Studying the Explanatory Capacity of Artificial Neural Networks for Understanding Environmental Chemical Quantitative Structure−Activity Relationship Models

Lei Yang,[†,‡] Peng Wang,*[,†] Yilin Jiang,[†] and Jian Chen[†]

Research Center for Green Chemistry and Technology, School of Municipal and Environmental Engineering, Harbin Institute of Technology, Harbin 150090, China, and Department of Applied Chemistry, Harbin Institute of Technology, Harbin 150001, China

Received March 5, 2005

Although artificial neural networks (ANNs) have been shown to exhibit superior predictive power in the study of quantitative structure−activity relationships (QSARs), they have also been labeled a "black box" because they provide little explanatory insight into the relative influence of the independent variables in the predictive process so that little information on how and why compounds work can be obtained. Here, we have turned our interests to their explanatory capacities; therefore, a method was proposed for assessing the relative importance of variables indicating molecular structure, on the basis of axon connection weights and partial derivatives of the ANN output with respect to its input, which can identify variables that significantly contribute to network predictions, and providing a variable selection method for ANNs. We show that, by extending this approach to ANNs, the "black box" mechanics of ANNs can be greatly illuminated, thereby making it very useful in understanding environmental chemical QSAR models.

1. INTRODUCTION

Modeling the relationships between chemical structure and various biological effects or activities of interest is the target of structure−activity studies. As the knowledge in the field of quantitative structure−activity relationships (QSARs) accumulates, it confirms the very high complexity and nonlinear character of most of such relationships. Artificial neural networks (ANNs) are among the best available tools to generate such nonlinear models.[1−2] A large number of authors have underlined the interest of using ANNs instead of linear statistical models,[3−5] and the rapidly increasing number of successful QSAR studies based on neural networks is the best proof of a very productive cooperation between artificial intelligence and chemistry.

A very important advantage of neural networks is that they are adaptive; that is, they can take data and learn from it. This ability does not depend on the prior knowledge of rules. Perhaps the most important advantage of all is that neural networks allow for going directly from factual data to the models without any human subjective interference, that is, without tainting the result with oversimplification or preconceived ideas. But a neural network model describes associations, not causes, and therefore, it does not explain its decisions. Many view this as a drawback. Another subjective objection against the use of neural networks is connected with model transparency. Most opponents of neural networks are promoting the idea that the use of neural networks is a "black box" approach to modeling the relationships between structure and various biological effects.[6−7]

This view stems from the fact that the contribution of the input variables in predicting the value of the output is difficult to disentangle within the network. Consequently, input variables indicating molecular structure are often entered into the network and an output value is generated without gaining any understanding of the inter-relationships between the variables and, therefore, providing no explanatory insight into the underlying mechanisms being modeled by the network.[8−10] The "black box" nature of ANNs is a major weakness compared to traditional statistical approaches that can readily quantify the influence of the independent variables in the modeling process. This may be true for the novice user, but it is very far from reality.

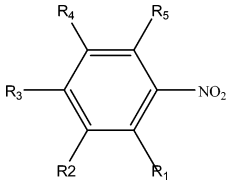
Convinced by the predictive power of ANNs and their ability to analyze nonlinear relationships, we consider them interesting to study from their explanatory point of view. In fact, starting from input variables, ANNs have the capacity to predict the output variable, but the mechanisms that occur within the networks are often ignored. Some authors have explored this problem and proposed algorithms to illustrate the role of variables in ANN models. Nevertheless, in most works, these methods are used to eliminate irrelevant input and are, therefore, called pruning methods.[11−14]

In this paper, we propose and apply a method for assessing the relative importance of variables on the basis of axon connection weights and partial derivatives of the ANN output with respect to its input and make an application of seven quantum chemical descriptors (e.g., Q_C, Q_N, $-E_{LUMO}$, $-E_{HOMO}$, etc.) and three kinds of recently introduced autocorrelation topological indices, A, B, and C,[15] to predict the toxicities of 35 nitro-aromatic compounds ($ArNO_2$) to the fathead minnow. This approach can identify variables that significantly contribute to network predictions and provide a variable selection method for ANNs. We show that, by extending this approach to ANNs, the "black box" mechanics of ANNs can be greatly illuminated, thereby making it very useful in finding out how nitro-aromatic compounds affect the fathead minnow.

* Corresponding author. Tel.: 86-0451-86283801. E-mail: pwang73@vip.sina.com.
† Research Center for Green Chemistry and Technology.
‡ Department of Applied Chemistry.

