

Prediction of Flash Point Temperature of Organic Compounds Using a Hybrid Method of Group Contribution + Neural Network + Particle Swarm Optimization

Juan A. Lazzús*

Department of Physics, University of La Serena, Casilla 554, La Serena, Chile

Abstract The flash points of organic compounds were estimated using a hybrid method that includes a simple group contribution method (GCM) implemented in an artificial neural network (ANN) with particle swarm optimization (PSO). Different topologies of a multilayer neural network were studied and the optimum architecture was determined. Property data of 350 compounds were used for training the network. To discriminate different substances the molecular structures defined by the concept of the classical group contribution method were given as input variables. The capabilities of the network were tested with 155 substances not considered in the training step. The study shows that the proposed GCM+ANN+PSO method represent an excellent alternative for the estimation of flash points of organic compounds with acceptable accuracy (AARD = 1.8%; AAE = 6.2 K).

Keywords flash point, group contribution method, artificial neural networks, particle swarm optimization, property estimation

1 INTRODUCTION

One of the most important physical properties used to determine the potential hazard of fire and explosion for industrial materials is the flash point temperature [1]. The flash point (T_f) is defined as the lowest temperature at which a liquid produces enough vapour to ignite in the presence of a source of ignition. Like the boiling point of a substance, the flash point depends on the ambient pressure: the more volatile a solvent is at normal temperature, the greater is the fire hazard it poses [2]. T_f is often used as a descriptive characteristic of liquid fuels, but it is also used to describe liquids that are not used intentionally as fuels [1].

The knowledge of T_f of materials is essential in

many pharmaceutical and chemical unit operations. For example, in some processes powdered materials are handled at low relative humidity (weighing, mixing, fluidization, drying, storage, transportation), in which frictional electricity may be generated [2].

There are many methods for prediction of T_f in the literature. Vidal *et al.* have presented a review of the most important methods for the prediction of flash point [3]. Mainly, prediction methods for this property can be categorized as group contribution method (GCM) and quantitative structure-property relationship (QSPR). Table 1 lists selected applications to estimate T_f [1, 4–12]. Studies on mixture flash points have also been developed (see Ref. [3]).

Artificial neural network (ANN) is accepted as the

Table 1 Reported GCMs for predicting flash points of organic compounds

Authors	Year	Method	Compounds	No. data
Suzuki <i>et al.</i> [4]	1991	Pseudo-QSPR	Hydrocarbons	400
Satyanarayana and Rao [5]	1992	Correlation	Diverse	1221
Tetteh <i>et al.</i> [6]	1999	QSPR	Diverse	400
Katritzky <i>et al.</i> [7]	2001	QSPR	Diverse	271
Zhokhova <i>et al.</i> [8]	2003	QSPR	Diverse	398
Zhokhova <i>et al.</i> [8]	2003	QSPR	Diverse	525
Albahri [9]	2003	GCM	Hydrocarbons	—
Vazhev <i>et al.</i> [10]	2006	IRSM ^①	Alkanes	85
Pan <i>et al.</i> [11]	2007	GCM	Alkanes	92
Katritzky <i>et al.</i> [12]	2007	QSPR	Diverse	758
Gharagheizi and Alamdari [1]	2008	QSPR	Diverse	1030

① Infrared spectra method.

Received 2009-09-18, accepted 2010-05-20.

* To whom correspondence should be addressed. E-mail: jlazzus@dfuls.cl

most powerful nonlinear technique in QSPR applications [2, 13]. The neural network modeling in QSPR has been applied to most physicochemical properties, for which suitable experimental data can be found in the literature. The GCM uses linear and non-linear regression techniques to represent the relations among the variables of a given system [14–16]. The relationship between the physical and thermodynamic properties is highly non-linear, and consequently an ANN may be a suitable alternative to model the underlying thermodynamic properties [13–16]. ANN is an especially efficient algorithm to approximate any function with finite number of discontinuities by learning the relationships between input and output vectors. Thus, ANN is an appropriate technique to model the non-linear behavior of thermophysical properties [2, 13–17].

In this work, the flash point of organic compounds is estimated using a simple GCM implemented in an ANN replacing standard backpropagation with particle swarm optimization (PSO), which is one of the most recently developed evolutionary algorithms [18].

