Research Article

Developing an Extreme Learning Machine-Based Model for Estimating the Isothermal Compressibility of Biodiesel

Yue Wang 1 and Hamid Heydari 2

1School of Intelligent Manufacturing, Xinxiang Vocational and Technical College, Xinxiang, Henan 453003, China
2Institute of Petroleum Engineering, University of Tehran, Tehran, Iran

Correspondence should be addressed to Yue Wang; wangyue20210601@126.com and Hamid Heydari; h.heydari.2014@ut.ac.ir

Received 3 June 2021; Revised 26 June 2021; Accepted 5 July 2021; Published 10 July 2021

Academic Editor: Alireza Baghban

Copyright © 2021 Yue Wang and Hamid Heydari. 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.

Nowadays, the high consumption of fossil fuels has caused many pollutants and environmental problems. Biodiesel has recently been considered as a clean and renewable alternative to fossil fuels. They are found in some molecular structures including fatty acid ethyl esters (FAEEs) and also fatty methyl esters (FAMEs), having various thermophysical characteristics. Thus, it appears essential to select the suitable methods for a particular diesel engine to estimate the ester characteristics. The current research sets out to develop a new and robust method predicting isothermal compressibility of long-chain fatty acid methyl and ethyl esters directly from several basic efficient parameters (pressure, temperature, normal melting point, and molecular weight). Therefore, as a novel and prevailing mathematical method in this field, an extreme learning machine was implemented for isothermal compressibility on the massive dataset. According to statistical evaluations, this novel established model had high accuracy and applicability ($R^2 = 1$ and RMSE = 0.0018714) which is more accurate than previous models presented by former researchers. Among various factors of the sensitivity analysis, temperature and pressure had the greatest effect on the output values, so that the output parameter has a direct relationship with temperature and an inverse relationship with pressure with relevancy factors of 22.44% and −79.81%.

1. Introduction

Conventional hydrocarbon fuels have limited resources and lead to environmental pollution, asphaltene deposition, and formation damage. The rise in the use and contamination of fossil fuels in 2008 by up to 2500 million tons draws the attention of researchers to the quest for better substitutes for common fuels. Biodiesel, among different fuel sources, is more desirable since it may be extricated from available agrarian yields, effortlessly mix (with all amounts) with gasoline, have more agreeable transport and storage, and release a lesser amount of greenhouse gas than usual fossil fuels [1–5].

Often, biodiesel is composed of long-chain FAMEs and FAEEs, generated via transesterification of industrial and natural fatty acid resources, such as vegetable oil and animal fat, besides other fatty feedstock. The long-chain fatty acids’ esterification system by short-chain alcohols, including ethanol or methanol, is indicated by the following reaction in the base catalyst existence:

$$ R – \text{COOH} + R’ – \text{OH} \xrightarrow{\text{base}} R – \text{COO} – R’ + \text{H}_2\text{O}, \quad (1) $$

where $R$ and $R’$ are the long-alkyl and short-alkyl chains, respectively. In an industrial sector, this reaction outlet, in a reactor unit, is a mixture of different kinds of long-chain FAMEs and FAEEs, typically related to impurities. The mixture is well-known as biodiesel. Typical biomass sources of fatty acids in generating biodiesel are rapeseed, soybean, palm oils, and sunflower [6–8]. Various types of long-chain fatty acids can be found in fatty feedstocks from which biodiesel can be obtained, resulting in different molecular structures of fatty acid esters in biodiesel. Different thermophysical characteristics result from various molecular
structures, which impact the fuel performance, the exhaust gas content, and combustion efficacy in diesel engines [9–11].

Fatty acid esters have four basic thermophysical properties that control the combustion behavior of biodiesel. These characteristics are density, speed of sound, isentropic compressibility, and isothermal compressibility [12, 13]. Appropriate fuel combustion necessitates accurate timing for fuel injection, being thoroughly associated with speed of sound within the biodiesel and compressibility. The higher the speed of sound and compressibility of the biodiesel, the earlier the injection timing, which leads to the release of more nitrogen gas oxide [14–16]. The characteristics studied (density, sound speed, isentropic compressibility, and isothermal compressibility) are the roles of fatty acid esters’ pressure, temperature, and molecular structure, which are the major biodiesel components.

To define the biodiesel’s performance, information on the stated thermophysical characteristics’ values for long-chain FAEEs and FAMEs is necessary for a variety of pressure and temperature settings because of their significance and efficient parts. In laboratories, the real values of the measures are provided by experimental measurement. In most cases, these measurements, however, are problematic or could be tedious and costly. Therefore, in this scope, it is essential to use empirical models and mathematical relations.

To establish computational correlations, various statistical-mathematical algorithms (machine-learning methods) exist [17, 18]. The well-known instruments to develop correlation include the adaptive neurofuzzy inference system (ANFIS), genetic programming (GP), artificial neural networks (ANNs), and stochastic gradient boosting (SGB) trees [18–23].

