Detailed Composition-Based Model for Predicting the Cetane Number of Diesel Fuels

Prasenjeet Ghosh* and Stephen B. Jaffe†

Compositional Modeling Group, ExxonMobil Process Research Laboratories, Paulsboro Technical Center, Paulsboro, New Jersey 08066

We have developed a simple composition-based model for predicting the cetane number of diesel fuels with general applicability to any diesel fuel regardless of the refining process it originates from. The cetane number is correlated to a total of 129 different hydrocarbon lumps determined by a combination of supercritical fluid chromatography, gas chromatography, and mass spectroscopic methods. A total of 203 diesel fuels are considered in this study derived from various diesel-range refinery process streams and their commercial blends. Across the multitude of such process streams and blends, the model predicts the cetane number with a standard error of 1.25 numbers, which is well within the experimental error of the measurement.

1. Introduction

Cetane number (CN) is a measure of the ignition quality of the diesel fuel and is determined by a standard engine test as specified by ASTM (ASTM D613). The ignition quality is quantified by measuring the ignition delay, which is the period between the time of injection and the start of combustion (ignition) of the fuel. A fuel with a high CN has a short ignition delay period and starts to combust shortly after it is injected into an engine. The ignition quality of the diesel fuel depends on its molecular composition. Some of the simpler molecular components such as the \( n \)-paraffins can ignite in a diesel engine with relative ease, but others like aromatics have more stable ring structures that require higher temperature and pressure to ignite.

The ASTM D613 test method defines the CN of a diesel fuel as the percentage by volume of normal cetane (\( C_{16}H_{34} \)), in a blend with 2,2,4,4,6,8,8-heptamethylnonane (sometimes called HMN or isocetane), which matches the ignition quality of the diesel fuel being rated under the specified test conditions. By definition, normal cetane has been assigned a CN of 100, whereas HMN has a CN of 15. The ASTM D613 method involves running the fuel in a single-cylinder compression ignition engine with a continuously variable compression ratio under a fixed set of conditions. Although it has been the traditional test method for CN, it was soon recognized that the test suffered from many disadvantages, some of which include a relatively large fuel sample volume requirement (~1 L), significant time consumption (approximately a few hours), and a relatively high reproducibility error. ASTM documents a maximum reproducibility error of 3–4 numbers for this test. Consequently, there have been many attempts to develop alternate tests to replace the ASTM D613 method. These include devising better engine tests and developing correlative models to predict the CN from bulk properties of the fuel that may be measured more quickly and reliably. Among the different engine tests developed, the Ignition Quality Tester (IQT) has been the most promising for reliable measurement of CNs of middle distillates and alternate fuels. It uses much smaller sample volumes (~50 mL) and requires much shorter times (~10 min) for completion of the test. The IQT consists of a constant-volume combustion chamber that is electrically heated, a fuel injection system, and a computer that is used to control the experiment, record the data, and provide interpretation of the data. The basic concept of the IQT is the measurement of the time from the start of fuel injection into the combustion chamber to the start of combustion. The thermodynamic conditions in the combustion chamber are influenced by the charge pressure (combustion air), chamber temperature, and fuel injection quantity. These conditions are precisely controlled to provide consistent measurement of the ignition delay time. The measured ignition delay time is correlated by an equation to a derived CN. We call it derived to highlight the fact that the CN was not measured in the actual ASTM engine. However, in terms of actual numbers, the IQT-based derived CN is the same as what we would have measured if we had run the fuel through the actual ASTM engine. The reproducibility errors on IQT-based derived CN are tighter than the those of ASTM D613 and vary between 1 and 2 numbers. The present work is therefore based on IQT measurements on the various diesel fuels and their commercial blends.

