19th Brazilian Congress of Thermal Sciences and Engineering
November 6th–10th, 2022, Bento Gonçalves - RS - Brazil

ENC-2022-0618
Predicting the heat of combustion of alcohols and carboxylic acids using particle swarm optimization

Bruno Roberto Pasa
Mikael Maraschin
Nina Paula Gonçalves Salau
Laboratório de Engenharia de Processos Assistida por Computador - LEPAC
Departamento de Engenharia Química - Universidade Federal de Santa Maria - UFSM
Prédio 9B – Centro de Tecnologia – Av. Roraima, n. 1000 – Camobi – Santa Maria, RS
bruno.pasa@acad.ufsm.br, maraschin.mikael@acad.ufsm.br, ninasalau@ufsm.br

Abstract. *The heat of combustion is the amount of energy released when one mole of a particular substance is completely burning. It is a key property for fuels and food. This work aims at predicting the heat of combustion of alcohols and carboxylic acids through a new proposed linear correlation and other five linear correlations found in literature. In this investigation, a dataset with 67 substances composed of 34 alcohols and 33 carboxylic acids was collected from previous works to fit all the correlations for heat of combustion. The estimated parameters as well as their respective elliptic confidence intervals were obtained using the iterative method particle swarm optimization (PSO), by considering each individual solution of PSO that satisfied a 95% of confidence interval for the joint confidence region of parameters. When compared to the other five correlations found in the literature, the new proposed correlation to determine the heat combustion showed the highest performance (coefficient of determination (R^2) of 0.997 and the sum square error (SSE) of* 1.198×10^5.

Keywords: *heat of combustion, correlation, particle swarm optimization*

1. INTRODUCTION

The heat of combustion is the energy released when a substance reacts in complete oxidation with oxygen, resulting in products, which are no more able to be oxidized. If the reaction occurs at 298.15 K, 1 bar, and constant pressure, the heat of combustion is equivalent to the standard enthalpy of combustion. The high heating value (HHV) is the amount of energy released when combustion takes place under certain conditions, or say more precisely when the water is present as liquid (Meyers, 2002).

The heating value is expressed as gross and net calorific value, depending on the status of water present in the exhaust. If water is present as liquid, then the heating value is called gross calorific value. If water is present as vapor, then the heating value is called net calorific value.

The importance of the determination of HHV covers a lot of fields. The knowledge of this measure is necessary from when you are dealing with fuels to food. For fuels, the use of HHV is mainly to find the most economical fuel to use since the greater the amount of heat produced by the fuel given a certain cost, the greater the cost-benefit of the fuel (Duchowicz *et al.*, 2007). In the food sector, it is used to express the energy content of nutrients in a diet or the energy necessary for the metabolism of foods, such as fats and sugars. Another wide application of HHV is the comparison of the stability of chemical compounds. When two isomeric hydrocarbons burn in equal amounts to produce the same amount of carbon dioxide and water, the one that releases more energy is the less stable because it is the more energetic in its compounded form (Gharagheizi, 2008).

High heating value is commonly determined by the use of an oxygen bomb calorimeter (Shen *et al.*, 2012), which demands time and costs. Another approach to determine the heat of combustion is through quantum chemistry calculations (Canneaux *et al.*, 2014), which demands a high computational cost. Thus, the most widely used approach to determine the heat of combustion is through linear correlations. One of the first linear correlations is the formula proposed by Dulong, Eq. 1, which takes into account the mass percentage of each element (Carbon, Hydrogen, and Oxygen) (Selvig, 1945), after the formulation of the Dulong equation, several modified versions of the equation appeared. One of these is the one proposed by Steuer (1926), which proposes a modified equation to calculate the calorific value of solid fossil fuels from elemental analysis, expressed in Eq. 2, others modified versions of Dulong's equation can be found in the works of D'Huart (1930); Strache and Lant (1924). But, there are many more works that predict the heat of combustion from the number of atoms of the compound. Like the correlation proposed by Boie (1953), Eq. 3, which is a correlation derived from the properties of hydrocarbon fuels, the one proposed by Elliott *et al.* (1990), Eq. 4, which was derived for oils derived from biomass and Gumz (1938), Eq. 5, which was derived for coals.

