LA-UR-16-25529
Approved for public release; distribution is unlimited.

Title: A Group Contribution Method for Estimating Cetane and Octane Numbers

Author(s): Kubic, William Louis

Intended for: Report

Issued: 2016-07-28 (rev.1)

LA-UR-16-25529
Approved for public release;
distribution is unlimited.

A Group Contribution Method for Estimating Cetane and Octane Numbers

William L. Kubic, Jr.

July 2016

Acknowledgements

This work was performed as part of a Laboratory Directed Research and Development Exploratory Research (LDRD-ER) engineering project for the Development of a Continuous Flow Reactor for the Conversion of Biomass Hydrolystates to Fuels and Feedstocks. The author would like to acknowledge the Principle Investigator, Andrew Sutton, for including him in the project and providing both technical and financial support for the endeavor. The author would also like to acknowledge Troy Semelsberger, Rhodri Jenkins, and Cameron Moore for their helpful discussions and review comments. Final, the author would like to thank Adam Davis, Travis Moulton, and Drew Kornreich of the authors own group, AET-2, for their help discussions and reviews of the methodology.

Table of Contents

Abstract **5**

1.0 Introduction **5**

2.0 Data Set **8**

2.1. Description of Data Set 8
2.2. Variability and Uncertainty in the Data 10
2.3. Comparison with Other Data Sets 12

3.0 Constructing the Artificial Neural Network **12**

3.1. Background Information Concerning Artificial Neural Networks 12
3.2. Structure of the Artificial Neural Network for Cetane and Octane Numbers 14
3.3. The Objective Function 15
3.4. Constructing the Artificial Neural Network 15
3.5. Discussion of Quality of Fit and Errors 19
3.6. Comparisons to Other Methods 22

4.0 Cross Validation **24**

4.1. 10-Fold Cross Validation 24
4.2. Extrapolation Test 26

5.0 Discussion **27**

6.0 Conclusions **28**

References **29**

Appendix A - Cetane Number versus Derived Cetane Number **33**

Reference 30

Appendix B - Octane and Cetane Number Data Tabulation **35**

Appendix C - Parameters for the Artificial Neural Network **55**

C.1 INTRODUCTION 55
C.2 PARAMETERS FOR HIDDEN NODE INPUTS 55
C.3 PARAMETERS FOR OUTPUT NODE INPUTS 59
C.4. SCALING FACTORS FOR OUTPUT PARAMETERS 60

Appendix D - Parameters for a Conventional Group Contribution Methods **61**

D.1. INTRODUCTION 61
D.2. POLYNOMIAL FUNCTION 61
D.3. BOUNDED FUNCTION 63
References 65

List of Figures

1. CN data for n-alkanes and primary alcohols of n-alkanes as a function of carbon number 7
2. CFD for deviations of RON measurements from the average 11
3. CDF for deviations of MON measurements from the average 11
4. CDF for combined deviations of CN measurements from the average 11
5. DF for deviations of CN measurements from the average for hydrocarbons and oxygenated hydrocarbons 11
6. A simple ANN 13
7. An intermediate or output node in an ANN 14
8. Initial ANN structure 16
9. A comparison of data to CN predictions for the initial ANN 16
10. The modified ANN for the initial working data set 17
11. A comparison of data to CN predictions for the modification of the initial ANN 17
12. Final ANN structure for estimating CN, RON, and MON functional group contributions 18
13. MON as a function RON 19
14. CN as a function RON 19
15. ANN results for CN of n-alkanes and primary alcohols 20
16. ANN results for CN and RON of terminal alkenes 20
17. Predicted versus measured values of CN for hydrocarbons 21
18. Predicted versus measured values of CN for oxygenated hydrocarbons 21
19. Predicted versus measured values of RON 22
20. Predicted versus measured values of MON 22
21. CDFs for errors in CN predictions for hydrocarbons and oxygenated hydrocarbons 22
22. CDFs for errors in RON predictions 23
23. CDFs for errors in MON predictions 23
24. CDFs for errors in CN predictions obtained from the regression analysis and the 10-fold cross validation 25
25. CDFs for errors in RON predictions obtained from the regression analysis and the 10-fold cross validation 25
26. CDFs for errors in MON predictions obtained from the regression analysis and the 10-fold cross validation 26
A-1. DCN as a function of CN for hydrocarbons and oxygenated hydrocarbons 33
C-1. Structure of the artificial neural network representing cetane and octane numbers 55
D-1. Simple ANN used to generate the bound function for a group contribution correlation represented by Eq. (D-3) 63

List of Tables

1. Data set references 8
2. Summary of compounds included in the data set 9
3. Summary of amount of data included in the final data set 10
4. A comparison of data sets used to develop structure-property relationships for cetane number 12
5. A comparison of data sets used to develop structure-property relationships for octane numbers 13
6. List of functional groups in the ANN. 20
7. Bias, standard deviation, and correlation coefficients for the ANN 21
8. Comparison of correlation coefficients for different group contribution methods 23
9. Comparison of correlation coefficients for the ANN group contribution method of this study to published group contribution methods 24
10. Comparison of standard deviation in data, regression analysis errors, and 10-fold validation errors 26
11. Results of F-test for extrapolation validation sets 27
12. Estimation errors for the ANN-based group contribution method 28
B-1. Cetane and Octane Number Data for Paraffins 35
B-2. Cetane and Octane Number Data for Olefins 38
B-3. Cetane and Octane Number Data for Alkynes 41
B-4. Cetane and Octane Number Data for Naphthenes 41
B-5. Cetane and Octane Number Data for Aromatics 43
B-6. Cetane and Octane Number Data for Oxygenates Other Than Esters 46
B-7. Cetane and Octane Number Data for Esters 49
C-1. Group contribution parameters for hidden nodes 56
C-2. Parameters for output node inputs 59
C-3. Upper and lower bounds for output variables 60
D-1. Group factors for conventional a group contribution correlation with a polynomial function 61
D-2. Constants for Eq. (D-2) 62
D-3. Standard deviation of errors for the polynomial group contribution correlation 62
D-4. Group factors for conventional a group contribution correlation with a bounded function 64
D-5. Constants for Eq. (D-4) 65
D-6. Standard deviation of errors for the bound function group contribution correlation 65

A Group Contribution Method for Estimating Cetane and Octane Numbers

William L. Kubic, Jr.
Process Modeling and Analysis Group
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

July 2016

ABSTRACT

Much of the research on advanced biofuels is devoted to the study of novel chemical pathways for converting nonfood biomass into liquid fuels that can be blended with existing transportation fuels. Many compounds under consideration are not found in the existing fuel supplies. Often, the physical properties needed to assess the viability of a potential biofuel are not available. The only reliable information available may be the molecular structure. Group contribution methods for estimating physical properties from molecular structure have been used for more than 60 years. The most common application is estimation of thermodynamic properties. More recently, group contribution methods have been developed for estimating rate dependent properties including cetane and octane numbers. Often, published group contribution methods are limited in terms of types of function groups and range of applicability. In this study, a new, broadly-applicable group contribution method based on an artificial neural network was developed to estimate cetane number research octane number, and motor octane numbers of hydrocarbons and oxygenated hydrocarbons. The new method is more accurate over a greater range molecular weights and structural complexity than existing group contribution methods for estimating cetane and octane numbers.

1.0 INTRODUCTION

First-generation biofuels, which include ethanol and biodiesel, are derived from agricultural products such as starch, sugar, animal fat, and vegetable oil.[1] These biofuels have been criticized because of the stress that their production places on food commodities. These problems have prompted interest in fuels produced from lignocellulosic materials, algae, crop residue, or other non-food feedstocks. Fuels produced from these alternative feedstocks are called second-generation biofuels or advanced biofuels.[1]

The feedstocks for advanced biofuels are much harder to convert into liquids than starches, sugars, fats, and oils. Overcoming this difficulty has prompted interest in novel chemical pathways, which yield the products that are not found in current fuels supplies. Because they currently have no practical or commercial value, key physical properties may not been measured. The only thing that may be known with certainty is the chemical structure.

Quantitative structure-property relationships have been studied extensively.[2] Since Lydersen published his report on estimating the critical properties of organic compounds,[3] group contribution methods have been, perhaps, the most widely used approach for correlating physical properties with molecular structure. Group contribution correlations have been developed for thermodynamic properties including critical properties,[3,4] freezing and boiling points,[4] liquid heat capacities,[5] gas heat capacities,[6,7] enthalpies of formation for gases,[6,7] entropies of gases,[6,7] and activity coefficients.[8] Group contribution methods are not

limited to thermodynamic properties. Orrik and Erbar published a group contribution method for estimating liquid viscosity.[9]

Group contribution methods also have been used to correlate combustion properties of fuels including auto ignition temperature,[10] delayed ignition time,[11] octane number (ON),[12,13] and derived cetane number (DCN).[11,14,15,16] Structure-property relationships for ON and Cetane Number (CN) are much more complex than thermodynamic properties of liquids and gases because they are related to combustion kinetics. Smolenskii[17] proposed a measure of the complexity of a property based on the variability of the property for subsets of isomers and total variability. Smolenskii et al.[18] report that models of properties with low complexity have reasonable accuracy while models of properties with high complexity are "senseless." Smolenskii et al. found that cetane number is a property with a moderate to high degree of complexity. Therefore, standard methods of correlating properties with structures are inadequate.

Group contribution methods for estimating ON and CN have been moderately successful. DeFries et al.[14] developed a correlation for estimating CN using data for 99 paraffinic and aromatic hydrocarbons with molecular sizes ranging from C_5 to C_{25}. Their correlation has the following functional form.

$$CN = \sum_{i \in groups} f_i \cdot cn_i \ , \tag{1}$$

where CN is the CN, f_i is the fraction of group i that comprises the molecule, and cn_i is the group contribution parameter for group i. This formulation of a group contribution method is well suited for estimating CN from carbon-13 nuclear magnetic resonance spectroscopy data. Although their study included a discussion of data for olefinic and naphthenic compounds, DeFries et al. did not define groups for ringed compounds or carbon-carbon double bonds.

Albahri[13] developed group contribution methods for research ON (RON) and motor ON (MON) using a data set consisting of 188 pure hydrocarbons, which included paraffins, naphthenes, olefins, and aromatics. Molecular sizes ranged from C_4 to C_{11}. Albahri used a functional form similar to other, well-known group contribution correlations.[4,5,6,7]

$$ON = H\left(\sum_{i \in groups} n_i \cdot x_i ; P \right) \ , \tag{2}$$

where ON is either RON or MON, n_i is the number of group i in the molecules, x_i is the group contribution parameter for i, and H is function of the summation of the functional groups and P, which is a vector of model parameters.

Dahmen and Marquardt[11] took a different approach when developing a group contribution method for estimating CN. Instead of correlating CN with structure, they developed a group contribution method for estimating ignition delay; and then they used to formulas in ASTM D6890[19] to determined DCN. Dahmen and Marquard's data set consisted of 154 compounds that included hydrocarbons, alcohols, esters, and other oxygenated compounds. The correlation includes more complex nonlinearities and interactions than DeFries et al. and Albahri's methods

$$\tau = A \cdot p_{sat}^{m} \cdot \exp\left(\sum_{i \in groups} g_i\left(n_i\right) \cdot x_i \right) \ , \tag{3}$$

where A is an empirical constant, p_{sat} is the vapor pressure at 298 K, m is an empirical constant, n_i is the number of group i in the molecules, g_i is a function characteristic of group i, x_i is the group contribution parameter for i. Vapor pressure is determined from Joback's group contribution correlations for critical properties and boiling point.[4]

Published group contribution methods and other structure-property correlations for CN and ONs are reported to be very accurate. Correlation coefficients (r^2) for hydrocarbons are reported to be 0.97 to greater than 0.99. Correlation coefficients for oxygenated compounds are lower, but significant. One study[20] reported correlation coefficients for alcohols and esters to be 0.90. However, when the published correlations were tested against a more extensive data set, the resulting errors were much greater than reported in the original papers. The published group contribution methods for CN and ON gave good results when applied to fuels with molecular sizes and complexities similar to those in the data sets used to develop the model, but the predictions were poor when these models were extrapolated.

One of the problems with existing group contribution methods for estimating CN and ONs is that they do not account for complex, nonlinear interactions between functional groups. Figure 1 is a plot of CNs for n-alkanes and primary alcohols of n-alkanes. If existing group contribution methods are correct, then the CN data for primary alcohols should form a curve having approximately the same shape as the curve for n-alkanes. However, Fig. (1) shows that data for primary alcohols follow a qualitatively different curve than n-alkanes indicating that a level of complexity that cannot be captured by standard group contribution methods.

