

Research Article

Parametric and Nonparametric Approaches of Reid Vapor Pressure Prediction for Gasoline Containing Oxygenates: A Comparative Analysis Using Partial Least Squares, Nonlinear, and LOWESS Regression Modelling Strategies with Physical Properties

Hayder M. Issa 🗅

Department of Chemistry, College of Science, University of Garmian, Kalar 46021, Iraq

Correspondence should be addressed to Hayder M. Issa; hayder.mohammed@garmian.edu.krd

Received 23 December 2023; Revised 17 February 2024; Accepted 22 February 2024; Published 11 March 2024

Academic Editor: Chen Lin Soo

Copyright © 2024 Hayder M. Issa. 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.

This study provides insights into the challenges involved in predicting the Reid vapor pressure (RVP) of gasoline-oxygenate blends (GOB), which is an important indicator of fuel quality and compliance with environmental and performance standards. Given the enormous variety of gasoline compositions and ratios available, there is a significant demand for a fast, straightforward, and cost-effective technique to predict RVP without relying on costly instruments or complicated spectral measurements that involve numerous input variables. A comparative performance analysis has been performed for different regression modelling strategies for predicting RVP in GOB, which is valuable for researchers and practitioners in the petroleum industry for saving time and money. Parametric and nonparametric approaches were compared using partial least squares regression (PLSR), nonlinear regression (NLR), and nonparametric regression (NPR) models. Locally weighted scatterplot smoothing (LOWESS) approach was applied to the NPR model. The gasoline's physical characteristics (distillation curves and density) formed the basis for the analysis of these models' performances. Acceptable error metrics have been reached for root mean square error of calibration and prediction (RMSEC and RMSEP) values, for the PLSR, NLR, and NPR models, which are 4.790, 6.235, 4.739, 6.149, 3.968, and 6.029, respectively, which are close for those reported in literature. The NPR model eliminates parametric constraints and allows for a different kind of data structure to emerge. The established models here demonstrate a sound ability to overcome barriers by omitting the use of inconvenient spectral measurements to save expense and simplify data calibration, making them a promising approach for RVP detection of GOB. This finding aids in the development of more accurate RVP prediction models and contributes to the optimization of fuel formulations.

1. Introduction

Gasoline is the main product of petroleum industry, and its chemical composition might change based on the refining methods. Gasoline is commonly produced from processes such as fractional distillation, isomerization, reforming, cracking, and alkylation, though these processes may not always work in tandem [1]. Additives such as antiknock agents, dispersants, detergents, and oxygenates are frequently used to increase the efficiency of gasoline [2]. Oxygenates are chemical compounds containing oxygen-containing functional groups and are added to gasoline to improve its properties and combustion performance by providing the proper amount of oxygen needed for the combustion of gasoline [3]. Various ethers and alcohols, such as ethanol, methyl tertiary butyl ether (MTBE), tertiary amyl methyl ether (TAME), tertiary butyl alcohol (TBA), ethyl tertiary butyl ether (ETBE), di-isopropyl ether (DIPE), and tertiary amyl ethyl ether (TAEE), are regularly used as oxygenates in gasoline [4].

Reid vapor pressure (RVP) is among few distinctive parameters that are frequently used for better handling and controlling of petroleum products during different stages of processing, transportation, or storage [5]. Furthermore, for gasoline, RVP is a key indicator to determine compliance with environmental and performance standards and regulations like vapor lock, percolation, fuel vaporization, and pollutant emission [6]. RVP testing is routinely used at 37.78°C (100°F) to determine the vapor pressure characteristic of gasoline-oxygenate blends (GOB) [7]. Dhamodaran and Esakkimuthu [8] determined that the uncertainties of different instruments used to investigate RVP amount to approximately 1.03% of the RVP that was tested in a given sample.

Measuring the RVP of GOB can be difficult, making it hard to confirm that anticipated RVP values are correct. Distillation curves for gasoline are greatly distorted when oxygenates with very diverse characteristics are added, which can cause a substantial change in RVP. It is also challenging to anticipate the RVP of GOB that contain several oxygenates because the RVP of oxygenates might vary greatly based on the specific oxygenate being employed. For GOB material development, when there are many different composition and ratio possibilities, and for quality assessment purposes, a fast, simple, and inexpensive method for RVP determination is still much needed. Industry-standard protocols such as [9, 10] are now used to measure RVP for GOB. Although these approaches can be deemed fast, especially for D5191, however, they are not performing up to par. So, more and more often, studies are turning to models that can anticipate future outcomes. Due to the influence of many factors (model sophistication, lack of data, and uncertainty of input variables), and the potential for changes in fuel composition over time, developing a model to reliably predict the RVP of GOB can be a complex and challenging process. Empirical and semiempirical methods are used in the creation of these models. These models can be used to predict RVP of GOB, including linear regression and multiple linear regression for gasoline blends containing ethanol, MTBE, and ETBE or other oxygenates. In order to use semiempirical models like UNIFAC, UNIQUAC, and SAFT, one must have indepth familiarity with the intricate vapor pressure thermodynamics of the tested GOBs [7]. The RVP of the GOB blends can be predicted using a UNIFAC-based method that takes into account the interaction between the blend components and temperature. Empirical models, including chemometric methods, are created through mathematical approximation in order to account for all potential sources of variance in a dataset [11]. Multivariate calibration, which takes into account a substance's spectral or physical properties, is commonly used in chemometrics. As one of several multivariate data analysis tools, regression methods are extensively utilized because of the valuable insight they provide into a variety of gasoline quality factors with a relatively small sample size.

Several linear and nonlinear multivariate regression techniques, including partial least squares regression (PLS), artificial neural network (ANN), support vector machine (SVM), and principal components regression (PCR) [12–15], have been used to successfully predict the RVP of gasoline based on data from spectral analysis or physical properties. Although these studies predict the RVP of gasoline by the regression calibration methods, using either spectroscopic analysis or physical properties has reached good standard error values, but still, it is required to explore other regression methods to overcome the difficulties that might arise due to various processes, adulteration, and blending causing tremendous variability of gasoline types. Sophisticated regression methods might serve to predict the RVP in complicated cases. Therefore, it may be helpful to use more advanced regression approaches to generate trustworthy predictive models for a crucial parameter like the RVP for gasoline.