Numerous attempts have been made in the past to correlate the CN with various physical and chemical attributes of the diesel fuel. These include correlations based on either bulk properties such as API gravity, boiling points, and aniline points or based on molecular composition, measured by gas chromatography (GC) or inferred from different spectroscopic methods such as Fourier transform infrared spectroscopy, nuclear magnetic resonance (\(^1\)H NMR, \(^{13}\)C NMR), etc. An exhaustive compilation of the correlations based on bulk properties can be found in ref 3, among which the most commonly used correlation is the ASTM D4737 equation. The D4737 equation correlates the API gravity and the \( T_{10}, T_{50}, \) and \( T_{90} \) points on the boiling curve to the CN of the fuel and is summarized in eq 1

\[
CN = 45.2 + 0.0892T_{10} + (0.131 + 0.901B_N)T_{50} + (0.0523 - 0.42B_N)T_{90} + [0.00049(T_{10}^2 - T_{90}^2)] + 107B_N + 60B_N^2
\]

where \( d \) is specific gravity at 60 °F, \( B_N = e^{-3.5(d-0.85)} - 1 \), \( T_{10} = T_{10} - 215 \), \( T_{50} = T_{50} - 260 \), \( T_{90} = T_{90} - 310 \), and \( T_c \) is the temperature (in °C) at which \( x \) vol % of the sample has distilled off. Despite its empirical nature, this equation has enjoyed enviable success in describing the CN of various diesel fuels and their blends and is used extensively within the petroleum industry. However, with the increasing trends toward
low sulfur diesel fuels and the addition of cetane improvers in the fuel, the results of this equation have become progressively less satisfactory, and newer and better models are sought. In contrast to bulk property correlations, many predictive models based on molecular composition have also been proposed to predict CN. These include models based on NMR and FTIR spectra and the molecular composition measured by a combination of liquid chromatography and GC—mass spectroscopy (MS) techniques. Although composition-based, most of these models are linear (and occasionally nonlinear) principal component regression models and do show encouraging results in predicting CN. However, the lack of a phenomenologically motivated mathematical structure for the functional form and the assumption of linearity (in most cases) make such models inadequate for extrapolation. Further, these models are typically refinary or application specific and require frequent updating, thus restrictive in their application for any diesel fuel. The underlying common thesis of all of the above models is the implicit or partial acknowledgment of the molecular composition of the diesel fuel, whether it is through density, the boiling curves, or the NMR and FTIR spectra. Because CN depends on the molecular composition of the fuel, we believe that an explicit and complete acknowledgment of the detailed molecular composition is necessary to build a predictive model. This is a prerequisite to make the model quantitatively superior and robust in its predictions, especially when used in extrapolation. Our objective in this study is to develop such a predictive model as a function of the molecular composition of the fuel. This model would be used to support various product quality predictions for diesel fuels in ExxonMobil’s refineries worldwide. Because the model depends purely on composition, it would also be integrated into our existing network of detailed process kinetic models. It will also be used in identifying the important hydrocarbon molecules that have a significant impact on the diesel fuel quality and that are important for meeting future transportation fuel specifications. The details of such a model development are presented in the next section.

2. Development of the Cetane Model

The development of the cetane model is analogous to that of the octane model for gasoline fuels reported in ref 13. Each diesel fuel, regardless of the process stream (e.g., kero cut, heavy cat naphtha, light cycle oil, heavy gas oil, etc.) is a mixture of many hydrocarbon molecules, each contributing to the CN of the diesel fuel. Let CN denote the measured cetane number for the diesel fuel while CN, represent the pure-component cetane number of each molecule i in the fuel. Because a molecule may not necessarily always contribute its pure-component cetane number to the diesel fuel, each molecule’s contribution toward the fuel cetane number is quantified by its blend value, denoted by CN, i. The blend value of a molecule depends on the overall composition of the fuel it is part of. Figure 1 shows the difference between CN, and CN using the example of a binary blend. By definition, CN is a linear volumetric blend of the blend contributions of all of the different molecules present in the diesel fuel; thus, it follows that

\[ \text{CN} = \sum_i v_i \text{CN}_i \]  

where \( v_i \) is the volume fraction of molecule \( i \) in the fuel. We assume that the blend value (CN,) of molecule \( i \) in a diesel fuel varies linearly with the diesel CN. Therefore, \( B_{i}^{\text{CN}} = a_i^{(0)} + a_i^{(1)} \text{CN} \)

Equation 3 is analogous to the equation for any partial molar property, where the partial molar property (CN, i) depends on the composition of the diesel fuel it blends into. CN, is parametrized by two parameters, a slope \( a_i^{(1)} \) and an intercept \( a_i^{(0)} \), thereby implying a linear variation of blend values with diesel composition. Equation 3 is based on the experimental observation made for octane numbers and gasoline fuels in our laboratory and by analogy extended for CNs. For instance, in ref 14, it was shown that the blend numbers of various n-paraffins varied almost linearly with the overall octane number of the gasoline fuel. Because both octane numbers and CNs are dictated by similar combustion chemistries (autoignition reactions) of the fuel, we believe that the observations for one may be extended to the other.

