energies

Article

Molecular Reconstruction Method Based on NIR Spectroscopy for Reformates

Mingyuan Cheng and Linzhou Zhang *

State Key Laboratory of Heavy Oil Processing, Petroleum Molecular Engineering Center (PMEC),
China University of Petroleum, Beijing 102249, China
* Correspondence: lzz@cup.edu.cn

Abstract: With the increasing attention to environmental protection and strict national standards, the production of high-quality clean gasoline is more and more required in modern refineries. At present, the molecular-level online modeling of the gasoline blending process is rarely reported due to the lack of a proper characterization method. In this study, we have developed a molecular reconstruction method based on near-infrared (NIR) spectroscopy. We established a library that comprises an NIR spectrum and molecular composition of known reformates. We developed the sample selection algorithm that finds the sample with an NIR spectrum close to the target sample. The method-solving algorithm based on multiple linear regression was used to obtain the molecular composition of gasoline quickly. The method combines online measurement methods and molecular management technology to achieve online composition reconstruction. Seventy-five groups of reformates were collected to verify the feasibility of the method, and the results showed that the method predicted well.

Keywords: near-infrared spectroscopy; molecular reconstruction method; sample selection algorithm; multiple linear regression; online measurement method

check for updates

Citation: Cheng, M.; Zhang, L. Molecular Reconstruction Method Based on NIR Spectroscopy for Reformates. *Energies* **2023**, *16*, 2982. https://doi.org/10.3390/en16072982

Academic Editor: Pavel A. Strizhak

Received: 1 March 2023
Revised: 22 March 2023
Accepted: 23 March 2023
Published: 24 March 2023

1. Introduction

Automotive gasoline occupies an important position in transportation fuels, while the combustion of gasoline in internal combustion engines emits a large number of pollutants and is an important source of environmental pollution. As the attention to environmental protection increases and national standards become stricter, more stringent production standards for motor gasoline are proposed in China. Table 1 shows the latest production standards for motor gasoline in China. Among them, the VIA standard was officially implemented on 23 December 2016, and the implementation date of the nationwide VIA standard was 1 July 2020. The implementation date of VIB is 1 July 2023, and the difference between the two is mainly reflected in the olefin content. From the VIA and VIB standards, we can see that the quality requirements for motor gasoline in China are very strict, which puts very high demands on the current gasoline production process in modern refineries.

Gasoline blending, the final process in gasoline production, has been a hot topic of research due to its importance and complexity. Gasoline blending is necessary to meet the ever-escalating quality standards of gasoline and maximize the efficiency of refineries. However, the non-linear superposition of the bulk properties of gasoline in the blending process makes gasoline blending very difficult, so a blending method is needed to guide the blending process to improve the success rate.

The traditional blending method is based on the bulk properties of feedstocks and products [1–7]. Taking the Ethyl-RT70 method as an example, it is based on the sensitivity of feedstocks and calculates the octane number of products by using the olefin content and aromatic content of feedstocks. The traditional blending method needs to obtain the bulk properties of feedstocks, and most of the bulk properties are obtained by experimental methods, which are time-consuming and labor-intensive. In actual applications, the method parameters have to be revised and trained continuously, which cause a great inconvenience for blending.

Table 1. The China motor gasoline VI standard.

Bulk Property	VIA	VIB	Bulk Property	VIA	VIB
RON	89 92 95	89 92 95	Residual amount (volume fraction)/%	2	2
(RON + MON)/2	84 87 90	84 87 90	Sulfur content/(mg/kg)	≤10	≤10
Vapor pressure, kPa (winter)	45–85	45–85	Aromatic content (volume fraction)/%	≤35	≤35
Vapor pressure, kPa (summer)	40–65	40–65	Benzene content (volume fraction)/%	≤0.8	≤0.8
Distillation range:			Olefin content (volume fraction)/%	≤18	≤15
T10/°C	≯70	≯70	Manganese content /(mg/L)	≤2	≤2
T50/°C	≯110	≯110	Oxygen content (mass fraction)/%	≤2.7	≤2.7
T90/°C	≯190	≯190	Methanol content (mass fraction)/%	≤0.3	≤0.3
Final distillation point/°C	≯205	≯205	Residual amount (volume fraction)/%	2	2

Instead of blending calculation based on measured bulk properties, near-infrared (NIR) spectroscopy has been widely used in online blending optimization. The infrared spectrum can be broadly divided into the near-infrared (NIR) spectrum in the wavelength range of 4000–12,800 cm^{-1} and the mid-infrared (MIR) spectrum in the wavelength range of 400–4000 cm^{-1}. The MIR spectrum reflects molecular structure information as "fingerprinting", which can be used in combination with chemometric methods to predict the bulk properties of petroleum fractions [8,9]. However, industrial applications of MIR spectroscopy in the gasoline blending process are rarely reported, mainly due to the inability to apply quartz optical fibers with good transmission performance at low cost. With the development of chemometric methods, NIR spectroscopy has been gradually applied to the analysis of petroleum samples [10–21].

The advantage of NIR spectroscopy is the ability to apply quartz optical fibers for the transmission of information, so NIR spectrometers can be placed in laboratories far from the field. This advantage allows NIR spectroscopy to be adapted to complex and harsh environments, and the online measurement method by NIR spectroscopy is widely used in the gasoline blending process [22–28]. The online measurement method based on NIR spectroscopy mainly includes three aspects:

(1) The NIR spectrum of the gasoline stream is detected by the probe, installed in the field gasoline stream pipeline;

(2) The obtained NIR spectrum is transmitted to the NIR spectrometer located in the laboratory through the quartz fiber;

(3) The NIR spectrometer analyzes and processes the obtained NIR spectrum.