10.1021/ci050079x CCC: $30.25
Published on Web 10/13/2005

Table 1. Nitro-Aromatic Compounds and Their Toxicities



number	compound	R_1	R_2	R_3	R_4	R_5	lg1/LC_{50} (obsd)[a]
1	2-nitrotoluene	CH_3	H	H	H	H	3.57
2	3-nitrotoluene	H	CH_3	H	H	H	3.63
3[b]	4-nitrotoluene	H	H	CH_3	H	H	3.76
4	1,2-dinitrobenzene	NO_2	H	H	H	H	5.45
5	1,3-dinitrobenzene	H	NO_2	H	H	H	4.38
6	1,4- -dinitrobenzene	H	H	NO_2	H	H	5.22
7[b]	2,3-dinitrotoluene	NO_2	CH_3	H	H	H	5.01
8	2,4-dinitrotoluene	CH_3	H	H	NO_2	H	3.75
9	2,5-dinitrotoluene	CH_3	H	NO_2	H	H	5.15
10	2,6-dinitrotoluene	CH_3	NO_2	H	H	H	3.99
11[b]	3,4-dinitrotoluene	NO_2	H	CH_3	H	H	5.08
12	3,5-dinitrotoluene	H	CH_3	H	NO_2	H	3.91
13	1,3,5-trinitrobenzene	H	NO_2	H	NO_2	H	5.29
14	nitrobenzene	H	H	H	H	H	3.02
15	2-nitroaniline	NH_2	H	H	H	H	3.70
16[b]	2,4-dinitroaniline	NH_2	H	H	NO_2	H	4.07
17	4-nitrophenol	H	H	OH	H	H	3.36
18	4-fluoronitrobenzene	H	H	F	H	H	3.70
19[b]	2,4,6-trinitrotoluene	CH_3	NO_2	H	NO_2	H	4.88
20	2,3,6-trinitrotoluene	NO_2	CH_3	NO_2	H	H	6.37
21	2-methyl-3-nitroaniline	CH_3	NH_2	H	H	H	3.48
22[b]	2-methyl-4-nitroaniline	H	CH_3	NH_2	H	H	3.24
23	2-methyl-5-nitroaniline	H	NH_2	CH_3	H	H	3.35
24	2-methyl-6-nitroaniline	NH_2	CH_3	H	H	H	3.80
25	3-methyl-6-nitroaniline	NH_2	H	CH_3	H	H	3.80
26	4-methyl-2-nitroaniline	NH_2	H	H	CH_3	H	3.79
27	3-nitro-4-hydroxyaniline	OH	H	H	NH_2	H	3.65
28[b]	4-methyl-3-nitroaniline	CH_3	H	H	NH_2	H	3.77
29	2,4-dinitrophenol	OH	H	H	NO_2	H	4.04
30	2-methyl-3,5-dinitroaniline	CH_3	NH_2	H	NO_2	H	4.14
31	2-methyl-3,6-dinitroaniline	CH_3	NH_2	NO_2	H	H	5.34
32	3-methyl-2,4-dinitroaniline	NH_2	H	H	NO_2	CH_3	4.26
33	3-methyl-2,6-dinitroaniline	NH_2	NO_2	CH_3	H	H	4.21
34	4-methyl-2,6-dinitroaniline	NH_2	H	NO_2	CH_3	H	4.18
35	4-methyl-3,5-dinitroaniline	CH_3	NO_2	H	NH_2	H	4.46

[a] The data lg1/LC_{50} come from ref 16. [b] The data of the testing set.

In addition, the predictive and explanatory capacity of stepwise regression will be compared with those of the contribution procedures associated with ANNs.

2. DATABASE

2.1. Compounds and Toxicity Data. The toxicity 96h-LC_{50} (mmol/L) data of the fathead minnow caused by nitro-aromatic compounds used here were reported by Hall.[16] The 35 nitro-aromatic compounds mainly consist of methylbenzenes, anilines, and phenols with different numbers of nitro groups. The basic structures of these compounds are shown in Table 1, along with lg1/LC_{50}.

2.2. Molecular Descriptor Data. Molecular descriptors originate from the molecular structures and can be related to physicochemical properties of molecules. A wide variety of descriptors have been reported for use in QSAR analysis.[17−19] As descriptor selection was limited to stereo-electronic and energy parameters previously found to be important in modeling toxic potency,[16,25] here, seven quantum chemical descriptors have been calculated, using the HyperChem version 6.03 software, and five autocorrelation topological descriptors belonging to three types A, B, and C also have been obtained,[15] using a program written in Borland C++ Builder 5.0 by our research center. All 12 of the descriptors are listed in Table 2, and relationships between them are shown in Table 3.

3. METHODS

3.1. Preparation of the Data. Prior to neural network optimization, the data set must be transformed so that the dependent variables exhibit particular distributional characteristics. The dependent variable must be converted to the range [0...1] so that it conforms to the demand of the transfer function used (sigmoid function) in the building of the neural network. This is accomplished by using the formula as follows:

$$r_n = \frac{y_n - \min(Y)}{\max(Y) - \min(Y)} \quad (1)$$

where r_n is the converted response value for observation n,

Table 2. Quantum Chemical and Topological Descriptors and Toxicity Data of Nitro-Aromatic Compounds

number	variable	type	characteristics
1[a]	Q_C	i	net positive charge of C atom on the benzene ring joined to nitro group
2[a]	Q_N	i	net charge of N atom in the nitro group on the benzene ring
3[a]	Q_{NO_2}	i	maximal net positive charge among nitro groups
4[a]	HF	i	heat of formation (kal/mol)
5[a]	m	i	molecular dipole moment (C m)
6[a]	$-E_{LUMO}$	i	energy of lowest unoccupied molecular orbital (eV)
7[a]	$-E_{HOMO}$	i	energy of highest occupied molecular orbital (eV)
8[b]	A1	i	sum of volume information between adjacent atoms
9[b]	A2	i	sum of volume information between spacing atoms
10[b]	B1	i	sum of electronic information between adjacent atoms
11[b]	B2	i	sum of electronic information between spacing atoms
12[b]	C1	i	sum of covalently linked information between atoms
13	$lg1/LC_{50}$	d	96-h half lethal concentration of the fathead minnow

[a] Descriptors calculated by HyperChem version 6.03 software. [b] Descriptors defined in ref 15.

y_n is the original response value for observation n, and min(Y) and max(Y) represent the minimum and maximum values, respectively, of the response variable Y.

