ARTICLE

Neural Network Based on Quantum Chemistry for Predicting Melting Point of Organic Compounds

Juan A. Lazzús*

Departamento de Física, Universidad de La Serena, Casilla 554, La Serena, Chile

(Dated: Received on September 9, 2008; Accepted on January 6, 2009)

The melting points of organic compounds were estimated using a combined method that includes a backpropagation neural network and quantitative structure property relationship (QSPR) parameters in quantum chemistry. Eleven descriptors that reflect the intermolecular forces and molecular symmetry were used as input variables. QSPR parameters were calculated using molecular modeling and PM3 semi-empirical molecular orbital theories. A total of 260 compounds were used to train the network, which was developed using MatLab. Then, the melting points of 73 other compounds were predicted and results were compared to experimental data from the literature. The study shows that the chosen artificial neural network and the quantitative structure property relationships method present an excellent alternative for the estimation of the melting point of an organic compound, with average absolute deviation of 5%.

Key words: Melting point, Quantitative structure-property relationship, Artificial neural network, Quantum chemistry

I. INTRODUCTION

The melting point (mp) is the temperature at which a solid becomes a liquid and the freezing point is the temperature at which a liquid solidifies, and for all practical purposes the two can be considered to be identical [1]. From a thermodynamics point of view, at mp the change in Gibbs free energy (ΔG) is zero, because the transition temperature is related to the enthalpy ($\Delta H_{\rm tr}$) and entropy of transition ($\Delta S_{\rm tr}$) by the following relationship:

$$T_{\rm tr} = \frac{\Delta H_{\rm tr}}{\Delta S_{\rm tr}} \Rightarrow T_{\rm m} = \frac{\Delta H_{\rm m}}{\Delta S_{\rm m}} \qquad (1)$$

The enthalpy of melting ($\Delta H_{\rm m}$) of an organic molecule is assumed to be dependent upon the interactions between its molecular fragments. The entropy of melting ($\Delta S_{\rm m}$) of an organic molecule is the sum of its positional, rotational, and conformational entropies [2,3]. Melting phenomena happen when the Gibbs free energy of the liquid becomes lower than the solid for that substance.

From the scientific and industrial point of view, a fundamental understanding of the chemical, physical and thermodynamic properties of substances should be known before their application to several processes. For instance, knowledge of some basic properties is useful in the area of fluid properties estimation, thermodynamic property calculations, and phase equilibrium, among others.

The melting point is one of the most widely used fundamental physical properties. It finds applications in chemical identification, purification, and calculation of a number of other physicochemical properties [2,3]. Also, several correlations of physicochemical properties make use of melting temperature [4].

Melting point prediction has a long history, starting in 1884 [1]. Mainly, prediction methods for this property can be categorized as group contribution methods (GCM) and quantitative structure-property relationship (QSPR).

Joback and Reid [5] reevaluated Lydersen's GCM [6], added several new functional groups, and determined new contribution values. For the case of mp the method of Joback and Reid includes one of the simplest methods available. Later, Constantinou and Gani developed an advanced GCM method based on the UNIFAC groups but enhanced by allowing for more sophisticated functions of the desired properties and by providing contributions at a second order level [7].

Yalkowsky *et al.* have explored connections between melting point and normal boiling point as well as have proposed correlations for mp [2,3]. The method consists of both group contributions which are additive and molecular descriptors which are not additive [4].

In all these methods, the property of a compound is calculated by summing up the contributions of certain defined groups of atoms, considering at the same time

*Author to whom correspondence should be addressed. E-mail: jlazzus@dfuls.cl, Tel.: +56-51-204128, FAX: +56-51-206658

TABLE I Reported QSPR models for predicting melting points of organic compounds.

Author	Compounds	Method[a]	N	R^2	Deviation[b]
Bhattacharjee *et al.*[8]	Halomethanes	LR	30	0.676	47.8
Dearden [9]	Anilines	MLR	42	0.897	23.7
Medić-Sărić *et al.* [10]	Pyrazolinones	MLR	17	0.891	15.0
Katritzky and Gordeeva [11]	Aldehydes	MLR	72	0.838	
	Amines	MLR	54	0.795	
	Ketones	MLR	52	0.865	
Charton and Charton [12]	Diverse	MLR	178	0.937	24.0
Cherqaoui *et al.* [13]	Alkanes	ANN	150	0.956	8.1
Tu [14]	Hydrocarbons	MLR	307		7.4
Murugan *et al.* [15]	Pyridines and piperidines	MLR	141	0.831	
Charton and Charton [16]	Diverse	MLR	366	0.919	17.9
Todeschini *et al.* [17]	Polycyclic aromatic hydrocarbons	MLR	79	0.895	
Piazza *et al.* [18]	Halobenzenes	LR	25	0.870	21.0
Pogliani [19]	Alkanes	MLR	17	0.901	14.0
	Caffeines	MLR	12	0.964	20.5
Pogliani [20]	Alkanes	MLR	17	0.895	13.6
	Amino acids	MLR	20	0.760	22.5
Todeschini and Gramatica [21]	Polycyclic aromatic hydrocarbons	MLR	79	0.887	35.1
Todeschini and Gramatica [22]	Chlorobenzenes	MLR	13	0.934	18.1
Chiorboli *et al.* [23]	Halobenzenes and halotoluenes	MLR	92	0.862	22.3
Pogliani [24]	Amino acids	MLR	20	0.929	12.7
Katritzky *et al.* [25]	Benzenes	MLR	443	0.837	30.2
Todeschini *et al.* [26]	Diverse	MLR	94	0.834	32.8
Gramatica *et al.* [27]	Polychlorinated biphenyls	MLR	82	0.820	21.3
Jain *et al.* [3]	Aromatics, non hydrogen-bonded	MLR	338		23.1
Pogliani [28]	Diverse	MLR	62	0.856	18.0
Pogliani [29]	Alkanes	MLR	17	0.693	20.0
Gironés *et al.* [30]	Alcohols	LR	15	0.947	14.9
Bergström *et al.* [31]	Drugs	PLS	277		
Karthikeyan *et al.* [32]	Diverse	ANN	4173	0.661	37.6
Modarresi *et al.* [33]	Drugs	MLR	323	0.673	40.4
	Drugs	GA	323	0.660	41.1