$$HHV = 0.3383C\% + 1.4443(H\% - (O\%/8)) + 0.0942S\% \left[\frac{MJ}{kg}\right] \tag{1}$$

$$HHV = 0.3391(C\% - ((3/8)O\%)) + 0.2386((3/8)O\%) + 1.444(H\% - ((1/16)O\%)) + 0.1047S\% \left[\frac{MJ}{kg}\right] \tag{2}$$

$$HHV = 0.3517C\% + 1.1626H\% + 0.1046S\% + 0.111O\% \left[\frac{MJ}{kg}\right] \tag{3}$$

$$HHV = 0.352C\% + 0.944H\% + 0.105(S\% - O\%) \left[\frac{MJ}{kg}\right] \tag{4}$$

$$HHV = 0.3403C\% + 1.2432H\% + 0.0628N\% + 0.1909S\% - 0.0984O\% \left[\frac{MJ}{kg}\right] \tag{5}$$

It's important to realize that a successful parameter estimating must supply both estimated parameter values and uncertainty, and this uncertainty is evaluated through the confidence intervals (Kelley and Lai, 2011). However, the correlation proposed in this work has three and four parameters, so evaluation of parameter uncertainties must preferably be performed with help of confidence regions (Tolazzi *et al.*, 2018). The importance of the uncertain is to infer whether the parameters are significant or not.

The parameter values that preserve the statistical meaning of the model fit to the available experimental data are found in the confidence region of the estimated model parameters (Tolazzi *et al.*, 2018). The confidence region of parameter estimates defines a hyper-ellipsoid in the parameters space, with the point estimate of model parameters put at the center of the hyper-ellipsoid, assuming that variances between expected and experimental data follow the normal probability distribution (Draper and Smith, 1998).

The main goal of this paper is to propose a linear correlation for HHV using the number of carbon, hydrogen, oxygen and nitrogen atoms in the molecules and respective uncertainties of the model through the analysis of shape of join confidence region of parameters obtained with particle swarm optimization (PSO).

2. Methodology

2.1 Data

The data of heat of combustion was obtained from (Domalski, 1972), which has a total of 740 chemical substances with the heat of combustion and entropy of formation evaluated in standard condition for thermodynamic experiments, the temperature of 298.15 K and absolute pressure of 1 bar, but just the substance that are alcohol and carboxylic acids were used in this work, resulting in 66 chemical substance. These chemical substances were composed of carbon, hydrogen, oxygen, and nitrogen.

Figure 1 present the plot of the dependent variable versus the independent ones. By inspecting the relation of each atom with HHV it is possible to notice high correlation between independent variables (C, H, O, N) and the dependent one (heat of combustion), mainly concerning carbon and hydrogen.

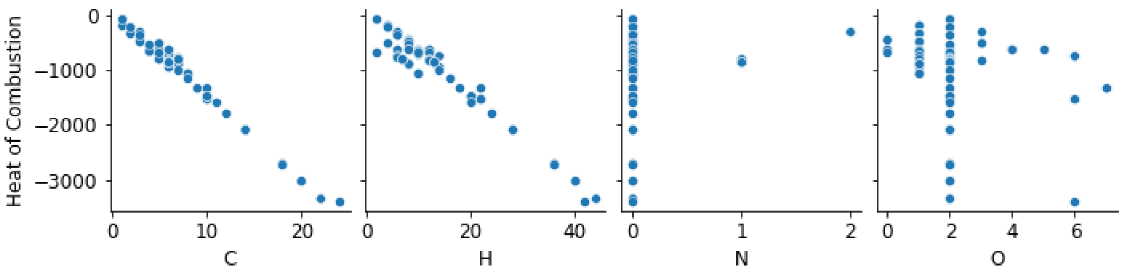

Figure 1. Data correlating heat of combustion and elemental composition of the samples

2.2 Parameters estimated value and confidence interval

The Eq. 6 is the structure of the new proposed linear correlation. The parameters to be estimated, β_C, β_H, β_O, β_N, β_0, are, respectively, the coefficients that multiply the number of carbon atoms, the number of hydrogen atoms, the number of oxygen atoms and the number of nitrogen atoms.

