Machine Learning Approaches for Prediction of Phase Equilibria in Poly (Ethylene Glycol) + Sodium Phosphate Aqueous Two-Phase Systems

Pirdashti, Mohsen*[+]***; Taheri, Mojtaba***
Chemical Engineering Department, Faculty of Engineering, Shomal University, PO Box 731 Amol, I.R. IRAN

Dragoi, Niculina; Curteanu, Silvia
Faculty of Chemical Engineering and Environmental Protection "Cristofor Simionescu", "Gh. Asachi" Technical University, Bld Mangeron 73, 700050, Iasi, ROMANIA

ABSTRACT: *In this research, liquid-liquid equilibrium (LLE) data were experimentally obtained for the ternary systems of (water + carboxylic acid + dipropyl ether) at T = 298.2 K and P = 101.3 kPa. The carboxylic acids used in this study were isobutyric acid, valeric acid and isovaleric acid. All these systems are according to Treybal classification, Type-2 systems because the two binary subsystems are partially miscible. The lowest distribution coefficients and separation factors were calculated for isobutyric acid (40 and 329, respectively). The authenticity of the experimental equilibrium data was identified from Hand and Othmer-Tobias correlations. The experimental tie line data were correlated by using the nonrandom two-liquid (NRTL) and universal quasi-chemical (UNIQUAC) activity coefficient models. RMSD values are between 0.0112 and 0.0155 for NRTL model; and are between 0.0083 and 0.0153 for UNIQUAC model.*

KEYWORDS: *Liquid-liquid equilibrium; Carboxylic acid; Dipropyl ether; NRTL; UNIQUAC.*

INTRODUCTION

Aqueous Two-Phase System (ATPS) is a liquid-liquid extraction technique that has gained interest due to its great potential for the extraction, separation, purification and enrichment of various biological materials such as proteins, membranes, viruses, recombinant proteins and enzymes, nucleic acids [1-7] and separation of metal ions, such as Chromium [8], Gold (III) [9, 10], Cobalt (II), Iron (III), Nickel (II) [11]], Hg(II) [12], Cd (II) [12, 13], Zn(II) [14, 15], Copper [14-16], Gallium [17], Tungsten (VI) [18], Molybdenum (VI) [19], Zirconium and Hafnium [20]. ATPSs have many advantages such as a high yield [21], a relatively high load capacity [22], scaling up feasibility [23], low material cost [24], low energy consumption [25], low interfacial tension [26], short processing time [27], good resolution [28], ease of continuous process [29], biologic compatibility [30], selective extraction [31].

The phase diagram is one of the most important elements for designing, optimizing and scale up of ATPSs.

** To whom correspondence should be addressed.*
+ E-mail: pirdashti@yahoo.com
1021-9986/2020/6/185-197 13/$/6.03

It is a "fingerprint" unique to ATPS under set conditions of, for example, polymer molecular weight and temperature [32]. Measurement of the phase equilibrium data is very sensitive to the experimental equipments and conditions, which makes it time consuming, costly and difficult to measure. For this reason, providing a mathematical model that can accurately predict phase equilibrium data can lead to a reduction of the number of experiments and thus it contributes to saving time, materials and energy. This predicted data can be of great help in designing the separation process based on ATPS.

Some researchers applied Artificial Neural Networks (ANNs) for prediction of the phase equilibria of PEG /salt ATPSs [33,34] and ionic liquid/salt ATPSs[35, 36]. The present work is devoted to obtaining the phase equilibria of five poly (ethylene glycol) (PEG) - sodium dihydrogen phosphate, ATPSs at different temperatures (298.15, 308.15 and 318.15 K). Also, phase equilibrium data of PEG1000 (and 6000) + NaH_2PO_4 + water [37] are both for reliability of the data and to compare the accuracy of the data, has also been used. After that, the system was modeled using a combination of artificial intelligence approaches, the scope being to determine the optimal model that can efficiently capture the dynamic of the system and can provide accurate predictions for different combinations of parameters, thus eliminating the need to manually select an acceptable thermodynamic model. The present approach consists in: i) a neuro-evolutive technique combining ANNs with Differential Evolution (DE) algorithm; ii) Support Vector Machines (SVM); and iii) SVM optimized with DE. A comparison between the three approaches was performed, and the variant generating the best predictions was considered.

In this work, the ANNs and SVMs were chosen as alternatives to the standard thermodynamic models because the main advantage of efficiently modeling complex input-output relations without the need of including the chemical and physical laws that govern the system. In addition, the two procedures, ANN and SVM, avoid a series of difficulties associated with thermodynamic models such as: i) are inaccurate (as the majority are based on the ideal gas equation of state); ii) multiple models exists with different efficiencies and complexities; and iii) choosing the one that provides a desired level of accuracy is, in fact, a trial – and - error approach, which does not guarantee that an optimum is reached. Many predictions can be made simultaneously with ANNs or SVMs.

SVM is based on the concept of structural risk minimization [38]. Origininally, SVMs were developed as classfiers (as they use hyperplanes to separate data into clasess) but they were futher extended to also perform regression [39]. Consequently they were applied to solve various problems including chemical engineering systems such as: CO_2 equilibrium absorption in Piperazine aqueous solutions [40], aerobic phase length for partial nitrification processes in lab-scale sequencing batch reactor [41], polypropylene fibres reinforced self-compacting composites incorporating nano-CuO [42], saturation pressure of crude oils [43], clathrate hydrates of methane, carbon dioxide, nitrogen, and hydrogen + water soluble organic promoters [44]. The main problem encountered in SVMs is the influence of the parameters on its performance and the difficulty in finding the optimal values [45] . Consequently, in this work, the idea of neuro-evolution (evolving ANN parameters to find good models) is adapted to SVMs, DE algorithm being applied to evolve the SVM parameters.

The novelty of the current study consists in the experimental analysis of the phase equilibrium behavior of five PEG - sodium di hydrogen phosphate ATPSs at different temperatures (298.15 and 308.15 K). Consequently, the effects of temperature and molecular weight on the phase behavior of the studied systems were investigated and the system was modeled using kto extend the understanding of the inner relations governing the process and to generate predictions in specific cases, thus reducing the need for high cost laboratory experiments and setting the base for reaction optimization.

