

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/231377615>

Corresponding States Method for Evaluation of the Solubility Parameters of Chemical Compounds

ARTICLE *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · FEBRUARY 2012

Impact Factor: 2.59 · DOI: 10.1021/ie202633m

CITATIONS

12

READS

49

5 AUTHORS, INCLUDING:



Farhad Gharagheizi

Texas Tech University

168 PUBLICATIONS **2,910** CITATIONS

SEE PROFILE



Mehdi Sattari

University of KwaZulu-Natal

40 PUBLICATIONS **325** CITATIONS

SEE PROFILE



Amir H. Mohammadi

557 PUBLICATIONS **4,805** CITATIONS

SEE PROFILE



Dominique Richon

Aalto University

533 PUBLICATIONS **6,594** CITATIONS

SEE PROFILE

Corresponding States Method for Evaluation of the Solubility Parameters of Chemical Compounds

Farhad Gharagheizi,[†] Ali Eslamimanesh,[‡] Mehdi Sattari,[§] Amir H. Mohammadi,^{*,‡,||} and Dominique Richon^{‡,||}

[†]Department of Chemical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

[‡]MINES ParisTech, CEP/TEP—Centre Énergétique et Procédés, 35 Rue Saint Honoré, 77305 Fontainebleau, France

[§]Saman Energy Giti Company, Postal Code 3331619636, Tehran, Iran

^{||}Thermodynamics Research Unit, School of Chemical Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa

Supporting Information

ABSTRACT: In this study, our objective is to apply the gene expression programming mathematical algorithm to propose a correlation based on the corresponding states method to determine the solubility parameters of 1641 pure compounds (mostly organic ones) at 298.15 K and atmospheric pressure. The studied compounds are from the 80 chemical families. The parameters of the method include the critical temperature, critical pressure, molecular weight, and acentric factor. Around 1477 solubility parameter data are randomly selected for developing (training + optimization) the correlation, and about 164 data are used for checking its prediction capability. The obtained statistical parameters, including average absolute relative deviation of the results from the applied data (about 6%), show the accuracy of the proposed method along with its simplicity and wide range of applicability.

1. INTRODUCTION

In our previous works,^{1–3} we have stated that reliable predictive tools with parameters defined for representative chemicals related to representative chemical families are of drastic significance. Moreover, detailed discussions^{1–5} about the necessity of application of mathematical algorithms such as gene expression programming (GEP)^{6,7} to evaluate the optimal values in the aforementioned methods (or tools) have already been presented.^{1–5} Application of the genetic algorithm (GA)⁸ as a population-based evolutionary optimization technique has generated considerable attention in engineering fields for several decades. In later years, an improvement of the original GA⁸ was developed, in which the solutions are treated as nonlinear structures of parse trees (treated as functions) instead of fixed length binary solutions (genetic programming (GP)).^{9,10} Modification of this new algorithm was continued until its reformulation by Ferreira,^{6,7} who presented the so-called “GEP” method.

The GEP strategy^{6,7} totally encodes ramified structures of different sizes and shapes (parse trees) in linear solutions of fixed length, which contributes to greater probability of obtaining the global optimum values of the model parameters.^{1–3,6,7} It has been currently proven that GEP is a generally acceptable mathematical strategy for this purpose.^{1–3,11–15} The details of this technique can be found elsewhere.^{1–3,6,7}

The concept of the solubility parameter was first introduced by Scatchard¹⁶ as a physicochemical parameter defining solvent affinity for solving a definite kind of solute.^{17–23} Wide ranges of applications of solubility parameter values in coating and paint technologies, complex extraction operations, polymer processes, etc. have been investigated so far.¹⁸ Furthermore, several

thermodynamic models developed for prediction of the amounts/conditions of precipitations/depositions of heavy petroleum fractions such as asphaltenes or wax^{24–32} are based on the application of solubility parameters of the species present in the framework of, for instance, regular solution theory.³³

The general definition of the solubility parameter can be expressed as follows:¹⁷

$$\delta = \left(\frac{\Delta E_v}{v} \right)^{1/2} = \left(\frac{\Delta U_{\text{vap}}}{v} \right)^{1/2} = \left(\frac{\Delta H_{\text{vap}} - RT}{v} \right)^{1/2} \quad (1)$$

where δ is the Hildebrand one-component solubility parameter, ΔE_v denotes the cohesive energy, which is introduced as the energy required for separating a molecule from its surrounded neighbors, v is the molar volume, ΔU_{vap} represents the energy change upon isothermal vaporization of the saturated liquid to the ideal gas state (energy of a complete vaporization), R and T represent universal gas constant and temperature, respectively, and ΔH_{vap} stands for the enthalpy of vaporization.¹⁷

