Calculating Hansen Solubility Parameters of Polymers with Genetic Algorithms

Guilherme Cañete Vebber,[1] Patricia Pranke,[2] Cláudio Nunes Pereira[3]

[1]Instituto de Física, Universidade Federal do Rio Grande do Sul (UFRGS), Avenida Bento Gonçalves 9500, Caixa Postal 15051, CEP 91501-970, Porto Alegre, Rio Grande do Sul, Brazil

[2]Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Avenida Ipiranga 2752, CEP 90610-000, Porto Alegre, Rio Grande do Sul, Brazil

[3]Tecnano Pesquisas e Serviços Ltda, ITCientec - Incubadora Tecnológica Cientec - Fundação de Ciência e Tecnologia, Rua Washington Luiz 675, CEP 90010-460, Porto Alegre, Rio Grande do Sul, Brazil

Correspondence to: C. N. Pereira (E-mail; claudio@tecnano.com.br).

ABSTRACT: In materials engineering, it is often essential to know what are the best solvents to process any polymer, and employing methods based on Hansen solubility parameters are an effective way to find them. In this work the Hansen solubility parameters of polyether sulfone, lignin, and bitumen have been calculated by an alternative optimization procedure. It has been shown that, by applying an evolutionary strategy to Hansen's correlation method, it is possible to improve the fitting of solubility spheres. Compared with previous calculations, most *quality-of-fit* parameters are optimized. As a result, the sphere radii are reduced and, except for lignin, at least one of the solubility parameters is considerably changed (by 0.9–1.5 MPa$^{1/2}$). Shortcomings of the correlation method are also pointed out, such as lack of data quality evaluation on set of solvents and uncertain character for partially solving solvents. At least the former could be handled by a proposed parameter called *fill factor*. © 2013 Wiley Periodicals, Inc. J. Appl. Polym. Sci. 000: 000–000, 2013

INTRODUCTION

In polymer science and engineering, preparing stable solutions is a fundamental step to succeed in the processing of advanced materials. Besides various physical and chemical properties involved (viscosity, surface tension, density, etc.), the fine selection of a set of solvents that are able to solubilize the solute of interest, represents a huge step towards the achievement of a thermodynamically stable solution. In this context, a computational tool for helping in this selection has been developed by calculating the Hansen solubility parameters of any kind of solute based on an empirical solubility map.

Hansen solubility parameters (HSPs) are numerical values associated with energies or intermolecular cohesive forces binding molecules in a substance. They have direct applications in several scientific disciplines, such as surface science, where they have been used to characterize the wettability of various surfaces and the adsorption properties of pigment surfaces.

In this work, genetic algorithms were applied to calculate HSPs of polymers and compare results with that of the literature. As no well-established standard, besides Hansen's is available to qualify results obtained by different methods and different researchers, a discussion is made about which criteria are good enough to guarantee reliable data.

Hansen Solubility Parameters

Before introducing concepts and the theory developed by Charles Hansen, it is beneficial to give a short physical argument about the solvation process.[1] For any solution to occur spontaneously, thermodynamics requires that the Gibbs free energy change in the mixing process be zero or negative. In a simplified description, the Gibbs free energy ΔG^M can be divided in two parts:

$$\Delta G^M = \Delta G^M_{\text{comb}} + \Delta G^M_{\text{noncomb}} \qquad (1)$$

where the combinatorial part ΔG^M_{comb} is related to the combinatorial entropy change ΔS^M_{comb} (that results by simply mixing components) by:

$$\Delta G^M_{\text{comb}} = -T\Delta S^M_{\text{comb}} \qquad (2)$$

and the noncombinatorial part $\Delta G^M_{\text{noncomb}}$ refers to all other free energy effects, including the heat of mixing, and can be calculated by:

$$\Delta G^M_{noncomb} = \varphi_1 \varphi_2 V_M (\delta_1 - \delta_2)^2 \qquad (3)$$

In the above equations, T is the absolute temperature, ϕ_1 and ϕ_2 are volume fractions of solvent and polymer, V_M is the volume of the mixture and δ_1 and δ_2 are called solubility parameters of solvent and polymer.

By analysing eqs. (1–3), one can see that the difference in solubility parameters for the solvent-solute pair is important in determining the system energy equilibrium. It is clear that a match in solubility parameters leads to a zero change in noncombinatorial free energy and the positive entropy change (the combinatorial entropy change), found in simple mixing to result in a disordered mixture compared to the pure components, will ensure that a solution is possible from a thermodynamic point of view. The maximum difference in solubility parameters that can be tolerated for a solution that still occurs corresponds to the case $\Delta G^M = 0$. From eqs. (1–3), this condition can be expressed as:

$$(\delta_1 - \delta_2)^2 = (T\Delta S^M_{comb})/(\varphi_1 \varphi_2 V_M) \qquad (4)$$

Equation (4) clearly shows that an alternate view of the solubility situation at its limit is that it is the entropy change that dictates how closely the solubility parameters must match each other for the mixture to stabilize.

Solubility parameters are sometimes called cohesion energy parameters as they derive from the energy required to convert a liquid to a gas state. All types of bonds holding molecules together are broken in the vaporization process. Thus the energy of vaporization is a direct measure of the total cohesive energy holding the liquid molecules together and both can be considered as being identical to each other.

The term solubility parameter was first used by Hildebrand in 1949.[2] The Hildebrand solubility parameter is defined as the square root of the total cohesive energy density:

$$\delta = (E/V)^{1/2} \qquad (5)$$

where V is the molar volume of the pure solvent and E is its (measurable) energy of vaporization.

The Hildebrand solubility parameters are applicable to the regular solutions, which implies strictly nonpolar systems. As this is not the case for most real systems, other researchers have tried to extend Hildebrand's original idea, splitting the Hildebrand solubility parameter into two or more components. The approach proposed by Charles Hansen in 1967 is one of the most proeminent in this regard.[3]