EXPERIMENTAL SECTION

Materials

To prepare the materials, PEG with averge of 1500, 2000, 3000, 4000, 8000 g mol^{-1} were obtained from Merck (Darmstadt, Germany) and used without further purification, sodium dihydrogen phosphate (NaH_2PO_4 , mass purity > 0.99%). Distilled deionized water (conductivity = 17.5 μs · cm^{-1}) has also been used. All materials had an analytical grade. All chemicals were dried for at least 24 h (433.15 K for salts and 313.15 K for polymers) to eliminate the adsorbed water.

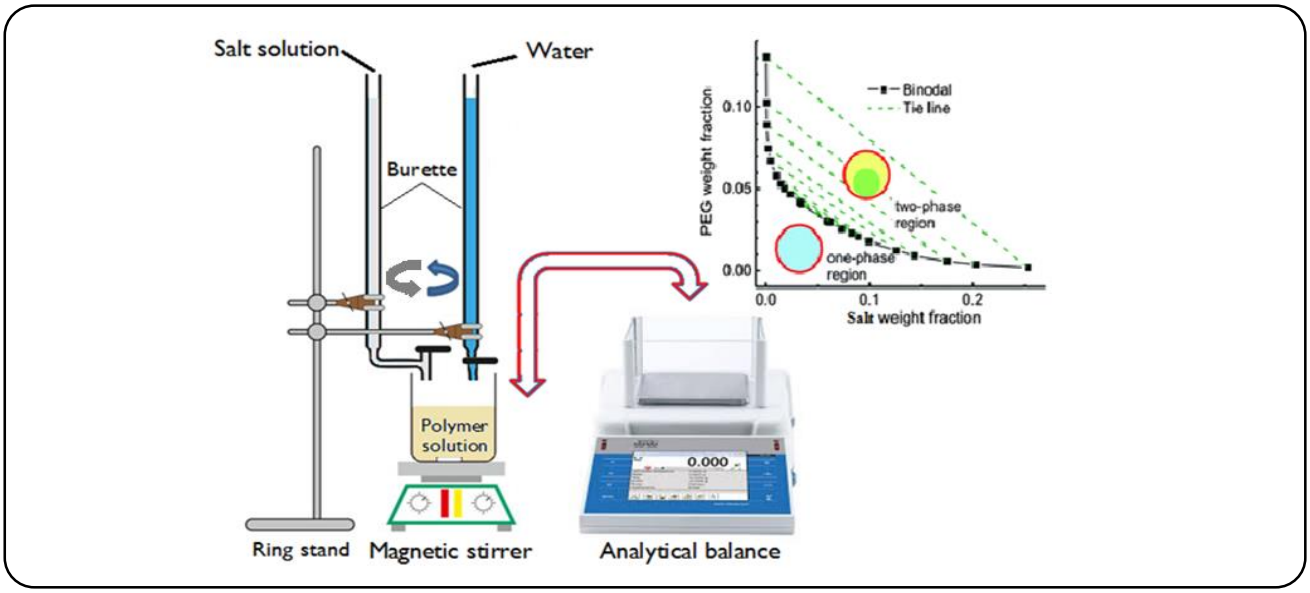

Fig. 1: Simple graphical representation of cloud point method used to determine the bimodal.

Apparatus and Procedure

Binodal Curve

The cloud-point titration method [32] was used to obtain the binodal curve. All the devices used in these experiments were essentially the same as in our previous works [46-48]. 5 g of a stock solution of PEG into a 25 cm^3volume jacketed glass vessel was weighted while its temperature within $\pm$0.05 K has been fixed by water circulation through an external jacket. Then the vessel was weighted and a stock solution of salt was added, drop-wise, until the first sign of turbidity was observed, i.e., the cloud point. By weighting the amounts of salt and polymer, the first point on the binodal is obtained. After that, a known weight of water was added below the cloud point and the step of adding salt stock solution was repeated. This method is illustrated graphically in Fig. 1.

Modelling and optimization

In this work, in order to model the considered systems, three distinct approaches were applied. They consists in: i) classical SVM; ii) SVM optimized with DE; and iii) ANN optimized with DE. As the classical SVM does not include a procedure to automatically set-up its control parameters and these control parameters influence its performance, DE was considered as optimizer for SVM (second approach). In order to determine if SVM (simple, or in combination with DE) is a good approach for the considered process, a combination seldom used was also employed (the third procedure) as a mean to compare the predictions.

The DE version used for the two models (cases two and three) previously mentioned is the same, the only difference consisting in the length of the individuals that contain the encoded models. Consequently, the observed difference in performance of estimations is only due to the model characteristics and specificity.

Support Vector Machine

Having a set of n examples in the form (X_i, Y_i) where X_i represents the input and Y_i the output (which in case of classification is a bipolar label) and using a non-linear maping $\theta()$, X_i is mapped into a high dimensional feature space F in which the optimum decision function $f(X_i)=\omega\theta(X_i)+b$ is constructed, where $\omega=(\omega_1, \omega_2, ..., \omega_n)$ is the vector of wheights in F. Using the structural risk minimization principle, the optimization problem becomes:

$$Minimize\ R = \frac{1}{2}\|\omega\|^2 + CRe \quad (1)$$

where R is the error term and C is the regularization parameter.

In the standard SVM regression, the optimization objective is formulated as:

$$min\, J(\omega,\xi) = \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \quad (2)$$
$$Y_i - \omega\theta(X_i) - b \le \varepsilon + \xi_i$$
$$\omega\theta(X_i) + b - Y \le \varepsilon_i + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0,\ i=1,\ldots, l$$

Where ξ_i, ξ_i^* are slack variables and ε is the accuracy.

The solution of the optimization is given by the saddle point of the Lagrangian and based on the Mercer condition; the kernels are defined as $K(X_i, X_i) = \theta(X_i)\theta(X_j)$ [48]. There are several types of kernel functions, the most used being represented by: linear, polynomial and radial basis function (RBF).

The parameters of the SVM which control its performance include *C*, the kernel type and its parameters. As there is no systematic manner to determine them, in this work, two approaches were performed. First, the SVM is run with the classical settings specific to the SVM library. Second, all the possible parameters are encoded into individuals which are evolved by an Evolutionary Algorithm (EA), called Differential Evolution (DE). DE is based on the Darwinian principle of evolution and performed a global serch by employing mutation crossover and selection for a number of iterations until a stop criterion is reached.

