**Molecule-Based Octane Number Blending Model**

Guangqing Cai and Linzhou Zhang\*

*State Key Laboratory of Heavy Oil Processing & Petroleum Molecular Engineering Center (PMEC),* *China University of Petroleum, Beijing 102249, People’s Republic of China*

# Abstract

In this work, we developed an octane number blending model based on molecular interaction. Using molecular composition as input, the bulk octane number can be predicted by the ideal mixed octane number and the excess octane number calculated by the proposed method. A molecular-interaction function was proposed to simulate the non-linear octane number variation during mixing. The binary interaction parameters were adopted to illustrate the interaction between each pair of molecular classes. The parameters were estimated through a combination of globally and locally optimized regression from a database containing several hundred ternary mixtures and gasolines. The mean absolute errors of the training and test set for RON were 0.99 and 0.91 (1.03 and 0.85 for MON), demonstrating that the model has a good predictive capacity. We also showed an application of the proposed method on gasoline production with research octane number ranging from 92 to 108 solely using the gas chromatography data.

**Keywords:** octane number; molecular interaction; non-linear; molecular-interaction function; binary interaction parameters

# 1. Introduction

Gasoline is one of the most important petrochemical products. Commercial gasoline is a blended stream from different refining processes, including catalytic cracking, reforming, alkylation, isomerization, straight-run distillation. The compositions and properties of different streams vary significantly. The optimization of blending recipe requires an accurate prediction of the bulk property of each “blending component” and finished blends. Among all the properties involved, the prediction of octane number is of most difficult for a long time. Many researchers studied the optimization process of gasoline blend scheduling, in which the gasoline blend planning focused on the optimal mixing of various intermediate distillations from the refinery and some additives to meet product quality specifications and demand requirements. Li et al. developed a slot-based continuous-time model for integrated scheduling of gasoline blending operations in a refinery.1, 2 Castillo et al. presented a new inventory pinch-based, two-level decomposition approach, which incorporates non-linear blending models into integrated planning and approximate scheduling of gasoline blends.3 In the gasoline blend planning, the accurate prediction of octane number is vital to the recipe determination.

The difficulty in predicting octane number lies in two main aspects. The octane number is highly sensitive to molecular structure, in which the differences in methyl positions can lead to significant changes in octane numbers. The octane number is a measure of antiknock performance, which results from the premature combustion of gasoline fuels due to compression in the engine.4 As the fuel/air mixture is compressed in the internal combustion engine, certain molecules in gasoline tend to self-ignite before reaching the ignition spark, resulting in resistive expansion motion during the compression stroke of the engine. The combustion rate depends on the thermal stability of molecules, which is highly correlated with the molecular structure.5 Therefore, the octane number is closely related to the molecular composition of the gasoline fuel. To accurately predict the octane number, it is necessary to establish the quantitative relationship between the structure of pure components and the octane number, thereby comprehending the contribution of different kinds of structures to the octane number. The other is that the octane number can exhibit intense nonlinearity (synergistic and antagonistic) during mixing. Due to the high nonlinearity, we also need to investigate the interaction relationship and determine the blend regularity between different kinds of gasoline components. The blend regularity can help us make full use of the synergistic effects of the components to improve the quality of gasoline.

Many researchers used quantitative structure-property relationship (QSPR) models to predict the octane number of pure components. Albahri et al. coupled the group contribution (GC) method and an artificial neural network (ANN) to predict the octane number of hydrocarbons. Kubic et al. used a similar method to estimate the octane number of hydrocarbons and oxygenated compounds. Our previous work presented the octane number prediction model of pure components using the molecular descriptor + principal component analysis (PCA) + ANN combination method, which has a wide application range and strong extrapolation. In general, the octane number prediction of pure components can get satisfactory results through the database establishment and the structure-property prediction method.

The prediction methods of the octane number for gasoline mixtures are roughly separated into three categories: octane number prediction based on bulk property, spectroscopy information, and molecular lumps. Traditional octane number prediction approaches of gasoline mixtures are usually based on the mixing of bulk properties of different streams, such as the octane number, olefin, and aromatic content.6-10 Schoen and Mrstik presented a graphical correlation to predict the octane number of a series of binary systems according to the octane rating and volumetric olefin content.11 The correlation could produce different results depending upon the order of calculation. Stewart et al. improved the method to make it suitable for multi-component mixtures with consistent consequences.12 Morris et al. used the binary interaction coefficient to describe the non-linear behavior resulting from gasoline blending.13 The interaction parameter depended on component types, octane levels, and octane differences. Twu and Coon proposed an octane number prediction model based on the interaction of hydrocarbons, which uses the octane rating, the content of olefins, aromatics, and saturates as inputs.14 The traditional gasoline octane number prediction model based on bulk properties does not dig into the molecular-level information of the gasoline. Whereas, the essence of the non-linear behavior of gasoline blending is derived from the molecular interaction. Although the octane number of different types of gasoline may be equal, the molecular composition can be very different. The model parameters need to be retrained when the composition of the gasoline changes, which restricts the application range of the traditional model.

The traditional octane number prediction model has the demand for the information of octane rating, olefin, and aromatic content. Measuring these bulk properties of gasoline streams will consume a lot of time and effort. The development of spectroscopy technology provides a way to quickly obtain the bulk properties of gasoline fuels. In brief, spectroscopy technology is to use the characteristic information in the spectrum to establish the mapping relationship between functional groups and bulk properties by non-linear fitting methods.15-23 The approach is very suitable for online real-time analysis and prediction due to the fast detection of the spectrum, which has been applied in most refineries. Kelly et al. developed a gasoline octane number prediction model based on infrared spectroscopy in the range of 660-1215 nm.24 The model adopted the partial least squares to build the relationship between structural groups and octane number, in which the structural groups comprise methyl, methylene, double bond, and so on. Cooper et al. established the correlation between structural features in the Raman spectra and octane number.25 The model was cross-validated by the leave-one-out method, and the standard errors of RON and MON were 0.535 and 0.415, respectively. The advantage of spectroscopy technology is the rapid detection speed and high prediction accuracy, thus achieving great success in industry. However, the drawback is that the bulk property estimation of gasoline blending still relies on the traditional mixing rules and lack of exploration on the molecular information of gasoline. Besides, When the process streams change, the model needs to be retrained to ensure the prediction precision of the octane number.