Due to the ultra-high detection speed of NIR spectroscopy, the time required to obtain the bulk properties of the feedstocks is significantly reduced by the combination with chemometric methods. However, the current method of calculating the bulk properties of the product still follows the mathematical algorithm of the traditional blending method, therefore, the disadvantage of poor parameter universality still cannot be solved.

Recently, molecular-level modeling of the gasoline blending process has attracted growing attention. Unlike the bulk properties, the molecular composition of the feedstocks shows a linear additivity. If the relationship between bulk property and molecular composition is developed, the molecular model shows higher universality and can be linked to other molecular-level process models. In previous research, the gasoline molecular modeling based on detailed molecular composition from gas chromatography (GC) analysis was developed [29–32]. Various petroleum molecule characterization frameworks have been proposed to convert petroleum molecules into a computer language that facilitates mathematical modeling at the molecular level [33–36]. Based on these frameworks, various molecular-level bulk property prediction methods have been developed [37–42]. Therefore, the rapid method to obtain the molecular composition of gasoline is a key challenge for the development of online molecular-level modeling.

Due to the long measurement time for GC, the online gasoline molecular compositional modeling based on GC analysis is difficult. If the molecular composition can be inferred from NIR spectroscopy, the online molecular-level optimization of the gasoline blending process can be achieved. This study developed a molecular reconstruction method based on NIR spectroscopy to solve this problem. A sample library containing samples' NIR spectra and molecular composition was established and a sample selection algorithm based on Euclidean distance was developed. The method finally uses the method-solving algorithm based on multiple linear regression to obtain the molecular composition of the target samples. A molecular reconstruction method based on NIR spectroscopy provides a method that can combine the online measurement method and gasoline molecular management technology, and the method can meet the needs of the online blending process in refineries. This facilitates the modeling of subsequent gasoline blending methods and solves the drawback of poor generalizability of parameters.

The reforming process is one of the main processes in the petrochemical industry. Its main feedstock is naphtha. Under certain temperature and pressure conditions, a bifunctional catalyst is used to rearrange the structure of hydrocarbons and one of the main products is high-octane gasoline, also known as reformates. The process mainly converts alkanes to aromatics through a reforming reaction, thus converting low-octane naphtha to high-octane gasoline. The research octane number (RON) of reformates is between 95 and 105, with low content of sulfur, nitrogen, oxygen, and other impurities and olefin, which is helpful to improve the octane number of products and reduce the olefin content. It is a very important high-octane number clean gasoline blending component at present. Therefore, reformates are chosen as the main object of this paper.

2. Method

2.1. Fundamentals

To establish the molecular reconstruction method based on NIR spectroscopy, the following hypothesis is proposed: each gasoline is mixed from several base gasolines. We preliminarily believe that the linear superposition of NIR spectra can represent the mixing process of the corresponding gasolines. Therefore, multiple linear regression is a good choice for obtaining the mixing ratio of the underlying sample.

The method can use multiple linear regression to fit the known NIR spectra of test samples by the known NIR spectra of base samples, and the parameters obtained by the multiple linear regression can be defined as the mixing ratio of base samples mixed with test samples. Based on the known molecular composition of base samples, the method can easily find the molecular composition of test samples by the mixing ratio.

Based on the above basic assumptions, first of all, the method needs to have a sample library consisting of base samples, which need to have known NIR spectra and molecular composition, and then preprocess the spectra and select suitable base samples according to the characteristics of test samples, after which the mixing ratio is obtained by multiple linear regression, and finally, the molecular composition of test samples is obtained. Figure 1 shows the model-building process.

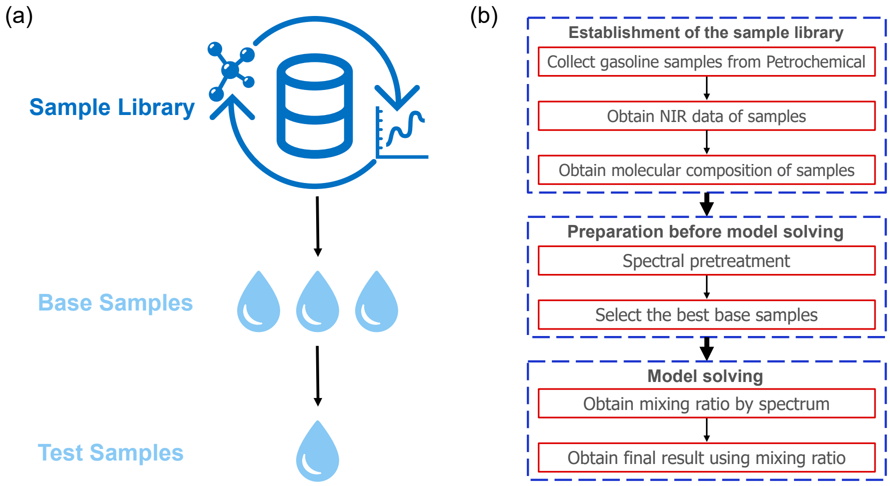

Figure 1. The diagram of predicting molecular composition from NIR spectrum. (**a**) Sample selection and test sample composition prediction. (**b**) schematic diagram for the algorithm used in the model.

2.2. Sample Library

To meet the basic requirements of the method, a total of 75 sets of reformates from Jingbo Petrochemical were collected, and the molecular composition of gasoline was obtained by using the gasoline molecular compositional model based on GC analysis. The gasoline was analyzed by NIR spectroscopy with a resolution of about 7 cm^{-1}, and the wave number region was about 4300–14,000 cm^{-1}. To verify the feasibility of the method, the collected data were divided into base samples and test samples in the ratio of 9:1 according to the collection time, with the earlier historical samples collected as the sample library of base samples and the later samples collected as the test samples. Compared with the MIR spectrum, the bands of the NIR spectrum overlap significantly, and a single band may be composed of multiple and combined frequencies of several fundamental frequencies, so it is difficult to accurately attribute them to the specific molecular structure or composition. Figure 2 shows the NIR spectrum in the sample library. Since the reformates are mainly composed of aromatics, the NIR spectrum in the figure is mostly the absorption peak of aromatics. From 5500 cm^{-1} to 6000 cm^{-1} is the first order frequency doubling absorption peak of the C-H group in the aromatic ring. The absorption peak at 4600 cm^{-1} is the combined frequency absorption peak of C-C stretching vibration and C-H stretching vibration on the aromatic ring. The double peak at 4060 cm^{-1} is usually the absorption peak of the substituted benzene C-H group's combined frequency, which is the combined frequency of its fundamental frequency.