For both SVM and ANN parameters, the same DE variant was used. It was first proposed in [49] and represented a hybridization of DE in combination with ANNs (hSADE-NN approach). Although initially applied on the depollution of some volatile compounds, due to the flexibility of the approach, it can be efficiently used to any problem that can be sufficiently described by a set of experimental data [50, 51]. In the hybridized DE version: i) initialization uses the principle of opposition based learning; ii) mutation uses 2 differential terms and it is based on the best so far individual; iii) a local search procedure, combining Backpropogation (BK) and Local Search (LS) algorithms, is applied to the best solution. More details about the motivation and inner workings of the hSADE-NN approach can be found in [49].

Since in this work DE is combined with both SVM (denoted SVM-DE) and ANN (denoted ANN-DE) and since the application of BK is only possible for ANNs, the local improvement procedure is modified to include only LS. In this manner, a comparison between the DE-SVM and DE-ANN combinations will show the efficiency of DE in providing good parameters.

Artificial Neural Networks

The ANNs have the capability of modelling any type of relation. However, their optimal parameters are difficult to set and, consequently, in this work, the training and identification of the best architecture is performed with an evolutionary algorithm named DE. The combination of ANNs with EAs is called neuro-evolution and the literature shows that it can efficiently be used for various problems, the range of applicability being large. Examples include: Cu(II) removal [51], PVC/clay nanocomposite foams [53], fault diagnosis [54]. The major disadvantage of neuro-evolution is its computational cost wich is closely related to the effectiveness of the EA used [55]. Consequently, this work uses an improved variant of the DE algorithm, the same applied to evolve the SVM parameters. In order to determine an optimal ANNs for the process studied, the individuals are structures containing the following ANNs parameters: number of hidden layers, number of neurons in each hidden layer, wheights, biases and activation functions (type and parameters). As it can be observed, DE simulatenously performs the training and the topology determination. In the majority of cases, the activation function type is set per layer, but in this case, each neuron can have one from the following list of activation functions: Linear, Hard Limit, Bipolar Sigmoid, Logarithmic Sigmoid, Tangent Sigmoid, Sinus, Radial Basis with a fixed centre at 0, and Triangular Basis functions. The main difference between the DE-SVM and DE-ANN combination is represented by the number, type and meaning of parameters included in the evolutionary process.

A general schema of the modelling procedure is provided in Fig. 2, where Model can be represented by either SVM or ANN. In case of SVM, the parameters considered include: the SVM type (μ-SVR, ε-SVR), kernel type (linear, polynomial, RBF and sigmoid), degree (applicable only for polynomial kernel type), γ (a coefficient of polynomial and sigmoid kernels) and *C*, the cost parameter. DE encodes all these parameters into a vector containing real values and when computing the fitness in order to determine the performance indicators, only the necessary parameters are extracted to construct the SVM. In this work, the SVM module is represented by the LIBSVM library [52], which was integrated into the DE procedure.

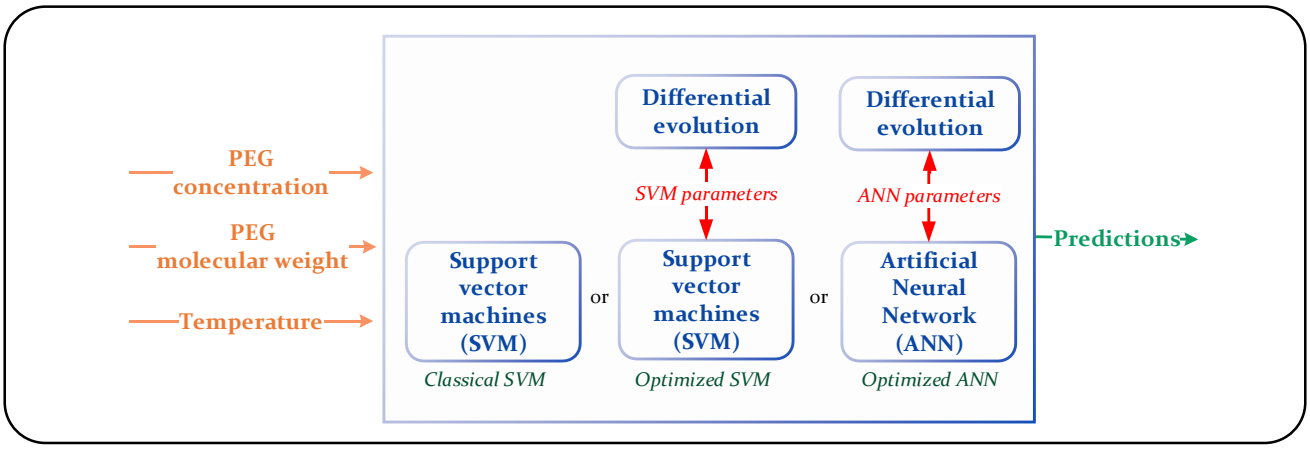

Fig. 2: General schema of the modelling procedure.

The weight fraction of polymer was modelled as function of weight fraction of salt, PEG molecular weight and temperature. The choice of these variables is based on their importance in polymer-salt ATPS design.

RESULT AND DISCUTION

Binodal Curve

The binodal experimental data plotted in Fig. 3 are shown in Table 1.

Effect of Molecular Weight on the Binodal Curve

According to Table 1 and Figs. 3 and 4, as the PEG molecular weight is increased, the binodal curves shift to lower PEG concentration. For the studied systems, the effect of polymer molecular weight on increasing the two-phase region follows the order: 8000 > 4000 > 3000 > 2000 > 1500. For complete comparison and investigation of the effect of polymer molecular weight, the binodal curve of the PEG1000(and 6000) + NaH_2PO_4 + water [37] are also shown in Fig. 3. This is because the solubility of PEG in water decreases with increasing PEG molecular weight. In fact, at high PEG molecular weights, the PEG-rich phase saturates at relatively low PEG concentrations, because the size of PEG molecules is quite large. This trend is in agreement with other experimental results [37, 46, 47,53-55].

