

## Research Article

# On the Investigation of Effective Factors on Higher Heating Value of Biodiesel: Robust Modeling and Data Assessments

Shicheng Wang <sup>1</sup>, Wei Li <sup>1,2</sup> and Issam Alruyemi <sup>3</sup>

<sup>1</sup>School of Economics and Management, Southwest Petroleum University, Chengdu, Sichuan 610500, China

<sup>2</sup>Petroleum Engineering School, Southwest Petroleum University, Chengdu, Sichuan 610500, China

<sup>3</sup>Fouman Faculty of Engineering, College of Engineering, University of Tehran, Fouman, Iran

Correspondence should be addressed to Shicheng Wang; [shicheng\\_wang@hotmail.com](mailto:shicheng_wang@hotmail.com) and Issam Alruyemi; [essamkhudur@ut.ac.ir](mailto:essamkhudur@ut.ac.ir)

Received 15 June 2021; Revised 22 June 2021; Accepted 1 July 2021; Published 12 July 2021

Academic Editor: Alireza Baghban

Copyright © 2021 Shicheng Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Higher heating value (HHV) is one of the properties of biomass fuels which is essential in investigating their special characteristics and potentialities. In this paper, various techniques based on Gaussian process regression (GPR) were utilized to assess this value for biomass fuels, including several kernel functions, i.e., exponential, Matern, rational quadratic, and squared exponential. An extensive databank was collected from literature. The findings were compared, and the results indicated that Exponential-based model was more accurate, with the coefficient of regression ( $R^2$ ) of 0.961 and the mean relative error (% MRE) of 3.11 for total data. Compared to former models presented by previous researchers, the model proposed in this study showed a higher ability to predict output values. With various analyses, it can be concluded that the proposed method has a high rate of efficiency in assessing the HHV of various biomass.

## 1. Introduction

The use of fossil fuels has problems and disadvantages such as environmental pollution, asphaltene deposition, and limited resources [1–3]. There have been growing attempts at reducing the use of conventional fossil fuels and finding suitable replacements to use in a world with an ever-increasing population and industrial expansion, a compromised environment, and steadily depleting energy sources. Among these alternatives, biomass has become of particular interest due to its carbon neutrality and ease in being processed (e.g., chemically, thermally, and biochemically) to produce energy [4]. In recent years, coal-fired power stations have turned to use biomass to replace part of their fuel. This way, without needing to change any of their equipment, they can lower their use of coal and thus contribute to environmental and economic prosperity [5, 6].

Characteristics of biomass fuel, before being incorporated as a renewable source of energy, must be fully identified. Among these characteristics, the higher heating value is fundamental for allocating the feedstock for specific uses. The conventional method of measuring the HHV for liquid and solid fuel is adiabatic oxygen bomb calorimetry, which is, however, expensive and inefficient [7]. There are two methods of finding correlations for HHV: ultimate and proximate. The former is capable of identifying the composition of the fuel and its elements [8] but is more expensive than the latter method and cannot function without specific prior experiments. This has led to the widespread use of the proximate method of analysis [9]. This method works by first determining the changes in the enthalpy of products and reactants of a specific type of fuel. The procedure is not complicated but it takes a long time and requires equipment that might sometimes be unavailable. As a result, calculations are

made by other empirical methods using the data from proximate or elemental analysis. Proximate analysis, which is simpler and faster, has more widespread use for measuring the HHV. From the gathered data, fixed carbon (FC), volatile matter (VM), and ash are the factors incorporated in calculations [10, 11].

In previous years, the use of artificial intelligence (AI) methods has many applications in various fields, and researchers have investigated complete and close analyses to develop empirical methods (which mostly involve linear and nonlinear models) to reliably approximate the HHV of different types of biomass fuel [12]. Despite the efforts to estimate biomass, the complications associated with its structure make understanding the relationship between HHV and the data from proximate or ultimate analysis problematic. As a result, attention has recently been turned to artificial intelligence and its high potential to solve complicated problems. Mesroghli et al. utilized ANN models to assess the HHV of coal [13]. Ghugare et al. assessed the HHV of solid biomass fuel utilizing MPL-ANN and GA-based models and used ultimate analysis to find correlation [14]. Another attempt at estimating HHV of biomass was undertaken by Hosseinpour et al. [15] using iterative neural network-adapted partial least squares. The data gathered by the proximate analysis were incorporated into an ANFIS model by Akkaya to estimate the heating value (HV) of biomass [16]. Uzun et al. experimented with various ANN structures to estimate the HHV of biomass [17]. Finally, Estiati et al. utilized ANN together with a few linear models [18].

The present study involves expanding models of estimating the HHV for biomass fuels to replace the ultimate analysis with the proximate analysis, which is both cheaper and faster. Innovative models are introduced based on Gaussian process regression modeling including four kernel functions, i.e., exponential, Matern, rational quadratic, and squared exponential. To design the models, the data regarding the HHV of various biomass were gathered from 382 studies. A comparison is drawn of these models with those studied and published in the past. The new models were further studied for their efficacy and usefulness in six types of biomass fuel.

## 2. Materials and Methods

**2.1. Data Collection.** The independent variables of volatile matter (VM), ash (A), and fixed carbon (FC) content on dry basis are the inputs in the present study. The output is the data regarding the HHV of biomass. Here, the aim is to find the most practical  $y$  or function  $f$  for the input data  $x_1, x_2, x_3$ , i.e., FC, VM, and A, and  $y$  or function  $f$  indicates the HHV of biomass fuels.

The data from 382 proximate analyses regarding biomass and their HHVs were gathered from open literature. The data collected have been reported elsewhere [19]. The data regarding HHV were categorized into the following six groups:

- (1) Byproducts of fruits

- (2) Agri-wastes
- (3) Wood chips and/or tree species
- (4) Grasses, leaves, and fibrous materials
- (5) Other waste materials
- (6) Briquettes, charcoals, and pellets

Learning from literature, 30% of the data were randomly set apart as a test set to prevent overtraining [20]. Designing and training the nonlinear regression and AI models were performed using the remaining 70%. The test dataset helped examine the precision of the results and generalize the newly proposed models.