Although the variation of the blend value of the individual molecules with the CN of the diesel fuel is taken to be linear, this does not imply a linear relationship of the fuel composition to its CN. Using the definition for CN from eq 2 in eq 3, it is easy to see that eq 3 captures multicomponent interaction. Equation 3 requires two parameters, \( a_i^{(0)} \) and \( a_i^{(1)} \), for each molecule; however, it is possible to eliminate one of these parameters using the special case when the diesel fuel is a pure component. For instance, if the diesel fuel is pure n-cetane, then the CN of the fuel is the same as the pure-component number for n-cetane, which, in turn, is the same as its blend value. This boundary condition is necessary to ensure that the model returns the pure-component value when the fuel is a pure-component stream. Application of this boundary condition yields

\[ \text{CN}_i = a_i^{(0)} + a_i^{(1)} \text{CN}_i \]

Further, if we define \( \beta_i = 1 - a_i^{(1)} \), eq 4 may be rewritten as

\[ a_i^{(0)} = \beta_i \text{CN}_i \]

Rearranging eqs 2, 3, and 5, we get

\[ \text{CN} = \frac{\sum_i v_i \beta_i \text{CN}_i}{\sum_i v_i} = \frac{\sum_i v_i \beta_i \text{CN}_i}{\sum_i v_i \beta_i} \]

This is the final form of the cetane prediction model, where the...
summation index $i$ runs over all of the molecules present in the
diesel fuel. However, a typical diesel fuel contains innumerable
different hydrocarbon molecules. Therefore, some amount of
molecular lumping is necessary to build any realistic model.
This is reflected in the second part of eq 6, where we replace
the summation variable $i$ with “lumps” to indicate that the sum
runs over only the molecular lumps relevant for the cetane
model. A lump defines the compositional level of abstraction
that the model uses. It could be a single molecule like $n$-cetane
or a group of similar molecules such as $C_{18}$ $i$-paraffins. Note
that, in light of this lumping, the term pure-component CN for
lump $i$ may not always necessarily reflect the CN of a single
molecule. For molecular lumps that correspond to single
molecules, $CN_i$ is the pure-component CN of that single
molecule, but for molecular lumps that correspond to a group
of similar molecules, $CN_i$ is the average of the pure-component
CNs of all of the different molecules in that lump. These lumps
are described in greater detail in the next section.

$\beta_i$'s are the adjustable parameters of the model and represent
whether a molecule contributes beneficially or detrimentally to
the CN of the diesel fuel. A contribution is considered beneficial
if the blend value of the molecule/lump ($B_i^{\text{CN}}$) is greater than
its pure-component cetane number ($CN_i$). Likewise, the
contribution is detrimental if $B_i^{\text{CN}} < CN_i$. For the two cases (a) $\beta_i < 1$ and $CN_i < CN$ and (b) $\beta_i > 1$ and $CN_i > CN$, the lump $i$
contributes beneficially to the fuel CN. For the other two cases,
(a) $\beta_i > 1$ and $CN_i < CN$ and (b) $\beta_i < 1$ and $CN_i > CN$, it
contributes detrimentally to the fuel CN, $\beta_i = 1$ is the special
case where the lump contributes equally to its pure-component
CN.

The model depicted in eq 6 is simple and demonstrates the
contribution of each molecular lump toward the fuel CN. Its
inputs are the volumetric composition of the fuel ($x_i$), the pure-
component CNs of the different molecules (or molecular lumps)
($CN_i$), and their blend contributions parametrized by $\beta_i$. The
model structure is consistent in predicting both the CN of a
typical diesel fuel and that of a pure component. In the limiting
case when the diesel fuel is a pure component ($x_i = 1$), the
model returns the pure-component cetane number ($CN_i$).

