Comparison between artificial neural network and density-based models for the correlation of the solubility of some pharmaceutical compounds in supercritical carbon dioxide.

C. Si-Moussa[1,2], S. Hanini[2], R. Derriche[1], M. Kraouche[2], A. Abdallah Al Hadj[2].

[1]*Département de Génie Chimique, Ecole Nationale Polytechnique, Alger, Algérie.*
[2]*LBMPT, Université de Médéa, Quartier Ain D'Heb, Médéa, Algérie.*

simoussa_cherif@yahoo.fr

Abstract – This work compares Artificial Neural Networks (ANN) to four density based semi empirical modelling of the solubility of seven non-steroidal anti-inflammatory (NSAID), two anti-Cancer and two anti-AIDS drugs in supercritical carbon dioxide ($scCO_2$). Experimental literature data for the eleven drugs were used for training (152 data points) and validation (75 data points) of the ANN model. The model has five inputs (two intensive state variables and three pure drug properties) and one output (solubility of solid drug in $scCO_2$ in mole fraction). Statistical analysis of the predictability of the neural networks model shows excellent agreement with experimental data. Furthermore, the comparison in terms of average absolute relative deviation (AARD) between the ANN predicted results for each binary for the whole temperature and pressure range and results predicted by four density based models (Chrastil, Kumar and Johnston, Bartle et al., and Méndez-Santiago and Teja) show that the ANN model correlates far better the solubility of the eleven solid drugs in $scCO_2$.
Keywords: Solubility; Artificial neural networks; Density based models, solid drugs; Supercritical Carbon dioxide.

Introduction

The extensive experimental data of solubility of biomolecules and pharmaceutical compounds in supercritical fluids that are being published every year together with recent literature related to applications of supercritical fluid extraction of naturaceuticals, and pharmaceutical particle formation processes, such as RESS, SAS, SEDS and PGSS, are clear indications of the increasing level of interest in supercritical fluid (SCF) technology in the pharmaceutical industry. The accurate knowledge of the solubility of solid drugs in supercritical fluid phase and the detailed understanding of the corresponding phase equilibrium are of paramount importance for the effectiveness and the correspondent technical and economical success of SCF processes. Attempts of modelling the solubility of solid solutes in SCF phase for the purpose of correlation and or prediction have followed suite though they have not been yet satisfactory to the desired level of accuracy. Most commonly, the solubility of a solid in a SCF phase is correlated using empirical and semi-empirical models based on density of the SCF, pressure and temperature and equations of state (EoS)[1,2].

The EOS approach, whether using simple semi-empirical cubic EOS or more complex and theoretical-sound state of the art EOS [3, 4], employs the solute properties that are not always available and cannot easily determined experimentally. The cubic EOS representation of solid-SCF equilibrium, which is used most often due to the relative computational simplicity, requires critical properties, acentric factor, sublimation pressure and molar volume of the solute. In the absence of experimental values of these parameters, which is more often the case particularly for complex pharmaceutical compounds, they have to be estimated using group contribution methods, thus adding a factor of uncertainty to the approach. As an example, Valderrama and Zavaleta [5] showed that variations of 10 % in the sublimation pressure of the solute may produce deviations between 5 and 19 % in solubility calculations. Coimbra et al. [2] have also demonstrated the sensitivity of solubility correlation to critical properties when using cubic EOS.

The use of empirical and semi-empirical models has been extensively cited in literature. These models are based on simple error minimisation using least square method to determine the adjustable parameters of the model, and for most of them there is no need to use physical properties of

the solute. All these models assume a solute solubility dependence on the solvent density, temperature and pressure. This simple representation is the reason of the poor correlation that can be obtained for some systems which is why modification are being made, continuously, to previous models in order to increase the level of accuracy of the predictions. A recent and brief review of density-based models is given by Sparks et al. [6].