**2.2. Gaussian Process Regression.** To establish Gaussian process regression (GPR), it is required to select random training dataset  $L = \{x_{L,i}, Y_{L,i}\}_{i=1}^n$  and testing dataset  $T = \{x_{T,i}, Y_{T,i}\}_{i=1}^n$  from a particular distribution. The training dataset is employed to set the tuning parameters of the model [21, 22]. The testing dataset, which includes the excluded observations of the previous stage, is utilized to perform the approximate justification of the extended model. Also,  $x$  is the input variable, while  $y$  denotes the target variable. They are impacted by noise. The general form of GPR modeling is formulated as [22]:

$$y_{L,i} = f(x_{L,i}) + \varepsilon_{L,i}, \quad i = 1.2.3 \dots \dots n, \quad (1)$$

in which  $x_L$  is the independent variable of the learning dataset,  $y_L$  is the learning dataset target, and  $\varepsilon \sim N(0, \sigma_{\text{noise}}^2 I_n)$  represents the observation noise of an independent Gaussian distribution (where  $\sigma_{\text{noise}}^2$  stand for the noise variance, while  $I_{\text{noise}}$  represent the unit array variance). Then, the measured targets are connected to the function  $f(x)$  by using a Gaussian noise model [23, 24]. It is worth mentioning that  $f$  values are assumed to be random variables in the GP. Likewise,

$$y_{T,i} = f(x_{T,i}) + \varepsilon_{T,i}, \quad i = 1.2.3 \dots \dots n, \quad (2)$$

in which  $x_T$  is the testing dataset independent variable, while  $y_T$  is the testing dataset target. Also,  $f(x)$  is a latent parameter and has a GP distribution with a mean of  $m(x)$  and covariance of  $k(x, x')$  [23].

$$f(x_{T,i}) \sim Gp\left(m(x), K\left(x, x'\right)\right). \quad (3)$$

To specify the mean function  $m(x)$ , one can utilize an explicit basis function, even though it would lead to a complex specification of a fixed  $m(x)$ . To simplify the calculations, one can let  $m(x)$  be zero [25–27]:

$$f(x_{L,i}) \sim Gp\left(0, K\left(x, x'\right)\right). \quad (4)$$

One can combine Equation (1) and Equation (4) to obtain

the prior distribution of  $y$  [25]:

$$y \sim N\left(0, K(x, x') + \sigma_{\text{noise}}^2 I_n\right). \quad (5)$$

The above equations could be collected as [27]:

$$\begin{bmatrix} \overrightarrow{f_l} \\ \overrightarrow{f_T} \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(x_L, x_L) & K(x_L, x_T) \\ K(x_T, x_L) & K(x_T, x_T) \end{bmatrix}\right), \quad (6)$$

$$\begin{bmatrix} \overrightarrow{\varepsilon_l} \\ \overrightarrow{\varepsilon_T} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_{\text{noise}}^2 I_n & 0 \\ 0 & \sigma_{\text{noise}}^2 I_n \end{bmatrix}\right).$$

These equations can be summed up into a Gaussian formulation as [21]:

$$\begin{bmatrix} \overrightarrow{y_l} \\ \overrightarrow{y_T} \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(x_L - x_L) + \sigma_{\text{noise}}^2 I_n & K(x_L, x_T) \\ K(x_T, x_L) & K(x_T, x_T) + \sigma_{\text{noise}}^2 I_n \end{bmatrix}\right). \quad (7)$$

Then, the Gaussian conditioning rule could be applied to find the posterior distribution of  $y_T$ , [27]:

$$\left(\overrightarrow{y_T} \middle| \overrightarrow{y_L}\right) \sim N\left(\mu_T, \Sigma_T\right), \quad (8)$$

where the mean value and covariance are written as:

$$\begin{aligned} \mu_T &= m(\overrightarrow{y_T}) = K(x_T, x_L) (K(x_L, x_L) + \sigma_{\text{noise}}^2 I_n)^{-1} \overrightarrow{y_L}, \\ \Sigma_T &= K(x_T, x_T) K(x_T, x_T) + \sigma_{\text{noise}}^2 I_n - K(x_T, x_L) (K(x_L, x_L) + \sigma_{\text{noise}}^2 I_n)^{-1} K(x_T, x_L). \end{aligned} \quad (9)$$

The theoretical GPR modeling concept is implemented. It is possible to predict the testing dataset outputs through the independent variable and training dataset [28]. These formulations are supportive of the claim that the mean function and covariance could provide a complete GP description through the introduction of the Gaussian distribution. It is important to select a Kernel function (i.e., a strong covariance function) in the training phase. The Kernel matrix has a symmetric, invertible matrix. This contributes to GPR model robustness in target prediction. To identify the optimal Kernel function, the present study manipulated four common Kernel functions, namely, (1) rational quadratic, (2) exponential, (3) squared exponential, and (4) Matern functions, to perform the learning process. The rational quadratic covariance function is defined as:

$$K_{RQ}(x, x') = \sigma^2 \left(1 + \frac{x - x'}{2\alpha\ell^2}\right)^{-\alpha}, \quad (10)$$

in which  $\sigma$  denotes the amplitude,  $\sigma^2$  is the variance,  $\ell$  represents the length scale, and  $\alpha > 0$  is the scale mixture that ascertains the change weights at both small and large scales. The exponential covariance function is formulated as:

$$K_E(x, x') = \sigma^2 \exp\left(-\frac{x - x'}{\ell}\right). \quad (11)$$

The squared exponential covariance function is expressed as:

