



QSAR Prediction of Half-Life, Nondimensional Effective Degradation Rate Constant and Effective Péclet Number of Volatile Organic Compounds

Mohammad Hossein Fatemi*, Afsane Heidari, Hanieh Malekzadeh

Chemometrics Laboratory, Faculty of Chemistry, University of Mazandaran, Babolsar,

* Corresponding author E-mail address: mhfatemi@umz.ac.ir

Received 28 November 2011|Received in revised form 12 March 2012|Accepted 17 March 2012

Abstract: In this work some quantitative structure activity relationship models were developed for prediction of three bioenvironmental parameters of 28 volatile organic compounds, which are used in assessing the behavior of pollutants in soil. These parameters are; half-life, non dimensional effective degradation rate constant and effective Péclet number in two type of soil. The most effective descriptors which were selected by stepwise multiple linear regression method reflect some information about structure and polarity of molecules. Linear and non-linear models were established by multiple linear regressions and least-squares support vector machine respectively. The statistical parameters of developed models demonstrated that both of them are satisfactory. However the models established with least-squares support vector machine are more stable and reliable than multiple linear regressions model.

Key words: Volatile organic compounds, Molecular Descriptor, degradation rate constant, Quantitative structure properties relationship

©2012 Published by University of Mazandaran. All rights reserved.

1. Introduction

Volatile organic compounds (VOCs) are one of the most important environmental concerns. They may enter to the soil from a variety of sources and may contaminate the ground water [1-3]. Thus the study of behavior of VOCs in soils has attracted considerable attention for the evaluation of their possible risks to

humans and the environment. Modeling of the environmental fate and transport of VOCs has been carried out by number of researchers in recent years [4-7]. They used different indices and parameters to screen the VOCs in environment. For example, half-life ($\log T_{1/2}$) and effective parameters can be used as criteria for assess the behavior of VOCs in soils. Effective parameters are essential properties of

homogeneous zone, which describe the flux processes in homogeneous zone of soil. One effective parameter is non dimensional effective degradation rate constant, which is defined as:

$$\gamma_{\text{eff-nd}} = \gamma_{\text{eff}}L / V_{\text{eff}} \quad (1)$$

where γ_{eff} is effective degradation rate constant, V_{eff} is effective velocity, and L is characteristic soil length. This quantity represents effects of degradation losses within the soil column, and volatilization losses at the soil surface as well. One another effective parameter is the effective Péclet number. It is a dimensionless number relevant in the study of transport phenomena in fluid flows and is defined to be the ratio of the rate of advection of a physical quantity by the flow to the rate of diffusion of the same quantity. For mass diffusion by using method of moments, it is defined as:

$$P_{\text{eff}} = V_{\text{eff}}L / D_{\text{eff}} \quad (2)$$

where D_{eff} is the effective mass diffusion coefficient [7]. Diffusion increases with decreasing Péclet number and reduces biodegradation losses, consequently leading to more mass being available for leaching. In this case; the probability of contamination of ground water will be increased. These parameters have been computed using temporal moment analysis (TMA), which is a non-parametric, statistical procedure that also provides useful information on solute transport characteristics in soils [8].

Since the experimental determination of these parameters needs time-consuming experiments and

complex calculations therefore they have been measured for small fraction of VOCs. Moreover different compounds will often have different Péclet numbers and nondimensional effective degradation rate constants in the same flow condition. Whereas, the values of these parameters are affected by structural features of compounds therefore, it is possible to calculate them by quantitative structure properties relationships (QSAR) approaches. In this method relevant molecular structural parameters are correlated mathematically to interested properties or activities of chemicals [9].

There are some reports about QSAR modeling of half-life of environmental chemicals [10-12]. In a former paper, Gramatica et al. developed principle component analysis (PCA) model for estimation of degradation half-life data of 250 persistent organic pollutants (POPs) [10]. Their 6-parameters model has the statistics of square correlation coefficient of $R^2=0.85$ and root mean square error of $RMSE=0.76$ for training set and $R^2=0.79$ and $RMSE=0.78$ for prediction set. Also Papa et al. established a multiple linear regression (MLR) in combination with the genetic algorithm variable subset selection procedure to predict the tropospheric half-life of 166 VOCs by three parameters model. Their best QSAR models showed RMSEs in range of 0.41-0.67 and 0.84-0.65 for training and prediction data sets, respectively. [11]. In a recent work, Luan et al. reported a QSAR model for prediction of the mean and the maximum atmospheric degradation half-life values of persistent organic pollutants by using nonlinear gene expression programming (GEP) [12].

