ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Correlation between the Molecular Structure and Viscosity Index of CTL Base Oils Based on Ridge Regression

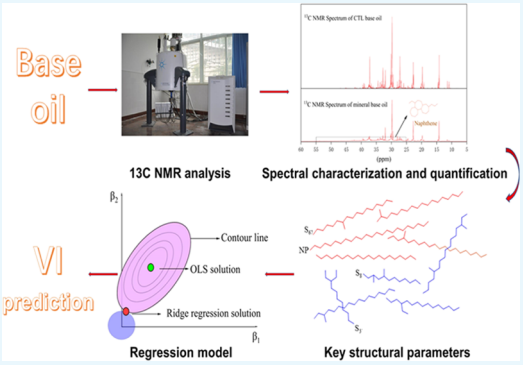
Chunhua Zhang, Hanwen Wang,* Xiaowen Yu, Chaolin Peng, Angui Zhang, Xuemei Liang, and Yinan Yan

Cite This: *ACS Omega* 2022, 7, 18887−18896

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In China, coal-to-liquid (CTL) lube base oils with ultrahigh viscosity index (VI) are very popular. Since it consists of chain alkanes only and can be precisely characterized by molecular structures alone, quantitative ^{13}C nuclear magnetic resonance (NMR) data are used to generate the average structural parameters (ASPs) of CTL base oil. In this work, the ASPs and bulk properties of CTL base oils were tested and compared with those of mineral base oils. Based on the test results, the correlation between the unique property of CTL base oil VI and ASPs was analyzed. To eliminate the effect of significant multicollinearity among the input variables, statistical methods such as ordinary least-squares (OLS), stepwise regression, and ridge regression methods were used to build the VI prediction model. The main findings are as follows: according to the ^{13}C NMR spectrum, CTL base oils had a significantly higher content of isomeric chain alkanes (including several branching structures) than mineral base oil, while the content of cycloalkanes was zero; among several branched structures, the one with the largest difference in content is structure S_{67}, which has the highest percentage in the iso-paraffin structures, all above 25.5% in CTL base oils and below 21.39% in mineral oils; according to the distillation curve of the simulated distillation (SimDist) analysis, CTL base oils with similar carbon number distribution showed lower boiling points, narrower distillation ranges, and higher distillation efficiencies than mineral base oil; correlation analysis showed that the average chain length (ACL), normal paraffins (NPs), and structure S_{67} caused the CTL base oil to exhibit a higher VI; and from ^{13}C NMR data, the ridge regression model was used to obtain regression coefficients consistent with reality, and the expected VI could be well predicted with a correlation coefficient of 0.935.



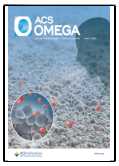
1. INTRODUCTION

The global distribution of energy mix and the expanding demand for high-performance lubricants have driven the development of coal-to-liquid (CTL), gas-to-liquid (GTL), and biomass-to-liquid (BTL) technologies.[1−4] In China, where coal mining resources are extensive, CTL base oil is a promising alternative to existing group II and III base oils. It is technically a lubricant component for machines and engines composed of straight-chain alkanes. These straight-chain alkanes are the main chemical species that provide lubricating properties similar to those of PAO (group IV).[1,5,6] CTL base oil has various advantages including high viscosity index (VI), excellent low-temperature performance, environment friendliness, and the absence of polyaromatic hydrocarbons, naphthenes, nitrogens, and sulfides.

CTL base oils are generally produced from feedstocks of coal through an indirect liquefaction process. In this method, synthesis gases (carbon monoxide and hydrogen) produced from coal are converted to paraffinic hydrocarbons by the Fischer−Tropsch (FT) reaction. Paraffinic hydrocarbons are processed by isomerization and fractionation, named CTL base oils. Pure synthetic CTL base oils, which have a high VI, are identified as group III base oils due to the use of traditional refining technologies such as hydrocracking and isomeric dewaxing.[4,7] Base oil properties such as VI, pour point, density, flash point, evaporation loss, and rotary pressure vessel oxidation test (RPVOT) affect the lubricating performance and service life of the lubricant.

VI, first proposed by standard oil's Dean and Davis in 1929,[8] is a method for describing the viscosity−temperature relationship between base oils and lubricants. The magnitude of its value is one of the most important indicators in determining the quality grade of a lube base oil. For example, a high VI value allows

Received: March 28, 2022
Accepted: May 12, 2022
Published: May 23, 2022



Fischer−Tropsch synthetic base oils to be considered group III base oils, which means that their kinematic viscosity (KV) is less sensitive to temperature. This smaller variation of KV with temperature is pleasing, as such base oils can be adapted for use in applications with a wide range of temperature variations to ensure effective lubrication, especially in the field of automotive engine lubricants. Furthermore, the VI of a lube base oil has been shown to be highly dependent on feedstocks, process conditions, etc., and is ultimately reflected in the molecular composition and structure of the base oil.[9]

Currently, the main approaches for molecular characterization of base oils are mass spectrometry (MS) and nuclear magnetic resonance (NMR).[10,11] MS provides information on the content of different types of compounds and the carbon number of each compound, which allows us to evaluate the changes in hydrocarbon structure types during the hydrogenation process. Nuclear magnetic resonance (NMR), on the other hand, is mainly used to make a certain degree of speculation on the structure of base oils by measuring the content of different types of carbon atoms and calculating some ASPs.[12−16] Compared to ^{1}H NMR, the chemical shifts of ^{13}C NMR are more sensitive to the molecular structure and can provide more valid information,[17] so the carbon-type distributions it provided are often used as input variables in models for VI calculations. Sarpal et al.[18,19] delineated the attribution of ^{13}C NMR spectral peaks and established the correlation between VI and carbon-type composition and the distribution of the branched structure. However, in a later study, it was found that the correlation was established based on qualitative results. Sharma et al.[20] analyzed the relationship between the average structural parameters of hydrogenated base oils and VI, such as the amount of normal paraffins (NPs), iso-paraffins (IPs), and average chain length (ACL), based on a simple linear regression approach. The results indicated that a single structural parameter cannot accurately predict the VI and that appropriate prediction accuracy requires the use of at least two or more structural parameters. In addition, the correlation indicated that a decrease in ACL caused an increase in VI; however, this conclusion did not work for other studies.[21,22] Verdier et al.[21] developed a correlation model for VI based on molecular structure data from ^{13}C NMR of 20 base oils and the correlation model was shown to predict the experimental data well with an R^2 of 0.9589. However, the significant correlation between molecular structures as input variables can make the regression model unstable. Recently, Noh et al.[11] developed two VI regression models for three types of base oils based on the molecular structure of hydrocarbons; they found that the constrained regression model that considers the physical significance of the regression components has better generalization ability than the stepwise regression method based on pure data. Despite the increasing production of CTL base oils in China and the fact that VI is a very important characteristic in determining the quality grade of base oils, so far, there is a lack of practical correlation models between VI and the molecule structure of CTL base oils.