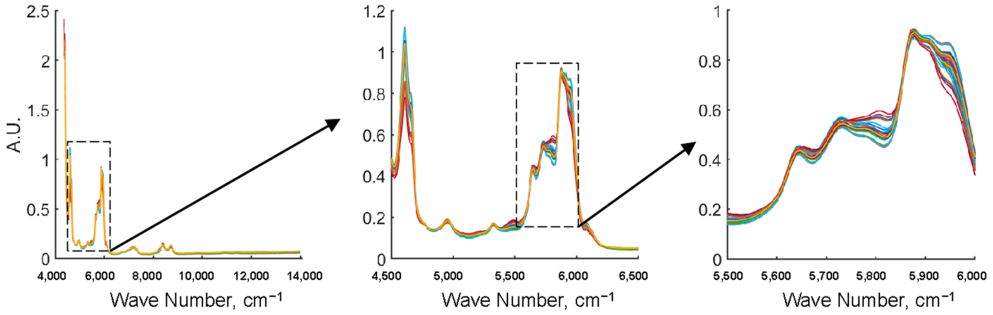

Figure 2. Original NIR spectra of all the samples in the library.

2.3. Preprocessing of Spectra

In addition to its chemical information, a spectrum contains other irrelevant information and noise, such as electrical noise, the background of the sample, and stray light. Therefore, a preprocessing method aimed at eliminating extraneous information and noise from spectra becomes critical and necessary [43]. It can be seen from Figure 2 that the NIR spectra of the base samples have serious baseline drift. In this study, the first-order derivative method was used to eliminate the baseline drift of the NIR spectra. To avoid the first-order derivative method amplifying the noise in the NIR spectra, the NIR spectra were smoothed and denoised before the first-order derivative method, to maximize the noise elimination in the NIR spectra. Therefore, the spectrum preprocessing method chosen in this study was a combination of smoothing denoising and the first-order derivative method, and the results of the sample library preprocessing are shown in Figure 3. From the figure, it can be seen that after first-order derivative treatment, the baselines of all the spectra were converted to values that were very close to zero (range 10,000 to 14,000 cm^{-1}). The spectrum preprocessing method chosen in this study solves the problem of baseline drift of the original spectra very well. Additionally, from the scaled-up figure of the spectra (left part of Figure 3), it can be seen that the difference in the NIR spectra of different samples is still retained, which can be used for sample selection and spectral manipulation.

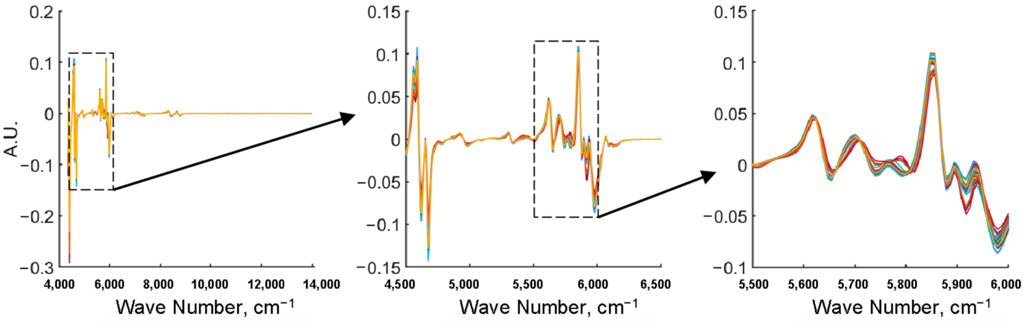

Figure 3. NIR spectra of all the samples in the library after preprocessing.

2.4. Sample Selection Algorithm

Due to a large number of base samples in the sample library, it is especially critical to select suitable base samples in the sample library according to the characteristics of the test samples, and the suitability of selected base samples for the test samples directly affects the final prediction results of the method. In this study, a sample selection algorithm based on Euclidean distance was developed, and Equation (1) is the calculation process of the Euclidean distance, where $Num.p$ is the total number of spectral points, x^{test} is the absorbance of the spectroscopy point of the test sample, x^{base} is the absorbance of the spectral point of the base sample. The method selects the base samples with the smallest Euc as the best base samples. In short, the sample selection method is used to select the samples in the library that have NIR spectra very close to the target sample. Since the molecular composition determines the corresponding NIR spectrum, the base sample that has a similar NIR spectrum is assumed to have similar molecular composition.

$$Euc = \sum_{i=1}^{Num.p} \left| x_i^{test} - x_i^{base} \right| \tag{1}$$

The number of best base samples should be appropriate to ensure the validity of the following base sample blending algorithm. If the number of base samples is too small, the variation of the molecular composition will be limited. It will lead to a smaller possible molecular compositional distribution space of test samples. If the number of base samples is too large, there is a possibility of interference terms, which greatly reduce the prediction accuracy of the method. After several trial-and-error tests, the number of base samples in this method was set to 5. The solution of the model was stable and the accuracy of the predicted molecular composition was acceptable.