[a] MLR=multiple linear regression, LR=linear regression, PLS=partial least squares, GA=genetic algorithms, ANN=artificial neural network.
[b] Standard, root mean square error, or mean absolute error.

the number frequency of each group occurring in the molecule.

Alternatives to GCM's have recently appeared. These other techniques of estimating of *mp* are based on molecular descriptors, which are not normally measurable. These QSPR parameters are usually obtained from on-line computation of the structure of the whole molecule using molecular mechanics or quantum mechanical methods. Thus, no tabulation of descriptor contributions is available in the literature even though the weighting factors for the descriptors are given. Estimates require access to the appropriate computer software to obtain the molecular structure and properties and then the macroscopic properties are estimated with the QSPR relations. It is common that different methods use different computer programs [4]. Table I lists selected works on QSPR applications to estimate *mp* that have been published in the literature during recent years.

Artificial neural networks (ANN) are accepted as the most powerful nonlinear technique in QSPR application [34]. The neural network modeling in QSPR has been applied to most physicochemical properties, for which suitable experimental data can be found in the literature.

Taskinen and Yliruusi presented a complete list of properties that have been analyzed in the literature using different approaches of ANNs [35]. Properties such

DOI:10.1088/1674-0068/22/01/19-26

as boiling point, critical temperature, critical pressure, vapor pressure, heat capacity, enthalpy of sublimation, heat of vaporization, density, surface tension, viscosity, thermal conductivity, and acentric factor were thoroughly reviewed. One notable exception according to these authors is the melting point [35]. To the best of the author's knowledge there is no application for melting point prediction that includes a heterogeneous set of compounds, such as the one presented here, and certainly there is no publication on the prediction of this property for diverse substances using an ANN+QSPR method.

In this work, the melting point of organic compounds was estimated using a combined method that includes a backpropagation neural network and a quantitative structure-property relationships (QSPR) method. Eleven parameters that reflect the intermolecular forces and molecular symmetry, were calculated using molecular modeling and PM3 semi-empirical molecular orbital theories, and were given for modeling this relevance property.

II. THE NEURAL NETWORK

Neural networks (NN) are accepted as the most powerful nonlinear technique in QSAR and QSPR modeling [34], and many models of neural networks have been used in the estimation of thermodynamic properties with QSPR methods [36-38]. In this work a feedforward backpropagation neural network was used, the type that is very effective to represent nonlinear relationships among variables. The network, programmed with the software MatLab, consists of a multilayer network in which the flow of information spreads forward through the layers while the propagation of the error flows backwards. In this process, the network uses some factors called "weights" (w_i) to quantify the influence of each fact and of each variable. There are two main states in the operation of an ANN: the learning and the validation. The learning or training is the process in which an ANN modifies the weights in response to entered information.

The most basic architecture normally used for this type of application involves a neural network consisting of three layers [35]. The input layer contains one neuron (node) for each QSPR parameter. The output layer has one node generating the scaled estimated value of the *mp*. The number of hidden neurons needs to be sufficient to ensure that the information contained in the data utilized for training the network is adequately represented. There is no specific approach to determine the number of neurons for the hidden layer, so many alternative combinations are possible. The optimum number of neurons was determined by adding neurons in systematic form during the learning process. The ANN was trained by the Levenberg-Marquardt algorithm. In a previous work, Karelson *et al.* show that the Levenberg-Marquardt method was superior over the conjugate gradient and delta rule methods as regarding convergence and prediction for application of QSAR/QSPR modeling [39].

This ANN program considers the reading of the necessary data organized in an Excel file: melting point experimental data for 260 compounds are used to train the network. To distinguish between the different physical and chemical properties of the substances considered in this study, so the network can discriminate and learn in optimum form, the following properties are considered QSPR parameters that reflect the intermolecular forces and molecular symmetry.