With small sample sizes, most studies have focused on simple linear statistical methods to determine regression models of dependent variables with multiple explanatory variables.[23−25] Based on experimental samples, the strongly correlated variables were selected as input features for modeling by measuring the linear correlation between the input structures and VI, which implied that the relationship between the representative variables with strong correlation coefficients and VI was confirmed. However, the coefficients of the regression model did not guarantee the consistency with the actual positive and negative correlations due to the significant multicollinearity among different molecular structures, which increased the variance of the coefficients of the input variables and made the prediction model unrepresentative of the true regulation.[26] For severe multicollinearity, the common possible solutions are (1) grouping input variables and combining or splitting highly correlated variables, (2) using stepwise regression methods to filter and eliminate input variables, and (3) using ridge regression methods to introduce a small amount of bias to reduce sensitivity to sample data.[27,28]

Studies on the relationship between base oils with various structures and bulk properties could provide the right direction for base oil production and processing. Although research on the structures and properties of mineral base oils has long been started, we know little about the differences in the properties of CTL base oils with different molecular structures. The current work, therefore, a study of five CTL base oils and four mineral base oils with different molecular structures, was carried out, and the results showed that ^{13}C NMR could accurately characterize the structure of CTL base oils. In addition, the established VI model has excellent predictive ability.

2. EXPERIMENTAL DEVICES AND METHODS

2.1. Samples. Nine different oil samples were analyzed in this study: C#1−C#5 were CTL base oils and M#1−M#4 were hydrotreated and hydrocracked mineral base oils. CTL base oils were supplied by a Chinese coal-to-oil production plant. Mineral base oils M#1 and M#2 were group III base oils and M#3 and M#4 were group II base oils, purchased from Ssangyong in Korea and Hainan Lian in China, respectively. The macroscopic physical properties of the individual base oils ensure a certain degree of variability to meet adequate representation.

2.2. Measurement of Bulk Properties. The different properties of base oils were tested according to ASTM standards. The pour point and flash point were measured using methods ASTM D-5949 and ASTM D-93, respectively, and evaporation loss was measured according to Din51.581 (noack). The oxidation stability was analyzed by the rotary pressure vessel oxidation test (RPVOT) according to ASTM D-2272. KV and density were measured using an Anton Par SVM3001 Stabinger viscometer at 40 and 100 °C according to ASTM D-7042 and ASTM D-4052, respectively. VI was calculated using the ASTM D-2270 method. Gas chromatograph (Agilent 8890) equipped with a DB-HT-SIMDIS (5 m, 0.53 mm, 0.1 μm) column was used to execute the simulated distillation of the base oil samples according to the ASTM D-6352 method. The retention time was converted to boiling point by Agilent SimDis software. The COC inlet was set to ramp up at 35 °C/min to a final temperature of 400 °C. The flame ionization detector temperature was set to 450 °C, and the flow rates of hydrogen, air, and nitrogen were 32, 400, and 24 mL/min, respectively. In addition, the carbon number distribution of the base oil was determined based on the retention times of *n*-alkanes. The precision of the Agilent 8890 GC system performance has been evaluated using 5010 standards before testing the carbon number distribution on nine oil samples. The retention time and calibration model showed excellent separation and detection capabilities, and the results of 10 runs showed an average deviation of less than 2%. Considering all sources of uncertainty causing the measurements, three repeatability tests were performed on different batches of

https://doi.org/10.1021/acsomega.2c01877
ACS Omega 2022, 7, 18887−18896

Table 1. Properties of the Nine Oil Samples

sample	VI	pour point (°C)	flash point (°C)	KV@40 °C (mm^2/s)	KV@100 °C (mm^2/s)	density (g/cm^3)	evaporation loss	oxidation stability (min)
C#1	142	−39	211.5	19.25	4.38	0.801	18.8	19
C#2	150	−33	294.7	72.1	11.35	0.818	0.6	21
C#3	133	−39	230.9	18.45	4.18	0.802	14.5	18
C#4	142	−30	278.8	43.76	7.60	0.813	1.3	22
C#5	151	−36	239.4	30.64	6.10	0.808	5	24
M#1	130	−15	235.2	31.4	5.9	0.814	10.7	48
M#2	121	−21	248.2	47.8	7.4	0.826	5.1	26
M#3	84	−12	274.1	461.6	28	0.861	1.5	35
M#4	109	−21	146.5	29.3	5	0.831	16.8	27

bottled oil samples to ensure adequate representativeness of the test data. The properties of these different oil samples are given in Table 1.

2.3. ^{13}C NMR Spectroscopy. For mineral oils, ^{13}C NMR cannot distinguish between the same type of carbon atoms in different molecules, such as in chain alkanes and in naphthenic side chains. In addition, since cycloalkane carbon usually appears in ^{13}C NMR spectra as an envelope peak with chemical shifts ranging from 24 to 60 ppm, the choice of integration method has a significant impact on the calculation results of the naphthenic carbon (C_n) content. However, for CTL base oils, since they do not contain C_n, no baseline drift occurs and the high intermolecular similarity makes it easier and more accurate to determine the quantitative data on molecular structure. In this study, quantitative ^{13}C NMR spectra of base oil samples were performed on a Bruker AVANCE spectrometer at a resonance frequency of 400 MHz. ^{13}C NMR spectra were equipped with a 5 mm dual resonance broad-band inverse probe. The oil samples were diluted in $CDCl_3$ with 0.1 M $Cr(acac)_3$. $Cr(acac)_3$ was used as a relaxant to induce spin−lattice relaxation times, and TMS was used as an internal standard to measure chemical shifts. The application used an inverse gating decoupling scheme with a pulse width of 2.7 μs, a relaxation delay of 5 s, and an acquisition time of 1.5 s. A total of 20 000−24 000 scans were acquired for each spectrum. MestReNova software was used to collect and analyze the data, and each spectrum was processed three times and the average values were reported.

2.4. Regression Model. In petroleum science with small samples, the primary forces of statistical analysis are OLS and stepwise regression. OLS estimates the unknown parameters of an equation by minimizing the sum of squares of the differences between the sample values and the predicted values. OLS regression produces unstable results when there is a high degree of multicollinearity among the input variables, so it can be used as a means of testing for multicollinearity.[29] Stepwise regression builds the model by screening and eliminating variables that cause multicollinearity, so that the variables that are ultimately retained in the model are both significant and have no significant multicollinearity. In addition, ridge regression also provides a way to address multicollinearity. The ridge regression algorithm is a regularization method that reduces the sensitivity of the results to the training data set by introducing a small amount of bias, which suppresses the adverse effects of covariance on the predictions. In this study, ordinary least-squares (OLS) and stepwise regression methods were developed using the Minitab statistical software version 19.0 (Minitab Inc., State College, PA), while ridge regression models were developed using SPSS 26.0 software (IBM Corp., Ltd., New York).

3. RESULTS AND DISCUSSION

The bulk properties of CTL and mineral base oils, including VI, pour point, flash point, KV@40 °C, KV@100 °C, density, evaporation loss, oxidation stability, and distillation properties, were studied and compared. In addition, the relative contents of carbon types and branched structures were compared based on ^{13}C NMR results. Finally, VI regression models were developed based on ASPs, and the applicability of OLS, stepwise regression models, and ridge regression models was investigated in turn. Meanwhile, the physical significance of the effect of each structure on the overall VI was considered.