### 3. Defining the Molecular Lumps

We now need to define the various molecular (or composi-
tional) lumps that will be used in the cetane model. The
molecular lumps are based on the following two criteria: first,
significant CN difference and, second, analytical resolution.
Thus, two molecules (or molecular lumps) are considered
distinct and included separately in the model if there is a
significant difference in their pure-component CNs and if the
analytical resolution between these two molecules (or molecular
lumps) is also available. We have chosen a total of 129
molecular lumps in this work belonging to 9 different hydro-
carbon classes. These include 12 lumps for normal paraffins,
21 lumps for each of the monobranched and multibranched
$i$-paraffins, 11 lumps for alkylcyclohexanes, 5 lumps for
decalins, 15 lumps for alkylbenzenes, 14 lumps for naphthenales,
16 lumps for tetrains, and 14 lumps for olefins. These lumps
are summarized in Table 1. In choosing these lumps, a number
of additional assumptions were made as noted below:

(i) The branching position is ignored in the lumping of mono-
and multibranched $i$-paraffins because such detail is difficult
to measure by the present analytical capabilities in diesel-range
fuels.

(ii) Multibranched $i$-paraffins represent all $i$-paraffins with
two or more branches.

<table>
<thead>
<tr>
<th>Lumps</th>
<th>Molecular class</th>
<th>Carbon number range</th>
</tr>
</thead>
<tbody>
<tr>
<td>nc5</td>
<td>$n$-paraffins</td>
<td>$nC_5$ to $nC_{16+}$</td>
</tr>
<tr>
<td>nc6</td>
<td>monobranch i-paraffins</td>
<td>$iC_5$ to $iC_{25+}$</td>
</tr>
<tr>
<td>nc7</td>
<td>multibranch i-paraffins</td>
<td>$iC_5$ to $iC_{25+}$</td>
</tr>
<tr>
<td>nc8</td>
<td>alkylcyclohexanes</td>
<td>cyclohexane to $C_{10}$ cyclohexanes</td>
</tr>
<tr>
<td>nc9</td>
<td>decalins</td>
<td>decalin to $C_4$ decalins</td>
</tr>
<tr>
<td>nc10</td>
<td>naphthenales</td>
<td>naphthalene to $C_{13+}$ naphthenales</td>
</tr>
<tr>
<td>nc11</td>
<td>alkylbenzenes</td>
<td>benzene to $C_{14+}$ benzenes</td>
</tr>
<tr>
<td>nc12</td>
<td>tetrains</td>
<td>tetralin to $C_{13+}$ tetrains</td>
</tr>
<tr>
<td>nc13</td>
<td>olefins</td>
<td>$C_3$ to $C_{18+}$ olefins</td>
</tr>
</tbody>
</table>

### Table 2. Pure-Component CN for the Various Lumps Considered in the Model

<table>
<thead>
<tr>
<th>Lumps</th>
<th>Molecular class</th>
<th>Carbon number range</th>
</tr>
</thead>
<tbody>
<tr>
<td>nc5</td>
<td>$n$-paraffins</td>
<td>$nC_5$ to $nC_{16+}$</td>
</tr>
<tr>
<td>nc6</td>
<td>monobranch i-paraffins</td>
<td>$iC_5$ to $iC_{25+}$</td>
</tr>
<tr>
<td>nc7</td>
<td>multibranch i-paraffins</td>
<td>$iC_5$ to $iC_{25+}$</td>
</tr>
<tr>
<td>nc8</td>
<td>alkylcyclohexanes</td>
<td>cyclohexane to $C_{10}$ cyclohexanes</td>
</tr>
<tr>
<td>nc9</td>
<td>decalins</td>
<td>decalin to $C_4$ decalins</td>
</tr>
<tr>
<td>nc10</td>
<td>naphthenales</td>
<td>naphthalene to $C_{13+}$ naphthenales</td>
</tr>
<tr>
<td>nc11</td>
<td>alkylbenzenes</td>
<td>benzene to $C_{14+}$ benzenes</td>
</tr>
<tr>
<td>nc12</td>
<td>tetrains</td>
<td>tetralin to $C_{13+}$ tetrains</td>
</tr>
<tr>
<td>nc13</td>
<td>olefins</td>
<td>$C_3$ to $C_{18+}$ olefins</td>
</tr>
</tbody>
</table>

*mono-iC means a monomethyl $i$-parafin with a total of $x$ carbon atoms. $C_i$ cyclohexanes represent alkylcyclohexanes only with a total of $x$ carbon atoms in the alkyl chain. A similar convention is followed for the other molecular classes.

