja V., Sihtmäe M., Kahru A., Dubourguier H. *Toxicol. Lett.*, 2009, **189**:192

[3]. Al-Awadhi J.M., Al-Awadhi A.A. *J. Arid Environ.*, 2009, **73**:987

[4]. Duchowicz P.R., Giraudo M.A., Castro E.A., Pomilio A.B, *Chemom. Intell. Lab. Syst.*, 2011, **107**:384

[5]. Goodarzi M., Chen T., Freitas M.P. *Chemom. Intell. Lab. Syst.*, 2010, **104**:260

[6]. Liu T., Nicholls I.A., Oberg T., *Anal. Chim. Acta*., 2011, **702**:37

[7]. Kaliszan R., Wiczling P., Markuszewski M.J., Al-Haj M.A. *J. Chromatogr. A*., 2011, **1218**:5120

[8]. Flieger J. *J. Chromatogr.* A., 2010, **1217**:540

[9]. Bucinski A., Wnuk M., Gorynski K., Giza A., Kochanczyk J., Nowaczyk A., Baczek T., Nasal A. *J Pharm Biomed Anal*., 2009, **50**:591

[10]. Lammerhofer M. *J. Chromatogr. A*., 2010, **1217**:814

[11]. Shahpar M., Esmaeilpoor Sh. *Asian J. Nano. Mat.,* 2018*,* **1***:1*

[12]. Ren S., Frymier P.D., Schultz T.W., *Schultz, Ecotox. Environ. Safe*, 2003, **55**:86

[13]. Sarıpınar E., Geçen N., Şahin K., Yanmaz E., *Eur. J. Med. Chem*., 2010, **45**:4157

[14]. Sagrado S., Cronin M.T.D., *Anal. Chim. Acta.,* 2008, **609**:169

[15]. Hernández-Caraballo E.A., Rivas F., Pérez A.G., Marcó-Parra L.M. *Anal. Chim. Acta.,* 2005, **533**:161

[16]. Gupta V.K., Khani H., Ahmadi-Roudi B., Mirakhorli Sh., Fereyduni E., Agarwal Sh. *Talanta,* 2011, **83**:1014

[17]. Bolanča T., Cerjan-Stefanović S., Regelja M., Regelja H., Lončarić S. *J. Chromatogr. A,* 2005, **1085**:74

[18]. Chamjangali M.A., Beglari M., Bagherian G. *J. Mol. Graphics Modell*., 2007, **26**:360

[19]. Shahpar M., Esmaeilpoor Sh. *Asian J. Green Chem*., 2017, **129**:116

[20]. Jia R., Mao Zh., Chang Y., Zhang Sh, *Chemom. Intell. Lab. Syst.,* 2010, **100**:91

[21]. Jalali-Heravi M., Kyani A., *Eur. J. Med. Chem.,* 2007*,* **42**:649

[22]. Noorizadeh H., Noorizadeh M*., Med Chem Res*., 2013, **11**:5442

[23]. Chen H., *Anal. Chim. Acta*., 2008, **609**:24

[24]. Shahpar M., Esmaeilpoor Sh*. Chem. Method*., 2017, **2**:105

How to cite this manuscript: Mehrdad Shahpar*, Sharmin Esmaeilpoor. Quantitative structure-retention relationships applied to chromatographic retention of ecotoxicity of anilines and phenols. *Asian Journal of Green Chemistry*, 2018, 2, 144-159. DOI: 10.22631/ajgc.2018.100313.1023