The examination of their model on their test set provided the statistic of R^2 ; 0.80 and 0.81 and RMSE of 0.448 and 0.426 for the mean and maximum half-life values, respectively.

Recently Parashar et al. used a sensitivity analysis to study the importance of different physical/chemical processes on volatilization, degradation, and leaching losses of some VOCs in two typical types of soil [7]. They calculated the effective nondimensional parameters by TMA technique and then the sensitivity of losses was examined with respect to their non dimensional parameters.

Since our literature survey indicate that there are not any published report about QSAR prediction of effective Péclet numbers and effective nondimensional degradation rate constants of VOCs, therefore we decided to develop some QSAR models to predict these parameters and half-life of some VOCs.

2. Methods

2.1. Data set

The experimental data shown in Table 1 are half-life, effective Péclet numbers and effective nondimensional degradation rate constants of 28 volatile organic compounds that were taken from Parashar's paper [7]. In this table half-life data represented in logarithmic units due to their abroad range. VOCs were evaluated under the initial condition of a uniform concentration of organic chemical located at a certain depth in sandy and clayey soil. The compounds were assumed to move by vapor or liquid diffusion and by mass flow while undergoing first-order degradation and linear

equilibrium adsorption. Compounds in data set were sorted according to the values of each parameter and test set was selected from this set by desirable distance from each other (y -ranking method). In each case the training set consists of 22 compounds while test set has 6 members.

2.2. Descriptors generation and selection

To obtain a QSAR model, the compounds must represented by the molecular descriptors that encode the structural features of chemicals.

Therefore, the chemical structures of molecules were drawn by using Hyperchem package (V.7) [13] and optimized by the AM1 semi-empirical method. Then a more precise optimization was done with semi-empirical AM1 method by MOPAC package (Ver. 6) [14]. Dragon (Ver.3) [15] and CODESSA softwares [16] were used to calculated molecular descriptors by using the Hyperchem and MOPAC output files. All calculated descriptors were divided to five classes: constitutional, topological, geometrical, electronic and quantum chemical descriptors. Constitutional descriptors reflect only the molecular composition of the compound.

Topological descriptors describe the atomic connectivity in a molecule. Geometrical descriptors are calculated from three-dimension atomic coordinates of the molecule. Electronic and quantum chemical descriptors reflect characteristics of the charge distribution in the molecule and add important information to the conventional descriptors.

Table 1: Data set and corresponding observed MLR and LS-SVM predicted values of P_{eff} and $\gamma_{\text{eff-nd}}$ in sandy and clayey soil and $\log T_{1/2}$

