12º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS

TÍTULO DO TRABALHO:

SURFACE TENSION OF HYDROCARBONS: AN ARTIFICIAL NEURAL NETWORK APPROACH

AUTORES:

Lucas Silva Queiroz[1], Vinícius Ferreira da Silva Bueno[2], Maria Fernanda Salomon Bernardo[1*], Hyago Braga dos Santos[1], Larissa Maria Fernandes Gatti[2], Victor Rolando Ruiz Ahón[1], Troner Assenheimer de Souza[1]

[1]Departamento de Engenharia Química e de Petróleo, Universidade Federal Fluminense, Niterói [2]Departamento de Engenharia Química, Universidade Tecnológica Federal do Paraná, Londrina

INSTITUIÇÃO:

[1] Chemical and Petroleum Engineering Department, Federal Fluminense University, Niterói, RJ, 156, 24210-420, Brazil.

[2] Chemical Engineering Department, Federal Technological University of Parana, Londrina, PR, 3131, 86036-370, Brazil.

SURFACE TENSION OF HYDROCARBONS: AN ARTIFICIAL NEURAL NETWORK APPROACH

Abstract

Surface tension represents a significant thermodynamic property of fluids that impact several processes in petroleum engineering and science. Accurate estimates of its values are required for the correct application and development of processes such as flow processes and oil recovery. In this regard, predictive models for surface tension represent an alternative to parameter estimation. Considering the nonlinear nature of surface tension, artificial neural networks (ANN) were used in the current study to obtain a predictive model for surface tension estimation through 75 groups of organic pure compounds, encompassing 21,444 data points, and using PVT properties as temperature (T), critical temperature (T_c), critical pressure (P_c), and molecular weight (MW). The development of the model was carried out applying Python programming, using a multilayer perceptron (MLP) architecture with seven hidden layers. The model accuracy was assessed by using the mean square error (MSE), and the coefficient of determination (R^2) was used to assess the accuracy of the experimental data fit. In addition, a K-fold cross-validation (CV) method was applied to validate the model and improve its robustness. The obtained results showed an average MSE value for CV of $1.26x10^{-4}$. By the fine-tuning of the hyperparameters, the MSE value was further reduced to $8.14x10^{-5}$ and an R^2 value of 0.9766 was achieved. The obtained average relative deviation (ARD) was 5.73%, shorter than the obtained for the correlation of Brock and Bird, indicating the accuracy of the proposed model in predicting surface tension of a wide range of compounds, highlighting the potential of ANN models in understanding fluid behaviors.

Keywords: Machine learning, PVT properties, data analysis, thermodynamics properties.

Introduction

Surface tension is defined as the amount of work required to increase the surface area of the liquid by a unit amount, through an isothermal and reversible process. This work resulted from the differences between the molecular interactions of the vapor and the liquid phases, as well as the unbalance of these forces in the interface liquid-vapor (Ahmed, 2007). In the thermodynamic field, the surface tension (γ) is defined as the variation of Gibbs free energy of a system (δG) with area unit (δA) variation, given fixed conditions of temperature (T) and pressure (P) (Ball, 2011).

This property plays a crucial role in both industrial and natural phenomena (Ramírez-Verduzco, Romero-Martínez e Trejo, 2006). More specifically, the determination of this property in organic compounds is accounting for effects that impact several areas such as pharmaceutical sciences (Fathi Azarbayjani, Jouyban e Chan, 2009), food industry (Adhikari *et al.*, 2007; Bhandari e Howes, 2005), energy efficiency (Pucko, Racar e Faraguna, 2022), and petroleum industry (Cevada *et al.*, 2018). Therefore, accurate measuring and estimation of surface tension are necessary to enable its applicability for obtain precise results and reliably predict fluid behavior (Miqueu *et al.*, 2008).

In this regard, different methods have been proposed to determinate surface tension of pure materials and substances (Brock e Bird, 1955; Pitzer e Curl, 1957), non-polar and weakly polar pure fluids and mixtures (Zuo e Stenby, 1997), and volatile compounds in petroleum fluids (Miqueu *et al.*, 2008). More recently, research and studies have been conducted to create correlations to predict surface tension of ionic liquids ((Ghatee *et al.*, 2010)), saturated hydrocarbons (Tian *et al.*, 2017), such as other sorts of compounds (Mulero, Cachadiña e Bautista, 2021; Nicola, Di, Coccia e Pierantozzi, 2016).