The basis of Hansen solubility parameters (HSPs) is that the total cohesive energy (E) of a liquid consists of three major intermolecular interactions: (nonpolar) dispersion forces, (polar) permanent dipole–permanent dipole forces, and (polar) hydrogen bonding. The most general is the nonpolar cohesive energy (E_D), derived from induced dipole forces and also called atomic or dispersion interactions in scientific literature. All molecules contain these types of attractive forces. The second type is the polar cohesion energy (E_P), which results from inherently molecular interactions and is essentialy found in polar (noncentrosymmetric) molecules. The molecular dipole moment is the primary parameter used to calculate it. The third
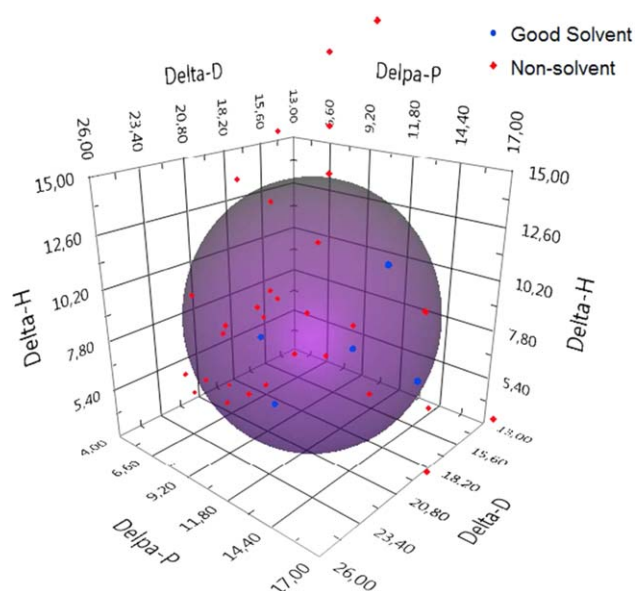
Figure 1. Representation of solvents (points) and a polymer (solubility sphere) in Hansen space. Delta-D, Delta-P, and Delta-H are, respectively, the dispersion (nonpolar), permanent-dipole, and hydrogen bonding solubility parameters. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

major cohesive energy source (E_H) comes from hydrogen bonds which, according to a modern definition,[4] are "attractive interactions between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H and an atom or a group of atoms in the same or a different molecule in which there is evidence of bond formation".

Therefore, the basic equation governing the assignment of Hansen parameters is that E must be the sum of the individual energies that make it up:

$$E = E_D + E_P + E_H \qquad (6)$$

Dividing this by the molar volume gives the square of the total (or Hildebrand) solubility parameter as the sum of the squares of Hansen components:

$$E/V = (E_D/V) + (E_P/V) + (E_H/V) \qquad (7)$$

$$\delta^2 = \delta_D^2 + \delta_P^2 + \delta_H^2 \qquad (8)$$

As shown by Figure 1, any molecular substance can be represented by a point in a tridimensional space, called Hansen space, whose orthogonal axes are HSPs ($x = \delta_d$, $y = \delta_p$, $z = \delta_h$).[1] Within Hansen space, a solute is represented not only by their HSPs, as with solvents, but also by an interaction radius (R_0), thus defining a solubility sphere whose centre coordinates are the HSPs (δ_d, δ_p, δ_h). All substances qualified as good solvents for the solute should stay within this sphere and all considered bad (nonsolvents) should lie outside. A useful parameter for comparing two substances is the solubility parameter distance (R_a), based on their respective HSP components:

$$R_a = \sqrt{4(\delta_{D2} - \delta_{D1})^2 + (\delta_{P2} - \delta_{P1})^2 + (\delta_{H2} - \delta_{H1})^2} \qquad (9)$$

The constant "4" is justified in theory as a convenient factor for visualizing spherical, rather than ellipsoidal regions of solubility.
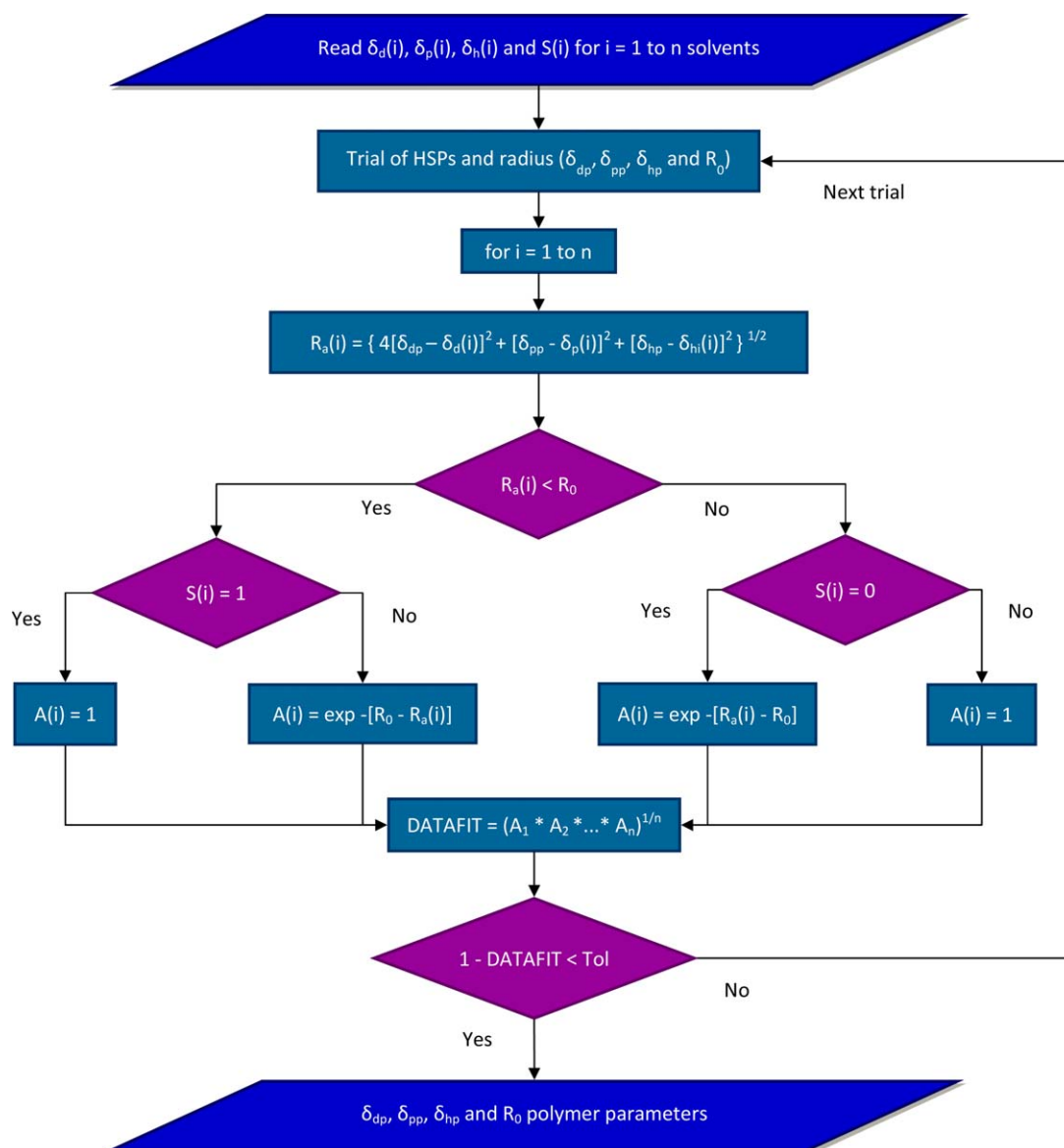
Figure 2. Basic flowchart for the implemented algorithm (adapted from reference [7]). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

It is obvious that solubility, or high affinity, requires $R_a < R_0$, so a RED (Relative Energy Difference) number is often used to quantify distances R_a relative to the interaction radius R_0:

$$\mathrm{RED} = \frac{R_a}{R_b} \qquad (10)$$

By this definition, $\mathrm{RED} = 0$ is equivalent to no energy difference, $\mathrm{RED} < 1$ indicates high solute-solvent affinity, $\mathrm{RED} > 1$ indicates low affinity and $\mathrm{RED} = 1$ (or around 1) reflects a boundary condition.