In the absence of efficient predictive models, artificial neural network (ANN) modelling has been suggested in a previous work [7] as a feasible and reliable alternative for correlating the solubility of solid solutes in scCO2. This work extends the application of ANN modelling of the solubility of solid drugs in $scCO_2$ to seven NSAID (Flurbiprofen, Ibuprofen, Ketoprofene, Nabumetone, Naproxen, Phenylbutazone, Salicylamide), two anti-Cancer (5-fluorouracil, Thymidine) and two anti-AIDS drugs (Azodicarbonamide, 2-Phenyl-4H-1,3-benzoxazin-4-one). The results in terms of average absolute relative deviation are compared to those obtained by four of the most commonly used density-based models (Chrastil [8], Kumar and Johnston [9], Bartle et al.[10], and Méndez-Santiago and Teja[11]). The experimental data of the solubility of solid drugs in sc CO_2 used in this work were those reported by Stassi et al.[12] for Ketoprophen; Ting et al.[13] for Naproxen; Charoenchaitrakool et al.[14] for Ibuprophen; Duarte et al.[15] for Flurbiprohpen; Su and Chen [16] for Nabumetone, Phenylbutazone, and Salicylamide; and Suleiman et al.[17] for 5-fluorouracil, Thymidine, Azodicarbonamide and 2-Phenyl-4H-1,3-benzoxazin-4-one. The PE software [18] was used for the estimation of critical properties and acentric factors of the solid drugs and the density of the solvent. The critical temperature and pressure of the solutes were calculated with Joback group contribution method, the acentric factor of solutes was determined by the Lee–Kesler correlation, and the CO_2 density was estimated with the BACK EOS.

Solubility modelling using density-based models

Most of the empirical models show a dependence of $\ln y_2$ on $1/T$ because solubility is closely connected to solute sublimation pressure, which can be expressed by a Clausius–Clapeyron equation. The models considered in this work are:

- Chrastil (CH)[8] model in which a dependence of $\ln y_2$ on $\ln \rho_1$ is considered:

$$\ln y_2 = a_0 + a_1 \ln \rho_1 + \frac{a_2}{T} \quad (1)$$

- Kumar and Johnston (KJ)[9] model in which a dependence of $\ln y_2$ on ρ_1 is proposed:

$$\ln y_2 = b_0 + b_1 \rho_1 + \frac{b_2}{T}, \quad (2)$$

- Bartle et al. (BR)[10] model in which a dependence of $\ln(yB_2P)$ on ρ_1 is considered:

$$\ln \frac{y_2 P}{P^{ref}} = c_0 + c_1(\rho_1 - \rho_1^{ref}) + \frac{c_2}{T}, \quad (3)$$

- Mendez-Santiago and Teja (MT)[11] model in which a dependence of $\ln(y_2P)$ on ρ_1 is considered:

$$T \ln y_2 P = d_0 + d_1 \rho_1 + d_2 T \quad (4)$$

Solubility modelling using ANN

In order to describe the phase behavior of the eleven CO_2(1)-solid drug(2) binaries by one ANN model a total of six variables have been selected in this work: three intensive state variables (equilibrium temperature, equilibrium pressure and the solubility of the solid drug in SCF phase) and three pure component properties of the NSAID (critical temperature, critical pressure and acentric factor). The equilibrium temperature (T), the equilibrium pressure (P) and the pure component properties of the solid drug (T_c, P_c and ω) have been selected as input variables and the solubility of the solid drug in the SCF phase (y_2) as the output variable (Fig.1). A detailed description of the strategy used for the application of ANN modeling of phase equilibria has been described in a previous work [19].

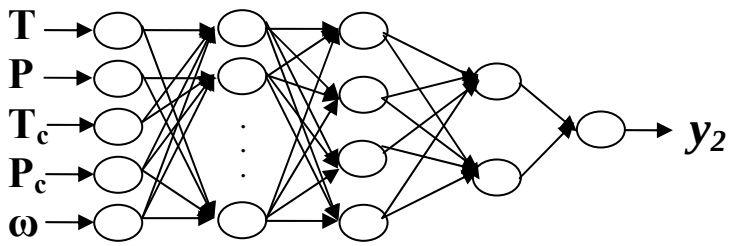

Figure 1: Multi-layer feedforward neural network for the prediction of the solubility of the solid drug in the SCF phase.