3.1. Measurement of Properties. *3.1.1. Physicochemical Property Analysis.* A comparison of the physicochemical properties of CTL and mineral base oils is presented in Table 1. In general, the physicochemical properties of CTL base oils differ from those of mineral base oils, while a similar isomerization process makes them similar in some respects. The KV of CTL base oils at 100 °C is similar to that of mineral base oils, but at 40 °C, the KV of mineral base oils is higher, resulting in a lower VI than that of CTL base oils. In addition, the higher pour points of the mineral base oils reflect poorer low-temperature fluidity than those of CTL base oils. Another characteristic of CTL base oil is its lower density compared to mineral base oil. In contrast, flash point and evaporation loss show no difference in these two types of base oils, with C#2, C#4, and M#3 having the lowest evaporation loss and M#4 having the lowest flash point. In addition, it was observed that evaporation loss and flash point showed a negative correlation; in general, the lower the evaporation loss, the higher the flash point. Among them, C#1 has the lowest flash point and the largest evaporation loss among CTL base oils, while the evaporation loss of mineral oil M#4 is similar to that of C#1, but the flash point is 65 °C lower. The reason for this is that the flash point indicates the lowest temperature at which a flash fire occurs and burns immediately when the mixture comes into contact with the flame, as shown in Figure 1; the content of light fraction (C14−C19) in M#4 is more than that of C#1, so the flash point is lower. The oxidation stability of mineral base oils is generally better than that of CTL base oils. The overall difference in the oxidation induction time in CTL base oil is small, and the difference between the longest and shortest times is only 5 min.

3.1.2. Simulated Distillation (SimDist) Analysis. The nine oil samples are divided from the SimDist analysis (by % wt) into light (C14−C23), medium (C33−C45), heavy (C46−C66), and super heavy (>C66) according to the number of carbons. Figure 1 shows that the carbon number distribution of CTL base oils is more concentrated overall compared to that of mineral base oils, showing more than 48% of the characterized fractions,

https://doi.org/10.1021/acsomega.2c01877
ACS Omega 2022, 7, 18887−18896

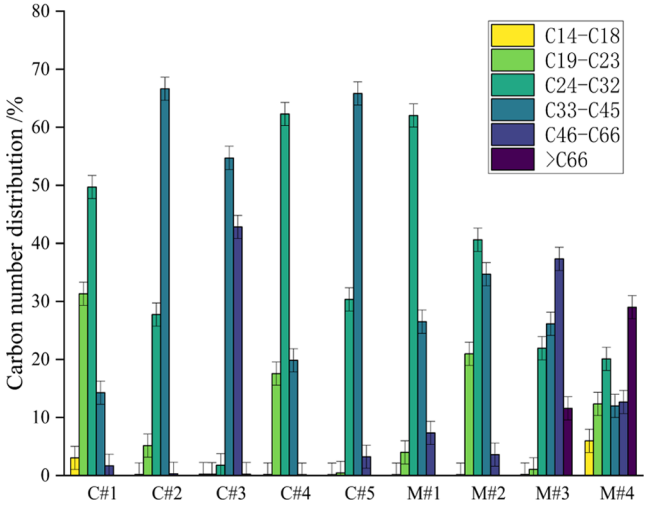


Figure 1. Carbon number distribution of base oils.

similar to M#1 and M#2 but much higher than the highest content fraction in M#3 and M#4. Among them, C14−C23 represents the relatively lightest fraction of nine base oils, and the lighter fraction with a large variation in content is considered to be the main factor affecting evaporation losses.

The actual cumulative yield of the base oil and the cumulative yield distribution curve obtained by the n-alkane boiling point calculation are shown in Figure 2, and the temperature

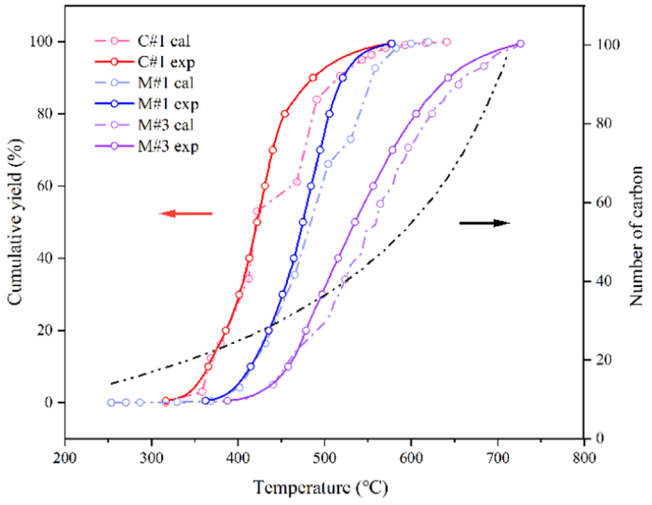


Figure 2. Cumulative yield of base oil with increasing temperature (% wt) and carbon number of *n*-paraffins corresponding to the boiling temperature.

corresponding to the n-alkane boiling point is derived from the study of Kudchadker et al.[30] It can be observed that the isomerization reduces the overall boiling point of the base oil and increases the efficiency of distillation, which is more obvious in the fractions with higher carbon numbers.

3.2. Measurement of ASPs. The ASPs of the base oil can be characterized in detail based on the peak positions of different types of carbon atoms provided by the ^{13}C NMR spectra. As shown in Table 2, the chemical shift assignments for the various carbon types were taken from Sarpal et al.[18,22]

The NMR spectra of CTL base oils differed significantly from those of mineral base oils, as shown in Figure 3. Taking C#4 and M#4 as an example, a severe drift of the baseline was observed in

Table 2. Algorithms for ASPs

^{13}C NMR parameter	chemical shift (ppm)
naphthenic carbons (C_n)	hump in region (60−24) ppm
paraffinic carbons (C_p)	(60−5) ppm - C_n
n-paraffinic α carbon (NP_α)	14.1 ppm
n-paraffinic β carbon (NP_β)	22.7 ppm
n-paraffinic γ carbon (NP_γ)	32.0 ppm
n-paraffinic δ or higher carbon (NP_n)	29.4 and 29.9 ppm
normal paraffins (NPs)	$NP_\alpha + NP_\beta + NP_\gamma + NP_n$
iso-paraffins (IPs)	C_p−NP
average chain length (ACL)	$2 \times NP/NP_\alpha$
various branched structures	
2-methyl-substituted (S_2)	28.2 ppm
3-methyl-substituted (S_3)	11.4 ppm
ethyl-substituted ($S_{3'}$)	10.7 ppm
4-methyl-substituted (S_4)	14.2 ppm
5-methyl-substituted (S_5)	14.3 ppm
6- or 7-methyl-substituted (S_{67})	27.0 ppm
2 or more methyl-substituted (S_8)	24−25.6 ppm

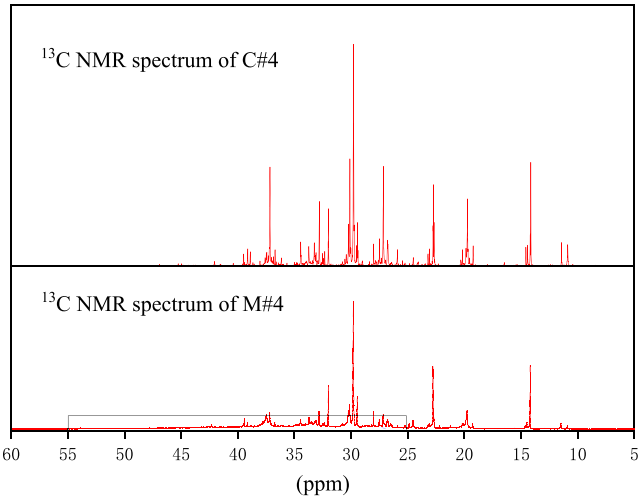


Figure 3. ^{13}C NMR spectrum of M#4 and C#4 base oils.

the M#4 spectrum compared with that in C#4, indicating the presence of relatively abundant C_n absorption peaks. In addition, the resonance signals of C#4 were stronger at 10−12, 14−15, 24−27, 32−34, and 36−40 ppm compared to that of M#4, which implies a higher content of branched structures. Based on the earlier study of peak assignments,[22] it was possible to distinguish specific branched structures, as shown in Figure 4, where the relative content of the carbonaceous units was recorded by normalizing the total carbon integral area, based on which the molar fractions of the different branched structures were calculated by the contribution to the IP. The relative contents of specific carbon types and branched structures were calculated and are shown in Table 3.