A wide number of recent research for predicting properties of petroleum fluids – also known as PVT (Pressure-Volume-Temperature) properties – have been carried out using artificial neural networks (ANN) models (Lashkarbolooki e Bayat, 2018; Pazuki, Nikookar e Sahranavard, 2011). In terms of improving the results obtained, the application of ANN models offers significant advantages over traditional empirical models and theoretical approaches, as these models allow precise identification of non-linear relationships and interactions between molecular features compared to conventional methods, taking into account that they have the flexibility to manage large and diverse data sets and allow the integration of multiple input variables to improve prediction accuracy (Chen *et al.*, 2023).

In this context, given the limited number of robust models found in the literature to predict the surface tension of a wide range of components, the objective of the present work was to propose a new model using artificial neural networks approach for estimating the surface tension of 75 pure organic compound groups, encompassing over than 1,600 components, providing a valuable tool for assessing this property across many compounds. In addition, the influence of different input features was explored, evaluating different ANN architectures to determine the optimal configuration, improving the surface tension prediction and achieving an accurate adjustment for the required parameters. The effectiveness of the developed models was assessed through rigorous validation against a large number of experimental data points, demonstrating the consistency and reliability of the new proposed model in comparison with existing literature.

Methodology

In the first moment, an extensive database was collected from literature, encompassing 75 different organic pure groups through 21,444 experimental data points. Several thermodynamic variables such as temperature (T), critical temperature (T_c), reduced temperature (T_r), normal boiling point temperature (NBP), critical pressure (P_c), acentric factor (ω), critical volume (V_c), and molecular weight (MW) were processed and used as input features for the development of the model. The data was inputted in a Python environment and processed using *Scikit-learn*, *Pandas*, *Keras* and *TensorFlow* libraries. The database was categorized into input and output datasets comprehending the features and the surface tension values, respectively. In addition, the data frames were subdivided into training, testing, and validation subsets to ensure consistency throughout the model development.

Different metrics were employed on the analysis of the model to assess performance, improve results and optimize computational resources. The mean squared error (MSE), represented by Equation 1, was selected to evaluate the quality of the data regression in the developed model (loss function).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\sigma_{i,exp} - \sigma_{i,rep./pred.}\right)^2 \quad (1)$$

$\sigma_{i,exp}$ and $\sigma_{i,rep./pred.}$ represents experimental surface tension and predicted surface tension, respectively.

A 5 K-fold cross-validation (CV) methodology was also applied to provide consistency to the proposed model. In addition, the coefficient of determination (R^2) – Equation 2 – provides a perspective of how accurate the outcomes were estimated by the ANN based model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\sigma_{i,exp} - \sigma_{i,rep./pred.}\right)^2}{\sum_{i=1}^{n}\left(\sigma_{i,rep./pred.} - \bar{\sigma}\right)^2} \quad (2)$$

The average relative deviation (ARD) – Equation 3 – was employed with objective of measure the divergence between the experimental and predicted data and to compare the obtained results of the current study against different literature models.

$$ARD\% = \frac{100}{N}\sum_{i}^{N}\frac{\left|\sigma(i)^{pred} - \sigma(i)^{exp}\right|}{\sigma(i)^{exp}} \quad (3)$$

The impact of different features on the performance of the model were evaluated through the Pearson correlation coefficient (PCC), aiming to analyze linear relationships among variables, and evaluate the influence of linearity inside the features of the model. *PairPlot* is a graphically visualization which represents pairwise relationships in a data set by a plot of each feature against every other feature and explore potential non-linear behaviors between them, supporting the identification of patterns not evidenced during PCC study and bringing robustness for the model.