$$HHV = \beta_C \cdot C + \beta_H \cdot H + \beta_O \cdot O + \beta_N \cdot N + \beta_0 \tag{6}$$

To estimate the parameters of linear regression, the minimization of the sum square error was used as an objective function, illustrated in the equation 7. Where β represents a vector of model parameters, *NE* is the number of experiments, f is the function that relates the model parameters and independents variables with predicted values and y_i^e is the experimental data.

$$SSE(\beta) = \sum_{i=1}^{NE} (f(\beta, x_i) - y_i^e)^2 \tag{7}$$

The analytical solution of the general case for a linear model is demonstrated with a linear algebra representation. Considering a $\overrightarrow{Y}$ a vector of observations (n x 1), $\overrightarrow{X}$ a matrix of independent variable (n x p), $\overrightarrow{\beta}$ a vector of observations (p x 1). The SSE could be rewritten as equation 8.

$$SSE(\beta) = \left(\overrightarrow{Y} - \overrightarrow{X}\overrightarrow{\beta}\right)^T \left(\overrightarrow{Y} - \overrightarrow{X}\overrightarrow{\beta}\right) = \overrightarrow{Y}^T\overrightarrow{Y} - 2\overrightarrow{\beta}^T\overrightarrow{X}^T\overrightarrow{Y} + \overrightarrow{\beta}^T\overrightarrow{X}^T\overrightarrow{X}\overrightarrow{\beta} \tag{8}$$

After differentiating the Eq. 8 with respect to $\overrightarrow{\beta}$, the result was set to zero to find a point of minimum. Resulting in the Eq. 9, finally the $\hat{\beta}$, which denotes the estimated parameters at the point of minimum in the objective function is obtained by solving equation 10.

$$\overrightarrow{X}^T\overrightarrow{X}\overrightarrow{\beta} = \overrightarrow{X}^T\overrightarrow{Y} \tag{9}$$

$$\hat{\beta} = \left(\overrightarrow{X}^T\overrightarrow{X}\right)^{-1}\overrightarrow{X}^T\overrightarrow{Y} \tag{10}$$

With an estimated value of parameters, it is possible to calculate the confidence interval of parameters. Which is important to determine whether each parameter is correct or prone to error, enable the estimation of confidence intervals of dependent variables, and also the significance or not for the parameter for the model.

This work made the construction of the join confidence regions of the parameters, for that was utilized the Equation 11, which was proposed in (Beale, 1960). Where $SSE(\beta)$ is the objective function, described in the linear regression section, p is the number of parameters, n is the total number of data points and F represents the fisher distribution. But, when experimental errors fluctuate according to an arbitrary probability distribution, Eq. 11 should be rewritten as the equation 12, where c is a constant that depends on the required confidence level.

$$SSE(\beta) \geqslant SSE(\hat{\beta})(1 + \frac{p}{n-p}F_{1-\alpha}^{p,n-p}) \tag{11}$$

$$SSE(\beta) \geqslant cSSE(\hat{\beta}) \tag{12}$$

Various methods can be used for objective function minimization, such as genetic algorithms (Holland, 1992), simulated annealing (Van Laarhoven and Aarts, 1987) and particle swarm optimizer (Kennedy and Eberhart, 1995). In this work, is used the PSO, because, as described in Schwaab (2005), PSO outperforms the other methods, particularly in parameter estimation problems.

The increased number of objective function evaluations necessitates longer calculation times when compared to standard approaches, which is a downside of employing this method. However, as described in Schwaab *et al.* (2008), by selecting points that satisfy the equation along the search path, this large number of iterations can be used to calculate the join confidence region. The confidence regions of the parameters are created by plotting these points.

For the cited reasons the PSO algorithm was chosen to estimate the parameters value and at same time the confidence region. Particle swarm optimization runs were performed with tolerance of $1 \cdot 10^{-4}$ to accept a convergence for the solution, maximum number of iteration of 1000 iteration and 4000 particles in the swarm, using the *particleswarm* function provided by the Matlab software.