Various theoretical efforts have been undertaken to represent/predict the solubility parameter, including the one-component or three-component solubility parameter (refer to the works of Hansen^{18,19} for more details about the definition of one/three-component solubility parameters). Fedors³⁴

Received: November 15, 2011

Revised: January 5, 2012

Accepted: January 12, 2012

correlated the solubility parameters and molar volumes of several chemical liquids and polymers. He stated that, for low molecular weight liquids, the solubility parameter can be well-calculated by the original definition (eq 1); however, regarding high molecular weight polymers, he used Small's additive group "molar-attraction constants" for its evaluation mainly because the volatility of such polymers is too low for calculation of cohesive energy from conventional methods.³⁴ The Carnahan–Starling equation of state (EoS) was applied by Lozada et al.³⁵ to predict the one-component solubility parameters of several chemicals in their supercritical state. Later, group-contribution-based methods were developed for calculation/estimation of the components of the three-component solubility parameter for a few chemical compounds.^{36–38}

Allada,³⁹ Panayiotou,⁴⁰ and Williams et al.⁴¹ are other researchers who determined the solubility parameter/hydrogen-bonding portion of the three-component solubility parameter using different theoretical approaches, e.g., the LFT (lattice fluid theory) or semiempirical correlations. Pursuing a theoretically correct algorithm, Bozdogan⁴² expressed that the partial molar entropy change of a polymer for mixing at a given temperature is proportional to the hydrodynamic volume or segment number of the polymer. Consequently, he calculated the solubility parameter of a high molar mass polymer at a specified temperature by extrapolating solubility parameter values of polymer fractions to high molar mass by applying the solubility parameter–segment number relation of the polymer fraction.⁴² Applying the Simha and Somcynsky (S–S) lattice-hole theory, Utracki and Simha⁴³ acceptably represented the pressure–volume–temperature (PVT) surface of chain molecular melts and consequently their solubility parameters.

Gharagheizi and co-workers²³ reduced the deviations of the Hansen sphere model predictions from the corresponding experimental values through study of the effects of the presented strategy on the values of the obtained solubility parameter. A model based on an equation of state was developed using the configurational partition function for determination of the partial solubility parameters of several compounds and polymers by Stefanis et al.⁴⁴ In addition, the PC SAFT (perturbed-chain statistical associating fluid theory) was used by Zeng et al.⁴⁵ to calculate the three-component solubility parameters of *n*-alkanes and 1-alcohols.

A quantitative structure–property relationship (QSPR)-based model was also proposed by Code et al.⁴⁶ to represent/predict the solubility parameters of small organic molecules. In a different approach, the one-third rule was utilized for calculation of the solubility parameters of hydrocarbons and crude oil systems by Vargas and Chapman.⁴⁷ In addition, Eslamimanesh and Esmailzadeh²⁰ have been the only researchers who evaluated the solubility parameters of different organic compounds at high temperatures using cubic equations of state. It should be noted that the aforementioned methods so far may not be taken into account as general models since they have been developed and checked on a limited number of chemical compounds/groups. Furthermore, the calculation procedures of some of these theories are not normally easy to follow.

Very recently, we presented several accurate models on the basis of group contributions²¹ and QSPR²² theories for calculation/estimation of the solubility parameters of around 1600 and 1400 compounds (mostly nonelectrolyte organic compounds because some inorganic compounds may be found

in the chemical families named as "other compounds"), respectively. As a consequence, we observed the average absolute relative deviations (AARDs) of the results from the corresponding data to be within the 1.5–4.6% range. However, although our previously proposed methods^{21,22} are accurate, they require computer programs for their usage.

Our literature survey demonstrates that there may be no corresponding states method available in the open literature for determination of the solubility parameters of pure chemicals. In this work, we present a corresponding states correlation for this purpose developed through application of the GEP strategy.^{6,7}

2. DATABASE

The solubility parameter values of about 1600 chemical compounds (mostly organic ones because there are some inorganic compounds present in the database) at 298.15 K and atmospheric pressure provided in the DIPPR 801 database⁴⁸ have been used for developing and validating the method.