### Table 1. Various Molecular Lumps Considered in the Cetane Model

<table>
<thead>
<tr>
<th>Molecular class</th>
<th>Carbon number range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$-paraffins</td>
<td>$nC_5$ to $nC_{16+}$</td>
</tr>
<tr>
<td>monobranch $i$-paraffins</td>
<td>$iC_5$ to $iC_{25+}$</td>
</tr>
<tr>
<td>multibranch $i$-paraffins</td>
<td>$iC_5$ to $iC_{25+}$</td>
</tr>
<tr>
<td>alkylcyclohexanes</td>
<td>cyclohexane to $C_{10}$ cyclohexanes</td>
</tr>
<tr>
<td>decalins</td>
<td>decalin to $C_4$ decalins</td>
</tr>
<tr>
<td>naphthenales</td>
<td>naphthalene to $C_{13+}$ naphthenales</td>
</tr>
<tr>
<td>alkylbenzenes</td>
<td>benzene to $C_{14+}$ benzenes</td>
</tr>
<tr>
<td>tetrains</td>
<td>tetralin to $C_{13+}$ tetrains</td>
</tr>
<tr>
<td>olefins</td>
<td>$C_3$ to $C_{18+}$ olefins</td>
</tr>
</tbody>
</table>

*mono-iC means a monomethyl $i$-parafin with a total of $x$ carbon atoms. $C_i$ cyclohexanes represent alkylcyclohexanes only with a total of $x$ carbon atoms in the alkyl chain. A similar convention is followed for the other molecular classes.

(iii) Alkyl side chains in all ring molecules are assumed to be $n$-alkyl.

(iv) Alkylcyclohexanes are 1-ring cyclohexanes with $n$-alkyl side chains.

(v) Decalins represent multiring naphthenes ($\geq 2$ rings) with $n$-alkyl side chains.
(vi) Alkylbenzenes are 1-ring aromatics with \( n \)-alkyl side chains.

(vii) Naphthalenes, like decalins, are multiring aromatics (\( \geq 2 \) rings) with \( n \)-alkyl side chains. Biphenyls are also included in this group.

(viii) Tetralins represent multiring (\( \geq 2 \) rings) naphthenoaromatics with \( n \)-alkyl side chains. Indanes are included in this group.

(ix) Olefins represent all three types of olefins: linear olefins, cycloolefins, and benzoolefins. Indenes are also included in this group.

(x) All heteroatom molecules containing sulfur and nitrogen are lumped into naphthalenes because most such structures contain two or more aromatic rings.

Parts a and b of Figure 2 show the pure-component CNs for the 129 molecular lumps considered in this work. The data are based on an exhaustive compilation of the CNs for various hydrocarbon molecules published in the literature. The abscissa represents the total number of carbon atoms in the molecule, while the ordinate indicates its pure-component CN. \( n \)-Paraffins have the highest CNs, and it increases with the carbon number; i.e., the longer the backbone carbon chain, the higher the CN. The addition of branches decreases the CN, which explains the lower CNs for \( i \)-paraffins compared to those of \( n \)-paraffins. The greater the degree of branching, the lower the CN. The position of the branch also plays a role (although secondary) in the CN. If the branching is concentrated at one end of the molecule, leaving a long chain at the other end, then such \( i \)-paraffins tend to have higher CNs. Compared to \( i \)-paraffins, olefins have higher CNs and follow closely the \( n \)-paraffin trend. Aromatics and naphthenes typically have lower CNs, unless they have a long \( n \)-paraffinic side chain. Multiring structures have lower CNs compared to the corresponding single-ring structures, which can be seen by comparing the alkylcyclohexanes with alkyldecalins or alkylbenzenes with the corresponding alkyltetralins and naphthalenes. Although decalins are slightly higher in CN than both tetralins and naphthalenes, the difference is marginal. The actual numerical values for these CN trends, as shown in Figure 2a,b, are summarized in Table 2. These numbers will be used for CN, in eq 6. Beyond a certain carbon number, the molecules