Regardless of octane number prediction methods based on the bulk property and spectroscopy technology, they neither analyze the molecular composition information of gasoline, nor do they investigate the chemical nature of the non-linear behavior. Therefore, it is impossible to construct a widely applicable octane number prediction model. The traditional gasoline octane prediction model requires repeated parameter training when the gasoline stream changes, resulting in a very high cost for maintaining the model. The molecular-level gasoline blending model digs into the molecular composition information and accurately predicts the octane number of gasoline mixtures. Meanwhile, it can avoid the tedious work of repeated parameter training. In recent years, many researchers have developed octane prediction models based on molecular composition information.

Researchers began to use gas chromatography to analyze the chemical composition of gasoline and tried to establish the relationship between the chemical composition and octane number.26-36 Anderson et al. employed 31 structural lumps to describe the gasoline composition on the basis of gas chromatography equipped with a flame ionization detector, in which they assigned an ‘effective’ octane number to each lump.37 Then the contribution of each lump to the octane number was summed linearly to calculate the octane number of gasoline mixtures. Leeuwen et al. applied nonlinear regression methods to correlate chemical composition with octane numbers, such as projection pursuit regression and artificial neural networks.38 Brudzewski et al. used the same way to build the octane number prediction model.39 Lugo et al. followed Anderson’s 31 structural lumps and considered the nonlinear interaction between the components.40 The model was applied to the octane number prediction of catalytic cracking naphtha, which obtains good results with the deviation of ± 0.6 for RON and ± 0.4 for MON. Ghosh et al. described the gasoline composition using 57 molecular lumps and proposed a mixing rule of octane number.5 They adopted 1,471 gasoline fuels from many naphtha process streams to regress the parameters in the mixing equation, in which the standard errors of RON and MON were 1.01 and 1.05 units. The molecular lump is the representation of one class of molecules. Due to the sensitivity of the octane number to the structure, the octane number of the isomers in gasoline varies greatly. Hence, the molecular lumps cannot represent the contribution of each molecule in gasoline to the bulk octane number. Another shortcoming is that it cannot also study the interaction between gasoline molecules.

Octane number is determined by comparing the behavior of the tested fuel with that of mixtures of n-heptane and iso-octane defined by their liquid volume fractions. Since actual gasoline fuels are complex mixtures of hundreds of compounds. Gasoline surrogates based on a restricted set of compounds are defined to emulate actual fuels while allowing a detailed description of their behaviors.41 Knop et al. presented a linear-by-mole mixing rule to evaluate the octane number of ternary mixtures of n-heptane, iso-octane, and toluene.42 Fioroni et al. predicted the octane number of various synergistic blendstocks in a four-component surrogate using kinetic simulations with the Co-Optimization of Fuels & Engines (Co-Optima) mechanism.43 Li et al. used machine learning method to simultaneously estimate ONs of pure fuel compounds and TPRF mixtures.44 Badra et al. reported the blending characteristics of methyl tert-butyl ether (MTBE) with gasoline surrogates and fuels.45 Since the actual fuel was replaced with the gasoline surrogates, the predicted octane numbers were in good agreement with the experimental values. However, larger errors may occur in the application of gasoline blending due to the complex composition and mixing effects.

So far, many prediction approaches for the octane number of gasoline mixtures have been reported in the literature, which is mainly partitioned into methods based on bulk properties, spectroscopic techniques, and molecular lumps. A common weakness of these methods is that they ignore the molecular-level information of gasoline and do not explore the interaction between gasoline molecules. With the development of instrumental analysis and computer technology, the molecules contained in gasoline can be analyzed by gas chromatography to obtain qualitative and quantitative information. We used the simplified molecular input line entry specification (SMILES) to express the molecules and built the octane number prediction model based on the detailed molecular composition data of gasoline. The developed model can consider the contribution of each molecule to the octane number of gasoline and determine the interaction between different kinds of molecules. A mixing rule was presented, comprising linear mixing and molecular-interaction function. We used two different ways to verify the predictive ability and generalization performance of the model. Finally, the model was applied to the blending process of different grades of commercial gasoline.

# 2. Octane Number Mixing Rule Based on Molecular Interaction

Gasoline is composed of hundreds of different types of compounds, containing paraffins, isoparaffins, olefins, naphthenes, aromatics, and a small number of compounds with heteroatoms. Each compound has significantly different ONs. Knowing the ON of each molecule does not mean the ability to the accurate ON calculation of mixture. Synergistic or antagonistic effects were frequently occurred during gasoline mixing, resulting in either increase or decrease of actual ON compared to the molecularly averaged number. The mixing relationship is very complex, forming a great challenge for commercial gasoline production.

In the previous works, Ghosh et al. from ExxonMobil presented the interaction relationship between paraffins and olefins, as well as paraffins and naphthenes. However, they neglected the interactions between other hydrocarbons, such as paraffins and aromatics. In this paper, we studied the interaction at the molecular level to propose the octane number prediction model. We collected the experimental data of octane number of ternary mixtures from API Research Project 45, which consist of 48% volume-based 2,2,4-trimethylpentane, 32% n-heptane, and 20% other compounds. We regarded 2,2,4-trimethylpentane and n-heptane as a whole to investigate their interaction with other compounds. The comparison between the linear mixing and the experimental ON of them with paraffins, olefins, naphthenes, and aromatics was calculated.