2 NEURAL NETWORK USED

In this study a feed-forward neural network is used to represent non-linear relationships among variables [2, 13–17]. The network programmed with the software MATLAB [19] consists of a multilayer network, in which the flow of information spreads forward through the layers. 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 ANN: learning and validation. The learning or training is the process in which an ANN modifies the weights in answer to entrance information [13–17].

This ANN program reads the necessary data organized in an Excel file. To distinguish different physical and chemical properties of substances so that the network can discriminate and learn in optimum form, properties derived from the molecular structure are considered.

The input layer contains one neuron (node) for each variable. The output layer has one node generating the scaled estimated value of T_f . The ANN is trained with particle swarm optimization [16, 17]. Some researchers have used PSO to train neural networks and found that PSO-based ANN has a better training performance, faster convergence rate, and a better prediction ability than the standard backpropagation algorithm [20].

PSO is a population-based optimization tool, where the system is initialized with a population of random particles and the algorithm searches for optima by updating generations [21]. In a PSO system, each particle is “flown” through the multidimensional search space, adjusting its position in search space according to its own experience and that of neighboring particles. The particle therefore makes use of the best position encountered by itself and its neighbors to position it-

self toward an optimal solution. The performance of each particle is evaluated using a predefined fitness function, which encapsulates the characteristics of the optimization problem [20].

In each iteration, the velocity for each particle is calculated according to the following formula:

$$v_i^p(t+1) = \omega v_i^p(t) + c_1 r_1 (\psi_i^p(t) - x_i^p(t)) + c_2 r_2 (\psi_g(t) - x_i^p(t)) \quad (1)$$

where t is the current step number, ω is the inertia weight, c_1 and c_2 are the acceleration constants, r_1 and r_2 are elements from two random sequences in the range (0, 1), $x_i^p(t)$ is the current position of the particle, ψ_i^p is the best solution for this particle, and ψ_g is the best solution for all particles. In general, the value of each component in v can be restricted to the range $[-v_{\max}, v_{\max}]$ to control excessive roaming of particles outside the search space [21]. After calculating the velocity, we obtain the new position of every particle

$$x^p(t+1) = x^p(t) + v^p(t+1) \quad (2)$$

The PSO algorithm repeats the application of the update equations above until a specified number of iterations is exceeded, or until the velocity updated is close to zero.

The steps to calculate the output parameter (T_f), using the input parameters, are as follows.

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

$$N_j^h = \sum_i^n w_{ij}^h p_i + b_j^h \quad (3)$$

where p corresponds to the vector of the inputs for 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 neuron j of the hidden layer, reached in its activation. Starting from these inputs, the outputs (y) of the hidden neurons are calculated, using a transfer function f^h associated with the neurons of this layer.

$$y_j^h = f_j^h \left(\sum_i^n w_{ij}^h p_i + b_j^h \right) \quad (4)$$

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

To minimize the error, the transfer function f should be differentiable. In the ANN, two types of transfer function are used: the hyperbolic tangent function (*tansig*) in the hidden layer, defined by the equation

$$f(N_{jk}) = \frac{e^{N_{jk}} - e^{-N_{jk}}}{e^{N_{jk}} + e^{-N_{jk}}} \quad (5)$$

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

$$f(N_{jk}) = (N_{jk}) \quad (6)$$

All the neurons of the ANN have an associated activation value for a give input pattern, and the algorithm continues finding the error for each neuron, except those of the input layer. After finding the output values, the weights of all layers of the network are actualized by PSO, using Eqs. (1) and (2) [16].

The PSO algorithm is very different from any of the traditional methods for training [16, 21]. Each neuron contains a position and velocity. The position corresponds to the weight of a neuron. The velocity is used to update the weight and control how much the position is updated. If a neuron is further away (the position is further from the global best position), it will adjust its weight more than a neuron closer to the global best. PSO initializes all weights to random values and starts training each one. When it passes through a data set, PSO compares the fitness of its weight. The network with the highest fitness is considered as the global best. The other weights are updated based on the global best network rather than on their personal error or fitness [16]. Fig. 1 shows a block diagram of the program developed and written in MATLAB M-file.