The architecture configuration was assessed by conducting tests to analyze the number of hidden layers, and epochs. To compare the obtained results with previous correlations from the literature, the equations of Brock and Bird (1955), Curl and Pitzer (1958), and Zuo and Stenby (1997) were employed in the current study and were shown in Table 1. On Brock and Bird (1955) equation, parameter α_c represents the Riedel parameter (Reidel, 1954) at the critical point. On Zuo and Stenby (1997) equation, the σ_r parameter for the fluid of interest is related to σ_r for methane (1) and n-Octane (2), the reference fluids.

Table 1. Literature equations used in this work

Author	Equation
Brock and Bird (1955)	$\sigma P_c^{-\frac{2}{3}} T_c^{-\frac{1}{3}} = (0.132\alpha_c - 0.279)(1 - T_r)^{\frac{11}{9}}$ (4)

The model tuning was performed through an evaluation on an Intel® Core™ i7 11th processor with 16 GB of RAM, assessing model loss performance, R^2 coefficient, and time yielded during model execution on different ANN model execution. Fig. 2 provides an overview of the methodological approach employed in the current study.

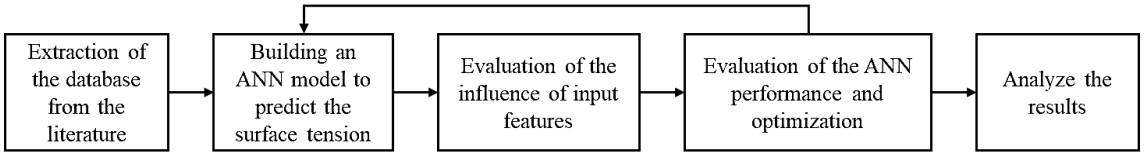

Fig. 2. Methodological approach workflow

Results and Discussions

Dataset was initially split into training, validation, and testing subsets, allocating 79% of the data for training, 11% for the validation, and 10% reserved for testing, ensuring unbiased evaluation and validation of the developed model. The first ANN developed employed an MLP architecture composed of seven hidden layers, with *ReLu* as the activation function and *Adam* as the optimizer algorithm. The model reliability and precision were tracked by MSE loss function, and 5-fold CV technique was applied to ensure model robustness and consistency. ANN specifications are presented on Table 2.

Table 2. ANN configuration parameters

Parameter	Value	Parameter	Value
Hidden layers	7	Epochs	1,000
Neurons per layer (1)	81	Batch Size	48
Neurons per layer (2 to 7)	97	Train and Test Random State	32
Learning Rate	0.001	CV Random State	42

The results obtained demonstrated a MSE value of 8.14×10^{-5} and R^2 score of 0.9766, validating the efficacy of chosen configuration. The performed CV assures the robustness of the method employed, in which was observed an average MSE value of 1.26×10^{-4} and an R^2 score of 0.9766 reflecting the performance achieved, assuring the absence of model overfitting. The Figs. 3 and 4 represent the R^2 score and the losses evolution, respectively, throughout the epochs of the evaluated model. By analyzing both Figures, it is possible to observe the convergence of the training and validation values, which leads to the absence of overfitting in the model.