To standardize the measurement scales of network inputs, the independent variables are converted to z scores (i.e., mean = 0, standard deviation = 1) using the following formula:

$$z_n = \frac{x_n - \bar{X}}{\sigma_x} \tag{2}$$

where z_n is the standardized value of observation n, x_n is the original value of observation n, and $\bar{X}$ and σ_x are the mean and standard deviation of the variable X. It is essential to standardize the input variables so that the same percentage change in the weight sum of the inputs and partial derivatives of the ANN output with respect to its input can cause a similar percentage change in the unit output.

In the process of modeling artificial neural networks, generally, the number of compounds in the learning set was about 4−5 times that in the testing set. And for testing data, their structures must be similar to those of the learning set. So the 35 nitro-aromatic compounds were divided into a learning set (28 compounds) and a testing set (7 compounds) (see Table 1).

3.2. Multiple Linear Regression Modeling (MLR). With MLR being the method most frequently used in QSAR, a comparison to ANNs was made in order to judge their explanatory capacities. The stepwise multiple regression technique was computed especially to define the significant variables and their contribution order. In fact, the influence of each variable can be roughly assessed by checking the final values of the regression coefficients.

3.3. Neural Network Modeling. We shall refrain from detailing the specifics of multilayer feed-forward network optimization and design and, instead, refer the reader to the extensive coverage provided in the literature by Smith[20] and Ripley,[21] as well as articles by Cheng and Titterington.[22] The network used consisted of three layers: one input layer with 12 neurons (one for each input variable), one hidden layer with different neurons, and one output layer with one neuron which was the output variable, $lg1/LC_{50}$ (Chart 1). To assess the relative importance of input variables, the distribution and contribution of each variable to the ANN output must be known. By partitioning the connecting weight for each variable, we can get an idea of the distribution of each input in the ANN models. And by calculating the partial derivatives of the ANN output with respect to each input, we are also able to evaluate the effect of each input on the output. On the basis of the two aspects, we can understand how the variables influence output and, furthermore, provide explanatory insight into how nitro-aromatic compounds poison the fathead minnow.

3.3.1. Partial Derivative Method. Two results can be obtained by partial derivatives of the ANN output with respect to its input. The first is a profile of the output variations for small changes of each input variable, and the second is a classification of the relatively important contributions of each variable to the network output.

For the ith input variable and the kth learning sample, the calculational method for the partial derivative can be obtained as below (assuming that a logistic sigmoid function is used for the activation).

Given that

$$f(u) = \frac{1}{1 + e^{-U}},\ U = \sum_{h=1}^{N_h} W_{ho} H_h - \theta_o,\ H_h(V) = \frac{1}{1 + e^{-V}},\ \text{and}\ V = \sum_{i=1}^{N_i} W_{ih} X_{ik} - \theta_h$$

$$\frac{\partial f}{\partial X_{ik}} = \frac{\partial f}{\partial U} \sum_{h=1}^{N_h} \frac{\partial U}{\partial H_h} \frac{\partial H_h}{\partial V} \frac{\partial V}{\partial X_{ik}} = \frac{e^{-U}}{(1 + e^{-U})^2} \sum_{h=1}^{N_h} W_{ho} \frac{e^{-V}}{(1 + e^{-V})^2} W_{ih} = f(u)\,[1 - f(u)] \sum_{h=1}^{N_h} W_{ho}\, H_h(V)\,[1 - H_h(V)] W_{ih} \tag{3}$$

where f is the value of the output neuron with respect to a learning sample, H_h is the response of the hth hidden neuron, W_{ho} and W_{ih} are the weights between the output neuron and hth hidden neuron and the weights between the ith input neuron and the hth hidden neuron, respectively, and N_i and N_h are the number of input layer neurons and hidden layer neurons.

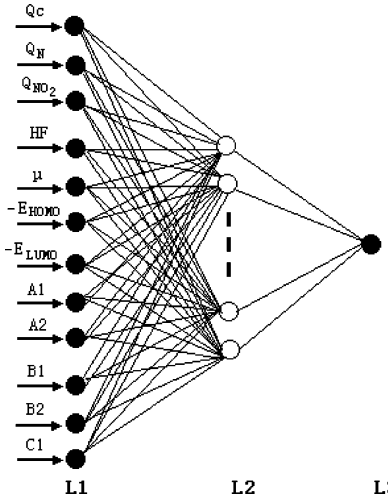
A set of graphs of the partial derivatives versus each corresponding input variable can then be plotted, enabling direct access to the influence of the input variable on the output. One example of an interpretation of these graphs is that, if the partial derivatives are negative, then, for this value of the studied variable, the output variable will tend to decrease while the input variable increases. Inversely, if the

Table 3. Correlation Matrix of All Variables[a]

	Q_C	Q_N	Q_{NO_2}	HF	m	$-E_{LUMO}$	$-E_{HOMO}$	A1	A2	B1	B2
Q_N	−0.222										
Q_{NO_2}	−0.378	0.308									
HF	0.623	−0.259	−0.711								
m	−0.421	0.545	0.678	−0.703							
$-E_{LUMO}$	0.365	−0.552	−0.642	0.461	0.811						
$-E_{HOMO}$	−0.489	0.279	0.336	0.632	−0.547	−0.112					
A1	0.732	0.057	0.564	0.687	0.727	−0.632	0.336				
A2	0.632	0.494	0.632	0.664	0.758	−0.588	0.479	0.736			
B1	−0.421	0.446	0.391	−0.559	0.536	0.614	0.553	−0.663	−0.715		
B2	−0.213	0.353	0.089	−0.415	0.446	0.677	0.211	−0.562	−0.703	0.775	
C1	0.583	0.372	0.449	0.531	0.666	−0.785	0.389	0.776	0.754	−0.669	−0.632
$lg1/LC_{50}$	0.446	0.669	0.586	0.241	−0.558	0.379	0.439	0.214	0.510	−0.222	−0.621

[a] Table 3 reveals that all the descriptors were not correlated and no single molecular descriptor accurately described the toxicity.