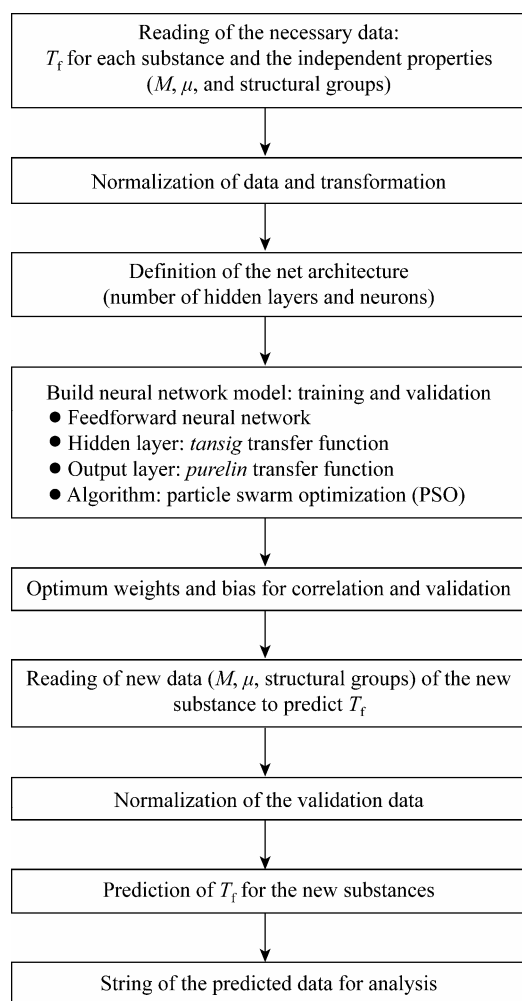


Figure 1 Flow diagram for the ANN+PSO program developed for this work

3 DATABASE USED AND TRAINING

In this study, 350 organic compounds were used to train the network and then values of T_f of 155 substances not used in the training process were predicted. Molecular mass M (size), dipole moment μ (polarity), and the structure of the molecules, represented by the number of well-defined groups forming the molecule, were provided as input variables. Molecular mass and dipole moment were chosen to characterize different molecules [13–16]. Note that, all the properties of interest (M , μ , T_f) are from the DIPPR database [22] that includes estimated uncertainties for the experimental data.

To distinguish different substances considered in this study so that the net can discriminate and learn in optimum form, the properties used cover wide ranges: 209 K to 716 K for T_f of organic compounds. In addition, the substances included in the study have very different physical and chemical characteristics. Low molecular weight substances such as acetonitrile ($M=41$) to high molecular weight substances such as diolein ($M=621$), or non-polar substances ($\mu=0$) such as benzene and anthracene to highly polar substances such as glycine ($\mu=12.8$). Thus, the problem is not straightforward and probably it is one of the reasons why the value of T_f has not been treated by using ANN as proposed in this paper.

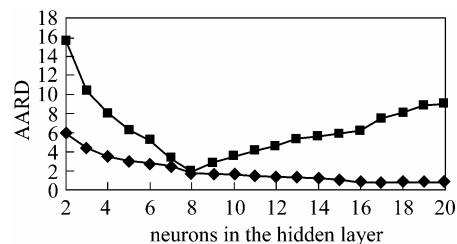
The molecular structure is represented by the number of well-defined groups forming the molecule. The value associated to the structural group is defined as: 0, when the group does not appear in the substance and n , when the group appears n times in the substance [14, 16]. For instance, for ethyl vanillin the property data are: $M=166.2$, $\mu=4.2$, and the structure of the molecule $[-CH_3]=1$, $[-CH_2-]=1$, $[-O-]=1$, $[-CHO]=1$, $[=CH-(ring)]=3$, $[=C<(ring)]=3$, and $[-OH(ring)]=1$. Table 2 shows the 44 groups used as entrance variables.

Several network architectures were tested to select the most accurate scheme. The most basic architecture normally used for this type of application involves a neural network consisting of three layers [2, 13–16]. 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 of the hidden layer, and many alternative combinations are possible. The optimum number of neurons is determined by adding neurons in systematic form and evaluating the average absolute deviations of the sets in the learning process [13–16]. Fig. 2 shows the average absolute relative deviation in correlating T_f of all compounds as a function of the number of neurons in the hidden layer (44–8–1). The optimum number of neurons in the hidden layer is between 7 and 10. The network with the lowest deviation during training is the one with 44 parameters in the input layer, 8 neurons in the hidden layer, and one neuron in the output layer (44–8–1). Its average deviation during training is 1.7% and during prediction is 1.9%.