In light of the fact that gasoline formulations are so complicated, it is hard to expect a simple correlation between the fuel's physical properties and its RVP. Therefore, linear regression models may not be enough for predicting gasoline attributes in the presence of more complicated and nonlinear interactions [16]. Thus, nonlinear and nonparametric regression approaches can be considered as a potential alternative strategy for RVP prediction in gasoline. Most of current developed regression models of RVP determination in gasoline in the literature depend on linear approaches like partial least square regression (PLSR) and multiple linear regression (MLR), in which they assume a linear formula for dependent variable correlation with independent variables. Even while nonlinear regression (NLR) models may improve prediction performance a little better, nonparametric regression (NPR) models can be more useful when dealing with relations that are hard to visualize or nonlinearity cases [17, 18].

This research is aimed at comparing the predictive performance of PLSR, NLR, and NPR, for NPL locally weighted scatterplot smoothing (LOWESS) regression that has been used, methods in order to gain a better understanding of the advantages and disadvantages of each method when used to develop a multivariate calibration approach for establishing a regression model to predict RVP of gasoline based on its physical properties. Density (S-Dens) at 15°C, initial boiling point (IBP), and final boiling point (FBP) are utilized as explanatory, independent variables in conjunction with distillation curves boiling temperatures at quantities of 10%, 50%, and 90% of recovered condensate (T10, T50, and T90, respectively). The chemometrics' use of machine learning algorithms to predict RVP in gasoline has not yet integrated NLR and NPR into multivariate calibration regression. As a result, the developed predictive models supply new chemometrics tools that address the need for streamlining the analytical scheme using physical properties to overcome difficulties of anomalous conditions generated from various gasoline types and compositions, and those are not previously included in calibration sets of existing prediction models.

2. Materials and Methods

2.1. Sampling, Measurements, and Instrumentations. The current study has been conducted using 913 commercial gasoline samples, of both premium and regular. The samples were taken from a number of gas stations in northern Iraq; the gasoline is produced in oil refineries both inside and outside of Iraq using a variety of crude oil sources. Oxygenates,

which include methanol, ethanol, TAME, MTBE, and ETBE, are in a varied range of quantities among GOB samples as stated in Table 1. Densities were measured using a portable automatic density tester, distillation curves were measured utilizing an automatic distillation analyzer, and RVP was determined utilizing a portable RVP tester. According to the standard test method [19], the collected samples were kept in sealed polyethylene containers at a temperature of less than 8°C.

To obtain the temperatures along the distillation curve (IBP, T10, T50, T90, and FBP), an automated microdistillation analyzer (Model: PMD 110: PACLP, USA) has been employed in accordance with ASTM standard test method [20] that correlates precisely with the ASTM standard test method [21]. Samples of cooled liquid gasoline, 100 ml in size, were first introduced to the apparatus. Under ambient pressure, the boiling points of the collected condensate were determined. To ensure the accuracy of the automated distillation process, 10 randomly selected samples of gasoline were tested using a manual distiller designed in accordance with the standard test procedure [21] where there was not a noticeable dissimilarity in both cases.

The Reid vapor pressure (RVP) was measured as airsaturated total vapor pressure using a portable automated RVP tester (Model: ERAVAP: eralytics GmbH, Austria) of mini method (single expansion method) according the standard test method of [10].

The portable tester (Model: ERAVAP: eralytics GmbH, Austria) equipped with a built-in high precision density meter was used to measure the density (S-Dens) of gasoline samples, according to ASTM standard test methods [22]. The tested samples were placed in clean, dry testing containers connected to the apparatus; the results were then converted to the density of gasoline at 15°C.

Testing devices utilized in this study are reliable since they adhere to the standard technique [23] followed by the Garmian Directorate of Oil and Minerals/Ministry of Natural Resources in the Kurdistan Region of Iraq, the laboratory where the testers were placed. In order to assess the repeatability and reproducibility of the testers utilized, a normal procedure called for random 10 gasoline samples to be examined seven times by three separate personnel [24]. The prediction accuracy of the RVP model was evaluated by comparing the repeatability and reproducibility values obtained from the analysis of three randomly selected samples by three different individuals with ten replicates per sample to the repeatability and reproducibility values obtained from the reference method [25].

2.2. Regression Analysis and Model Development. The standardized geometrical distance to the PLS model in the explanatory variables (X variables) and the RVP dependent variable (y variable), abbreviated DModX and DModY, respectively, was used to perform an outlier analysis on the dataset of 913 gasoline samples with 95% confidence level, where the samples lie outside of the tolerance volume around the model, beyond the data range of the model, as determined by the critical distance (D_{crit}) corresponding to the 0.05 probability level [26]. As reported by Silva et al. [27], the DModX value for each sample was determined using Equation (1). The DModX was standardized by dividing it by the $D_{\text{crit}}(x)$ value for *x* variables, as shown in Equation (2).

DModX =
$$\sqrt{\frac{\sum_{k=1}^{k} (x_{j,k} - x_{\text{model},j,k})^2}{K - A}}$$
, (1)

Standarized DModX =
$$\frac{DModX}{D_{crit}(x)}$$
, (2)

where k represents an X variable, K is the total number of X variables, j is for latent variable, and A is the number of latent variables in the PLS model. The DModY for the dependent variable y was calculated in the same way as the DModX.

Regression models were created for the independent y variable (RVP) and the explanatory X variable (physical characteristics), with PLSR, NLR, and NPR as the underlying regression approaches. Following the Kennard-Stone algorithm [28], the gathered gasoline samples were first divided into two primary groups: calibration (609) and prediction (304) for PLSR and NLR. Calibration (571), validation (38), and prediction (304) are the three primary categories into which NPL's samples were divided.

Multiple linear regression (MLR) has the potential for producing collinearity, which can raise standard errors and call into question the reliability of the model's coefficients. Partial least squares regression (PLSR) and other regression techniques are used to reduce the number of highly correlated independent (physical properties) variables and thereby eliminate the collinearity problem [29].