Table 3 shows the relative contents of carbon types and branched structures in base oil components. A certain fraction range of C_n significantly reduces the relative content of IP, which makes the biggest difference in the structure between CTL and mineral base oils, where CTL base oil has a higher branched-chain content than mineral base oils, such as structures S_3−S_7. First, structure $S_{3'}$ of CTL base oil has an average value of 7.904% at 10.9 ppm, which is higher than that of mineral base oils. The average value of structure S_{67} at 27 ppm is 29.866%, which is also higher than that of mineral base oil. Second, the

https://doi.org/10.1021/acsomega.2c01877
ACS Omega 2022, 7, 18887−18896

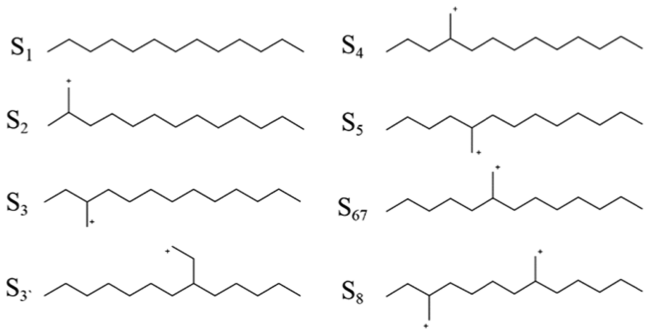


Figure 4. Various branched structures possible in base oils from ^{13}C NMR spectra (S_1 = 14.1 ppm, S_2 = 28.2 ppm, S_3 = 11.4 ppm, $S_{3'}$ = 10.7 ppm, S_4 = 14.2 ppm, S_5 = 14.3 ppm, S_{67} = 27.0 ppm, S_8 = 24.0−25.6 ppm) (+ indicates the atoms detected from the structure).

average values of structures at 11.4 (S_3), 14.4 (S_4), and 14.5 (S_5) ppm in CTL base oils are slightly higher than those of mineral base oils. In addition, ACL indicates the amount of carbon in the chain alkanes and also as part of the side chain attached to the cycloalkene ring, so the ACL of mineral base oils containing some of the cycloalkanes is lower than actual.

3.3. Relationship between VI and Base Oil Structures. VI reflects the variation of the viscosity of the base oil with temperature. *n*-Paraffins have the highest VI, and iso-paraffins have a slightly lower VI than *n*-paraffins. The presence of cycloalkanes and aromatic hydrocarbons negatively affects the overall VI of the base oil, and VI decreases as the number of rings increases.[31] In mineral base oils with a certain content of cycloalkanes, high VI depends on the relative content of monocyclic alkane and IP.[32] In addition, different levels of branching of iso-alkanes have different effects on the VI, and the key parameters for high VI point to molecules with the methyl branching structure at the center of the carbon chain or without ethyl branching.[21] Based on the collected ASPs of CTL base oils and mineral base oils, one-dimensional linear regression equations of VI and characteristic structure fractions are constructed as a way to determine whether there is a tight linear correlation.

In Figure 5, each graph plots the data points of VI versus the fraction of ASPs and gives the correlation coefficient (R^2). Among them, ACL, NPs, and structure S_{67} are positively correlated with VI, while other branched structures are negatively correlated with VI. The coefficients for the same molecular structure are in the same direction in both base oils, indicating a consistent effect of ASPs on VI. In addition, the values of the individual correlation coefficients are closer to 1 in the CTL base oil, indicating a stronger linear correlation between the ASPs and VI. However, in mineral base oils, the correlation between molecular structures and VI is not significant, such as C_n (R^2 = 0.118), NPs (R^2 = 0.184), $S_{3'}$ (R^2 = 0.0306), and S_8 (R^2 = 0.0297).

Among the various types of methyl structures of iso-paraffin, structure S_{67} is the only one that positively correlated with VI, and R^2 is 0.9005 in CTL base oils and 0.4629 in mineral base oils. This finding is in good agreement with previous studies that methyl branched chains in the center of carbon chains possess the ability to restrict molecular diffusion at high temperatures and thus exhibit high VI.[22] ACL also shows a similar correlation to VI in both base oils, with R^2 = 0.7597 in CTL base oil and R^2 = 0.6043 in mineral base oil. However, the positive effect of ACL on VI has been underestimated in some studies;[20] in mineral base oils, ACL does not only represent the length of carbon chains in normal and isoparaffinic chains but also represent part of the side chain that includes the linkage to the naphthenic hydrocarbons, so the ACL of mineral base oils is affected by the carbon content of naphthenic hydrocarbons, thus underestimating its positive effect.

3.4. Development of the Prediction Model. The CTL base oil ASPs obtained from the ^{13}C NMR quantification technique were used to build the regression model for predicting VI. In Table 4, the data points in the first five rows are the experimental results for CTL base oils shown above. In addition, the data in the last six rows of Table 4 were obtained from the six base oils synthesized using the Fischer−Tropsch technique.[33] Thus, in total, data from 11 experiments are used to construct the regression model for VI.

3.4.1. Variable Selection. It is increasingly critical how to select representative molecular structures. Presently, the focus is on reducing the covariance among the input key molecular structures, and grouping the input variables is a good way to reduce the covariance.[34] Since most of the chain alkanes constituting IP are lightly branched, mainly with one methyl group, a few with one ethyl group or two or more methyl groups located in the side chain, and their carbon chain lengths may be close to *n*-paraffins. Therefore, we treat ACL as an independent group, which represents the average size of base oil molecules. NPs, as a separate group, represents the molecular fraction of *n*-paraffins in the base oil. For the multiple branching structures in IP, two grouping strategies are considered. On the one hand, the position of the methyl branches in the chain alkanes has different effects on the rigidity of the molecular structure. For example, structure S_2 is near the end of the carbon chain and in a nonequilibrium position; such a structure is easily deformed at low temperatures, while structure S_{67} is near the center of the carbon chain and maintains a rigidity similar to that of *n*-