2.5. Base Oil Blending Algorithm

It is assumed that the unknown target samples are a blend of the chosen base samples. After the base sample selection step, the final step is to calculate the blending ratio of the base sample to obtain a new blend, which has an NIR spectrum that is close to the measured data of the target sample. If the NIR spectrum is almost the same, the molecular composition of the blend is close to the target sample. Since the molecular composition of the base sample is known, the molecular composition of the blend can be calculated by mass balance equations under the calculated blending ratio. The method-solving algorithm to calculate the blending ratio was developed based on multiple linear regression, and the principle of the algorithm for multiple linear regression is shown in Equations (2)–(6).

Suppose there are now m groups of samples, each of which has n-1-dimensional characteristics. The mathematical expression of group m samples is shown in Equation (2), where h_m represents the output variable of group m samples, and x represents the input variable, in which $x_{m,0}$ can be set to 1. In real problems, if there is no constant term, $x_{m,0}$ can be set to 0. The purpose of multiple linear regression is to solve each parameter θ_{n-1} by an algorithm.

$$h_m = \theta_0 * x_{m,0} + \theta_1 * x_{m,1} + \theta_2 * x_{m,2} + \ldots + \theta_{n-1} * x_{m,n-1} \tag{2}$$

For the convenience of expression, the mathematical information contained in group m samples can be summarized in the form of Equation (3), where H is a matrix of $m * 1$, X is a matrix of $m * n$, and θ is a vector of $n * 1$ dimensions.

$$H = X * \theta \tag{3}$$

The parameter-fitting process of multiple linear regression is actually a process of minimization. Here, we define the target loss function as shown in Equation (4), where Y is a matrix of $m * 1$, representing the dependent variable of the sample.

$$f(\theta) = (X * \theta - Y)^2 \tag{4}$$

If the target loss function is the smallest, the fitting parameter is the best. Then, the multiple linear regression can be reduced to the minimum value of the target loss function. According to mathematical knowledge, the extreme value of the function appears at the place where its derivative is zero, so we can calculate the partial derivative of the target loss function, and the calculation process is shown in Equation (5)

$$\frac{\partial}{\partial\theta} f(\theta) = 0 \tag{5}$$

The optimal parameters can be obtained through simplification, and the detailed calculation process is shown in Equations (6) and (7).

$$\frac{\partial}{\partial\theta} f(\theta) = 2 * X^T * X * \theta - 2 * X^T * Y = 0 \tag{6}$$

$$\theta = \left(X^T * X\right)^{-1} * X^T * Y \tag{7}$$

In this method, multiple linear regression can help the method to solve the problem of obtaining the mixing ratio, in which the number of base samples can be regarded as columns of the input vector, and the NIR spectral vector can be regarded as rows of the input vector. The parameters obtained by the final solution can be defined as the mixing ratio of base samples.

Equations (8)–(11) show the detailed calculation procedure of the method-solving algorithm.

$$X = \left[x_1 \ldots x_j \ldots x_{Num.p}\right] \tag{8}$$

$$\textstyle\sum_{i=1}^{5} Mix_i * X_i^{base} = X^{test} \tag{9}$$

constraint:

$$Mix \geq 0 \tag{10}$$

$$\sum_{i=1}^{5} Mix_i = 1 \tag{11}$$

Mix is the mixing ratio obtained by multiple linear regression. The molecular composition of test samples can be obtained by the mixing ratio. Equations (12) and (13) represent the detailed calculation process.

$$MOLE = \left[mole_1 \ldots mole_j \ldots mole_{Num.m}\right] \tag{12}$$

$$\sum_{i=1}^{5} Mix_i * MOLE_i^{base} = MOLE^{test} \tag{13}$$

$MOLE^{base}$ is the molecular composition of base samples, $MOLE^{test}$ is the molecular composition of test samples, and $Num.m$ is the number of molecular species in the samples. It is determined by the gasoline molecular compositional model based on GC analysis to obtain the molecular composition of samples, which is 305 in this method.

3. Application on Typical Reformates

Due to a large number of test samples, this paper cannot show the detailed calculation process of all test samples. In this section, a detailed calculation process for a typical reformate (test sample 2) is discussed to reveal the result of each step and the model performance. Firstly, the method selects the best five base samples in the sample library by the sample selection algorithm, and the indices of base samples in the sample library are 48, 39, 40, 41, and 11. The comparison of the NIR spectra of base samples and target sample is shown in Figure 4a. Blue and red lines represent the NIR spectra of base samples and test sample, respectively. The NIR spectra of base samples are similar to that of the target sample, which indicates the effectiveness of the sample selection algorithm. However, a difference in NIR spectra can still be observed. A further improvement should be carried out by the following mixing ratio calculation step.

We applied multiple linear regression to solve for the mixing ratio, and the calculation results and process of multiple linear regression are shown in Figure 4. Figure 4b is the mixing ratio calculated by multiple linear regression. The ratio of base samples 11, 39, and 48 occupied more than 99 m%, while those of base samples 40 and 41 summed up to less than 1 m%. Figure 4c is the comparison of fitted NIR and measured NIR spectra of test samples. The NIR spectrum is almost the same between the calculated blend and target sample.

After obtaining the blending ratio, the molecular composition of the target sample is predicted. Figure 5 shows the final method solution results for test sample 2, where ABS represents the absolute error of each molecule composition predicted value compared with the gasoline molecular compositional model based on GC analysis, MABS represents the mean absolute error of molecular composition of test sample 2, and R represents the correlation coefficient. The horizontal error bar at each point in the graph represents the fluctuating range of the composition of that molecule in the sample library. The molecular compositional distribution in the sample library is large, but the concentration of each molecule can be accurately predicted by the proposed method. From the figure, it can be obtained that the method has a mean absolute error of 0.899×10^{-4} for the prediction of the molecular composition of test sample 2, and the correlation coefficient is 0.9998. The absolute errors of key molecules such as toluene and m-xylene are 0.0034 and 0.0024, respectively, with good prediction results.