According to the results of their mixing with paraffins, it can be observed that the linear mixing is basically in agreement with the experimental values, which indicates that there is no interaction between paraffin molecules. Due to the complex variation of structures for olefins, we divided them into three categories: olefins with a single double bond, olefins with two or more double bonds, cycloolefins with one or two double bonds. The experimental points are mainly above the linear mixing values, demonstrating that the blending of paraffins and olefins exhibits the synergistic effect. Furthermore, we can also discover that the distance between each experimental point and the linear mixing value is not the same, that is to say, the magnitude of the non-linear effect is diverse. If we set the interaction between each of the two molecules as an independent parameter, the molecular-interaction function will produce plenty of parameters that need to be regressed. Hence, we assumed that there existed only one kind of synergy between paraffins and olefins.

Similarly, the naphthenes were partitioned into cycloalkanes with five or fewer carbon numbers and cycloalkanes with six or more carbon numbers. We found that the cycloalkanes with three carbon numbers and paraffins would have a weak positive effect. But the content of three-carbon cycloalkane detected from gas chromatography in gasoline is very small in practice. So, we did not consider the interaction between them and paraffins here. Almost all of the naphthenes with six or more carbon numbers belong to linear mixing, except for cyclohexane and methyl-cyclohexane. Therefore, we supposed that there was nearly no interaction between paraffins and naphthenes. Finally, the aromatics were separated into single-branched, two-branched, and three- or more-branched aromatic hydrocarbons. The results denote that the experimental values are mostly above linear mixing, which is similar to the olefins. The interaction between paraffins and aromatics was considered to be a synergistic effect. Since there is no experimental data between olefins, naphthenes, and aromatics, we cannot investigate the interaction among them. We made a hypothesis that there are interactions between olefins and naphthenes, olefins and aromatics, as well as naphthenes and aromatics. Moreover, the produced gasoline always consists of a small number of oxygenated additives. Due to the synergistic or antagonistic effects between oxygenates and hydrocarbons, we have added the interaction parameters between them.

The interaction relationship between the molecules in the gasoline was determined through the analysis of the octane number of ternary mixtures. In this section, a novel mixing rule for the octane number was proposed, which contains the parts of linear mixing and molecular-interaction function. Figure 1 displays the schematic diagram of the octane mixing rule. We can observe that the first part is the linear mixing as shown in Equation (1),

|  |  |
| --- | --- |
|  | (1) |

where is the volume fraction of molecule , is the octane number of molecule . The of molecules contained in the mixture can be experimental data or those calculated from the QSPR model. In our previous work, several QSPR models were established to predict gasoline molecular properties, including research octane number (RON), motor octane number (MON), Reid vapor pressure (RVP), yield sooting index (YSI), and combustion heat.46 For the compound that has not been experimentally tested, its can be calculated from the QSPR model. is a linear volumetric blend of the of all the molecules present in the gasoline fuel.

The second part is the molecular-interaction function, which is the core part of the mixing rule we proposed. First, the molecules of gasoline were separated into P/I/O/N/A/OXY families. It should be noted that an excess of classification will result in a dramatic increase of model parameters, leading to model overfitting and weak extrapolation performance. Therefore, the molecular-interaction function is composed of six pairs of interactions, including interactions between paraffins and olefins, paraffins and aromatics, olefins and naphthenes, naphthenes and aromatics, olefins and aromatics, as well as oxygenates and hydrocarbons. The molecular-interaction function was proposed referring to the Rayleigh distribution function, where the variation of the curve was adjusted by two parameters. The function was shown in equation (2),

|  |  |
| --- | --- |
|  | （2） |

where  and are the binary interaction parameters, which control the shape of the variation curve. and correspond to the volume fraction of two kinds of families. In Figure 1 (b), The molecular-interaction function can fit the positive, negative effect, or linear mixing that occurs in the gasoline blending process. If is greater than 0, the interaction contributes beneficially to the ON of mixtures. Moreover, the variation curve will gradually become convex as the increases. On the contrary, if is less than 0, the interaction contributes detrimentally to the ON of mixtures. The special case is that is equal to 0, indicating the interaction has no contribution to the ON of mixtures. The function has 12 parameters that need to be fitted in total. The final mixing rule of octane number is as shown in Equation (3).

|  |  |
| --- | --- |
|  | （3） |

The predictive stability of the mixing equation we developed is relatively good with no large deviations. The mixing rule presented by Ghosh used the factor β in front of the PONA family, which represents the contribution of PONA class to the gasoline octane number. This makes the mathematical expression have high variability. In practical application, the mathematical formula can fit the parameters very well even without enough octane number data. Nevertheless, while estimating the octane number of gasoline samples not in the training set, sometimes there will be relatively large deviations. The equation we proposed is a modification to the octane number on the basis of linear mixing, which can decrease the maximum error and improve the stability of the model.



Fig 1. Schematic diagram of octane number mixing rule, comprising the linear mixing and molecular-interaction function.