Many studies have been performed for predicting biodiesel characteristics thus far. Several approaches, including the ANFIS and ANN, have been used to investigate and model biodiesel density at a variety of pressures and temperatures [24–29]. Some researchers also have studied sound speed and compressibility [16, 30]. Nevertheless, the current research aims at proposing a new and appropriate model to predict isothermal compressibility of several long-chain fatty acid esters based on modest and basic parameters, such as pressure, temperature, normal melting point, and molecular weight. In this research, an extreme learning machine (ELM) was applied as a prevailing mathematical modeling method, the use of which appears to be new in this field. The applicability and prediction performance of this novel model has been increased due to using great deals datasets with various temperature and pressure settings. Furthermore, some other benefits of the models are their simplicity, and basic parameters, the high precision.

2. Materials and Methods

2.1. Data Collection. In order to build the new model, a large dataset containing 310 data samples of isothermal compressibility was collected from the literature. The relevant collected data are given elsewhere [31]. In practice, this dataset was first divided into two subsets, randomly, as a tool for the development of powerful model based on the data, among which 70% was regarded as training data, and the remaining parts were considered testing data. The training data were then applied to design the novel model, and the testing data were used to assess the established model’s estimation ability. Similar trends for more precision were seen in both data subsets.

2.2. Extended Learning Machine. The extended learning machine or ELM is a hidden-layer single counter neural network that in the updated version may not make the hidden layer neurons homogeneous [32]. ELM, despite its special features, avoids many problems such as the amount of learning and training courses and the selection of stop criteria [33–35].

The ELM algorithm is developed from biological learning systems and BP learning algorithms that have been proposed to solve difficult problems. Its main application is clustering, regression, classification, and extraction of features. Depending on the characteristics of biological learning, a neural network can have random neurons whose settings are independent of the situation [36, 37]. In the extended learning machine method, all hidden neurons are independent of training samples and each other, which is in contrast to crash training methods/network models [38, 39].

Learning nodes do not need to be regulated, and hidden weights can be determined before the learning phase. Also, when the hidden neurons are large enough for problems, the ELM architecture is strong, but it does not work well for highly data-dependent learning methods.

The extended learning machine method does not require adjustment for the connection between input and hidden layers. The ELM output is a simple model with an output node such as follows [40]:

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta.$$  

In the above formula, \(h(x) = [h_1(x) \ldots h_L(x)]\) represents the output vector of the hidden layer, which is dependent on the input of \(x\). This function plots the input \(d\)-dimensional data into the output \(L\)-dimensional data, and \(\beta = [\beta_1 \ldots \beta_L]^T\) is the output weight vector between the output node and the hidden layer. The ELM decision function is described as a classification of binary applications as follows [41]:

$$f_L(x) = \text{sign}[h(x)\beta].$$

The training error in ELM is the lowest compared to the old methods and even tries to reach the smallest norm of output weight.

From Bartlett’s theory, we can see that for a similar training error compared to training neural networks, lower weights are required to achieve better generalization performance. It should be noted that if \(h(x)\) is an infinitely variable activation function, then there is no need to adjust the ELM weight [42].
At the beginning of the training phase, the input weight can be selected randomly. Finally, in the training phase, the weight of the input layer is fixed. Also, the weight that connects the hidden and output layers is the weight that searches the least squares of the following target function. The ELM method reduces output weight and learning error norms as follows [40]:

\[
\minimize: \|H\beta - T\|^2 \quad \text{and} \quad \|\beta\|,
\]

where \(H\) represents the output layer activation matrix of the hidden layer.

Simultaneous minimization of the weight connection norm \(b\) and learning error is the main task of the extended learning machine method. From solving the least squares problem, the output weights of the hidden layer can be easily calculated as follows [43, 44]:

\[
\beta = H^T T.
\]

In the above equation, \(H^T\) stands for the pseudoinverted Moore–Penrose \(H\) matrix. The ELM method consists of several steps, which are summarized in the following:

1. Randomly set hidden node parameters for hidden nodes, input weight, biases \(b_i\), and \(a_i\).
2. Obtain the output layer of the hidden layer
3. Calculate the output weight vector

\[
\beta = H^T T.
\]

\(T\) is defined as follows:

\[
T = [T_1 \ldots T_L]^T.
\]

Analyzing the quasi-inverse Moore–Penrose of a matrix can be performed in several ways, including the orthogonalization and iterative methods, the ultrasonic projection method, and unit value analysis. The orthogonal projection method is possible in the following two ways: for times when \(HH^T\) is nonsingular and \(H^T = H^T (HH^T)^{-1}\) or for times \(H^T H\) and \(H^T = (H^T H)^{-1}\) is nonsingular. This solution works more accurately and its generalized performance is better as well [45].

From ELM theory, it can be concluded that many \(h(x)\) feature mapping functions may be used in ELM design. As a result, any continuous operation can be approximated by ELM.