3. DEVELOPING THE CORRESPONDING STATES METHOD

The GEP computational steps^{6,7} described in our previous works^{1–3} have been followed herein to develop the corresponding states method. The introduced corresponding states principle parameters, including critical pressure (P_c), critical temperature (T_c), reduced temperature (T_r), critical volume (V_c), acentric factor (ω), reduced normal boiling point temperature (T_{br}), and molecular weight (M_w), have been introduced as input parameters to the algorithm (decision variables of the optimization procedure). It is first assumed that the solubility parameter value can be formulated as a function of the aforementioned properties as follows:

$$\delta = f(T_c, T_r, P_c, V_c, \omega, T_{br}, M_w) \quad (2)$$

Having defined the probable input parameters of the method, the following computational steps have been pursued:^{1–3}

- (1) Initialization of the population, i.e., randomly generating the chromosomal structures of the individuals by setting various correlations presented as parse trees by applying the operators ($-$, $+$, \times , $/$) and terminals as functions of the input data and the output desired results (B values).
- (2) Calculation of the fitness value for every individual of the generated population by the following objective function (OF):

$$OF(i) = \frac{100}{N - n} \sum_i \frac{|\delta(i)^{rep/pred} - \delta(i)^{exp}|}{\delta(i)^{exp}} \quad (3)$$

where N stands for the number of data points used in the GEP^{6,7} algorithm, n denotes the number of correlation parameters, and the superscripts "rep/pred" and "exp" are the represented/predicted solubility parameter values by the final developed correlation and the applied data, respectively.

- (3) Selection of the individuals to stand for proper parents for replacement, which were evaluated from the fitness values. In this work, the tournament technique^{49,50} has been used to provide an acceptable diversity of the population in each generation.
- (4) Use of the genetic operators, including replication, mutation, and inversion, for gene reproduction with modification computational steps.

- (a) The replication operator copies exactly the chromosomes of the individuals chosen in the selection step (step 3).^{6,7}
 - (b) The mutation operator results in efficient adaptation of populations of individuals.^{6,7} In this study, the point mutation has been applied, in which a random node (in the structures of the chromosomes) is selected and the stored information is replaced with a different random primitive of the same arity taken from the initial (old) set.^{6,7} Having defined the mutation rate (p_m), the mutation can happen everywhere in the structural organization of the chromosomes, however with preservation of the original structure.^{6,7} The mutation can be normally performed through changing the heads of the gene symbols and terminals of the tails.^{6,7}
 - (c) The inversion operator is applied to create new individuals through modification of the heads of randomly selected genes. It has already been argued that all the new individuals created by inversion are considered as correct programs.^{6,7} The performance of this operator can be adapted by choosing a value for the inversion rate (p_i).^{6,7}
- (5) Transposition and insertion sequence elements: The transposable elements of gene expression programming are a part of the genome that can be activated and jumped to another place in the chromosome and include three types as implemented by Ferreira:^{6,7} short fragments with either a function or a terminal in the first position that transpose to the heads of genes, short fragments with a function in the first position that transpose to the roots of genes, and entire genes that transpose to the beginning of chromosomes.
- (6) Recombination: This step, which is conducted in three manners, including one-point recombination, two-point recombination, and gene recombination,^{6,7} randomly chooses two chromosomes to exchange specific materials with each other, leading to the appearance of two new chromosomes.^{6,7} Consequently, a new generation is created. The preceding procedure is repeated until the defined stopping criteria (can be user-defined convergence criteria or a maximum number of generations) are satisfied. The details of this procedure along with comprehensive examples are provided by Ferreira.^{6,7}

4. RESULTS AND DISCUSSION

An easy-to-use corresponding states correlation has been obtained by undertaking the described computational algorithm. In this work, the main data set⁴⁸ has been randomly divided into three sub data sets, the “training” set (1313 data points, about 80% of the whole data set), the “validation (optimization)” set (164 data points, about 10% of the whole data set), and the “test (prediction)” set (164 data points, about 10% of the whole data set), to develop and test the method. It is worth pointing out that the GEP^{6,7} technique determines the required parameters which yield the most accurate correlation from the introduced parameters (T_c , P_c , V_c , ω , T_{br} , M_w). Therefore, one can introduce several independent parameters for a particular problem and obtain the ones which have the

most positive effects on the desired output results. The final correlation can be reported as follows:

$$\delta = 10^{-3} \left(21.2483T_c - M_w - 5P_c + 67.2154 \left(\frac{768P_c - 64M_w + 2048P_c\omega}{T_c} \right) \right) \quad (4)$$

In eq 4, δ is calculated in $(\text{J}\cdot\text{cm}^{-3})^{0.5}$ or $(\text{MPa})^{0.5}$, P_c and T_c are in bar and kelvin, respectively, and M_w is in grams per mole. The numbers of digits in the coefficients in this correlation are normally determined through simultaneously performing sensitivity analysis of the calculated/estimated results of these values during the optimization processes.