2.3 Particle Swarm Optimization

Particle swarm optimization was developed by Kennedy and Eberhart (1995), is a meta-heuristic global optimization method, which belongs to the family of algorithms based on the concept of swarm intelligence. In analogy to the behavior of bird flocks and fish schools, in PSO the set of candidate solutions to the optimization problem is defined as a swarm of particles that move along the search space and exchange information with other particles, by the equation 13 and equation 14. In image 2 is illustrated how the PSO works, first, start with a random position with a random direction, represent in the image a, after a few iterations the particle starts heading towards the optimal point, as shown in images b and c, and after a few more iterations the points converge to the best solution, demonstrated in image d.

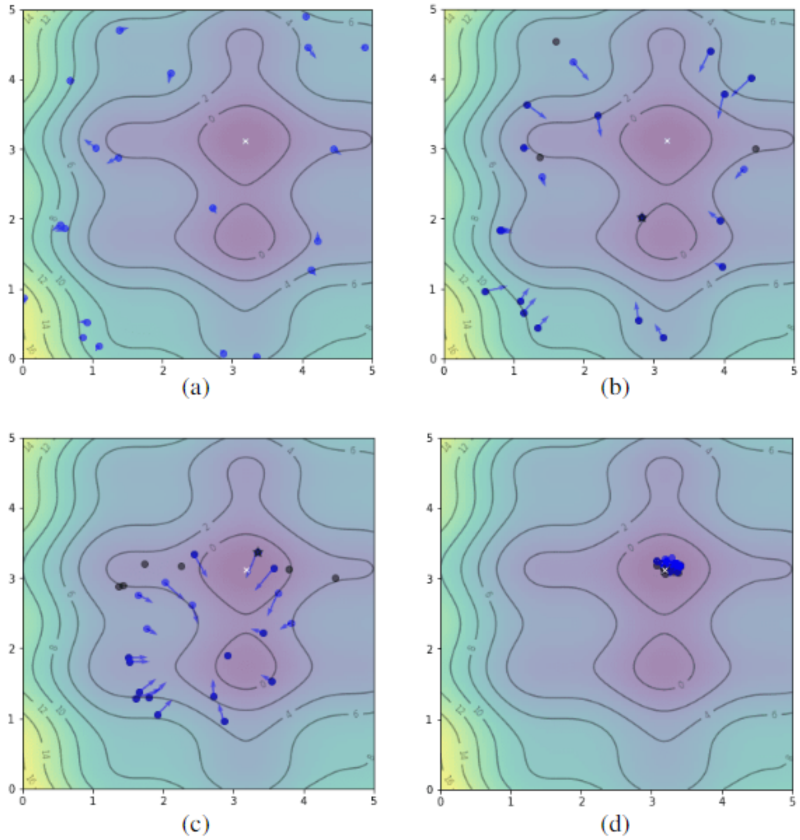

Figure 2. Particle movement of a PSO method. Font: https://machinelearningmastery.com/a-gentle-introduction-to-particle-swarm-optimization/

$$v_{p,d}^{k+1} = wv_{p,d}^{k} + c_1 r_1 (x_{p,d}^{ind} - x_{p,d}^{k}) + c_2 r_2 (x_{d}^{glo} - x_{p,d}^{k}) \quad (13)$$

$$x_{p,d}^{k+1} = x_{p,d}^{k} + v_{p,d}^{k+1} \quad (14)$$

Where p denotes the particle, d is the search direction, k represents the iteration number, v is the velocity (or pseudo-velocity) of the particle and x is the position of the particle. The search space where the objective function attains low values are represented by the x_{ind} and the x_{glo}. Where the x_{ind} is the best position found by the particle itself, while the x_{glo} is the best position found by the whole swarm. The r_1 and the r_2 are two random numbers with uniform distribution in the range [0, 1]. The search parameters are w, c_1 and c_2. The parameters c_1 and c_2 are the cognition and the social parameters. The parameter w is called inertial weight and was not present in the original form of the algorithm.