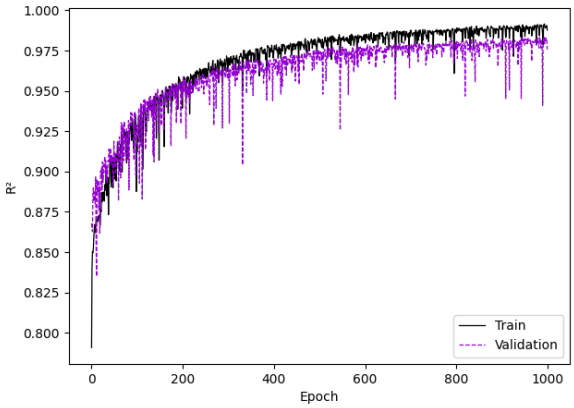

Fig. 3. R^2 score evolution across the epochs

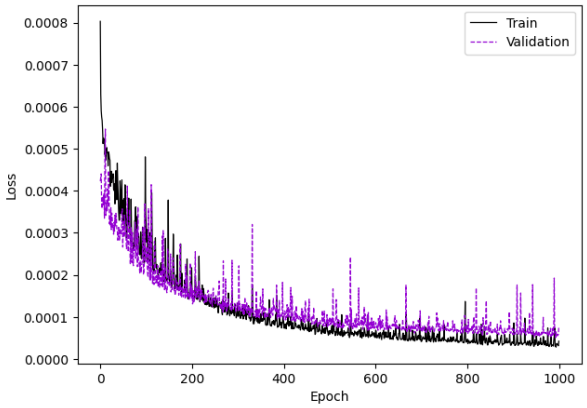

Fig. 4. Loss evolution across the epochs

According to Fig. 5, it is possible to observe that the fitting of the ANN through test dataset close matches the experimental data points, demonstrating a precise fit, reflecting the predictive accuracy and reliability achieved by the developed model along the different organic groups and components evaluated.

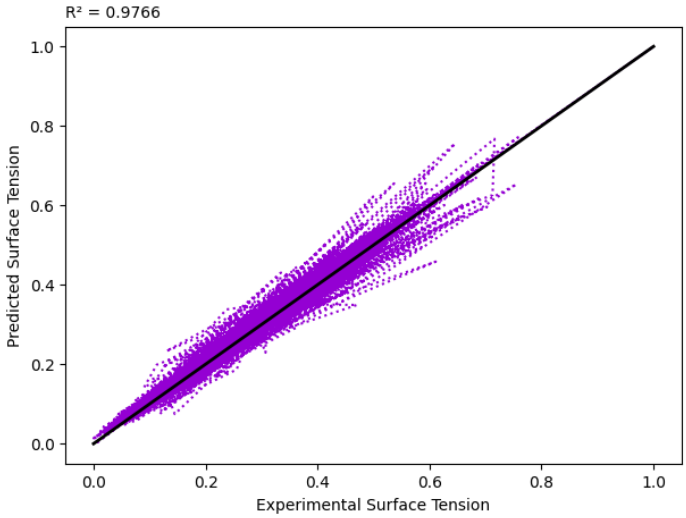

Fig. 5. Model evaluation through the test dataset

As observed in Table 3, the ARD results of the developed model overcome the results obtained by using an established literature equation, indicating substantial superior accuracy in fitting the evaluated data compared to the equation of Brock and Bird (1955).

Table 3. %ARD comparison entire data set

Model	%ARD
Present study	5.73
Brock and Bird (1955)	18.39

The developed model achieved comparable values to Lashkarbolooki and Bayat (2018) and Pazuki et al. (2011), despite it were slightly shorter than their results. Although, it is highlighted that the utilized dataset in the current study was higher than the above-mentioned studies, ensuring that the proposed model can accurately predicts the surface tension behavior.

To evaluate linear correlations between input features, the Pearson correlation coefficient (PCC) analysis was carried out, providing a standardized measurement of covariance. The heatmap representation observed in Fig. 6 illustrates the obtained results for PCC study. The *PairPlot* analysis was also conducted, aiming to explore potential non-linear behaviors among the features. In this regard, Fig. 7 presents a matrix of graphs that enables the visualization of relationships between each pair of variables analyzed. The diagonal represents the values distribution of each feature, and dashed lines were added to facilitate the comparison with the linear behavior.