The reliability of the spherical characterization and the need to divide the total cohesion energy into at least three parts has been confirmed systematically by locating nondissolving solvents that are able to dissolve a given polymer when mixed. They only need to be located on opposite sides of the polymer sphere of solubility. This result also demonstrates that the mixing of the solvents follows a geometric mean rule.

The Correlation Method

An advanced method to compute HSPs and the interaction radius of solutes is based on solubility/swelling tests with many solvents. There are similar simpler methods which rely on pure statistical analysis[5] or correlation with intrinsic viscosities[6] but the computational ones can be much more effective. The HSPs of most substances are determined by experimental procedures combined with theoretical models, or simply group contribution techniques, which have been optimized by fitting and regression methods as the amount of collected data increases. The HSPs of solutes, on the other hand, are usually calculated by testing the degree of interaction with a large variety of solvents (qualitatively evaluated by means of
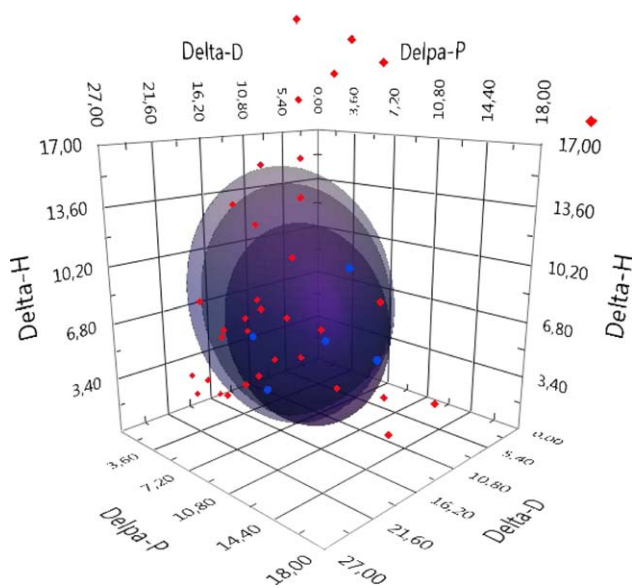
Figure 3. Solubility spheres for polyether sulfone calculated with downhill-simplex, Hansen, and genetic optimization algorithms (bigger to smaller, respectively). Blue sphere points indicate good solvents and red diamond points bad solvents. Data values from Table I. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

solubility, swelling or some other physical change) and applying computer algorithms to correlate empirical tests with solvent HSPs.

In this work a general algorithm is applied, similar to the one proposed by Hansen[1] and used by other groups[7,8] for calculating HSPs of polymers. Figure 2 illustrates a flowchart of this algorithm. The input data to the computer program is a list of n solvents for which there is experimental data in the correlation ($i = 1, 2, ..., n$), their known HSPs ($\delta_{di}, \delta_{pi}, \delta_{hi}$) and solubility factors ($S_i = 0, 1$). The solubility factor indicates the degree of interaction between polymer and solvents. Zero stands for insoluble or no swelling, and 1 for soluble.

The program starts with a set of random trials for the polymer HSPs and interaction radius ($\delta_d, \delta_p, \delta_h$, and R_0), all limited to a suitable range. For each set of HSPs and radius, the program evaluates the input data using a quality-of-fit function (originally callled Desirability Function) named as the *DATAFIT* function by Hansen, which has the form:

$$DATAFIT = \sqrt[n]{(A_1 \times A_2 \times \cdots \times A_n)} \quad (11)$$

where n is the number of solvents and the A_i quotients are given by:

$$A_i = e^{-[\text{ERROR DISTANCE}\,(i)]} \quad (12)$$

where ERROR DISTANCE(i) is the error distance of the solvent to the sphere boundary, so it depends on R_0, R_{ai}, and S_i. The distance R_{ai} is the already introduced solubility parameter distance, calculated by eq. (9), which in this case measures the distance from a given solvent point to the sphere center.

The A_i quotient for any good solvent ($S_i = 1$) inside the sphere ($R_{ai} < R_0$) or any bad solvent ($S_i = 0$) outside the sphere ($R_{ai} > R_0$)

equals 1, since the error distance is 0. For a good solvent ($S_i = 1$) outside the sphere ($R_{ai} > R_0$), the A_i quotient is given by:

$$A_i = e^{-(R_{ai} - R_0)} \quad (13)$$

Similarly, a bad solvent ($S_i = 0$) inside the sphere ($R_{ai} < R_0$) gives the following contribution to DATAFIT:

$$A_i = e^{-(R_0 - R_{ai})} \quad (14)$$

Thus all A_i quotients are within the range (0–1), limiting the DATAFIT function to the same range, according to eq. (11). DATAFIT approaches 1 as the fitting improves during optimization, reaching this end value just when all the good solvents are inside and all the bad ones are outside the sphere. It is common that the DATAFIT doesn't reach 1, so it is considered a tolerance for the maximum difference to 1 (or minimum DATAFIT) below which optimization stops.