Effect of Temperature on the Binodal Curve

As shown in Table 1 and Fig. 5, as the temperature rises, an increase of the PEG concentration in the top phase and a decrease of the salt concentration in the bottom phase were observed. The increase of the two-phase region with increasing temperature has also been observed in previous works [37, 46, 47, 53-56]. When the temperature increases, the attraction between the PEG molecules increases and, therefore, with increasing temperature, the attraction between the PEG and the water molecules will decrease in the PEG-rich region.

Modeling Results

In order to model the process, three approaches were considered: i) SVM, ii) SVM-DE; and iii) ANN-DE. These combinations were implemented in Visual Studio. NET Framework 4.0 (C# environment) and the simulations run on a desktop computer with Pentium 4 Dual Core and 250 GB SSD. From the dataset including 266 exemplars, 200 were randomly chosen to be used in the training phase and 66 in the testing phase. This splitting and random assignation to the two classes is valid for all the three approaches.

Due to the intrinsic nature of the DE algorithm, the variants that use it to determine the optimal values for the parameters (ANN-DE and SVM-DE) are repeated 50 times. In all three cases, the results obtained are presented in Table 2, where Fitness is invers proportional with mean squared error (MSE) testing and represents the measure of performance used internally to guide the DE evolution and to identify the optimal control parameters. R^2 represents the coefficient of determination. The DE parameters used were: number of generations = 200, individuals in the population = 50. The *F* and *Cr* parameters of the DE algorithm were internally set through the process of self-adaptability, their limits being between 0 and 1.

As it can be observed, the best model is the one based on SVM and DE because it has the highest fitness and

Table 1: Binodal curve data of the poly (ethylene glycols) (1500, 2000, 3000, 4000 and 8000)+ sodium di hydrogen phosphate + water system at (298.15 and 308.15) K and 0.1 MPa.

PEG1000+NaH_2PO_4 [37]				PEG1500+ NaH_2PO_4			
298.15K		308.15K		298.15K		308.15K	
w_s	w_p	w_s	w_p	w_s	w_p	w_s	w_p
30.86	0.39	26.63	1.15	6.59	31.03	6.57	31.23
28.87	0.70	25.29	1.87	6.83	30.26	6.89	30.46
25.95	1.61	23.69	2.88	7.01	29.44	7.11	29.16
24.89	2.59	22.32	4.05	14.29	13.45	14.31	13.56
23.20	3.94	22.07	4.30	14.59	12.69	14.59	12.56
21.57	5.70	21.70	4.71	14.78	12.27	14.55	12.27
19.86	7.98	21.14	5.36	17.21	8.61	17.14	8.61
18.00	10.75	20.31	6.40	17.54	8.02	17.30	7.89
16.29	13.24	19.73	7.21	17.88	7.83	17.45	7.77
13.70	17.45	19.14	8.06	18.67	6.32	18.29	6.13
12.88	18.79	18.28	9.36	20.02	4.44	19.71	4.14
12.19	19.19	17.34	10.91	20.88	3.74	20.28	3.55
10.52	23.03	16.41	12.5	21.01	3.44	20.77	3.12
8.64	26.86	14.63	15.57			21.21	2.52
7.22	30.72	13.11	18.41				
5.96	33.88	12.73	19.76				
4.86	37.27	11.18	22.06				
4.05	39.94	9.98	24.53				
3.29	42.65	8.76	27.26				
2.63	45.00	7.22	31.12				
2.08	47.73	5.99	34.65				
		4.75	38.86				
		3.45	43.31				
		2.58	46.76				
		2.02	49.61				
PEG2000+ NaH_2PO_4				PEG3000+ NaH_2PO_4			
298.15K		308.15K		298.15K		308.15K	
w_s	w_p	w_s	w_p	w_s	w_p	w_s	w_p
6.29	28.55	6.01	28.17	6.25	27.99	5.63	28.87
6.72	27.12	6.44	27.29	6.39	27.43	5.82	28.37
6.99	26.56	6.56	27.01	6.66	26.55	5.94	28.03

Table 1: Binodal curve data of the poly (ethylene glycols) (1500, 2000, 3000, 4000 and 8000)+ sodium di hydrogen phosphate + water system at (298.15 and 308.15) K and 0.1 MPa.

14.24	12.41	13.12	13.26	12.84	13.21	8.90	20.12
14.54	11.63	13.21	13.06	13.00	13.00	9.45	18.76
14.72	11.29	13.31	12.83	13.06	12.82	9.87	18.03
17.11	7.51	15.95	8.42	15.50	8.45	10.21	17.06
17.45	7.02	16.08	8.26	15.55	8.38	11.76	14.15
17.56	6.81	16.14	8.14	15.59	8.33	11.95	13.77
18.80	5.14	16.57	7.64	16.94	6.17	13.01	11.73
18.93	4.96	17.43	6.24	16.89	6.10	13.28	11.06
18.98	4.86	17.93	5.53	16.92	6.08	15.80	7.10
19.02	4.81	18.02	5.46	16.91	6.05	16.33	6.21
		18.61	4.45	17.81	4.80	16.96	5.12
				17.82	4.73	18.04	3.66
				17.84	4.69		
				18.45	3.89		
				18.54	3.77		
				18.55	3.69		
				19.09	3.15		
				19.18	3.05		
				19.29	2.92		
				19.72	2.57		
				19.89	2.41		
PEG4000+NaH_2PO_4				PEG6000+ NaH_2PO_4 [37]			
298.15K		308.15K		298.15K		308.15K	
w_s	w_p	w_s	w_p	w_s	w_p	w_s	w_p
5.53	28.77	4.92	29.57	18.46	0.21	3.41	37.61
5.73	27.93	5.09	28.99	16.31	1.5	2.86	40.57
5.76	27.83	5.21	28.54	15.23	2.83	2.27	43.78
11.2	13.91	8.99	17.26	14.19	4.68	14.99	0.27
11.28	13.73	9.11	17.02	13.69	5.62	14.36	0.52
11.31	13.62	9.21	16.79	12.99	7.2	13.32	1.35
13.72	8.99	10.95	13.23	12.03	9.26	12.51	2.55
13.74	8.90	11.04	13.00	10.99	11.63	11.54	4.56
13.73	8.86	11.15	12.74	9.93	14.17	10.75	6.65
14.95	6.60	12.49	9.88	9.04	16.51	9.93	8.94

Table 1: Binodal curve data of the poly (ethylene glycols) (1500, 2000, 3000, 4000 and 8000)+ sodium di hydrogen phosphate + water system at (298.15 and 308.15) K and 0.1 MPa.