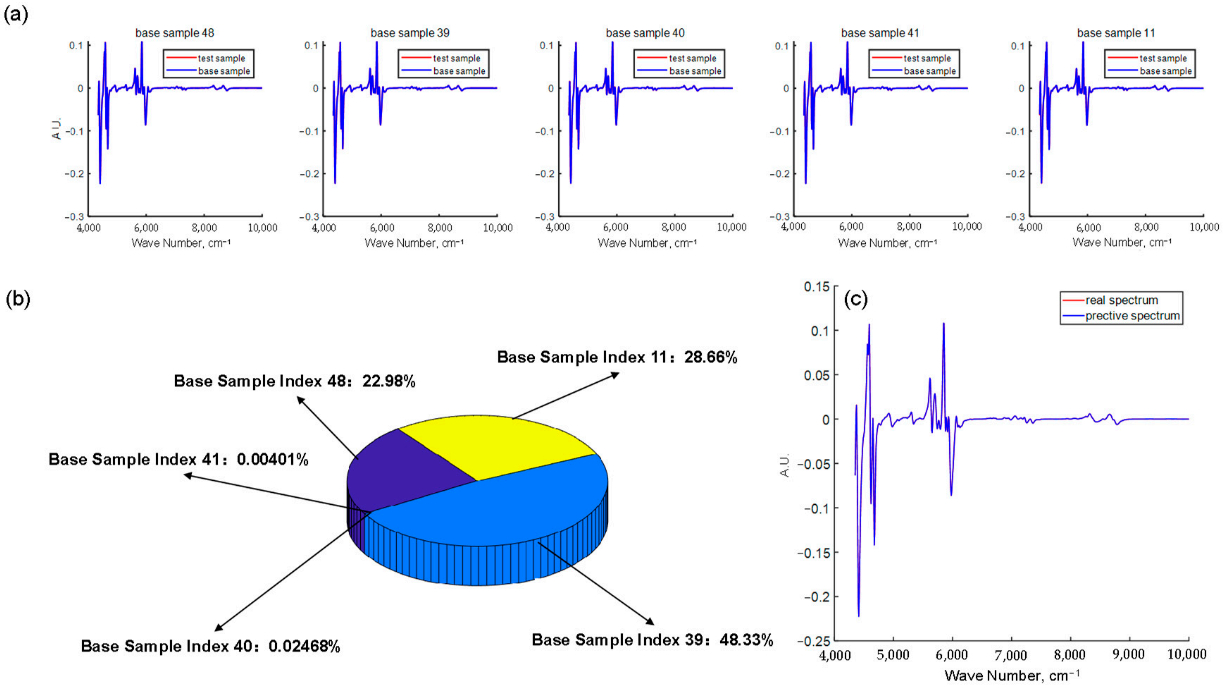

Figure 4. (**a**) NIR spectra of test samples and the selected base samples. Blue and red lines represent the NIR spectra of base samples and test sample, respectively. (**b**) The mixing ratio calculated by multiple linear regression. (**c**) Comparison of fitted NIR and measured NIR spectrum of test sample.

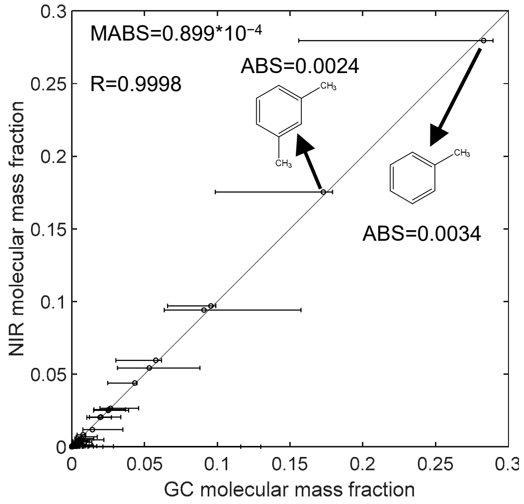

Figure 5. The comparison between predicted and measured molecular composition for test sample 2. The horizontal error bar at each point in the graph represents the fluctuating range of the composition of that molecule in the sample library.

Due to the large variety of molecules contained in test sample 2, the predictions for molecules with molecular mass fractions greater than 1% are shown in this paper. The molecular composition and n-alkane (P), iso-alkane (I), olefin (O), cycloalkane (N), and aromatic (A) (PIONA) composition predictions of test sample 2 are shown in Table 2. From the table, it can be seen that reformates are mainly composed of aromatics. The method predicts the total aromatic content with a high accuracy. It should be noted that the prediction of the content of individual molecules is also achieved. The prediction accuracy of the method for test sample 2 can agree with the gasoline molecular compositional model based on GC analysis. The result proved that predicting the detailed molecular composition from an NIR spectrum is possible. This information is quite useful for molecular-based property prediction, and can be potentially used for thermodynamic-based separation process calculation and mechanistic kinetic model development.

Table 2. The comparison between predicted and measured molecular composition and PIONA composition for sample 2.

Classification of Group and Molecular Composition	Predicted	GC-FID	ABS
paraffins (P)	0.0058	0.0062	0.0004
isoparaffins (I)	0.0125	0.0132	0.0007
olefins (O)	0.0001	0.0000	0.0001
naphthenics (N)	0.0046	0.0049	0.0003
aromatics (A)	0.977	0.9756	0.0014
methylbenzene	0.2796	0.2831	0.0035
ethylbenzene	0.0595	0.0577	0.0018
m-xylene	0.1753	0.1729	0.0024
p-xylene	0.0438	0.0432	0.0006
o-xylene	0.0969	0.0956	0.0013
n-butylbenzene	0.0202	0.0194	0.0008
1-methyl-3-ethylbenzene	0.0542	0.0532	0.001
1-methyl-4-ethylbenzene	0.025	0.025	0.000
1-methyl-2-ethylbenzene	0.0256	0.0254	0.0002
1,3,5-trimethylbenzene	0.0204	0.0199	0.0005
1,2,4-trimethylbenzene	0.0941	0.0908	0.0033
1,2,3-trimethylbenzene	0.0264	0.0265	0.0001
C10 aromatics	0.0118	0.0141	0.0023

4. Applications on Different Reformates

In order to show the applicability of the proposed method on different reformates, in this section, the molecular compositions of various reformates that were not included in the training sample library are described. The detailed result of each step is not shown and only the final result of the molecular composition prediction is shown.