Chemical name	$\log T_{1/2}$ (exp.)	$\log T_{1/2}$ (pred.) MLR	$\log T_{1/2}$ (pred.) LS- SVM	P_{eff} (exp.)	Sandy soil					Clayey soil						
					P_{eff} (pred.) MLR	P_{eff} (pred.) LS- SVM	$\gamma_{\text{eff-nd}}$ (exp.)	$\gamma_{\text{eff-nd}}$ (pred.) MLR	$\gamma_{\text{eff-nd}}$ (pred.) LS- SVM	P_{eff} (exp.)	P_{eff} (pred.) MLR	P_{eff} (pred.) LS- SVM	$\gamma_{\text{eff-nd}}$ (exp.)	$\gamma_{\text{eff-nd}}$ (pred.) MLR	$\gamma_{\text{eff-nd}}$ (pred.) LS- SVM	
					1	1,1,1-Trichloroethane	2.563	2.547	2.694	1.670	2.273	1.893	0.014	-0.021	0.012	0.370
2	1,1-Dichloroethane	1.653*	2.197	2.297	3.260	2.889	2.822	0.054	-0.017	0.008	0.990*	1.701	1.346	0.675	0.275	0.295
3	1,2-Dichloroethane	1.954	1.820	1.958	5.640	4.301	5.141	0.016	-0.024	0.004	3.470	1.512	1.688	0.178*	0.421	0.335
4	2,4-Dichlorophenol	2.204	2.640	2.418	6.550*	7.968	5.869	0.119	0.101	0.159	8.220*	7.815	7.506	1.630	1.551	1.550
5	2-Chloronaphthalene	3.158	3.184	3.115	6.290	4.794	5.357	0.033	0.305	0.169	6.130	5.095	6.029	0.460	0.371	0.413
6	Acrolein	2.477	2.244	2.310	7.730	6.402	7.427	0.002*	0.108	0.040	8.010	7.734	8.163	0.014*	-0.549	0.075
7	Benzene	2.563	2.323	2.653	3.370	4.254	3.860	0.011	0.221	0.105	1.070	1.650	1.134	0.135	1.116	0.191
8	Bromoethane	1.477*	1.437	1.767	2.720	3.024	2.930	0.030	-0.025	0.003	0.800	1.579	1.299	0.289	-0.222	0.128
9	Chlordane	2.000	2.409	2.029	6.390	6.085	6.253	0.204*	0.298	0.226	7.160	7.177	7.056	2.860*	1.301	1.066
10	Chloroethane	1.477	1.728	1.619	2.150	2.876	2.753	0.052	-0.035	0.002	0.520	1.011	0.794	0.614*	0.259	0.259
11	Chloroethene	1.477	1.588	1.571	1.110	1.815	0.846	0.022	0.048	0.020	0.230	0.365	0.637	0.175	-0.082	0.166
12	Chloroform	2.000	1.818	1.987	3.980*	2.734	3.280	0.017	-0.040	0.001	1.510	1.742	1.789	0.201	0.480	0.251
13	Chloromethane	2.079	1.970	1.949	2.470	2.505	2.435	0.006	-0.060	-0.004	0.700	-0.007	0.911	0.058	0.699	0.094
14	DBCP	3.000	2.932	2.962	6.610	7.829	6.625	0.006	-0.006	0.019	7.580	8.842	7.507	0.077	-0.845	0.056
15	Dichlorodifluoromethane	4.000*	3.840	3.631	0.120*	1.128	1.893	0.001	-0.054	0.006	0.020*	-0.155	0.250	0.007	0.155	0.085
16	Dichloromethane	2.000*	2.763	2.913	4.130	3.343	4.277	0.011	-0.035	-0.003	1.680	0.334	1.435	0.109	0.724	0.126
17	EPTC	1.477	1.836	1.747	6.210	6.037	6.126	0.404	0.419	0.355	6.160	6.729	6.126	6.030	5.097	5.908
18	Ethylene dibromide	3.563	3.845	3.628	6.190*	4.401	5.236	0.001	-0.027	0.004	4.910*	2.825	0.951	0.008	-0.662	0.100
19	Heptachlor	3.342	3.212	3.342	6.210	5.934	6.214	0.128	0.145	0.166	5.850	5.156	6.002	1.800	2.249	1.833
20	Hexachlorocyclohexane	2.860	3.003	2.835	6.220	7.611	6.698	0.134*	0.101	0.110	5.950	6.723	6.173	1.880	1.811	1.846
21	Methyl ethyl ketone	2.000	1.758	1.669	7.390*	6.393	7.472	0.009	0.114	0.067	8.600	8.790	8.372	0.092	-0.197	0.146
22	Methyl isobutyl ketone	2.000*	1.683	1.646	6.970	5.610	6.516	0.014*	0.129	0.092	7.700	5.094	7.300	0.160	1.253	0.207
23	Phorate	1.914	2.231	2.112	6.260	7.149	6.207	0.340	0.456	0.387	6.620*	6.224	5.528	5.010*	3.206	3.530
24	Toluene	0.699	0.979	0.732	2.270	3.054	2.486	0.974*	0.497	0.374	0.980	1.982	1.054	4.860	4.050	4.737
25	Toxaphene	3.563	2.912	3.151	6.680	5.875	6.541	0.007*	0.122	0.209	8.970	8.167	8.797	0.100*	0.894	0.782
26	Trichloroethylene	2.863*	3.749	3.727	2.470*	3.375	3.673	0.009	0.039	0.017	0.630*	2.094	0.325	0.113	0.169	0.292
27	Trillate	2.000	1.743	2.044	5.010	5.593	5.834	1.550	1.396	1.386	0.410	0.948	0.474	0.021	-0.227	0.114
28	Xylene	2.041	1.770	1.886	3.120	3.237	3.246	0.116	0.044	0.063	0.780	1.432	0.856	1.710	2.623	1.810