The goal of this study was to develop a reliable correlation for estimating cetane and octane numbers of hydrocarbons and oxygenated hydrocarbons from chemical structure. While such a correlation will not replace the need for actual measurements, it would be a very useful screening tool for evaluating candidate biofuels early in the research and development process. When CN and ON estimates are combined with blending models[21,22] the results could be used to evaluate whether a proposed biofuel can be blended with current fuels. CN and ON are indicative of the quality of the fuel; and hence, its economic value.

The specific objective of this study is to develop a group contribution method for estimating CN, RON, and MON that

- is applicable to a wide range of hydrocarbons and oxygenated hydrocarbons typical of advanced biofuels,
- accounts for nonlinear interactions among the functional groups that make up the fuel, and
- can be extrapolated with a reasonable degree of confidence.

To accomplish these objectives, a group contribution method was developed using an artificial neural network (ANN). ANN-based group contribution methods have been developed for estimating densities,[23] auto-ignition temperature,[10] and octane number.[13] These studies indicate that an ANN can account for nonlinear interactions among function groups.

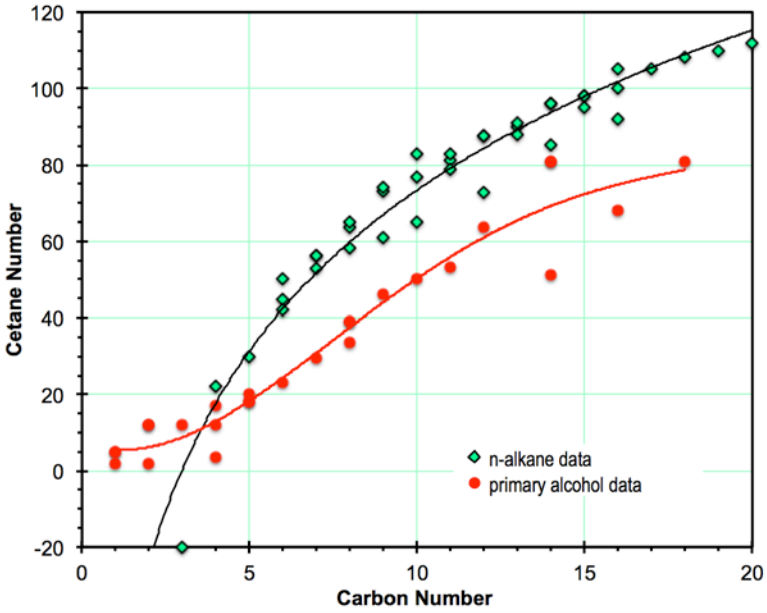

Fig. 1. CN data for n-alkanes and primary alcohols of n-alkanes as a function of carbon number.

2.0 DATA SET

2.1. Description of Data Set

Developing a reliable group contribution method for estimating CN and ON of hydrocarbons and oxygenated hydrocarbons requires comprehensive data set that includes compounds with a broad range of molecular weights and molecular complexity. The data set used in this study was assembled from a variety of sources including published data compilations, data set used to develop other structure-property relationships, and primary sources. CN and DCN we found to be nearly equivalent, so the CN data included both CN and DCN measurements. Appendix A contains additional discussion of CN versus DCN. Blend CN and ONs were not included in the data set. Table 1 is a summary of the references used in this study and the types of data obtained from those references. Table 2 lists the number of each class of compound in the data set with the corresponding range of carbon numbers.

The data obtained from the references listed in Table 1 were not always consistent. Reported values of CN differed by more than 50. The spread in RON and MON measurements can be as a great as 40. The variability could be the result of differences in test procedures, impurities in the test material, and experimental errors. The reference listed in Table 1 includes 85 compounds with duplicate CN measurements, 57 compounds with duplicate RON measurements, and 36 compounds with duplicate MON measurements. The initial compilation included all measured values. No judgment was made concerning the accuracy of the data.

Table 1. Data set references.

Reference	Property			Class of Class									
	CN	RON	MON	Paraffins	Olefins	Alkynes	Naphthenes	Aromatics	Alcohols	Ethers and Furans	Aldehydes and Ketones	Carboxylic Acids	Esters and Triglycerides
Yanowitz et al.[24]	X			X	X	X	X	X	X	X	X	X	X
Creton at al.[16]	X			X	X		X	X					
Salanda et al.[25]	X			X	X		X	X	X	X	X	X	X
Smolenskii et al.[18]	X			X			X						
Dahmen and Marquardt[11]	X			X	X		X	X	X	X	X		X
Kamei and Singal[26]	X									X			
Nagdeote and Deshmukh[27]	X								X	X			
Naegeli et al.[28]	X									X			X
Freedman and Bagby[29]	X								X				X
Knothe et al.[30]	X												X
API [31]		X	X	X	X	X	X	X					
da Silva and Bozzelli[32]		X						X					
Masum et al.[33]		X							X				
Hamadi[34]		X	X					X	X		X		
Egloff and Van Arsdell[35]		X	X	X	X	X	X	X	X		X		
Yanowitz et al.[36]		X	X						X	X			X

Table 2. Summary of compounds included in the data set.

Compound Type	Cetane Number		Octane Number	
	Number of Compounds	Carbon Number Range	Number of Compounds	Carbon Number Range
Paraffins	77	C_3 – C_{24}	45	C_2 – C_{12}
Olefins and Alkynes	39	C_5 – C_{24}	68	C_2 – C_{11}
Naphthenes	47	C_5 – C_{26}	43	C_5 – C_{11}
Aromatics	65	C_6 – C_{26}	37	C_6 – C_{10}
Alcohols	28	C_1 – C_{18}	13	C_1 – C_6
Esters and Triglycerides	130	C_5 – C_{57}	3	C_6 – C_9
Other Oxygenates	63	C_2 – C_{18}	9	C_2 – C_6
Total	449	C_1 – C_{57}	218	C_1 – C_{12}

Picking "best" values of CN, RON, and MON measurements for each compound can bias the correlation, and it results in under estimating the uncertainty. Averaging multiple measurements to obtain a single value also results in under estimation ofuncertainty. Using all measurements to develop the correlation also can result in bias. For a few compounds, a large number of a large number of duplicate measurements have been reported. For example, the references in Table 1 include 11 duplicate CN measurements for *n*-decane. Including all of these measurements in the data set would give more weight to *n*-decane than other compounds, which could bias the correlation. To avoid giving excess weight to a few compounds that have been studied extensively while retaining information on variability and uncertainty, the number of measurement included in the final data set for a given compound was limited to three – the "best estimate," the minimum, and the maximum.

In operations research, the probability distribution for activity durations is commonly estimated from three points – a "pessimistic," an "optimistic," and a "most likely" value.[37] Restricting the number of measurements for a given compound to three resembles this practice. Based on the data summarized in Ref. 24, a reasonable three-point estimate of the distribution of CN data can be obtained using the Swanson-Megil approximation.[37] This ability to approximate the observed distribution of measurements from the minimum, maximum, and best estimate values indicates that limiting the number of measurements for a given compound resembles data compression rather than data censoring.

A limited amount of data censoring was used when compiling the data set. Obvious outliers were eliminated. For example, dimethyl ether is reported to be a poor substitute for gasoline. Therefore, a reported RON of 170 for dimethyl ether is clearly incorrect. Values reported as inequalities also were eliminated. In some cases data compilations include compounds, but give ambiguous structural information. For example, specifying a compound as dimethylcyclohexane fails to indicate the relative positions of the methyl groups on the cyclohexane ring. Compounds specified with an ambiguous structure were eliminated from the data set if there was insufficient information to determine the group composition of that compounds. Table 3 lists the number of measurements included in the final data set.

Appendix B contains a complete tabulation of the data used in this study.

Table 3. Summary of amount of data included in the final data set.

Compound Class	Number of Measurements		
	Cetane Number	Research Octane Number	Motor Octane Number
Paraffins	115	66	63
Olefins and Alkynes	45	75	67
Naphthenes	63	51	36
Aromatics	83	42	40
Alcohols	40	26	22
Esters and Triglycerides	186	3	3
Other Oxygenates	73	8	10
Total	604	271	241

2.2. Variability and Uncertainty in the Data

As noted in the previous section, CN and ON data from different sources rarely agree. Huber and Hauber[38] compared CN measurements obtained from different laboratory in a "round robin" test and found substantial differences in measured CNs. Compounds for which multiple measurements were available were used to estimate variability or uncertainty in the measured values of CN, RON, and MON. For each compound with multiple measurements, the average value of the property was valuated, and deviations for the average were determined. For a each property, the deviations were aggregated into a single set, and order statistics were used to estimate the cumulative distribution functions (CDFs) for the combined set of deviations. The value of the CDF for each point was estimated using the following equation.

$$F_i = \frac{i - \frac{1}{2}}{N} \quad , \tag{4}$$

where i is the index of the data point in the ordered set of deviations, F_i is the CDF for the i^{th} point, and N is the total number of points in the ordered data set. The index i runs from 1 to N.

Figures 2 and 3 are the CDF from deviations in the RON and MON measurements. The ordinate in Figs. 2 and 3 is a normal probability scale. The CDF for a normal probability distribution is a straight line on these plots with a positive slope equal to the reciprocal of standard deviation. The deviation data for both RON and MON lie on a straight line indicating that the distributions of deviations are normal. Linear fits of the data for both RON and MON gave correlation coefficients (r^2) of 0.99. Based on the slope of these lines, the standard deviation for RON is 3.4 and the standard deviation for MON is 3.3. The fact that the combined set of deviations can be represented by the same normal probability distribution suggests that the combined sets of deviations are characteristic of the uncertainty in the experimental data for RON and MON.

Figure 4 is the CDF of CB deviations plotted on a normal probability scale. The deviations do not lie on a straight line indicating that the distribution of deviations is not normal. The tails of this distribution are dominated by high molecular-weight esters suggesting that the variability in CN measurements for oxygenated hydrocarbons may differ from the variability for hydrocarbons. To test this hypothesis, the set

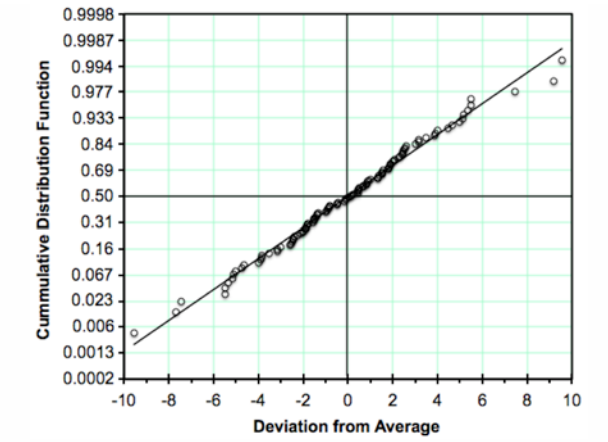

Fig. 2. CFD for deviations of RON measurements from the average.

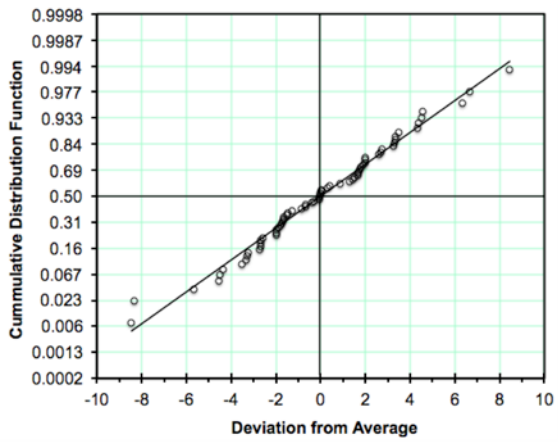

Fig. 3. CDF for deviations of MON measurements from the average.

of CN deviations were divided into two populations – hydrocarbons and oxygenated hydrocarbons. CDFs for these two populations are plotted in Fig. 5. Both CDFs are approximately linear. The correlation coefficient for hydrocarbons is 0.96, and the correlation coefficient for oxygenated hydrocarbons is 0.97. Variability in CN for hydrocarbons and oxygenated hydrocarbons can be approximated by two different normal probability distributions. The standard deviation for hydrocarbons is 4.1 and the standard deviation for oxygenated hydrocarbons is 7.5. As was the case for ONs, variability in CN is assumed to be characteristic of uncertainty. For reasons that are not understood, the CN data for hydrocarbons are more accurate than CN data for oxygenated hydrocarbons.