Improving Optimization with Genetic Algorithms

The main difference between the approach followed here and the others is related to the optimization process. Hansen's deterministic approach starts at a definite point and evolutes by averaging points at the corners of cubes (of varying sizes) and choosing the best corners as new centers.[1] Gharagheize[7] and Redelius[8] apply different deterministic numerical methods in MATLAB environment, the *Nelder-Mead (Downhill Simplex)* algorithm[9] and the *Sequential Quadratic Programming*.[10] On the other hand, in this work a stochastic evolutionary strategy is used, which spans the subspace of interest in Hansen space and is considered more effective, since it is less susceptible to converge to local optima.[11]
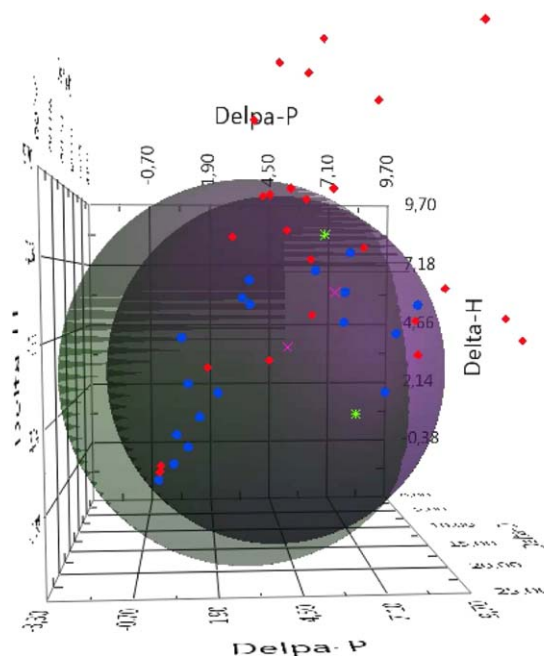
Figure 4. Solubility spheres for bitumen 2 calculated with Hansen and genetic optimization algorithms (bigger to smaller, respectively). Points: blue spheres locate good solvents, red diamonds bad solvents, green asterisks and violet X outliers. Data values from Table I. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Genetic algorithms are heuristic stochastic global optimization methods successfully applied in diverse fields, like aerodynamic design problems,[12] data mining,[13] vehicle routing,[14] climate control,[15] and quantum control of atomic/molecular dynamics.[16] In analogy to evolution in nature, they mimic natural processes of biological evolution, such as selection of individuals, mutation and recombination of genes.[17] In the struggle for life, the genetic merit of individuals of a species influence their chance of survival. After successive generations, each species evolves following a natural selection (the selection of the fittest), in which genetic variants more adapted to environmental conditions reproduce preferentially in following generations.

Any genetic algorithm can promote competition between, say, different sets of polymer HSPs and interaction radius (the species individuals). Selection of the best individual after several algorithm iteration cycles (the evolution), is guided according to the purpose for which it is applied; in the case in hand, DATAFIT maximization. Therefore, DATAFIT value is the chosen individual fitness parameter, a comparative quality factor between different HSPs and radius sets (like strength, beauty, intelligence and immunity, among others, which are fitness factors for animal species). Product of interaction between genetic qualities and the environment, the fitness factor is what determines whether the individual is good enough to reproduce.

EXPERIMENTAL

In the present work, the implemented genetic algorithm starts with a large initial population (usually 10,000), from which the best (1,024) are selected to be the first parent set. A percentage (95%) of parents has a percentage (100%) of the 4 genes (3 HSPs and 1 R_0) mutated by a random top-limited (1) amount before they recombine genes (2 crossover points ramdomly chosen) and generate the next population (8,192). From this generation, again the best are selected, having their genes mutated and recombined to produce the following generation. This evolutionary process repeats cyclically until DATAFIT is higher than a tolerance factor (based on previous tests). At the end, the best individual gives ideal HSPs and radius for the polymer.

In order to keep the sphere radius as small as possible, it is multiplied to DATAFIT a size factor function, which acts to decrease the DATAFIT value for larger radii R_0. So the real DATAFIT used for optimization is slightly different from the one presented above. The higher the R_0, the lower the DATAFIT. In this way, as in Hansen's program, the output is for the least radius allowing the maximum DATAFIT. Although this feature affects results decisively, the way it is implemented is not discussed by Hansen and not even mentioned by other authors. In the actual implementation, DATAFIT is multiplied by the function:

$$\text{Size Factor} = (R_0)^{-1/m} \tag{15}$$

where m is an empirical degree of the mth root, which in most calculations performed, has been set to around 20.

Applying this optimization approach to the correlation method, the solubility spheres of polyether sulfone, lignin, and bitumen were calculated and compared with results from the literature. As there are not many tested polymers, the chosen ones were selected by considering the completeness of published data.

RESULTS AND DISCUSSION

The main results with calculated and compared HSPs for the selected polymers are shown in Table I. DATAFIT is the original function, given by eq. (11). Unlike comparative works, two extra criteria of quality are exhibited: *sensibility* and *specificity*. The former evaluates the ratio of good solvents in the expected region (inside the solubility sphere) to the total good solvents; likewise, the latter represents the ratio of bad solvents outside the sphere to the total bad solvents. The parameter *outliers* corresponds to the ratio of solvents out of the expected region (i.e., good solvents outside the sphere and bad solvents inside it) to the total solvents, and is complementary for both good and bad solvents.

For each polymer presented in Table I, deviations between comparative results are analyzed by generating another table with all tested solvents, their HSPs and solubility factors and comparing

Table I. Comparative Results Between Different Calculation Methods for the Selected Polymers

Polymer	Algorithm	δ_D (MPa$^{1/2}$)	δ_P (MPa$^{1/2}$)	δ_H (MPa$^{1/2}$)	R_0 (MPa$^{1/2}$)	DATAFIT	Nr.	Sensibility (%)	Specificity (%)	Outliers (%)
Polyether sulfone	Hansen[18]	19.6	10.8	9.2	6.2	0.999	41	100.0	97.2	2.4
	Downhill-Simplex[7]	20.09	10.63	9.57	6.72	1.000	41	100.0	100.0	0.0
	Genetic	18.84	11.22	7.95	5.42	1.000	41	100.0	100.0	0.0
Bitumen 1	Sequential Quadratic Programming[8]	18.40	3.90	3.60	5.76	0.980	48	94.7	93.1	6.2
	Genetic	18.66	4.79	3.45	5.94	0.989	48	94.7	96.6	4.2
Bitumen 2	Hansen[19]	18.6	3.0	3.4	6.3	0.977	48	85.0	85.7	14.6
	Genetic	18.33	3.99	3.49	5.71	0.980	48	90.0	92.9	8.3
Lignin	Hansen[20]	21.9	14.1	16.9	13.7	0.990	82	87.5	98.5	3.7
	Genetic	21.71	14.18	16.93	13.45	0.990	82	93.8	98.5	2.4

Table II. List of Solvents in the Correlation for HSP Calculation of Polyether Sulfone