In the above table "*" indicate the test set compounds

Table 2: Descriptors and specification of multiple linear regression models

Dependent variable	Name of descriptor	Notation	Classification	Constant	Coeff.	SE
logT _{1/2}	average molecular weight	AMW	constitutional	0.653	0.13	0.014
	HA dependent HDCA2 ZefirovsPC	HAdependent.	electronic		-4.56	0.710
	number of double bonds	nDB	constitutional		0.32	0.061
	3D-MORSE-signal 18 weighted by atomic masses	Mor18m	geometrical		0.58	0.162
P _{eff} (sandy)	R maximal autocorrelation of lag 4/weighted by atomic sanderson electronegativities	R4e_A	topological	-3.310	17.45	5.321
	Mean information index of atomic composition	AAC	topological		4.78	0.957
	R maximal autocorrelation of lag 2/weighted by atomic sanderson electronegativities	R2e_A	topological		-7.15	2.277
	Geary autocorrelation - lag 2 / weighted by atomic polarizabilities	GATS2p	topological		1.88	0.831
P _{eff} (clayey)	R maximal autocorrelation of lag 4/weighted by atomic sanderson electronegativities	R4e_A	topological	3.370	53.12	6.523
	3D-MORSE-signal 29/weighted by atomic vander Waals volumes	Mor29v	geometrical		-12.08	2.444
	3D-MORSE-signal 12/ unweighted	Mor12u	geometrical		5.61	1.242
	R maximal autocorrelation of lag 2/unweighted	R2u_A	topological		-10.73	3.222
γ _{eff-nd} (sandy)	number of multiple bonds	nBM	constitutional	-0.100	0.06	0.012
	Qyy COMMA2 value/weighted by atomic Sanderson	QYYe	geometrical		0.01	0.001
	Moran autocorrelation-lag 6/ weighted by atomic Sanderson electronegativities	MATS6e	topological		0.20	0.070
	Mean information content index	Kierflexibilityindex	topological		0.02	0.014
γ _{eff-nd} (clayey)	3D-MORSE-signal 30/ unweighted	Mor30u	geometrical	-1.300	21.89	3.910
	Moran autocorrelation -lag 3 /weighted by atomic masses	MATS3m	topological		3.11	0.588
	R maximal autocorrelation of lag 4/unweighted	R6u_A	topological		31.07	7.946
	structural information content (neighborhood symmetry of 3-order).	SIC3	topological		4.21	1

After calculation of the molecular descriptors, descriptors that show high correlation ($R > 0.90$) with each other were identified and only one of them in each correlated pair was considered in developing the model because of these descriptors encoded similar information about the molecule of interest.

Also descriptors that are constant or near-constant for all molecules were eliminated from pool of descriptors. Then, the stepwise multiple linear regression method was used for the selection of the most relevant descriptors [17]. Table 2 shows the names and classes of selected descriptors. These descriptors would be used as inputs for construction of Least-squares support vector machine (LS-SVM) and MLR models.

2.3. Least-squares support vector machine

Least-squares support vector machine (LS-SVM), which is a modified algorithm of SVM, has the capability of dealing with linear and non-linear multivariate calibration and resolving these problems in a relatively rapid way. This novel type of machine learning method was developed by Suykens et al. [18] with the implementation of a least squares version for support vector machine.

The details of LS-SVM algorithm could be found in the references of [19-21]. In principle, LS-SVM always fits a linear relation between the regressors (x) and the dependent variable (y) follows:

$$y = \sum_{i=1}^N \alpha_i x_i^T x + b \quad (3)$$

and

$$\alpha_i = 2\gamma e_i \quad (4)$$

In the above equations, α_i is the Lagrange multipliers called support value, b is the bias term and γ (gam) is the relative weight of the error term, which would be optimized by the user. By introducing the kernel function, following nonlinear regression function is achieved.

$$y = \sum_{i=1}^N \alpha_i K(x_i, x) + b \quad (5)$$

where $K(x_i, x)$ is the kernel function. The most general kernel function is radial basis function (RBF):

$$K(x_i, x) = \exp(-\|x_i - x\|^2 / 2\sigma^2) \quad (6)$$

where σ^2 is the width of the RBF function. Generalization capability of SVM depends on the proper selection of its parameters. After training, the machine is used to predict or estimate target values for objects where these values are unknown.