Table 3. Structural Parameters by ^{13}C NMR (in %) of Base Oils

		carbon-type compositions (%)			branched structures (mol % C)						
sample	ACL	C_n	NPs	IPs	S_2	S_3	$S_{3'}$	S_4	S_5	S_{67}	S_8
C#1	22.22	0	32.64	67.36	5.02	5.07	7.67	5.07	4.25	30.67	9.60
C#2	27.66	0	35.42	64.58	4.03	3.63	7.55	3.91	3.97	32.99	8.50
C#3	19.21	0	32.47	67.53	5.74	5.45	7.94	5.85	5.91	25.50	11.13
C#4	21.73	0	32.94	67.06	4.69	4.82	8.21	4.82	6.36	28.10	10.06
C#5	23.78	0	31.56	68.44	4.88	4.53	8.15	4.70	4.36	32.07	9.75
M#1	22.43	13.54	35.78	50.68	5.19	3.57	4.27	4.04	3.29	21.39	8.94
M#2	22.95	19.23	35.24	45.53	4.18	3.32	3.26	3.26	2.18	19.99	9.34
M#3	27.12	36.8	24.94	38.26	5.38	2.19	1.46	2.19	1.93	10.71	14.37
M#4	20.32	18.22	30.68	51.10	5.80	3.57	2.74	3.89	2.68	17.65	14.78

https://doi.org/10.1021/acsomega.2c01877
ACS Omega 2022, 7, 18887−18896

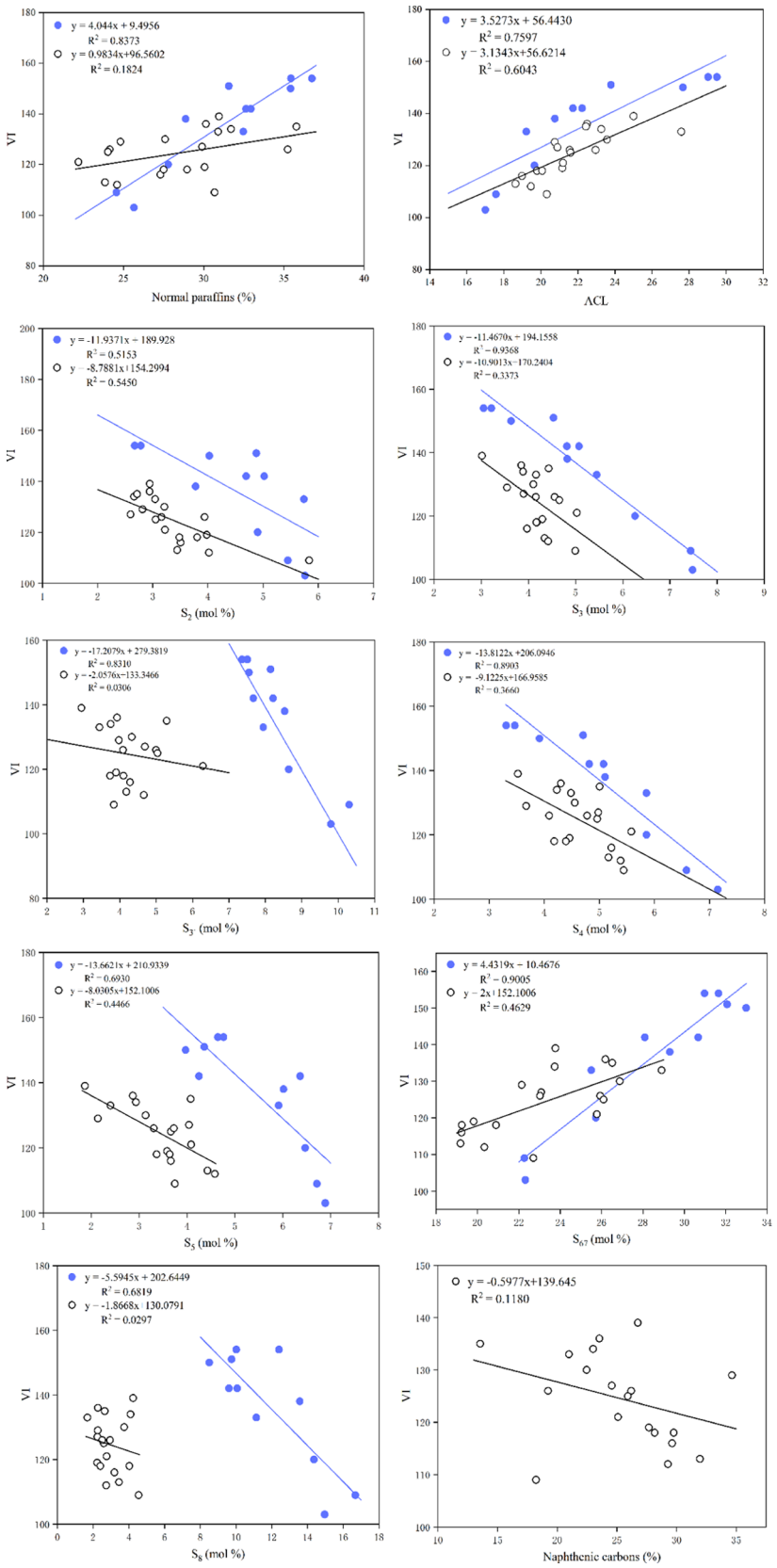


Figure 5. Correlation between VI and the fraction of molecular structure type (CTL and mineral base oils are indicated by filled blue circles and unfilled circles, respectively).

paraffins, which limits the mobility of the molecule at high temperatures. On the other hand, the grouping is based on the number and type of branching structures since they have different strengths of influence on VI. As shown in Figure 5, the coefficients of the input variables of the monomethyl branching equation (except S_{67}) are very different from structures $S_{3'}$ and

https://doi.org/10.1021/acsomega.2c01877
ACS Omega 2022, 7, 18887−18896

Table 4. Experimental Data Sets Used for Multivariable Regression Models of VI

	NPs	ACL	S_2	S_3	$S_{3'}$	S_4	S_5	S_{67}	S_8	VI
C#1	32.64	22.22	5.02	5.08	7.67	5.08	4.25	30.67	9.6	142
C#2	35.42	27.66	4.03	3.63	7.55	3.91	3.97	32.99	8.5	150
C#3	32.47	19.21	5.74	5.45	7.94	5.85	5.91	25.5	11.13	133
C#4	32.94	21.73	4.69	4.82	8.21	4.82	6.36	28.1	10.07	142
C#5	31.56	23.78	4.88	4.53	8.15	4.70	4.36	32.07	9.75	151
FT#1	25.65	17.00	5.76	7.48	9.80	7.15	6.89	22.31	14.96	103
FT#2	27.79	19.65	4.90	6.26	8.64	5.85	6.47	25.73	14.36	120
FT#3	36.75	29.5	2.78	3.22	7.36	3.46	4.76	31.66	10.02	154
FT#4	24.56	17.57	5.45	7.44	10.30	6.58	6.71	22.27	16.68	109
FT#5	28.88	20.75	3.78	4.83	8.53	5.10	6.01	29.3	13.57	138
FT#6	35.44	29.03	2.67	3.05	7.51	3.31	4.64	30.98	12.40	154

S_8, which means that structures $S_{3'}$ and S_8 have a weaker ability to reduce VI. Ultimately, the IPs are divided into four groups, structures S_2, S_3, S_4, and S_5 combined into one group (S_{2345}) and $S_{3'}$, S_{67}, and S_8 into their own groups.