2.4 Statistical Parameters

To verify if the proposed model was properly fitted to the experimental data, 2 statistical parameters as performance criteria were applied: the coefficient of determination (R^2) and the sum of squared error (SSE). These parameters statistics are calculated according to the equations 15 and 16, where v_{est} is the value estimated by the model, v_{exp} is the experimental value and m_{exp} is the average of the experimental values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(v_{exp} - v_{est})^2}{\sum_{i=1}^{n}(v_{exp} - m_{exp})^2} \quad (15)$$

$$SSE = \sum_{i=1}^{n} (v_{exp} - v_{est})^2 \tag{16}$$

3. RESULTS AND DISCUSSION

Table 1 presents the fitted value of parameters for the new proposed linear correlation and the respective statistical evaluation for each of estimated values with R^2 of 0.997 and SSE of $1.198 \cdot 10^5$. The Eq. 17 is the new proposed linear correlation, achieved by replacing the values of the estimated parameters, that are illustrated in the second column of the table 1, in the Eq. 6. It is possible to accept a null hypothesis for the coefficient β_N, by considering the p-value for β_N as a non-statistically significant difference from zero. This lack of significance occurred only to nitrogen contribution in the heat of combustion, it does not repeat in the other coefficients. The absence of significance to nitrogen parameters in the correlation was already expected because most of the data set is composed of substances without nitrogen in the molecular composition.

Table 1. Parameters of the linear correlation.

Coefficient	Estimated Value	Standard Error	t value	p value
β_C	-0.458	0.013	-34.943	$2.200 \cdot 10^{-16}$
β_H	-0.0969	0.006	-14.836	$2.200 \cdot 10^{-16}$
β_O	0.118	0.017	6.838	$4.090 \cdot 10^{-9}$
β_N	0.0835	0.059	1.418	0.161
β_0	0.165	0.046	3.276	0.002

$$HHV = -0.458C - 0.0969H + 0.118O + 0.0835N + 0.165 \left[\frac{MJ}{mol}\right] \tag{17}$$

After removing the number of nitrogen and the respective coefficient, Table 2 presents the estimated value for only statistically significant parameters, that said, the Eq. 18 refers to the linear correlation proposed in this work, just with the significant parameters. By analyzing the predicted values, it is showed a negative contribution of carbon and hydrogen, however, the presence of oxygen provides a positive contribution, this is expected because the presence of O-C bonds in alcohols and O-C and C=O bonds are molecular more oxidized than compared with hydrocarbons in the absence of these bonds.

Table 2. Significant parameters of the linear correlation.

Coefficient	Estimated Value	Standard Error	t value	p value
β_C	-0.457	0.013	-34.638	$2.200 \cdot 10^{-16}$
β_H	-0.0969	0.006	-14.949	$2.200 \cdot 10^{-16}$
β_O	0.117	0.017	6.964	$2.307 \cdot 10^{-9}$
β_0	0.164	0.046	3.600	0.001

$$HHV = -0.457C - 0.0969H + 0.117O + 0.164 \left[\frac{MJ}{mol}\right] \tag{18}$$

The PSO approach to predict the elliptical confidence interval is presented in figure 3, this figure shows the possible binary parameter plots, in which the β_C with β_H presented an expected ellipse with a well-defined shape. Since, the shape of the joint confidence region could be interpreted as a qualitative measurement of the quality and linearity of parameters. In other words, weakly nonlinear relations will present slight deviations from elliptical shape, instead, strongly nonlinear relations will present obliqueness and asymmetric shapes.

