The Journal of Supercritical Fluids 133 (2018) 466–478

Contents lists available at ScienceDirect

The Journal of Supercritical Fluids

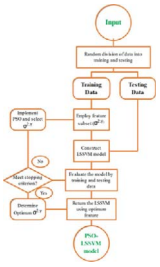
journal homepage: www.elsevier.com/locate/supflu

# Efficient modeling of drug solubility in supercritical carbon dioxide

Alireza Baghban[a,**], Ali Jalali[a], Amir H. Mohammadi[b,c,*], Sajjad Habibzadeh[a,d,**]

[a] Department of Chemical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Mahshahr Campus, Mahshahr, Iran
[b] Institut de Recherche en Génie Chimique et Pétrolier (IRGCP), Paris Cedex, France
[c] Discipline of Chemical Engineering, School of Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa
[d] Department of Chemical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Keywords:*
Drug
Solubility
Supercritical $CO_2$
Least square support vector machine
Particle swarm optimization
Correlation

## ABSTRACT

Development of new models and correlations to determine the drug solubility in supercritical fluids, i.e., supercritical $CO_2$, helps us to avoid time-consuming computations with poor outcomes while improving the supercritical technology. In this communication, a supervised technique, namely, the least square support vector machine (LSSVM) was employed to estimate the solubility of 33 different drug compounds in supercritical $CO_2$. The corresponding solubility was determined as a function of temperature, pressure, density of supercritical $CO_2$, and two physical properties of drug, i.e. molecular weight and melting point. The results obtained from the employed LSSVM model were compared with eight correlations. The obtained average absolute relative deviation and the square of regression coefficient for the testing group of LSSVM model were found to be 5.61% and 0.9975, respectively. Therefore, the model developed in this work can be reliably applied to the studied drugs by only knowing their physical properties.

## 1. Introduction

Supercritical fluids (SCF) as solvents with temperatures and pressures above their particular critical values are reliable alternatives for organic solvents [1]. This is due to their adjustable properties with temperature and pressure. In addition, SCFs have been extensively employed in different areas of separation processes, reactions, purification and particle sizing of pharmaceuticals, etc [2–4]. Among various types of SFCs, supercritical carbon dioxide has captured much interest in pharmaceutical processing since it is economical, inflammable, environmentally safe, and inherently non-toxic.

Furthermore, accurate prediction of solid solute solubility in an SFC has the major contribution to any SCF technology development. Therefore, the corresponding thermophysical properties must have been modeled and determined at the early stage, leading to studying solubility of pharmaceutical compounds at different temperatures and

* Corresponding author at: Institut de Recherche en Génie Chimique et Pétrolier (IRGCP), Paris Cedex, France.
** Corresponding authors at: Department of Chemical Engineering, Amirkabir University of Technology.
*E-mail addresses:* Alireza_baghban@alumni.ut.ac.ir (A. Baghban), a.h.m@irgcp.fr, amir_h_mohammadi@yahoo.com (A.H. Mohammadi), sajjad.habibzadeh@aut.ac.ir (S. Habibzadeh).

https://doi.org/10.1016/j.supflu.2017.10.032
Received 30 July 2017; Received in revised form 26 October 2017; Accepted 26 October 2017
Available online 21 November 2017
0896-8446/ 

pressures [5–7]. Nevertheless, in order to avoid disturbance to the equilibrium, it would be needed to develop reliable experimental solubility in the supercritical region [6]. As a result, a correlation for estimating the solubility of drugs in SC-$CO_2$ might be beneficial for the preparatory design of related processes.

Several methods have been employed for correlations of the experimental solid solubility in SC-$CO_2$ [8–15]. They are typically classified into either equation-of-state based correlations or density based correlations, e.g., semi-empirical models; however, the semi-empirical models compared to equation-of-state correlations are often utilized. This is partly due to their relative ease of usage; thus, it is not needed to employ physicochemical properties, such as critical properties and sublimation pressure that cannot be directly obtained experimentally and should be calculated by various methods [7].

Recently, artificial intelligence approaches have broadly been applied in numerous fields, especially in oil and gas areas. Baghban et al. have modeled solubility of some gases, including $CO_2$, $H_2S$, and $NH_3$ in ionic liquids and amine solutions by using different embranchments of artificial intelligence, i.e., artificial neural network (ANN), fuzzy logic system, adaptive neuro-fuzzy inference system (ANFIS), and support vector machine (SVM) [16–19].

The current study deals with the usability of least square-SVM paradigm, as a simplification of conventional SVM, to estimate the solubility of 33 different drugs in supercritical $CO_2$. A comprehensive actual data bank was used for developing this method [20–30]. In addition, particle swarm optimization (PSO) technique was coupled with the aforementioned LSSVM as an optimization theme for indicating of hyperparameters. The model is proposed based on temperature, pressure, density of supercritical $CO_2$, and two physical features of drugs, namely, molecular weight and melting point. The outcomes has been compared with Chrastil, Sung and Shim, Adachi and Lu, Spark et al., Gordillo et al., Mendez-Santiago and Teja, and Keshmiri et al. correlations for different drugs [8–15]. Results show that the proposed drug solubility model works much better than the others. To the best of our knowledge, no record for modeling of drug solubility in supercritical $CO_2$ by LSSVM approach is found in the literature.