Solvent	δ_D	δ_P	δ_H	Solubility	RED1	RED2 [7]	RED3 [18]
Methyl-2-pyrrolidone	18.0	12.3	7.2	1	0.393	0.758	0.655
Dimethyl_formamide	17.4	13.7	11.3	1	0.934	0.958	0.915
Acetophenone	19.6	8.6	3.7	1	0.962	0.936	0.955
Methylene_dichloride	18.2	6.3	6.1	1	0.997	1.000	0.990
Dimethyl_sulfoxide	18.4	16.4	10.2	0[a]	1.054	1.000	0.996
g-Butyrolactone	19.0	16.6	7.4	1	0.999	1.000	0.998
Ethylene_dichloride	19.0	7.4	4.1	0	1.002	1.000	1.007
Isophorone	16.6	8.2	7.4	0	1.001	1.147	1.094
o-Dichlorobenzene	19.2	6.3	3.3	0	1.255	1.165	1.204
Tetrahydrofuran	16.8	5.7	8.0	0	1.265	1.246	1.237
Diacetone_alcohol	15.8	8.2	10.8	0	1.357	1.340	1.321
Methyl_ethyl_ketone	16.0	9.0	5.1	0	1.240	1.409	1.368
Acetone	15.5	10.4	7.0	0	1.252	1.420	1.371
2-Nitropropane	16.2	12.1	4.1	0	1.215	1.433	1.387
Ethylene_glycol_monoethyl_ether	16.2	9.2	14.3	0	1.567	1.372	1.395
Propylene_carbonate	20.0	18.0	4.1	0	1.500	1.367	1.429
Cyclohexanol	17.4	4.1	13.5	0	1.747	1.389	1.467
Chloroform	17.8	3.1	5.7	0	1.600	1.433	1.483
Trichloroethylene	18.0	3.1	5.3	0	1.605	1.431	1.485
1,4-Dioxane	19.0	1.8	7.4	0	1.741	1.392	1.493
Ethyl_acetate	15.8	5.3	7.2	0	1.570	1.545	1.547
Ethylene_glycol_monobutyl_ether	16.0	5.1	12.3	0	1.736	1.525	1.563
Chlorobenzene	19.0	4.3	2.0	0	1.684	1.505	1.576
Nitroethane	16.0	15.5	4.5	0	1.457	1.606	1.580
Ethylene_glycol_monomethyl_ether	16.2	9.2	16.4	0	1.874	1.556	1.618
Butyl_acetate	15.8	3.7	6.3	0	1.808	1.713	1.741
1-Butanol	16.0	5.7	15.8	0	2.056	1.698	1.777
Methyl_isobutyl_ketone	15.3	6.1	4.1	0	1.760	1.776	1.782
Nitromethane	15.8	18.8	5.1	0	1.866	1.886	1.899
Toluene	18.0	1.4	2.0	0	2.139	1.883	1.978
Ethanol	15.8	8.8	19.4	0	2.431	1.962	2.077
Diethylene_glycol	16.6	12.0	20.7	0	2.496	1.967	2.101
Benzene	18.4	0.0	2.0	0	2.347	2.007	2.129
Diethyl_ether	14.5	2.9	5.1	0	2.278	2.131	2.183
Ethanol_amine	17.0	15.5	21.2	0	2.655	2.091	2.241
Carbon_tetrachloride	17.8	0.0	0.6	0	2.502	2.181	2.301
Propylene_glycol	16.8	9.4	23.3	0	2.947	2.274	2.457
Methanol	15.1	12.3	22.3	0	2.990	2.421	2.575
Hexane	14.9	0.0	0.0	0	2.922	2.632	2.745
Ethylene_glycol	17.0	11.0	26.0	0	3.397	2.614	2.837
Formamide	17.2	26.2	19.0	0	3.485	2.844	3.044

In the solubility column [a] denotes an outlier for one of the comparative results. RED1 exhibits results from this work, while RED2 and RED3 those found in the literature.[7,18] (HSP and solubility factors of solvents were extracted from the same reference.)

REDs for each solvent. This is shown for polyether sulfone, bitumen 1, bitumen 2, and lignin in Tables (II–V), respectively. When calculated solubility spheres have a reasonable variation in size or center position, it is helpful to reproduce them as 2D or 3D plots in Hansen space. From the resultant spheres compared in Table I, the most prominent changes are seen for polyether sulfone and bitumen 2, which are shown in Figures 3 and 4, respectively.

Table III. List of Solvents in the Correlation for HSP Calculation of Bitumen 1

Solvent	δ_D	δ_P	δ_H	Solubility	RED1	RED2 [8]
2-Butanol	15.8	5.7	14.5	0	2.101	2.128
2-Butyl octanol	16.1	3.6	9.3	0	1.327	1.276
Butyraldehyde	15.6	10.1	6.2	0	1.442	1.522
Caprolactone (epsilon)	19.7	15.0	7.4	0	1.870	2.096
1-Chloropentane	16.0	6.9	1.9	1[a]	1.000	1.022
Chloroform	17.8	3.1	5.7	1	0.560	0.445
Cyclohexanol	17.4	4.1	13.5	0	1.749	1.762
Cyclohexanone	17.8	6.3	5.1	1	0.479	0.540
Cyclohexylamine	17.2	3.1	6.5	1	0.770	0.670
Cyclopentanone	17.9	11.9	5.2	0	1.257	1.434
cis-Decahydronaphthalene	18.8	0.0	0.0	1	0.990	0.925
1,4-Dichlorobutane	18.3	7.7	2.8	1	0.515	0.678
Dichloromethyl methyl ether	17.1	12.9	6.5	0	1.549	2.080
1,1-Diethoxy ethanol (acetal)	15.2	5.4	5.3	0	1.214	1.177
Diethylene glycol monoethyl ether acetate	16.2	5.1	9.2	0	1.278	1.258
Diisopropylamine	14.8	1.7	3.5	0	1.402	1.301
1,2-Dimethoxybenzene	19.2	4.4	9.4	0	1.020	1.059
Ethyl acetate	15.8	5.3	7.2	0	1.159	1.126
Ethyl benzene	17.8	0.6	1.4	1	0.836	0.710
Ethyl lactate	16.0	7.6	12.5	0	1.831	1.877
2-Ethyl-hexanol	15.9	3.3	11.8	0	1.706	1.676
Ethylene glycol dibutyl ether	15.7	4.5	4.2	0[a]	1.009	0.945
Hexadecane	16.3	0.0	0.0	0	1.271	1.165
Hexyl acetate	15.8	2.9	5.9	0	1.099	1.000
Isopropyl acetate	14.9	4.5	8.2	0	1.501	1.459
Lauryl alcohol	17.2	3.8	9.3	0	1.116	1.080
Mesityl oxide	16.4	6.1	6.1	0[b]	0.914	0.906
Methyl acetate	15.5	7.2	7.6	0	1.339	1.353
Methyl benzoate	17.0	8.2	4.7	1	0.831	0.914
Methyl ethyl ketone	16.0	9.0	5.1	0	1.177	1.245
1-Methyl naphthalene	20.6	0.8	4.7	1	0.955	0.959
Methyl oleate	14.5	3.9	3.7	0	1.412	1.349
3-Methyl-2-butanol	15.6	5.2	13.4	0	1.969	1.979
Methylene dichloride	18.2	6.3	6.1	1	0.540	0.615
Nitrobenzene	20.0	8.6	4.1	1	0.786	1.000
Oleyl alcohol	14.3	2.6	8.0	0	1.699	1.630
Pyrrolidine	17.9	6.5	7.4	1	0.772	0.827
Salicylaldehyde	19.4	10.7	14.7	0	2.150	2.299
Tetrahydrofuran	16.8	5.7	8.0	1[a]	1.005	1.000
Tetrahydronaphthalene	19.6	2.0	2.9	1	0.569	0.546
1,2,3,5-Tetramethylbenzene	18.6	0.5	0.5	1	0.873	0.794
Toluene	18.0	1.4	2.0	1	0.659	0.524
2-Toluidine	19.4	5.8	9.4	0	1.046	1.127
1,1,2-Trichloroethane	18.2	5.3	6.8	1	0.595	0.620
Tricresyl phosphate	19.0	12.3	4.5	0	1.278	1.489
1,2,4-Trimethylbenzene	18.0	1.0	1.0	1	0.790	0.681
2,2,4-Trimethylpentane	14.1	0.0	0.0	0	1.828	1.746
o-Xylene	17.8	1.0	3.1	1	0.704	0.544

In the solubility column [a] denotes an outlier for one of the comparative results and [b] for both. RED1 exhibits results from this work and RED2 those found in the literature.[8] (HSP and solubility factors of solvents were extracted from the same reference.)