Separate matrices X of computed physical properties and a vector y of measured RVP were used for multiple linear regression in PLSR model development. According to Equations (3) and (4), the variables X and y are transformed into "loading" matrices consisting of scores, crucial data about gasoline samples, and the original variables as stated by Geladi and Kowalski [30] and Issa [31].

$$X = \mathrm{TP}' + E,\tag{3}$$

$$y = \mathbf{U}\mathbf{q}' + f. \tag{4}$$

E and *f* are the nonmodulated part of calibration dataset. Then, RVP for new samples (y_{pr}) can be predicted with a regression coefficient of b_{PLS} by using PLS regression to construct a linear relation model between *X* and *y*. The relative loading weight of these variables can be measured using Equation (5) [31, 32]

$$y_{\rm pr} = b_{\rm PLS} X. \tag{5}$$

When modelling a correlation condition that is too complex for linear models, nonlinear regression is often employed [33]. Due to the nonlinear nature of the correlations between the variables comprising gasoline's physical properties and RVP, estimating the RVP is a challenging

TABLE 1: Descriptive statistics of 913 (GOB samples used for RVP	prediction using PLS, NLR,	and NPR regression models.
	· · · · · · · · · · · · · · · · · ·		

Parameter	Min.	Max.	1st quartile	Median	3rd quartile	Mean	Standard deviation $(n-1)$	Sample counts
RVP (kPa @ 37°C)	38.70	75.30	51.70	56.40	61.60	56.62	6.47	913
S-Dens (g/cm ³ @ 15°C)	0.71	0.76	0.73	0.73	0.74	0.73	0.01	913
IBP (°C)	28.70	64.10	35.50	36.90	38.60	39.28	7.25	913
T10 (°C)	20.80	60.80	44.90	46.40	47.80	44.64	6.12	913
T50 (°C)	34.90	111.30	86.40	89.70	93.10	86.38	13.00	913
T90 (°C)	134.40	191.40	152.50	155.40	161.60	158.15	9.29	913
FBP (°C)	163.40	212.60	184.20	187.90	190.40	186.28	7.42	913
MTBE (Vol.%)	0.13	23.5	1.780	4.34	7.130	5.541	4.399	445
ETBE (Vol.%)	0.95	9.85	1.050	1.07	1.100	1.362	1.428	186
TAME (Vol.%)	0.8	7.63	1.058	1.33	2.003	1.725	1.316	28
Methanol (Vol.%)	0.32	25.1	1.283	4.95	10.648	7.366	7.849	82
Ethanol (Vol.%)	0.6	24.1	3.080	3.42	4.150	3.544	1.233	467
Oxygenates (Vol.%)	1.26	29.9	3.150	4.55	5.520	5.585	4.455	913

task. The nonlinear regression process is known for its ability to determine the parameter values of a selected model in order to get the most appropriate fit with the observed data [34]. Nonetheless, when comparing nonlinear models with the goal of producing the least root mean square (RMS) error [35], multiple nonlinear regression (NLR) revealed that the exponential (power law) model provides more fair representation, but still not enough when compared with more complicated methods, of the nonlinear system. Equation (6) shows that the employed power model of the six nonlinear physical property components is necessary to predict the RVP. The power law regression model is advantageous due to its capacity to accurately capture and depict long-tailed distributions in data, making it especially preferred for systems defined by irregular yet significant data [36]. Equation (6) represents the power law regression model, which shows a proportionate relationship between the value of response variable and the values of explanatory variables raised to certain power values.

$$y_{\rm pr} = b_0 * X_i^{b_i}.$$
 (6)

 $y_{\rm pr}$ represents predicted RVP values, both b_0 and b_i are model's coefficients, and X_i represents explanatory variables in the NLR model. Assuming a model (Equation (7)) that eliminates the parametric constraints on the regression curve, nonparametric regression (NPR) makes place for a different kind of structure to emerge, one in which the *x* variables do not have a fixed form but are built instead using the data. Here, the response variable is related to the covariates x_i by [37, 38]

$$y_{\rm pr} = m(x_i) + \epsilon_i. \tag{7}$$

 ϵ_i is the error, or sometimes, it is called random deviation term, and $m(x_i)$ is the regression function term; if it is smooth enough, a particular parametric form can be determined. Nonparametric regression (NPR) models the

expected (conditional) value $E(y_{pr} | x_1 \cdots x_k)$ of y_{pr} depending on the covariates x_i . As a result, the response value that is expected is a function of the variables, $E(y_{pr} \mid x_1 \cdots, x_k) =$ $f(x_1 \cdots, x_k)$. Assuming that the X_i values of the explanatory variables for the *i*th sample at any part of the dataset are represented by an average of y values corresponding to X values in a region close to those particular values of X, by treating the X_i values as random and $m(X_i)$ is then interpreted as a mean of y_{pr} conditional on $X_i = x_i$, $m(x_i) = E(y_{pr} | X_i = x_i)$, which will lead to good predictions of RVP, rather than estimating the RVP from the given physical property dataset using probable different regression models (linear, quadratic, or polynomial) [39], those we are not sure about. The utilized NPR, LOWESS method in this work, gives a considerable effective prediction approach and solves the problem of dispersion of experimental dataset [40], which renders any normal parametric regression method ineffective, because the new model that is constructed numerically for each observation is obtained. The used algorithm here, locally weighted scatterplot smoothing (LOWESS), was originally developed by Cleveland and Devlin [41] in depending on an original work of Cleveland in [42], which proposed to deal with noisy and dispersed datasets. The procedure presumes that it is possible to successfully fit smooth curves using a statistical procedure called "local regression," which makes no assumptions on the shape or form of the curve being smoothed [43]. The LOWESS technique fits a regression model on the k nearest samples using moving nonparametric regression [44].

The explored simple PLSR, NLR, and NPR regression models with easy to obtain physical explanatory variables have been selected for their cost-effectiveness and timesaving characteristics, as opposed to the sophisticated and costly spectral and ANN methods. Their capability for RVP prediction was evaluated by employing realistic error metrics (Equations (8)–(10)).

Root mean square error of calibration (RMSEC) and prediction (RMSEP) [31, 45] were used to assess the precision of the utilized regression methods in comparison to the average error estimate of the true RVP in gasoline samples (Equations (8) and (9). The root mean square error of crossvalidation (RMSECV), as reported by Kehimkar et al. [46] and Issa [31] as shown in Equation (10), was employed using the leave-one-out (LOO) and k-fold cross-validation approaches to compare the performance of PLSR, NLR, and NPR approaches, with and without outliers. The LOO technique extrapolates the results for a left-out sample by fitting calibration model to the number (n-1) of all the datasets, while k-fold technique divides the dataset into a number of folds. The model's performance was assessed using cross-validation on the dataset following the exclusion of one sample and the model's training on the remaining training samples (n-1). Once all samples have been eliminated once, the process is iterated. In general, the estimation of the cross-validation error is achieved by calculating the average of the observed error for each sample.

More error metrics, such as root mean square error for total data set (RMSE), mean absolute percentage error (MAPE) which is equal to average absolute error (AAD) when deviation is around the mean, and mean absolute deviation (MAD), taken from [47], have been employed to compare with prior investigations and validate RVP prediction models in this work for the tested GOB samples (Equations (11)–(13)).