The steps to calculate the output parameter, using the input parameters, are the following ones.

The net inputs are calculated (N) for the hidden neurons coming from the input neurons. For a hidden neuron:

$$N_j^h = \sum_i^n w_{ij}^h p_i + b_j^h \tag{2}$$

where the p corresponds to the vector of the inputs of the training, j is the hidden neuron, w_{ij} is the weight of the connection among the input neurons with the hidden layer, and the term b_j corresponds to the bias of the neuron j of the hidden layer, reached in its activation. Starting from these inputs the outputs are calculated (y) of the hidden neurons, using a transfer function f^h associated with the neurons of this layer.

$$y_j^h = f_j^h \left(\sum_i^n w_{ij}^h p_i + b_j^h \right) \tag{3}$$

Similar calculations are carried out to obtain the results of each neuron of the following layers until the output layer.

To minimize the error the transfer function f should be differentiable. In the ANN two types of transfer function were used.

The hyperbolic tangent function (*tansig*) in the hidden layer, defined by the equation:

$$f\left(N_{jk}\right) = \frac{\mathrm{e}^{N_{jk}} - \mathrm{e}^{-N_{jk}}}{\mathrm{e}^{N_{jk}} + \mathrm{e}^{-N_{jk}}} \tag{4}$$

and the lineal function (*purelin*) in the output layer, defined as:

$$f\left(N_{jk}\right) = \left(N_{jk}\right) \tag{5}$$

All the neurons of the ANN have an associate activation value for a given input pattern, the algorithm continues finding the error that is presented for each neuron, except those of the input layer. After finding the value of the gradient of the error the weights of network are actualized, for all layers:

$$w_{kj}\left(t+1\right) = w_{kj}(t) - 2\alpha\delta_k^{\mathrm{o}} \tag{6}$$

$$b_k\left(t+1\right) = b_k(t) - 2\alpha\delta_k^{\mathrm{o}} \tag{7}$$

$$w_{ji}\left(t+1\right) = w_{ji}(t) - 2\alpha\delta_j^h p_i \tag{8}$$

DOI:10.1088/1674-0068/22/01/19-26

$$b_j\,(t+1) = b_j(t) - 2\alpha\delta_j^h \tag{9}$$

This process repeats for the total number of patterns used for training. For a successful process the objective of the algorithm is to adapt all of the weights and biases of the ANN minimizing the total mean squared error.

$$Ep^2 = \frac{1}{2}\sum_k {\delta_k}^2 \tag{10}$$

$$E^2 = \frac{1}{2}\sum_{p=1} (Ep)^2 \tag{11}$$

Figure 1 presents a block diagram of the program developed and written in MatLab M-file.

A. Data used and training

In this work, 260 experimental data of *mp* were used to train the ANN, introduced as entrance parameters that reflect the intermolecular forces and molecular symmetry. The output parameter was *mp*. Several molecular descriptors for each compound were determined from knowledge of the chemical structure, and used for discriminating among the different substances.

Descriptors are defined as numerical characteristics associated with chemical structures. These are derived proceeding from the chemical constitution, topology, geometry, wave function, potential energy surface, or some combination of these items for a given chemical structure.

Molecular topological descriptors, generated using molecular modeling, included valence molecular connectivity indices ($^m\chi^v$), Wiener index (W), Balaban's J index and the Kappa shape indices ($^m\kappa$).

Quantum chemical descriptors, derived from the PM3 semi-empirical molecular orbital theory were also calculated; these included average molecular polarizability, dipole moment, ionization potential, molecular orbital levels, molecular weight, moment of inertia, heat of formation, total energy, electronic energy, nuclear-nuclear energy, and energy components partitioned into the individual one-center and two-center terms.

The next step with the data sets was to determine a reduced set of parameters that provide the effective model. The selection of the optimal set of input parameters was carried out based on a heuristic method of analysis. This heuristic method for descriptor selection proceeds with a pre-selection of descriptors by sequentially eliminating descriptors that do not match any of the following criteria: (i) Fisher F-criteria greater than 1.0; (ii) R^2 value less than a value defined at the start; (iii) Student's t-criterion less than a defined value; (iv) duplicate descriptors having a higher squared intercorrelation coefficient than a predetermined level (retaining the descriptor with higher R^2 with reference to the property). The descriptors that remain are then listed