Figure 6 shows a comparison between predicted and measured molecular composition for different reformates. As can be seen from the figure, the reformates collected in this paper consist mainly of toluene and m-xylene, and the sum of the two compositions exceeds 50%.

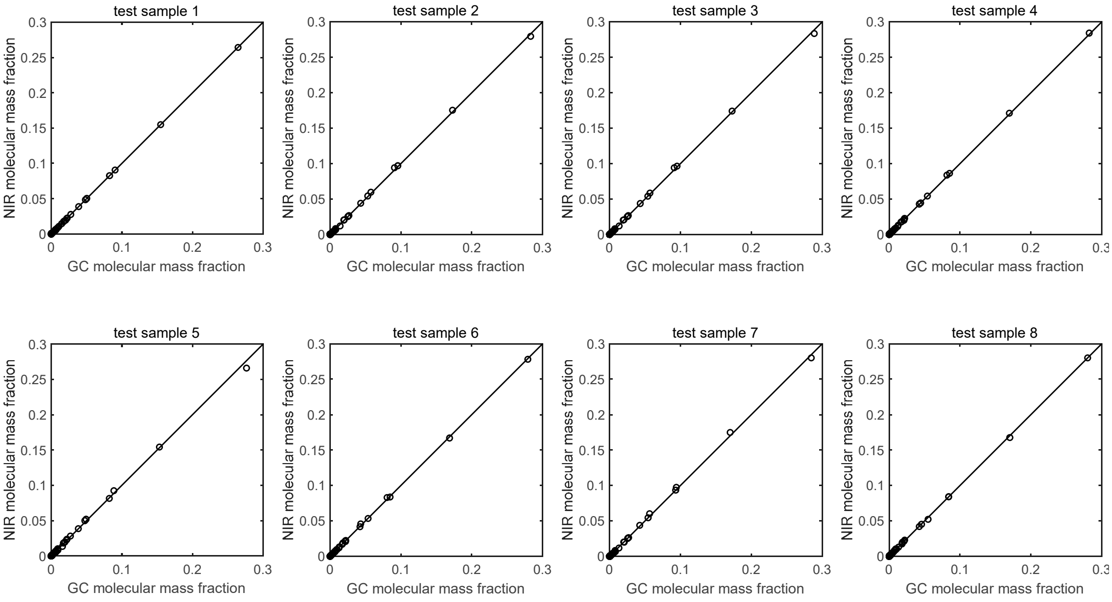

Figure 6. A comparison between predicted and measured molecular composition for different reformates.

Table 3 summarizes the error analysis of molecular composition predicted from NIR spectroscopy for different reformates. It can be seen that the spectral residuals of the test samples are all within 4×10^{-5}, indicating that the multiple linear regression is a very good fit for the spectra of the test samples. In terms of error, the method has good predicted results for reformates. The R values are all above 0.999 and the MABS values are all around 2×10^{-4}, which can prove that the prediction accuracy of this method can be consistent with the detection results based on GC analysis. From the prediction stability of the method, the fluctuation of MABS is around 1.5×10^{-4}, which can indicate that the prediction of reformates by this method is more stable.

Table 3. Error analysis of molecular composition predicted from NIR spectroscopy for different reformates.

Sample Index	MABS/$\times 10^{-4}$	R	Spectral Residual/$\times 10^{-5}$
1	0.261	0.9999	1.491
2	0.899	0.9998	3.156
3	0.717	0.9998	3.555
4	0.639	0.9999	2.754
5	1.515	0.9994	4.584
6	0.591	0.9999	4.054
7	1.0069	0.9997	3.356
8	0.616	0.9999	3.115
mean	0.781	0.9998	3.258

The predicted results of this method are good and can prove that the method is feasible for reformates rich in aromatics. We believe that this method is also valid for different types of gasoline.

5. Conclusions

In this paper, a molecular reconstruction method based on near-infrared (NIR) spectroscopy for reformates was described. In order to establish the method, we propose a hypothesis that "all gasoline is obtained from the mixture of multiple base gasolines". We propose that the mixing process can be represented by mixing between NIR spectra, and the mixing ratio can be obtained by multiple linear regression. The molecular composition of target gasoline can be obtained by the known molecular composition and mixing ratio of base gasoline. To verify the method, 75 sets of reformate samples from Jingbo Petrochemical and their NIR spectra were collected, and their molecular compositions were obtained by the gasoline molecular compositional model based on gas chromatography (GC) analysis. The samples were divided into base samples and test samples according to the collection time, and the ratio of the two was 9:1. The samples collected earlier were the base samples, and the samples collected later were the test samples. The method uses a combination of smoothing denoising and first-order derivative to preprocess the collected NIR spectra, eliminating the noise and baseline drift in the original NIR spectra. In this study, a sample selection algorithm based on Euclidean distance is developed to select the best five base samples based on the characteristics of test samples. The method uses multiple linear regression to obtain the mixing ratios of base samples mixed with test samples. In multiple linear regression, the input variable is the NIR spectrum of the base sample, and the output variable is the NIR spectrum of the test sample. It can be seen from the prediction results of the method that the accuracy of this method can reach the accuracy of the gasoline molecular compositional model based on GC analysis. The establishment of the molecular reconstruction method based on NIR spectroscopy enables the online measurement method to be applied to gasoline molecular management technology, which is of great significance to the online blending method of refineries.