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_{\text{meas},i} - y_{\text{cal},i} \right)^2}{n}},$$
(8)

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_{\text{meas},i} - y_{\text{pred},i} \right)^2}{n}},$$
(9)

RMSECV =
$$\sqrt{\frac{\sum_{i=1}^{N} (y_{\text{meas},i} - y_{\text{CV},i})^2}{N}}$$
, (10)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(\mathcal{Y}_{\text{meas},i} - \mathcal{Y}_{\text{pre},i} \right)^2}{N}}, \qquad (11)$$

MAPE =
$$\frac{1}{N} \sum_{1}^{N} \left| \frac{(y_{\text{meas},i} - y_{\text{pred},i})}{y_{\text{meas},i}} \right| * 100\%,$$
 (12)

$$MAD = \sum_{i=1}^{N} \frac{y_{\text{meas},i} - y_{\text{pred},i}}{N}.$$
 (13)

For *i*th observation, y_{meas} is the measured *y* value, y_{cal} is the calibrated *y* value calculated using the derived models in the calibration set, y_{pred} is the predicted *y* value, y_{CV} is the cross-validated *y* value, *n* is the number of *y* values of the concerning dataset, and *N* is the total number of *y* values. To validate the data readjustment of the calibration and prediction sets, the regression coefficient (R^2) was determined between the actual and predicted values. ISO criteria [24] were used to the chosen model to assess its repeatability and reproducibility.

Using Equation (14) as reported by Guan et al. [48] for the calculation of predicted residual error sum of squares (PRESS) value for internal validation, the leave-one-out with cross-validation approach was applied to determine the number of LV assumed for the PLSR model. By fitting calibration models with different LV to n-1 of the data, it was possible to infer findings for the missing sample. All of the data set samples in the calibration set went through this process. The "leave-one-out" method was used while the number of latent variables was selected based on the lowest PRESS value.

$$PRESS = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2.$$
(14)

3. Results and Discussion

Table 1 shows some descriptive statistics for the 913 gasoline samples dataset. Since several different gasoline kinds and grades were used, the boiling temperatures of the samples used in this investigation vary greatly, the T50 boiling temperature of 50% evaporated gasoline ranges from 34.9°C to 111.3°C. Therefore, T50 becomes greatly dispersed around the mean with a high standard deviation. Similarly, the temperature range between the upper and lower limits of T10 and IBP reveals significant diversity in sample grades. RON and MON values in this investigation ranged from 89 to 120.1 and 84 to 97.8, respectively.

The outlier analyses DModX and DModY for detecting outliers are applied to the results to identify samples that are fairly far from the mean of the dataset by calculating critical distance (D_{crit}) at a 95% confidence interval. The D_{crit} values were 2.0718 and 2.3398 for X and y variables, respectively. As seen in Figure 1, for a significance level (alpha value) of 5%, using DModX and DModY analyses, 9 outliers have been detected for both y and X variables, and they are subsequently excluded in the study's subsequent calculations.

For the purpose of evaluating the prediction performance of the models, PLSR, NLR, and NPR, and their scores to reflect significant variance from the physical property measurements, the statistical indices DModX and DModY are utilized. These indices are defined in Equation (1) and represent the residuals of the developed models. These indices can be utilized to assess the accuracy of predictions and identify instances where the input variables deviate from the norm [49]. They aid to find outliers or unexpected behavior by measuring the distance between the observed data of GOB's explanatory physical characteristics and the RVP predictions. This allows to detect when the process is operating outside of the expected RVP range and take appropriate corrective actions. DModX and DModY help identify the measurements that cause variance from the model and assist in evaluating prediction performance at an identified level of confidence of 95%. The standardized DModX and DModY that are concerned in evaluating the distance from an accepted range of response variable RVP can be calculated using Equation (2).



FIGURE 1: Results of for outlier analysis at a confidence level of 95% for 913 gasoline samples: (a) DModX and (b) DModY.

Figure 2 illustrates the impact of removing outliers on the various regression models examined here. For all datasets, preprocessing by removing outliers used the leaveone-out method of root mean square error for crossvalidation (RMSECV), which builds a model with a calibration set and prediction set. For NPR, the calibration, validation, and prediction sets were used.

The RVP regression models were constructed using the dataset without outlier samples because, as Figure 2 illustrates, removing outliers reduced the RMSECV for the PLSR, NLR, and NPR models. Figure 2 shows that when two procedures of LOO and *k*-fold (here, 10-fold was utilized) of cross-validation assessment are applied to predict RVP for

the analyzed physical qualities, the NPR technique performs marginally better than the PLSR and NLR. Eliminating outlying data has enhanced the performance of all three models, but NPR is slightly higher. There was a close assessment between the two cross-validation techniques.

Before going any further, it is possible to use the Pearson correlation analysis and the linear correlation coefficient (R), to check the correlation between the RVP and the physical property variables (x) in the dataset [50]. The correlation coefficients (R) range from -1 to 1, indicating, respectively, negative and positive linear relationships between the two variables in question. The correlation analysis for the physical property variables used in this study is shown in Table 2,



FIGURE 2: Error internal validation using RMSECV value comparison of the three used regression methods (PLSR, NLR, and NPR) for two different dataset cases: with outliers and without outliers using leave-one-out (LOO) and *k*-fold cross-validation.

TABLE 2: Correlation matrix for RVP and physical properties of GOB samples.

Variable	S-Dens	IBP	T10	T50	T90	FBP	RVP
S-Dens	1.000	-0.362	0.527	0.571	-0.096	0.458	-0.651
IBP	-0.362	1.000	-0.860	-0.921	0.718	-0.835	0.336
T10	0.527	-0.860	1.000	0.948	-0.775	0.755	-0.462
T50	0.571	-0.921	0.948	1.000	-0.688	0.855	-0.462
T90	-0.096	0.718	-0.775	-0.688	1.000	-0.448	0.157
FBP	0.458	-0.835	0.755	0.855	-0.448	1.000	-0.409
RVP	-0.651	0.336	-0.462	-0.462	0.157	-0.409	1.000

and it can be seen that the correlation coefficients with RVP are weak, with *R* values ranging from -0.651 to 0.157 and R^2 values between 0.424 and 0.025, suggesting that there is a nonlinear relationship between the RVP and the *x* variables.