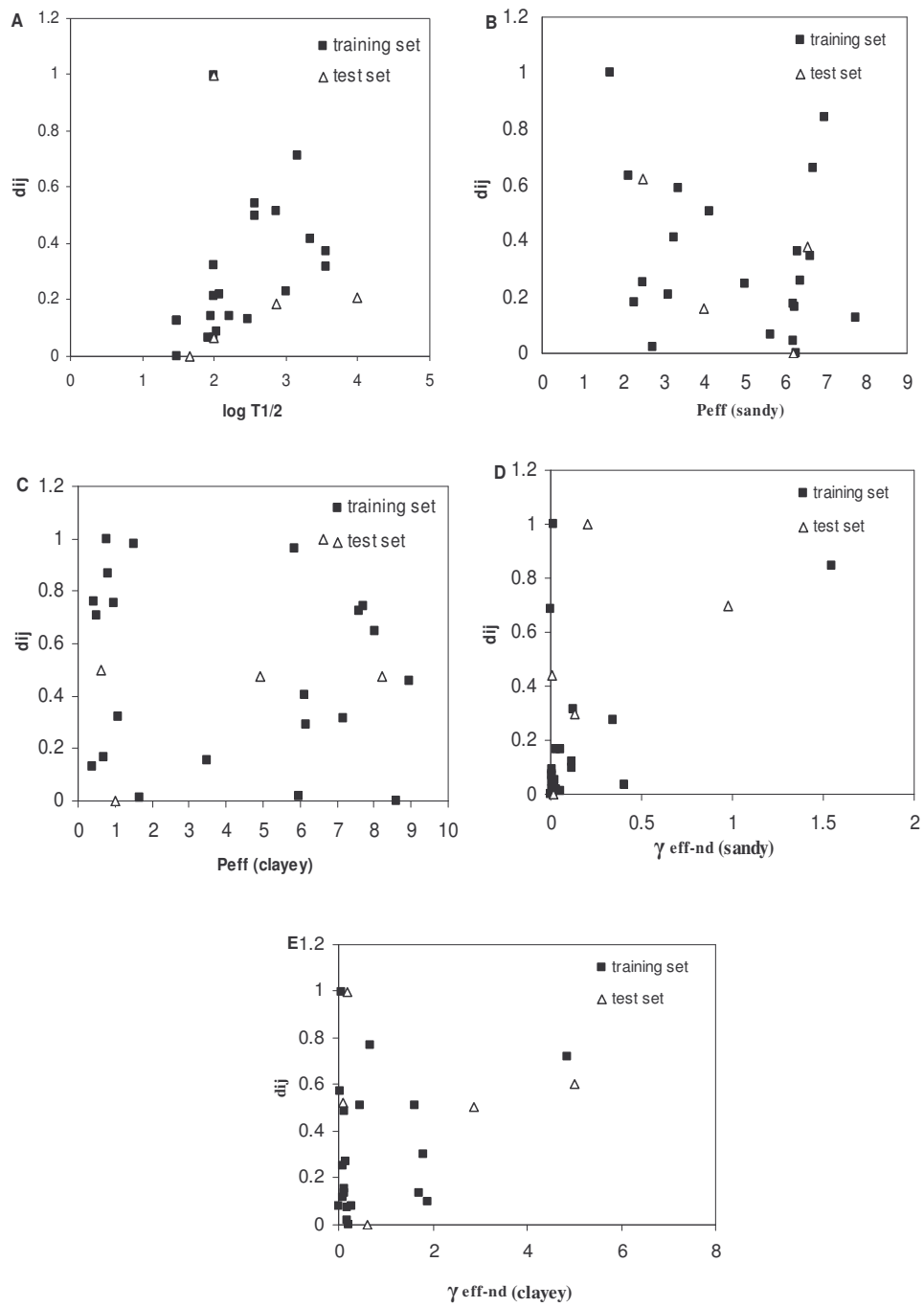


Figure 1: The results of diversity test

3. Results and discussion

3.1. Diversity analysis

Rational division of the experimental data set into training and test sets are an important part in the development and validation of reliable QSAR model.

In this study, diversity analysis was performed to make sure that the structures of the training and test cases can represent those of the whole ones. In this way, the mean distances of one sample to the remaining ones (\bar{d}_i) were computed from descriptor space matrix as follows:

$$\bar{d}_i = \frac{\sum_{j=1}^n d_{ij}}{n-1} \quad i=1, 2, \dots, n \quad (7)$$

where d_{ij} is a distance score for two different compounds, that can be measured by the Euclidean distance norm based on the compound's descriptors (X_{ik} and X_{jk}):

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (8)$$

Then the mean distances were normalized within the interval of zero to one and the result values were plotted against experimental parameters (Fig. 1). As can be seen from this figure, the structures of the compounds are diverse in all sets and the training set with a broad representation of the chemistry space was adequate to ensure the model's stability and the diversity of test set can prove the predictive capability of the model.

3.2. Model development

The method of stepwise multi-linear regression was used to select the most relevant descriptors. The results of calculation of correlation matrix for selected descriptors of each parameter showed that there is no high correlation among the selected descriptors. Then the MLR technique was used to build linear QSAR models to predict the effective Péclet numbers and effective nondimensional degradation rate constants of some VOCs in sandy and clayey soil and also to predict their half-life values. Table 2 shows the specification of these MLR models. For modeling of each parameter, the MLR technique was performed on the molecules of the training set by SPSS (V.13) and the best model was applied on the molecules of the test set to evaluate the prediction power of developed models.

The calculated values of these parameters by MLR models were shown in Table 1.

Since the bioenvironmental activities of VOCs is very complex phenomena that influenced by many environmental and structural parameters, therefore it was necessary to develop some non-linear models to investigate any non-linear relation between dependent variables and molecular descriptors.