# 3. Model Interaction Parameters Training

As mentioned above, we have proposed the binary interaction parameters and mixing rule of octane number. In this part, we need to train the binary interaction parameters in the equation to acquire the octane prediction model. The foundation of model construction is the establishment of molecular composition and the corresponding octane number database. The prediction accuracy of the model depends upon the quality, quantity, and diversity of data. Therefore, the construction of a comprehensive and reliable database is crucial to the success of the model. This work aims to predict the octane number of gasoline mixtures from the detailed molecular composition. Three principles should be followed when collecting data. One is to ensure the quality of molecular composition and the corresponding octane number data. The molecular composition is analyzed through gas chromatography equipped with a flame ionization detector. The RON and MON are measured using ASTM D2699 and ASTM D2700 standard methods, respectively.47, 48 The second is to gather enough data to train the model parameters. The training of the parameters between diverse compounds relies on the quantity of binary or ternary mixtures in the database. The third is to cover different types of gasoline streams in the library, such as straight run, catalytic cracking, reforming, alkylation, isomerization, and hydrogenation, and so forth. The diversity of data determines the generalization performance of the model.

Based on the above principles, we collected 231 sets of gasoline composition and the corresponding RON data (170 for MON), in which the molecular qualitative and quantitative data were detected through gas chromatography. In addition, The RON of 248 ternary mixtures (244 for MON) were gathered from API Research Project 45, and 90 sets of gasoline surrogate mixtures (50 for MON) from the literature.49, 50 Hence, the database we established comprised 569 RON and 464 MON experimental data. Figure 2 (a) displays the distribution of research octane number data. It can be seen that the octane number has a wide distribution range covering 30-120 units, which lays a good foundation for building the octane number prediction model with strong generalization performance.

The combination of genetic algorithm and local optimization algorithm was selected to optimize the parameters in the molecular-interaction function. The genetic algorithm accomplishes the choice of the optimal individual through operations such as selection, crossover, and mutation, which has the global search capacity in the solution space. First, we used the genetic algorithm to fit the parameters, in which the objective function was defined as the mean absolute error (MAE) between the predicted and experimental values. The optimized 12 interaction parameters ​​were obtained by minimizing the objective function. Although the genetic algorithm has the global searching ability, it is difficult to acquire an optimal solution in a huge space. Therefore, the parameters derived from the genetic algorithm were then input to the sequential quadratic programming (SQP) as the initial values. The SQP algorithm is a very effective tool for nonlinearly constrained optimization problems, which divides the original problem into a series of quadratic programming sub-problems to solve. A reasonable upper and lower bound was required in the SQP optimization algorithm. After repeated trials, it was determined that the upper and lower bound were set to the original parameter multiplied by 120% and 80%, respectively. The binary interaction parameters were continued to optimize in the specific space to seek better solutions. Even if the prediction accuracy of the octane number is increased by 0.1 units, it is a very big improvement for the model.

Figure 2 (b) and (c) display the parity plots of the predicted ​and experimental values for gasoline streams and ternary mixtures. It can be found that the mean absolute error is less than 1 unit for gasoline streams, illustrating that the model has a good training effect. However, the mean absolute error for ternary mixtures is 3.64 units. The larger error is acceptable because the ternary mixtures only serve to train the binary interaction parameters. The construction process of the model is not completed. We need to perform different ways of verifications on the prediction model to avoid overfitting. In the following section, we will elaborate on the specific process of the validation methods. Tables 1 and 2 list the specific values ​​of 12 parameters for RON and MON optimized by genetic and SQP algorithms. Figure S1 shows the distribution of MON experimental data and the comparison of the predicted and experimental values for gasoline streams and ternary mixtures.



Fig 2 (a) The distribution of RON experimental data collected from refineries and literature; (b-c) The parity plots of the predicted ​and experimental values for gasoline streams and ternary mixtures.

Table 1 The optimized binary interaction parameter matrix for RON using the combination of genetic algorithm and SQP optimization algorithm.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Parameters (RON) | P | O | N | A | OXY | Hydrocarbon |
| P | 0 | 45.9724 | 0 | 29.3924 | 0 | 0 |
| O | 0.0756 | 0 | 0.2451 | 12.3742 | 0 | 0 |
| N | 0 | 32.6996 | 0 | 58.8900 | 0 | 0 |
| A | -4.2176 | 0.0812 | -1.2812 | 0 | 0 | 0 |
| OXY | 0 | 0 | 0 | 0 | 0 | 1.7633 |
| Hydrocarbon | 0 | 0 | 0 | 0 | 20.3532 | 0 |

Table 2 The optimized binary interaction parameter matrix for MON.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Parameters (MON) | P | O | N | A | OXY | Hydrocarbon |
| P | 0 | 8.8320 | 0 | 28.8174 | 0 | 0 |
| O | -10.6951 | 0 | 0.1535 | 9.5020 | 0 | 0 |
| N | 0 | 27.4616 | 0 | -37.5355 | 0 | 0 |
| A | -0.6544 | 0.1407 | 0.0929 | 0 | 0 | 0 |
| OXY | 0 | 0 | 0 | 0 | 0 | -0.5256 |
| Hydrocarbon | 0 | 0 | 0 | 0 | -42.5730 | 0 |

# 4. Model Validation

The predictive ability is the key to whether the model can be applied to practical production. The overfitting phenomenon of parameters is a common problem in models, especially lacking enough experimental data. The overfitting means that the model performs quite well on the training data, but the consequence on the test set is often very poor. In this section, we employed ternary mixtures and gasoline streams to validate the extrapolation and generalization performance of the model.