Chart 1. Structure of the Neural Network Used in This Study[a]



[a] L1, input layer with neurons comprising 12 variables; L2, hidden layer with neurons whose number is determined empirically; L3, output layer with single neuron corresponding to the single dependent variable

partial derivatives are positive, the output variable will tend to increase while the input variable also increases.

The second result of the partial derivative of the ANN output with respect to its input concerns the relative importance of the contribution of the ANN output to the data set with respect to an input. It can be calculated by a sum of the square partial derivatives obtained per input variable (i.e., for the *i*th input variable):

$$S_i = \sum_{k=1}^{N} \frac{\partial f}{\partial X_{ik}} \tag{4}$$

where N is the number of training sample data points.

Then, the relative importance of *i*th input variable acting on all the output (PRI) is obtained.

$$\mathrm{PRI}(\%)_i = \frac{S_i}{\sum_{i=1}^{N_i} S_i} \times 100 \tag{5}$$

3.3.2. Weight Method. The procedure for partitioning the connection weights to determine the relative importance of the various inputs was proposed first by Garson[11] and repeated by Goh.[23] The method essentially involves partitioning the hidden−output connection weights of each hidden neuron into components associated with each input neuron. We suggested a simplification of this algorithm that gave results identical to those of the algorithm initially proposed:

(1) For each hidden neuron *h* with respect to *i*th input variable, divide the absolute value of the input−hidden layer connection weight by the sum of the absolute value of the input−hidden layer connection weights concerning *i*th input variable.

$$Q_{ih} = \frac{|W_{ih}|}{\sum_{i=1}^{N_i} |W_{ih}|} \tag{6}$$

(2) For each input neuron *i*, divide the sum of the Q_{ih} for each hidden neuron by the sum for each hidden neuron of the sum for each input neuron of Q_{ih} and multiply by 100. The relative importance of all output weights (WRI) attributable to the given input variable is then obtained.

$$\mathrm{WRI}(\%)_i = \frac{\sum_{h=1}^{N_h} Q_{ih}}{\sum_{h=1}^{N_h}\sum_{i=1}^{N_i} Q_{ih}} \times 100 \tag{7}$$

3.3.3. Assessing the Relative Importance of Variables. On the basis of the two methods above, we can finally identify the relative importance of each variable. To minimize the influence of network structure and raise the stability of the method assessing the relative importance of each input variable, we are first concerned with how much the partial derivatives of $lg1/LC_{50}$ with respect to each input and connection weight contribute to the relative importance of the input variable. Using different networks containing different hidden neuron modeling, various PRI_i and WRI_i values can be obtained. For each network, we can get an error of PRI_i/WRI_i with an average of PRI_i/WRI_i for all networks; then, the total errors of PRI/WRI for each network structure can be obtained. Thus, the variances of the two methods' errors can also be obtained.

Here, X and Y respectively represent the total errors of PRI/WRI of each network for the partial derivative method

and the weight method; a and b are the rates of contribution of these two methods to the relative importance of each variable; therefore, the sum is 1. To minimize the difference of relative importance of each variable in each network and raise the stability of the method, a and b must be determined as follows:

$$\begin{aligned}\mathrm{Var}(aX + bY) &= a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)\\ &= (1-b)^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)\\ &= [\mathrm{Var}(Y) + \mathrm{Var}(X)]b^2 - 2b\,\mathrm{Var}(X) + \mathrm{Var}(X)\\ &= [\mathrm{Var}(Y) + \mathrm{Var}(X)]\left\{b - \frac{\mathrm{Var}(X)}{[\mathrm{Var}(Y) + \mathrm{Var}(X)]}\right\}^2 + \frac{-\mathrm{Var}^2(X)}{[\mathrm{Var}(Y) + \mathrm{Var}(X)]} + \mathrm{Var}(X)\end{aligned}$$

$$b = \frac{\mathrm{Var}(X)}{\mathrm{Var}(Y) + \mathrm{Var}(X)} \qquad a = \frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y) + \mathrm{Var}(X)} \quad (8)$$

Then, the relative importance (RI) of the ith input variable could be gained.

$$\mathrm{RI}_i = a \times \mathrm{PRI}_i + b \times \mathrm{WRI}_i \quad (9)$$

In this paper, the partial method contributed 42% to the assessment of the relative importance of each variable, whereas the weight method contributed 58%.

4. RESULTS

4.1. Multiple Linear Regression Models. *4.1.1. Predictive Capacity.* Using all 12 available variables and the complete data set, the equation of the MLR model and the determination coefficient were

$\lg 1/LC_{50} =$	2.30	$-0.27Q_C$	$-2.28Q_N$	$-4.43Q_{NO_2}$
t		−0.423	−3.555	−1.912
sig.		0.6770	0.0020	0.0680
	$+0.13HF$	-0.37μ	$-0.13E_{LUMO}$	
t	−0.756	−3.01	0.565	
sig.	0.4580	0.0060	0.5770	
	$+0.47E_{HOMO}$	−0.001A1	−0.05A2	
t	−2.837	−0.023	−0.074	
sig.	0.090	0.8211	0.9420	
	−1.13B1	+0.55B2	+0.792C1	
t	−1.20	0.852	0.567	
sig.	0.2420	0.4030	0.5760	

$$R^2 = 0.898 \qquad \mathrm{RMSE} = 0.2155 \quad (10)$$

Using forward−backward stepwise MLR, four variables were retained by the model: Q_N, μ, −B2, and $-E_{HOMO}$. The equation and the determination coefficient became

$\lg 1/LC_{50} =$	1.209	$-2.63Q_N$	-0.294μ
t		−4.852	−2.975
sig.		0.0000	0.0002
	+0.393B2	$+0.484E_{HOMO}$	
t	2.029	−4.175	
sig.	0.0060	0.0060	

$$R^2 = 0.863 \qquad \mathrm{RMSE} = 0.2112 \quad (11)$$

4.1.2. Explanatory Capacity. MLR partial coefficients generally give an indication of environmental reality. Each coefficient is the partial derivative of the response of the model with respect to the variable of that coefficient; therefore, the influence of each variable can be assessed by checking the final values of the regression coefficients.