Table IV. List of Solvents in the Correlation for HSP Calculation of Bitumen 2

Solvent	δ_D	δ_P	δ_H	Solubility	RED1	RED2 [19]
Benzophenone	19.6	8.6	5.7	1[a]	1.000	1.012
2-Butanol	15.8	5.7	14.5	0	2.143	2.019
2-Butyl octanol	16.1	3.6	9.3	0	1.285	1.231
Butyraldehyde	15.6	10.1	6.2	0	1.511	1.541
Caprolactone (epsilon)	19.7	15.0	7.4	0	2.102	2.038
1-Chloro pentane	16.0	6.9	1.9	1[b]	1.002	1.059
Chloroform	17.8	3.1	5.7	1	0.457	0.445
Cyclohexanol	17.4	4.1	13.5	0	1.783	1.657
Cyclohexanone	17.8	6.3	5.1	1	0.527	0.642
Cyclohexylamine	17.2	3.1	6.5	1	0.677	0.663
Cyclopentanone	17.9	11.9	5.2	0	1.425	1.458
cis-Decahydronaphthalene	18.8	0.0	0.0	1	0.943	0.723
1,4-Dichlorobutane	18.3	7.7	2.8	1	0.661	0.758
1,1-Diethoxy ethanol (acetal)	15.2	5.4	5.3	0	1.168	1.184
Ethylene glycol monoethyl ether acetate	16.2	5.1	9.2	0	1.263	1.241
Diisopropylamine	14.8	1.7	3.5	0	1.300	1.224
1,2-Dimethoxybenzene	19.2	4.4	9.4	0	1.081	0.996
Ethyl acetate	15.8	5.3	7.2	0	1.123	1.135
Ethyl benzene	17.8	0.6	1.4	1	0.722	0.557
Ethyl lactate	16.0	7.6	12.5	0	1.886	1.817
2-Ethyl-hexanol	15.9	3.3	11.8	0	1.690	1.586
Ethylene glycol dibutyl ether	15.7	4.5	4.2	0[b]	0.934	0.959
Hexadecane	16.3	0.0	0.0	0	1.169	1.025
Hexyl acetate	15.8	2.9	5.9	1	1.000	0.974
Isopropyl acetate	14.9	4.5	8.2	0	1.460	1.420
Laurylalcohol	17.2	3.8	9.3	0	1.092	1.044
Mesityl oxide	16.4	6.1	6.1	0[b]	0.896	0.956
Methyl acetate	15.5	7.2	7.6	0	1.348	1.363
Methyl benzoate	17.0	8.2	4.7	1	0.898	0.991
Methyl ethyl ketone	16.0	9.0	5.1	0	1.231	1.289
1-Methyl naphthalene	20.6	0.8	4.7	1	0.995	0.753
Methyl oleate	14.5	3.9	3.7	0	1.342	1.310
3-Methyl-2-butanol	15.6	5.2	13.4	0	1.993	1.884
Methylene dichloride	18.2	6.3	6.1	1	0.612	0.689
Nitrobenzene	20.0	8.6	4.1	0	1.003	1.000
Oleyl alcohol	14.3	2.6	8.0	0	1.636	1.549
Pyrrolidine	17.9	6.5	7.4	1	0.828	0.872
Salicylaldehyde	19.4	10.7	14.7	0	2.319	2.185
Tetrahydrofuran	16.8	5.7	8.0	1[a]	1.000	1.021
Tetrahydronaphthalene	19.6	2.0	2.9	1	0.574	0.364
1,2,3,5-Tetramethylbenzene	18.6	0.5	0.5	1	0.810	0.608
Toluene	18.0	1.4	2.0	1	0.536	0.388
2-Toluidine	19.4	5.8	9.4	0	1.146	1.081
1,1,2-Trichloroethane	18.2	5.3	6.8	1	0.625	0.664
Tricresyl phosphate	19.0	12.3	4.5	0	1.485	1.492
1,2,4-Trimethylbenzene	18.0	1.0	1.0	1	0.691	0.531
2,2,4-Trimethylpentane	14.1	0.0	0.0	0	1.748	1.600
o-Xylene	17.8	1.0	3.1	1	0.560	0.409

In the solubility column [a] denotes an outlier for one of the comparative results and [b] for both. RED1 exhibits results from this work and RED2 those found in the literature.[19] (HSP and solubility factors of solvents were extracted from the same reference.)