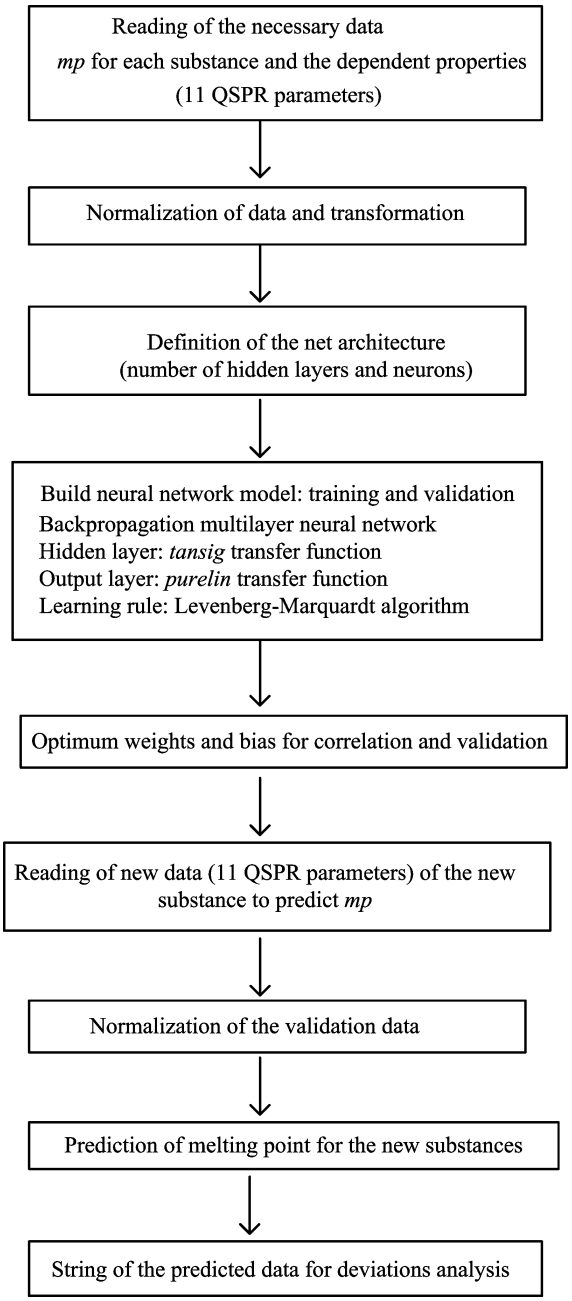

FIG. 1 Flow diagram for the ANN program developed for this work.

in decreasing order of correlation coefficients when used in global search for 2-parameter correlations. Each significant 2-parameter correlation by F-criteria is recursively expanded to an n-parameter correlation till the normalized F-criteria remains greater than the startup value. The best N correlations by R^2, as well as by F-criterion, are saved. With this heuristic method, elimination of ill-defined, inter-correlated and poorly correlated descriptors led to a set of eleven QSPR descriptors [40].

The final set of 11 input parameters that includes:

DOI:10.1088/1674-0068/22/01/19-26

molecular weight (M)[41], dipole moment as sum of both point charge μ [42], average molecular polarizability α [42,43], resonance energy $E_{\rm R}$ [44], exchange energy $E_{\rm exc}$ [44], electron-nuclear attraction $E_{\rm ne}$ [44], nuclear-nuclear repulsion $E_{\rm nn}$ [44], and the first-, second-, third- and fourth-order valence connectivity indices ${}^1\chi^v$, ${}^2\chi^v$, ${}^3\chi^v$, ${}^4\chi^v$ [45,46]. (The values of the parameters and mp for all substances considered in the study can see the supporting information). The definition for these descriptors is given in Eq.(12)-Eq.(20).

$$M = \sum_i m_i \tag{12}$$

m_i is atomic weights of constituent atoms of a molecule.

$$\mu = -\sum_{i=1}^{\rm occ} \int_V \phi_i \hat{r} \phi_i {\rm d}v + \sum_{a=1}^{M} Z_a \vec{R}_a \tag{13}$$

ϕ_i is molecular orbitals, $\hat{r}$ is electron position operator, Z_a is a-th atomic nuclear charge, $\vec{R}_a$ is position vector of a-th atomic nucleus.

$$\mu' = \mu + \alpha \mathbf{E} + \frac{1}{2}\beta\ \mathbf{E}^2 + ... \tag{14}$$

α is molecular polarizability, μ is permanent dipole moment of the molecule, μ' is induced dipole moment of the molecule, $\mathbf{E}$ is external electric field.

$$E_{\rm R}({\rm AB}) = \sum_{\mu\in {\rm A}} \sum_{\nu\in {\rm B}} P_{\mu\nu}\beta_{\mu\nu} \tag{15}$$

A and B are atomic species, $P_{\mu\nu}$ is density matrix elements over atomic basis $\{\mu\nu\}$, $\beta_{\mu\nu}$ resonance integrals on atomic basis $\{\mu\nu\}$.

$$E_{\rm exc}({\rm AB}) = \sum_{\mu,\nu\in {\rm A}} \sum_{\lambda,\sigma\in {\rm B}} P_{\mu\lambda}P_{\nu\sigma} \langle \mu\lambda \mid \nu\sigma \rangle \tag{16}$$

$P_{\mu\lambda}$ and $P_{\nu\sigma}$ are density matrix elements over atomic basis $\{\mu\lambda\nu\sigma\}$, $\langle \mu\nu \mid \lambda\sigma \rangle$ are electron repulsion integrals on atomic basis $\{\mu\lambda\nu\sigma\}$.