Concerning the complete model, only three variables contribute significantly. They were, in order of importance, Q_N, μ, and $-E_{HOMO}$. The relationship between $\lg 1/LC_{50}$ and $-E_{HOMO}$ was positive, while for Q_N and μ, the relationships with $\lg 1/LC_{50}$ were negative. The stepwise model retained four significant variables, which were Q_N, μ, B2, and $-E_{HOMO}$. So the stepwise procedure did not lead to a very different conclusion from the complete one.

4.2. Artificial Neural Network Models. *4.2.1. Predictive Capacity.* Using ANNs with different structures to model, the results were $R^2 \geq 0.962$ and RMSE = 0.1627 for the learning set and $R^2 \geq 0.972$ and RMSE = 0.1512 for the testing set. The results were as good in the learning set as in the testing set. The ANN structures could then be used in the complete database for a relativel importance analysis. The results were $R^2 \geq 0.974$ and RMSE = 0.1812, affirming the predictive quality of the ANN model.

4.2.2. Explanatory Capacity. Figure 1 presents the derivative plots of the ANN output with respect to each variable (the x coordinate was indicated by original data of each variable), and from these, we can know the effect of each variable on the outputs.

The partial derivative values of $\lg 1/LC_{50}$ with respect to Q_C were all positive except for those of ortho-nitroaniline and 2,4-dinitroaniline and were increasingly high for the high values of Q_C; an increase of Q_C led to an increase of $\lg 1/LC_{50}$. The higher the value of Q_C was, the more important its change contributed to $\lg 1/LC_{50}$ (Figure 1a).

The partial derivative values of $\lg 1/LC_{50}$ with respect to Q_N were all positive, but the total trend was not clear (Figure 1b). Two clusters appear, a smaller one in the left upper side of the plot and, then, a series of points at about 0.45 on the x axis with absolutely no trend on the y. But, points appearing on the left side are all nitroanilines, which might suggest a mechanism different from that of other compounds.

The partial derivative values of $\lg 1/LC_{50}$ with respect to Q_{NO_2} were all positive as well and were near zero for high values of Q_{NO_2} (Figure 1c). $\lg 1/LC_{50}$ increased with the increase of Q_{NO_2} until it became constant at high values of Q_{NO_2}.

The partial derivative values of $\lg 1/LC_{50}$ with respect to HF were almost all positive; an increase of HF led to an increase of $\lg 1/LC_{50}$ (Figure 1d).

The partial derivative values of $\lg 1/LC_{50}$ with respect to μ were all negative, but the total trend was not clear (Figure