3.1. Partial Least Squares Regression (PLSR). In this work, the popular linear PLSR method served as a starting point due to its adaptability and ease of implementation [51]. Given that this is the case for RVP prediction using gasoline's physical properties, the high dimensionality and irregularity of the dataset makes it difficult to PLSR for developing a highly reliable regression model. Moreover, nonlinear or polynomial PLSR algorithms were not included in this investigation to investigate the linear PLSR performance when the dataset is highly irregular [52]. The selected technique is preferred to be both successful and easy to use, so as to avoid the need to develop a time-consuming and difficult approach to achieving the desired results.

Based on the minimum PRESS value employing leaveone-out internal cross-validation to select an appropriate number of latent variables (LVs) [53], the PLSR model was built using two LVs to predict the changes in the response variable RVP from the variance of independent variables of physical properties of gasoline (density and distillation curve temperatures). The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed subsections.

Figure 3 shows the results of PLSR, where the value of coefficient of determination (R^2) for the calibration set is 0.438, using two LVs. From Figure 3(a), the plotted predicted RVP values against reference RVP values give an indication of a high nonlinearity of the dataset for both calibration and prediction sets. As a fitted regression line, Figure 3(a) shows that the R^2 of the PLSR model is unconvincing. Figure 3(b) indicates that applying PLSR results with prediction set equals 4.805, rendering the approach can achieve reasonable prediction. The residual plot in Figure 3(b), generated using the PLSR model, provides an evaluation of prediction errors. It demonstrates that relying



FIGURE 3: Prediction performance using PLSR: (a) experimental RVP values against predicted RVP values in calibration and prediction sets and (b) the residual of predicted RVP for calibration set.

on R^2 as the sole determinant factor for assessing model performance in prediction results is inaccurate. It agrees with the proposition that the implementation of R^2 has several limitations in assessing model fit when dealing with complex models and multiple outlier cases [54, 55].

The PLSR model's inspection of the residuals' normality distribution, which was developed from the RVP prediction, is shown in Figure 4. The histogram of generated residuals is displayed in Figure 4(a). The residual distribution is nearly symmetrical and has minimal skewing to the right, indicating that the residuals are primarily normally distributed. The results for the probability plot of the residuals in Figure 4(b), which show that the close line-up of ordered residuals on the normality recognizing line confirms the residuals' normal distribution, provide support for this finding.

Figure 4(c) shows the relationships for the first and second latent variables (LV1 and LV2) and between the



FIGURE 4: Normality assumption plots for PLSR prediction performance: (a) histogram plot for residuals and (b) probability plot of residual for semistudentized residuals ($e_{(k)}/S$), where S is the standardized residuals, k is the kth ordered residual, and $Z_{(k)}$ is the inverse normal distribution of ordered residuals, and (c) two-dimensional plot of PCA latent variables LV1 and LV2 loading with PLSR model for response and explanatory variables.

response variable (RVP) and density, and distillation curve temperatures of the GOBs that were studied in the PLSR model. There is a robust positive enriched relation between RVP and IBP and T90 with positive loading, and the first latent variable (LV1) accounts for 26.03% of the total variance. A negative and statistically significant loading exists between RVP, S-Dens, T10, T50, and FBP. The negative correlation between FBP and RVP is slightly less. S-Dens, IBP, and T90 have negative loading with RVP for the second latent variable (LV2), while FBP, T50, and T10 of the independent variables are positively associated with RVP. The correlation between T90 and RVP for LV2 is slightly lower. 16.83% of the total variation is explained by LV2. 3.2. Nonlinear Regression (NLR). Here, another model was developed utilizing the NLR calibration method, and its performance is shown in Figure 5; this model predicts RVP based on 904 data and six independently observable gasoline physical parameters. As shown in Figure 5(a), the fitting method has not been realistically improved, but the R^2 value is 0.452. The residual errors in NLR model predictions are displayed in Figure 5(b), with the RSE value dropping to 4.747.

Figure 6 shows the results of verifying the normality dis tribution of the residuals produced by the NLR model for RVP prediction. Figure 6(a) displays the analyzed histogram of the generated residuals. The distribution of the residuals is



FIGURE 5: Prediction performance using NLR: (a) experimental RVP values against predicted RVP values in calibration and prediction sets and (b) the residual of predicted RVP for calibration set.



FIGURE 6: Normality assumption plots for NLR prediction performance: (a) histogram plot for residuals and (b) probability plot of residual for semistudentized residuals.

nearly symmetrical, with a relative skew to the right, indicating that the residuals mostly follow a normal distribution. The results shown in Figure 6(b) for the probability plot of the residuals corroborate this finding. It is clear that the normal distribution of the ordered residuals is generally aligned with the normality matching line. 3.3. Nonparametric Regression (NPR). The NPR calibration method is used to establish a model for the response variable RVP, using the same observations and independent variable as the PLSR and NLR methods. Figure 7 displays the results of the NPR (LOWESS) regression model's fitting and residual evaluation. As shown in Figure 7(a), the R^2 value has



FIGURE 7: Prediction performance using NPR (LOWESS) regression model: (a) experimental RVP values against predicted RVP values in calibration and prediction sets and (b) the residual of predicted RVP for calibration set.



FIGURE 8: Normality assumption plots for NPR prediction performance: (a) histogram plot for residuals and (b) probability plot of residual for semistudentized residuals.

increased to 0.612, indicating that the fitting process has been enhanced as the model partially overcomes the difficulty of dataset dispersion and reduces the impact of outliers. In Figure 7(b), we can see that the RSE value for the predictions made by the NLR model at the calibration set has decreased to 3.971, from a previous value of 4.805 made by PLSR. Figure 8 illustrates the check-up of normality distribution of residuals generated from RVP prediction using the NPR model. Figure 8(a) shows the examined histogram of generated residuals, where as it can be seen the residuals, and distribution is close to be symmetrical with relative skewing to the right, suggesting that residuals are generally normally distributed. This outcome is supported with the

Investigation approach	MAPE (%)	MAD (kPa)	RMSE (kPa)	No. of samples	No. of input variables
PLSR (this work)	7.752	4.131	5.321	904	6
NLR (this work)	7.661	4.096	5.253	904	6
NPR (LOWESS) (this work)	7.020	3.933	5.163	904	6
PC-SAFT ^a	4.030			7	80
PSRK ^a	7.610			7	80
Wilson EqLS ^b	5.640			24	24
Wilson EqLI ^b	4.410			24	24
UNIFAC ^c	1.0 - 10.0			5	16
LSSVM ^d (without oxygenates)			9.231	362	23
ANN ^e (without oxygenates)		2.280		362	27
SAFT- γ EoS ^f	4.36-10.59			7	18
CPA ^g	5.500			4	226

TABLE 3: Accuracy of PLSR, NLR, and NPR models for predicting gasoline RVP and comparison with previous studies.