$$K_{SE}(x, x') = \sigma^2 \exp\left(-\frac{x - x'}{\ell^2}\right). \quad (12)$$

Finally, the Matern covariance function is represented as:

$$K_M(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{x - x'}{\ell}\right) K_\nu\left(\sqrt{2\nu} \frac{x - x'}{\ell}\right), \quad (13)$$

where  $\Gamma$  is the gamma function,  $K_\nu$  represents the modified Bessel function, and  $\ell$  and  $\nu$  are positive variables. In fact, the exponential covariance function and squared exponential covariance function are two particular forms of the Matern covariance function. Setting  $\nu$  to 0.5 converts the Matern covariance function into the exponential covariance function. Also, the Matern covariance function transforms into the squared exponential covariance function at a  $\nu$  approaching infinity. In light of its additional parameter (i.e.,  $\nu$ ) as a larger degree of freedom, the Matern covariance function could make more accurate estimates as compared to the exponential and squared exponential covariance functions.

### 3. Results and Discussion

**3.1. Analysis of Validity and Reliability.** For the accuracy and reliability evaluation of the developed GPR models in the higher heating value prediction of biodiesels, the present study performed a multivariable statistical test. This work coupled some typical statistical measures and some graphical depictions.

**3.2. Statistical Variables.** For the performance evaluation of the proposed models, the present study exploited the mean square error (MSE), the mean of relative error (MRE), standard deviation (STD), root mean square error (RMSE), and coefficient of determination ( $R^2$ ).

TABLE 1: A comparison of the models in MRE, RMSE, MSE, STD, and  $R$ -squared for the training, testing, and total data under various Kernel functions.

Model	Phase	$R^2$	MRE (%)	MSE	RMSE	STD
GPR (exponential)	Train	0.973	2.70	0.45	0.67	0.46
	Test	0.907	4.53	1.03	1.02	0.63
	Total	0.961	3.11	0.58	1.02	0.52
GPR (Matern)	Train	0.951	3.57	0.78	0.88	0.59
	Test	0.932	4.07	0.80	0.90	0.52
	Total	0.944	3.82	0.83	0.90	0.59
GPR (squared exponential)	Train	0.941	3.40	0.72	0.85	0.58
	Test	0.947	4.91	1.18	1.09	0.69
	Total	0.940	3.99	0.89	1.09	0.62
GPR (rational quadratic)	Train	0.946	3.94	0.92	0.96	0.64
	Test	0.928	3.65	0.59	0.77	0.40
	Total	0.943	3.86	0.84	0.77	0.59

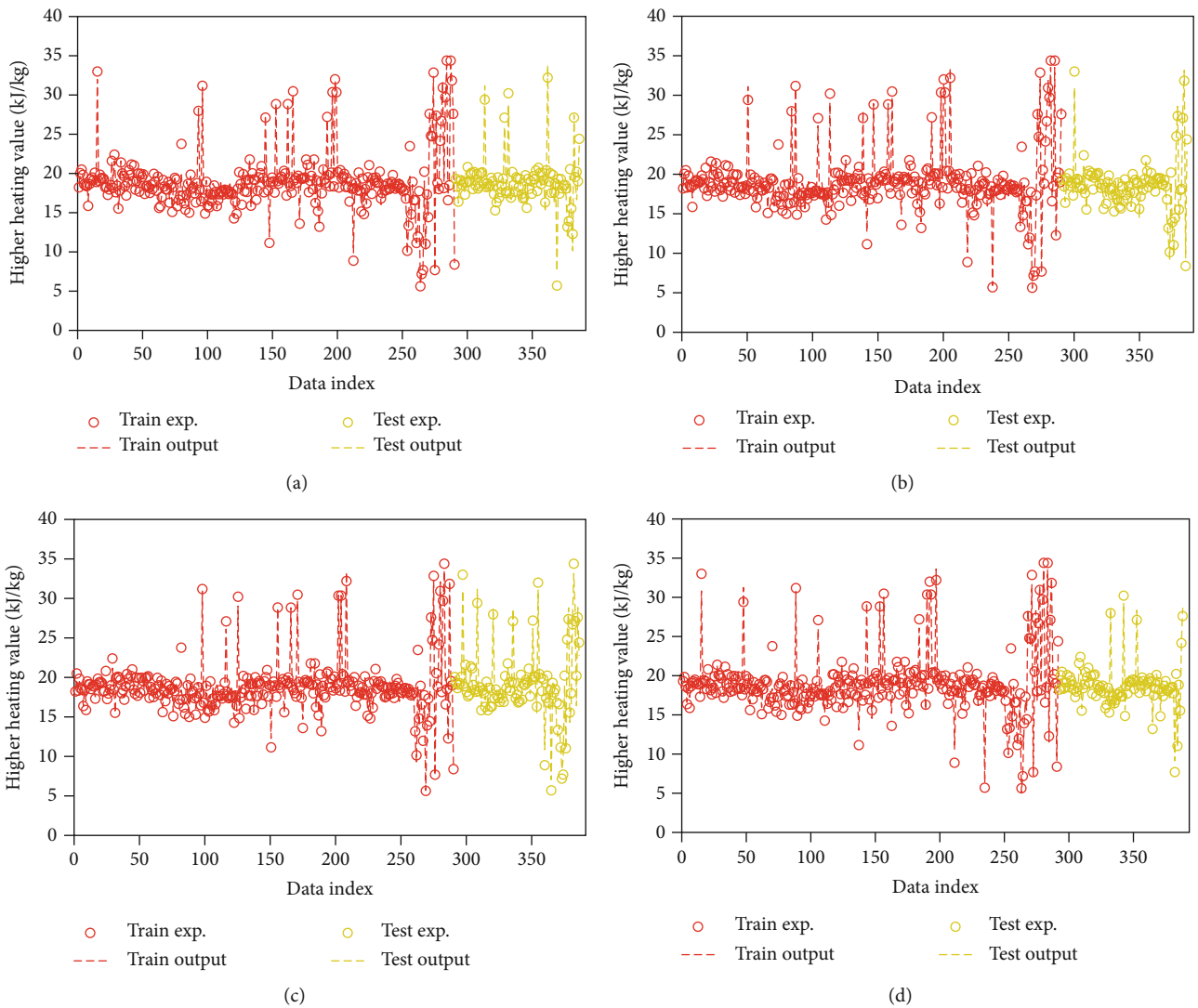


FIGURE 1: A point-by-point comparison of the modeled estimates to experimental quantities under the (a) exponential, (b) Matern, (c) squared exponential, and (d) rational quadratic kernel functions.