Author Contributions: Conceptualization, M.C. and L.Z.; software, M.C. and L.Z.; writing—original draft preparation, M.C.; writing—review and editing, L.Z.; supervision, L.Z.; funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Area Research and Development Program of Guangdong Province (2021B0101220004).

Data Availability Statement: Data can be made available by the corresponding author by request.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Healy, W.C.; Maassen, C.W.; Peterson, R.T. A new approach to blending octane. *Proc. Am. Inst.* **1959**, *39*, 132–136.
2. Stewart, W.E. Predict octanes for gasoline blending. *Pet. Refin.* **1956**, *38*, 135–139.
3. Muller, A. New method produces accurate octane blending values. *Oil Gas J.* **1992**, *90*, 80–90.
4. Morris, W.E. Optimum blending gives best pool octane. *Oil Gas J.* **1986**, *63*, 63–68.
5. Gary, J.H.; Handwerk, G.E.; Kaiser, M.J. *Petroleum Refining Technology and Economics*; CRC Press: Boca Raton, FL, USA, 2007.
6. Rusinm, H.; Chungh, S.; Marshallj, F. A "transformation" method for calculating the research and motor octane numbers of gasoline blends. *Ind. Eng. Chem. Fundam.* **1981**, *20*, 195–204. [CrossRef]
7. Twu, C.H.; Coon, J.E. Predict octane numbers using a generalized interaction method. *Hydrocarb. Process.* **1996**, *75*, 51–56.
8. Sivaramakrishnan, K.; Nie, J.J.; de Klerk, A.; Prasad, V. Least Squares-Support Vector Regression for Determining Product Concentrations in Acid-Catalyzed Propylene Oligomerization. *Ind. Eng. Chem. Res.* **2018**, *57*, 13156–13176. [CrossRef]
9. Lopez-Zamora, S.; de Lasa, H. A Mid-Infrared Region (MIR) lumped Group Contribution based method for monitoring light gases and gasolines in Fluid Catalytic Cracking. *Chem. Eng. Sci.* **2020**, *212*, 115324. [CrossRef]
10. Balabin, R.M.; Safieva, R.Z. Gasoline classification by source and type based on near infrared (NIR) spectroscopy data. *Fuel* **2008**, *87*, 1096–1101. [CrossRef]
11. Balabin, R.M.; Safieva, R.Z.; Lomakina, E.I. Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. *Chemom. Intell. Lab.* **2007**, *88*, 183–188. [CrossRef]
12. Balabin, R.M.; Safieva, R.Z.; Lomakina, E.I. Wavelet neural network (WNN) approach for calibration model building based on gasoline near infrared (NIR) spectra. *Chemom. Intell. Lab.* **2008**, *93*, 58–62. [CrossRef]
13. Balabin, R.M.; Safieva, R.Z.; Lomakina, E.I. Gasoline classification using near infrared (NIR) spectroscopy data: Comparison of multivariate techniques. *Anal. Chim. Acta* **2010**, *671*, 27–35. [CrossRef] [PubMed]
14. Da Silva, N.C.; Cavalcanti, C.J.; Honorato, F.A.; Amigo, J.M.; Pimentel, M.F. Standardization from a benchtop to a handheld NIR spectrometer using mathematically mixed NIR spectra to determine fuel quality parameters. *Anal. Chim. Acta* **2017**, *954*, 32–42. [CrossRef] [PubMed]
15. Leal, A.L.; Silva, A.M.S.; Ribeiro, J.C.; Martins, F.G. Data driven models exploring the combination of NIR and H-1 NMR spectroscopies in the determination of gasoline properties. *Microchem. J.* **2022**, *175*, 107217. [CrossRef]
16. Lee, S.; Choi, H.; Cha, K.; Chung, H. Random forest as a potential multivariate method for near-infrared (NIR) spectroscopic analysis of complex mixture samples: Gasoline and naphtha. *Microchem. J.* **2013**, *110*, 739–748. [CrossRef]
17. Pereira, C.F.; Pimentel, M.F.; Galvao, R.K.H.; Honorato, F.A.; Stragevitch, L.; Martins, M.N. A comparative study of calibration transfer methods for determination of gasoline quality parameters in three different near infrared spectrometers. *Anal. Chim. Acta* **2008**, *611*, 41–47. [CrossRef] [PubMed]
18. Wang, S.T.; Liu, S.Y.; Zhang, J.K.; Che, X.G.; Wang, Z.F.; Kong, D.M. Feasibility study on prediction of gasoline octane number using NIR spectroscopy combined with manifold learning and neural network. *Spectrochim. Acta A* **2020**, *228*, 117836. [CrossRef]
19. Yu, H.J.; Wang, X.J.; Shen, F.F.; Long, J.; Du, W.L. Novel automatic model construction method for the rapid characterization of petroleum properties from near-infrared spectroscopy. *Fuel* **2022**, *316*, 123101. [CrossRef]
20. Femandes, H.L.; Raimundo, I.M.; Pasquini, C.; Rohwedder, J.J.R. Simultaneous determination of methanol and ethanol in gasoline using NIR spectroscopy: Effect of gasoline composition. *Talanta* **2008**, *75*, 804–810. [CrossRef]
21. Wu, Y.S.; Liu, Y.S.; Li, X.L.; Huang, Z.; Han, D. Gasoline octane number prediction from near-infrared spectroscopy with an ANN-based model. *Fuel* **2022**, *318*, 123543. [CrossRef]
22. Da Silva, N.C.; Massa, A.R.C.D.; Domingos, D.; Amigo, J.M.; Reboucas, M.D.; Pasquini, C.; Pimentel, M.F. NIR-based octane rating simulator for use in gasoline compounding processes. *Fuel* **2019**, *243*, 381–389. [CrossRef]
23. He, K.X.; Qian, F.; Cheng, H.; Du, W.L. A novel adaptive algorithm with near-infrared spectroscopy and its application in online gasoline blending processes. *Chemom. Intell. Lab.* **2015**, *140*, 117–125. [CrossRef]
24. He, K.X.; Zhong, M.Y.; Du, W.L. Weighted incremental minimax probability machine-based method for quality prediction in gasoline blending process. *Chemom. Intell. Lab.* **2020**, *196*, 103909. [CrossRef]
25. Wang, K.; He, K.X.; Du, W.L.; Long, J. Novel adaptive sample space expansion approach of NIR model for in-situ measurement of gasoline octane number in online gasoline blending processes. *Chem. Eng. Sci.* **2021**, *239*, 116672. [CrossRef]