^aPC-SAFT is perturbed-chain statistical associating fluid theory, and PSRK is predictive Soave-Redlich-Kwong equation adopted from Vella and Marshall [56], and they used data of gasoline and methanol blend taken from the experimental work of Andersen et al. [57]; ^bLI is Lagrange interpolating polynomial statistical method, and LS is least squares statistical fitting method adopted from Pumphrey et al. [58] for gasoline and isopropanol blend samples; ^cUNIFAC is universal quasichemical functional group activity coefficients adopted from Hatzioannidis et al. [59] for gasoline containing MTBE, methanol, ethanol, and isopropanol blends; ^dLSSVM is least squares support vector machine adopted from Kamari et al. [14] (considered for gasoline only); ^eANN is artificial neural network adopted from Albahri et al. [60] and considered for gasoline only; ^fSAFT-*y* EoS is statistical associating fluid theory-Mie model equation of state adopted from Landera et al. [61]; ^gCPA is cubic plus association model adopted from Gaspar et al. [7].

results presented in Figure 8(b) for probability plot of residuals, and it can be observed that the normal distribution of ordered residuals is mostly lined up on normality recognizing line, except minor diversion and the ends.

3.4. Comparison of Calibration Techniques. A comparison between the developed models, PLSR, NLR, and NPR, provides information about model performance on unseen data. To accomplish this purpose, various error detection criteria of calibration (RMSEC) and prediction (RMSEP) were used on calibration and prediction sets for testing their prediction capacity. The results of RMSEC are 4.790, 4.739, and 3.968 and for RMSEP are 6.235, 6.149, and 6.029 for PLSR, NLR, and NPR, respectively. The NPR model has lower prediction errors, in calibration and in prediction sets.

The density of gasoline is generally correlated with the molecular size of the mixture. However, other factors, such as temperature, pressure, and the presence of oxygenates, can also alter the density of a hydrocarbon mixture, so the relationship between the two is not always straightforward. In addition, the volatility of a GOB, which is typically correlated to the intensity of the intermolecular interactions within the mixture and which can be represented by the parameters of the distillation curve (IBP, T10, T50, T90, and FBP), is greatly diverted by the impacts of oxygenates. As a result, there are restrictions on what can be accomplished when attempting to predict RVP using linear regression, PLSR, or even NLR. The results of a comparison of the PLSR, NLR, and NPR models under study make this quite obvious.

Table 3 provides a summary of the outcomes of the used, PLSR, NLR, and NPR (LOWESS) regression models in terms of prediction performance and error evaluation. Results were compared with those of prior studies found in the literature to get a clearer picture of the efficacy of the approaches used. In order to achieve a significant RVP detection of GOB mixtures, as shown in Table 3, comparing with previous works [7, 14, 56–61] for RVP of gasoline containing oxygenates, the used prediction techniques demonstrate reasonable ability to overcome the described barriers despite the fact that just elementary input variables are employed and a higher degree of credibility due to the large number of samples used. In the case of extremely dispersed and scattered data, the results show that the R^2 value is not a determining factor for prediction evaluation.

Model performance on the identified data can be obtained by comparing the developed PLSR, NLR, and NPR models against existing studies in literature. To validate the results of the generated models, a number of error detection metrics, including RMSE, MAPE, and MAD, were applied to total data sets. Significant and realistic RVP prediction potentials are shown by the outcome values for PLSR, NLR, and NPR in Table 3. When compared to prior results, the constructed models, particularly the LOWESS model, are able to produce predictions that are on par with or better than those found in the literature in terms of error metrics. Despite the fact that some of the selected studies were conducted on gasoline without oxygenates, those of ANN and SVM methods, the results can still reach reasonable predictions when evaluating the performance even though there is a great deal of complexity and difficulty associated with gasoline-oxygenate blends. The constructed models can be considered simple and dependable when considering the simplicity of the input data. Rather than relying on complex spectrum, activity group, and chemical composition input data, which are necessary for the predictions, this is proven

by the use of straightforward physical properties. The larger size of dataset used in this study lends more credence to the proposed models.

4. Conclusions

Comparative analysis of regression models for predicting RVP in gasoline-oxygenate blends has been made in this study using PLSR, NLR, and NPR models. The regression models showed promising results and provide a cheap and straightforward alternative for RVP prediction. NPR model showed more accurate results as it overcame dataset dispersion by reducing the impact of outliers, eliminating parametric constraints, and allowing for a new data structure. The research compared the PLSR, NLR, and NPR models for forecasting RVP of GOB from a parametric and nonparametric perspective. DModX and DModY analyses were used to remove anomalies from the data set. The statistical analysis and error detection indicators such as residual quantification RSE, RMSEC, and RMSEP were used to compare the three models' performance and prediction capability. The NPR (LOWESS) regression model showed slightly more accurate results in predicting RVP, as it overcame the difficulty of dataset dispersion. Parametric limitations are removed in the NPR (LOWESS) regression model, allowing for a new structure to form for X variables that are derived from the data. The limitation faced in conducting this work is briefed by high dimensionality and irregularity of the gasoline compositions and ratio dataset, which is difficult to any model for developing a viable regression and generalizing of the findings. The outlooks developed here suggest that the established models, slightly higher for NPR, can be a useful choice for consistently predicting RVP in gasoline-oxygenate blends by giving a cheap and straightforward alternative to the existing complex approaches, since a minimal set of inputs is required for more credible findings in highly irregular datasets. Further investigation into the use of other algorithms is needed to improve the prediction accuracy of RVP in GOB and explore other regression modelling strategies that can handle high-dimensional and irregular datasets to develop comprehensive and more accurate prediction models by implementing other modelling strategies and assumptions on the potential limitations and underlying data structure.

Data Availability

Data generated or analyzed during this study are available from the author upon reasonable request.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The author expresses his gratitude to the Oil and Metals Directorate of the Garmian Region, particularly General Dir. Mr. Yadgar Fattah and Chief Engr. Habib Faraj, for providing the required laboratories and facilities to complete the experimental portion of this study.