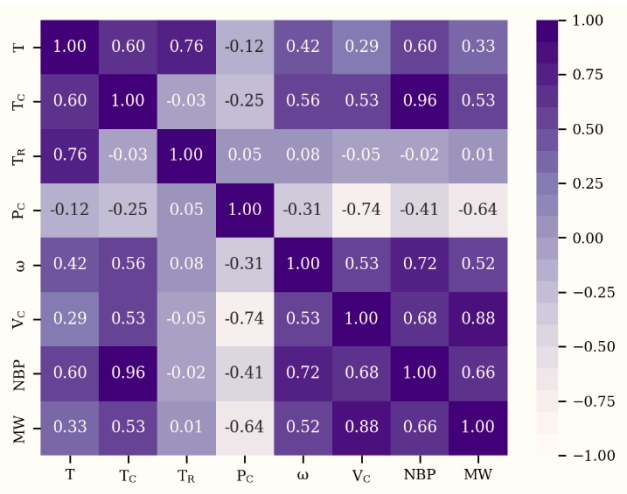

Fig. 6. Heat map of PCC analysis

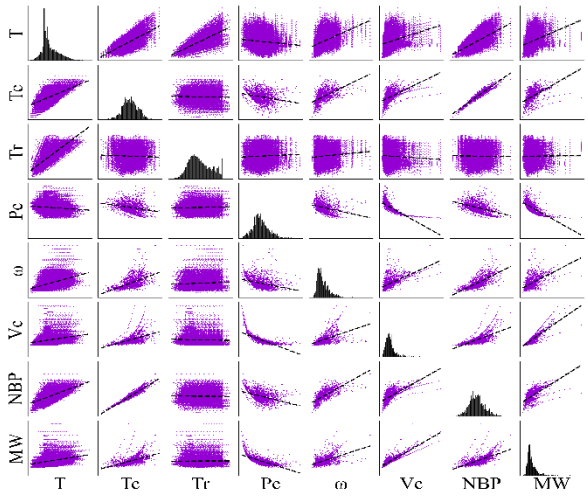

Fig. 7. PairPlot analysis

From the Fig. 6 analysis, it was observed that T_c and NBP properties exhibit a PCC value of 0.96, and V_c and MW properties exhibit a PCC value of 0.88, indicating a significant linear correlation between them, and potentially redundant information input to the model. From the analysis of Fig. 7, it was observed that the critical volume (V_c) exhibits a likely exponential correlation compared to the critical pressure (P_c) and critical temperature (T_c) properties, which is consistent with the thermodynamic aspects of these variables found in the literature, such as the correlations between them, as evidenced by Brock and Bird (1955).

Despite the possible redundancy between the evaluated input features evaluated, different architectures were assessed. However, the results obtained were not satisfactory compared to the first model, which best reproduces the experimental values of surface tension.

Conclusions

The developed study provided a broad analysis by evaluating the utilization of an ANN-based model to predict the surface tension behaviors of a wide number of pure organic components. The developed model demonstrated consistency by the cross-validation, with an average MSE value of $1.26x10^{-4}$. Further optimization of the model hyperparameters provided a reduction of the MSE value to $8.14x10^{-5}$, yielding an R^2 score of 0.9766, which demonstrates the consistency and accuracy of the developed model for the surface tension prediction of a large databank of pure organic components across different temperatures and pressures conditions. The PPC and PairPlot analyses allow to verify which features are linear correlated or present non-linear behaviors, providing insights for future research. Additionally, the developed ANN achieved an average relative deviation (ARD) of 5.73%, indicating consistent accuracy compared to the analyzed literature equation. Furthermore, despite the obtained results were similar but slightly shorter than the compared research, however the utilized data set in the current study was higher, making the developed model a valuable tool for predicting surface tension behaviors along a wide range of pure compounds.

Acknowledgments

The authors would like to acknowledge the institutions that supported the present study, namely the Brazilian National Agency of Petroleum, Natural Gas and Biofuels (ANP), the PRH-ANP, and the National Council for Scientific and Technological Development (CNPq). In addition, the authors would also like to acknowledge all those who contributed to the development of this work.

References

ADHIKARI, B.; HOWES, T.; SHRESTHA, A.; BHANDARI, B. R. Effect of surface tension and viscosity on the surface stickiness of carbohydrate and protein solutions. **Journal of Food Engineering**, v. 79, n. 4, p. 1136–1143, abr. 2007.

AHMED, T. **Equations of State and PVT Analysis – Applications for Improved Reservoir Modeling**. Houston, Texas, USA: Gulf Publishing Company, 2007.

BALL, D. W. **Physical Chemistry**. USA: Cengage Learning, 2011.

BHANDARI, B.; HOWES, T. Relating the Stickiness Property of Foods Undergoing Drying and Dried Products to their Surface Energetics. **Drying Technology**, v. 23, n. 4, p. 781–797, 1 abr. 2005.

BROCK, J. R.; BIRD, R. B. Surface tension and the principle of corresponding states. **AIChE Journal**, v. 1, n. 2, p. 174–177, 17 jun. 1955.