## 4.1 Comparison for Ternary Mixtures

We used ternary mixtures to investigate the interaction relationship between different types of molecules and regress the binary interaction parameters. In this section, the training effect for ternary mixtures was displayed. Figure 3 shows the comparison among the experimental, linear, and nonlinear mixing values for ternary mixtures. It can be seen from Figure 3 (a), (e), and (f) that the non-linear mixing and the linear mixing values are overlapped, indicating that there is no interaction between paraffins and paraffins, as well as paraffins and naphthenes. Moreover, the non-linear values ​​are in good agreement with the experimental points. From Figure 3 (b) we can find that the nonlinearity is closer to the experimental point compared to linear mixing, which demonstrates that the model has a better training effect on olefins with one double bond. In Figure 3 (c) and (d), The non-linear values have a larger error from the experimental values. The content of olefins and cycloolefins containing two double bonds is less in gasoline, thereby contributing less to the bulk octane number of gasoline. Therefore, we did not set the independent parameters for these two types of olefins. From Figure 3 (g) we observed that the experimental value of aromatics with one side chain is close to the linear mixing. We have done a special treatment here, which regards the blending of paraffins and aromatic with one side chain as linear, except for aromatic hydrocarbons with double bonds in the branch. In Figure 3 (h) and (i), the non-linear fitting values are closer to the experimental points after adding the binary interaction parameters. Figure S2 displays the comparison among the experimental, linear, and non-linear mixing values for MON.

The model fits the octane number of ternary mixtures well, but some molecules have larger errors. Because the model trained the binary interaction parameters between P/I/O/N/A/OXY families, rather than each molecule. It is worth noting that the prediction model we developed does not place too much emphasis on the precision of the octane number of ternary mixtures, which only helps us optimize the parameters. The key point we focused on is the accuracy of the octane number of different types of gasoline streams and blended fuels.



Fig 3. The comparison among the experimental, linear, and non-linear mixing values for ternary mixtures, where the red dot represents the experimental value, the green dotted line signifies linear mixing values, the blue solid line represents the non-linear mixing value.

## 4.2 Validation for Overfitting Phenomenon

We used the octane number variation as the volume fraction increases to validate whether the model was over-fitted. All the pictures for the ternary mixtures were plotted. Due to the limited space, we selected different types of hydrocarbons to demonstrate the verification results. Figure 4 schematizes the variation trend of the octane number with the volume fraction for paraffins (2,2,4-trimethylpentane and n-heptane) mixing with different types of hydrocarbons, including paraffins, olefins, naphthenes, and aromatics. Figure 4 (a), (b), (c), and (d) plot the paraffins blending with paraffins, where the blue dots signify the experimental points of paraffins at 20% volume fraction. The predicted values are in line with the experimental data as well as the linear mixing. Figure 4 (e), (f), (g), and (h) are the results of paraffins mixing with olefins of diverse structures. As the volume fraction of olefins increases, the octane number curve of the ternary mixtures varies smoothly. Meanwhile, the errors calculated between the predicted and experimental values at the 20% volume fraction are very small. These outcomes illustrate that the model performs very well on the trade-off between high precision and overfitting. During the verification process, we found that if the model was over-fitted, the variation curve of the ternary mixtures would present an "S" shape, which is not in accord with our knowledge. The reason for appearing the "S" curve is that the model is forced to fit the experimental points at the 20% volume fraction at the expense of values at other volume fractions, thereby resulting in overfitting. Figure 4 (i), (j), (k), and (l) display the consequences of paraffin mixing with naphthenes. The predicted values almost overlap with the experimental points and the linear mixing values, which is similar to the paraffins mixing with paraffins. Figure 4 (m), (n), (o), and (p) show the results of paraffins blending with aromatics, which is similar to olefins. The mixed octane number changes smoothly with the increase of the volume fraction. The variation curve is more convex compared to the olefins. Moreover, the predicted octane numbers are in good agreement with the experimental values at 20% volume fractions. In summary, even though there is only one experimental point for each ternary mixture, the model can still predict the variation tendency of the octane number with the volume fraction, which indicates that the model does not exist over-fitting phenomenon and has a strong predictive capacity. Figure S3 shows the variation tendency of the MON as the volume fraction increases.



Fig 4. The variation curve of the RON with the volume fraction for paraffins (2,2,4-trimethylpentane and n-heptane) mixing with different types of hydrocarbon molecules, including paraffins, olefins, naphthenes, and aromatics.

## 4.3 Validation for Gasoline Streams

To verify the predictive ability for gasoline fuels that are not in the training set, the different process streams were selected for validation, which consists of straight-run, alkylation, hydrogenation, catalytic cracking, reforming, and blended gasoline. First, the molecular composition of different gasoline samples was analyzed by gas chromatography equipped with a flame ionization detector. The previously constructed QSPR model was employed to predict the octane number of pure components. Then we used the trained octane number prediction model to obtain the final results using the molecular composition and the octane number of each molecule as input. Figure 5 shows the comparison between the experimental and predicted values of different gasoline streams, where the red dot represents the training set and the blue dot represents the test set. In the training set, the predicted octane numbers of gasoline streams are in accordance with experimental values, in which the mean absolute errors are all about 1 unit. In the test set, the predicted values are also in good agreement with the experimental values, in which the maximum and minimum mean absolute errors are 1.22 and 0.35. The maximum error is derived from the reforming gasoline. The reason is that there may be not enough experimental data of reforming streams in the training set to optimize the octane number. The other may be that the content of aromatic compounds with complex structures accounts for the majority, which makes it difficult to train the binary interaction parameters. In the database, the catalytic cracking gasoline accounts for the majority, while the experimental data for other types of gasoline accounts for the minority. If the amount of data for other types of gasoline can be increased, the effect of model training will be better and have stronger generalization performance. Table 3 lists the specific values of PIONA families, experimental and predicted ONs for several representative gasoline fuels. The PIONA families of different types of gasoline vary greatly. The model we developed performed well on these representative gasoline fuels. Figure S4 displays the parity plots of the experimental and predicted values for MON, including different types of gasoline streams.



Fig 5. Parity plots of the experimental and predicted values using the optimized parameters for RON, including different types of gasoline streams.