3.4.2. OLS. It has been shown that the different findings are attributed to inconsistencies in the methodology and model selection. Among them, ordinary least-squares estimation is the simplest and most widely used regression estimation method.[35,36] In this work, to see whether grouping different molecular structures could resolve the effects caused by covariance, we use ordinary least squares for VI prediction modeling. Results of the OLS regression are shown in Table 5, with a p-value less than 0.01 and an R^2 of 0.977, indicating that the results of the OLS regression model are significant. NPs and structure S_{67} have a positive effect on VI; however, the coefficient (B) for the corresponding variable shows that the increase in NPs, structure S_{67}, and ACL decrease the VI, which is not consistent with the results in Section 3.3. The reason is that variance inflation factor (VIF) values of all variables are in the range of 33.213-inf, which indicates strong multicollinearity between these variables. Therefore, the regression results of OLS do not reflect the true relationship between the molecular structure and VI, and new estimation methods are needed to overcome this deficiency.

Table 5. Least-Squares Regression Coefficient[a]

	B	β	VIF	R^2	F	P
constant	0.1204		inf	0.977	42.936	0.0004
NPs	4.1059	0.929	inf			
ACL	−3.6288	−0.897	33.213			
S_{2345}	−3.3136	−0.842	inf			
$S_{3'}$	7.2571	0.384	inf			
S_{67}	2.9245	0.626	inf			
S_8	1.0670	0.158	inf			

[a]Inf stands for infinity.

3.4.3. Stepwise Regression. The stepwise regression method reduces the degree of multicollinearity by eliminating variables that are less important and highly correlated with other variables.[37,38] The model is performed with NPs, ACL, and structures S_{2345}, $S_{3'}$, S_{67}, and S_8 as independent variables and VI as a dependent variable. The result shows that the model constructed by NPs and structural S_{67} passes the F-test and the R^2 is 0.951, implying that the NP and structure S_{67} explained 95.1% of the variation in VI. From Table 6, the coefficient (B) of the corresponding variable for NP is 1.796, while that for structure S_{67} is 2.850, indicating that the effect of structure S_{67} on VI is greater than that of NPs. Although the VI of n-paraffins with the same carbon number is higher than iso-paraffins, the low carbon number distribution of n-paraffins may be responsible for this phenomenon. However, ACL also has a strong positive correlation with VI, as shown in Figure 5; when there is a large change in the ACL value, the VI may show a large deviation and poor generalization performance.

3.4.4. Ridge Regression. An alternative regression model is developed to determine the contribution of each molecular structure to the VI, in other words, to decompose the VI into a weighted sum of the individual structures. To alleviate the effect of multicollinearity on the model and decrease the error in data processing, taking logarithms for each variable of the model is a common treatment. The model is as follows

$$\ln \mathrm{VI} = \beta_0 + \beta_1 \ln \mathrm{NP} + \beta_2 \ln \mathrm{ALC} + \beta_3 \ln S_{2345} + \beta_4 \ln S_{3'} + \beta_5 \ln S_{67} + \beta_6 \ln S_8 \quad (1)$$

Considering the effect of multicollinearity among molecular structures, a ridge regression approach was used for modeling. Compared with the stepwise regression method, although the ridge regression analysis is a biased estimation method, it does not require the elimination of variables, so the obtained regression coefficients are more realistic and reliable than stepwise regression. To more intuitively represent the multicollinearity among variables, we calculated the Pearson correlation coefficients among the characteristic structural variables. The results of the correlation analysis are shown in Table 7. In statistics, the Pearson correlation coefficient is a measure of vector similarity. The output ranges from −1 to +1,

Table 6. Stepwise Regression Coefficient

	model	B	SE(B)	β	t	sig F	VIF	R^2
1	constant	10.473	14.023		0.747	0.474		0.901
	S_{67}	4.432	0.491	0.949	9.027	0.000	1.000	
2	constant	−0.899	11.177		−0.080	0.938		0.951
	S_{67}	2.850	0.662	0.610	4.303	0.003	3.277	
	NPs	1.796	0.627	0.406	2.865	0.021	3.277	

https://doi.org/10.1021/acsomega.2c01877
ACS Omega 2022, 7, 18887−18896

Table 7. Pearson Correlation Coefficients

variables	NPs	ACL	S_{2345}	$S_{3'}$	S_{67}
ACL	0.886				
S_{2345}	−0.900	−0.971			
$S_{3'}$	−0.948	−0.804	0.867		
S_{67}	0.834	0.860	−0.928	−0.863	
S_8	−0.853	−0.645	0.693	0.854	−0.822

with 0 representing no correlation, negative values representing negative correlation, and positive values representing positive correlation. As seen in Table 7, the correlation coefficient values between the variables range from 0.693 to 0.886 and −0.948 to −0.645, which indicates that the multicollinearity between them is significant. Based on the ridges shown in Figure 6, 0.147 is used as the ridge parameter (k) because the coefficients seem to stabilize around this value.

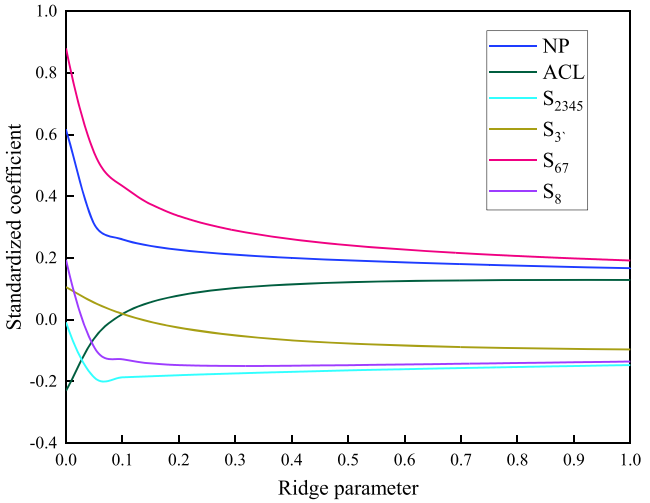


Figure 6. Ridge trace of the coefficient estimates of the ridge regression.

The results of the ridge regression are shown in Table 8, and the significance p-value of the regression model is 0.022, which presents a level of significance and rejects the original hypothesis, indicating the existence of a regression relationship between the independent and dependent variables. Meanwhile, the goodness-of-fit R^2 and ANOVA table reflect that the model is adequate and reasonable. The experimental data versus the model regression values are shown in Figure 7.

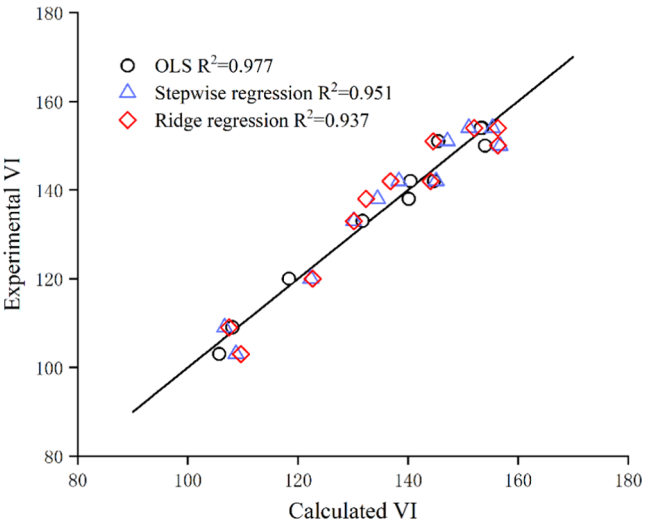


Figure 7. Deviation between the experimental and predicted VI values by least squares, stepwise regression, and ridge regression.