The significant parameters of the applied GEP^{6,7} technique are reported in Table 1, including the numbers of treated

Table 1. Parameters of the GEP^{6,7} Algorithm Applied in the Computational Steps

GEP ^{6,7} algorithm param	value
number of chromosomes	20
head size	6
number of genes	6
linking function	+
number of generations without change	2000
fitness function	AARD, ^a %
mutation	0.044
inversion	0.1
IS transposition	0.1
RIS transposition	0.1
one-point recombination	0.3
two-point recombination	0.3
gene recombination	0.1
gene transposition	0.1
number of constants per gene	2
operators used	+, −, ×, /, √, exp ^b , log _e , power

^aAARD (%) = $(100/(N - n)) \sum_i^N (|\text{rep}(i)/\text{pred}(i) - \exp(i)|/\exp(i))$, where N is the number of applied data and n is the number of parameters. ^bexp = exponential.

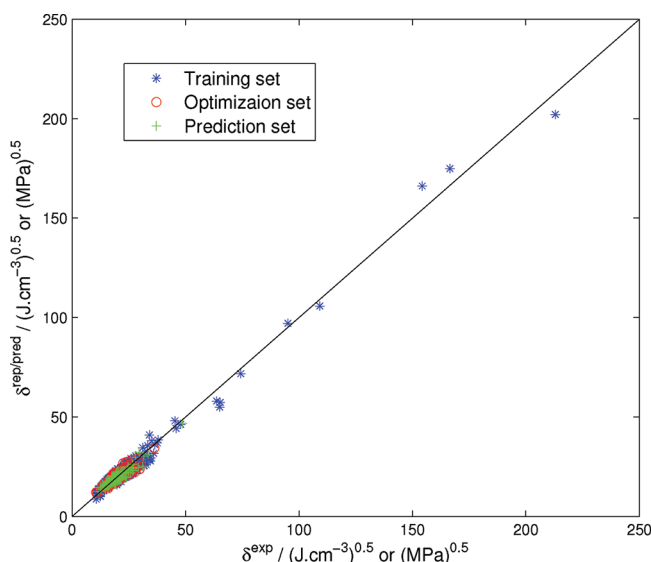
chromosomes and genes, the mutation and inversion coefficients, and the applied operators. The statistical parameters of the obtained results (Table 2) indicate that the average absolute relative deviation of the calculated/estimated solubility parameter values from the applied data⁴⁸ is about 6%. Therefore, it is found that the developed corresponding states method brings about acceptable accuracy for determination of the solubility parameters of the investigated chemical compounds at 298.15 K and atmospheric pressure. Comparisons between the determined solubility parameter values and the data and also the percent relative deviations of the obtained results are better illustrated in Figures 1 and 2. The detailed results using the developed method for all of the studied compounds accompanied by the values of the input parameters to the correlation are reported as Supporting Information.

In addition, the capability of the proposed method for determination of the solubility parameter values for different chemical families is reported as Supporting Information. Reasonable deviations from the applied data can be observed. This issue argues about the acceptable applicability of the proposed correlation for many of the chemical families (mostly

Table 2. Statistical Parameters of the Developed Corresponding States Method (Eq 4)

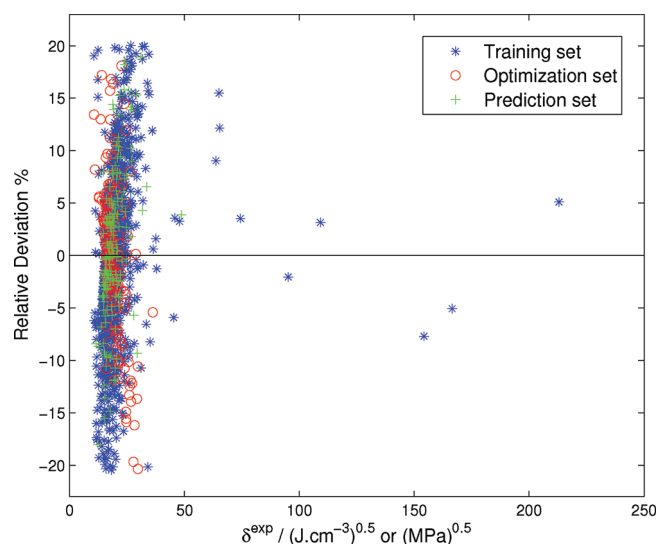
statistical param	value
Training Set	
AARD, ^a %	5.8
std dev error	1.67
root mean square error	1.68
N ^b	1313
Optimization Set	
AARD, %	5.8
std dev error	1.64
root mean square error	1.65
N	164
Prediction Set	
AARD, %	5.7
std dev error	1.50
root mean square error	1.53
N	164
Training Set + Optimization Set + Prediction Set	
AARD, %	5.7
std dev error	1.65
root mean square error	1.67
N	1641

^aAARD (%) = $(100/(N - n)) \sum_i^N (|rep(i)/pred(i) - exp(i)|/exp(i))$, where n is the number of model parameters. ^bNumber of data points.