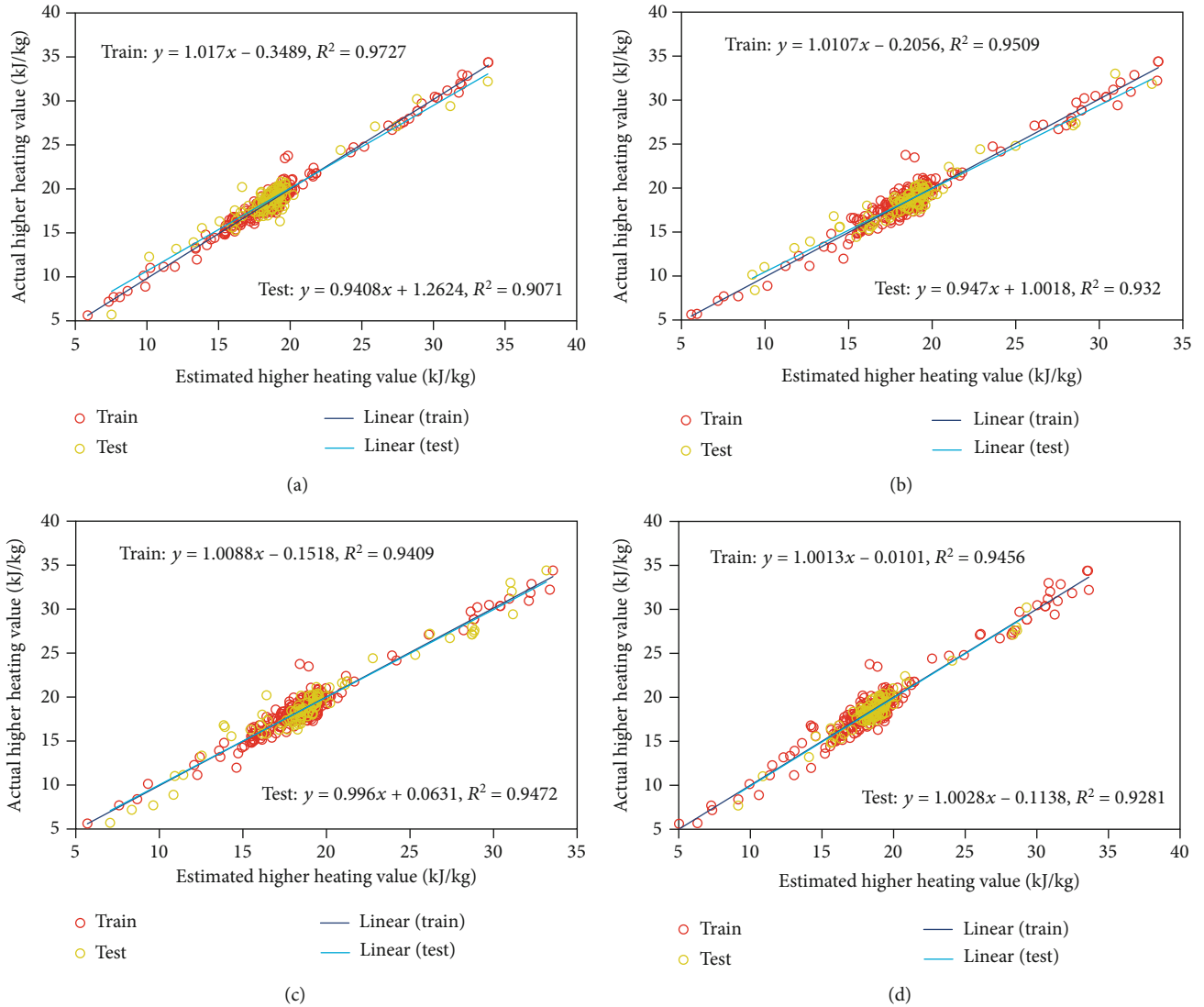


FIGURE 2: A cross plot comparison of the modeled estimates to the experimental data under the (a) exponential, (b) Matern, (c) squared exponential, and (d) rational quadratic kernel functions.

$$\begin{aligned}
 \text{MRE} &= \frac{1}{n} \sum_{i=1}^n \frac{|y_{\text{exp},i} - y_{\text{pred},i}|}{y_{\text{pred},i}}, \\
 \text{MSE} &= \frac{1}{n} \sum_{i=1}^n (y_{\text{exp},i} - y_{\text{pred},i})^2, \\
 \text{RMSE} &= \sqrt{\text{MSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{\text{exp},i} - y_{\text{pred},i})^2}, \\
 \text{STD} &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n \left( \frac{y_{\text{exp},i} - y_{\text{pred},i}}{y_{\text{exp},i}} \right)^2}, \\
 R^2 &= 1 - \frac{\sum_{i=1}^n (y_{\text{pred},i} - y_{\text{exp},i})^2}{\sum_{i=1}^n (y_{\text{pred},i} - \overline{y_{\text{exp},i}})^2}.
 \end{aligned} \tag{14}$$

The statistical parameters related to the mentioned models are calculated and given in Table 1. Dashti and his

colleagues used different models to predict the HHV data [19]. The input and output data used in our paper are similar to their work. The most powerful model they presented was the GARBF model, which has ability to estimate the target values with  $R^2$  and MSE equal to 0.9500 and 0.7401, respectively. However, according to the values obtained in Table 1 of our paper, the GPR (exponential) model has the ability to estimate these values with an accuracy of 0.961 and 0.58, respectively.