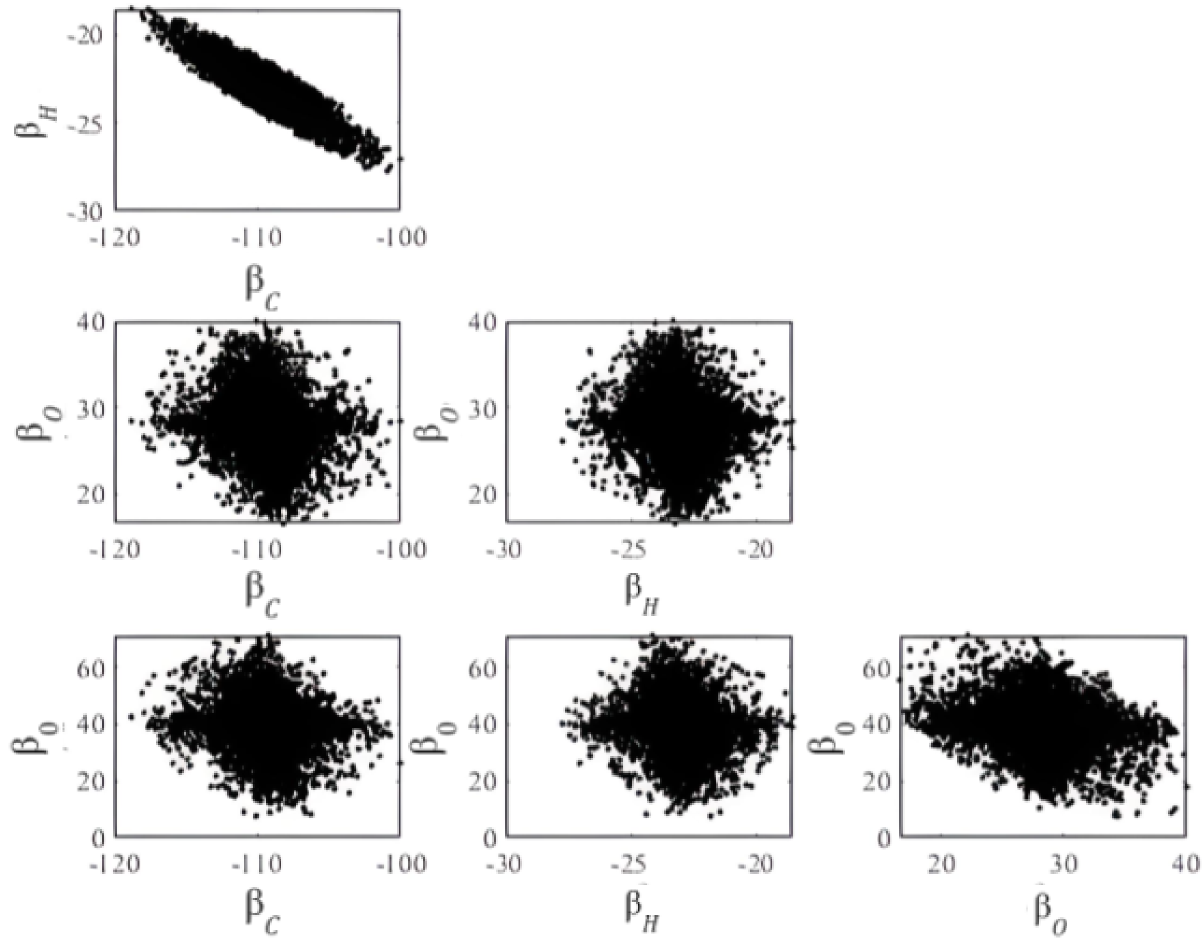

Figure 3. Join confidence region of parameters obtained through particle swarm optimization

Considering the parameters β_O and β_0 the shape of the joint confidence region deviates from a characteristic elliptical shape as was shown in β_C with β_H, however, the shapes did not have obliqueness or asymmetric relations, in this sense the produced joint confidence region by these parameters are elliptical with small differences in minor and major diameters. For this reason, from a qualitative point of view, there could be a weakly nonlinear contribution of oxygen to the heat of combustion and by excluding the relations of nitrogen atoms in the correlation it is possible to affect the quality of bias in the correlation even if the value still be statistically significant.

To gain a better insight into the prediction success of the correlation, figure 4 shows the actual value plotted against the predicted values. It is possible to see that the values predicted to match the experimental values very well.

The new proposed correlation was compared to 5 correlations found in the literature. The table 3 summarizes these comparisons performed. As can be seen, the model that obtained the highest performance in both R^2 and SSE is the model proposed in this article. However, it is important to note that all correlations were developed to predict the HHV of specific substances, such as the one from Elliott *et al.* (1990) which was developed for oils derived from biomass. That said, these models should have superior performance when done with the right type of substance. So the reason the model proposed in this article performed better is due to the dataset used, with just alcohols and carboxylic acids.