In addition to experimental uncertainty, a large data set will most likely contain erroneous values as a result of gross experimental errors or transcription errors. Stating that a data set contains erroneous values in no way disparages the researchers who performed the original measurements or compiled the data. Rather, it is recognition that the probability of human errors is not zero. Williams[39] gives unreliability estimates for various generic tasks. A complex task requiring a high level of comprehension and skill has

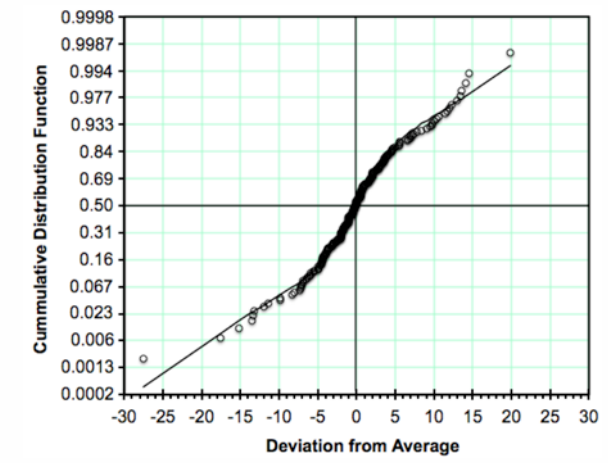

Fig. 4. CDF for combined deviations of CN measurements from the average.

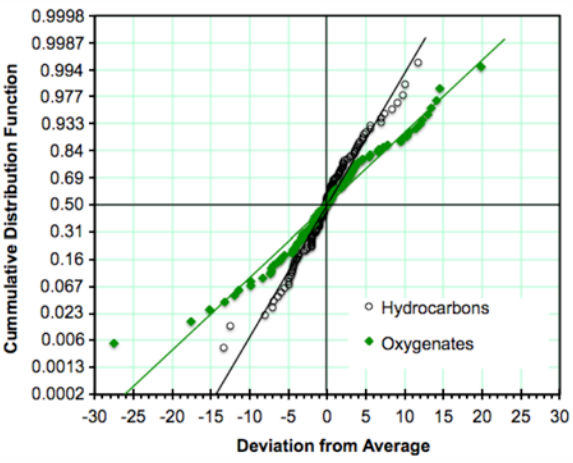

Fig. 5. DF for deviations of CN measurements from the average for hydrocarbons and oxygenated hydrocarbons.

a nominal human unreliability of 0.16 with a range of 0.12 to 0.28. The ASTM tests for determining ON and CN could be considered such a complex task. A routine, highly practiced task involving a relatively low level of skill has a nominal human unreliability of 0.02 with a range of 0.007 to 0.045. Transcribing experimental results would fit this category. Based on Williams unreliability estimates, it is reasonable to assume that between 2 and 16 values out of every 100 values in a compilation of experimental results are erroneous.

2.3. Comparison with Other Data Sets

Table 4 contains a comparison of the CN data set used in this study with the data sets used in three other studies of structure-property relationships for CN. Overall, the data set used in this study contains CN data more compounds than other published studies. Only this study and Dahmen and Marquardt's[11] study included oxygenated compounds other than alcohols, esters, and triglycerides. Although the data set used study includes more compounds with CN data than other studies, it contains fewer olefins than data set of Saldana et al.[25] and fewer other oxygenates than the data set of Dahmen and Marquardt.[11] These small differences are primarily the result of excluding data for fuels with ambiguous molecular structures.
`

Table 5 contains a comparison of the ON data sets used with the data sets used in three other studies of structure-property relationships for ON. As was the case for CN, the data set used in this study contains ON data for more compounds than other published studies. It is the only data set used to develop property-structure relationships for ON that includes oxygenated compounds. The data set used by Albahri contains ON data for seven more olefins than this study. The difference is primarily the result of eliminating data for which the structure of the compound is not specified unambiguously.

The data set used in this study are more comprehensive than the data set used to develop other structure-property relationships for CN and ON.

3.0 CONSTRUCTING THE ARTIFICIAL NEURAL NETWORK

3.1. Background Information Concerning Artificial Neural Networks

An ANN is an acyclic directed graph consisting of multiple layers of interconnected nodes. Figure 6 is an illustration of a simple ANN. The lowest layer of nodes (black) are the input nodes, and they correspond to the values of the independent variables. The middle layers (blue) are the hidden nodes. An ANN may

Table 4. A comparison of data sets used to develop structure-property relationships for cetane number.

Compound Type	Number of Compounds			
	Lapidus et al.[40]	Saldana et al.[20]	Dahmen & Marquardt[11]	This Study
Paraffins	32	65	24	77
Olefins and Alkynes	0	46	18	39
Naphthenes	31	31	10	47
Aromatics	0	60	8	65
Alcohols	0	22	10	28
Esters and Triglycerides	0	61	18	130
Other Oxygenates	0	0	66	63
Total	**63**	**284**	**154**	**449**

Table 5. A comparison of data sets used to develop structure-property relationships for octane numbers.

Compound Type	Number of Compounds			
	DeFries et al.[14]	Albahri[13]	Lapidus et al.[40]	This Study
Paraffins	35	44	32	45
Olefins and Alkynes	0	75	0	68
Naphthenes	0	32	31	43
Aromatics	64	37	0	37
Alcohols	0	0	0	13
Esters and Triglycerides	0	0	0	3
Other Oxygenates	0	0	0	9
Total	**99**	**188**	**63**	**218**

have one or more layers of hidden nodes. The number of hidden nodes does not necessarily equal the number of input nodes. The simple example shown in Fig. 6 has only a single layer of hidden nodes. The top layer consists of the output nodes (red). The simple ANN shown in Fig. 1 contains only a single output node, but ANN often includes multiple output nodes. The output nodes correspond to the dependent variables. The edges in an ANN connect adjacent layers of nodes in the graph. Edges connect nodes in nonadjacent layers are not usually included in an ANN. The edges are directed from the lower layer of nodes to the next higher layer. Figure 6 shows edges connecting a node in a lower level of the ANN to all nodes in the next higher layer. This architecture is not a necessary feature of an ANN. A node need not be connected to all nodes in the next layer.

The edges in an ANN correspond to variables in the model – independent variables or input parameters, dependent variables or output parameters, and intermediate variables. Variables are typically defined on a finite interval, which is most often [0,1] or [-1,1]. Input and output parameters are scaled to fit on the interval.

Nodes in an ANN correspond to functions. For the input nodes, the function is a simple identity function

$$O_k = X_k \ , \tag{5}$$

where O_k is the output variable for input node k and X_k is the input parameter for input node k. Intermediate and output nodes are a scalar-valued function of a vector, as illustrated in Fig. 7. O_k is the output variable for node k and $\boldsymbol{I}$ is the vector of input variables. The input variables are combined linearly to obtain an intermediate scalar variable .

$$z_k = w_{o,k} + \sum_{j=1}^{n} w_{j,k} \cdot I_j \tag{6}$$

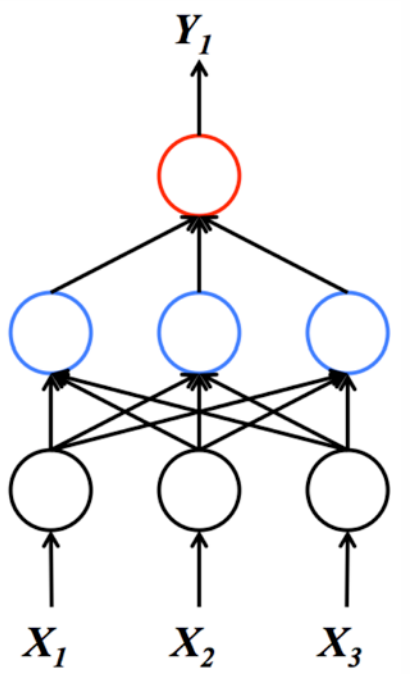

Fig. 6. A simple ANN.

where z_k is the intermediate variable and $w_{o,k}$ and $w_{j,k}$ are empirical parameters or weighting factors determined from a regression analysis. O_k is typically a simple, sigmoidal function of z_k. The logistic function is one of the most commonly used function to represent nodes in an ANN.

$$O_k = \frac{1}{1+e^{-z_k}} \quad . \tag{7}$$

Other functions, such as $\tanh^{-1}$, could be used in an ANN.

The parameters for an ANN are determined by minimizing the error in the output parameters or dependent variables. The back propagation algorithm is often used to estimate the parameters. The back propagation algorithm is a gradient search that takes advantage of the structural properties and functional forms that are used for an ANN. Because it is an application specific algorithm, it is a very efficient method of determining ANN parameters.

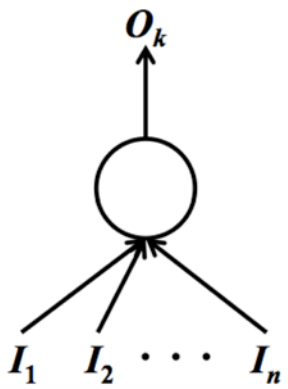

Fig. 7. An intermediate or output node in an ANN.

Much good work has been done developing and applying ANN technology. Unfortunately, much of the published literature is mired in the jargon of artificial intelligence. Some authors state that an ANN simulates the structure of the human brain. Others call ANN "machine learning." When the hype and jargon is put aside, ANN technology can be seen as a useful method of nonlinear, multivariate regression analysis. In this study, ANNs were viewed as nonlinear regression analysis, so this report favors the language of statistics and regression analysis over the language of artificial intelligence.

3.2. Structure of the Artificial Neural Network for Cetane and Octane Numbers

The ANN used to correlate CN and ON data consisted of three layers of nodes – the input nodes, which represent functional groups that make up the molecule; the hidden nodes; and the output nodes, which represent CN, RON, and MON. The ANN has one input node representing each type of functional group. The value of the input node is the number of the corresponding functional groups in the molecule. Functional groups can be defined with differing levels of specificity. For example, all aliphatic alcohols could be represented by a single –OH group or they could be represented by three groups corresponding to primary, secondary, or tertiary alcohols. The number and type of functional groups used in the ANN was not predetermined. The need to add functional groups and to make finer distinctions between functional groups was determined as the network evolved. Likewise, the number of hidden nodes and connections between the hidden nodes and the output nodes was not predetermined.

The number of a particular functional group in a molecule is defined on the set of nonnegative integers. To scale the number of functional groups in a molecule to fit on a scale of [0,1] requires the division of the number of functional groups by a predefined maximum. Such a simple arithmetic operation would transform Eq. (6) into the following equation.

$$z_k = w_{o,k} + \sum_{j=1}^{n} w_{j,k} \cdot \frac{n_j}{n_{max,j}} \quad , \tag{8}$$

where n_j is the number of functional group j in the molecule and $n_{max,j}$ is the maximum value of n_j. The ratio $n_j / n_{max,j}$ is I_j in Eq. (6). Because $n_{max,j}$ is a constant, the quotient $w_{j,k} / n_{max,j}$ is also a constant. Equation (8), therefore, can be transformed into the following equation.

$$z_k = G_{o,k} + \sum_{j=1}^{n} G_{j,k} \cdot n_j \quad , \tag{9}$$

where $G_{o,k} = w_{o,k}$ and $G_{j,k} = w_{j,k} / n_{max,j}$. From a practical and mathematical point of view, Eq. (9) is equivalent to Eq. (8). Scaling the input parameters for a group contribution method introduces an arbitrary and unnecessary parameter into the model. Therefore, the input parameters, which are the number of each type of functional group in the molecule, were not scaled in this study. Equation (9) was used with the parameters $G_{j,k}$ being the group contribution factors for hidden node k.

3.3. The Objective Function

The parameters for the ANN were determined using an unweighted least squares regression analysis. The objective function for an unweighted least squares regression analysis is simply the sum of squared errors.

$$\begin{aligned} SSE = & \sum_{i \in CN\ Data} \left(CN_{calc,i} - CN_{data,i} \right)^2 + \sum_{i \in RON\ Data} \left(RON_{calc,i} - RON_{data,i} \right)^2 \\ & + \sum_{i \in MON\ Data} \left(MON_{calc,i} - MON_{data,i} \right)^2 \end{aligned} \tag{10}$$

where SSE is the sum of squared errors; $CN_{calc,i}$, $RON_{calc,i}$, and $MON_{calc,i}$ are the values of CN, RON, and MON calculated from the ANN; and $CN_{data,i}$, $RON_{data,i}$, and $MON_{data,i}$ are the measured values of CN, RON, and MON. Use of the sum of squares as the objective function can be problematic when applied to the ANN architecture used in this study. The parameters in Eqs. (6) and (9) are unbounded, which could lead to numerical problems when combined with the logistics function. The logistic function (Eq. (7)) used for hidden and output nodes is approximately zero when $z_k < -7$ and approximately unity when $z_k > 7$. This type of behavior can result in trivial solutions involving parameters with large absolute values.

To avoid trivial solutions, a modified least squares objective function was used.

$$\begin{aligned} MSSE = & \sum_{i \in CN\ Data} \left(CN_{calc,i} - CN_{data,i} \right)^2 + \sum_{i \in RON\ Data} \left(RON_{calc,i} - RON_{data,i} \right)^2 \\ & + \sum_{i \in MON\ Data} \left(MON_{calc,i} - MON_{data,i} \right)^2 + \alpha \cdot \left(\sum_j \sum_k G_{j,k}^2 + \sum_j \sum_k w_{j,k}^2 \right) \end{aligned} \tag{11}$$

where $G_{j,k}$ are the group contribution parameters applied to the inputs to the hidden nodes and $w_{j,k}$ are the weighting factors applied to the inputs to the output nodes. The value of α was set arbitrarily to 0.01. Including the squares of the parameters in the search helped to stabilize the numerical minimization without altering the solution significantly. The difference between *SSE* and *MSSE* was less than 1%.

3.4. Constructing the Artificial Neural Network

The ANN for CN and OC was developed in a stepwise fashion. The process began by selecting an initial working data set, which was subset of the overall data set consisting of simple molecules containing only a few functional groups. The optimum number of hidden nodes, connections between nodes, and parameters were determined for this initial set of compounds . Data were added to the working data set a few molecules at a time. The additional data represented either molecules with a new functional group or molecules with the same functional groups but increasing complexity. ANN structure and parameters were optimized each time the working data set was expanded. The process resembles a step regression analysis.

The initial working data set consisted of n-alkanes and terminal alcohols of n-alkanes. The structure of these molecules could be represented by three functional groups – the methyl group ($–CH_3$), the methylene group ($–CH_2–$), and the alcohol group (-OH). The search for the optimal ANN structure for this small working data set began by assuming a simple structure consisting of one hidden node with each hidden node connected to only one output node. Figure 8 is an illustration of the initial ANN structure.

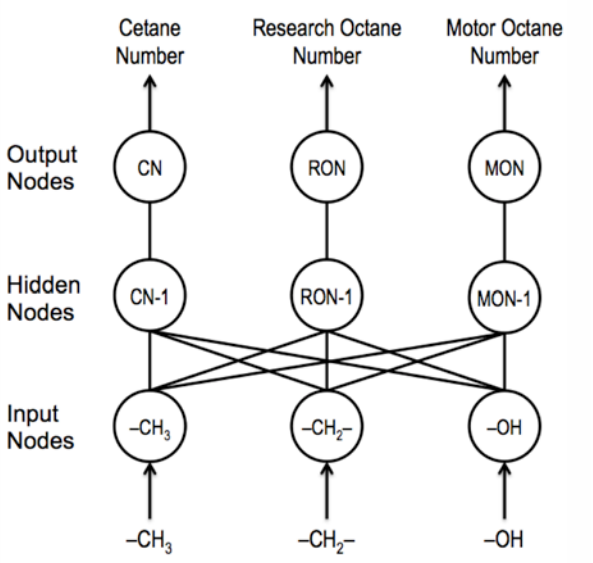

Fig. 8. Initial ANN structure.

Calculations were performed using Microsoft Excel with Visual Basic for Applications (VBA). The ANN was implemented with VBA. Excel functioned primarily as a convenient input/output interface.

The optimum model parameters were obtained by minimizing Eq. (11) using the Solver add-in in Excel. The particular algorithm used was the GRG Nonlinear Solver, which is a conditional gradient method.[41] Equation (11) is a non-convex function. Conditional gradient methods, like most minimization algorithms, do not guarantee global convergence for non-convex functions. Therefore, the constrained multistart option in the Solver add-in was used. The multistart option solves the minimization problem multiple times with random initial estimates, which increases the likelihood of convergence to a global minimum.

Figure 9 shows the CN results for the initial ANN. The initial network gives reasonable results for CN of primary alcohols and n-alkanes with carbon numbers greater than 5. Failure of the initial ANN to capture CN behavior at low carbon numbers indicates that the initial ANN is inadequate and needs to be modified. The following procedure was used to determine the optimum modifications to the ANN.

1. The network was modified by adding an additional hidden node or by adding an additional connection between a hidden node and an output node.
2. Parameters for the modified network were optimized by minimizing the modified sum of squared errors.
3. An F-test was used to determine if the modifications improved the model significantly.
   A. If Prob(F) ≥ 0.9, the modification was accepted because it resulted in a significant improvement.
   B. If Prob(F) < 0.5, the modification was rejected because it resulted in no improvement.

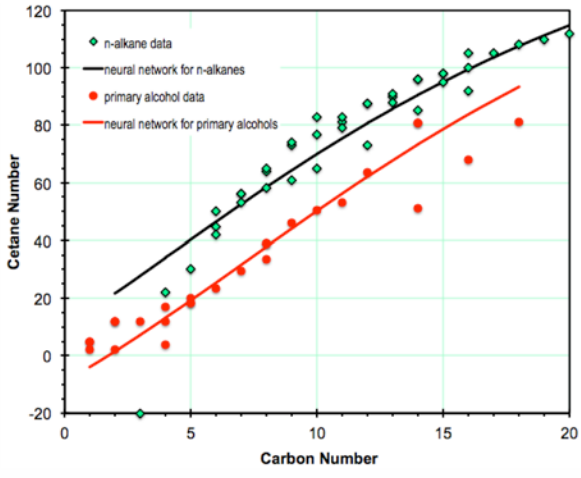

Fig. 9. A comparison of data to CN predictions for the initial ANN.

C. If $0.5 \leq \text{Prob(F)} < 0.9$, the residuals were examined. If the modification eliminated one or more outliers, it was accepted. An outlier was defined as a measurement that exhibited large deviations from the model. If no outliers were eliminated, the modification was rejected.

4. Steps 1 – 3 were repeated until no additional improvements could be identified.

In the second step of the procedure, separate F-tests were performed for CN, RON, and MON. The following equation defines the F statistic.

$$F = \frac{\left(SSE_2 - SSE_1\right)/\left(DOF_1 - DOF_2\right)}{SSE_2/DOF_2} \tag{12}$$

where SSE_1 is the sum of squared errors for the current model, SSE_2 is the sum of squared errors for proposed model modification, DOF_1 is the degrees of freedom for the current model, and DOF_2 is the degrees of freedom for the proposed modification. The degrees of freedom are the number of measurements minus the number of parameters. The number of parameters is not the total number of parameters in the model, but the number of parameters that affect the given output. For example, the number of parameters for MON in the ANN shown in Fig. 8 only includes the parameters associated with hidden node MON-1 and output node MON because they are the only parameters that affect the predicted value of MON.

Figure 10 shows the modified ANN for the initial working data set. The modifications include an additional node for CN and additional connections between the hidden nodes and output nodes for RON and MON. Figure 11 shows the CN results for the modified network. The additional hidden node for CN accounts for the qualitatively different behavior of n-alkanes and primary alcohols with low carbon numbers. The additional hidden node increases the number of parameters in the model, but it also improves the quality of the fit significantly. The additional links between RON and MON suggests that these qualities are dependent on common underlying molecular structures.

The ANN illustrated in Fig. 10 was expanded in a deliberate, step-wise fashion to include more compounds and functional groups. Each step of the network development process was initiated by expanding the compounds in the working data set one function group at a time (e.g. adding

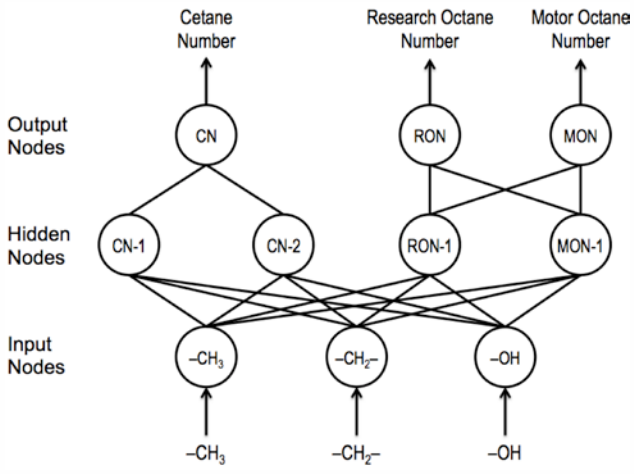

Fig. 10. The modified ANN for the initial working data set.

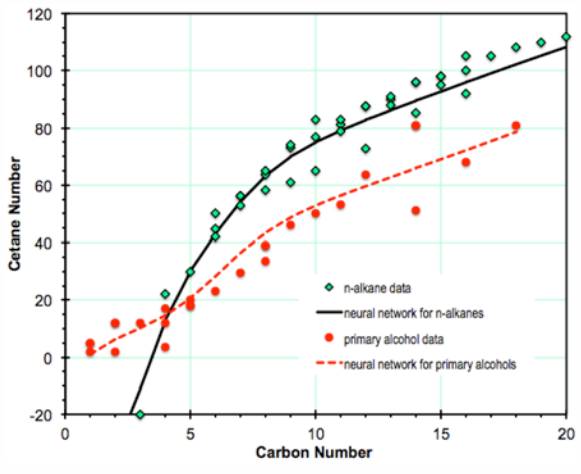

Fig. 11. A comparison of data to CN predictions for the modification of the initial ANN.

monosubstituted alkanes to the initial working data set) or expanding the complexity of the compounds in the working data set one level at a time (adding polysubstituted alkanes to a working data set that contains monosubstituted and disubstituted alkanes). The following procedure was used to expand and optimize the ANN.

1. Data for fuels containing one additional functional group or one additional level of complexity were added to the working data set.
2. If a new functional group was added to the working data set, a new input node was added to the ANN, and that node was connected to all hidden nodes in the network.
3. Parameters for the expanded network were optimized using the expanded data set.
4. The network was modified by
   A. making finer distinctions in the functional groups (e.g. defining functional groups for primary, secondary, and tertiary alcohols versus using a single generic alcohol group),
   B. adding connections between hidden nodes and output nodes, or
   C. adding additional hidden nodes.
5. Parameters for the modified network were optimized.
6. An F-test was used to determine if the modifications improved the model significantly.
   A. If Prob(F) $\geq$ 0.9, the modification was accepted because it resulted in a significant improvement.
   B. If Prob(F) < 0.5, the modification was rejected because it resulted in no improvement.
   C. If 0.5 $\leq$ Prob(F) < 0.9, the residual were examined. If the modification eliminated one or more outliers, it was accepted. If no outliers were eliminated, the modification was rejected.
7. Steps 4 – 6 were repeated until no additional improvements could be identified.
8. Steps 1 – 7 were repeated until all the data were included in the working data set.

The final ANN consisted of 38 input nodes corresponding to 38 separate functional groups, 4 hidden nodes for CN and 4 hidden nodes for RON and MON. Figure 12 is an illustration of final ANN. Effectively, the final solution consists of two independent ANN – the first for CN and the second for RON and MON. The dependence of RON and MON on the same set of hidden nodes is not surprising. Both quantities measure a fuels performance in a spark-ignited

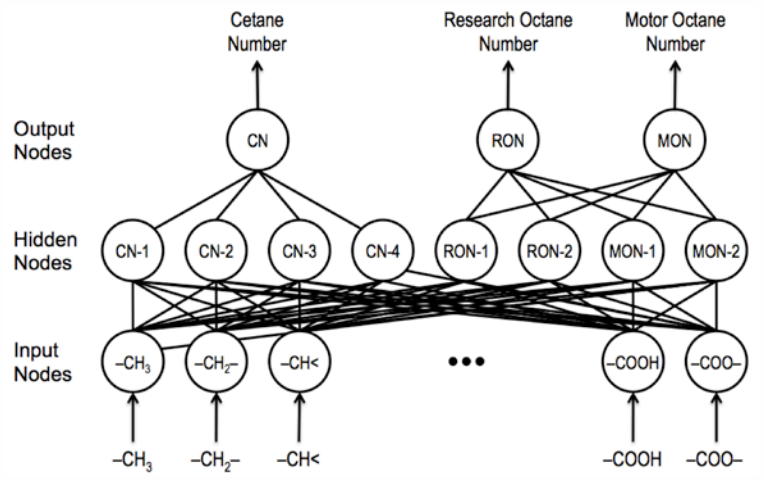

Fig. 12. Final ANN structure for estimating CN, RON, and MON functional group contributions.

engine; and therefore, both should depend on the same underlying chemical structures. This dependence on common structures is supported by the high degree of correlation between RON and MON ($r^2 = 0.94$) as illustrated in Fig. 13. CN depends on a different set of hidden nodes than RON and MON, which suggests that CN depend different chemical structures. The lack of dependence on common structures contradicts correlations between CN and ON a reported in the literature.[42,43,44] However, these correlations were developed from limited data for gasoline and simple hydrocarbon mixtures. These data are not representative of the pure hydrocarbons and oxygenated hydrocarbons included in the data set used for this study. The RON and CN data compiled for this study are plotted in Fig. 14, which shows a poor correlation ($r^2 = 0.45$) between CN and RON. The ANN structure shown in Fig. 12 accurately reflects the observed correlations between RON and MON and CN and RON.

The functional groups used in the ANN are listed in Table 6. Some functional groups depend on the position of the group in the carbon chain or on the ring. This approach is similar to that used by Albahri.[13] The symbol C# in Table 6 indicates carbon number. Carbon number is determined by locating the longest chain on which the functional group is located and counting from the nearest terminal carbon or oxygen atom in the chain. The CN and ONs for ethers and esters were found to have a nonlinear dependence on the number of groups in the molecule. The nonlinearities were accounted by defining different function groups for the first, second, and third and higher ether group in the molecule and defining different functional groups for the first, second, and third ester groups.

All parameters for the ANN-based group contribution model are given in Appendix C. Appendix C also includes eight examples illustrating how the group structure of a molecule is determined.

3.5. Discussion of Quality of Fit and Errors

Assessing the quality of fit for the ANN is difficultndue, in part, to the difficulty of summarizing complex multivariate data on a few simple plots. Results for straight-chain alkanes, alcohols, and terminal alkenes can be plotted as a function of carbon number as illustrated in Figs. 15 and 16. However, plots showing limited results can be deceiving. Figure 15 indicates that the ANN fits the CN data for n-alkanes and primary alcohols within measurement errors. Figure 16 shows very good agreement between the ANN and the data for CN and RON for terminal alkenes. However, these simple plots are not representative of the full range of regression errors. Furthermore, data for branched alkanes and alcohols cannot be plotted as a function of a single, representative parameter. Different statistics and plots must be used to represent the data and the quality of the fit.

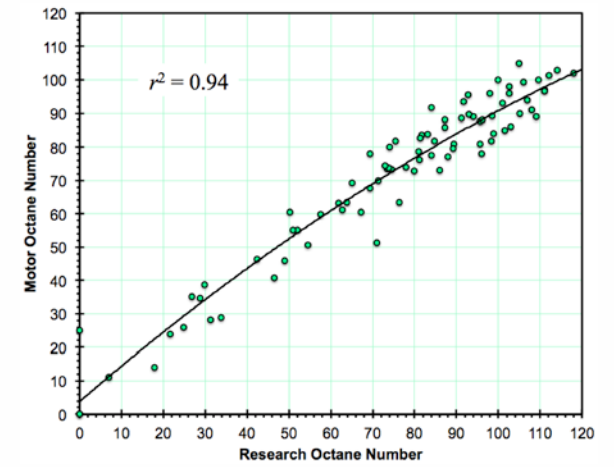

Fig. 13. MON as a function RON.

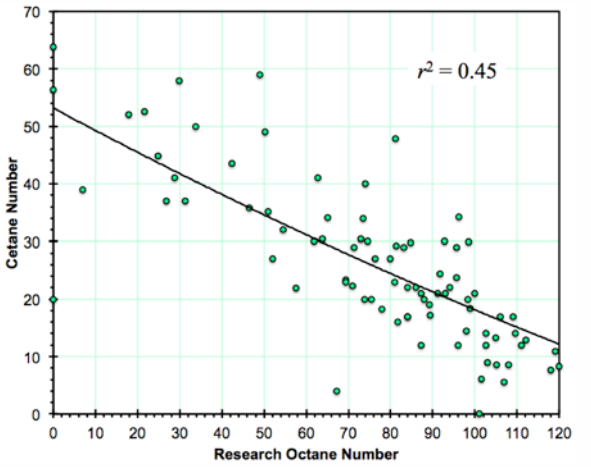

Fig. 14. CN as a function RON.

Table 6. List of functional groups in the ANN.

Class	Group	Class	Group	Class	Group
Paraffins	$-CH_3$	Alkynes	≡CH	Aromatics	=C< (para)
	$-CH_2-$		≡C–	Oxygenates	–OH
	–CH< (C# = 2)	Naphthenes	$-CH_2-$		=O (aldehyde)
	–CH< (C# = 3)		–CH<		=O (ketone)
	–CH< (C# ≥ 4)		–CH< (ortho)		–O– (1^{st})
	>C< (C# = 2)		>C<		–O– (2^{nd}+)
	>C< (C# ≥ 3)		=CH–		–O– (ring)
Olefins	$=CH_2$		>C=		–O– (aromatic)
	=CH– (C# = 2)	Aromatics	=CH–		–COOH
	=CH– (C# = 3)		=C< (ring/ring)		–COO– (1^{st})
	=CH– (C# ≥ 4)		=C<		–COO– (2^{nd})
	>C= (C# = 2)		=C< (ortho)		–COO– (3^{rd})
	>C= (C# ≥ 3)		=C< (meta)		

Bias, standard deviation, and correlation coefficient (r^2) are quantitative measure of goodness of fit. Table 7 summarizes these quantities for the ANN-based group contribution method. Estimates exhibit a small, insignificant bias for CN and RON and no bias for MON. Standard deviations for errors in the ANN predictions are larger than the estimated variability in the data. A t-test confirms that the differences are significant. Therefore, the ANN does not fit the data within experimental errors. Correlation coefficients for CN of hydrocarbons, RON and MON are greater than 0.9. The correlation coefficient for CN of oxygenated hydrocarbons is 0.86. The lower correlation coefficient for oxygenated hydrocarbons is due in part to the greater uncertainty in the data. Data uncertainty accounts for 32% of variance in CN errors for hydrocarbons, but it accounts for 43% of the variance in errors for oxygenated hydrocarbons.

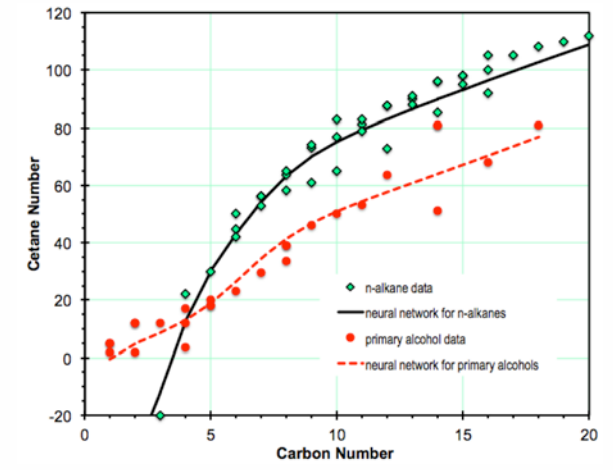

Fig. 15. ANN results for CN of n-alkanes and primary alcohols.

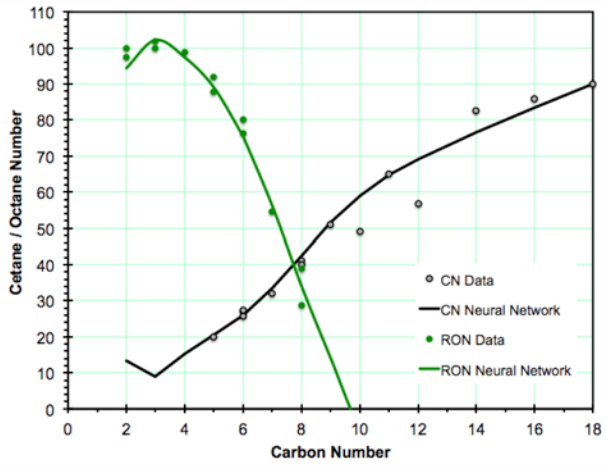

Fig. 16. ANN results for CN and RON of terminal alkenes.

Table 7. Bias, standard deviation, and correlation coefficients for the ANN.

Property	Bias in Correlation	Standard Deviation		Correlation Coefficient
		Data Uncertainty	Correlation Error	
CN - Hydrocarbons	0.1	4.1	7.2	0.93
CN - Oxygenates	-0.2	7.5	11.4	0.86
RON	-0.1	3.4	6.9	0.93
MON	0.0	3.3	6.1	0.91

Smolenskii et al.[18] state the accuracy of a structure-property relationship decrease with increasing "complexity" of the property. Reasonably accurate models can be developed for properties with a low level of complexity, such as density and critical properties. Models with a moderate degree of complexity, such as CN, are often inaccurate. Therefore, correlation coefficients between 0.86 and 0.93 are indicative of a reasonably successful model of moderately complex properties.

Figures 17 – 20 are plots of predicted values versus measured values for CN and ONs. The scatter about the diagonal for CN of hydrocarbons, RON, and MON indicate that the absolute model uncertainty is constant for these properties. However, Fig. 18 indicates that uncertainty in CN for oxygenated hydrocarbons increases with increasing CN. A maximum likelihood fit of the model errors to a zero-mean normal distribution confirm that absolute model uncertainty for CN of hydrocarbons, RON, and MON are constant and that uncertainty in CN for oxygenated hydrocarbons increase with CN according to the following linear equation

$$\sigma_{CN} = 6.3 + 0.083 \cdot CN \tag{13}$$

A plot of the CDF for model errors is another useful tool for understanding the nature of uncertainty. Figure 21 shows the CDFs for errors in CN. The CDFs for both hydrocarbons and oxygenated hydrocarbons are nearly linear (r^2 = 0.97 for hydrocarbons and 0.96 for oxygenated hydrocarbons) on a

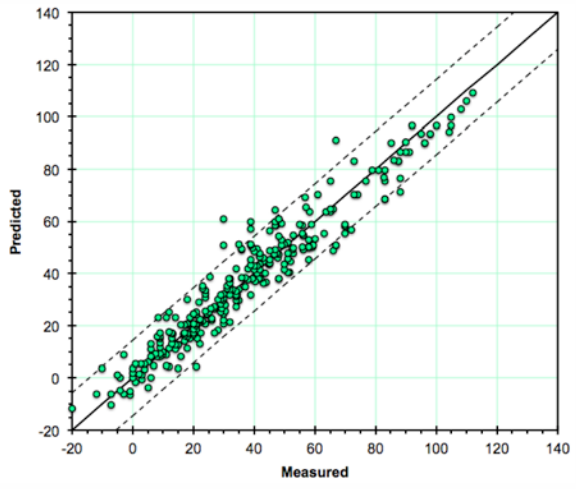

Fig. 17. Predicted versus measured values of CN for hydrocarbons.

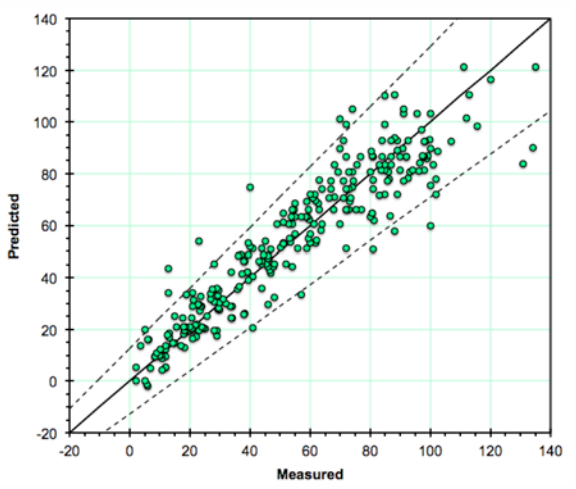

Fig. 18. Predicted versus measured values of CN for oxygenated hydrocarbons.

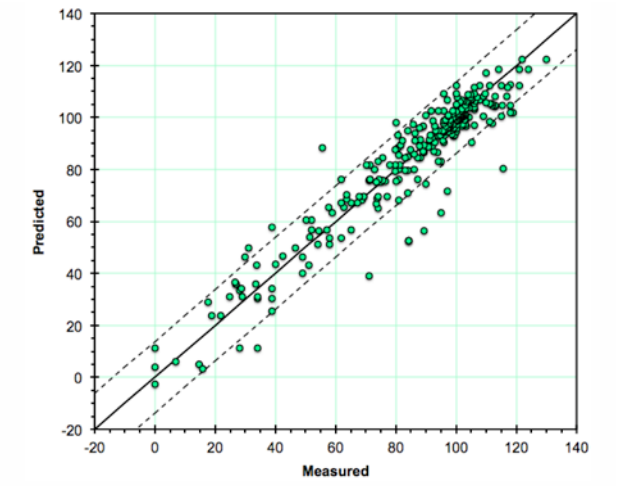

Fig. 19. Predicted versus measured values of RON.

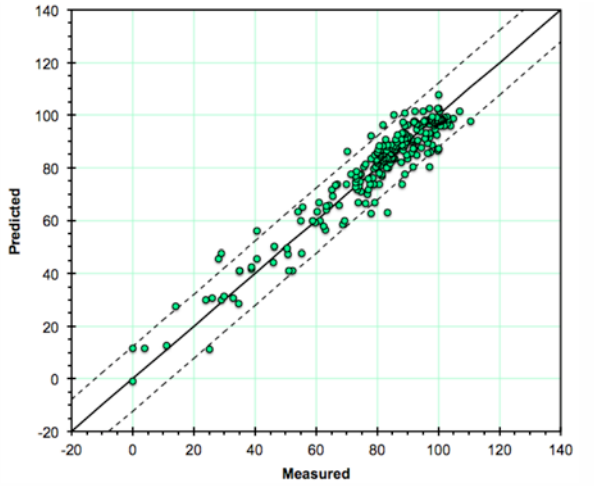

Fig. 20. Predicted versus measured values of MON.

normal probability scale indicating that the distributions of errors are approximately normal. The plot shows outliers in the tails of the distribution for the CN of oxygenated hydrocarbons. Outliers are points with large deviations from the normal distribution. Outliers constitute about 8% of the data for oxygenated hydrocarbons. They have a higher average carbon number than the other compounds (25 versus 14) and a higher average CN (99 versus 54). Outliers consist primarily of high molecular-weight esters and triglycerides. The deviations for the normal distribution are more likely the result of increasing uncertainty with CN rather than a problem with the correlation or data.

Figures 22 and 23 show the CDFs for RON and MON errors. Both CDFs are nearly linear on a normal probability scale (r^2 = 0.96 for RON and 0.98 form MON) indicating normal distributions. The tails of the CDF for RON include a few outliers, but these only constitute less than 3% of the data. The CDF for MON has no outliers.

3.6. Comparisons to Other Methods

To evaluate whether an ANN-based group contribution method is superior to a more standard approach, two additional group contribution methods were developed as part of this study. The first method used an approach similar to Albahri.[13] CN and ONs were correlated as a polynomial function of the sum of group contributions. In the second approach, neural networks consisting of a single hidden node and a single output node were developed for CN, RON, and MON. This second method is essentially the same as the first except that CN and ONs are correlated using a combination of logistic functions rather than a polynomial. Parameters for both correlations were determined using the complete data set developed for this study. Descriptions of both methods and their associated parameters are given in Appendix D.

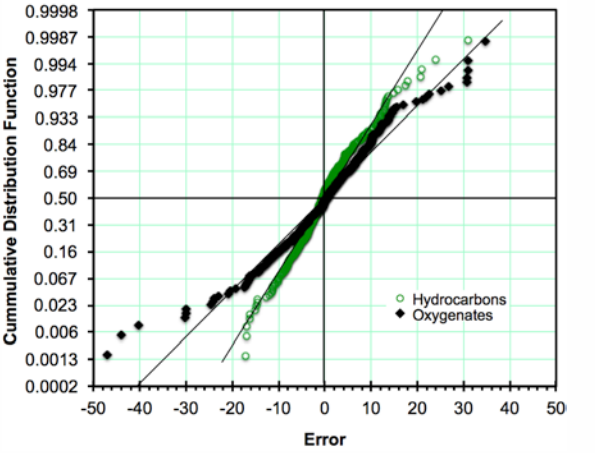

Fig. 21. CDFs for errors in CN predictions for hydrocarbons and oxygenated hydrocarbons.

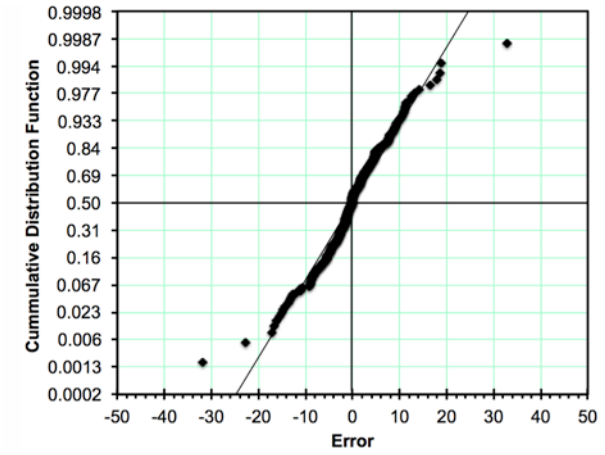

Fig. 22. CDFs for errors in RON predictions.

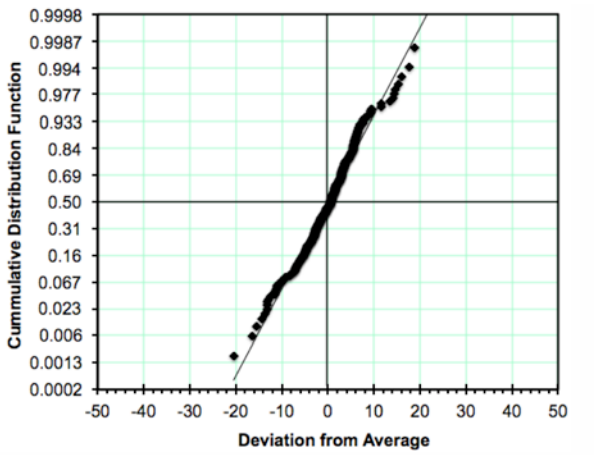

Fig. 23. CDFs for errors in MON predictions.

Correlation coefficients for the ANN-based method and the two alternative group contribution methods are summarized in Table 8. Because all three correlations were developed using the same data set, a comparison of correlation coefficients is a fair comparison of methods. Correlation coefficients for the group contribution method based on an ANN are greater than the correlation coefficients for the other two methods indicating that the ANN is superior to the other two methods. A simple comparison of correlation coefficients is a qualitative test. Statistical F-tests confirmed this conclusion for CN, RON, and MON a confidence level greater than 0.99.

Comparisons were also made to published group contribution methods for estimating CN, RON, and MON. These other methods were used to compute the CN, RON, and MON for all applicable compounds in the data set developed for this study. Correlation coefficients for the each method and each class of compounds are listed in Table 9. Because these published correlations did not include all classes of compounds considered in this study, correlation coefficients are given for each class of compound. If the method was not applicable to a particular class, the entry in Table 10 was left blank.

The correlation coefficients in Table 9 shows that the ANN-based group contribution method developed in this study is superior to other group contribution methods for both hydrocarbons and oxygenated hydrocarbons. However, correlation coefficients tell only part of the story. All of the group contribution methods listed in Table 9 perform well for compounds similar to those in the data sets used in their development. However, predicted values of CN and ONs are poor when the correlations are extrapolated. The group contribution method of DeFries et al. performs well for n-alkanes and alkanes with single methyl side groups, but it performs poorly for polysubstituted alkanes and aromatics. Damen and

Table 8. Comparison of correlation coefficients for different group contribution methods.

Method	Correlation Coefficients (r^2)		
	CN	RON	MON
ANN	0.90	0.93	0.91
Simplified ANN	0.80	0.84	0.88
“Standard” Method	0.80	0.85	0.80

Table 9. Comparison of correlation coefficients for the ANN group contribution method of this study to published group contribution methods.

Compound Class	CN			RON		MON	
	This Study	DeFries et al.[14]	Dahmen & Marquardt[11]	This Study	Albahri13	This Study	Albahri[13]
Paraffins	0.91	0.73	0.53	0.94	0.86	0.95	0.87
Olefins and Alkynes	0.90	-	-0.48	0.90	0.53	0.65	-1.55
Naphthenes	0.81	-	0.25	0.85	0.75	0.89	-0.40
Aromatics	0.87	0.44	0.58	0.76	-3.28	*	*
Oxygenates	0.85	-	0.41	0.62	-	0.56	-
Total	0.90	0.64	0.53	0.93	0.55	0.91	-1.16

*Insufficient data for meaningful evaluation of r^2.

Marquardt's method performs best when applied to straight chain hydrocarbons and oxygenated hydrocarbons containing more the four carbon atoms. However, their method gives poor results when applied to low molecular weight compounds and polysubstituted molecules. Albahri's group contribution method gives good results when applied to compounds with carbons numbers between four and nine. Poor results are obtained when the method is applied to compounds containing a higher or lower number of carbon atoms or heavily substituted hydrocarbons.

The poor overall performance of the published group contribution methods relative the ANN-based method is not simply the result of differences in methodology. Table 9 shows correlation coefficients for the two alternative group contribution methods developed as part of this study are mathematical similar to the published studies, but they were devel`oped using the data set compiled for this study. A comparison of the correlation coefficients in Tables 9 and 10 shows that the standard group contribution methods developed in this study are in much better agreement with the data than the published group contribution methods. Thus, the accuracy of the ANN-based group contribution method is not solely the result of using a more complex mathematical formulation. The use of a more extensive data set is a major factor, if not the most important factor, contributing to the improved accuracy of ANN-based group contribution method developed in this study.

4.0 CROSS VALIDATIONS

The discussion of errors in Section 3.5 addresses the issue of goodness of fit, or how well the ANN fits the data. Goodness of fit is related to, but not identical to, the predictive power of a statistical model. Because an ANN model contains a large number of parameters than other group contribution methods, the possibility exists that good results were obtained because model over fits the data. Models that over fit the data often have low predictive power, and they tend to be unreliable when extrapolated. The predictive power of a group contribution method describes the accuracy of the method when applied to molecules not included in the data used to develop the correlation. Two cross validation studies were done to assess the predicative power of the ANN-based group contribution method.

4.1. 10-Fold Cross Validation

In a 10-fold cross validation, the original data set is divided into ten random, disjoint subsets.[45] Each subset contains approximately the same number of members,[1] and the union of these ten subsets contains all the elements of the original data set. The data set for this study consists of a list of compounds with associated measurements of CN, RON, or MON. If more than one property measurement is available for a given compound, that compound is listed multiple times in the data set. The data set used in this study contained 743 elements. For the 10-fold cross validation, the original data set was divided into subsets containing 74 or 75 elements. Only 106 of the 743 entries in the data set have values for CN, RON, and MON. The majority of entries include values for only one of these three properties. When the data was divided into ten random subsets, no attempt was made to ensure that each subset contained equal measurements for CN, RON, or MON. The average number of measurements in each subset was 58 for CN, 28 for RON, and 25 for MON. The maximum deviation from these mean was ±7.

Of the 10 subsets, one was selected as the validation set for testing the model. Data for the remaining 9 subsets were combined, and used to determine a new set of model parameters. These new model parameters were then used to predicted CN, RON, and MON for the compounds in the validation set; and prediction errors were evaluated. This process was repeated until all ten subsets were used as validation sets.

Figures 24 – 26 show the CDFs for errors in the regression analysis and predictive errors obtained from the 10-fold cross validation. The CDFs for the regression analysis and the 10-fold cross validation are nearly identical for CN, RON, and MON indicating that the ANN has good predictive power and does not over fit the data. Standard deviations for the data, the regression analysis, and the 10-fold cross validation are given in Table 11. The standard deviation for errors in the regression analysis and 10-fold cross validation are both significantly greater than the variability of the data indicating that the model does not fit or predict the data within in experimental errors. Table 11 also includes the results of an F-test comparing the standard deviations of the validation test to the standard deviations of the regression analysis. The standard deviations for the validation test are larger, but the significance level is low. Also, the differences in standard deviations are less than 5%. The low significance of differences in standard deviations confirms the qualitative conclusion based on comparison of the CDFs for errors.

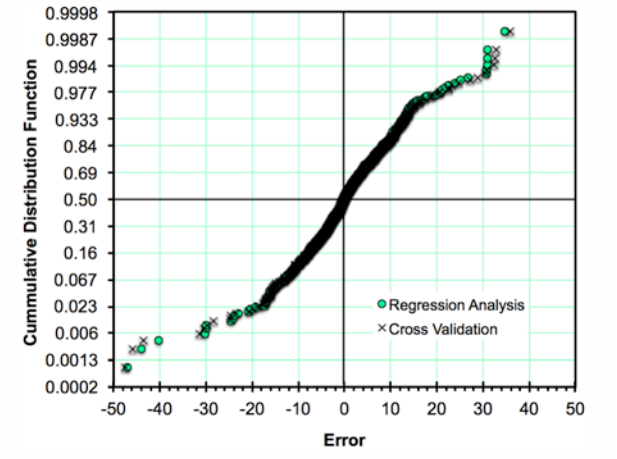

Fig. 24. CDFs for errors in CN predictions obtained from the regression analysis and the 10-fold cross validation.

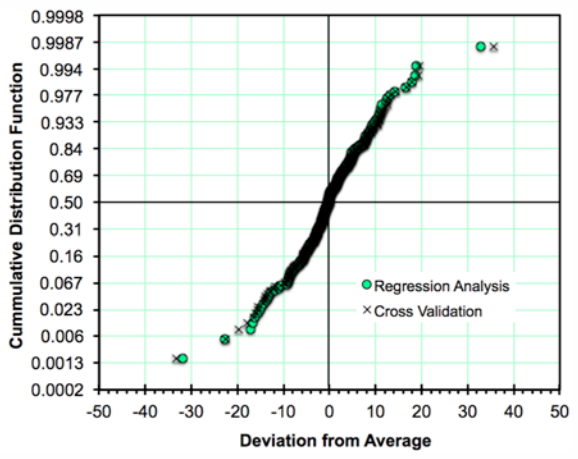

Fig. 25. CDFs for errors in RON predictions obtained from the regression analysis and the 10-fold cross validation.

[1] The subsets can only contain equal number of elements if the number of elements in the original data set is a multiple of 10.

4.2. Extrapolation Test

As noted in Section 3.6, published group contribution methods for estimating CN and ONs gave poor results when extrapolated. A second cross validation study was devised for this study to determined whether the ANN-based group contribution method could be extrapolated with a reasonable degree of confidence.

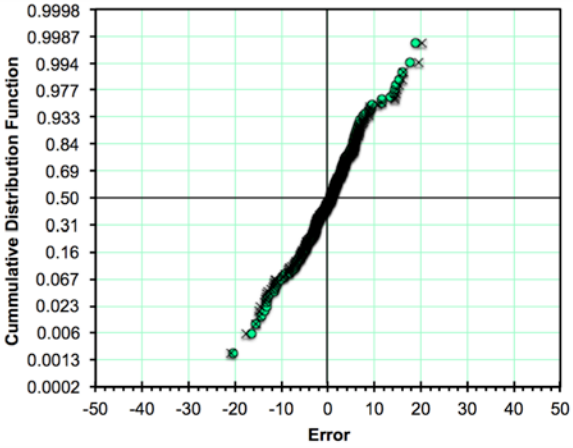

Fig. 26. CDFs for errors in MON predictions obtained from the regression analysis and the 10-fold cross validation.

In the extrapolation test, CN and ONs were evaluated separately. For CN, the original data set was divided into two subsets based on carbon number. The first subset contained compounds ranging from C_6 to C_{16}. Model parameters were determined from this subset. The validation set included all compounds not contained in the first data set. CN was determined for this data set, and the errors in the model prediction were evaluated. This calculation is an extrapolation of the model based on carbon number. A similar procedure was used to evaluate extrapolation of ONs. The first subset used to determine parameters for ON contained compounds ranging from C_5 to C_9. The validation set consisted of all other compounds in the original data set.

To determine the ANN could be extrapolated beyond the data set, the errors in the validation sets were compared to the errors in the regression analysis for the same set of compounds. A statistical F-test was used for the comparison. The null hypothesis was that the errors for the validation sets are significantly greater than the errors in the regression analysis for the same set of compounds. The probabilities that the null hypothesis is true are given in Table 11. The F-test results indicate a lack of strong evidence supporting they null hypothesis. Based on the lack evidence to the contrary, it was concluded that moderate extrapolations of the ANN-based group method does not result in large errors or unrealistic results.

Table 10. Comparison of standard deviation in data, regression analysis errors, and 10-fold validation errors.

Property	Standard Deviation			Probability that Validation Errors Greater Than Regression Errors
	Data Uncertainty	Regression Error	10-Fold Validation Error	
CN - Hydrocarbons	4.1	7.2	7.3	0.64
CN - Oxygenates	7.5	11.4	11.8	0.71
RON	3.4	6.9	7.2	0.77
MON	3.3	6.1	6.4	0.80

Table 11. Results of F-test for extrapolation validation sets.

Property	Probability Null Hypothesis is True
CN	0.54
RON	0.48
MON	0.48

Although the results of the extrapolation test were positive, the scope of the test was limited to moderate extrapolation of carbon number. This test does not prove that the ANN-based group contribution method is universally applicable. For example, the extrapolation test does not imply that the ANN-based method will give accurate results if applied to compounds containing large numbers of oxygenated functional groups, such as sugars or polyalcohols. The test, however, does provide some assurance that moderate extrapolation will result in physically reasonable results.

5.0 DISCUSSION

The accuracy of a structure-property relationship depends on the complexity of the property.[18] Enthalpy of formation in the vapor phase is a simple property that can be approximated by summing bond energies. Such a property can be readily correlated with a standard group contribution method. Empirical models of complex properties, such as freezing point, are senseless. Freezing depends on details of the molecular shape and short range interaction among function groups which are difficult or impossible to capture in a group contribution method or other structure-property relationship. CN and ONs are moderately complex properties. These properties depend on reaction kinetics involving complex interaction among function groups.

Several group contribution methods and other structure-property relationships have been published for estimating CN and ONs. Most of these methods are applicable only to hydrocarbons and they are based on limited set of data in terms of molecular size and complexity. The group contribution methods evaluated in this study were found to be reasonably accurate when applied to compounds similar to those in the data set used to develop the correlation. However, extrapolating these group contribution methods resulted in inaccurate or unrealistic predictions.

CN and ONs are moderately complex properties that cannot be accurately represented by a standard group contribution method. A successful structure-property relationship for CN and ONs requires a mathematical formulation that captures a greater degree of complexity than the standard methods. In this study, an ANN-based group contribution method was investigated as method for correlating CN, RON, and MON to molecular structure of hydrocarbons and oxygenated hydrocarbons. The data set used to develop the ANN-based group contribution method included compounds with a broad range of carbon numbers (C_1 – C_{57}), molecular complexity, and functional groups. To ensure that error estimates accurately reflected the inherent variability in CN and ON data,[38] multiple measurements were included in the data set when they were available.

The statistically informed approach used to develop the ANN-based group contribution method resulted in what is effectively two separate ANNs – one for CN and the other for RON and MON. Although attempts have been made to correlate octane and cetane numbers,[42,43,44] the data compiled for this study does not support such a relationship. Separate networks for CN and ONs reflect the lack of a significant correlation between CN and RON and CN and MON. The statistically informed approach led to the position-dependent function groups similar to the functional groups used in Albahri's[13] group contribution method for RON and MON. The position dependent function groups are indicative of the complex nature of CN, RON, and MON.

The ANN-based group contribution method developed in this study was successful in the sense that could represent broad range of data with a reasonable degree of accuracy, it has good predictive power, and it can be extrapolated beyond the base data set with a reasonable degree of confidence. Furthermore, it is more accurate than published group contribution methods and applicable to a broader range of compounds. The ANN-based group contribution method is only moderately successful in the sense that it could not represent the data within the estimated variability of the data.

When using the ANN-based group contribution method to estimate CN, RON, or MON, uncertainty in the predicted value should be based on the standard deviations obtained from the 10-fold validation because they are indicative of the prediction errors rather than the regression errors. Estimation errors for the ANN, which correspond to 2 standard deviations, are given in Table 12. Errors are constant for the CN of hydrocarbons, RON of all compounds, and MON of all compounds. Errors for CN of oxygenated hydrocarbons are a linear function of the predicted CN. These results are similar to the correlation error bands shown in Figs. 17 – 20.

Table 12. Estimation errors for the ANN-based group contribution method.

Property	Estimation Error
CN – Hydrocarbons	±15
CN – Oxygenates	±(9 + 0.26·CN)
RON	±14
MON	±13

6.0 CONCLUSIONS

The ANN-based group contribution method developed in this study for estimating CN, RON, and MON is superior to other group contribution methods including published group contribution methods and standard group contribution methods developed as part of this study. The cross validation tests demonstrate the predictive power of the new ANN-based method. Prediction errors are only slightly greater than the regression errors. Also, the correlation gives reasonable results when extrapolated. The positive cross validation results shows that, in spite of having significantly more empirical parameters than standard group contribution methods, the ANN-based method does not over fit the data.The success of the ANN-based group contribution method can be attributed to three factors:

- mathematic formulation – the more complex mathematic structure of the ANN is able to account for the complex interactions among groups better than standard group contribution methods;
- methodology – the statistically informed approach ensured that subtleties in the function dependence of CN and ONs on structure were account for without over fitting the data; and
- data – the large, more comprehensive data set ensured a more robust model.

Because the ANN-based group contribution method for estimating CN, RON, and MON is applicable to a broad range of hydrocarbons and oxygenated hydrocarbons, it is a reliable tool for evaluating new advanced biofuels during the early stages of the development. First, the method could be used to evaluate whether a proposed fuel is worth pursuing. A compound with a low CN would not make a good diesel fuel, while compounds with low RON and MON would be poor substitutes for or additives to gasoline. When several viable fuels are being considered, the ANN-based method could be used to rank the alternative to focus research efforts on the best option. Finally, CN and ONs are qualitative indicators of economic value. Fuels with high CN or ONs have greater economic value than fuels with low CN or ONs. Fuels with RON greater than 120 are potential octane boasters and, therefore have a value greater than their fuel value. Early evaluation of potential biofuels will help focus research efforts and funds on the best options.

REFERENCES

1. S. Naik, V, Gould, P, Rout, A. Dalai, "Production of First and Second Generation Biofuels: A Comprehensive Review," *Renewable and Sustainable Energy Reviews*, 14: 578 – 597 (2010).

2. A. Katritzky, M. Kuanar, S. Slavov, C. Hall, M Karelson, I. Kahn, and D. Dobchev, "Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction," *Chemical Reviews*, 110: 5714−5789 (2010).

3. A. Lydersen, *Estimation of Critical Properties of Organic Compounds*, University of Wisconsin College of Engineering Experimental Station report 3, Madison, WI (April 1955).

4. K. Joback, A Unified Approach to Physical Property Estimation Using Multivariate Statistical Techniques" S.M. thesis, Department of Chemical Engineering, Massachusetts Institute of Technology (1982).

5. C. Chueh and A. Swanson, "Estimation of Liquid Heat Capacity," *The Canadian Journal of Chemical Engineering*, 51:596 – 60 (1973).

6. Y. Yoneda, "An Estimation of the Thermodynamic Properties of Organic Compounds in the Ideal Gas State. I. Acyclic Compounds and Cyclic Compounds with a Ring of Cyclopentane, Cyclohexane, Benzene, or Naphthalene," *Bulletin of the Chemical Society of Japan*, 52: 1297 – 1314 (1979).

7. T. Thin and T. Trong, "Estimation of Standard Heats of Formation ΔH * fT , Standard Entropies of Formation ΔS * fT , Standard Free Energies of Formation ΔF * fT and Absolute Entropies S * fT of Hydrocarbons from Group Contributions: An Accurate Approach," *The Canadian Journal of Chemical Engineering*, 54: 344 – 357 (1976).

8. Aa. Fredenslund, J. Gmehling, and P, Rasmussen, *Vapor-Liquid Equilibria Using UNIFAC*, Elsvier, New York (1977).

9. R. Reid, J. Prausnitz, and B. Poling, *The Properties of Gases and Liquids*, 4th Ed. McGraw-Hill Book Company, New York (1987).

10. T. Albahri and R. George, "Artificial Neural Network Investigation of the Structural Group Contribution Method for Predicting Pure Component Auto Ignition Temperature," *Industrial and Engineering Chemistry Research*, 42: 5708 – 5714 (2003).

11. M. Dahmen and W. Marquardt, "A Novel Group Contribution Method for the Prediction of the Derived Cetane Number of Oxygenated Hydrocarbons," *Energy Fuels*, 29: 5781 – 5801 (2015).

12. R. Meusinger, and R. Moros, "Determination of Quantitative Structure Octane Rating Relationships of Hydrocarbons by Genetic Algorithms," *Chemometrics and Intelligent Laboratory Systems*, 46: 67−78 (1999).

13. T. Albahri, "Structural Group Contribution Method for Predicting the Octane Number of Pure Hydrocarbons and Their Mixtures," *Industrial and Engineering Chemistry Research*, 42: 657 – 662 (2003).

14. T. DeFries, R. Kastrup, and D. Indritz, "Prediction of Cetane Number by Group Additivity and Carbon-13 Nuclear Magnetic Resonance," *Industrial and Engineering Chemistry Research*, 26: 188−193 (1987).

15. H. Yang, C. Fairbridge, and Z. Ring, "Neural Network Prediction of Cetane Numbers for Isoparaffins and Diesel Fuel," *Petroleum Science and Engineering*, 19: 573 – 586 (2001).

16. B. Creton, C. Dartiguelongue, T. de Bruin, and H. Toulhoat "Prediction of the Cetane Number of Diesel Compounds Using the Quantitative Structure Property Relationship," *Energy Fuels* 24: 5396 – 5403 (2010).

17. E. Smolenskii, "On Some Aspects of the Structure-Property Problem," *Doklady Chemistry* (English translation), 365: 93 – 98 (1999).

18. E. Smolenskii, V. Bavykin, A. Ryzhov, O. Slovokhotova, I. Chuvaeva, and A. Lapidus, "Cetane Numbers of Hydrocarbons: Calculations Using Optimal Topological Indices," *Russian Chemical Bulletin*, 57: 461− 467 (2008).

19. ASTM D6890, Standard Test Method for Determination of Ignition Delay and Derived Cetane Number (DCN) of Diesel Fuel Oils by Combustion in a Constant Vol. Chamber; American Society for Testing and Materials (ASTM): West Conshohocken, PA (2011).