Due to some advantage of LS-SVM over ANN technique (such as the absence of overtraining and ability to work on small data set) it was decided to use LS-SVM technique with RBF kernel function as a non-linear feature mapping method. In the model development by using LS-SVM technique, the values of γ and σ^2 parameters were a manageable task as a two-dimensional problem. To find out the optimal

values of γ and σ^2 , a search was performed based on cross validated root mean square error (RMSECV) of the original training set for all combinations of γ and σ^2 from 1 to 80, with increment steps of 1.

A robust model is achieved by selecting parameters that give the lowest RMSE. The obtained response

graph of RMSECV of effective nondimensional degradation rate constants in clayey soil as a function of γ and σ^2 is presented in Fig. 2. The same graph for other parameters was achieved in the same way. By inspection to these graphs, optimal values of γ and σ^2 are collected in Table 3.

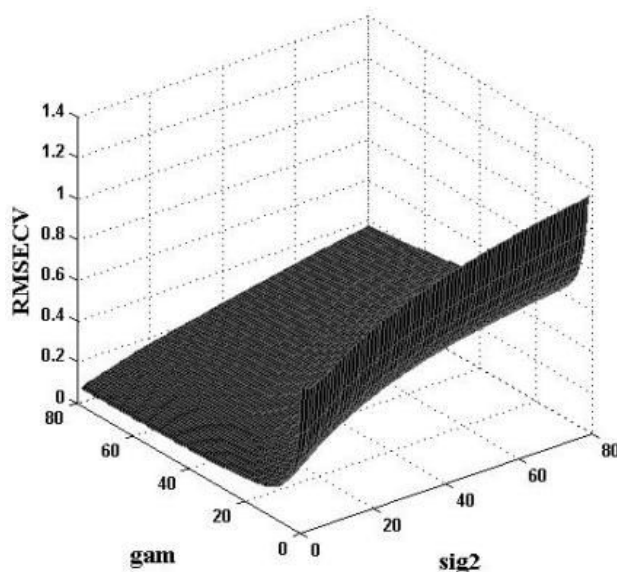


Figure 2: Response graph of LS-SVM model of $\gamma_{\text{eff-nd}}$ (clayey)

Table 3: Optimal values of γ and σ^2 for LS-SVM models

Dependent variable	γ	σ^2
$\log T_{1/2}$	80	20
P_{eff} (sandy)	80	19
P_{eff} (clayey)	30	2
$\gamma_{\text{eff-nd}}$ (sandy)	15	50
$\gamma_{\text{eff-nd}}$ (clayey)	75	10

The developed models were evaluated by test sets. The calculated values of $\log T_{1/2}$, P_{eff} and $\gamma_{\text{eff-nd}}$ in sandy and clayey soil of VOCs by LS-SVM models are shown in Table 1. Leave one out cross validation test on developed LS-SVM models showed that these models are more stable and reliable than MLR models.

The cross validated correlation coefficient (R_{cv}^2), standardized predicted error sum of squares (SPRESS) and other statistical parameters for the MLR and LS-SVM models are shown in Table 4.

The residual of LS-SVM models calculated values of each parameter are plotted against their experimental values in Fig. 3. The propagation of residuals on both sides of zero line shows that no systematic error exists in the developed models. As can be seen in Table 2 sixteen descriptors appear in QSAR models of P_{eff} and $\gamma_{\text{eff-nd}}$ in sandy and clayey soils, which eleven of them are topological and four of them are geometrical.

Inspection to these descriptors revealed the effect of atomic connectivity and three-dimension structure of VOCs on their bioenvironmental parameters. Four descriptors appear in QSAR model of $\log T_{1/2}$. Among of these descriptors average molecular weight (AMW) and number of double bonds (nDB) are constitutional. AMW depends on number and type of atoms. nDB encodes the information about number of attack sites for electrophilic compounds. HA dependent HDCA2 ZefirovsPC describes the potency of molecules in formation of hydrogen bonds. Mor18m represent 3D-structure of molecules based on electron diffraction. Among of these descriptors AMW and Mor18m represent size and topology of molecules and nDB and HA dependent HDCA2 ZefirovsPC encode the effect of electronic aspect on molecular degradability of chemicals.