## 2. Review of empirical correlations

Several empirical correlations have been developed to predict/estimate the solid solubility in $SCCO_2$. These models are typically density based methods in which the effects of temperature and pressure are both considered, leading to a linear relationship between solvent density and experimental data. The first model that is commonly used to correlate the solid solubility in $SCCO_2$ was developed by Chrastil et al. [8]. This equation is given as follows:

$$\ln S = A_1 + \frac{A_2}{T} + A_3 \ln \rho \tag{1}$$

where S is the solubility (kg $m^{-3}$) of the species in $SCCO_2$, ρ is the solvent density (kg $m^{-3}$), T is the temperature (K), $A_1$, $A_2$ and $A_3$ are characteristic constants for the binary system that depend on the molar mass of the solute.

Bartle et al. semi-empirical model to correlate the solubility of the solute in SCF is given by [9]:

$$\ln\left(\frac{yP}{P_{ref}}\right) = A_1 + \frac{A_2}{T} + A_3(\rho - \rho_{ref}) \tag{2}$$

where $P_{ref}$ (1 bar) and $\rho_{ref}$ (700 kg $m^{-3}$) are the reference pressure and the density, respectively. $A_1$, $A_2$ and $A_3$ are empirical constants. Furthermore, Sung and Shim proposed an equation to correlate the solubility of solute in $SCCO_2$ by adding a term to Chrastil's equation. The resulting equation is expressed in the form of the following empirical formula [10]:

$$\ln y = A_1 + \frac{A_2}{T} + \left(A_3 + \frac{A_4}{T^2}\right)\ln \rho \tag{3}$$

where y is the solubility in mole fraction and ρ is the $CO_2$ density (kg $m^{-3}$).

The next equation that is considered herein was developed by Adachi and Lu [11]. They modified Chrastil's equation led to the following expression:

$$\ln S = A_1 + (A_2 + A_3\rho + A_4\rho^2)\ln \rho + \frac{A_5}{T} \tag{4}$$

The fifth equation that is considered herein was suggested by Spark et al. based on Adachi–Lu model who provided an improved correlation [12]:

$$\ln S = A_1 + (A_2 + A_3\rho + A_4\rho^2)\ln \rho + \frac{A_5}{T} + \frac{A_5}{T^2} \tag{5}$$

In contrast to Eqs. (1)–(5) in which the effect of pressure was not considered, Gordillo et al. presented a model including pressure by the following equation [13]:

$$\ln y = A_1 + A_2P + A_3P^2 + A_4PT + A_5T + A_6T^2 \tag{6}$$

where P is the pressure (bar), y is the mole fraction of solute and T is the temperature (K). In addition, Mendez-Santiago and Teja introduced the equation below based on dilute solutions theory [14]:

$$T \ln(yP) = A_1 + A_2T + A_3\rho \tag{7}$$

Finally, the last empirical equation to correlate solute solubility was developed by Keshmiri et al. with the input of temperature, pressure, and density of pure supercritical $CO_2$ [15]. This equation is given as follows:

$$\ln y = A_1 + \frac{A_2}{T} + A_3P^2 + \left(A_4 + \frac{A_5}{T}\right)\ln \rho \tag{8}$$

In all of the semi-empirical models, $A_i$ coefficients are determined through data regression using a conventional least square method.

## 3. Theory

### 3.1. Least square support vector machine (LSSVM)

Support vector machine (SVM), as an intelligent approach was first introduced by Vapnik [31]. SVM is a quite effective method used widely for classification, regression and pattern recognition. The cornerstone of SVM transforms the nonlinear input area to a high-dimensional properties area while finding a hyperplane via a nonlinear mapping. This novel approach is based upon the statistical learning theory (SLT) and the structural risk minimization (SRM) concepts [32,33]. Moreover, SVMs lead to the solutions by solving the quadratic programming (QP), and the solution returned is global (or even unique) instead of many local ones. In addition, unlike other regression techniques such as neural networks, QP problem is considered as a convex function. However, this approach may be time-intensive and difficult since it functions based on the solution of a set of nonlinear equations (quadratic program). Suykens and Vandewalle proposed least square support vector machine (LSSVM) model as an alternate formulation of SVM regression [34]. LSSVM not only takes the advantages of SVM, but also requires solving a set of linear equations (linear programming) instead of a quadratic programming (QP) problem. This renders LSSVM computationally simpler, thereby, makes the problem easier to deal with. The formulation of LSSVM for nonlinear function estimation is expressed as follows. Given training set $\{x_k, y_k\}$, $k = 1,2, ...,N$, where $x_k \in R^n$ is the $k$th input data in input space and $y_k \in R^n$ is output value for given value of specific input variable (i.e. $x_k$) and $N$ refers to the number of the training samples. Using nonlinear function (.), which maps the training set in input space to the high (and possibly infinite) dimensional space, the following regression model is constructed: