Fluid Phase Equilibria 412 (2016) 228–234

Contents lists available at ScienceDirect

Fluid Phase Equilibria

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

Modeling of the vaporization enthalpies of petroleum fractions

Arash Kamari [a], Mehdi Sattari [a], Amir H. Mohammadi [a, b, c, *], Deresh Ramjugernath [a, **]

[a] Thermodynamics Research Unit, School of Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa
[b] Institut de Recherche en Génie Chimique et Pétrolier (IRGCP), Paris Cedex, France
[c] Département de Génie des Mines, de la Métallurgie et des Matériaux, Faculté des Sciences et de Génie, Université Laval, Québec, QC, G1V 0A6, Canada

ARTICLE INFO

Article history:
Received 2 August 2015
Received in revised form
17 December 2015
Accepted 18 December 2015
Available online 22 December 2015

Keywords:
Vaporization enthalpy
LSSVM
Empirical correlation
Comparative study
Outlier detection

ABSTRACT

The vaporization enthalpies of petroleum fractions are required in processing and transportation applications, as well as to optimize and design oil and gas production, and in heat flux calculations. A model was developed in this study for determining the vaporization enthalpies of pure hydrocarbon components and petroleum fractions. The model uses a least squares support vector machine (LSSVM) algorithm which is adjusted using a coupled simulated annealing (CSA) tuning tool. The inputs into the model are the boiling point temperature, specific gravity, and molecular weight. A comparative study was also undertaken between the method developed in the current study, neural network models, and some previously published empirical correlations. The Leverage approach was used to identify probable outlier data. It is found that only one data point from the vaporization enthalpy database in literature is an outlier. The results obtained indicate that the proposed CSA-LSSVM method is more rapid, accurate, and effective than available empirical correlations and neural network modeling for determination of the vaporization enthalpies of pure hydrocarbon components, as well as petroleum fractions. The LSSVM mathematical algorithm provides good results, as are indicated by an overall average absolute relative deviation of the predicted property from the real data of 1.16%, and an overall squared correlation coefficient of 0.9982.

1. Introduction

A simple definition of vaporization enthalpy is the difference between the enthalpies of the vapor and liquid phases at the same equilibrium pressure and temperature. Vaporization enthalpy (ΔH^{vap}) is the energy needed to transform a quantity of liquid substance into a vapor phase at its boiling point temperature [1]. As a result, the vaporization enthalpy for petroleum fractions and hydrocarbon components is a property that is used in many chemical disciplines, as well as is in the oil and gas industries. From a thermodynamic perspective, vaporization enthalpies can be applied in processing and transportation facilities for the optimization and design of oil and gas production and for heat flux calculations, as well as the estimation of some physical phenomena like the solubility parameters of hydrocarbons [2]. The vaporization enthalpy of pure components, in particular hydrocarbons, and also petroleum fractions is a key and fundamental thermodynamic property which is related to the specific gravity (S), boiling point temperature (T_b), and molecular weight (M) through various thermodynamic relationships.

The vaporization enthalpy is important from both an experimental and theoretical perspective because of its use in engineering optimization and design, and thus experimental techniques, correlations, and estimation models have been developed to provide greater thermodynamic insight to assist petroleum and chemical engineers in their work [3]. For hydrocarbon components, Vetere [4,5] developed two empirical correlations for the calculation of vaporization enthalpy using two variables, viz. molecular weight and normal boiling temperature. Riazi and Daubert [6] proposed an empirical correlation for predicting the vaporization enthalpy as a function of T_b and S. Both the Vetere and Riazi and Daubert correlations show an estimation error of approximately 7%.

Mohammadi and Richon [2] developed a simple correlation for

---

* Corresponding author. Thermodynamics Research Unit, School of Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa

** Corresponding author.

E-mail addresses: a.h.m@irgcp.fr, amir_h_mohammadi@yahoo.com (A.H. Mohammadi), ramjuger@ukzn.ac.za (D. Ramjugernath).

http://dx.doi.org/10.1016/j.fluid.2015.12.033

the vaporization enthalpy which is a function of the T_b and S. It is capable of calculating the vaporization enthalpies of pure hydrocarbon components and petroleum fractions. They also proposed an artificial neural network (ANN) tool for comparison of the results obtained from their empirical correlation. The results showed good agreement between their empirical correlation, the ANN model, and experimental values. Parhizgar et al. [1] proposed an empirical method for determination of vaporization enthalpies of pure hydrocarbon components and petroleum fractions using genetic programming which is a function of the T_b and S. Their results indicated that their correlation can calculate the vaporization enthalpy of both pure hydrocarbon components and petroleum fractions with an average absolute relative deviation (AARD) of approximately 1.35%.

Most of the available techniques for the prediction of vaporization enthalpy have been based on laboratory measured data for hydrocarbon components, and hence, may not be accurate and reliable for petroleum fluids [2]. Furthermore, the estimation of the enthalpy of vaporization using conventional thermodynamic methods requires values for many adjustable parameters. In other words, to predict thermo-physical properties, such as vaporization enthalpy, existing thermodynamics models may have some shortcomings. The thermodynamic methods normally require accurate characterization of reservoir fluids and petroleum fractions. Thermodynamic models also require reliable multiphase-multicomponent flash calculation algorithms and accurate properties. Furthermore, some of them may have convergence problems. Consequently, there is still a need for a fast, yet robust, predictive method for the determination of the vaporization enthalpy for both petroleum fractions and hydrocarbon components. To this end, smart techniques such as least square support vector machines (LSSVM), ANN approaches, genetic algorithms, particle swarm optimization, *etc.*, have been successfully employed in recent years for solving regression and classification problems in petroleum and chemical engineering [7–16]. Therefore, in the current study a new model has been proposed which uses a reliable technique, namely, the least square support vector machines (LSSVM) algorithm for the determination of the vaporization enthalpy of pure hydrocarbon components, as well as petroleum fractions as a function of M, S and T_b using approximately 122 data values.

2. Vaporization enthalpies data

Vaporization enthalpy data are normally determined at the normal T_b through an appropriate technique and are then calculated at the required temperature. There are two classes of empirically derived methods for determining the vaporization enthalpy. The first class of correlations relate the vaporization enthalpy at the normal boiling point temperature to the critical properties and the normal boiling point [17]. The next class of correlations relate vaporization enthalpy to the specific gravity, molecular weight, and the normal boiling point temperature [2]. As a result, the selection of the most appropriate input/predictor variables, in other words, the efficiency of the databank used for building the LSSVM models plays a significant role in the model efficiency, accuracy and reliability [18,19]. Therefore, to accurately predict the vaporization enthalpy (ΔH^{vap}, *kJ/g-mol*) of pure hydrocarbon and petroleum fractions the parameters which most influence the property, viz. the boiling point temperature (T_b, *K*), specific gravity, and molecular weight (*M, g/g-mol*) are gathered from literature [2,3]. The collected databank covers an extensive range of the vaporization enthalpy, from 19.0 to 80.1 *kJ/g-mol* for both petroleum fractions and pure hydrocarbon components. Distribution of the collected data in terms of minimum, and maximum, as well as averages are summarized in Table 1, with the input variables being T_b, S, and M, and the output parameter, ΔH^{vap}. In the current study, the collected vaporization enthalpy databank is randomly divided into two sub-data sets comprising of the “Training” set (80% of the data, 98 data points) and the “Test” set (20% of the data, 24 data points).

Table 1
Distribution of the data used in this study for forecasting the vaporization enthalpies of pure hydrocarbon components and petroleum fractions.

Parameter	Min.	Avg.	Max.	Type
T_b, K	231.1	451.9	722.8	Input
S	0.5	0.7	0.8	Input
M, g/g-mol	44.1	160.2	422.8	Input
ΔH^{vap}, kJ/g-mol	19.0	42.1	80.1	Output

3. Methodology

The aim of this study was to develop a nonlinear relationship between data reported in literature as inputs of the model (specific gravity, boiling point temperature, and molecular weight) and its corresponding output (vaporization enthalpies of petroleum fractions and pure hydrocarbons). For this task, an appropriate predictive tool which is consistent and precise is needed. The support vector machine (SVM) computing strategy which is based on statistical machine-learning approach which uses the structural risk minimization principle [20] was used in this study to accomplish the task. This mathematical algorithm is based on approaching the upper bound of the structural risk or minimum estimation/representation errors, and it differs from back-propagation neural networks in its approach to empirical risk or minimum learning errors [21]. A SVM is a novel intelligent technique which identifies patterns and analyses data. It is also used for solving regression problems and is additionally characterized as a non-probabilistic binary linear classifier.

Suykens and Vandewalle [20] presented a modified form of the classical SVM to support the solutions of the classical SVM system set of nonlinear formulas (quadratic programming). The resulting least-squares SVM (LSSVM) [20] strategy benefits from the advantages of the classical SVM system, although it only requires the solving of a set of linear programming (linear equations) which results in a quicker technique to the classical form/version of SVM methodologies. As a consequence, the regression deviation for the LSSVM mathematical methodology is computed as the difference between the determined/estimated output and the actual values, and is considered in addition to the tuning problem limitation. Normally, the value of the regression deviation is adjusted during the computations in most commonly-applied SVM plans, while it is mathematically expressed in the LSSVM system [20,22,23]. The penalized cost function for the LSSVM strategy employed in the current study is expressed as follows:

$$Q_{LSSVM} = \frac{1}{2}w^T w + \gamma \sum_{k=1}^{N} e_k^2 \tag{1}$$

with the following constraints [20,22,23]:

$$y_k = w^T \varphi(x_k) + b + e_k \quad k = 1, 2, 3, \ldots N \tag{2}$$

where *w* expresses the regression weight or the slope of linear regression, *b* expresses the intercept of the linear regression in the LSSVM approach, γ stands for the relative weight of the summation of the regression errors in comparison with *w*, *x* denotes the input vector for the parameters of the LSSVM approach, *y* is the output, e_k shows the regression deviation for *N* training objects, φ stands for