Based on the unnormalized coefficient B, the model obtained is shown in eq 2. It can be seen that the coefficients of the variables NPs, ACL, and structure S_{67} are positive, while the coefficients of the branched structures S_{2345}, $S_{3'}$, and S_8 are negative, which is consistent with the conclusions obtained from the correlation analysis. The unstandardized coefficient B for each structure highlights the extent to which the variation in the content of that structure affects the VI value and reflects the

Table 8. Model Result of Ridge Regression

ridge regression with k = 0.147				
Mult.R: 0.9682305829				
R-square: 0.9374704617				
Adj. R-square: 0.8436761542				
SE: 0.0554879472				
ANOVA table				
	df	SS	MS	
regress	6	0.185	0.031	
residual	4	0.012	0.003	
F value: 9.994961173				
sig F: 0.021549626				
variables in the equation				
	B	SE(B)	β	B/SE(B)
NPs	0.186150287	0.109814748	0.178210938	1.695130117
ACL	0.044820653	0.075792941	0.061647876	0.591356559
S_{2345}	−0.095218243	0.051409628	−0.159994372	−1.240327990
$S_{3'}$	−0.205859636	0.165971934	−0.158168683	−1.852148058
S_{67}	0.339286723	0.135637772	0.342602710	2.501417701
S_8	−0.057932194	0.083993669	−0.090185024	−0.689720957
constant	3.855484766	0.423333814	0.00000000	9.107433989

https://doi.org/10.1021/acsomega.2c01877
ACS Omega 2022, 7, 18887−18896

intrinsic reason for developing CTL base oil products with desirable molecular structures

$$\ln VI = 3.855 + 0.186 \ln NP + 0.045 \ln ACL - 0.095 \ln S_{2345} - 0.206 \ln S_{3'} + 0.339 \ln S_{67} - 0.058 \ln S_8 \quad (2)$$

4. CONCLUSIONS

In this study, the differences between CTL base oils and mineral base oils in terms of chemical structures and conventional properties were investigated. The main objective of this study was to determine the effect of each structural feature of CTL base oils on VI, to combine the characteristic molecular structures with similar functions, and to develop a VI prediction model consistent with the physical significance of each structure. The main conclusions can be drawn as follows:

(1) The molecular structure of CTL base oils is simpler than that of mineral base oils, and the main components are iso-paraffins. In this regard, the content of structure S_{67} is most different between the two base oils, and its average content in the iso-paraffins of CTL base oil is 29.866%, while it is 17.435% in the mineral oil.

(2) CTL base oils have a higher VI, lower flash point, density, and oxidation induction period than mineral base oils. At the same time, according to the results of high-temperature simulated distillation tests, CTL base oil has a narrower distillation range, lower distillation temperature, and higher distillation efficiency for similar carbon number distribution.

(3) From the correlation analysis, NPs, ACL, and structure S_{67} are the key factors for the high viscosity index of CTL base oils, and the increase of other branched-chain structure contents will reduce the viscosity index; structure $S_{3'}$ has the greatest impact.

(4) From the analytical data of ^{13}C NMR, the stepwise regression model has an R^2 of 0.951 and the ridge regression model has an R^2 of 0.937, but we considered that ridge regression is more reliable because it takes into account the physical significance of molecular structure and therefore obtains more realistic regression coefficients, which makes the model more powerful in generalization.

AUTHOR INFORMATION

Corresponding Author

Hanwen Wang − *Key Laboratory of Shaanxi Province for Development and Application of New Transportation Energy, Chang'an University, Xi'an 710064, China;* orcid.org/0000-0001-5790-0024; Phone: +86-18220171002; Email: whwenchd@126.com

Authors

Chunhua Zhang − *Key Laboratory of Shaanxi Province for Development and Application of New Transportation Energy, Chang'an University, Xi'an 710064, China;* orcid.org/0000-0002-0349-390X

Xiaowen Yu − *Key Laboratory of Shaanxi Province for Development and Application of New Transportation Energy, Chang'an University, Xi'an 710064, China*

Chaolin Peng − *Key Laboratory of Shaanxi Province for Development and Application of New Transportation Energy, Chang'an University, Xi'an 710064, China*

Angui Zhang − *CHN Energy Ningxia Coal Industry Co., Ltd., Yinchuan 750411, China*

Xuemei Liang − *CHN Energy Ningxia Coal Industry Co., Ltd., Yinchuan 750411, China*

Yinan Yan − *CHN Energy Ningxia Coal Industry Co., Ltd., Yinchuan 750411, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c01877

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was supported by the Key Research and Development Project of Ningxia Hui Autonomous Region (2018BDE02057), the Innovation Capability Support Program of Shaanxi (2021TD-28), and the Special Fund for Basic Scientific Research of Central Colleges, Chang'an University (300102221512, 300102221510, 300102222502).

REFERENCES

(1) Sosna, M. K.; Goltibeva, I. A.; Kononenko, A. A.; Zaichenko, V. A.; Grishina, I. N.; Korolev, E. V. Prospects of base oil production by GTL technology in Russia. *Chem. Technol. Fuels Oils* **2019**, *54*, 751−758.

(2) Zhang, Y.; Li, J.; Yang, X. Comprehensive competitiveness assessment of four coal-to-liquid routes and conventional oil refining route in China. *Energy* **2021**, *235*, No. 121442.

(3) Zhou, H.; Qian, Y.; Kraslawski, A.; Yang, Q.; Yang, S. Life-cycle assessment of alternative liquid fuels production in China. *Energy* **2017**, *139*, 507−522.

(4) Jenčík, J.; Honig, V.; Obergruber, M.; Hajek, J.; Vrablik, A.; Cerny, R.; Schlehofer, D.; Herink, T. Advanced biofuels based on Fischer-Tropsch synthesis for applications in diesel engines. *Materials* **2021**, *14*, No. 3077.

(5) Wang, L.; Li, G.; Guo, Q.; Xia, G. The hydrogenation upgrading industrial test to produce lube and wax from F-T waxes. *Pet. Process. Petrochem.* **2016**, *47*, 5−9.

(6) Zhang, Y.; Wang, Y.; Zhang, Z. Comparative analysis of products from Fischer-Tropsch oil and petroleum based oil. *Chem. Indus. Engin. Prog.* **2018**, *37*, 3781−3787.

(7) Neuner, P.; Graf, D.; Mild, H.; Rauch, R. Catalytic hydroisomerisation of Fischer−Tropsch waxes to lubricating oil and investigation of the correlation between its physical properties and the chemical composition of the corresponding fuel fractions. *Energies* **2021**, *14*, No. 4202.

(8) Dean, E. W.; Davis, G. H. B. Viscosity variations of oils with temperature. *Chem. Met. Eng.* **1929**, *36*, No. 618.

(9) Sperber, O.; Kaminsky, W.; Geißler, A. Structure analysis of paraffin waxes by 13C-NMR spectroscopy. *Petrol. Sci. Technol.* **2005**, *23*, 47−54.