Table 4: Statistical parameters obtained using MLR and LS-SVM

Dependent variable	MLR						LS-SVM					
	training set		test set		R_{cv}^2	SPRESS	training set		test set		R_{cv}^2	SPRESS
	R	RMSE	R	RMSE			R	RMSE	R	RMSE		
$\log T_{1/2}$	0.92	0.28	0.87	0.48	0.63	0.50	0.97	0.16	0.80	0.62	0.65	0.35
P_{eff} (sandy)	0.90	0.89	0.87	1.40	0.69	1.07	0.98	0.39	0.94	1.03	0.87	0.75
P_{eff} (clayey)	0.94	1.18	0.94	1.10	0.75	1.20	0.99	0.46	0.89	1.7	0.82	0.72
$\gamma_{\text{eff-nd}}$ (sandy)	0.95	0.09	0.94	0.13	0.57	0.13	0.99	0.05	0.85	0.26	0.73	0.14
$\gamma_{\text{eff-nd}}$ (clayey)	0.93	0.65	0.92	0.84	0.60	0.79	0.99	0.12	0.93	1.00	0.70	0.52

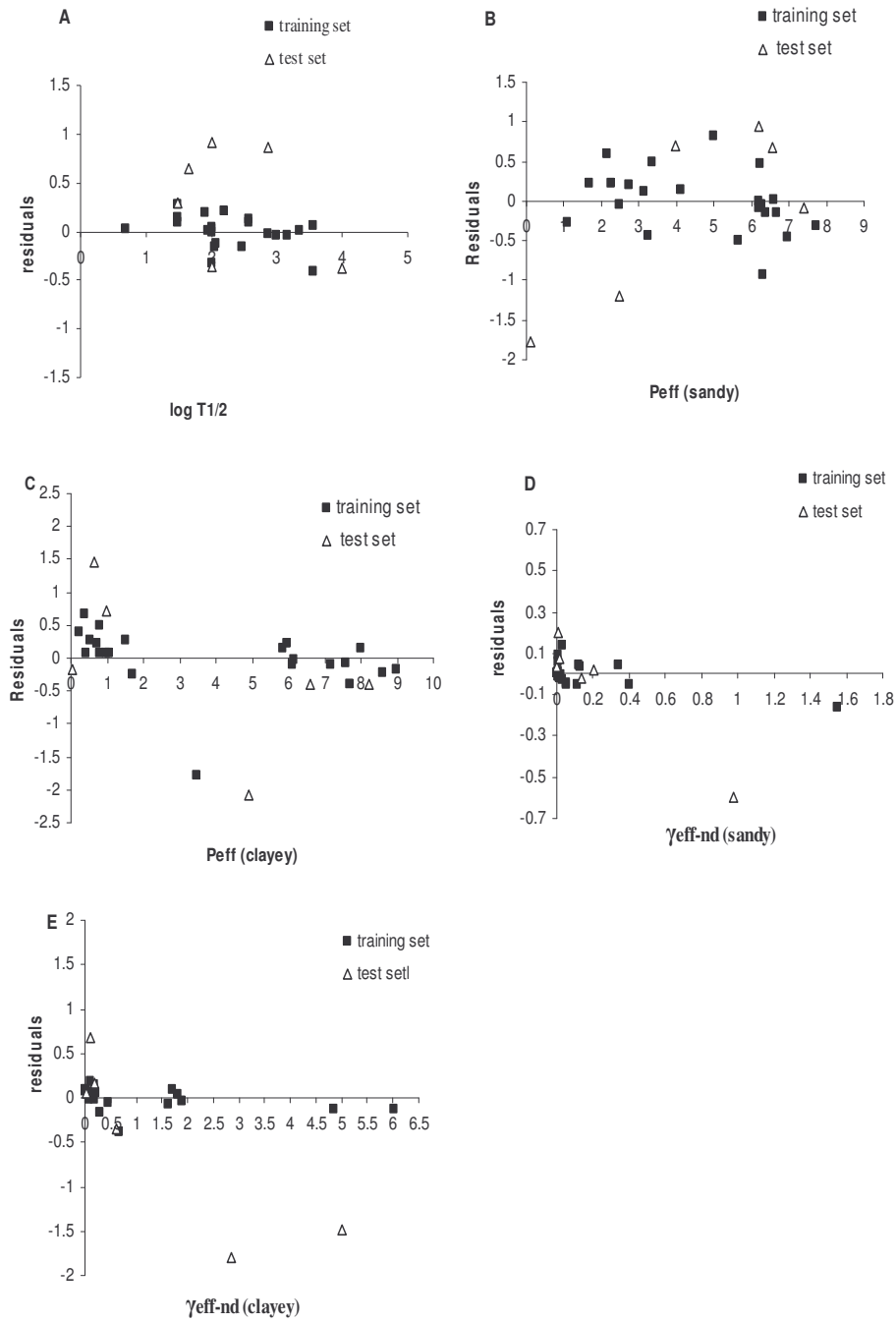


Figure 3: Plots of residuals versus their experimental factor

Estimation of the relative importance of each descriptor in the LS-SVM models is important for chemical interpretation. In this study, sensitivity analysis was used for this purpose. This method is performed based on the sequential setting to zero of each descriptor.