### 3. Results and Discussion

In Table 1, we used different methods including STD, \(R^2\), MSE, RMSE, and MRE% to analyze the estimated isothermal compressibility and compare it with the real value. To calculate the statistical parameters, we have used the following equations.

\[
MRE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_{exp,i} - y_{pred,i}}{y_{pred,i}} \right|,
\]

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{y_{exp,i} - y_{pred,i}}{y_{exp,i}} \right)^2,
\]

\[
RMSE = \sqrt{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{y_{exp,i} - y_{pred,i}}{y_{exp,i}} \right)^2,
\]

\[
STD = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{y_{exp,i} - y_{pred,i}}{y_{exp,i}} \right)^2},
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_{exp,i} - y_{pred,i})^2}{\sum_{i=1}^{n} (y_{exp,i} - \bar{y}_{exp})^2}.
\]

In equations (8)–(12), \(y_{exp,i}\) and \(y_{pred,i}\) represent the desired value and the estimated values, respectively, and the parameter \(n\) represents the number of experimental data. Models with more \(R^2\) and less STD, MSE, and MRE% are more accurate in evaluating output values. Looking at Table 1, it is easy to see that the GPR model is an accurate model in estimating isothermal compressibility. Figure 1 also shows that according to the relative deviation analysis, this model has a high accuracy in predicting the target values, and as shown in the figure, most relative deviation points are close to zero and its maximum value is equal to 2.5%.

One method of evaluating the accuracy of the model in predicting target values is to visually compare real data and estimated data. Figure 2 deals with such comparison. Accordingly, the prediction of the majority of data related to the testing and training phases is done by the ELM model with high precision.

Figure 3 shows the cross-plot of estimated isothermal compressibility related to the ELM model and experimental ones. Since the slope of the line is near to 45°, the correctness of the presented model is high, and \(R^2\) values related to this model for training and testing phases were obtained as 0.9999 and 0.9998, respectively.

By examining the statistical parameters, it can be seen that the performance of the current model is better in predicting target values compared to the previously suggested models. By comparing the results of this study with the study of Abooalai et al. that used the same data to estimate the isothermal compressibility, the superiority of the ELM model can be concluded [31]. Table 2 proves the accuracy and efficiency of the models. The results of estimation of the isothermal compressibility of diesel oil by ELM, SGB, and GP methods are given. As shown, the developed model has better accuracy compared to the previous models in the...
the value of $R^2$ is 0.9998 for the developed model and 0.99931 and 0.99885 for the previous models.

3.1. Sensitivity Analysis. To investigate the input parameters that affect the isothermal compressibility, the following equation can be used to calculate the relevancy factor.

$$
    r = \frac{\sum_{i=1}^{n} (X_{ki} - \bar{X}_k)(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_{ki} - \bar{X}_k)^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}} \quad (13)
$$

In equation (13), $\bar{Y}$ represents the mean output, $Y_i$ abbreviated $i^{th}$ output and $X_{ki}$ is the $k^{th}$ input, and the input is represented by $X_k$. The value of $r$ is consistently less than unity. Figure 4 shows the impact of each of the input parameters on the isothermal compressibility. It can be seen from the presented figure that the pressure and melting point have a negative effect on the output. Unlike the negative effect of them, the two characteristics of temperature and molecular weight have a positive effect on the target values. In fact, the isothermal compressibility increases with increasing temperature ($T$) and molecular weight (MW) of biodiesel oils. Also, it can be seen that temperature has the most effect and pressure has the least effect on the isothermal compressibility.

3.2. Detecting Available Outlier Datapoints in the Dataset. Data measured in the laboratory are always subject to an error. To check the accuracy of real data, we used the leverage limit method. Due to its critical impact in evaluating the proposed model, based on soft computing, the approach presented in this method is underlying and useful. The main purpose of this method is to ensure that the real output datapoints are authentic. In this research, Williams’s design has been used to evaluate the feasibility and reliability of the developed ELM model to predict target values [46]. According to this concept, the data that are found in the

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Dataset</th>
<th>$R^2$</th>
<th>MRE (%)</th>
<th>MSE</th>
<th>RMSE</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isothermal compressibility (1/GPa)</td>
<td>Train</td>
<td>1.000</td>
<td>0.13</td>
<td>0.0000019</td>
<td>0.0013666</td>
<td>0.0011408</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>1.000</td>
<td>0.18</td>
<td>0.0000084</td>
<td>0.0029067</td>
<td>0.0026475</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1.000</td>
<td>0.14</td>
<td>0.0000035</td>
<td>0.0018714</td>
<td>0.0016566</td>
</tr>
</tbody>
</table>
The modeling procedure was designed on the basis of analogous independent parameters, including temperature (T), pressure (P), normal melting point (Tm), and molecular weight (Mw) of fatty acid esters. The current research used the prevailing statistical-mathematical machine-learning frameworks of an extreme learning machine. The results show that the established model was appropriate and valid in estimating the hinted characteristics concerning statistical variables. Nevertheless, more evaluations indicated the precision of the ELM technique to be higher than that of previously proposed model. This model offers novel uses in simulating controlling biodiesel systems, besides estimating the characteristics for newly developed long-chain fatty acid esters.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