14. 96	6.58	12.75	9.48	8.19	18.94	9.15	11.31
14.97	6.57	12.99	9.11	7.46	21.26	8.19	14.34
15.92	5.11	14.54	6.33	6.15	25.5	7.18	17.92
15.94	5.08	14.76	5.90	5.01	30.22	6.68	20.28
15.90	5.07	15.57	4.71	4.14	33.86	6.06	22.95
16.47	4.22	16.02	3.97			5.39	26.18
16.49	4.18					4.79	29.02
16.52	4.15					4.1	32.97
16.98	3.54					3.44	36.95
17.00	3.52					2.87	41.23
17.03	3.51					2.53	44.12
17.44	3.06						
17.47	3.01						
17.50	3.00						

PEG8000+ NaH_2PO_4			
298.15K		308.15K	
w_s	w_p	w_s	w_p
4.33	30.11	3.46	31.83
4.46	29.58	3.62	30.56
4.64	28.79	3.78	29.77
8.41	13.54	6.94	14.72
8.52	13.35	7.05	14.58
10.07	9.5	7.17	14.25
10.13	9.38	8.8	9.27
10.14	9.21	9	8.91
11.28	5.71	9.17	8.41
11.68	4.84	10.57	5.07
12.04	4.11	10.78	4.75
12.09	4.01	11.04	4.08
12.56	3.35	11.9	2.25
12.55	3.09	12.13	1.83
12.91	2.56	12.32	1.41
12.93	2.46		
13.12	2.19		
13.25	2.07		

Standard uncertainties: u (w_i) = 0.02; u (T) = 0.05 K; u (pH) = 0.01 and u (P) = 5 kPa.

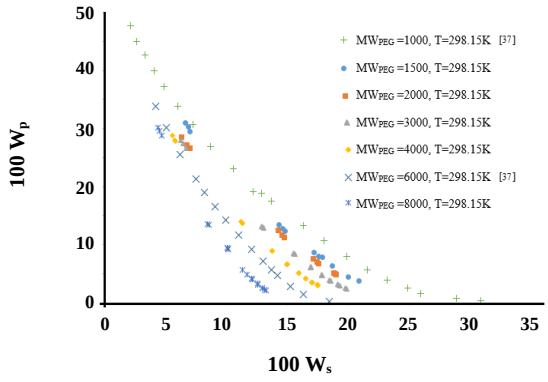

Fig. 3: The experimental binodal curves of the aqueous PEG1000 (+)[37],PEG1500(●), PEG2000(■),PEG3000(▲), PEG4000(◆), PEG6000(×)[37],PEG 8000(*)+ NaH2PO4 two-phase system at 298.15K.

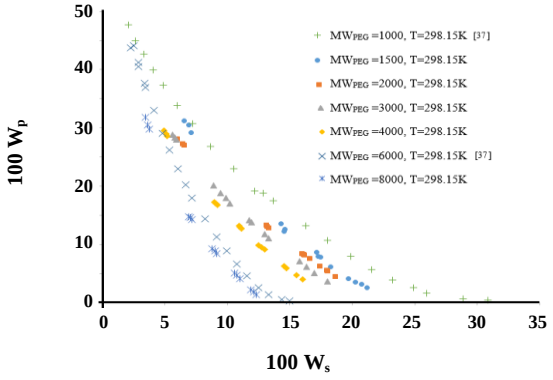

Fig. 4: The experimental binodal curves of the aqueous PEG1000 (+)[37],PEG1500(●), PEG2000(■),PEG3000(▲), PEG4000(◆), PEG6000(×)[37],PEG 8000(*)+ NaH2PO4 two-phase system at 308.15K.

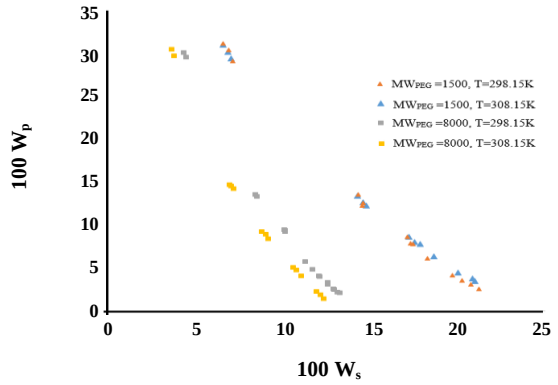

Fig. 5: The experimental binodal curves of the aqueous PEG1500, PEG 8000 + NaH_2PO_4 two-phase system at different temperatures (298.15 and 308.15K).

the lowest MSE in both training and testing phases. A further anaysis of the best results obtained with each method indicates that the same trend is kept for average absolute error, in the testing phase the following errors beeing obtained: 29.85% (SVM), 12.99% (ANN-DE) and 6.71% (SVM-DE). Fig. 6 presents a point by point comparison between predictions generated by the three approaches for the testing set and the experimental data.

The analysis of Table 2 indicates two aspects:

i) The use of DE for parameter determination considerably improves the performance of SVM. This is due to the fact that DE adapts the landscape of the search space and identifyes the near optimum position (combination of parameters). In contrast, the manual set-up does not guarantee finding a good solution and, in addition, it is computationally costly, as there are a multitude of parameters that can be tested.

ii) DE provided better parameters when is in combination with SVM than with ANNs. This can be explained by two factors: a) in this case, SVM has a better generalization capability; b) the number of parameters to optimized for ANNs were 3 x 20 + 20 x 1 + 3 x 24 = 144, while for SVM, the total number of parameters to optimized was 12; consequently, for ANNs, the search space beeing bigger, more computational resources are required to find the optimum.