Table 2 Structural groups used in the GCM+ANN+PSO model^①

No.	Parameter		No. occurrence		
	Group	Max. value	Training set	Predicting set	Total set
1	<i>M</i>	621.0	350	155	505
2	μ	12.8	350	155	505
3	—CH ₃	8	233	100	333
4	—CH ₂ —	34	177	80	257
5	>CH—	6	44	18	62
6	>C<	2	20	5	25
7	=CH ₂	2	23	7	30
8	=CH—	4	30	4	34
9	=C<	1	12	2	14
10	=C=	1	4	1	5
11	≡CH	1	2	0	2
12	≡C—	2	3	0	3
13	—OH	5	32	12	44
14	—O—	2	12	3	15
15	>C=O	2	11	7	18
16	—CHO	2	27	8	35
17	—COOH	2	38	8	46
18	—COO—	2	25	12	37
19	HCOO—	1	5	1	6
20	=O	1	2	1	3
21	—NH ₂	2	13	10	23
22	—NH—	1	12	4	16
23	>N—	1	4	3	7
24	=N—	2	1	1	2
26	—CN	2	10	3	13
25	—NO ₂	2	7	12	19
27	—F	6	2	0	2
28	—Cl	6	6	15	21
29	—Br	4	8	5	13
30	—I	2	3	3	6
31	—SH	2	19	6	25
32	—S—	2	5	5	10
33	—CH ₂ —(ring)	10	32	11	43
34	>CH—(ring)	6	11	6	17
35	=CH—(ring)	20	134	76	210
36	>C<(ring)	2	5	4	9
37	=C<(ring)	6	128	74	202
38	—O—(ring)	3	12	2	14
39	—OH(ring)	2	25	4	29
40	>C=O(ring)	2	9	1	10
41	—NH—(ring)	2	5	2	7
42	>N—(ring)	2	4	1	5
43	=N—(ring)	1	4	9	13
44	—S—(ring)	1	3	2	5

① Substances and properties used in the GCM+ANN+PSO model and an Excel file that can be used for further calculations can be provided. The material is available *via* the author's e-mail.

**Figure 2** Average absolute relative deviations (AARD) for correlating the flash points of all substances as a function of the number of neurons in the hidden layer

◆ during the training step; ■ during the prediction step

The accuracy of the chosen final network is checked using the average relative deviation (ARD), average absolute relative deviation (AARD), and average absolute error (AAE) between the calculated value of T_f after training and the data in literature. The deviations are calculated as

$$\text{ARD} = \frac{100}{N} \sum_{i=1}^N \left(\frac{T_f^{\text{calc}} - T_f^{\text{exp}}}{T_f^{\text{exp}}} \right) \quad (7)$$

$$\text{AARD} = \frac{100}{N} \sum_{i=1}^N \left| \frac{T_f^{\text{calc}} - T_f^{\text{exp}}}{T_f^{\text{exp}}} \right| \quad (8)$$

$$\text{AAE} = \frac{100}{N} \sum_{i=1}^N |T_f^{\text{calc}} - T_f^{\text{exp}}| \quad (9)$$

4 RESULTS AND DISCUSSION

Table 3 shows the overall minimum, maximum, and average deviations for all the substances using the proposed network 44-8-1. The results show that the ANN can be accurately trained and that the chosen architecture can estimate T_f of organic compounds with enough accuracy. It gives lower deviations than others with GCM available in literature: AARD less than 1.7% and AAE less than 5.9K for the 350 organic compounds used in the training. AARD less than 1.9% and AAE less than 6.9K for the 155 organic compounds used in the prediction step. For all substances (505 organic compounds) the AAE is a little higher than 6K, the AARD is below 9%, and for 466 compounds the AARD is below 5%.

Once the best architecture was determined, the optimum weights required to estimate T_f of organic compounds were obtained. Table 4 shows the optimum weight and biases for the ANN 44-8-1.