$$E_{\rm ne}({\rm tot}) = \sum_{\rm A} E_{\rm ne}({\rm A}) \tag{17}$$

$E_{\rm ne}({\rm A})$ is electron-nuclear attraction energy for atom A.

$$E_{\rm nn}({\rm tot}) = \sum_{\rm A} E_{\rm nn}({\rm A}) \tag{18}$$

$E_{\rm nn}({\rm A})$ is nuclear-nuclear repulsion energy for atom A.

$$^m\chi^v = \sum_{i=1}^{N_s} \prod_{k=1}^{m+1} \left(\frac{1}{\delta_k^v}\right)^{1/2} \tag{19}$$

$$\delta_k^v = \frac{(Z_k^v - H_k)}{(Z_k - Z_k^v - 1)} \tag{20}$$

δ_k^v is valence connectivity for the k-th atom in the molecular graph, Z_k is the total number of electrons in the k-th atom, Z_k^v is the number of valence electrons in the k-th atom, H_k is the number of hydrogen atoms directly attached to the k-th non-hydrogen atom, m=1, 2, 3, and 4 are atomic valence connectivity indices.

The experimental data of mp was taken from the DIPPR data base [47]. In this work, mp covered a wide range: 85 K to 531 K for organic compounds. In addition to that, the substances included in the study have very different physical and chemical characteristics: from low molecular weight substances such as ethene (M=28) to high molecular weight substances such as 1,1,2,2-tetrabromoethane (M=346); or from non-polar substances (μ=0) such as benzene and naphthalene, to highly polar substances such as 2-chloro-1-nitrobenzene (μ=5.4). Thus, the problem is not straightforward and most probably is one of the reasons why the melting point has not been treated previously using neural network as proposed in this work.

Once the training was successfully done and the optimum network architecture was determined, input data (11 QSPR parameters) of 73 compounds not used in the training process were fed to the ANN and the melting point was predicted. Several network architectures were tested to select the most accurate scheme. Since no additional information about the recommended number of neurons has been found for the calculation of properties for any type of substances, the optimum number of neurons was determined by trial and error. Figure 2 shows the average absolute deviation found in correlating the melting point of all compounds as a function of the number of neurons in the hidden layer. As observed in the figure, the optimum number of neurons in the hidden layer is between 9 and 12. The network that gave the lowest deviation during training was one with 11 parameters in the input layer, 11 neurons in the hidden layer, and one neuron in the output layer. For this architecture the average deviation during training

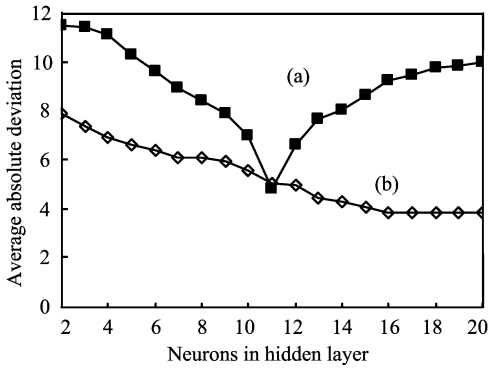

FIG. 2 Average absolute deviation found in correlating the melting point of all substances as function of the number of neurons in hidden layer. (a) During the prediction and (b) during training.

DOI:10.1088/1674-0068/22/01/19-26

TABLE II Overall minimum, maximum, and average deviations for the calculated melting point for all compounds using the backpropagation artificial neural network model.

ANN model	Training set	Prediction set	Total set
N° substances	260	73	333
$\%\Delta mp_{\min}$	0.0	0.0	0.0
$\%\Delta mp_{\max}$	19.3	−19.0	19.3
$\%\Delta mp$	0.4	0.1	0.3
$\lvert\%\Delta mp\rvert$	5.2	4.8	5.0
$N^\circ\lvert\%\Delta mp\rvert<10$	226	69	295
$N^\circ\lvert\%\Delta mp\rvert>10$	34	4	38

is 5.2% and during prediction is 4.8%.

The accuracy of the chosen final network was checked using the average relative deviation $\%\Delta mp$ and average absolute deviation $|\%\Delta mp|$ between the calculated value of melting point after training and the data from the literature. The deviations were calculated as:

$$\%\Delta mp = \frac{100}{N}\sum_{i=1}^{N}\left(\frac{mp^{\text{calc}} - mp^{\text{exp}}}{mp^{\text{exp}}}\right)_i \tag{21}$$

$$|\%\Delta mp| = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{mp^{\text{calc}} - mp^{\text{exp}}}{mp^{\text{exp}}}\right|_i \tag{22}$$

III. RESULTS AND DISCUSSION

Table II shows the overall minimum, maximum, and average deviations for all the substances using the proposed network 11-11-1. The results show that the ANN can be accurately trained and that the chosen architecture can estimate the melting point of organic compounds with enough accuracy. It gives lower deviations than any other model yet available in the literature: absolute average deviations less than 5.2% for the 260 compounds used in the training and absolute average deviations of less than 4.8% for the 73 compounds in the prediction step. For all substances (333 organic compounds) the deviations are below 20% and for 295 compounds of the total set the deviations are below 10%.