(10) Jones, H. E.; Palacio Lozano, D. C.; Huener, C.; Thomas, M. J.; Aaserud, D. J.; DeMuth, J. C.; Robin, M. P.; Barrow, M. P. Influence of biodiesel on base oil oxidation as measured by FTICR mass spectrometry. *Energy Fuels* **2021**, *35*, 11896−11908.

(11) Noh, K.; Shin, J.; Lee, J. H. Change of hydrocarbon structure type in lube hydroprocessing and correlation model for viscosity index. *Ind. Eng. Chem.* **2017**, *56*, 8016−8028.

(12) Poveda, J. C.; Molina, D. R. Average molecular parameters of heavy crude oils and their fractions using NMR spectroscopy. *Pet. Sci. Eng.* **2012**, *84*−*85*, 1−7.

(13) Strahan, G. D.; Mullen, C. A.; Boateng, A. A. Characterizing biomass fast pyrolysis oils by 13C NMR and chemometric analysis. *Energy Fuels* **2011**, *25*, 5452−5461.

(14) Zhang, P.; Lu, S.; Li, J.; Chen, C.; Xue, H.; Zhang, J. Petrophysical characterization of oil-bearing shales by low-field nuclear magnetic resonance (NMR). *Mar. Pet. Geol.* **2018**, *89*, 775−785.
(15) Abdul Jameel, A. G.; Khateeb, A.; Elbaz, A. M.; Emwas, A. H.; Zhang, W.; Roberts, W. L.; Sarathy, S. M. Characterization of deasphalted heavy fuel oil using APPI (+) FT-ICR mass spectrometry and NMR spectroscopy. *Fuel* **2019**, *253*, 950−963.
(16) AlHumaidan, F. S.; Hauser, A.; Rana, M. S.; Lababidi, H. M. S. NMR Characterization of asphaltene derived from residual oils and their thermal decomposition. *Energy Fuels* **2017**, *31*, 3812−3820.
(17) Mäkelä, V.; Karhunen, P.; Siren, S.; Heikkinen, S.; Kilpeläinen, I. Automating the NMR analysis of base oils: Finding napthene signals. *Fuel* **2013**, *111*, 543−554.
(18) Sarpal, A. S.; Kapur, G. S.; Chopra, A.; Jain, S. K.; Srivastava, S. P.; Bhatnagar, A. K. Hydrocarbon characterization of hydrocracked base stocks by one- and two-dimensional NMR spectroscopy. *Fuel* **1996**, *75*, 483−490.
(19) Sarpal, A. S.; Kapur, G. S.; Mukherjee, S.; Jain, S. K. Characterization by 13C NMR spectroscopy of base oils produced by different processes. *Fuel* **1997**, *76*, 931−937.
(20) Sharma, B. K.; Adhvaryu, A.; Perez, J. M.; Erhan, S. Z. Effects of hydroprocessing on structure and properties of base oils using NMR. *Fuel Process. Technol.* **2008**, *89*, 984−991.
(21) Verdier, S.; Coutinho, J. A. P.; Silva, A. M. S.; Alkilde, O. F.; Hansen, J. A. A critical approach to viscosity index. *Fuel* **2009**, *88*, 2199−2206.
(22) Sarpal, A. S.; Sastry, M. I. S.; Bansal, V.; Singh, I.; Mazumdar, S. K.; Basu, B. Correlation of structure and properties of groups I to III base oils. *Lubr. Sci.* **2012**, *24*, 199−215.
(23) Haus, F.; Boissel, O.; Junter, G. A. Multiple regression modelling of mineral base oil biodegradability based on their physical properties and overall chemical composition. *Chemosphere* **2003**, *50*, 939−948.
(24) Sastry, M. I. S.; Chopra, A.; Sarpal, A. S.; Jain, S. K.; et al.*et al.* Determination of physicochemical properties and carbon-type analysis of base oils using mid-IR spectroscopy and partial least-squares regression analysis. *Energy Fuels* **1998**, *12*, 304−311.
(25) Adhvaryu, A.; Erhan, S. Z.; Sahoo, S. K.; Sing, I. D. Thermo-oxidative stability studies on some new generation API group II and III base oils. *Fuel* **2002**, *81*, 785−791.
(26) Heredia-Langner, A.; Cort, J. R.; Grubel, K.; O'Hagan, M. J.; Jarman, K. H.; Linehan, J. C.; Albrecht, K. O.; Polikarpov, E.; King, D. L.; Smurthwaite, T. D.; Bays, J. T. Methodology for the development of empirical models relating 13C NMR spectral features to fuel properties. *Energy Fuels* **2020**, *34*, 12556−12572.
(27) Xie, C.; Hawkes, A. D. Estimation of inter-fuel substitution possibilities in China's transport industry using ridge regression. *Energy* **2015**, *88*, 260−267.
(28) Liu, H.; Miao, E. M.; Wei, X. Y.; Zhuang, X. D. Robust modeling method for thermal error of CNC machine tools based on ridge regression algorithm. *Int. J. Mach. Tools Manuf.* **2017**, *113*, 35−48.
(29) Rahiminezhad, A. A Comparison of Partial Least Squares (PLS) and Ordinary Least Squares (OLS) regressions in predicting of couple. *Procedia Soc. Behav. Sci.* **2010**, *5*, 1459−1463.
(30) Kudchadker, A. P.; Zwolinski, B. J. Vapor pressures and boiling points of normal alkanes C21 to C100. *J. Chem. Eng. Data* **1966**, *11*, 253−255.
(31) Lee, S. K.; Rosenbaum, J. M.; Hao, Y.; Lei, G. D. Premium lubricant base stocks by hydroprocessing. *Springer Handb.Pet. Technol.* **2017**, 1015−1042.
(32) Wang, Q.; Ling, H.; Shen, B. X.; Li, K.; Ng, S. Evaluation of hydroisomerization products as lube base oils based on carbon number distribution and hydrocarbon type analysis. *Fuel Process. Technol.* **2006**, *87*, 1063−1070.
(33) Zhang, Z.; Zhang, Y.; Gao, j; Liu, D. NMR characterization of lube base oil structure. *Pet. Process. Petrochem.* **2019**, *50*, 91−96.
(34) Miao, E. M.; Gong, Y. Y.; Niu, P. C.; Ji, C. Z.; Chen, H. D. Robustness of thermal error compensation modeling models of CNC machine tools. *Int. J. Adv. Manuf. Technol.* **2013**, *69*, 2593−2603.
(35) Lin, B.; Benjamin, I. N. Causal relationships between energy consumption, foreign direct investment and economic growth for MINT: Evidence from panel dynamic ordinary least square models. *J. Clean. Prod.* **2018**, *197*, 708−720.
(36) Mirezi, B.; Kaçıranlar, S.; Özbay, N. A minimum matrix valued risk estimator combining restricted and ordinary least squares estimators. *Commun. Stat. Theory Methods* **2021**, 1−11.
(37) Wang, M.; Wright, J.; Brownlee, A.; Buswell, R. A comparison of approaches to stepwise regression on variables sensitivities in building simulation and analysis. *Energy Build.* **2016**, *127*, 313−326.
(38) Liao, X.; Li, Q.; Yang, X.; Zhang, W.; Li, W. Multiobjective optimization for crash safety design of vehicles using stepwise regression model. *Struct. Multidiscip. Otim.* **2008**, *35*, 561−569.

https://doi.org/10.1021/acsomega.2c01877
ACS Omega 2022, 7, 18887−18896