CONCLUSIONS

The main objective of this study was the development and construction of a model that faithfully predicts the phase equilibria of seven PEG- sodium dihydrogen phosphate salt ATPSs. For this purpose, new expremental LLE data for five ATPSs containing sodium dihydrogen phosphate and poly (ethylene glycols) (PEG) (1500, 2000, 3000, 4000 and 8000) have been determined at different temperatures (298.15 and 308.15 K). PEG1000(or 6000) + NaH_2PO_4 + water [37] were used as to indicated the reliability of the new data and to compare its accuracy. The main trend observed is that with the increase of temperature and molecular mass, the two-phase region increases. In order to model the system, three cases were considered: SVM, SVM-DE and ANN-DE. As to have a reliable comparison, for the SVM-DE and ANN-DE, the same DE version with the same parameters was used. The simulation data indicated that the use of DE for parameter estimation improves performance and that SVM-DE

Table 2: Centralized results obtained with the 3 modelling approaches.

Approach		Fitness	MSE training	MSE testing	R^2 training	R^2 testing
ANN-DE	Best	884.28	1.13E-03	2.51E-03	0.9897	0.9912
	Worst	194.01	5.15E-03	1.18E-03	0.6381	0.4183
	Average	417.468	2.65E-03	5.51E-03	0.8027	0.7535
SVM-DE	Best	2962.391	3.38E-04	3.19E-04	0.9984	0.9983
	Worst	2739.599	3.65E-04	2.83E-04	0.9985	0.9983
	Average	2834.884	3.53E-04	3.06E-04	0.9984	0.9993
SVM	-	-	4.96E-03	4.99E-03	0.9840	0.9808

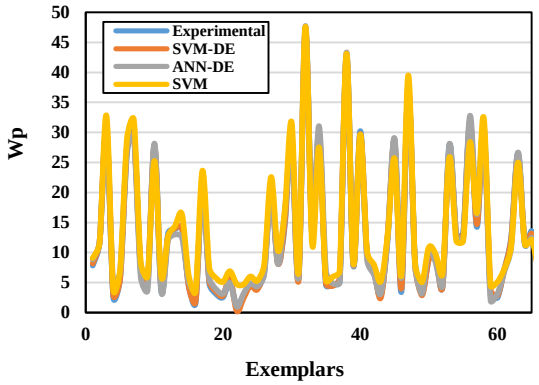

Fig. 6: Comparison between the predictions generated by the three approaches and the experimental data for the testing set.

provides the best results, its predictions beeing in close agreement with the experimental data. The average absolute error of SVM-DE for the testing dataset is 6.71%, which, for the considered process, is in an aceptable interval.

Received : Jun. 4, 2019 ; Accepted :Sep. 16, 2019

REFERENCES

[1] Rahimpour F., Feyzi F., Maghsodi S., Kaul R.H., Purification of Plasmid DNA with Polymer-Salt Aqueous Two-Phase System: Optimization Using Response Surface Methodology, *Biotech. Bioeng.* **95(4)**: 627-637(2006).

[2] Rahimpour F., Feyzi F., Maghsodi S., Kaul R.H., Optimization Refolding and Recovery of Active Recombinant Bacillus Halodurans Xylanase in Polymer-Salt Aqueous Two-Phase System Using Surface Response Analysis, *J. Chromatogr. A.*, **1141(1)**: 32-40 (2007).

[3] Zaveckas M, Zvirblieńe A., Zvirblis A., Chmieliauskaite V., Bumelis V., Pesliakas H., Effect of Surface Histidine Mutations and Their Number on the Partitioning and Refolding of Recombinant Human Granulocyte-Colony Stimulating Factor (Cys17Ser) in Aqueous Two-Phase Systems Containing Chelated Metal Ions, *J. Chromatography B.*, **852(1-2)**: 409-419 (2007).

[4] Shahbaz Mohamadia H., Omidinia E., Purification of Recombinant Phenylalanine Dehydrogenase by Partitioning in Aqueous Two-Phase Systems, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, **854(1-2)**: 273-278 (2007).

[5] Shahbaz Mohamadia H., Omidinia E., Process Integration for the Recovery and Purification of Recombinant Pseudomonas Fluorescens Proline Dehydrogenase Using Aqueous Two-Phase Systems, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, **929(1)**: 11-17 (2013).

[6] Lan Ch-W J., Yeh C-Y., Wang C-C., Yang Y-H., Wu H-S., Partition Separation and Characterization of the Poly Hydroxyal Kanoates Synthase Produced from Recombinant Escherichia Coli Using an Aqueous Two-Phase System, *J. Biosci. Bioeng.*, **116(4)**: 499-505 (2013).

[7] Ibarra-Herrera C., Aguilar O., Rito-Palomares M., Application of an aqueous two-phase system strategy for the potential recovery of a recombinant protein form Alfalfa (Medicago Sativa), *Sep. Purif. Technol.*, **77(1)**: 94–98 (2011).

[8] Wang Z.H., Song M., Ma Q., Two-Phase Aqueous Extraction of Chromium and its Application to Speciation analysis of Chromium in Plasma, *Mikrochim. Acta*., **134(1-2)**: 95–99 (2000).

[9] Gao Y.T., Wang W.W., Distribution Behavior and Extraction Mechanism of Gold(III) in Polyethylene Glycol Ammonium Sulphate Aqueous Biphasic System, *Chin. J. Appl. Chem*., **19(6)**: 578–580 (2002).

[10] Zheng Y., Tong Y., Wang Sh., Zhang H., Yang Y., Mechanism of Gold (III) Extraction Using a Novel Ionic Liquid-Based Aaqueous Two Phase System without Additional Extractants, *Sep. Purif. Technol.*, **154(1)**: 123–127 (2015).

[11] Patrício P.R., Mesquita M.C., da Silva L.H.M., da Silva M.C.H., Application of Aqueous Two-Phase Systems for the Development of a New Method of Cobalt(II), Iron(III) and Nickel(II) Extraction: A Green Chemistry Approach, *J. Hazard. Mater.*, **193(1)**: 311-318 (2011).

[12] Bulgariu L., Bulgariu D., Selective extraction of Hg(II), Cd(II) and Zn(II) Ions from Aqueous Media by a Green Chemistry Procedure Using Aqueous Two-Phase Systems. *Sep. Purif. Technol.*, **118(1)**: 209-216 (2013).

[13] Bulgariu L., Bulgariu D., Sârghie I., Măalutan Th., Cd(II) Extraction in PEG-based Two-Phase Aqueous Systems in the Presence of Iodide Ions, Analysis of PEG-rich Solid Phases, *Cent. Eur.J.Chem.*, **5(1)**: 291-302 (2007).