---

**Figure 2.** Pure-component CNs for different molecular lumps considered in the cetane model. The curves indicate the general trend of cetane variation across each hydrocarbon class as a function of the number of carbon atoms in the molecule. The chemical structures associated with each class are shown only for visual clarity. Also, for ease of comparison across both parts a and b, \( n \)-paraffin curves are plotted in both parts.
Table 3. Parameter Values

<table>
<thead>
<tr>
<th>molecular class</th>
<th>molecular lumps</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-paraffins</td>
<td>nC5 to nC16</td>
<td>0.5212</td>
</tr>
<tr>
<td>i-paraffins</td>
<td>iC5 to iC25, mono- and multibranched</td>
<td>7.3717</td>
</tr>
<tr>
<td>naphthenes</td>
<td>cyclohexane to C10+ naphthenes,</td>
<td>0.0727</td>
</tr>
<tr>
<td></td>
<td>decalin to C4+- decalins</td>
<td></td>
</tr>
<tr>
<td>aromatics</td>
<td>benzene to C14+ aromatics,</td>
<td>3.1967</td>
</tr>
<tr>
<td></td>
<td>naphthalene to C13+, naphthenalenes,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>tetralin to C15+ tetralins</td>
<td></td>
</tr>
<tr>
<td>olefins/cycloolefins</td>
<td>C5 to C18: linear olefins, branched olefins, and cycloolefins</td>
<td>0.3597</td>
</tr>
</tbody>
</table>

are lumped into a single lump, e.g., nC16+. This is because there is hardly any difference in the pure-component CN of paraffins beyond nC16. Similar observations hold for the other lumps in the other molecular classes as well.

To apply eq 6, we must also specify the various adjustable parameters $\beta_i$. In principle, each molecular lump in eq 6 should be associated with a separate fitting parameter $\beta_i$ because it parametrizes its blend contribution to the fuel cetane. However, here we make the simplifying assumption that $\beta_i$ is the same for all of the molecules in a given molecular class. Therefore, all $n$-paraffins, regardless of their carbon chain length, will have the same $\beta_i$. Likewise, all aromatics, regardless of their carbon number, will have the same $\beta_i$, and so on. This simplification reduces the number of adjustable parameters to 9, one $\beta$ for each molecular class.

4. Experimental Program

A database of 203 diesel fuels was collected from 8 different refineries in this study. The fuels include 45 different diesel-range refinery process streams and 158 commercial blends of these process streams. They were so chosen to reflect a wide range of diesel properties with CNs varying from 20 to 60. The CNs on all of these fuels were measured using the IQT method. Detailed compositional analysis was also obtained on these fuels using a combination of supercritical fluid chromatography (SFC), GC, and MS methods. The SFC was used to measure the total paraffins, naphthenes, and aromatics by ring class, GC was used to measure the boiling point distribution, and MS was used to get the molecular weight distribution. The fuels collected spanned a wide range of molecular compositions from highly paraffinic to highly naphthenic to highly aromatic diesel fuels.

5. Results

5.1. Parameter Estimation: Constrained Least-Squares Formulation. A constrained least-squares minimization problem was solved using the Levenberg–Marquadt algorithm in order to regress the parameters of the model. The adjustable parameters $\beta_i$ were allowed to vary between lower and upper bounds of $-10$ and $+10$, respectively. The experimental dataset was partitioned into a training set (90% of the samples) and a testing set (10% of the samples), and many such partitions were considered to ensure robustness of the parameter estimates. Different initial guesses were assumed and the parameters reoptimized to ensure that the optimization problem was not trapped in an inferior local solution. Sensitivity analysis in the form of local perturbation of the parameters was also performed to see whether the parameters returned to the same optimal or not.