References

- E. Stauffer, J. A. Dolan, and R. Newman, "Chapter 7 flammable and combustible liquids," in *Fire Debris Analysis*, E. Stauffer, J. A. Dolan, and R. Newman, Eds., pp. 199–233, Academic Press, 2008.
- [2] V. S. Pinto, F. F. Gambarra-Neto, I. S. Flores, M. R. Monteiro, and L. M. Lião, "Use of 1H NMR and chemometrics to detect additives present in the Brazilian commercial gasoline," *Fuel*, vol. 182, pp. 27–33, 2016.
- [3] I. G. Petrisor, "Use of oxygenates to date a gasoline release," *Environmental Forensics*, vol. 7, no. 2, pp. 103-104, 2006.
- [4] S. A. Stout, G. S. Douglas, and A. D. Uhler, "Chemical fingerprinting of gasoline and distillate fuels," in *Standard Handbook Oil Spill Environmental Forensics*, S. A. Stout and Z. Wang, Eds., pp. 509–564, Academic Press, 2016.
- [5] H. M. Issa, "An investigation of the comparative performance of diverse vapor pressure prediction techniques of sour natural gas," *Petroleum Science and Technology*, vol. 33, no. 13-14, pp. 1443–1448, 2015.
- [6] P. Mello, F. Wildner, G. S. de Andrade, R. Cataluña, and R. da Silva, "Combustion time of the oxygenated and nonoxygenated fuels in an Otto cycle engine," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 36, no. 2, pp. 403–410, 2014.
- [7] D. J. Gaspar, S. D. Phillips, E. Polikarpov et al., "Measuring and predicting the vapor pressure of gasoline containing oxygenates," *Fuel*, vol. 243, pp. 630–644, 2019.
- [8] G. Dhamodaran and G. S. Esakkimuthu, "Experimental measurement of physico-chemical properties of oxygenate (DIPE) blended gasoline," *Measurement*, vol. 134, pp. 280–285, 2019.
- [9] ASTM D323 20a, Standard Test Method for Vapor Pressure of Petroleum Products (Reid Method), ASTM International, 2020.
- [10] ASTM D5191-20, Standard Test Method for Vapor Pressure of Petroleum Products and Liquid Fuels (Mini Method), ASTM International, 2020.
- [11] A. Biancolillo and F. Marini, "Chapter four chemometrics applied to plant spectral analysis," in *Comprehensive Analytical Chemistry*, J. Lopes and C. Sousa, Eds., pp. 69–104, Elsevier, 2018.
- [12] G. Mendes, H. G. Aleme, and P. J. S. Barbeira, "Reid vapor pressure prediction of automotive gasoline using distillation curves and multivariate calibration," *Fuel*, vol. 187, pp. 167– 172, 2017.
- [13] A. Tatar, A. Barati, A. Najafi, and A. H. Mohammadi, "Radial basis function (RBF) network for modeling gasoline properties," *Petroleum Science and Technology*, vol. 37, no. 11, pp. 1306–1313, 2019.
- [14] A. Kamari, M. Sattari, A. H. Mohammadi, and D. Ramjugernath, "Modeling of the Properties of Gasoline and Petroleum Fractions Using a Robust Scheme," *Petroleum Science and Technology*, vol. 33, no. 6, pp. 639–648, 2015.
- [15] F. A. Honorato, B. de Barros Neto, M. F. Pimentel, L. Stragevitch, and R. K. Galvão, "Using principal component analysis to find the best calibration settings for simultaneous spectroscopic determination of several gasoline properties," *Fuel*, vol. 87, no. 17, pp. 3706–3709, 2008.

- [16] J. Li, I. A. Karimi, and R. Srinivasan, "Recipe determination and scheduling of gasoline blending operations," *AICHE Journal*, vol. 56, no. 2, pp. 441–465, 2010.
- [17] I. Hassouneh, T. Serra, B. K. Goodwin, and J. M. Gil, "Nonparametric and parametric modeling of biodiesel, sunflower oil, and crude oil price relationships," *Energy Economics*, vol. 34, no. 5, pp. 1507–1513, 2012.
- [18] A. Donnelly, B. Misstear, and B. Broderick, "Real time air quality forecasting using integrated parametric and nonparametric regression techniques," *Atmospheric Environment*, vol. 103, pp. 53–65, 2015.
- [19] ASTM D4057 19, Standard Practice for Manual Sampling of Petroleum and Petroleum Products, ASTM International, 2019.
- [20] ASTM D7345-17, Standard Test Method for Distillation of Petroleum Products and Liquid Fuels at Atmospheric Pressure (Micro Distillation Method), ASTM International, 2017.
- [21] ASTM D86-20a, Standard Test Method for Distillation of Petroleum Products and Liquid Fuels at Atmospheric Pressure, ASTM International, 2020.
- [22] ASTM D4052 18a, Standard Test Method for Density, Relative Density, and API Gravity of Liquids by Digital Density Meter, ASTM International, 2018.
- [23] ISO 4259-3, Petroleum and Related Products Precision of Measurement Methods and Results — Part 3: Monitoring and Verification of Published Precision Data in Relation to Methods of Test, ISO – International Organization for Standardization, 2020.
- [24] ISO 5725-2, Accuracy (Trueness and Precision) of Measurement Methods and Results — Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a standard Measurement Method, ISO – International Organization for Standardization, 2019.
- [25] ASTM E1655-17, Standard Practices for Infrared Multivariate Quantitative Analysis, ASTM International, 2018.
- [26] L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, and P. Gramatica, "Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs," *Environmental Health Perspectives*, vol. 111, no. 10, pp. 1361–1375, 2003.
- [27] A. F. Silva, J. Vercruysse, C. Vervaet et al., "In-depth evaluation of data collected during a continuous pharmaceutical manufacturing process: a multivariate statistical process monitoring approach," *Journal of Pharmaceutical Sciences*, vol. 108, no. 1, pp. 439–450, 2019.
- [28] R. W. Kennard and L. A. Stone, "Computer aided design of experiments," *Technometrics*, vol. 11, no. 1, pp. 137–148, 1969.
- [29] R. E. L. Abreu, J. E. M. Paz, A. C. Silva, M. J. C. Pontes, and S. G. Lemos, "Ethanol fuel adulteration with methanol assessed by cyclic voltammetry and multivariate calibration," *Fuel*, vol. 156, pp. 20–25, 2015.
- [30] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986.
- [31] H. M. Issa, "Prediction of octane numbers for commercial gasoline using distillation curves: a comparative regression analysis between principal component and partial least squares methods," *Petroleum Science and Technology*, vol. 42, no. 10, pp. 1233–1249, 2024.
- [32] P. Gemperline, *Practical Guide to Chemometrics*, CRC Press, Boca Raton, USA, 2006.