[14] de Lemos L. R., Araújo Campos R., Rodrigues G.D, da Silva L.H. M., da Silva M.C. H., Green Separation of Copper and Zinc Using Triblock Copolymer Aqueous Two-Phase Systems, *Sep. Purif. Technol.*, **115(1)**: 107–113 (2013).

[15] de Lemos L. R., Araújo Campos R., Rodrigues G.D, da Silva L.H. M., da Silva M.C. H., Copper Recovery from ore by Liquid–Liquid Extraction Using Aqueous Two-Phase System, *J.Hazard. Mater*., **237**– **238(1)**: 209-214 (2012).

[16] Khayati Gh., Ghanadzadeh Gilani H., Safari Keyvani Z., Extraction of Cu (II) Ions from Aqueous Media Using PEG/Sulphate Salt Aqueous Two-Phase System, *Sep Sci. Technol.,* **51(4)**: 601-608 (2016).

[17] Chen Y., Liu X., Lu Y., Zhang X., Investigation of Gallium Partitioning Behavior in Aqueous Two-Phase Systems Containing Polyethylene Glycol and Ammonium Sulfate, *J. Chem. Eng. Data.*, **54(7)**: 2002–2004 (2009).

[18] Zhang Y., Sun T., Lu T., Yan Ch., Extraction and Separation of Tungsten (VI) from Aqueous Media with Triton X-100-Ammonium Sulfate-Water Aqueous Two-Phase System without any Extractant, *J Chromatogr A*., **1474(1)**: 40-46 (2016).

[19] Yongqiang Zh. , Tichang S., Qingxia H., Qing G., Tieqiang L., Yingchao G., Chunhuan Y., A Green Method for Extracting Molybdenum (VI) from Aqueous Solution with Aqueous Two-Phase System without any Extractant, *Sep. Purif. Technol.*, **169(1)**: 151–157 (2016).

[20] Smolik M., Jakóbik-Kolon A., Porański M., Extraction of Zirconium and Hafnium in Polyethylene Glycol-Based Aqueous Biphasic System, *Sep. Purif. Technol*., **42(8)**: 1831–1841 (2007).

[21] Rito-Palomares M.,Practical Application of Aqueous Two-Phase Partition to Process Development for the Recovery of Biological Products, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci*., **807(1)**: 3–11 (2004).

[22] Asenjo J.A., Andrews B. A., Aqueous Two-Phase Systems for Protein Separation: A Perspective, *J. Chromatogr. A.* **1281(49)**: 8826-8835 (2011).

[23] Rocha MV., Nerli BB., Molecular Features Determining Different Partitioning Patterns of Papain and Bromelain in Aqueous Two-Phase Systems, *Int. J. Bio. Macromol*., **61**(1):204-211 (2013).

[24] Rosa P.A., Azevedo A.M., Sommerfeld S., Bäcker W., Aires-Barros M.R., Aqueous Two-Phase Extraction as a Platform in the Biomanufacturing Industry: Economical and Environmental Sustainability. *Biotechnol. Adv.*, **29**(6):559–567 (2011).

[25] Naganagouda K., Mulimani V.H., Aqueous Two-Phase Extraction (ATPE): An Attractive and Economically Viable Technology for Downstream Processing of Aspergillus Oryzae α-Galactosidase, *Process Biochem*., **43(11)**: 1293–1299 (2008).

[26] Luechaua F., Ling T. Ch., Lyddiatt A., A descriptive Model and Methods for Up-Scaled Process Rutes for Interfacial Partition of Bioparticles in Aqueous Two-Phase Systems, *Biochem. Eng. J.*, **50(3)**: 122–130 (2010).

[27] Hu R., Feng X. , ChenP., Fu M., Chen H. , Guo L., Liu B-F., Rapid, Highly Efficient Extraction and Purification of Membrane Proteins Using a Microfluidic Continuous-Flow Based Aqueous Two-Phase System. *J. Chromatogr. A.*, **1218(1)**: 171–177 (2011).

[28] Rodrigues G.D., Teixeira L.D., Ferreira G.M.D., da Silva M.D.H., da Silva L.H.M., de Carvalho R.M.M., Phase Diagrams of Aqueous Two-Phase Systems with Organic Salts and F68 Triblock Copolymer at Different Temperatures, *J. Chem. Eng. Data*., **55(3)**: 1158–1165 (2010).

[29] Espitia-Saloma E., Vázquez-Villegas P., Aguilar O., Rito-Palomares M., Continuous Aqueous Two-Phase Systems Devices for the Recovery of Biological Products, *Food Bioprod. Process*., **92(2)**: 101–112 (2014).

[30] Raghavarao K, Ravganathan T., Srinivas N., Barhate R., Aqueous Two Phase Extraction—An Environmentally Benign Technique, *Clean Technol. Environ. Policy*., **5(2)**:136–141 (2003).

[31] Rodrigues G.D., de Lemos L. R., da Silva L.H.M., da Silva M.C.H., Application of Hydrophobic Extractant in Aqueous Two-Phase Systems for Selective Extraction of Cobalt, Nickel and Cadmium, *J. Chromatogr. A*., **1279(1)**:13-19 (2013).

[32] Hatti-Kaul R., "Methods in Biotechnology: Aqueous Two-Phase Systems: Methods and Protocols", Humana Press Inc., Totowa, NJ (2000).

[33] Lv h., Zheng y., A Newly Developed Tridimensional Neural Network for Prediction of the Phase Equilibria of Six Aqueous Two-Phase Systems, *J. Ind. Eng. Chem*., **57**(1):377-386 (2018).

[34] Kan P., Lee Ch-J., A Neural Network Model for Prediction of Phase Equilibria in Aqueous Two-Phase Extraction, *Ind. Eng. Chem. Res.*, **35(6)**: 2015-2023 (1996).

[35] Alvarez-Guerra E.,Ventura S.P.M., Alvarez-Guerra M., Coutinho J.A.P., Irabien A., Modeling of the Binodal Curve of Ionic Liquid/Salt Aqueous Systems, *Fluid Phase Equilib*., **426(1)**: 10-16 (2016).

[36] Shahriari Sh., Shahriari Sh., Predicting Ionic Liquid Based Aqueous Biphasic Systems with Artificial Neural Networks. *J. Mol. Liq*., **197(1)**: 65-72 (2014).