Table 3 The specific values of PIONA families, experimental and predicted ONs for several representative gasoline fuels.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gasoline | P | I | O | N | A | Linear |  | Exp. | Pre. | Error |
| Straight-run | 18.77 | 28.64 | 0 | 37.37 | 15.19 | 51.13 | 4.25 | 55.2 | 55.38 | 0.18 |
| Alkylation | 7.87 | 88.51 | 0.99 | 0.52 | 2.12 | 93.53 | 0.91 | 94.9 | 94.44 | -0.45 |
| Catalytic Cracking | 10.58 | 25.55 | 30.35 | 9.32 | 24.21 | 77.87 | 10.74 | 88.7 | 88.61 | -0.08 |
| Hydrogenation | 8.50 | 29.64 | 16.51 | 13.03 | 32.32 | 78.45 | 10.35 | 88.2 | 88.80 | 0.60 |
| Reforming | 4.41 | 9.12 | 0.42 | 1.59 | 84.45 | 97.92 | 4.71 | 103.5 | 102.63 | -0.86 |

# 5. Application on Gasoline Blending

Gasoline blending is the final process for commercial gasoline production, which is vital to raise the refinery's profit. If the blended octane number is higher than the gasoline standard, it will bring certain losses to the refinery. Therefore, accurate prediction of the octane number is of great significance to gasoline production. Figure 6 shows the flow chart of octane number prediction for application in refineries. The traditional gasoline blending process is to mix different types of component fuels according to a given recipe. Then an instrument is used to measure the octane number of blended gasoline, which costs a great quantity of manpower and material resources. In this work, the developed octane number prediction model can accurately estimate the octane number, thereby replacing the cumbersome detection process. First, the molecular qualitative and quantitative information of gasoline samples should be analyzed through gas chromatography equipped with a flame ionization detector. Second, the mass fraction of each molecule was summed linearly to obtain the molecular composition of blended gasoline according to the recipe. Third, the trained octane model was employed to predict the octane number of gasoline mixtures.



Fig 6. The flow chart of octane number prediction for application in refineries.

We selected four different grades of gasoline products to illustrate the prediction process of the octane number on the molecular level. Tables 4 lists the detailed recipes for the four blended gasoline, which involve the blending components and the corresponding mass fraction. For example, No. 92 gasoline is composed of 75.5% mass-based hydrogenation, 3.5% MTBE, 8.5% alkylation, 2% reforming, 2% aromatics (1), 8.5% aromatics (2). The molecular composition of different types of component oils varies greatly, leading to a large difference in octane number. Due to lack of the experimental data, it is impossible to compare the octane number of the middle distillate gasoline. Tables S1, S2, S3, and S4 list the P/I/O/N/A/OXY families and the predicted octane numbers for the four blended gasoline.

Then, the content of each molecule was linearly summed according to the mass fraction of each component gasoline. We used the trained model to estimate the ONs of the four blended gasoline. Table 5 lists the experimental and predicted values of octane number. The PIONA composition of the gasoline products is similar, while the content of oxygenated compounds has a larger difference. The absolute errors of the four brands of blended gasoline are 0.37, 0.66, 1.30, and 0.69. The deviations of NO. 92, 95, and 108 are relatively small, less than 0.7 units. But the error of NO. 98 gasoline reaches 1.30. The possible reason is that there exist errors in the GC-FID detection process or in the prediction of the octane number of middle distillate gasoline. In general, these results demonstrate that the octane number prediction model we developed is quite stable and reliable, which can guide the gasoline blending process in refineries. In addition, we also list the detailed results of each step from the molecular composition analysis of gas chromatography to the octane number prediction of the blended gasoline in the supporting information.

Table 4 The detailed recipes for NO. 92, 95, 98, and 108 gasoline, containing the blending components and the corresponding mass fractions.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| NO. 92 | Recipe | NO. 95 | Recipe | NO. 98 | Recipe | NO. 108 | Recipe |
| Hydrogenation | 75.5 | Hydrogenation | 66 | Hydrogenation | 30 | Alkylation | 30 |
| MTBE | 3.5 | MTBE | 9 | Alkylation | 20 | MTBE | 18 |
| Alkylation | 8.5 | Alkylation | 9.5 | Tuluene | 24 | Isopentane | 12 |
| Reforming | 2 | Reforming | 5 | MTBE | 11 | Tuluene | 40 |
| Aromatics 1 | 2 | Aromatics | 2.5 | Iso-pentane | 15 |  |  |
| Aromatics 2 | 8.5 | Tuluene | 8 |  |  |  |  |

Table 5 The experimental and predicted values of octane number for the four grades of blended gasoline.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gasoline | P | I | O | N | A | OXY | Linear |  | Exp. | Pre. | Error |
| NO. 92 | 5.35 | 39.77 | 13.46 | 7.87 | 30.20 | 3.35 | 83.45 | 8.78 | 92.60 | 92.23 | -0.37 |
| NO. 95 | 4.71 | 36.34 | 12.96 | 6.84 | 30.53 | 8.62 | 88.75 | 7.71 | 95.80 | 96.46 | 0.66 |
| NO. 98 | 4.14 | 41.86 | 7.97 | 3.38 | 31.65 | 11.00 | 94.76 | 4.89 | 98.30 | 99.65 | 1.30 |
| NO. 108 | 1.25 | 37.65 | 1.88 | 0.29 | 40.93 | 18.00 | 106.26 | 2.65 | 109.60 | 108.91 | -0.69 |

# 6. Conclusion

In this paper, we developed a gasoline octane number prediction model based on molecular interaction. The ternary mixtures were used to investigate the interaction between different types of molecules. We found that there was nearly no interaction between paraffins and paraffins, paraffins and naphthenes. Moreover, there existed a synergistic effect between paraffins and olefins, paraffins and aromatics. The molecular-interaction function between P/I/O/N/A/OXY families was proposed. Then, the genetic algorithm and the SQP optimization algorithm were combined to fit the 12 binary interaction parameters. The mean absolute errors of the RON and MON were 0.99 and 1.03 units through repeated parameter adjustments. The ternary mixtures were used to verify whether the model was over-fitted. In addition, we also selected different types of process streams and blended gasoline that are not in the training database to validate the predictive capacity of the model, in which the mean absolute errors were 0.91 and 0.85. Finally, the developed octane number prediction model was applied to guide the refinery blend process on the molecular level.