**Figure 1.** Comparison between the represented/predicted results of the developed correlation (eq 4) and the applied data⁴⁸ of the solubility parameters of the investigated compounds at 298.15 K and atmospheric pressure.

organic and nonelectrolyte). It is obvious from Figures 1 and 2 that there are no data points in the prediction set for high values of solubility parameters (i.e., from about 50 to around 200 (MPa)^{0.5} with respect to the data pertaining generally to the solubility parameters of metallic elements such as Na, Ca, etc.). As already mentioned, the divisions of the data into three sub data sets are completely random. Therefore, this phenomenon has not been performed on purpose or for a specific reason.

It seems to be very fruitful here to point out some discussions on the aforementioned data for which the correlation does not result in accurate solubility parameter

**Figure 2.** Relative deviations of the represented/predicted solubility parameter values of the studied compounds by eq 4 from the corresponding data.⁴⁸

values (generally in the 13–16% absolute relative deviation (ARD) range). We may be able to relate these results to the data mainly from special chemical families (inorganic halides, some polyfunctional organics, nitriles, polyfunctional nitriles).

First, it is worth knowing that the computational steps in each generation of the GEP^{6,7} strategy need parallel computing and consequently high amounts of time. For development of the method, we have defined a stopping criterion for the algorithm, which is the difference between the accuracy of the obtained results from the current generation and the previous one in each step. Consequently, it is probable to develop a more accurate correlation through continuation of the calculation steps by producing more generations from the subsequent populations. However, careful investigation using more powerful computers should be performed to verify this concept.

Second, the applied solubility parameter values⁴⁸ contain definite uncertainties, which generally lie between 1% and 25% (see the Supporting Information). These uncertainties affect, indeed, the prediction capability of the obtained correlation.¹ We are well aware that there are several methods to detect the outlier data for the model as follows:

- (1) It is possible to eliminate the doubtful data from the training set and develop a new correlation. Later, these omitted data can be predicted using the new correlation. If these predictions are again within the unacceptable deviations from the data (i.e., with almost the same ARDs (%) obtained from application of the previous correlation), we may conclude there is a high probability these solubility parameter data are outliers. On the contrary, if the new correlation acceptably predicts these data, we may not be able to report them as the experimental outliers of the model, and consequently, the correlation would not be more accurate than the current one through elimination of these points.
- (2) Numerical methods such as neural networks^{51,52} or support vector machines^{53–55} could be used on the whole database in the training step only to try to identify the experimental outliers of the correlation (the GEP algorithm can also be used for this method; however, it

may not be possible at this time because of the high computational time as already stated). Later, a new data set could be constructed by eliminating the outliers. This time, the new data set would be used for the training, optimization, and prediction test steps.

- (3) There are other statistical methods for definition of the outliers that are independent of the new divisions (data allocations) of the data in the three sub data sets, which need very detailed mathematical procedures.

Each of the preceding solutions can be pursued to determine the outliers of the developed method. However, our objective, in this work, has been to investigate the ability of the investigated corresponding states correlation for representation/prediction of the whole solubility parameter values from one of the most comprehensive data sets in the literature.⁴⁸ These phenomena reveal some broad recommendations to improve the correlation's predictive capability in future works. It should be noted that the uncertainties of the data⁴⁸ could be normally reduced by designing more appropriate equipment and careful calibration of the measuring devices.^{1–3}

5. CONCLUSION

In this study, the gene expression programming^{6,7} mathematical algorithm was utilized to develop an easy-to-use corresponding states method for determination of the solubility parameters of about 1600 chemical compounds (mostly nonelectrolyte organic compounds) at 298.15 K and atmospheric pressure. The critical temperature, critical pressure, molecular weight, and acentric factor are the input parameters of the method. A total of 1477 solubility parameter data (about 90% of the whole data set) and about 164 solubility parameter data (around 10% of the whole data set) were applied for developing and predictability testing the proposed corresponding states correlation, respectively. The statistical parameters of the obtained correlation show about 6% average absolute relative deviation of the results from the corresponding data.⁴⁸ Investigation of the capability of the proposed method for evaluation of the solubility parameter values of many groups of chemicals at 298.15 K and atmospheric pressure demonstrates its wide range of applicability compared with the available methods in the literature. Indeed, using more accurate experimental values (in the case of availability), considering more calculation time for the GEP^{6,7} algorithm to converge to a more probable global optimum of the objective function of the problem, and elimination of the probable outliers from the main data set shall contribute to developing more predictive tools for the same purpose.^{1–3}