Once the best architecture was determined, the optimum weights required to carry out the estimate of melting point of organic compounds, were obtained. Table III shows the optimum weight and bias for the ANN 11-11-1.

Figure 3 shows a comparison between experimental and calculated values of melting point for organic compounds. Figure 3(a) shows a comparison during training between correlated and literature values of melting point. The correlation coefficient R^2 is 0.9637 and the slope of the curve m is 0.9706 (expected to be 1.0). Figure 3(b) shows a comparison during prediction between

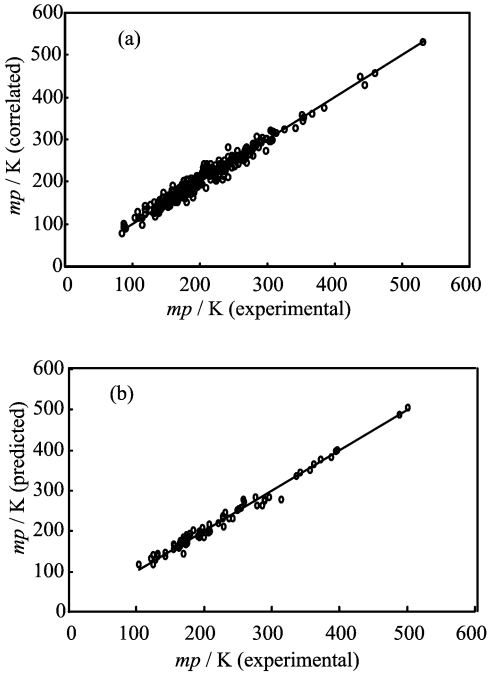

FIG. 3 Comparison between experimental (solid line) and calculated (circles) values of melting point. (a) During training and (b) during prediction.

predicted and literature values of mp. In this case, the correlation coefficient R^2 is 0.9847 and m (also expected to be 1.0) is 0.9718. For the total set: R^2 is 0.9703 and m is 0.9707.

Most published QSPRs relating mp to chemical structure are confined to limited and/or small sets of hydrocarbons, substituted aromatics, aldehydes, amines, and ketones [48]. In the literature there are very few works with data sets for diverse substances, because then mp is more difficult to predict [49]. In this work the heterogeneous set of compounds includes: aromatic and aliphatic hydrocarbons, halogens, polychlorinated biphenyls, mercaptans, sulfides, anilines, pyridines, alcohols, carboxylic acids, aldehydes, amines, ketones, and esters.

Table IV shows a comparison between some methods proposed in the literature for prediction of mp for diverse compounds and the ANN+QSPR method proposed in this work (this comparison considers only similar data sets). Charton and Charton [12-16], Todeschini *et al.* [26], and Pogliani [28], predicted mp with root mean square error (RMSE) higher than 10 K (17.9 K to 32.8 K), and the predictions with the proposed network show RMSE little higher than 10 K (11.5 K). The low deviations found with the proposed method can estimate the mp with better accuracy than other methods. These results represent a tremendous increase in accuracy for predicting this important property and show that not only the use of the optimum network architecture is crucial, but also the appropriate selection of the

DOI:10.1088/1674-0068/22/01/19-26

TABLE III Optimum weight and bias for ANN 11-11-1.

p_i	M	μ	α	$E_{\rm R}$	$E_{\rm exc}$	$E_{\rm ne}$	$E_{\rm nn}$	$^1\chi^v$	$^2\chi^v$	$^3\chi^v$	$^4\chi^v$	b
w_{ij}	1	2	3	4	5	6	7	8	9	10	11	b_j
1	−0.1433	−0.6734	1.0812	−17.159	13.358	−28.794	−39.749	8.7152	−1.0693	−1.1944	0.1636	−1.0540
2	−3.3818	−0.4121	2.1796	−1.4837	3.6596	130.69	134.03	−0.2938	2.1642	−1.8852	0.6712	−0.0212
3	−86.209	34.855	20.635	−670.20	756.11	−52.742	−206.94	384.16	15.810	−170.14	47.017	−71.952
4	3.3532	0.3984	−2.2587	1.9022	−3.9686	−129.44	−132.60	0.2985	−2.1965	1.8962	−0.6488	0.0240
5	−0.0955	0.6880	−1.0580	16.098	−12.393	22.415	32.953	−8.4204	1.1599	1.2804	−0.1366	1.0159
6	−4.2348	−1.4237	4.0776	1.7639	5.4603	335.16	346.54	−6.8322	3.4397	−0.2936	−0.1318	−1.0856
7	−1.7894	−0.1378	3.3400	−13.804	12.482	20.922	19.408	−2.2752	2.6254	−0.7174	−0.5356	−0.4514
8	−2.8944	−9.1957	3.8769	−41.593	37.548	−271.35	−281.40	2.0713	5.6188	1.9885	−0.4635	−9.8033
9	0.9947	6.0585	−0.0600	70.890	−65.675	151.15	154.03	−2.6646	−0.4381	3.3525	0.9846	4.0521
10	−2.8420	−9.1585	3.5476	−44.253	39.670	−306.45	−316.88	2.2500	5.4698	1.8922	−0.4806	−9.9373
11	−1.4480	−0.6089	−31.474	56.769	−25.540	488.71	500.97	62.252	−3.0178	32.989	−22.426	20.855
w_{jk}	1	2	3	4	5	6	7	8	9	10	11	b_k
1	20.200	−44.233	0.0807	−44.6350	20.403	0.8399	−1.4607	−24.170	−0.3431	24.366	−0.3927	0.2372

independent variables.