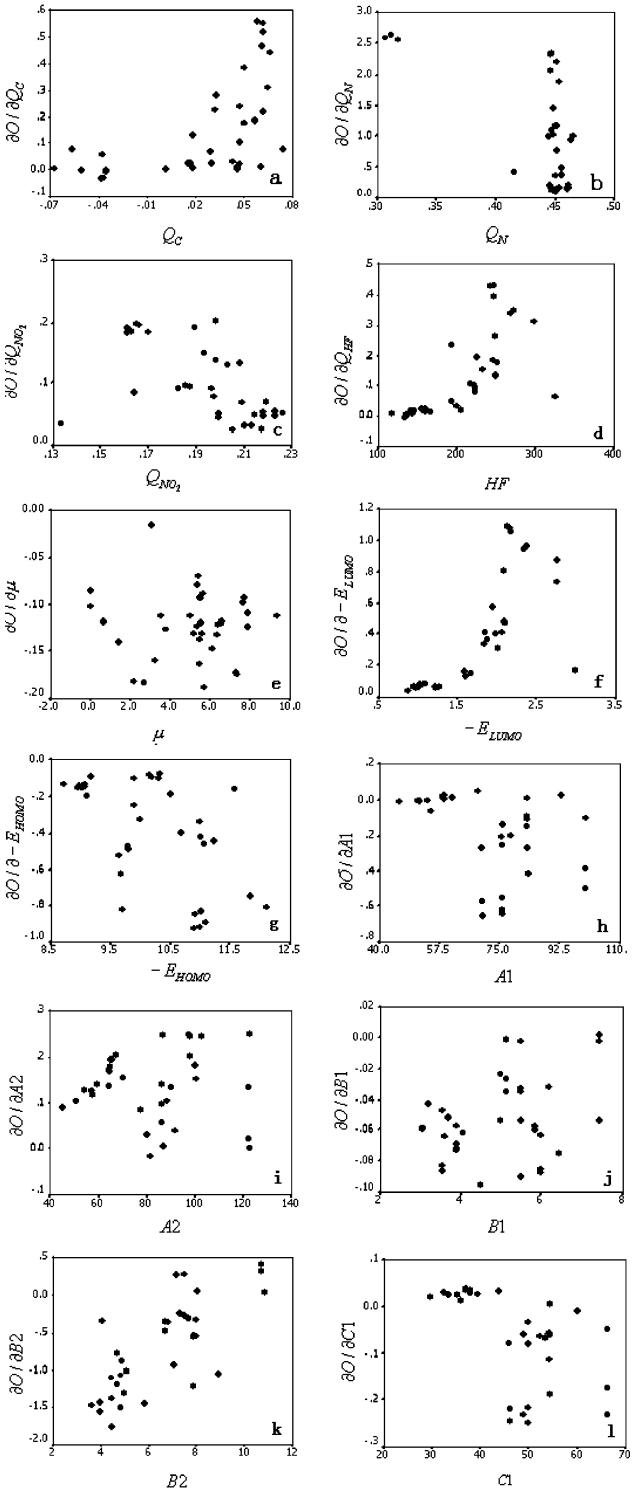


Figure 1. Partial derivatives of the ANN model response ($lg1/LC_{50}$) with respect to each independent variable. (a) Q_C, (b) Q_N, (c) Q_{NO_2}, (d) HF, (e) μ, (f) $-E_{LUMO}$, (g) $-E_{HOMO}$, (h) A1, (i) A2, (j) B1, (k) B2, (l) C1.

1e). Different values of μ had little influence on the output of ANNs, which suggests that μ is insignificant for the output, which is validated below.

The partial derivative values of $lg1/LC_{50}$ with respect to $-E_{LUMO}$ were all positive and increased with the increase of values of $-E_{LUMO}$ (Figure 1f). The higher the value of $-E_{LUMO}$ was, the more important its change contributed to $lg1/LC_{50}$.

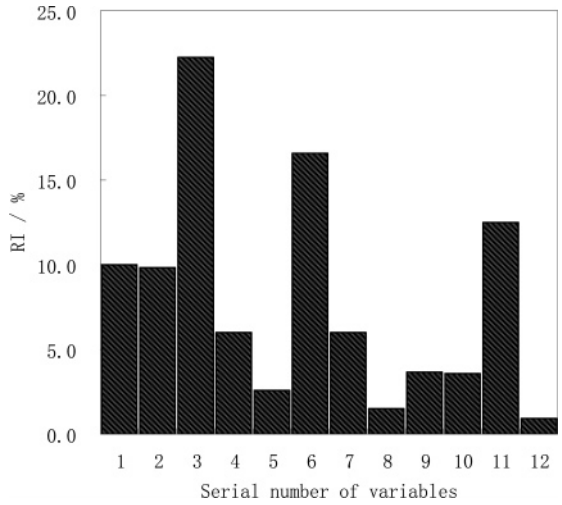


Figure 2. Contribution of the 12 independent variables (Q_C, Q_N, Q_{NO_2}, HF, μ, $-E_{LUMO}$, $-E_{HOMO}$, A1, A2, B1, B2, and C1) used in the ANN model for $lg1/LC_{50}$.

The partial derivative values of $lg1/LC_{50}$ with respect to $-E_{HOMO}$ were all negative; an increase of $-E_{HOMO}$ led to a decrease of $lg1/LC_{50}$ (Figure 1g).

The partial derivative values of $lg1/LC_{50}$ with respect to A1 were all negative except for that of nitroaniline (Figure 1h).

Most partial derivative values of $lg1/LC_{50}$ with respect to A2 were positive, but the total trend was not clear (Figure 1i). Different values of A2 had little influence on $lg1/LC_{50}$, which suggests that A2 is insignificant for the output, which is validated below.

The partial derivative values of $lg1/LC_{50}$ with respect to B1 were all negative, but the total trend was not clear (Figure 1j).

Almost all of the partial derivative values of $lg1/LC_{50}$ with respect to B2 were negative except for those of mononitrobenzene and nitroaniline; $lg1/LC_{50}$ decreased with an increase of B2 (Figure 1k).

The partial derivative values of $lg1/LC_{50}$ with respect to C1 were all positive except for that of nitroaniline (Figure 1l).

Figure 2 presents the relative contributions resulting from the application of the partial derivative method and the weight method [the x coordinate is indicated by the serial number (see Table 2)]. The method was then very stable, whatever the model, and had a low confidence interval.

Q_{NO_2} was the highest contributing variable (>22%), followed by $-E_{LUMO}$ (>16%) and B2 (>12%), and then came Q_C and Q_N with contributions as high as 10%. However, the contributions of the other variables were low, and the difference between HF, $-E_{HOMO}$, A2, and B1 was not significant; then came μ, A1, and at last C1. The contribution of each variable obtained by ANNs was very different from that of MLR.

5. DISCUSSION

MLR has been thoroughly statistically tested and is universally known. Its success comes from its easy use. However, its inability to take into account nonlinear relationships between the dependent variable and each independent

variable is its principal drawback. That is why the use of ANNs is wholly justified in QSAR modeling.

In this work, explanatory methods were developed, like contributions adding power to ANNs in their explanatory capacity, to clarify the "black-box" approach of ANNs. ANN models are able to make perfect predictions and explanations. The prediction results were satisfactory, affirming, then, a good prediction of the toxicities of nitro-aromatic compounds to the fathead minnow, which was better with ANNs than with MLR, confirming the nonlinearity of the relationship between the variables.

Nitro-aromatic compounds are a class of important genotoxic compounds. Early studies[24] suggested that nitro-aromatic compounds required a metabolic reduction of the nitro group to exhibit mutagenic activity; the ultimate mutagenic metabolites are arylhydroxylamine (ArNHOH) or the corresponding hydroxamic esters (ArNHOR), in both of which the nitrogen atoms are electrophilic centers that react with nucleophilic centers in the cellular DNA of the target organism to form an adduct.