Figure 3 shows a comparison for experimental (solid line) and calculated values (circles) of T_f for organic compounds. Fig. 3 (a) shows a comparison in the training step for correlated and literature values of T_f . The correlation coefficient R^2 is 0.9891 and the slope of the curve (m) is 0.9815 (expected to be 1.0). Fig. 3 (b) shows a comparison in the prediction step for predicted and literature values of T_f . The correlation coefficient R^2 is 0.9853 and m (also expected to be 1.0) is 0.9816. For the total set, R^2 is 0.9881 and m is 0.9814.

Table 3 Overall minimum, maximum, and average deviations for the calculated flash points for all compounds using the GCM+ANN+PSO model

Deviation	No. Substances	ARD _{min} /%	ARD _{max} /%	ARD/%	AAE/K	AAE _{max} /K	AARD/%	No. AARD<5%	No. AARD>9%
training set	350	-8.8	8.6	0.0	5.9	26.3	1.7	324	0
prediction set	155	-7.4	6.7	-0.1	6.9	26.6	1.9	142	0
total set	505	-8.8	8.6	0.0	6.2	26.6	1.8	466	0

Table 4 Optimum weights and biases for the GCM+ANN+PSO model 44-8-1

w_{ji}	1	2	3	4	5	6	7	8
1	5.8434	-2.2311	5.7657	-1.3795	23.898	4.2349	-8.5584	-7.2881
2	2.3359	-1.5198	6.8655	1.0406	-8.2926	-1.6984	3.0101	-0.4193
3	2.4578	1.9084	-4.4478	4.4605	-8.3091	-1.2154	-6.0922	0.9045
4	6.2536	4.0579	-0.9654	1.3302	-16.859	1.2607	6.1968	4.3105
5	16.110	2.4261	0.6878	-2.0116	2.0053	3.4806	-6.7674	1.2115
6	-4.6850	-4.5627	-1.4966	2.2717	13.976	-4.1419	1.4287	0.5730
7	0.1422	-0.2985	5.0051	2.6954	-0.2756	-4.7778	-1.6792	0.2242
8	-6.4103	-1.6893	6.8316	-1.4651	-0.5401	9.3862	-1.5031	0.5911
9	-8.8051	4.3388	3.5570	2.3151	2.5846	-3.8564	-4.8775	0.2164
10	0.9881	0.2980	0.4473	0.1364	0.3382	0.1330	1.4999	0.0423
11	-1.0091	0.2447	-1.1312	1.1389	3.3849	1.1650	3.0804	4.1582
12	2.2501	0.0152	-1.7426	1.3379	-3.3456	-3.1250	0.4005	-8.1713
13	1.9592	-11.822	0.4696	2.4056	-6.3104	8.3939	-2.3784	0.1183
14	0.9082	-0.6127	-5.9747	2.3377	-3.6832	-2.0599	9.8593	0.0241
15	1.6012	-1.3001	2.6361	3.3800	5.2458	-2.8983	3.1280	0.0938
16	10.551	1.2986	-2.6271	4.3050	7.1964	3.0557	12.554	0.5098
17	0.9304	-6.8914	-3.4135	-2.2351	0.9644	4.6809	-10.295	0.4804
18	-2.9416	2.4080	5.9233	-2.8952	-0.9688	1.1741	-3.0445	0.8931
19	-0.4679	-0.3948	-0.5454	1.2559	-5.0619	-1.5320	-0.8333	0.2629
20	1.2034	2.6897	1.6079	0.9361	-7.0835	-0.3788	-1.3620	-0.0114
21	-11.631	-0.5195	-0.9349	8.9506	-4.7688	-1.7638	5.0438	-0.1618
22	-0.0440	2.2616	11.527	2.4671	5.1147	-0.3517	-5.1998	0.9049
23	-0.6497	3.1264	3.4644	3.3032	2.5815	-0.4748	4.9758	-0.1138
24	-0.3628	0.0918	-0.7278	1.2464	0.1056	0.4641	0.6499	0.3373
25	-9.4054	0.4606	-2.1335	1.0998	7.1582	2.1947	7.4877	0.4487
26	-8.0639	1.4617	-0.5930	-4.9255	20.567	-0.0361	-2.7603	1.0151
27	-0.6909	1.8959	-0.1779	0.8050	-8.9110	-3.8063	-0.5399	1.1580
28	-1.2057	-0.9808	-2.4938	-0.3631	4.3261	-4.6949	3.9962	1.7815
29	2.9755	0.5922	-4.3800	0.1521	-12.729	-2.6100	-6.2742	2.9301
30	0.1983	1.2863	8.9326	0.1049	-10.256	-5.0972	-2.6431	2.3745
31	-8.2857	3.8574	3.7150	-2.1621	-4.7115	0.3903	7.0729	0.4994
32	0.2745	1.4118	-3.7693	2.0507	-6.9337	-0.6275	4.8414	0.2756
33	1.8373	0.3513	-3.0272	2.5451	-1.4811	1.6800	2.2665	1.3680
34	-1.0409	-5.0993	5.2380	-8.4811	-4.6895	4.7850	-4.3952	0.6447
35	-3.9664	2.6299	-0.6757	0.7209	-15.379	-1.3876	-8.1599	1.8742
36	0.7026	-4.7074	-2.5301	3.6644	2.6025	4.9796	-0.8372	0.9856
37	-4.7723	-5.4480	-1.6608	-2.0604	4.8109	2.2251	-9.7466	0.8932
38	5.3898	-2.8438	-2.3025	1.4395	11.368	-4.3243	-3.3353	0.3335
39	0.4067	3.3737	2.4209	-3.2168	10.764	-2.0749	-6.5604	0.0535
40	1.6635	3.0381	-1.5126	0.5535	1.0556	7.3535	10.359	0.3428
41	1.5427	-1.8020	-0.9856	1.3436	-3.8569	-3.4641	3.5500	-0.4619
42	-1.2357	-0.3713	-1.4211	0.7338	9.9229	1.2164	0.4163	0.1823
43	2.1771	1.8935	-4.2678	3.6563	-2.7002	-2.9109	4.1198	-0.1563
44	2.3566	-7.6840	5.0116	-0.9446	3.3542	-1.3482	-0.4722	0.6831
b_j	1.3765	1.4480	-0.3732	-1.2754	3.2431	1.1682	-0.8497	17.201
w_{jk}	-0.0653	-0.1986	0.3353	-0.2653	0.1253	0.1258	-0.0532	-0.4972
b_i								-0.1340