Results and discussion

In this work both the ANN modelling and the least square fitting of the parameters of density based models were carried out using MATLAB®. Table 1 shows the structure and the details of MATLAB functions used for the optimized ANN. For the least square fitting of the parameters o the density based models the *lsqcurvefit* MATLAB function was used with the option of the 'Levenberg-Marquardt' algorithm on.

A first and global comparison between the ANN models and density based models is shown through validation agreement plots of the estimated versus experimental solubility (Figs 2-6). It is clear from the coefficient of correlation and the distribution of data points on the y=x line that the ANN model correlates the experimental better than the four density based models. A more detailed comparison is considered in table 2, where the performance of each model, in terms of the average absolute relative deviation (AARD), is shown for each of the solid drugs. This table shows that to the exception of the CO_2-Azodicarbonamide system, where the Bartle et al. model performed best, the ANN model AARD for the remaining systems are the lowest and do not exceed 6.49%. Also, the table shows that none of the density-based models could correlate the data of the eleven systems better than the other models, though globally the Chrastil model performance is slightly better than the other three models. In order to illustrate further the correlating performance of the models the plots of solubility versus pressure for CO_2(1)-Nabumetone(2) at 328K , CO_2(1)-Phenylbutazone(2) at 328K and CO_2(1)-Azodicarbonamide(2) at 318K are shown in Fig. 7-9 respectively.

Table 1: Structure of the optimised ANN

Network type	'Feedforward backpropagation' : *newff*	
Layers	Number of neurons	Activation function
Input Layer	5	--------
1st Hidden Layer	24	Logarithmic Sigmoid : *logsig*
2nd Hidden Layer	14	Logarithmic Sigmoid : *logsig*
3rd Hidden Layer	9	Logarithmic Sigmoid' : *logsig*
Output Layer	1	Linear : *purelin*
Training Algorithm	'Levenberg-Marquardt backpropagation' : *trainlm*	

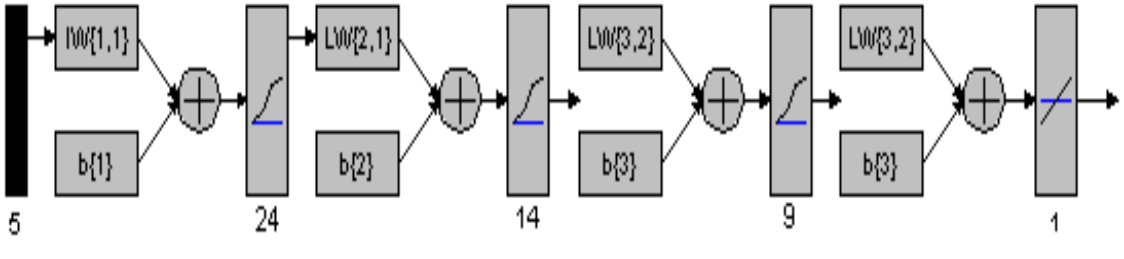

Table 2: Comparison of the AARD (%) of the predicted solubility of the solid drugs in $scCO_2$ obtained by ANN model and density-based models

System	AARD(%)				
	ANN	CH	KJ	BR	MT
Flurbiprofen	6.49	10.24	6.97	12.82	11.89
Ibuprofen	1.73	11.39	22.47	30.14	13.67
Ketoprofene	6.01	10.91	15.75	39.75	13.82
Nabumetone	1.54	10.70	23.77	17.41	19.55
Naproxen	2.69	6.55	4.36	5.39	5.31
Phenylbutazone	3.95	11.08	27.38	29.77	22.82
Salicylamide	3.32	7.72	16.45	26.53	12.71
5-fluorouracil	3.34	8.71	8.51	5.76	8.75
Thymidine	5.18	27.20	26.83	10.65	28.24
Azodicarbonamide	4.83	15.86	14.97	3.20	17.17
2-Phenyl-4H-1,3-benzoxazin-4-one	6.02	31.89	30.78	32.47	33.21
Global	4.35	14.52	18.57	19.27	17.61