Table V. List of Solvents in the Correlation for HSP Calculation of Lignin

Solvent	δ_D	δ_P	δ_H	Solubility	RED1	RED2 [19]
Acetic acid	14.5	8.0	13.5	0	1.194	1.195
Acetic anhydride	16.0	11.7	10.2	0	1.003	1.006
Acetone	15.5	10.4	7.0	0	1.215	1.212
Acetonitrile	15.3	18.0	6.1	0	1.280	1.277
Acetophenone	19.6	8.6	3.7	0	1.113	1.096
Aniline	19.4	5.1	10.2	1	0.908	0.897
Benzaldehyde	19.4	7.4	5.3	0	1.058	1.044
Benzene	18.4	0.0	2.0	0	1.608	1.582
1-Bromonaphthalene	20.3	3.1	4.1	0	1.278	1.254
1,3-Butanediol	16.6	10.0	21.5	1	0.888	0.895
1-Butanol	16.0	5.7	15.8	0	1.061	1.060
Butyl acetate	15.8	3.7	6.3	0	1.416	1.403
Butyl lactate	15.8	6.5	10.2	0	1.161	1.158
Butyric acid	14.9	4.1	10.6	0	1.345	1.337
gamma-Butyrolactone	19.0	16.6	7.4	1	0.835	0.833
Butyronitrile	15.3	12.4	5.1	0	1.304	1.298
Carbon disulfide	20.5	0.0	0.6	0	1.618	1.586
Carbon tetrachloride	17.8	0.0	0.6	0	1.710	1.683
Chlorobenzene	19.0	4.3	2.0	0	1.391	1.369
1-Chlorobutane	16.2	5.5	2.0	0	1.523	1.506
Chloroform	17.8	3.1	5.7	0	1.309	1.293
m-Cresol	18.0	5.1	12.9	1	0.922	0.917
Cyclohexane	16.8	0.0	0.2	0	1.787	1.761
Cyclohexanol	17.4	4.1	13.5	0	1.019	1.013
Cyclohexanone	17.8	6.3	5.1	0	1.206	1.193
Cyclohexylchloride	17.3	5.5	2.0	0	1.442	1.424
Diacetone alcohol	15.8	8.2	10.8	0	1.085	1.085
o-Dichlorobenzene	19.2	6.3	3.3	0	1.229	1.210
2,2-Dichlorodiethyl ether	18.8	9.0	5.7	0	1.016	1.006
Diethylamine	14.9	2.3	6.1	0	1.567	1.552
Diethylene glycol	16.6	12.0	20.7	1	0.826	0.836
Diethylene glycol monobutyl ether	16.0	7.0	10.6	0	1.108	1.105
Diethylene glycol monomethyl ether	16.2	7.8	12.6	1[a]	1.000	1.001
Diethyl ether	14.5	2.9	5.1	0	1.621	1.605
Diethyl sulfide	16.8	3.1	2.0	0	1.563	1.543
Di(isobutyl) ketone	16.0	3.7	4.1	0	1.496	1.480
Dimethylformamide	17.4	13.7	11.3	1	0.766	0.774
Dimethyl sulfoxide	18.4	16.4	10.2	1	0.721	0.727
1,4-Dioxane	19.0	1.8	7.4	0	1.229	1.211
Dipropylamine	15.3	1.4	4.1	0	1.650	1.631
Dipropylene glycol	16.5	10.6	17.7	1	0.821	0.831
Ethanol	15.8	8.8	19.4	1	0.983	0.988
Ethanolamine	17.0	15.5	21.2	1	0.775	0.788
Ethyl acetate	15.8	5.3	7.2	0	1.316	1.306
Ethylbenzene	17.8	0.6	1.4	0	1.640	1.615
2-Ethyl-1-butanol	15.8	4.3	13.5	0	1.173	1.169
Ethylene glycol	17.0	11.0	26.0	1[b]	1.001	1.002

TABLE V. *Continued*

Solvent	δ_D	δ_P	δ_H	Solubility	RED1	RED2 [19]
Ethylene glycol monobutyl ether	16.0	5.1	12.3	0	1.138	1.134
Ethylene glycol monoethyl ether	16.2	9.2	14.3	1	0.920	0.925
Ethylene glycol monoethyl ether acetate	15.9	4.7	10.6	0	1.210	1.204
Ethylene glycol monomethyl ether	16.2	9.2	16.4	1	0.900	0.906
Furan	17.8	1.8	5.3	0	1.390	1.372
Glycerol	17.4	12.1	29.3	0	1.132	1.128
Hexane	14.9	0.0	0.0	0	1.929	1.904
Isoamyl acetate	15.3	3.1	7.0	0	1.460	1.447
Isobutyl isobutyrate	15.1	2.9	5.9	0	1.530	1.516
Isooctyl alcohol	14.4	7.3	12.9	0	1.238	1.237
Isophorone	16.6	8.2	7.4	0	1.130	1.125
Mesityl oxide	16.4	6.1	6.1	0	1.278	1.268
Methanol	15.1	12.3	22.3	0	1.070	1.076
Methylal	15.0	1.8	8.6	0	1.492	1.479
Methyl ethyl ketone	16.0	9.0	5.1	0	1.282	1.274
Methyl isoamyl ketone	16.0	5.7	4.1	0	1.424	1.411
Methyl isobutyl carbinol	15.4	3.3	12.3	0	1.286	1.279
Methyl isobutyl ketone	15.3	6.1	4.1	0	1.476	1.464
Morpholine	18.8	4.9	9.2	1	0.997	0.986
Nitrobenzene	20.0	8.6	4.1	0	1.071	1.054
Nitroethane	16.0	15.5	4.5	0	1.259	1.254
Nitromethane	15.8	18.8	5.1	0	1.290	1.286
2-Nitropropane	16.2	12.1	4.1	0	1.267	1.260
1-Pentanol	15.9	4.5	13.9	0	1.147	1.143
1-Propanol	16.0	6.8	17.4	0	1.012	1.013
Propylene carbonate	20.0	18.0	4.1	0	1.027	1.015
Propylene glycol	16.8	9.4	23.3	0[b]	0.940	0.944
Pyridine	19.0	8.8	5.9	1	0.997	0.987
Styrene	18.6	1.0	4.1	0	1.444	1.421
Tetrahydrofuran	16.8	5.7	8.0	0	1.171	1.163
Tetrahydronaphthalene	19.6	2.0	2.9	0	1.417	1.392
Toluene	18.0	1.4	2.0	0	1.562	1.538
1,1,1-Trichloroethane	16.8	4.3	2.0	0	1.518	1.500
Trichloroethylene	18.0	3.1	5.3	0	1.316	1.298
Xylene	17.6	1.0	3.1	0	1.546	1.524

In the solubility column [a] denotes an outlier for one of the comparative results and [b] for both. RED1 exhibits results from this work and RED2 those found in the literature.[20] (HSP and solubility factors of solvents were extracted from the same reference.)

The results obtained in this work clearly show that the proposed optimization algorithm is not only valid to calculate polymer solubility spheres, but is also able to refine HSP values calculated by other searching methods applied in the past. From Table I it is observed that in all cases, at least one of the quality-of-fit parameters is improved.

The best refinement is found for polyether sulfone, where a substantial change in HSPs is validated by the improvement of most quality-of-fit parameters. Compared to the fitting achieved by Hansen's SPHERE program,[18] DATAFIT and specificity (or sensibility) are increased to the maximum (1 and 100%, respec-

tively), whereas outliers fraction is reduced to zero. If the sphere size (R_0) is considered as another quality-of-fit parameter, it is possible to conclude that the fitting of the solubility sphere is much better compared also to Gharagheize's result in MATLAB (Downhill-Simplex algorithm),[7] reducing the sphere radius from 6.72 to 5.42 $MPa^{1/2}$. By fitting this improved sphere, considerable differences are also found in HSP values (sphere center), especially for the hydrogen component, which decreased from 9.57 to 7.95 $MPa^{1/2}$.

Regarding obtained results for the two bitumen correlations, almost all quality-of-fit parameters are improved, especially

bitumen 2, for which all are improved, including sphere size. The most appreciable change in HSPs is found for polar component, which increased from 3.90 (3.0) to 4.79 (3.99) MPa$^{1/2}$.