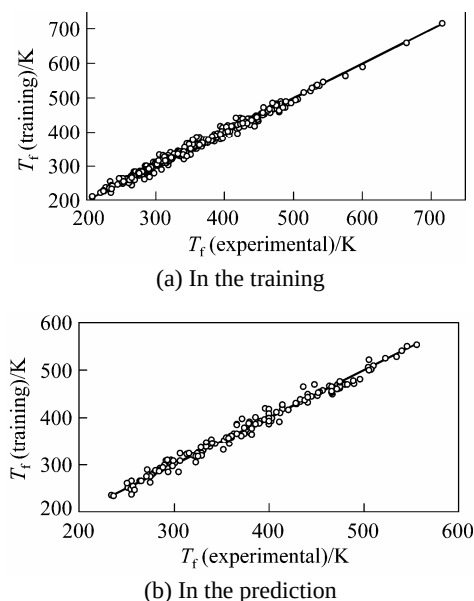


Figure 3 Comparison for experimental and calculated values of flash point

Most published methods relating T_f to chemical structure are confined to limited and/or small sets of hydrocarbons, substituted aromatics, and alkanes (see Table 1). There are very few studies with data sets for diverse substances in literature. In this work the compounds include aromatic and aliphatic hydrocarbons, halogens, polychlorinated biphenyls, mercaptans, sulfides, anilines, pyridines, alcohols, carboxylic acids, aldehydes, amines, ketones, and esters.

Table 5 shows a comparison for some methods proposed in literature for prediction of T_f for diverse organic compounds [1, 6–8, 12] and the GCM+ANN+PSO model proposed in this work. Tetteh *et al.* [6] used a radial basis neural network to predict T_f of 400 pure components with AAE higher than 12 K. Suzuki *et al.* [4] used a QSPR method to estimate T_f of 271 pure components with AAE of 14K, AARD of 3.8% and AARD_{max} higher than 21%. Zhokhova *et al.* [8] used fragmental descriptors in two approaches for the QSPR to obtain T_f of 525 compounds with AAE higher than 14 K. Katritzky *et al.* [12] repeated their work for T_f with 758 organic components with a QSPR method and ANNs with AAE of 14 K, AAE_{max} of 87 K, and

AAE higher than 3.5% with AARD_{max} higher than 24%. Recently, Gharagheizi and Alamdari [1] proposed a genetic algorithm-based multivariate linear regression to predict T_f of 1030 pure compounds using a QSPR model with AAE higher than 10 K.