3.3. Point-by-Point Agreement Plot. Figure 1 compares the HHV estimates of the GPR models to the measured values, in which “Data Index” represents the sample number, “Train Exp.” Represents the experimental training set, “Train Output” stands for the training set estimate, “Test Exp.” denotes the experimental testing dataset, and “Test Output” represents the testing dataset estimates. According to this figure, most estimates are in good agreement with the experimental data points in all the models. Also, the exponential approach



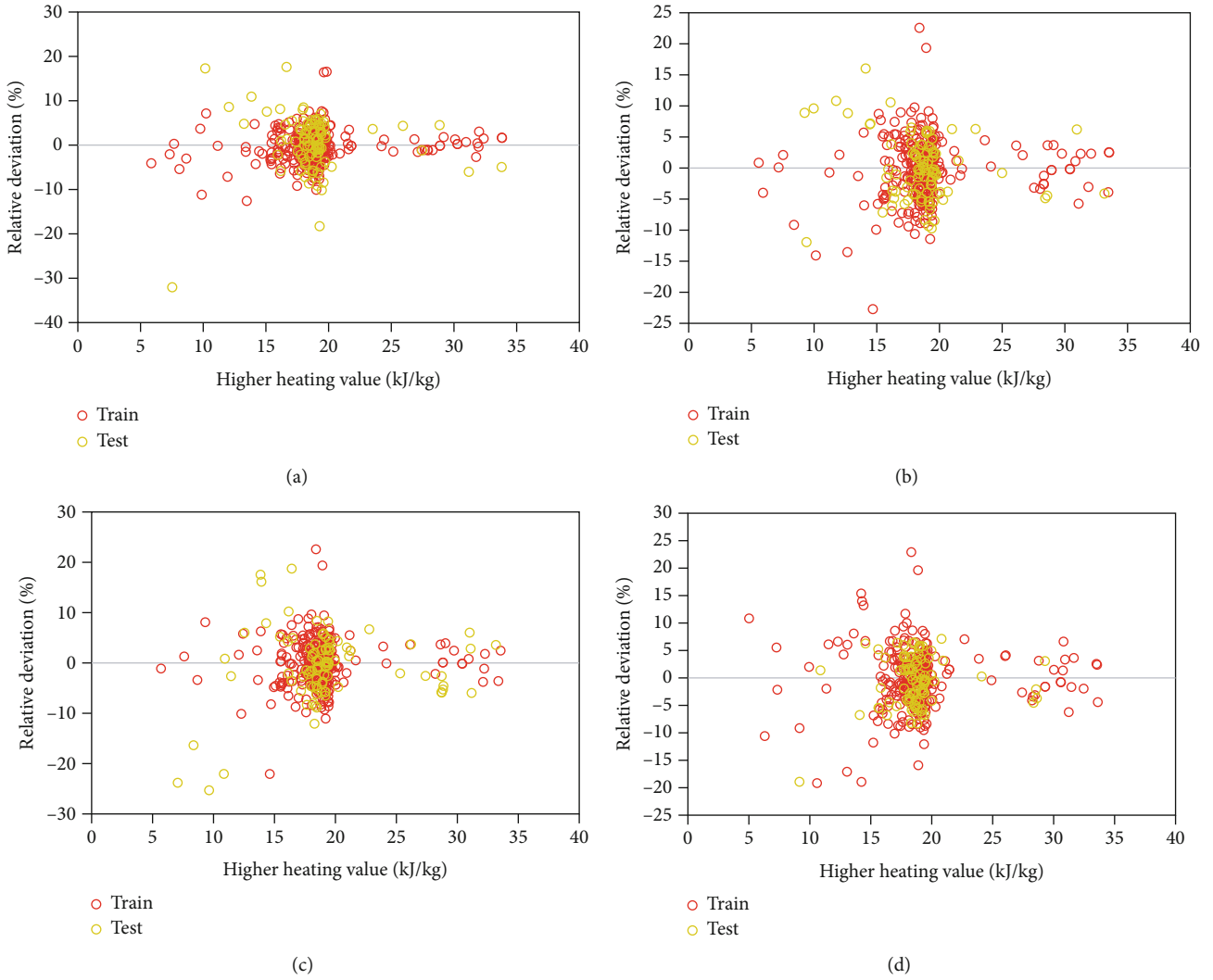


FIGURE 3: Relative deviations of the GPR models versus HHV measurements under the (a) exponential, (b) Matern, (c) squared exponential, and (d) rational quadratic kernel functions.

has the highest accuracy and lowest discrepancy. This is supportive of the statistical evaluation findings.

**3.4. Cross Plot.** Figure 2 illustrates the cross plots of experimental HHV quantities versus the corresponding estimates. It further supports the reliability of the proposed models. As can be seen, the linear trend with an  $R^2$  range of 0.90-0.97 demonstrates that the predictions and measurements are consistent for both the training and testing datasets. As can be seen in Figure 2(a), the most accurate results were obtained by the exponential kernel function.

**3.5.  $y \sim y \sim$  Relative Deviation Distribution.** Figure 3 depicts the relative deviation distributions of the HHV estimates of the developed GPR models. It should be noted that the relative deviation (RD) is calculated as:

$$RD(\%) = 100 \times \left( \frac{y_{\text{exp},i} - y_{\text{pred},i}}{y_{\text{exp},i}} \right). \quad (15)$$

These graphs help determine the degree to which the calculations are realistic based on the experimental quantities. The reliability of the estimates is described by locations of the training and testing data points concerning the horizontal zero-line. According to Figure 3, most relative deviations were found to be from -20% to 20%, which is a favorable range. Furthermore, the points are mostly resting near the horizontal line (Figure 3(a)), in particular those of the exponential kernel function.