The electronic effects of the additional substituents on the benzene ring are anticipated to affect the genotoxicity of studied compounds, which can be described by variables B1 and B2.

Because the metabolic activation procedure of $ArNO_2$ is a reduction course, the reducibility of $ArNO_2$ is also anticipated to affect the genotoxicity. $-E_{LUMO}$ and $-E_{HOMO}$ can respectively describe the capacity of accepting or offering electrons.

HF is the heat of formation of molecules per mole. The greater the value of HF is, the lower the stability of the compounds is, thus, resulting in higher reactivity.

To identify the action spot of compounds, Q_{NO_2}, Q_C, and Q_N were calculated to discuss their contributions to the toxicity of the compounds to the fathead minnow.

Furthermore, additional information concerning the structure of the compounds was also considered, such as A1, A2, μ, and C1, to gain further insight into the mechanism of toxicity occurring between the nitro-aromatic compounds and the fathead minnow.

On the basis of Figure 2, Q_{NO_2} was the highest contributing variable to the toxicity of the compounds. The greater its value, the greater the net positive charge of it, and Figure 1c showed that $lg1/LC_{50}$ increased with an increase of Q_{NO_2} until it was near zero at high values of Q_{NO_2}. It suggested that nitro-aromatic compounds with different numbers of nitro groups first reacted, the reaction occurring at the nitrogen atom of the nitro group with maximal net charge, acting as an electrophilic center, with the cellular DNA of the fathead minnow, acting as a nucleophilic center. It also suggested that for high values of Q_{NO_2}, the effect of the electrophilic center on the fathead minnow could be negligible. From a biochemical view, this was more reasonable than Q_N, the most important variable obtained by the MLR method, and the effect of Q_N on $lg1/LC_{50}$ was negative, contrary to reality.

$-E_{LUMO}$ was the second highest contributing variable. Figure 1f showed that $lg1/LC_{50}$ increased with an increase of $-E_{LUMO}$, and the higher the value of $-E_{LUMO}$ was, the more important its change contributed to $lg1/LC_{50}$. It suggested that as well as enhancing the capacity of accepting electrons, the nitro-aromatic compounds' toxicity to the fathead minnow was great. Also, it showed that nitro-aromatic compounds acted as the electrophile in the reaction with the cellular DNA of the fathead minnow. A higher $-E_{LUMO}$ could make the reaction easier and resulted in greater toxicity. On the contrary, in the MLR equation, it is not a significant variable; however, μ, not a significant variable from a biochemical view, was the second highest contributing variable and had provided poor explanatory capacity.

The third most important variable was B2, which described the sum of electronic information between spacing atoms. Figure 1k showed that $lg1/LC_{50}$ decreased with an increase of B2. It suggested that the electronic interaction of the substituents also influenced the reactivity with the cellular DNA of the fathead minnow. Electron-donating substituents such as NH_2, CH_3, and OH (Table 1) might delocalize the positive charge on the nitrogen atom of ArN^+; the energy of the transition state was increased, and the activation energy of the mutagenic reaction was raised.[25] Therefore, the reaction was decelerated and the genotoxic responses were declined, which provided support for the following supposed mechanism of a mutagenic reaction between DNA and the ultimate mutagenic compound ArNHOH (or ArNHOR): the reaction was initiated by the departure of the hydroxyl group (or alkoxyl group) to form ArN^+ as the transition state, which might immediately react with the nucleophilic centers in cellular DNA to form adducts.

Q_C and Q_N followed B2 in the order of relative importance in Figure 2. Q_C showed that the net positive charge of carbon atoms joined to the nitro groups increased the capacity of the reaction spot as well and had an active influence on the genotoxic responses. From the $lg1/LC_{50}$ values listed in Table 1, we can see that all the dinitro compounds are genotoxic, while the mononitro ones are less toxic. Obviously, the number of the nitro substituents plays an important role in the genotoxic activities of nitro-aromatic compounds. Q_N also contributed to the toxicity of compounds importantly, as well as Q_{NO_2}, as shown above. But in the MLR model, Q_C and Q_N both had negative responses to $lg1/LC_{50}$, but not active responses, which was very different from ANN and far from reality.

Furthermore, A1 and A2 were not significant variables in this study, which indicated that the volume of molecules played an unimportant role in the genotoxic activities of nitro-aromatic compounds, as did $-E_{HOMO}$, μ, HF, and C1.

On the basis of the partial derivative method and parts a, b, h, k, and l of Figure 1, nitroanilines differed from the other compounds in the way of effect on the genotoxic response to the fathead minnow. The existence of NH_2 might result in this consequence, and the mechanism of the mutagenic reaction with cellular DNA should be studied more.

The explanatory methods of ANN, thus, help to identify factors affecting the fathead minnows and how these factors contribute to them. Comparing the partial derivative method and the weight method, they are more coherent from a computational point of view. In fact, the two methods of contribution analysis of the input variables seem to be very interesting to use, but it is important to underline the need for a biochemical opinion regarding the ranking of importance of the inputs and their mode of action on the output.

In QSAR studies, it is important to predict phenomena that occur in the studied compounds. Finally, when ANNs with five input neurons, that is, Q_{NO_2}, $-E_{LUMO}$, B2, Q_C, and Q_N, and five hidden neurons were used to model, the results were $R^2 \geq 0.954$ for the learning set and $R^2 \geq 0.960$ for the testing set. Thus, the results were $R^2 \geq 0.958$ for the complete database, affirming the predictive quality of the ANN model.