The present model is also compared to the neural network with standard backpropagation (BPNN), similar architecture (44-8-1) and database. The BPNN shows AAE of 20 K, AAE_{max} higher than 100K, AARD higher than 5%, and AARD_{max} greater than 30%. Fig. 4 shows the AARD in the prediction of T_f using the BPNN and the present model. The low deviations with the proposed GCM+ANN+PSO model (AAE of 6 K with AAE_{max} a little higher than 25 K, AARD a little higher than 1%, and AARD_{max} of 9%) indicate that it can estimate T_f of organic compounds with better accuracy than other methods available in literature. These results represent a tremendous increase in accuracy to predict this important property, and shows that the incorporation of the descriptors, M (size) and μ (polarity and symmetry), for distinguish different physical and chemical characteristic of the substances is crucial.

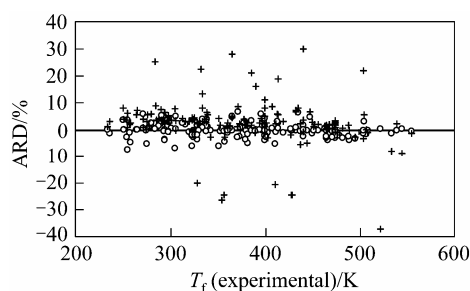


Figure 4 Average relative deviations (ARD) in the prediction of flash points

○ artificial neural network with particle swarm optimization;
+ artificial neural network with standard backpropagation

To reproduce the results presented in this paper and to obtain the flash point temperature of any organic compound using the proposed method, supplementary material can be provided.

5 CONCLUSIONS

This work presents a hybrid method that includes a

Table 5 Comparison of the proposed GCM+ANN+PSO model with other QSPR methods in literature to estimate the flash points of diverse organic compounds

Authors	Method	No. data	AAE	AARD	R^2
Tetteh <i>et al.</i> [6]	QSPR	400	10.2	—	0.9326
Katritzky <i>et al.</i> [7]	QSPR	271	14.0	3.8	0.9020
Zhokhova <i>et al.</i> [8]	QSPR	525	14.0	—	0.9343
Katritzky <i>et al.</i> [12]	QSPR	758	13.9	3.5	0.8780
Gharagheizi and Alamdari [1]	QSPR	1030	10.2	—	0.9669
This work	GCM	505	6.2	1.8	0.9881

simple group contribution method (GCM) implemented in a neural network (ANN) replacing standard back-propagation with particle swarm optimization (PSO), for the correlation and prediction of flash points (T_f) of organic compounds. The proposed GCM+ANN+PSO model is compared with other methods available in literature, demonstrating that it has a better capacity for prediction of this important property.

Based on the results and discussion presented in this study, the following main conclusions are obtained. (1) The great differences in structure chemical and physical properties of the organic compounds considered in the study impose additional difficulties to the problem, and the proposed ANN is able to handle it. (2) The ANN can be properly trained and the chosen architecture (44-8-1) can estimate T_f of organic compounds with low deviations. The consistency of the method is checked by comparing the calculated values with experimental values of T_f of organic compounds. (3) The low deviations with the proposed GCM+ANN+PSO model indicate that it can estimate T_f of organic compounds with better accuracy than other GCMs available in the literature. (4) The results of the GCM+ANN+PSO model represent a tremendous increase in accuracy for predicting T_f , and show that the incorporation of the descriptors M (size) and μ (polarity and symmetry) for distinguishing different physical and chemical characteristic of the substances is crucial. (5) The values calculated with proposed method are believed to be accurate enough for engineering calculations and for generalized correlations, among other uses.

ACKNOWLEDGEMENTS

The author thank the Direction of Research of the University of La Serena (DIULS), and the Department of Physics of the University of La Serena (DFULS), by the special support that made possible the preparation of this paper.