# Author Information

\*Corresponding Author Email: Lzz@cup.edu.cn (L. Zhang);

# Acknowledgments

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (U19B2002). The authors declare no competing financial interests.

# Supporting Information

The supporting information is available free of charge. The detailed training and test results for MON (Figures S1-S4) can be found in the supporting information, which is provided in Word format (.docx). The molecular composition and the corresponding RON of the middle distillate gasolines were displayed in the supporting information, which is provided in Excel format (.xlsx).

# Reference

1. Li, Jie, I. A. Karimi, and Rajagopalan Srinivasan. "Recipe determination and scheduling of gasoline blending operations." AIChE journal 56.2 (2010): 441-465.

2. Li, J.; Xiao, X.; Floudas, C. A., Integrated gasoline blending and order delivery operations: Part I. short-term scheduling and global optimization for single and multi-period operations. AIChE Journal 2016, 62 (6), 2043-2070.

3. Castillo, P. A. C.; Mahalec, V., Inventory pinch based, multiscale models for integrated planning and scheduling-part I: Gasoline blend planning. AIChE Journal 2014, 60 (6), 2158-2178.

4. Cornelius, W.; Caplan, J. D., Some Effects of Fuel Structure, Tetraethyl Lead, and Engine Deposits On Precombustion Reactions in a Firing Engine. 1952.

5. Ghosh, P.; Hickey, K. J.; Jaffe, S. B. J. I.; Research, E. C., Development of a Detailed Gasoline Composition-Based Octane Model. 2006, 45 (1), págs. 337-345.

6. Essien, Gabriel, and Ayoade Kuye. "Development and Validation of Linear Programming Models for Gasoline and Fuel Oil Blending." International Journal of Scientific & Engineering Research 2016, 7: 1204.

7. HEALY W C, JR C W M, PETERSON R T. A new approach to blending octane. Proc Am Inst, 1959, 39(3): 132-136.

8. Rusin, M. H., Chung, H. S., & Marshall, J. F. A" transformation" method for calculating the research and motor octane numbers of gasoline blends. Industrial & Engineering Chemistry Fundamentals, 1981, 20(3): 195-204.

9. Muller A. New method produces accurate octane blending values. Oil and Gas Journal;(United States), 1992, 90(12).

10. Pasadakis, N.; Gaganis, V.; Foteinopoulos, C. J. F. P. T., Octane number prediction for gasoline blends. 2006, 87 (6), 505-509.

11. Schoen, W. F.; Mrstik, A. V. J. I.; Chemistry, E., Calculating Gasoline Blend Octane Ratings. 2002, 47 (9), 1740-1742.

12. Zahed, A. H.; Mullah, S. A.; Processing, M. B. J. H., Predict octane number for gasoline blends. 1993, 72 (5), 85-87.

13. Snee, R. D., Morris, W. E., & Smith, W. E. Interaction blending equations enhance reformulated gasoline profitability. Oil and Gas Journal;(United States), 1994, 92(3).

14. Twu, C. H.; Coon, J. E. J. H. P., Predict octane numbers using a generalized interaction method. 1996, 75 (2).

15. Bueno, A.; Baldrich, C. A.; Daniel, M. V. J. E.; Fuels, Characterization of Catalytic Reforming Streams by NIR Spectroscopy. 2009, 23 (3), 3172-3177.

16. Alves, J. C. L., Henriques, C. B., & Poppi, R. J. Determination of diesel quality parameters using support vector regression and near infrared spectroscopy for an in-line blending optimizer system. Fuel, 2012, 97: 710-717.

17. Jeong, H. I., Lee, H. S., & Jeon, J. H. Determination of research octane number using NIR spectral data and ridge regression. Bulletin of the Korean Chemical Society, 2001, 22(1): 37-42.

18. Khanmohammadi, M.; Garmarudi, A. B.; Miguel, D. J. T., Feature selection strategies for quality screening of diesel samples by infrared spectrometry and linear discriminant analysis. 2013, 104, 128-134.

19. Paíga, P.; Mendes, L.; Albergaria, J. T.; Delerue-Matos, C. M. J. C. P., Determination of total petroleum hydrocarbons in soil from different locations using infrared spectrophotometry and gas chromatography. 2012, 66 (8), 711-721.

20. Pasquini, C.; Bueno, A. F. J. F., Characterization of petroleum using near-infrared spectroscopy: Quantitative modeling for the true boiling point curve and specific gravity. 2007, 86 (12-13), 1927-1934.

21. Dong, C.; Tan, J.; Chen, Z. J. S.; ANALYSIS, S., Prediction the Octane Number of Gasoline by Near Infrared Spectra Analysis. 1999, 19 (3), 314.

22. Maggard, S. M.; Welch, W. T. J. W., Determination of aromatics in hydrocarbons by near infrared spectroscopy. 1994.

23. Li, W. M., Wu, G. Y., & Lin, X. P. Study of the Gasoline Octane Number Model Using ANN Method with near Infrared Spectra Analysis. JOURNAL-JIANGSU POLYTECHNIC UNIVERSITY, 2005, 17(3): 16.

24. Kelly, J. J.; Barlow, C. H.; Jinguji, T. M.; Callis, J. B. J. A. C., Prediction of gasoline octane numbers from near-infrared spectral features in the range 660-1215 nm. 1989, 61 (4), 313-320.