- [33] T. Zeng, C. Zhang, M. Hu et al., "Modelling and predicting energy consumption of a range extender fuel cell hybrid vehicle," *Energy*, vol. 165, pp. 187–197, 2018.
- [34] W. F. C. Rocha, C. B. Prado, and N. Blonder, "Comparison of chemometric problems in food analysis using non-linear methods," *Molecules*, vol. 25, no. 13, p. 3025, 2020.
- [35] N. Maheshwari, C. Balaji, and A. Ramesh, "A nonlinear regression based multi-objective optimization of parameters based on experimental data from an IC engine fueled with biodiesel blends," *Biomass and Bioenergy*, vol. 35, no. 5, pp. 2171– 2183, 2011.
- [36] M. Kooi and A. A. Koelmans, "Simplifying microplastic via continuous probability distributions for size, shape, and density," *Environmental Science & Technology Letters*, vol. 6, no. 9, pp. 551–557, 2019.
- [37] J. S. Simonoff, "Nonparametric regression," in *Smoothing Methods in Statistics*, J. S. Simonoff, Ed., pp. 134–214, Springer, New York, 1996.
- [38] L. Fahrmeir, T. Kneib, S. Lang, and B. Marx, *Regression: Models, Methods and Applications*, Springer, Berlin Heidelberg, 2013.
- [39] A. Kibria, S. B. Akhundjanov, and R. Oladi, "Fossil fuel share in the energy mix and economic growth," *International Review* of Economics & Finance, vol. 59, pp. 253–264, 2019.
- [40] D. Alviso, G. Artana, and T. Duriez, "Prediction of biodiesel physico-chemical properties from its fatty acid composition using genetic programming," *Fuel*, vol. 264, article 116844, 2020.
- [41] W. S. Cleveland and S. J. Devlin, "Locally weighted regression: an approach to regression analysis by local fitting," *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 596–610, 1988.
- [42] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829–836, 1979.
- [43] W. S. Cleveland, E. Grosse, and W. M. Shyu, "Local regression models," in *Statistical models in S*, T. J. Hastie, Ed., pp. 309– 376, Routledge, 2017.
- [44] H. Wang and L. Chan, "Wavelet packet transform-assisted least squares support vector machine for gear wear degree diagnosis," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9889933, 9 pages, 2021.
- [45] M. Hilgemann, P. C. Nascimento, D. Dias et al., "Fast assessment of diene values in pyrolysis gasoline by UV-visible spectroscopy and multivariate calibration," *Journal of Petroleum Science and Engineering*, vol. 78, no. 2, pp. 283–287, 2011.
- [46] B. Kehimkar, B. A. Parsons, J. C. Hoggard, M. C. Billingsley, T. J. Bruno, and R. E. Synovec, "Modeling RP-1 fuel advanced distillation data using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry and partial least squares analysis," *Analytical and Bioanalytical Chemistry*, vol. 407, no. 1, pp. 321–330, 2015.
- [47] O. Sanni, O. Adeleke, K. Ukoba, J. Ren, and T. C. Jen, "Prediction of inhibition performance of agro-waste extract in simulated acidizing media via machine learning," *Fuel*, vol. 356, article 129527, 2024.
- [48] L. Guan, X. L. Feng, Z. C. Li, and G. M. Lin, "Determination of octane numbers for clean gasoline using dielectric spectroscopy," *Fuel*, vol. 88, no. 8, pp. 1453–1459, 2009.
- [49] Z. Li, Y. Yu, X. Pan, and M. N. Karim, "Effect of dataset size on modeling and monitoring of chemical processes," *Chemical Engineering Science*, vol. 227, article 115928, 2020.

- [50] I. Barra, M. Kharbach, E. M. Qannari, M. Hanafi, Y. Cherrah, and A. Bouklouze, "Predicting cetane number in diesel fuels using FTIR spectroscopy and PLS regression," *Vibrational Spectroscopy*, vol. 111, p. 103157, 2020.
- [51] D. D. Silalahi, H. Midi, J. Arasan, M. S. Mustafa, and J. P. Caliman, "Kernel partial least square regression with high resistance to multiple outliers and bad leverage points on near-infrared spectral data analysis," *Symmetry*, vol. 13, no. 4, p. 547, 2021.
- [52] J. Kohonen, S.-P. Reinikainen, K. Aaljoki, and A. Höskuldsson, "Non-linear PLS approach in score surface," *Chemometrics and Intelligent Laboratory Systems*, vol. 97, no. 2, pp. 159–163, 2009.
- [53] J. L. Martínez, H. Saulo, H. B. Escobar, and J. Leao, "A new model selection criterion for partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 169, pp. 64–78, 2017.
- [54] S. A. Book and P. H. Young, "The trouble with R2," *Journal of Parametrics*, vol. 25, no. 1, pp. 87–114, 2006.
- [55] H.-P. Piepho, "An adjusted coefficient of determination (R2) for generalized linear mixed models in one go," *Biometrical Journal*, vol. 65, no. 7, article 2200290, 2023.
- [56] J. R. Vella and B. D. Marshall, "Prediction of the distillation curve and vapor pressure of alcohol-gasoline blends using pseudocomponents and an equation of state," *Industrial & Engineering Chemistry Research*, vol. 59, no. 17, pp. 8361– 8373, 2020.
- [57] V. F. Andersen, J. E. Anderson, T. J. Wallington, S. A. Mueller, and O. J. Nielsen, "Vapor pressures of alcohol–gasoline blends," *Energy & Fuels*, vol. 24, no. 6, pp. 3647–3654, 2010.
- [58] J. A. Pumphrey, J. I. Brand, and W. A. Scheller, "Vapour pressure measurements and predictions for alcohol–gasoline blends," *Fuel*, vol. 79, no. 11, pp. 1405–1411, 2000.
- [59] I. Hatzioannidis, E. C. Voutsas, E. Lois, and D. P. Tassios, "Measurement and prediction of Reid vapor pressure of gasoline in the presence of additives," *Journal of Chemical & Engineering Data*, vol. 43, no. 3, pp. 386–392, 1998.
- [60] T. A. Albahri, "Specific gravity, RVP, octane number, and saturates, olefins, and aromatics fractional composition of gasoline and petroleum fractions by neural network algorithms," *Petroleum Science and Technology*, vol. 32, no. 10, pp. 1219–1226, 2014.
- [61] A. Landera, N. Mac Dowell, and A. George, "Development of robust models for the prediction of Reid vapor pressure (RVP) in fuel blends and their application to oxygenated biofuels using the SAFT- γ approach," *Fuel*, vol. 283, article 118624, 2021.