**3.6. Sensitivity Analysis.** The present study employed a sensitivity analysis to relate the exponential outputs to the independent input variables. Furthermore, this work employed the relevancy factor (RF) as Pearson's method as [29, 30]:

$$RF = \frac{\sum_{i=1}^n (\bar{x}_k - x_{k,i})^2 \sum_{i=1}^n (\bar{y} - y_i)}{\sqrt{\sum_{i=1}^n (\bar{x}_k - x_{k,i})^2 \sum_{i=1}^n (\bar{y} - y_i)^2}}, \quad (16)$$

in which  $k$  denotes the input type, while  $n$  represents the

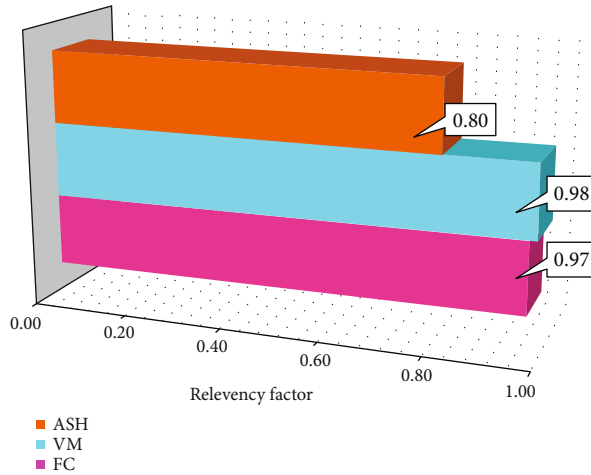


FIGURE 4: Sensitivity analysis results of the best GPR model.

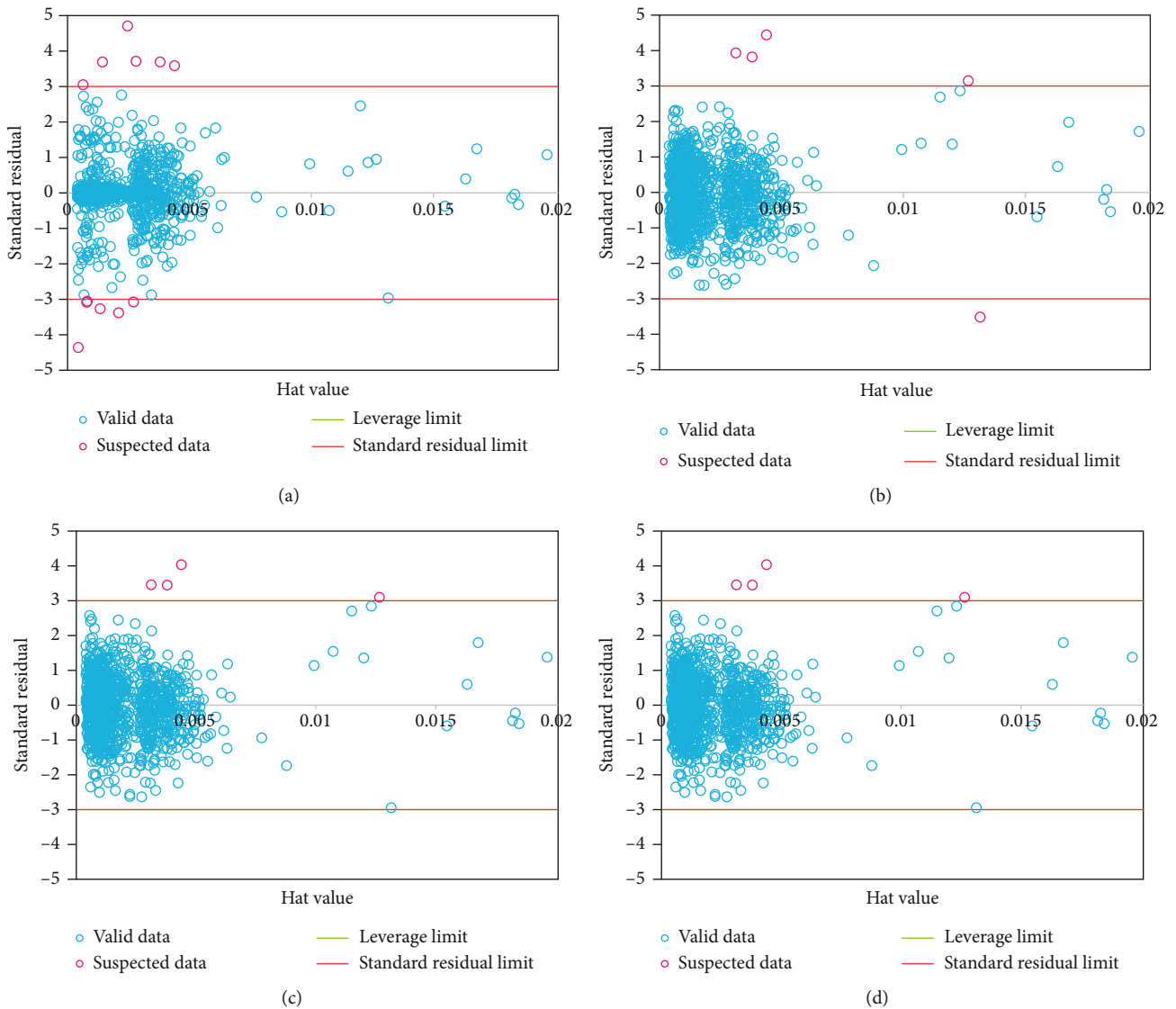


FIGURE 5: William plots of (a) exponential, (b) Matern, (c) squared exponential, and (d) rational quadratic kernel functions.

number of data points. Also,  $x$  is the input value,  $\bar{x}_k$  is the average value of input  $k$ ,  $y$  is the target, and  $\bar{y}$  is the average value of the target [31, 32]. RF varies in the range of  $[-1, 1]$ ; a negative RF represents an inverse relationship between the inputs and output, while a positive RF stands for a direct relationship. A smaller difference between RF and the above-mentioned limits would imply a stronger input-output relationship.