25. Cooper, J. B.; Wise, K. L.; Groves, J.; Welch, W. T. J. A. C., Determination of Octane Numbers and Reid Vapor Pressure of Commercial Petroleum Fuels Using FT-Raman Spectroscopy and Partial Least-Squares Regression Analysis. 1995, 67 (22), 4096-4100.

26. Institute, A. P.; University, P. S. J. A. P. I., Technical Data Book - Petroleum Refining, Metric Edition. 1976.

27. Scott, E. J. Y. Knock characteristics of hydrocarbon mixtures. Proc. API DiV. Refin, 1958, 38: 90.

28. Montero, E.; Ruiz-Gutierrez, V.; Villar, J.; Muñiz, O.; Puppo, A.; Gil-Peralta, A.; Stiefel, P. J. c. e. i. e. a., "Determinacion de la composicion de acidos grasos del tejido adiposo mediante cromatografia gaseosa con columnas capilares en pacientes con infarto de miocardio y AVC de origen isquémico". 1991, 3, 34.

29. Nikolaou, N., Papadopoulos, C. E., Gaglias, I. A., & Pitarakis, K. G. A new non-linear calculation method of isomerisation gasoline research octane number based on gas chromatographic data. Fuel, 2004, 83(4-5): 517-523.

30. Proticlovasic, G.; Jambrec, N.; Deursiftar, D.; Prostenik, M. V. J. F., Determination of catalytic reformed gasoline octane number by high resolution gas chromatography. 1990, 69 (4), 525-528.

31. Yoshio, S., MEASURING METHOD OF RESEARCH OCTANE NUMBER OF GASOLINE BY GAS CHROMATOGRAPH AND ITS APPARATUS. 1997.

32. Alexandrovna, S. J., & Tuyen, D. C. Development of a detailed model for calculating the octane numbers of gasoline blends. In International Forum on Strategic Technology 2010 (pp. 430-432). IEEE.

33. Cui, C.; Zhang, L.; Ma, Y.; Billa, T.; Hou, Z.; Shi, Q.; Zhao, S.; Xu, C.; Klein, M. T. J. E.; fuels, Computer-Aided Gasoline Compositional Model Development Based on GC-FID Analysis. 2018, 32 (8), 8366-8373.

34. Huang, S. W.; Zhao, X. F.; Guo, Z.; Wang, S. C. J. G. C. I., Determination of Research Octane Number of Gasoline by Gas Chromatography. 2018.

35. Liu, M. Y.; Zhou, P.; Ping, K.; Yang, C. G.; Mu, M. R. In Determination of octane number of gasoline by double ANN algorithm combined with multidimensional gas chromatography, Sixth International Conference on Natural Computation, 2010.

36. Telen, S.; Jambrec, N.; Romac, I. B. J. G. I. M., FCC GASOLINE RESEARCH OCTANE NUMBER DETERMINATION BY GAS CHROMATOGRAPHY. 2003, 42 (2), 105-115.

37. Anderson, P. C., Sharkey, J. M., & Walsh, R. P. Calculation of the research octane number of motor gasolines from gas chromatographic data and a new approach to motor gasoline quality control. J. Inst. Pet, 1972, 58(560): 83.

38. Leeuwen, J.; Jonker, R. J.; Gill, R. J. C.; Systems, I. L., Octane number prediction based on gas chromatographic analysis with non-linear regression techniques. 1994, 25 (2), 325-340.

39. Brudzewski, K.; Kesik, A.; Ko?Odziejczyk, K.; Zborowska, U.; Ulaczyk, J. J. F., Gasoline quality prediction using gas chromatography and FTIR spectroscopy: An artificial intelligence approach. 2006, 85 (4), 553-558.

40. Lugo, H. J.; Ragone, G.; Zambrano, J. J. I. e. c. r., Correlations between Octane Numbers and Catalytic Cracking Naphtha Composition. 1999, 38 (5), 2171-2176.

41. Correa Gonzalez, S.; Kroyan, Y.; Sarjovaara, T.; Kiiski, U.; Karvo, A.; Toldy, A. I.; Larmi, M.; Santasalo-Aarnio, A., Prediction of Gasoline Blend Ignition Characteristics Using Machine Learning Models. Energy & Fuels 2021, 35 (11), 9332-9340.

42. 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.

43. 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.

44. Li, R., Herreros, J. M., Tsolakis, A., & Yang, W. Machine learning regression based group contribution method for cetane and octane numbers prediction of pure fuel compounds and mixtures. Fuel, 2020, 280: 118589.

45. Badra, J., Alowaid, F., Alhussaini, A., Alnakhli, A., & AlRamadan, A. S. Understanding of the octane response of gasoline/MTBE blends. Fuel, 2022, 318, 123647.

46. Cai, G.; Liu, Z.; Zhang, L.; Shi, Q.; Xu, C. J. C. E. S., Systematic Performance Evaluation of Gasoline Molecules Based on Quantitative Structure-Property Relationship Models. 2020, 229, 116077.

47. Institution, B. S., Methods of test for petroleum and its products. Motor and aviation-type fuels. Determination of knock characteristics. Motor method.

48. Institution, B. S., Petroleum products. Determination of knock characteristics of motor fuels. Research method.

49. Lovell, W. G. J. I.; Chemistry, E., Knocking Characteristics of Pure Hydrocarbons. 1948, 40 (12), 2388-2438.

50. Knop, V.; Loos, M.; Pera, C.; Jeuland, N. J. F., A linear-by-mole blending rule for octane numbers of n-heptane/iso-octane/toluene mixtures. 2014, 115 (jan.), 666-673.