REFERENCES

- Gharagheizi, F., Alamdari, R.F., "Prediction of flash point temperature of pure components using a quantitative structure-property relationship model", *QSAR Comb. Sci.*, **27**, 679–683 (2008).
- Taskinen, J., Yliruusi, J., "Prediction of physicochemical properties based on neural network modeling", *Adv. Drug Deliv. Rev.*, **55**, 1163–1183 (2003).
- Vidal, M., Rogers, W.J., Holste, J.C., Mannan, M.S., "A review of estimations method for flash points and flammability limits", *Process Saf. Progress*, **23**, 47–55 (2004).
- Suzuki, T., Ohtaguchi, K., Koide, K., "A method for estimating flash points of organic compounds from molecular structures", *J. Chem. Eng. Jpn.*, **24**, 258–261 (1991).
- Satyanarayana, K., Rao, P.G., "Improved equation to estimate flash points of organic compounds", *J. Hazard. Mater.*, **32**, 81–85 (1992).
- Tetteh, J., Suzuki, T., Metcalfe, E., Howells, S., "Quantitative structure-property relationships for the estimation of boiling point and flash point using a radial basis function neural network", *J. Chem. Inf. Comput. Sci.*, **39**, 491–507 (1999).
- Katritzky, A.R., Petrukhin, R., Jain, R., Karelson, M., "QSPR analysis of flash points", *J. Chem. Inf. Comput. Sci.*, **41**, 1521–1530 (2001).
- Zhokhova, N.I., Baskin, I.I., Palyulin, V.A., Zefirov, A.N., Zefirov, N.S., "Fragmental descriptors in QSPR: Flash point calculations", *Russ. Chem. Bull. Int. Ed.*, **52**, 1885–1892 (2003).
- Albahri, T.A., "Flammability characteristics of pure hydrocarbons", *Chem. Eng. Sci.*, **58**, 3629–3641 (2003).
- Vazhev, V.V., Aldabergenov, M.K., Vazheva, N.V., "Estimation of flash points and molecular masses of alkanes from their IR spectra", *Petrol. Chem.*, **46**, 136–139 (2006).
- Pan, Y., Jiang, J., Wang, Z., "Quantitative structure-property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network", *J. Hazard. Mater.*, **147**, 424–430 (2007).
- Katritzky, A.R., Stoyanova-Slavova, I.B., Dobchev, D.A., Karelson, M., "QSPR modeling of flash points: An update", *J. Mol. Graph. Model.*, **26**, 529–536 (2007).
- Lazzús, J.A., "Neural network based on quantum chemistry for predicting melting point of organic compounds", *Chin. J. Chem. Phys.*, **22**, 19–26 (2009).
- Lazzús, J.A., "p-T-P prediction for ionic liquids using neural networks", *J. Taiwan Inst. Chem. Eng.*, **40**, 213–232 (2009).
- Lazzús, J.A., "Prediction of solid vapor pressures for organic and inorganic compounds using a neural network", *Thermochim. Acta*, **489**, 53–62 (2009).
- Lazzús, J.A., "Estimation of density as a function of temperature and pressure for imidazolium-based ionic liquids using a multilayer net with particle swarm optimization", *Int. J. Thermophys.*, **30**, 833–909 (2009).
- Lazzús, J.A., "Hybrid method to predict melting points of organic compounds using group contribution + neural network + particle swarm algorithm", *Ind. Eng. Chem. Res.*, **48**, 8760–8766 (2009).
- Luo, Q., Yi, D., "A co-evolving framework for robust particle swarm optimization", *Appl. Math. Comput.*, **199**, 611–622 (2008).
- MathWorks, MatLab version 6.5.0., The MathWorks Inc. (2002).
- Da, Y., Xiurun, G., "An improved PSO-based ANN with simulated annealing technique", *Neurocomputing*, **63**, 527–533 (2005).
- Jiang, Y., Hu, T., Huang, C., Wu, X., "An improved particle swarm optimization algorithm", *Appl. Math. Comput.*, **193**, 231–239 (2007).
- Daubert, T.E., Danner, R.P., Sibul, H.M., Stebbins, C.C., Physical and Thermodynamic Properties of Pure Chemicals. Data Compilation, Taylor & Francis, London (2000).