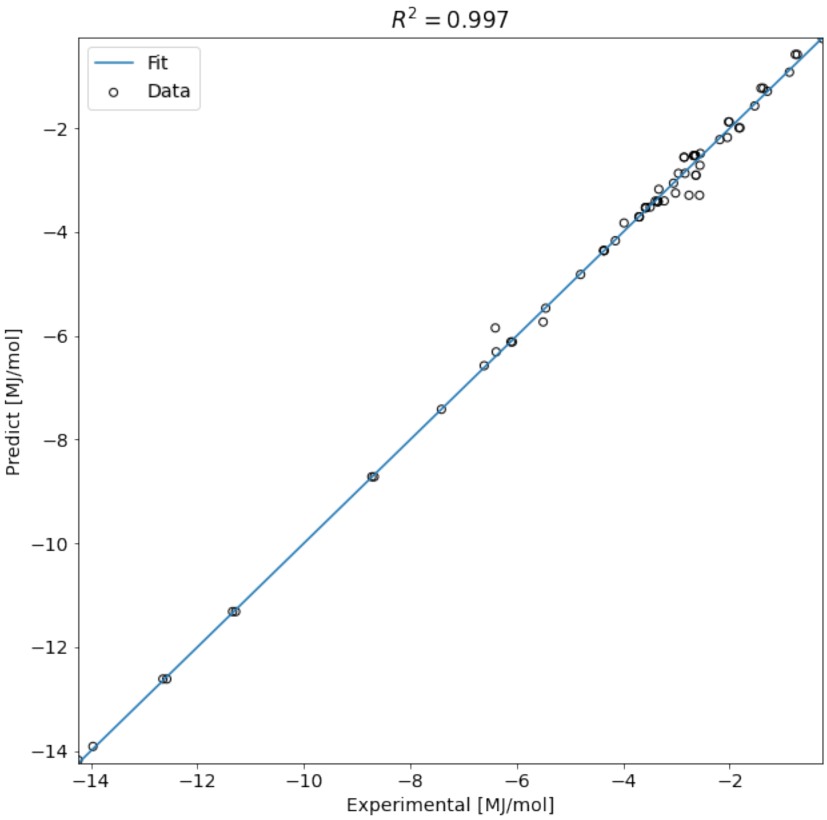

Figure 4. Linear model adjust to data

Table 3. Comparison with correlation from literature.

Correlation	R^2	SSE
Dulong's Equation	0.784	$8.246 \cdot 10^6$
Steuer (1926)	0.983	$6.441 \cdot 10^5$
Gumz (1938)	0.994	$2.200 \cdot 10^5$
Boie (1953)	0.922	$2.967 \cdot 10^6$
Elliott *et al.* (1990)	0.986	$5.144 \cdot 10^5$
New proposed	0.997	$1.198 \cdot 10^5$

4. CONCLUSION

The focus of this study is the application of a new proposed linear correlation for the estimation of the higher heating value of alcohols and carboxylic acids and validate the linearity of the parameters through the elliptical confidence region. For this prediction, the number of carbon, hydrogen, oxygen, and nitrogen atoms of each substance was used as input. For this, a dataset containing 67 data was used. The main conclusions of the work are mentioned below:

- The linear model was able to interpret the data very well, obtaining good statistical parameters, reaching an R^2 of 0.997 and SSE of 1.198×10^5. That was already expected by the high correlation between the data, shown in the correlation matrix.
- Through a statistical analysis of the p-value, it was found that nitrogen has no significance for predicting the HHV of the data used. This is because most substances used at work do not contain nitrogen.
- With the confidence interval of the parameters, a strong linear contribution of carbon and hydrogen was evident, due to the elliptical format of the interval.

- And also with the confidence interval of the parameters, it was able to interpret that oxygen has a weak nonlinear contribution to the HHV.

That said, the correlation proposed in this article can be used as a strong tool for estimating the HHV of alcohols and carboxylic acids.

5. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the FAPERGS and the CAPES for awarding the scholarships of the first and second authors, respectively.

6. REFERENCES

Beale, E., 1960. "Confidence regions in non-linear estimation". *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 22, No. 1, pp. 41–76.