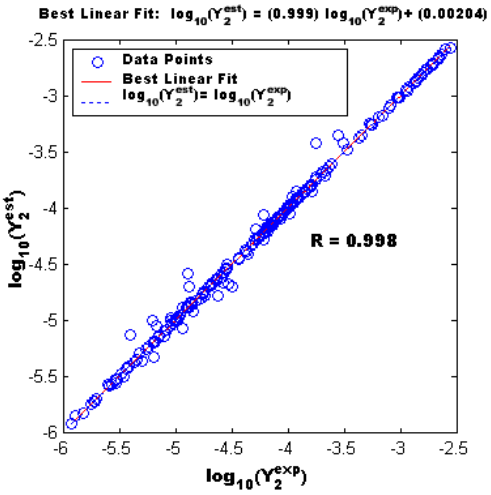

Figure 2: Validation agreement plot of the ANN model

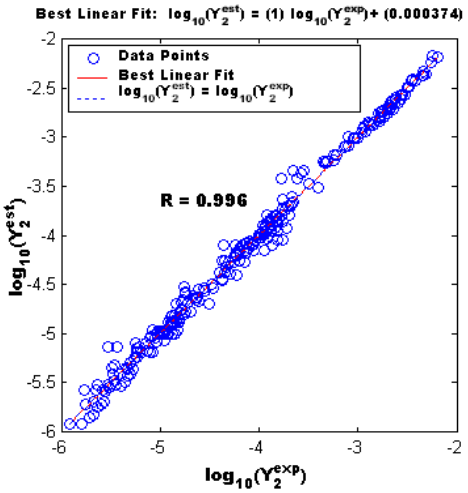

Figure 3: Validation agreement plot of Chrastil model

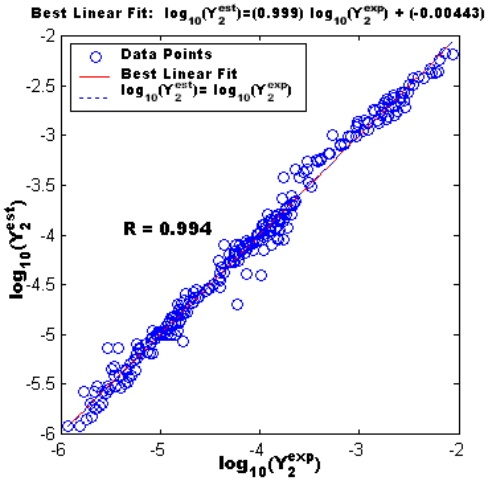

Figure 4: Validation agreement plot of Kumar and Johnston model

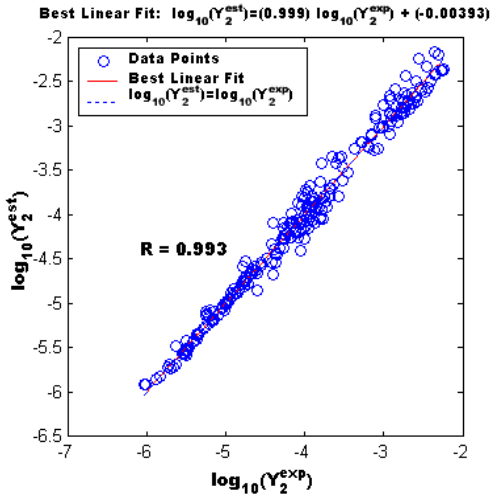

Figure 5: Validation agreement plot of Bartle et al. model

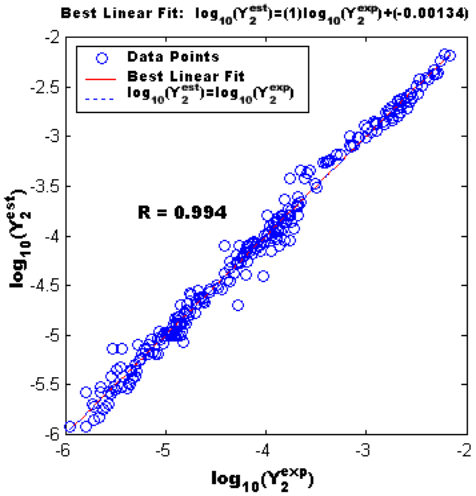