Comparative results for lignin, on the other hand, show minimum improvement in quality-of-fit parameters, with almost no change in HSPs. This is because the large number of solvents in the correlation (82, compared with 41 and 48) fulfills Hansen space and defines more precisely an optimum sphere. In other words, there are more boundary solvents limiting possible arrangements. In a numerical optimization perspective, it could be said that there are less local optimum configurations close to the global optimum.

Comparing this last result for lignin with the others, a shortcoming of this correlation method can be pointed out. As already mentioned by Hansen,[1] the approach of locating the polymer HSP as the centre of a sphere has a problem in that it is in reality, the poor solvents or nonsolvents located near the boundary of the sphere that fix the boundary (and centre) rather than the best solvents in the middle. This may present problems for small sets of solvents, or even large sets when they are badly distributed in Hansen's space.

On the basis of these problems, partially solved by testing a large set of solvents, a new parameter is proposed to validate, or at least, estimate the data quality in these correlations. This parameter, called "fill factor" (Φ), should be able to determine if any data set is large and spread in space enough to give a reliable solubility sphere. If we divide the filling volume (a minimum volume in Hansen's space that emcompasses all solvents in the correlation) by N equal volume units, Φ would give a measure of how Hansen's space is filled. It could simply have the following function form:

$$\Phi = n_f / N \qquad (16)$$

where n_f is the number of volume units filled by any solvent. If one calculates Φ for many good data sets, the average value should be a reliable estimate to validate any data set.

It is worth mentioning another aspect related to the quality of correlations and fittings. In this work two extra quality-of-fit parameters, sensibility and specificity, are proposed in order to measure the quality of fittings along with the two common parameters, DATAFIT and outliers. These are justified first, by noticing some cases in which DATAFIT and outliers alone don't characterize enough the features in the resulting sphere fitting. Second, because there is no standard in the literature to evaluate the quality of fittings, thus several published polymer HSPs are poor when qualifying data sets and fittings and cannot be safely used and referred to.[6,21–23] Last but no less an important fact is that some authors are not clear when considering boundary solvents (with partial solubility) in the correlation.[22,23] This is another shortcoming of the correlation method, which must be dealt with by assigning to them one of the two possible characters: good (1) or bad (0) solvents in the algorithm.

CONCLUSIONS

By using an evolutionary strategy as an alternative optimization procedure, it has been shown that it is possible to improve the fitting of solubility spheres while applying Hansen's correlation method. In other words, it can better characterize materials, thus providing refined knowledge for processing mixed compounds and reducing costs and time in laboratory tests. This is clearly seen for all selected samples. Compared with previous results, most quality-of-fit parameters are optimized, including two new suggested ones, sensibility and specificity, which are shown to be more sensitive to any fitting improvement.

The genetic algorithm optimization approach, therefore, combined to the correlation method, has proven to be an effective way to find global solutions for HSPs of polymers and other solutes. Although for convenience it has been performed in visual programming environment (LabVIEW platform), it is possible to be implemented with virtually any programming language and in any environment. The only shortcoming is that, due to its complexity, ready packages are usually no available to date.

It is expected that the addressed disadvantages of the correlation method, as well as the suggested fill factor parameter, motivate future researchers to increasingly develop improvements and establish standards in this field.

REFERENCES

1. Hansen, C. M. Hansen Solubility Parameters: A User's Handbook. CRC Press, Inc., Boca Raton FL, **2007**; Chapter 1, p 1.

2. Hildebrand, J. H. *Chem. Rev.* **1949**, *44*, 37.

3. Hansen, C. M. *Ind. Eng. Chem. Prod. Res. Dev.* **1969**, *8*, 2.

4. Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; Alkorta, I.; Clary, D. C.; Crabtree, R. H.; Dannenberg, J. J.; Hobza, P.; Kjaergaard, H. G.; Legon, A. C.; Mennucci, B.; Nesbitt, D. *J. Pure Appl. Chem.* **2011**, *83*, 1619.

5. Wisniewski, R.; Smieszek, E.; Kaminska, E. *Prog. Org. Coat.* **1995**, *26*, 265.

6. Segarceanu, O.; Leca, M. *Prog. Org. Coat.* **1997**, *31*, 307.

7. Gharagheizi, F. *J. Appl. Polym. Sci.* **2007**, *103*, 31.

8. Redelius, P. *Energ. Fuel.* **2004**, *18*, 1087.

9. Nelder, J. A.; Mead, R. *Comput. J.* **1965**, *7*, 308.

10. Fletcher, R. The Sequential Quadratic Programming Method. http://php.math.unifi.it/users/cimeCourses/2007/02/200723-Notes.pdf (accessed October 9, **2012**).

11. Cartwright, H. M. In An Introduction to Evolutionary Computation and Evolutionary Algorithms; Johnston, R. L.,

Ed.; Applications of Evolutionary Computation in Chemistry Series 110; Springer: Berlin, Heidelberg, **2004**; p 1–32.

12. Lian, Y. S.; Oyama, A.; Liou, M. S. *Prog. Aerosp. Sci.* **2010**, *46*, 199.

13. Freitas, A. A. In Soft Computing for Knowledge Discovery and Data Mining; Maimon, O; Rokach, L., Eds.; Springer: New York, **2008**; Part II, Chapter 3, p 79.

14. Braysy, O.; Dullaert, W.; Gendreau, M. *J. Heuristics.* **2004**, *10*, 587.

15. Moles, C. G.; Lieber, A. S.; Banga, J. R.; Keller, K. In Advances in Soft Computing: Engineering Design and Manufacturing; Benitez, J. M.; Cordon, O.; Hoffmann, F.; Roy, R., Eds.; Springer, **2003**; p 331.

16. Zeidler, D.; Frey, S.; Kompa, K.-L.; Motzkus, M. *Phys. Rev. A.* **2001**, *64*, 023420-1–023420-13.

17. Eiben, A. E.; Hinterding, R.; Michalewicz, Z. *IEEE T. Evolut. Comput.* **1999**, *3*, 124.

18. Hansen, C. M. Hansen Solubility Parameters: A User's Handbook. CRC Press, Inc., Boca Raton FL, **2007**; Chapter 5, p 95.

19. Hansen, C. M. Hansen Solubility Parameters: A User's Handbook. CRC Press, Boca Raton FL, **2007**, Chapter 9, p 151.

20. Hansen, C. M. Hansen Solubility Parameters: A User's Handbook. CRC Press, Boca Raton FL, **2007**, Chapter 15, p 269.

21. Archer, W. L. *Drug Dev. Ind. Pharm.* **1992**, *18*, 599.

22. Salaün, F.; Devaux, E.; Bourbigot, S.; Rumeau, P. *Carbohyd. Polym.* **2010**, *79*, 964.

23. Bordes, C.; Fréville, V.; Ruffin, E.; Marote, P.; Gauvrit, J. Y.; Briançon, S.; Lantéri, P. *Int. J. Pharm.* **2010**, *383*, 236.