For each zeroed descriptor, root mean square error (RMSE_i) was calculated. Generally RMSE_i value increases in this way. Then, the differences between RMSE_i and RMSE of established LS-SVM was calculated and shown as DRMSE. Each descriptor which causes greater value of DRMSE is more important. Results of the sensitivity analysis showed that the order of importance of descriptors in LS-SVM models for each interested bioenvironmental parameter are; AMW> HAdependentHDCA2ZefirovsPC> nDB> Mor18m for logT_{1/2}, AAC> R4e_A> R2e_A> GATS2p for P_{eff} (sandy soil), Mor29v> R2u_A> Mor12u> R4e_A for P_{eff} (clayey soil), QYYe> MATS6e> Kierflexibilityindex and nBM for $\gamma_{\text{eff-nd}}$ (sandy soil) and Mor30u> MATS3m> R6u_A> SIC3 for $\gamma_{\text{eff-nd}}$ (clayey).

4. Conclusion

The results of this study revealed that it was possible to predict the half-life, nondimensional effective degradation rate constants and effective Péclet numbers in two type of soil for volatile organic compounds from their structural features. Descriptors that were appeared in these models were topological and electronic types, which revealed the role of electronic and steric interactions in environmental fats of chemicals. nential of its oxidation reaction.

5. References

- [1] Ü. Niinemets, F. Loreto, M. Reichstein, Trends Plant Sci., 9(2004)180.
- [2] P. Tata, J. Witherspoon, C. Lue-Hing, VOC Emissions from Wastewater Treatment Plants: Characterization, Control and Compliance, Lewis Publishers, Boca Raton, Florida, 2003.
- [3] M. S. Diaz-Cruz, D. Barcelo, Chemosphere, 72(2008)333.
- [4] M. Minnich, Behavior and determination of volatile organic compounds in soil, US Environmental Protection Agency, Las Vegas, Nevada, 1993.
- [5] M. M. Hantush, R. S. Govindaraju, M. A. Marino, Z. Zhang, J. Hydrol., 260(2002)58.
- [6] W. A. Jury, D. Russo, G. Streile, H. E. Abd, Water Resour. Res., 26(1990)13.
- [7] R. Parashar, R. S. Govindaraju, M. M. Hantush, J. Environ. Engin., 133(2007)879.
- [8] F. Stagnitti, G. Allinson, M. Morita, M. Nishikawa, H. Il, T. Hirata, Environ. Model. Assess., 5(2000)229.
- [9] M. Jalali-Heravi, M. Asadollahi-Baboli, P. Shahbazikhah, Eur. J. Med. Chem., 43(2008)548.
- [10] P. Gramatica, F. Consolaro, S. Pozzi, Chemosphere, 43(2001)655.
- [11] E. Papa, P. Gramatica, SAR and QSAR Environ. Rcs., 19(2008)655.
- [12] F. Luan, H. Z. Si, H. T. Liu, Y. Y. Wen, X. Y. Zhang, SAR and QSAR Environ. Rcs., 19(2008)465.
- [13] HYPERCHEM. Release 7.0 for Windows, Hypercube, Inc, Florida, USA, 2002.

- [14] J. P. P. Stewart, MOPAC 6.0, Quantum Chemistry Program Exchange, QCPE, No. 455, Indiana University, Bloomington, IN, 1989.
- [15] R. Todeschini, V. Consonni, A. Mauri, and M. Pavan, DRAGON 3.0, Talete srl, Milan, Italy; <http://www.talete.mi.it/dragon.htm>.
- [16] A. R. Katritzky, V. S. Lobanov, M. Karelson, CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis 2.0), Reference Manual, 1994.
- [17] M. H. Fatemi, *Anal. Chim. Acta.*, 556(2006)355.
- [18] J. A. K. Suykens, T. van Gestel, J. de Brabanter, B. de Moor, J. Vandewalle, *Least-Squares Support Vector Machines*, World Scientific, Singapore, 2002.
- [19] M. F. Ferrao, S. C. Godoy, A. E. Gerbase, C. Mello, J. C. Furtado, C. L. Petzhold, R. J. Poppi, *Anal. Chim. Acta*, 595(2007)114.
- [20] W. Cui, X. Yan, *Chemom. Intell. Lab. Syst.*, 98(2009)130.
- [21] Q. Chen, J. Zhao, C. H. Fang, D. Wang, *Spectrochim. Acta Part A*, 66(2007)568.