Figure 6: Validation agreement plot of Mendez-Santiago and Teja model

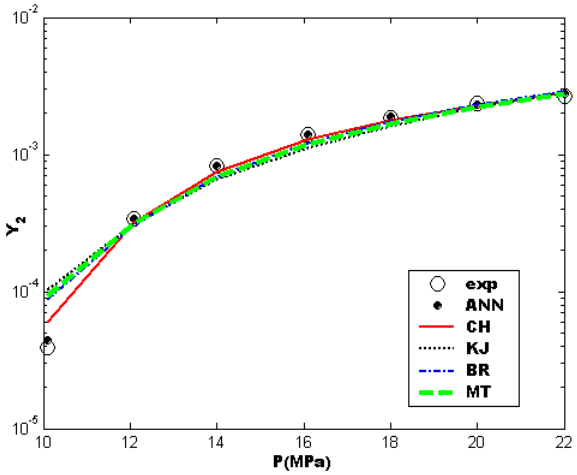

Figure 7: Solute solubility versus pressure plot for CO_2(1)-Nabumetone(2) at 328K.

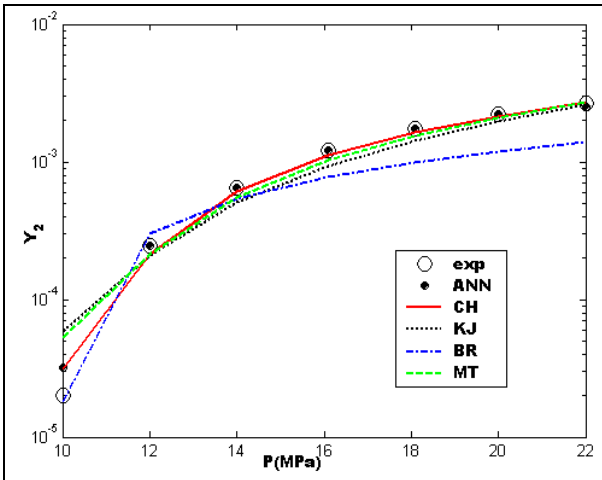

Figure 8: Solute solubility versus pressure plot for CO_2(1)-Phenylbutazone(2) at 328K.

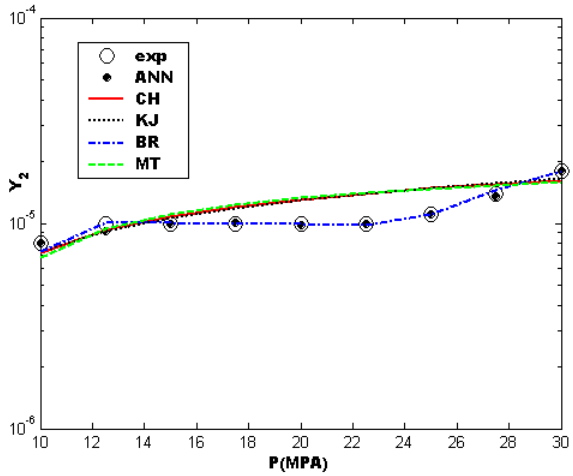

Figure 9: Solute solubility versus pressure plot for CO_2(1)-Azodicarbonamide(2) at 318K.