TABLE IV Comparison of the method proposed in this work with other QSPR methods found in the literature to determine the melting point.

Authors	Method	N	R^2	RMSE
Charton and Charton [12]	MLR	178	0.937	24.0
Charton and Charton [16]	MLR	366	0.919	17.9
Todeschini *et al.* [26]	MLR	94	0.834	32.8
Pogliani [28]	MLR	62	0.856	18.0
This work	ANN	333	0.970	11.5

An important observation that must be mentioned is the influential effects of the variables as: the molecular weight M (size) and the dipole moment μ (polarity and symmetry), for distinguishing between the different physical and chemical properties of the substances considered in this study. Also, the molecular connectivity indices, which provide the network of important information for obtain a best correlation and prediction. Katritzky and Gordeeva proposed that the use of molecular connectivity indices provides the best correlations in QSPR applications [11]. This shows that the eleven QSPR parameters used were optimum to reflect the intermolecular forces and molecular symmetry that influence the melting point. Using other QSPR parameters as the independent variables does not produce the low deviation that was reached using the proposed scheme.

IV. CONCLUSION

This work presents a combined method that includes ANN+QSPR method for the correlation and prediction of melting points of organic compounds. Using molecular modeling and PM3 semi-empirical molecular orbital theories, twelve QSPR parameters were calculated for modeling this property: molecular weight (M), dipole moment (μ), average molecular polarizability (α), resonance energy ($E_{\rm R}$), exchange energy ($E_{\rm exc}$), electron-nuclear attraction ($E_{\rm ne}$), nuclear-nuclear repulsion ($E_{\rm nn}$), and the first-, second-, third- and fourth-order valence connectivity indices ($^1\chi^v$, $^2\chi^v$, $^3\chi^v$, $^4\chi^v$).

Based on the results and discussion presented in this study, the following main conclusions are obtained: (i) The great differences in structure chemical and physical properties of the organic compounds considered in the study impose additional difficulties on the problem that the proposed ANN has been able to handle; (ii) The results show that the ANN can be properly trained and that the chosen architecture (11-11-1) can estimate the melting point of organic compounds with low deviations. The consistency of the method was checked using experimental values of melting point of organic compounds and comparing them with the calculated values of the ANN; (iii) The eleven used QSPR parameters were optimum to reflect the intermolecular forces and molecular symmetry. These results add to the growing support for the use of ANNs in QSPR studies; and (iv) The values calculated with the proposed method are believed to be accurate enough for engineering calculations, for generalized correlations, and for equation of state methods, among other uses.

V. ACKNOWLEDGMENT

The work was supported by the Department of Physics of the University of La Serena-Chile.