Although the final parameter values ($\beta$) for all of the individual molecular lumps used in this model cannot be disclosed in the publication for proprietary reasons, the average values of these parameters, grouped by P, I, O, N, and A classes, are reported in Table 3. The results based on these average parameter values are satisfactory, though we may point out that the use of individual $\beta$ values for the different molecular lumps enhances model predictions. The results presented below are based on the individual $\beta$ parameters.

5.2. Prediction Statistics. On the basis of the regression of the parameters $\beta$, the prediction results of the model are shown in Figure 3 for all of the 203 diesel fuels. The plot not only highlights the performance separately across the different diesel-range refinery process streams and commercial blends as indicated by the different symbols. Across all of these fuels, the overall standard error (SE) defined as $\sqrt{\sum (\text{meas} - \text{pred})^2/n}$ for the whole dataset is 1.25 numbers. The individual statistics for the training and the testing datasets are 1.15 and 1.3 numbers, respectively. Also shown in this plot are the ±1.5 number error lines which is the typical repeatability error of the cetane measurements in the IQT. The results shown in this figure clearly demonstrate quantitative agreement of the model predictions across this multitude of diesel fuels for both the process streams and the commercial blends. It is unlikely that the SE could be reduced further because the model predictions are well within the measurement error of the property itself.

An important requirement of any regression model is that its predictions are unbiased with respect to its inputs. To ensure this, the frequency distribution of the model errors is plotted in Figure 4. The abscissa in the figure represents the error level in the model predictions, while the ordinate represents the number of predictions.
of fuels predicted at these various error levels. Approximately 74% of the samples are predicted within ±1.5 CN, 95% are predicted within ±3.0 CN, and a very few samples (∼5%) show an error of more than 3.0 numbers. The frequency distribution of Figure 4 closely resembles a normal distribution, suggesting that the model is independent of any inherent bias. Further, the plot of the prediction error against each of the molecular lumps revealed no directional trend of the composition with the model error.

A comparison of the model predictions with the ASTM D4737 equation (eq 1) is shown in Figure 5. ASTM D4737, as indicated earlier, uses API gravity and the distillation curve to predict the CN. The SE for the present model is 1.25 numbers, compared to the SE of 3.32 numbers for ASTM D4737. The standard deviation (σ) values for the two models are 1.28 and 2.61 numbers, respectively. The superior prediction quality of the present model against the ASTM correlation is not surprising because the present model builds upon the direct and explicit knowledge of the detailed molecular composition of the diesel fuel. Such detailed molecular accounting is critical because it provides more robustness and reliability in the predictions, which is important because such models would be most frequently used in the extrapolative mode in the petroleum refinery.

6. Conclusions

A general composition-based predictive model for CN is presented here that can be universally applied across a wide variety of diesel fuels including process streams and their blends. Each diesel fuel was compositionally represented by 9 different hydrocarbon classes containing a total of 129 different molecular lumps measured by a combination of SFC, GC, and MS methods. The model predicts the CN with a standard error of ±1.25 numbers across the multitude of diesel fuels considered in this work. The model explicitly acknowledges the detailed molecular composition of the diesel fuel and incorporates the pure-component molecular trends of CN. As a result, the model predictions are expected to be robust and reliable during the use of the model in extrapolation.

Literature Cited

(3) Ladommatos, N.; Goucher, J. Equations for predicting the cetane number of diesel fuels from their physical properties. Fuel 1995, 74 (7), 1083.
(8) Cookson, D. J.; Illipoulos, P.; Smith, B. E. Composition—Property relations for jet and diesel fuels of variable boiling range. Fuel 1995, 74 (1), 70.
(10) Yang, H.; Ring, Z.; Brikzer, Y.; McLean, N.; Friesen, W.; Fairbridge, C. Neural network prediction of cetane number and density of diesel fuel from its chemical composition determined by LC and GC-MS. Fuel 2002, 81, 65.

Received for review July 11, 2005 
Revised manuscript received October 10, 2005
Accepted October 14, 2005
IE0508132

Figure 5. Comparison of the model predictions of the present model against the ASTM D4737 equation. Also shown in the parity plot are the ±1.5 CN error lines.