[37] Zafarani-Moattar M.T., Sadeghi R., Liquid–Liquid Equilibria of Aqueous Two-Phase Systems Containing Polyethylene Glycol and Sodium Dihydrogen Phosphate or Disodium Hydrogen Phosphate Experiment and Correlation, *Fluid Phase Equilib*.**181(1-2)**: 95–112 (2001).

[38] dos Santos G.S., Luvizotto L.G.J., Mariani V.C., Coelho L.D.S., Least Squares Support Vector Machines with Tuning Based on Chaotic Differential Evolution Approach Applied to the Identification of a Thermal Process, *Expert Syst Appl*. **39(5)**: 4805-4812 (2012).

[39] Rodriguez-Galiano V., Sanchez-Castillo M., Chica-Olmo M., Chica-Rivas M., Machine Learning Predictive Models for Mineral Prospectivity: An Evaluation of Neural Networks, Random Forest, Regression Trees and Support Vector Machines, *Ore Geol. Rev*., **71(1)**: 804-818 (2015).

[40 Yarveicy H., Ghiasi M. M., Mohammadi A. H., Performance Evaluation of the Machine Learning Approaches in Modeling of CO_2 Equilibrium Absorption in Piperazine Aqueous Solution," *J.Mol.Liq.* **255(1)**: 375-383 (2018).

[41] Jaramillo F., Orchard M., Muñoz C., Antileo C., Sáez D., Espinoza P., On-Line Estimation of the Aerobic Phase Length for Partial Nitrification Processes in SBR Based on Features Extraction and SVM Classification, *Chem. Eng. J.*, **331(1)**: 114-123 (2018).

[42] Naseri F., Jafari F., Mohseni E., Tang W., Feizbakhsh A., Khatibinia M., Experimental Observations and SVM-Based Prediction of Properties of Polypropylene Fibres Reinforced Self-Compacting Composites Incorporating Nano-CuO, *Constr.Build.Mater*.**143(1)** :589-598 (2017).

[43] Ansari H. R., Gholami A., An Improved Support Vector Regression Model for Estimation of Saturation Pressure of Crude Oils, *Fluid Phase Equilib*. **402(1)**: 124-132 (2015).

[44] Eslamimanesh A., Gharagheizi F., Illbeigi M., Mohammadi A. H., Fazlali A., Richon D., Phase Equilibrium Modeling of Clathrate Hydrates of Methane, Carbon Dioxide, Nitrogen, and Hydrogen+Water Soluble Organic Promoters using Support Vector Machine Algorithm, *Fluid Phase Equilib*. **316(25)**: 34-45 (2012).

[45] Marghny M., Abd El-Aziz R. M., Taloba A. I., Differential Search Algorithm-Based Parametric Optimization of Fuzzy Generalized Eigenvalue Proximal Support Vector Machine, *Int. J. Comput. Appl*., **108(19)**: 0975–8887 (2014).

[46] Barani A., Pirdashti M.., Rostami A.A.,Liquid-Liquid Equilibrium of Poly (Ethylene Glycol) 1500 + di-Potassium Tartrate +water at Different pH (6.41, 7.74 and 9.05), *Fluid Phase Equilib*., **459(1)**: 1-9 (2018).

[47] Barani A., Pirdashti M., Rostami A.A.,Density, Viscosity, Refractive Index and Excess Properties of Binary and Ternary Mixtures of Poly (Ethylene Glycol), Water, and di-Potassium Tartrate at 298.15 and Atmospheric Pressure, *J. Chem. Eng. Data*., **63(1)**: 127–137 (2018).

[48] Arzideh S.M. ., Movagharnejad K., Pirdashti M., Influence of the Temperature, Type of Salt and Alcohol on Phase Diagrams of 2-Propanol + Inorganic Salt Aqueous Two-Phase Systems: Experimental Determination and Correlation, *J. Chem Eng.Data*., **63(8)**: 2813–2824 (2018).

[49] Curteanu S., Suditu G., Buburuzan A. M., Dragoi E. N., Neural Networks and Differential Evolution Algorithm Applied for Modelling the Depollution Process of Some Gaseous Streams, *Environ. Sci. Pollut. Res. Int*., **21(22)**: 12856-12867 (2014).

[50] Dragoi E.-N., Curteanu S., Cascaval D., Galaction A.-I., Artificial Neural Network Modelling of Mixing Efficiency in a Split-Cylinder Gas-Lift Bioreactor for Yarrowia Lipolytica Suspensions, *Chem. Eng. Commun*., **203(12)**: 1600-1608 (2016) .

[51] Bleotu I., Dragoi E. N., Mureşeanu M., Dorneanu S.-A., Removal of Cu(II) Ions from Aqueous Solutions by an Ion-Exchange Process: Modeling and Optimization, *Environ Prog. Sustain Energy*, **37(1)**: 605-612 (2018).

[52] Fernandez J. C., Hervas C., Martinez-Estudillo F. J., Gutierrez P.A., Memetic Pareto Evolutionary Artificial Neural Networks to Determine Growth/ No-Growth in Predictive Microbiology, *Appl. Soft. Comput*., **11(1)**: 534-550 (2011).

[53] Forciniti D., Hall C., Kula M.-R., Influence of Polymer Molecular Weight and Temperature on Phase Composition in Aqueous Two-Phase Systems, *Fluid Phase Equilib*., **61(3)**: 243-262 (1991).

[54] Voros N., Proust P., Fredenslund A., Liquid-Liquid Phase Equilibria of Aqueous Two-Phase Systems Containing Salts and Polyethylene Glycol, *Fluid Phase Equilib*., **90(2)**: 333-353(1993).

[55] Gonzalez-Amado M., Rodil E., Arce A., Soto A., Rodríguez O., The Effect of Temperature on Polyethylene Glycol (4000 or 8000)+(sodium or ammonium) Sulfate Aqueous Two Phase Systems, *Fluid Phase Equilib*., **428(1)**: 95-101 (2016).

[56] Hartounian H., Floeter E., Ka E., Sandler S., Effect of Temperature on the Phase Equilibrium of Aqueous Two-Phase Polymer Systems, *AIChE J*., **39(12)**: 1976-1984 (1993).