■ ASSOCIATED CONTENT

● Supporting Information

All of the evaluated solubility parameter values of the investigated compounds accompanied by their absolute relative deviations from the data.⁴⁸ This material is available free of charge via the Internet at <http://pubs.acs.org/>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: amir-hossein.mohammadi@mines-paristech.fr. Tel.: + (33) 1 64 69 49 70. Fax: + (33) 1 64 69 49 68.

■ ACKNOWLEDGMENTS

A.E. is grateful to MINES ParisTech for providing him with a Ph.D. scholarship.

■ REFERENCES

- (1) Gharagheizi, F.; Eslamimanesh, A.; Sattari, M.; Mohammadi, A. H.; Richon, D. Prediction of the surface tension of chemical compounds through a novel corresponding states mode. Private communication, 2012.
- (2) Gharagheizi, F.; Eslamimanesh, A.; Sattari, M.; Mohammadi, A. H.; Richon, D. Corresponding states method for determination of the viscosity of gases at atmospheric pressure. *Ind. Eng. Chem. Res.* Article in press. DOI: 10.1021/ie202591f
- (3) Gharagheizi, F.; Eslamimanesh, A.; Sattari, M.; Mohammadi, A. H.; Richon, D. Corresponding states method for determination of the second virial coefficient using the gene expression programming strategy. Private communication, 2012.
- (4) Eslamimanesh, A.; Gharagheizi, F.; Mohammadi, A. H.; Richon, D.; Illbeigi, M.; Fazlali, A.; Forghani, A. A.; Yazdizadeh, M. Phase equilibrium modeling of structure H clathrate hydrates of methane + water “insoluble” hydrocarbon promoter using group contribution-support vector machine technique. *Ind. Eng. Chem. Res.* **2011**, *50*, 12807–12814.
- (5) Eslamimanesh, A.; Gharagheizi, F.; Mohammadi, A. H.; Richon, D. Phase equilibrium modeling of structure H clathrate hydrates of methane + water “insoluble” hydrocarbon promoter using QSPR molecular approach. *J. Chem. Eng. Data* **2011**, *56*, 3775–3793.
- (6) Ferreira, C. *Gene Expression Programming*, 2nd revised and extended ed.; Springer-Verlag: Dordrecht, The Netherlands, 2006.
- (7) Ferreira, C. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst.* **2001**, *13*, 87–129.
- (8) Holland, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*; University of Michigan Press: Ann Arbor, MI, 1975; 2nd ed.; MIT Press: Cambridge, MA, 1992.
- (9) Cramer, N. L. A representation for the adaptive generation of simple sequential programs. In *Proceedings of the First International Conference on Genetic Algorithms and Their Applications*, Pittsburgh, PA, July 24–26, 1985; Grefenstette, J. J., Ed.; Erlbaum: Mahwah, NJ, 1985; pp 183–187.
- (10) Koza, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*; MIT Press: Cambridge, MA, 1992.
- (11) Si, H.; Zhao, J.; Cui, L.; Lian, N.; Feng, H.; Duan, Y. B.; Hu, Z. Study of human dopamine sulfotransferases based on gene expression programming. *Chem. Biol. Drug Des.* **2011**, *78*, 370–377.
- (12) Azamathulla, H. M.; Ghani, A. A.; Leow, C. S.; Chang, C. K.; Zakaria, N. A. Gene-expression programming for the development of a stage-discharge curve of the Pahang River. *Water Resour. Manage.* **2011**, *25*, 2901–2916.
- (13) Özcan, F. Gene expression programming based formulations for splitting tensile strength of concrete. *Constr. Build. Mater.* **2012**, *26*, 404–410.
- (14) Dong, Z.; Zhu, Y. Transformer fault diagnosis based on gene expression programming classifier. *Adv. Mater. Res.* **2012**, *354–355*, 1022–1026.
- (15) Kisi, O.; Emin Emiroglu, M.; Bilhan, O.; Guven, A. Prediction of lateral outflow over triangular labyrinth side weirs under subcritical conditions using soft computing approaches. *Expert Syst. Appl.* **2012**, *39*, 3454–3460.
- (16) Scatchard, G. Equilibria in non-electrolyte solutions in relation to the vapor pressures and densities of the components. *Chem. Rev.* **1931**, *8*, 321–333.
- (17) Hildebrand, J. H.; Scott, R. L. *The Solubility of Nonelectrolytes*; Reinhold Publishing Corp.: New York, 1950.
- (18) Hansen, C. M. The three dimensional solubility parameter—key to paint component affinities: I. Solvents, plasticizers, polymers, and resins. *J. Paint Technol.* **1967**, *39*, 104–117.