CONCLUSION

A feed forward artificial neural network model has been used to predict the solubility of eleven solid drugs in $scCO_2$ given the equilibrium temperature, the equilibrium pressure and the critical temperature, the critical pressure and the acentric factor of the solid solute. The optimized NN consists of five neurons in the input layer, three hidden layers with 24, 14 and 9 neurons respectively and one neuron in the output layer. This was obtained by applying a strategy based on assessing the parameters of the best fit of the validation agreement plots (slope and y intercept of the equation of the best linear fit and the correlation coefficient R^2) for the validation data set as a measure of the predictive ability of the model. The same data set has been correlated using Chrastil, Kumar and Johnston, Bartle et al. and Mendez-Santigo and Teja density-based models. The comparison in terms of the AARD% shows that the ANN model correlates the eleven systems far better than the four density-based models, and that none of these could correlate all the eleven systems better than the other models, though globally Chrastil model performed slightly better than the other three models. Therefore, the ANN model can be reliably used to estimate the solubility of the eleven solute-$scCO_2$ binaries within the ranges of temperature and pressure considered in this work without the need neither to accurate estimates of critical properties and acentric factor of the solid solutes nor to the density of the solvent. This study also shows that a single ANN model could be developed for the prediction of the solubility of a mixed drugs (having different therapeutic effect) in $scCO_2$, provided reliable experimental data are available, to be used in supercritical fluid processes.

References

[1] I. Ashour, R. Almehaideb, S.-E. Fateen, G. Aly, *Fluid Phase Equilibria,* **2000**, 167, p.41.
[2] P. Coimbra,C.M.M. Duarte H.C. de Sousa, *Fluid Phase Equil.*, **2006,** 139, p.188.
[3] C. Zhong, H. Yang, *Ind. Eng. Chem. Res.,* **2002,** 41, p.4899.
[4] H. Yang, C. Zhong, *J. Supercrit. Fluids*, **2005**, 33, p.99
[5] J.O.Valderrama, J. Zavaleta, *Ind. Eng. Chem. Res.* **2005**, *44*, p.4824.
[6] D.L. Sparks, R. Hernandez, L.A. Estévez, *Chem. Eng. Sci.*, **2008,** 63, p.4292.
[7] C. Si-Moussa, R. Derriche, S. Hanini, in the 11th European Meeting on Supercritical Fluids: Reactions, Materials and Natural Products Processing, May 4-7 **2008**, Barcelona, Spain.
[8] J. Chrastil, *J. Phys. Chem.*, **1982**, 86 (15), p.3016.
[9] S.K Kumar, K.P. Johnston, *J. Supercrit. Fluids*, **1988**, 1, p.15
[10] K.D. Bartle, A.A. Clifford, S.A. Jafar, G.F. Shilstone, *J. Phys. Chem. Ref. Data,* **1991**, 20, p.713.
[11] J. Méndez-Santiago, A.S. Teja, *Fluid Phase Equilibr.*, 158, **1999**, p.501.
[12] A. Stassi, R. Bettini, A. Gazzaniga, F. Giordano A. Schiraldi, *J. Chem. Eng. Data* **2000,** *45,* p.161165.
[13] S.S.T. Ting, S.J. Macnaughton, D.L. Tomasko, N.R. Foster, *Ind. Eng. Chem. Res.* **1993,***32***,** p.1471.
[14] M. Charoenchaitrakool, F. Dehghani, N.R. Foster H.K. Chan, *Ind. Eng. Chem. Res.* **2000,** *39***,** p**.**4794.
[15] A.R.C. Duarte, P. Coimbra, H.C. de Sousa, C.M.M. Duarte, *J. Chem. Eng. Data* **2004,** *49, p.*449.
[16] C.-S. Su, Y.-P. Chen, *J. Supercritical Fluids,* **2008,** 43, p.438.
[17] D. Suleiman, L.A. Estevez, J.C. Pulido, J.E. Garcia, C. Mojica, *J. Chem. Eng. Data,* **2005,** 50, p.1234
[18] O. Pfohl, S. Petkov, R. Dohrn, G. in: 8th International Conference on Properties and Phase Equilibria for Product and Process Design, April 26-May 1, **1998,** Noordwijkerhout, Netherlands.
[19] C. Si-Moussa, S. Hanini, R. Derriche, M. Bouhedda, A Bouzidi, *Braz. J. Chem. Engng*, **2008,** 25, p.183.