Thus, this ANN property is now well-understood. Adding new methods to ANNs and allowing the analysis of the contributions of the different variables will lead to gaining a better predictive and explanatory capacity than MLR and will help in understanding the biochemical phenomena and, finally, in finding solutions to act on it, restore it, and improve the environmental conditions for life.

6. CONCLUSIONS

The method based on axon connection weights and partial derivatives of the ANN output with respect to its input can identify variables that significantly contribute to network predictions and provide a variable selection method for ANNs. We have shown that, by extending this approach to ANNs, the "black box" mechanics of ANNs can be greatly illuminated.

Compared with MLR, and coupling this new explanatory power of neural networks with its strong predictive abilities, ANNs have a better predictive and explanatory capacity and promise to be a valuable quantitative tool to evaluate, understand, and predict biochemical toxicity in QSAR studies.

Furthermore, the aforementioned results and discussion lead us to conclude that autocorrelation topological indices proposed by our research center can be used successfully for modeling and estimating the toxicity of nitro-aromatic compounds to the fathead minnow.

ACKNOWLEDGMENT

This work received financial support from the Scientific Research Foundation of Harbin Institute of Technology, under contract 2002-58.

Note Added after ASAP Publication. This article was released ASAP on October 13, 2005 without an acknowledgment section. The correct version was posted on October 19, 2005.

REFERENCES AND NOTES

(1) Lek, S.; Delacoste, M.; Baran, P.; Dimopoulos, I.; Lauga, J.; Aulagnier, S. Application of neural networks to modeling nonlinear relationships in ecology. *Ecol. Modell.* **1996**, *90*, 39−52.

(2) Klaus, L. E. The use of neural networks in QSAR for acute aquatic toxicological endpoints. *J. Molec. Struct.* **2003**, *622*, 85−95.

(3) González-Arjona, D.; López-Pérez, G.; Gustavo González, A. Nonlinear QSAR modeling by using multilayer perceptron feed-forward neural networks trained by back-propagation. *Talanta* **2002**, *56*, 79−90.

(4) Gagné, F.; Biaise, C. Predicting the toxicity of complex mixtures using artificial neural networks. *Chemosphere* **1997**, *35*, 1343−1363.

(5) Wienke, D.; Domine, D.; Buydens, L.; Devillers, J. *Neural Networks in QSAR and Drug Design*; Academic Press: London, 1996; pp 50−77.

(6) Lek, S.; Guégan, J. F.; Artificial neural networks as a tool in ecological modeling: an introduction. *Ecol. Modell.* **1999**, *120*, 65−73.

(7) Paruelo, J. M.; Tomasel, F. Prediction of Functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. *Ecol. Modell.* **1997**, *90*, 39−52.

(8) Anderson, J. A. *An Introduction to Neural Networks*; Cosmao, N., Ed.; MIT Press: Cambridge, MA, 1995; pp 31−55.

(9) Bishop, C. M. *Neural Networks for Pattern Recognition*; Balland, L., Ed.; Clarendon: Oxford, U. K., 1995; pp 123−143.

(10) Ripley, B. D. *Pattern Recognition and Neural Networks*; Arie, C., Ed.; Cambridge University Press: Cambridge, MA, 1996; pp 67−89.

(11) Garson, G. D. Interpreting neural-network connection weight. *Artif. Intell.* **2001**, *6*, 47−51.

(12) Guo, Z.; Uhrig, R. E. Using modular neural networks to monitor accident conditions in nuclear power plants, SPIE. *Appl. Artif. Neural Networks* **1992**, *1709*, 505−516.

(13) Kim, S. H.; Yoon, C.; Kim, B. J. Structural monitoring system based on sensitivity analysis and a neural network. *Comput.-Aided Civil Infrastruct. Eng.* **2000**, *155*, 309−318.

(14) Liong, S. Y.; Lim, W. H.; Paudyal, G. N. River stage forecasting in Bangladesh: neural network approach. *J. Comput. Civil Eng.* **2000**, *14*, 1−18.

(15) Yu, X. J.; Wang, P.; Long, M. C. Quantitative relationship between the biodegradeability and vertex degree autocorrelation topological indexes of organic chemicals. *J. Environ. Sci.* **2000**, *20* (supplement), 93−96 (in Chinese).

(16) Hall, L. H. The study for the toxicity of nitrobenzene to the fathead minnow. *Environ. Toxicol. Chem.* **1989**, *18*, 431−440.

(17) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*. June, C., Ed.; Research Studies: Chichester, U. K., 1987; pp 232−256.

(18) Karelson, M.; Lobanov, V. S. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, *96*, 1027−1043.

(19) Todechimi, R.; Consonni, V. *Handbook of Molecular Descriptors*; Weinheim: Weinheim, Germany, 2000; pp 45−97.

(20) Smith, M. *Neural Networks for Statistical Modeling*; Van Nostrand Reinhold: New York, 2000; pp 132−154.

(21) Ripley, B. D. *Neural Network: Artificial Intelligence and Industrial Applications*; Springer: London, 1995; pp 123−167.

(22) Cheng, B.; Titterington, D. M. Neural network: a review from a statistical perspective. *Stat. Sci.* **1994**, *9*, 32−54.

(23) Goh, A. T. Back-propagation neural networks for modeling complex systems. *Artif. Intell. Eng.* **1994**, *9*, 143−151.

(24) McCoy, E. C.; Rosencranz, H. S. Genotoxic study for nitro-aromatic compounds. *Biochem. Biophys. Res.* **1990**, *108*, 1362−1373.

(25) Yang, J. Genotoxicity of substituted nitro benzenes and the quantitative structure−activity relationship. *J. Environ. Sci.* **1996**, *8*, 103−109 (in Chinese).

CI050079X