Figure 4 shows the relative deviation results of the proposed models. As can be seen, all input variables have a direct effect on the HHV. Hence, the proposed models can be said to be able to emulate the effects of several inputs on the target.

**3.7. Outlier Detection.** Laboratory data values are always accompanied by uncertainty. The present work employed the Williams plot of standardized residuals ( $R$ ) versus leverage ( $H$ ) to shed some light on uncertain points. The diagonal entries represent the leverage values in the projection matrix  $H = X(X^T X)^{-1} X^T$ , in which  $X$  represents the explanatory variable matrix, while  $T$  stands for the transpose matrix operator [33, 34]. A leverage value above the threshold implies uncertainty and a high-leverage point. The leverage threshold is obtained as [35, 36]:

$$H^* = \frac{3(\text{number of inputs} + 1)}{\text{number of data points}}. \quad (17)$$

Figure 5 illustrates the William plots of the proposed models. One can qualify the data points based on the corresponding locations in the plots. The model applicability domain is represented by the squared area of  $-3 \leq R \leq 3$  and  $H < H^*$ . The area of  $\leq R \leq 3$  and  $h > H^*$  represents the good high leverage data. A question mark represents the model's ability to estimate data points resting in this area. The points that lie in the domains  $R > 3$  or  $R < -3$  are referred to as the bad high leverage data (i.e., outliers). According to Figure 5, a small number of points exist in the bad high leverage and good high leverage areas; the remaining points fall in the model applicability domain.

## 4. Conclusion

The present study adopted GPR and implemented a comprehensive modeling work on extensive data collected from the literature. HHV was modeled as a function of fixed carbon (FC), volatile matter (VM), and ash  $s$  by using four Kernel functions. The data were divided into training and testing datasets. This study utilized cross plots, relative deviation diagrams, sensitivity analyses, and Williams plots along with the parametric analysis of errors (including MRE, MSE, RMSE, and  $R^2$ ). The developed GPR models were found to have high performance in the HHV estimation of biodiesels. The exponential function exhibited the highest accuracy, while the squared exponential function showed the lowest accuracy—the MRE and adjusted  $R^2$  were calculated to be 3.11% and 0.961 for the exponential function, respectively, while they were obtained to be 3.99% and 0.94 for the squared exponential function, respectively. The cross plots and relative deviations demonstrated satisfactory consistency

between the HHV measurements and estimates. Finally, the outlier analysis was performed to evaluate data validity and GPR model reliability.

## Data Availability

The data used to support the findings of this study are provided within the paper.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Social Science Planning Project Fund of Sichuan Province (SC19C058) and Humanities Special Fund of Southwest Petroleum University (2018RW007).

## References

- [1] M. Simjoo, "Polyacrylamide gel polymer as water shut-off system: preparation and investigation of physical and chemical properties in one of the Iranian oil reservoirs conditions," *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, vol. 26, no. 4, pp. 99–108, 2007.
- [2] M. Tajmiri and M. R. Ehsani, "Wettability alteration of oil-wet and water-wet of Iranian heavy oil reservoir by CuO nanoparticles," *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, vol. 36, no. 4, pp. 171–182, 2017.
- [3] M. Yazdizadeh, H. Nourbakhsh, and M. R. Jafari Nasr, "A solution model for predicting asphaltene precipitation," *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, vol. 33, no. 1, pp. 93–102, 2014.
- [4] X. Junming, J. Jianchun, and K. Sun, "An investigation of wood-plastic novolac modified by biomass pyrolysis oils," *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, vol. 30, no. 1, pp. 83–87, 2011.
- [5] M. Le Page, *The Great Carbon Scam*, Elsevier, 2016.
- [6] D. R. Nhuchhen and P. A. Salam, "Estimation of higher heating value of biomass from proximate analysis: a new approach," *Fuel*, vol. 99, pp. 55–63, 2012.
- [7] A. MAJUMDER, R. JAIN, P. BANERJEE, and J. BARNWAL, "Development of a new proximate analysis based correlation to predict calorific value of coal," *Fuel*, vol. 87, no. 13–14, pp. 3077–3081, 2008.
- [8] A. Demirbaş, "Calculation of higher heating values of biomass fuels," *Fuel*, vol. 76, no. 5, pp. 431–434, 1997.
- [9] P. Basu, *Biomass Gasification and Pyrolysis: Practical Design and Theory*, Academic press, 2010.
- [10] J. Parikh, S. Channiwala, and G. Ghosal, "A correlation for calculating HHV from proximate analysis of solid fuels," *Fuel*, vol. 84, no. 5, pp. 487–494, 2005.
- [11] T. Cordero, F. Marquez, J. Rodriguez-Mirasol, and J. J. Rodriguez, "Predicting heating values of lignocellulosics and carbonaceous materials from proximate analysis," *Fuel*, vol. 80, no. 11, pp. 1567–1571, 2001.
- [12] R. Maachou, "Energy consumption modeling in activated sludge process using coupling PCA-ANFIS approach," *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, vol. 38, no. 6, pp. 261–273, 2019.