26. Long, J.; Jiang, S.Y.; Liu, T.B.; Wang, K.; He, R.C.; Zhao, L. Modified Hybrid Strategy Integrating Online Adjustable Oil Property Characterization and Data-Driven Robust Optimization under Uncertainty: Application in Gasoline Blending. *Energy Fuels* **2022**, *36*, 6581–6596. [CrossRef]
27. He, K.X.; Zhong, M.Y.; Fang, J.Z.; Li, Y.R. Biased Minimax Probability Machine-Based Adaptive Regression for Online Analysis of Gasoline Property. *IEEE Trans. Ind. Inform.* **2020**, *16*, 2799–2808. [CrossRef]
28. Chen, W.H.; Yang, J. A double loop optimization method for gasoline online blending. In Proceedings of the 2016 IEEE International Conference on Industrial Technology (ICIT), Taipei, Taiwan, 14–17 March 2016; pp. 622–627.
29. Cui, C.; Zhang, L.Z.; Ma, Y.J.; Billa, T.; Hou, Z.; Shi, Q.; Zhao, S.Q.; Xu, C.M.; Klein, M.T. Computer-Aided Gasoline Compositional Model Development Based on GC-FID Analysis. *Energy Fuels* **2018**, *32*, 8366–8373. [CrossRef]
30. Cui, C.; Billa, T.; Zhang, L.Z.; Shi, Q.; Zhao, S.Q.; Klein, M.T.; Xu, C.M. Molecular Representation of the Petroleum Gasoline Fraction. *Energy Fuels* **2018**, *32*, 1525–1533. [CrossRef]
31. Durand, J.P.; Gautier, S.; Robert, E.; Guilhem, M.C.; Phan-Tan-Luu, R. Experimental design optimization of the analysis of gasoline by capillary gas chromatography. *J. High Resolut. Chromatogr.* **1997**, *20*, 289–294. [CrossRef]
32. Bruckner, C.A.; Prazen, B.J.; Synovec, R.E. Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis. *Anal. Chem.* **1998**, *70*, 2796–2804. [CrossRef]
33. Jaffe, S.B.; Freund, H.; Olmstead, W.N. Extension of structure-oriented lumping to vacuum residua. *Ind. Eng. Chem. Res.* **2005**, *44*, 9840–9852. [CrossRef]
34. Quann, R.J.; Jaffe, S.B. Structure-oriented lumping: Describing the chemistry of complex hydrocarbon mixtures. *Ind. Eng. Chem. Res.* **1992**, *31*, 2483–2497. [CrossRef]
35. Feng, S.; Cui, C.; Li, K.Y.; Zhang, L.Z.; Shi, Q.; Zhao, S.Q.; Xu, C.M. Molecular composition modelling of petroleum fractions based on a hybrid structural unit and bond-electron matrix (SU-BEM) framework. *Chem. Eng. Sci.* **2019**, *201*, 145–156. [CrossRef]
36. Klein, M.T.; Hou, G.; Bertolacini, R.; Broadbelt, L.J.; Kumar, A. *Molecular Modeling in Heavy Hydrocarbon Conversions*; CRC Press: Boca Raton, FL, USA, 2005.
37. Brudzewski, K.; Kesik, A.; Kolodziejczyk, K.; Zborowska, U.; Ulaczyk, J. Gasoline quality prediction using gas chromatography and FTIR spectroscopy: An artificial intelligence approach. *Fuel* **2006**, *85*, 553–558. [CrossRef]
38. Ghosh, P.; Hickey, K.J.; Jaffe, S.B. Development of a detailed gasoline composition-based octane model. *Ind. Eng. Chem. Res.* **2006**, *45*, 337–345. [CrossRef]
39. Haas, F.M.; Dryer, F.L. Application of blending rules for ignition quality metrics: A comment on "A linear-by-mole blending rule for octane numbers of n-heptane/iso-octane/toluene mixtures". *Fuel* **2014**, *120*, 240–242. [CrossRef]
40. Knop, V.; Loos, M.; Pera, C.; Jeuland, N. A linear-by-mole blending rule for octane numbers of n-heptane/iso-octane/toluene mixtures. *Fuel* **2014**, *115*, 666–673. [CrossRef]
41. Fioroni, G.M.; Rahimi, M.J.; Westbrook, C.K.; Wagnon, S.W.; Pitz, W.J.; Kim, S.; McCormick, R.L. Chemical kinetic basis of synergistic blending for research octane number. *Fuel* **2022**, *307*, 121865. [CrossRef]
42. Li, R.Z.; Herreros, J.M.; Tsolakis, A.; Yang, W.Z. Machine learning regression based group contribution method for cetane and octane numbers prediction of pure fuel compounds and mixtures. *Fuel* **2020**, *280*, 118589. [CrossRef]
43. Fabe, N.K.M. Multivariate Sensitivity for the Interpretation of the Effect of Spectral Pretreatment Methods on Near-Infrared Calibration Model Predictions. *Anal. Chem.* **1999**, *71*, 557–565. [CrossRef]