Boie, W., 1953. "Fuel technology calculations". *Energietechnik*, Vol. 3, pp. 309–316.

Canneaux, S., Bohr, F. and Henon, E., 2014. "Kisthelp: A program to predict thermodynamic properties and rate constants from quantum chemistry results". *Journal of computational chemistry*, Vol. 35, No. 1, pp. 82–93.

Domalski, E.S., 1972. "Selected values of heats of combustion and heats of formation of organic compounds containing the elements c, h, n, o, p, and s." *Journal of Physical and Chemical Reference Data*, Vol. 1, No. 2, pp. 221–277.

Draper, N.R. and Smith, H., 1998. *Applied regression analysis*, Vol. 326. John Wiley & Sons.

Duchowicz, P.R., Garro, J.C., Andrada, M.F., Castro, E.A. and Fernández, F.M., 2007. "Qspr modeling of heats of combustion for carboxylic acids". *QSAR & Combinatorial Science*, Vol. 26, No. 5, pp. 647–652.

D'Huart, K., 1930. "The development of formulas for determining the heating value of solid fuels". *Die Warme*, Vol. 53, pp. 13–17.

Elliott, D., Baker, E., Beckman, D., Solantausta, Y., Tolenhiemo, V., Gevert, S., Hörnell, C., Östman, A. and Kjellström, B., 1990. "Technoeconomic assessment of direct biomass liquefaction to transportation fuels". *Biomass*, Vol. 22, No. 1-4, pp. 251–269.

Gharagheizi, F., 2008. "A simple equation for prediction of net heat of combustion of pure chemicals". *Chemometrics and Intelligent Laboratory Systems*, Vol. 91, No. 2, pp. 177–180.

Gumz, W., 1938. "Feuerungstech". In *Chemical Abstract*. Vol. 26, pp. 322–323.

Holland, J.H., 1992. "Genetic algorithms". *Scientific american*, Vol. 267, No. 1, pp. 66–73.

Kelley, K. and Lai, K., 2011. "Accuracy in parameter estimation for the root mean square error of approximation: Sample size planning for narrow confidence intervals". *Multivariate Behavioral Research*, Vol. 46, No. 1, pp. 1–32.

Kennedy, J. and Eberhart, R., 1995. "Particle swarm optimization". In *Proceedings of ICNN'95-international conference on neural networks*. IEEE, Vol. 4, pp. 1942–1948.

Meyers, R.A., 2002. *Encyclopedia of physical science and technology*. Academic.

Schwaab, M., 2005. "Evaluation of heuristic optimization algorithms in parameter estimation problems". *UFRJ*.

Schwaab, M., Biscaia Jr, E.C., Monteiro, J.L. and Pinto, J.C., 2008. "Nonlinear parameter estimation through particle swarm optimization". *Chemical Engineering Science*, Vol. 63, No. 6, pp. 1542–1552.

Selvig, W., 1945. "Calorific value of coal." *Chemistry of coal utilization*, pp. 132–144.

Shen, J., Zhu, S., Liu, X., Zhang, H. and Tan, J., 2012. "Measurement of heating value of rice husk by using oxygen bomb calorimeter with benzoic acid as combustion adjuvant". *Energy Procedia*, Vol. 17, pp. 208–213.

Steuer, W., 1926. "Allgemeine formel zur berechnunbdes heizwertes von festen fossilenbrennstoffen aus des elementaranalyse". *Brennstoff-Chem*, Vol. 7, pp. 344–347.

Strache, H. and Lant, R., 1924. *Kohlenchemie: Entstehg u. chem. Verhalten d. Kohlen u. ihrer Bestandteile*. Akadem. Verlagsgesellschaft.

Tolazzi, N., Steffani, E., Barbosa-Coutinho, E., Júnior, J.B.S., Pinto, J.C. and Schwaab, M., 2018. "Adsorption equilibrium models: computation of confidence regions of parameter estimates". *Chemical Engineering Research and Design*, Vol. 138, pp. 144–157.

Van Laarhoven, P.J. and Aarts, E.H., 1987. "Simulated annealing". In *Simulated annealing: Theory and applications*, Springer, pp. 7–15.

7. RESPONSIBILITY NOTICE