- [13] S. Mesroghli, E. Jorjani, and S. C. Chelgani, "Estimation of gross calorific value based on coal analysis using regression and artificial neural networks," *International Journal of Coal Geology*, vol. 79, no. 1-2, pp. 49-54, 2009.
- [14] S. B. Ghugare, S. Tiwary, V. Elangovan, and S. S. Tambe, "Prediction of higher heating value of solid biomass fuels using artificial intelligence formalisms," *Bioenergy Research*, vol. 7, no. 2, pp. 681-692, 2014.
- [15] S. Hosseinpour, M. Aghbashlo, M. Tabatabaei, and M. Mehrpooya, "Estimation of biomass higher heating value (HHV) based on the proximate analysis by using iterative neural network-adapted partial least squares (INNPLS)," *Energy*, vol. 138, pp. 473-479, 2017.
- [16] E. Akkaya, "ANFIS based prediction model for biomass heating value using proximate analysis components," *Fuel*, vol. 180, pp. 687-693, 2016.
- [17] H. Uzun, Z. Yıldız, J. L. Goldfarb, and S. Ceylan, "Improved prediction of higher heating value of biomass using an artificial neural network model based on proximate analysis," *Biore-source Technology*, vol. 234, pp. 122-130, 2017.
- [18] I. Estiati, F. B. Freire, J. T. Freire, R. Aguado, and M. Olazar, "Fitting performance of artificial neural networks and empirical correlations to estimate higher heating values of biomass," *Fuel*, vol. 180, pp. 377-383, 2016.
- [19] A. Dashti, A. S. Noushabadi, M. Raji, A. Razmi, S. Ceylan, and A. H. Mohammadi, "Estimation of biomass higher heating value (HHV) based on the proximate analysis: smart modeling and correlation," *Fuel*, vol. 257, p. 115931, 2019.
- [20] A. Amani, P. York, H. Chrystyn, B. J. Clark, and D. Q. Do, "Determination of factors controlling the particle size in nanoemulsions using artificial neural networks," *European Journal of Pharmaceutical Sciences*, vol. 35, no. 1-2, pp. 42-51, 2008.
- [21] H. Huang, A. Qin, H. Mao et al., "The prediction method on the early failure of hydropower units based on Gaussian process regression driven by monitoring data," *Applied Sciences*, vol. 11, no. 1, 2021.
- [22] M. Mahdaviara, A. Rostami, F. Keivanimehr, and K. Shahbazi, "Accurate determination of permeability in carbonate reservoirs using Gaussian process regression," *Journal of Petroleum Science and Engineering*, vol. 196, p. 107807, 2021.
- [23] C. Edward and K. Christopher, *Gaussian Processes for Machine Learning Adaptive Computation and Machine Learning*, MIT Press Cambridge, 2005.
- [24] H. Yu, "The gaussian process regression for TOC estimation using wireline logs in shale gas reservoirs," in *International Petroleum Technology Conference*, 2016.
- [25] Q. Fu, W. Shen, X. Wei, P. Zheng, H. Xin, and C. Zhao, "Prediction of the diet nutrients digestibility of dairy cows using Gaussian process regression," *Information Processing in Agriculture*, vol. 6, no. 3, pp. 396-406, 2019.
- [26] C. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 2, MIT press Cambridge, MA, 2006.
- [27] M. Zhu, S. Liu, and S. Gu, "Short-term tide level forecasting based on Gaussian process regression," in *The Eleventh ISOPE Pacific/Asia Offshore Mechanics Symposium 2014 International Society of Offshore and Polar Engineers*, 2014.
- [28] S. Asante-Okyere, C. Shen, Y. Yevenyo Ziggah, M. Moses Rulegeya, and X. Zhu, "Investigating the predictive performance of Gaussian process regression in evaluating reservoir porosity and permeability," *Energies*, vol. 11, no. 12, p. 3261, 2018.
- [29] M. H. Ahmadi, A. Baghban, M. Ghazvini, M. Hadipoor, R. Ghasempour, and M. R. Nazemzadegan, "An insight into the prediction of TiO<sub>2</sub>/water nanofluid viscosity through intelligence schemes," *Journal of Thermal Analysis and Calorimetry*, vol. 139, no. 3, pp. 2381-2394, 2020.
- [30] A. Dargahi-Zarandi, A. Hemmati-Sarapardeh, S. Hajirezaie, B. Dabir, and S. Atashrouz, "Modeling gas/vapor viscosity of hydrocarbon fluids using a hybrid GMDH-type neural network system," *Journal of Molecular Liquids*, vol. 236, pp. 162-171, 2017.
- [31] S. Hajirezaie, A. Hemmati-Sarapardeh, A. H. Mohammadi, M. Pournik, and A. Kamari, "A smooth model for the estimation of gas/vapor viscosity of hydrocarbon fluids," *Journal of Natural Gas Science and Engineering*, vol. 26, pp. 1452-1459, 2015.
- [32] A. S. Noushabadi, A. Dashti, M. Raji, A. Zarei, and A. H. Mohammadi, "Estimation of cetane numbers of biodiesel and diesel oils using regression and PSO-ANFIS models," *Renewable Energy*, vol. 158, pp. 465-473, 2020.
- [33] A. Kamari, A. H. Mohammadi, A. Bahadori, and S. Zندهboudi, "A reliable model for estimating the wax deposition rate during crude oil production and processing," *Petroleum Science and Technology*, vol. 32, no. 23, pp. 2837-2844, 2014.
- [34] A. Kamari, A. Khaksar-Manshad, F. Gharagheizi, A. H. Mohammadi, and S. Ashoori, "Robust model for the determination of wax deposition in oil systems," *Industrial & Engineering Chemistry Research*, vol. 52, no. 44, pp. 15664-15672, 2013.
- [35] D. C. Hoaglin and R. E. Welsch, "The hat matrix in regression and ANOVA," *The American Statistician*, vol. 32, no. 1, pp. 17-22, 1978.
- [36] A. Kamari, A. Safiri, and A. H. Mohammadi, "Compositional model for estimating asphaltene precipitation conditions in live reservoir oil systems," *Journal of Dispersion Science and Technology*, vol. 36, no. 3, pp. 301-309, 2015.