- (19) Hansen, C. M. The universality of the solubility parameter. *Ind. Eng. Chem. Prod. Res. Dev.* **1969**, *8*, 2–11.
- (20) Eslamimanesh, A.; Esmailzadeh, F. Estimation of solubility parameter by the modified ER equation of state. *Fluid Phase Equilib.* **2010**, *291*, 141–150.
- (21) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Group contribution-based method for determination of solubility parameter of nonelectrolyte organic compounds. *Ind. Eng. Chem. Res.* **2011**, *50*, 10344–10349.
- (22) Gharagheizi, F.; Eslamimanesh, A.; Farjood, F.; Mohammadi, A. H.; Richon, D. Solubility parameters of nonelectrolyte organic compounds: Determination using quantitative structure-property relationship strategy. *Ind. Eng. Chem. Res.* **2011**, *50*, 11382–11395.
- (23) Gharagheizi, F.; Sattari, M.; Angaji, M. T. Effect of calculation method on values of Hansen solubility parameters of polymers. *Polym. Bull.* **2006**, *57*, 377–384.
- (24) Mousavi-Dehghani, S. A.; Mirzayi, B.; Mousavi, S. M. H.; Fasih, M. An applied and efficient model for asphaltene precipitation in production and miscible gas injection. *Pet. Sci. Technol.* **2010**, *28*, 113–124.
- (25) Novosad, Z.; Costain, T. G. Experimental and modeling studies of asphaltene equilibria for a reservoir under CO₂ injection. *Proceedings of the 65th Annual Technical Conference and Exhibition of the SPE*, New Orleans, LA, Sept 23–26, 1990; Society of Petroleum Engineers: Richardson, TX, 1990; SPE 20530.
- (26) Nor-Azian, N.; Adewumi, M. A. Development of asphaltene phase equilibrium predictive model. *Proceedings of the Eastern Regional Conference and Exhibition of the SPE*, Pittsburgh, PA, Nov 2–4, 1993; Society of Petroleum Engineers: Richardson, TX, 1993; SPE 26905.
- (27) MacMillan, D. J.; Tackett, J. E.; Jesse, M. A.; Monger-McClure, T. G. A unified approach to asphaltene precipitation: Laboratory measurement and modeling. *Proceedings of the SPE International Symposium on Oilfield Chemistry*, San Antonio, TX, Feb 14–17, 1995; Society of Petroleum Engineers: Richardson, TX, 1995; SPE 28990.
- (28) Mohammadi, A. H.; Eslamimanesh, A.; Richon, D. Mono-disperse thermodynamic model based on chemical + Flory-Huggins polymer solution theories for predicting asphaltene precipitation. *Ind. Eng. Chem. Res.* **2012**, Article in press. DOI: DOI:10.1021/ie202737p.
- (29) Alboudwarej, H.; Akbarzadeh, K.; Beck, J.; Svercek, W. Y.; Yarranton, H. W. Regular solution model of asphaltene precipitation from bitumen. *AIChE J.* **2003**, *11*, 2948–2956.
- (30) Akbarzadeh, K.; Alboudwarej, H.; Svercek, H. Y.; Yarranton, H. W. A generalized regular solution model for the prediction of asphaltene precipitation from n-alkane diluted heavy oils and bitumens. *Fluid Phase Equilib.* **2005**, *232*, 159–170.
- (31) Yarranton, H. W.; Masliyah, J. H. Molar mass distribution and solubility modeling of asphaltenes. *AIChE J.* **1996**, *42*, 3533–3543.
- (32) Mohammadi, A. H.; Richon, D. The Scott-Magat polymer theory for determining onset of precipitation of dissolved asphaltene in the solvent + precipitant solution. *Open Thermodyn. J.* **2008**, *2*, 13–16.
- (33) Hildebrand, J. H. Solubility: XII. Regular Solutions. *J. Am. Chem. Soc.* **1929**, *51*, 66–80.
- (34) Fedors, R. F. A method for estimating both the solubility parameter and molar volumes of liquids. *Polym. Eng. Sci.* **1974**, *14*, 147–154.
- (35) Lozada, C. M.; Monfort, J. P.; Del Río, F. Calculation of solubility parameters from an equation of state. *Chem. Phys. Lett.* **1977**, *45*, 130–133.