DOI:10.1088/1674-0068/22/01/19-26

[1] J. C. Dearden, Environ. Toxicol. Chem. **22**, 1696 (2003).
[2] L. Zhao and S. H. Yalkowsky, Ind. Eng. Chem. Res. **38**, 3581 (1999).
[3] A. Jain, G. Yang, and S. H. Yalkowsky, Ind. Eng. Chem. Res. **43**, 7618 (2004).
[4] B. Poling, J. M. Prausnitz, and J. P. O'Connell, *The properties of Gases and Liquids*, New York: the McGraw-Hill Companies Inc., (2004).
[5] K. Joback and R. Reid, Chem. Eng. Commun. **57**, 233 (1987).
[6] A. L. Lydersen, Ph. D Dissertation. University of Wisconsin, Wisconsin (1955).
[7] L. Constantinou and R. Gani, AIChE J. **40**, 1697 (1994).
[8] S. Bhattacharjee, A. S. Rao, and P. Dasgupta, Comput. Chem. **15**, 319 (1991).
[9] J. C. Dearden, Sci. Total Environ. **109**, 59 (1991).
[10] M. Medić-Sărić, S. Nikolić, and J. Matijević-Sosa, Acta Pharm. **42**, 153 (1992).
[11] A. R. Katritzky and E. V. Gordeeva, J. Chem. Inf. Comput. Sci. **33**, 835 (1993).
[12] M. Charton and B. I. Charton, *Proceeding of the 27th Mid Atlantic Regional Meeting,* New York: American Chemical Society, (1993).
[13] D. Cherqaoui, D. Villemin, and V. Kvasnicka, Chemom. Intell. Lab. Syst. **24**, 117 (1994).
[14] C. H. Tu, J. Chin. Inst. Chem. Eng. **25**, 151 (1994).
[15] R. Murugan, M. P. Grendze, J. E. Toomey, A. R. Katritzky, M. Karelson, V. Lobanov, and P. Rachwal, Chemtech **24**, 17 (1994).
[16] M. Charton and B. I. Charton, J. Phys. Org. Chem. **7**, 196 (1994).
[17] R. Todeschini, P. Gramatica, R. Provenzani, and E. Marengo, Chemom. Intell. Lab. Syst. **27**, 221 (1995).
[18] R. Piazza, A. Pino, S. Marchini, L. Passerini, C. Chiorboli, and M. L. Tosato, SAR/QSAR Environ. Res. **4**, 59 (1995).
[19] L. Pogliani, J. Phys. Chem. **99**, 925 (1995).
[20] L. Pogliani, J. Phys. Chem. **100**, 18065 (1996).
[21] R. Todeschini and P. Gramatica, SAR/QSAR Environ. Res. **7**, 89 (1997).
[22] R. Todeschini and P. Gramatica, Quant. Struct. Act. Relat. **16**, 113 (1997).
[23] C. Chiorboli, P. Gramatica, R. Piazza, A. Pino, and R. Todeschini, SAR/QSAR Environ. Res. **7**, 133 (1997).
[24] L. Pogliani, Med. Chem. Res. **7**, 380 (1997).
[25] A. R. Katritzky, U. Maran, M. Karelson, and V. S. Lobanov, J. Chem. Inf. Comput. Sci. **37**, 913 (1997).
[26] R, Todeschini, M. Vighi, A. Finizio, and P. Gramtica, SAR/QSAR Environ. Res. **7**, 173 (1997).
[27] P. Gramatica, N. Navas, and R. Todeschini, Chemom. Intell. Lab. Syst. **40**, 53 (1998).
[28] L. Pogliani, J. Phys. Chem. A **103**, 1598 (1999).
[29] L. Pogliani, J. Phys. Chem. A **104**, 9029 (2000).
[30] X. Gironés, L. Amat, D. Robert, and R. Carbó-Dorca, J. Comput. Aided Mol. Des. **14**, 477 (2000).
[31] C. A. S. Bergström, U. Norinder, K. Luthman, and P. Artursson, J. Chem. Inf. Comput. Sci. **43**, 1177 (2003).
[32] M. Karthikeyan, R. C. Glen, and A. Bender, J. Chem. Inf. Model. **45**, 581 (2005).
[33] H. Modarresi, J. C. Dearden, and H. Modarress, J. Chem. Inf. Model. **46**, 930 (2006).
[34] B. Lučić, D. Amić, and N. Trinajstić, J. Chem. Inf. Comput. Sci. **40**, 403 (2000).
[35] J. Taskinen and J. Yliruusi, Adv. Drug Deliver. Rev. **55**, 1163 (2003).
[36] J. Tettech, E. Metcalfe, and S. L. Howells, Chemom. Intell. Lab. Syst. **32**, 177 (1996).
[37] G. Espinosa, D. Yaffe, A. Arenas, Y. Cohen, and F. Giralt, Ind. Eng. Chem. Res. **40**, 2757 (2001).
[38] D. Yaffe and Y. Cohen, J. Chem. Inf. Comput. Sci. **41**, 463 (2001).
[39] M. Karelson, D. A. Dobchev, O. V. Kulshyn, and A. R. Katritzky, J. Chem. Inf. Model. **46**, 1891 (2006).
[40] A. R. Katritzky, V. Lobalov, M. Karelson, *CODESSA Reference Manual*, Gainesville: University of Florida Press, (1996).
[41] J. R. De Laeter, J. K. Böhlke, P. De Bièvre, H. Hidaka, H. S. Peiser, K. J. Rosman, and P. D. Taylor, Pure App. Chem. **75**, 683 (2003).
[42] P. W. Atkins, *Quanta*, Oxford: Oxford University Press, (1991).
[43] D. F. V. Lewis, C. Ioannides, and D. V. Parke, Xenobiotica **24**, 401 (1994).
[44] E. Clementi, *Computational Aspects of Large Chemical Systems*, New York: Springer Verlag, (1980).
[45] L. B. Kier and L. H. Hall, Eur. J. Med. Chem. **12**, 307 (1977).
[46] L. B. Kier and L. H. Hall, J. Pharm. Sci. **70**, 583 (1981).
[47] T. E. Daubert, R. P. Danner, H. M. Sibul, and C. C. Stebbins, *Physical and Thermodynamic Properties of Pure Chemicals. Data Compilation*, London: Taylor & Francis, (2000).
[48] A. R. Katritzky, U. Maran, V. S. Lobanov, and M. Karelson, J. Chem. Inf. Comput. Sci. **40**, 1 (2000).
[49] L. D. Hughes, D. S. Palmer, F. Nigsch, and J. B. O. Mitchell, J. Chem. Inf. Model. **48**, 220 (2008).

DOI:10.1088/1674-0068/22/01/19-26