CEVADA, E.; ROOS, K.; ALVAREZ, F.; CARLOTTI, S.; VÁZQUEZ, F. High molar mass polyethers as defoamers of heavy crude oil. **Fuel**, v. 221, p. 447–454, jun. 2018.

CHEN, W.-H.; BISWAS, P. P.; UBANDO, A. T.; KWON, E. E.; LIN, K.-Y. A.; ONG, H. C. A review of hydrogen production optimization from the reforming of C1 and C2 alcohols via artificial neural networks. **Fuel**, v. 345, p. 128243, ago. 2023.

FATHI AZARBAYJANI, A.; JOUYBAN, A.; CHAN, S. Y. Impact of Surface Tension in Pharmaceutical Sciences. **Journal of Pharmacy & Pharmaceutical Sciences**, v. 12, n. 2, p. 218, 12 ago. 2009.

GHATEE, M. H.; ZARE, M.; ZOLGHADR, A. R.; MOOSAVI, F. Temperature dependence of viscosity and relation with the surface tension of ionic liquids. **Fluid Phase Equilibria**, v. 291, n. 2, p. 188–194, maio 2010.

LASHKARBOLOOKI, M.; BAYAT, M. Prediction of surface tension of liquid normal alkanes, 1-alkenes and cycloalkane using neural network. **Chemical Engineering Research and Design**, v. 137, p. 154–163, set. 2018.

MIQUEU, C.; MENDIBOURE, B.; GRACIAA, A.; LACHAISE, J. Petroleum mixtures: An efficient predictive method for surface tension estimations at reservoir conditions. **Fuel**, v. 87, n. 6, 2008.

MULERO, A.; CACHADIÑA, I.; BAUTISTA, D. Recommended Correlations for the Surface Tension of n-Alkanes. **Journal of Physical and Chemical Reference Data**, v. 50, n. 2, 1 jun. 2021.

NICOLA, G. DI; COCCIA, G.; PIERANTOZZI, M. Surface tension of silanes: A new equation. **Fluid Phase Equilibria**, v. 418, p. 88–93, jun. 2016.

PAZUKI, G. R.; NIKOOKAR, M.; SAHRANAVARD, L. Prediction of Surface Tension of Pure Hydrocarbons by An Artificial Neural Network System. **Petroleum Science and Technology**, v. 29, n. 22, p. 2384–2396, 19 set. 2011.

PITZER, K. S.; CURL, R. F. The Volumetric and Thermodynamic Properties of Fluids. III. Empirical Equation for the Second Virial Coefficient [1]. **Journal of the American Chemical Society**, v. 79, n. 10, p. 2369–2370, 1 maio 1957.

PUCKO, I.; RACAR, M.; FARAGUNA, F. Synthesis, characterization, and performance of alkyl methacrylates and tert-butylaminoethyl methacrylate tetra polymers as pour point depressants for diesel Influence of polymer composition and molecular weight. **Fuel**, v. 324, p. 124821, set. 2022.

RAMÍREZ-VERDUZCO, L. F.; ROMERO-MARTÍNEZ, A.; TREJO, A. Prediction of the surface tension, surface concentration, and the relative Gibbs adsorption isotherm of binary liquid systems. **Fluid Phase Equilibria**, v. 246, n. 1–2, p. 119–130, ago. 2006.

REIDEL, L. Eine neue universelle Dampfdruckformel Untersuchungen uber eine Erweiterung des Theorems der ubereinstimmenden Zustande. Teil I. **Chemie Tnq. Tech**, n. 2, 1954.

TIAN, J.; ZHANG, C.; ZHANG, L.; ZHENG, M.; LIU, S. Corresponding state-based correlations for the temperature-dependent surface tension of saturated hydrocarbons. **Modern Physics Letters B**, v. 31, n. 28, p. 1750259, 10 out. 2017.

ZUO, Y.; STENBY, E. H. Corresponding-states and parachor models for the calculation of interfacial tensions. **The Canadian Journal of Chemical Engineering**, v. 75, n. 6, p. 1130–1137, 27 dez. 1997.