- (36) Hansen, C. M.; Beerbower, A. Solubility parameters. In *Kirk-Othmer Encyclopedia of Chemical Technology*, Suppl. Vol., 2nd ed.; Standen, A., Ed.; Interscience: New York, 1971.
- (37) Beerbower, A. Environmental capability of liquid lubricants. *Interdisciplinary Approach to Liquid Lubricant Technology*; NASA Publication SP-318; National Aeronautics and Space Administration: Washington, DC, 1973.
- (38) Van Krevelen, D. W.; Hoftyzer, P. J. *Properties of Polymers: Their Estimation and Correlation with Chemical Structure*, 2nd ed.; Elsevier: Amsterdam, 1976.
- (39) Allada, S. R. Solubility parameters of supercritical fluids. *Ind. Eng. Chem. Prod. Res. Dev.* **1984**, *23*, 344–348.
- (40) Panayiotou, C. Solubility parameter revisited: An equation-of-state approach for its estimation. *Fluid Phase Equilib.* **1997**, *131*, 21–35.
- (41) Williams, L. L.; Rubin, J. B.; Edwards, H. W. Calculation of Hansen solubility parameter values for a range of pressure and temperature conditions, including the supercritical fluid region. *Ind. Eng. Chem. Res.* **2004**, *43*, 4967–4972.
- (42) Bozdogan, A. E. A method for determination of thermodynamic and solubility parameters of polymers from temperature and molecular weight dependence of intrinsic viscosity. *Polymer* **2004**, *45*, 6415–6424.
- (43) Utracki, L. A.; Simha, R. Statistical thermodynamics predictions of the solubility parameter. *Polym. Int.* **2004**, *53*, 279–286.
- (44) Stefanis, E.; Tsivintzelis, I.; Panayiotou, C. The partial solubility parameters: An equation-of-state approach. *Fluid Phase Equilib.* **2006**, *240*, 144–154.
- (45) Zeng, Z.-Y.; Xu, Y.-Y.; Li, Y.-W. Calculation of solubility parameter using perturbed-chain SAFT and cubic-plus-association equations of state. *Ind. Eng. Chem. Res.* **2008**, *47*, 9663–9669.
- (46) Code, J. E.; Holder, A. J.; Eick, J. D. Direct and indirect quantum mechanical QSPR Hildebrand solubility parameter models. *QSAR Comb. Sci.* **2008**, *27*, 841–849.
- (47) Vargas, F. M.; Chapman, W. G. Application of the one-third rule in hydrocarbon and crude oil systems. *Fluid Phase Equilib.* **2010**, *290*, 103–108.
- (48) *Evaluated Process Design Data*; Project 801, Public Release Documentation; Design Institute for Physical Properties (DIPPR), American Institute of Chemical Engineers (AIChE): New York, 2006.
- (49) Cai, W.; Pacheco-Vega, A.; Sen, M.; Yang, K. T. Heat transfer correlations by symbolic regression. *Int. J. Heat Mass Transfer* **2006**, *49*, 4352–4359.
- (50) Goldberg, D. E.; Deb, K. A comparison of selection schemes used in genetic algorithms. In *Foundations of Genetic Algorithms*; Rawlins, G. J. E., Ed.; Morgan Kaufmann: San Mateo, CA, 1991; pp 69–93.
- (51) Eslamimanesh, A.; Gharagheizi, F.; Mohammadi, A. H.; Richon, D. Artificial neural network modeling of solubility of supercritical carbon dioxide in 24 commonly used ionic liquids. *Chem. Eng. Sci.* **2011**, *66*, 3039–3044.
- (52) Hagan, M.; Demuth, H. B.; Beale, M. H. *Neural Network Design*; International Thomson Publishing: Boston, MA, 2002.
- (53) Suykens, J. A. K.; Van Gestel, T.; De Brabanter, J.; De Moor, B.; Vandewalle, J. *Least Squares Support Vector Machines*; World Scientific: Singapore, 2002.
- (54) Suykens, J. A. K.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300.
- (55) Pelckmans, K.; Suykens, J. A. K.; Van Gestel, T.; De Brabanter, D.; Lukas, L.; Hamers, B.; De Moor, B.; Vandewalle, J. *LS-SVMLab: A Matlab/C Toolbox for Least Squares Support Vector Machines*; Internal Report 02-44; ESAT-SISTA, Katholieke Universiteit Leuven: Leuven, Belgium, 2002.