20. D. Saldana, L. Starck, P. Mougin, B. Rousseau, L. Pidol, N. Jeuland, and B. Creton, "Flash Point and Cetane Number Predictions for Fuel Compounds Using Quantitative Structure Property Relationship (QSPR) Methods," *Energy and Fuels*, 25: 3900−3908 (2011).

21. P. Ghosh and S. Jaffe, "Detailed Composition-based Model for Predicting the Cetane Number of Diesel Fuels," *Industrial Engineering Chemistry Research*, 45: 346 – 351 (2006).

22. P. Ghosh, K. Hickey, and S. Jaffe, "Development of a Detailed Gasoline Composition-Based Octane Model," *Industrial Engineering Chemistry Research*, 45: 337 – 347 (2006).

23. M. Moosavi and N. Soltani, "Prediction of Hydrocarbon Densities Using an Artificial Neural Network-Group Contribution Method up to High Temperatures and Pressures," *Thermochimica Acta*, 556: 89 – 96 (2013).

24. J. Yanowitz, M. Ratcliff, R. McCormick, J. Taylor, and M. Murphy, "Compendium of Experimental Cetane Numbers," National Renewable Energy Laboratory technical report NREL/SR-540-36805 (September 2004).

25. D. Saldana, L. Starck, P. Mougin, B. Rousseau, L. Pidol, N. Jeuland, and B. Creton, "Flash Point and Cetane Number Prediction for Fuel Compounds Using Quantitative Structure Property Relationship (QSPR) Methods," *Energy and Fuels*, 25: 3900 – 3908 (2011).

26. W. Kamei and S. Singal, "Dimethyl Ether (DME) as Fuel for Automotive Diesel Engines," Petrotech-2010, New Delhi, India (October 31 – November 3, 2010).

27. D. Nagdeote and M. Deshmukh, "Experimental Study of Diethyl Ether and Ethanol Additives with Biodiesel-Diesel Blended Fuel Engine," *International Journal of Emerging Technology and Advanced Engineering*, 2: 195 – 199 (2012).

28. D. Naegeli, S. Moulton, E. Owens, and E. Frame, "Oxygenates for Advanced Petroleum-Based Diesel Fuels," Southwest Research Institute interim report TFLRF No. 351 (February 2001).

29. B, Freedman and M. Bagby, "Predicting Cetane Numbers of n-Alcohols and Methyl Esters from Their Physical Properties," *Journal of American Oil Chemists Society*, 67: 565 – 571 (1990).

30. G. Knothe, A. Matheaus, and T. Ryan, "Cetane Numbers of Branched and Straight-Chain Fatty Esters Determined in an Ignition Quality Tester," *Fuel* 82: 971 – 975 (2003).

31. *Technical Data Book: Petroleum Refining*, 11th Ed., American Petroleum Institute, Washington DC (1992).

32. G. da Solva and J. Bozzelli, "On Reactivity of Methylbenzenes," *Combustion and Flame*, 157: 2175 – 2183 (2010).

33. B. Masum, H. Masjuki, M. Kalam, S. Palash, M. Wakil, and S. Imtenan, "Tailoring the Key Fuel Properties Using Different Alcohols (C2 – C6) and Their Evaluation in Gasoline Engine," *Energy Conversion and Management* 88: 382 – 390 (2014).

34. A. Hamadi, "Selective Additives for Improvement of Gasoline Octane Number," *Tikrit Journal of Engineering Sciences*, 17 : 22 – 35 (2010).

35. G. Egloff and P. Van Arsdell, "Octane Rating Relationships of Aliphatic, Alicyclic, Mononuclear Aromatic Hydrocarbons, Alcohols, Ethers, and Ketones," *Journal of the Institute of Petroleum (London)*, 27: 121 – 138 (1941).

36. J. Yanowitz, E. Christensen, and R. McCormick, "Utilization of Renewable Oxygenates as Gasoline Blending Components," National Renewable Energy Laboratory technical report NREL/TP-5400-50791 (August 2011).

37. I. Premachandra, "An Approximation of the Activity Distribution in PERT," *Computers & Operations Research*, 28: 443 – 453 (2001).

38. K. Huber and J. Hauber, "Biofire – Biogenic Fuel Ignition Research," in *Sustainable Automotive Technologies 2012*, A. Subic, J. Wellnitz, M. Leary, and L. Koopmans Eds., Springer-Verlag, Berlin (2012).

39. J. Williams, "HEART – A Proposed Method for Assessing and Reducing Human Error," *Proceedings of the 9th Advances in Reliability Technology Symposium*, B3/R/1 – B3/R/13, University of Bradford (1986).

40. A. Lapidus, E. Smolenskii, V. Bavykin, T. Myshenkova, and L. Kondrat'ev, "Models for the Calculation and Prediction of the Octane and Cetane Numbers of Individual Hydrocarbons," *Petroleum Chemistry* 48: 277 – 286 (2008).

41. E. Levitin and B. Polyak, "Constrained Minimization Methods," *USSR Computational Mathematics and Mathematical Physics*, 3: 1 – 50 (1966).

42. J. Bowden, A. Johnston, and J. Russel, "Octane-Cetane Relationship," Southwest Research Institute Report number AD-779 512 (March 1974).

43. G. Kalgahatgi, “Auto-Ignition Quality of Practical Fuels and Implications for Fuel Requirements of Future SI and HCCI Engines,” SAE Technical Papers, DOI: 10.4271/2005-01-0239 (2005).

44. W. Morris, “Method Relates Diesel Cetane, Octane Ratings,” *Oil & Gas Journal*, 105 (45): 58 – 60 (December 3, 2007).

45. G. McLachlan, K-A Do, and C. Ambroise, *Analyzing Microarray Gene Expression Data*, John Wiley & Sons, Inc., Hoboken, NJ (2004).

APPENDIX A
CETANE NUMBER VERSUS DERIVED CETANE NUMBER

Several methods are available for determining the cetane number (CN) of diesel fuels. The standard test method is specified in ASTM D614.[1] A common alternative to the standard method is based on measurement of ignition delay and a correlation relating ignition delay to CN. The CN number determined from ignition delay is called derived CN (DCN). The procedures for measuring ignition delay determining DCN are given in ASTM D6890.[2]

When developing a structure-property relationship for CN, the question arises as to whether CN and DCN should be considered as separate properties, as is the case for research and motor octane number, or whether they should be considered the same property measured using different techniques? The National Renewable Energy Laboratory's (NREL's) compilation of CN data[3] contains several compounds with data for CN as determined using ASTM D614 and DCN determined using ASTM D6890. These data were used to determine what, if any, differences exist between CN and DCN.

Figure A-1 shows DCN as a function of CN. The plot includes data for hydrocarbons and oxygenated hydrocarbons. Three models of the relationship between DCN and CN were evaluated – a linear model, a linear model with and intercept of zero, and a constant bias. The linear model had the minimum sum of squares and the linear model with an intercept of zero had the maximum sum of squares. However, based on an F-test, none of the three models was significantly better than the others. Therefore, the simplest model, which is a constant bias, was used to relate CN to DCN.

The average deviation of the DCN from the CN or bias is +2.7. The standard deviation about the mean is 5.3. The bias in the DCN is smaller than the standard deviation; but according to a t-test, the bias is significant different from zero. The confidence level for this conclusion is greater than 99%. Using the methodology discussed in Section 2.2, the variability in CN data has a standard deviation of 5.8. The constant bias model relating DCN to CN, therefore, fits the data within the uncertainty of the CN data.

Based on the limited amount of data considered in this study, DCN is a biased estimate of CN. Although the bias is statistically significant it is small relative the variability in CN measurements. Furthermore, diesel fuels of practical interest have CNs greater than 40, so the bias is less than 7% of the absolute values of CNs of practical interest. From a practical point of view, the small bias is not significant. Therefore it is appropriate to include both CN and DCN in a data set used to develop a group contribution method for estimating CN.

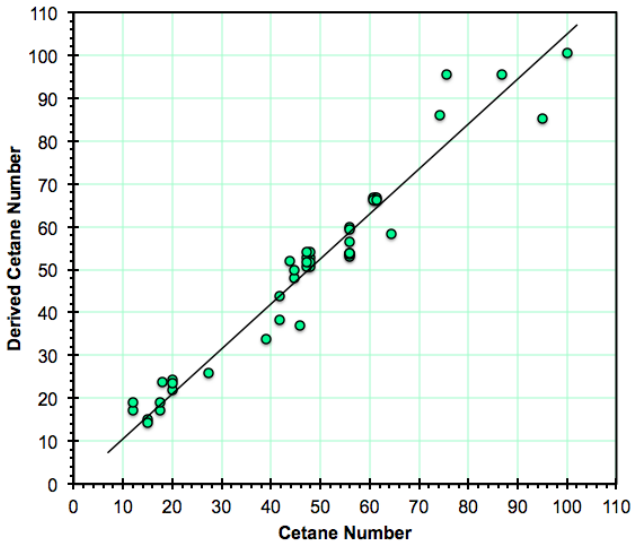

Fig. A-1. DCN as a function of CN for hydrocarbons and oxygenated hydrocarbons.

REFERENCES

1. ASTM D613 - 16, Standard Test Method for Determination for Cetane Number of Diesel Fuel Oil, American Society for Testing and Materials (ASTM): West Conshohocken, PA, (2016).

2. ASTM D6890, Standard Test Method for Determination of Ignition Delay and Derived Cetane Number (DCN) of Diesel Fuel Oils by Combustion in a Constant Vol. Chamber; American Society for Testing and Materials (ASTM): West Conshohocken, PA (2011).

3. J. Yanowitz, M. Ratcliff, R. McCormick, J. Taylor, and M. Murphy, "Compendium of Experimental Cetane Numbers," National Renewable Energy Laboratory technical report NREL/SR-540-36805 (September 2004).

APPENDIX B
OCTANE AND CETANE NUMBER DATA TABULATION

Table B-1. Cetane and Octane Number Data for Paraffins.

Compound Name	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Ethane	C_2H_6				114.9	100.0		99.0		
Propane	C_3H_8	-20.0			111.0	100.0		96.6		
2-Methylpropane	C_4H_{10}				102.1			97.0		
n-Butane	C_4H_{10}	22.0			94.0	95.0		89.1		
2-Methylbutane	C_5H_{12}	21.0			93.0	96.0		89.6		93
2,2-Dimethylpropane	C_5H_{12}				85.5			80.2		
n-Pentane	C_5H_{12}	30.0			61.8	58.0		63.2		
2-Methylpentane	C_6H_{14}	34.0	29.0	34.5	73.4	75.0		73.5		77
2,2-Dimethylbutane	C_6H_{14}	24.4	21.0		91.8	96.0		93.4		97
2,3-Dimethylbutane	C_6H_{14}				104.3			94.2		
3-Methylpentane	C_6H_{14}	30.0			74.5			73.3		
n-Hexane	C_6H_{14}	44.8	42.0	50.0	29.0	24.8	34.0	26.0		33.0
2-Methylhexane	C_7H_{16}	43.5			42.4			46.4		
2,2-Dimethylpentane	C_7H_{16}	30.0			92.8	89.0		95.6	89	
2,2,3-Trimethylbutane	C_7H_{16}	12.9	9.0		112.1	113.0		101.3		103
2,3-Dimethylpentane	C_7H_{16}	21.0	30.0		91.1	88.0		88.5	83	
2,4-Dimethylpentane	C_7H_{16}	29.0	28.0		83.1		88.0	83.8	83	
3-Ethylpentane	C_7H_{16}	34.1			65.0			69.3		
3-Methylhexane	C_7H_{16}	27.0			52.0		57.0	55.0		59
3,3-Dimethylpentane	C_7H_{16}				80.8			86.6		
n-Heptane	C_7H_{16}	56.3	53.0	56.3	0.0			0.0		4
2-Methyl-3-ethylpentane	C_8H_{18}	12.0			87.3	81.0		88.1	79	
2-Methylheptane	C_8H_{18}	52.6	47.0	56.2	21.7	19.0		23.8		29
2,2-Dimethylhexane	C_8H_{18}	40.0			74.0			80.0		
2,2,3-Trimethylpentane	C_8H_{18}	14.0			109.6	106.0		99.9		100
2,2,4-Trimethylpentane	C_8H_{18}	21.0			100.0	96.0		100.0	97	
2,3-Dimethylhexane	C_8H_{18}				71.3			78.9		
2,3,3-Trimethylpentane	C_8H_{18}	17.0			106.1		111.0	99.4		102
2,3,4-Trimethylpentane	C_8H_{18}	14.0	11.1	18.9	102.7			95.9		

Compound Name	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
2,4-Dimethylhexane	C_8H_{18}	29.0			71.3	62.0		69.9	67	
2,5-Dimethylhexane	C_8H_{18}				65.2			55.7		
3-Ethylhexane	C_8H_{18}				33.5			52.4		
3-Methyl-3-ethylpentane	C_8H_{18}				80.8			88.7		
3-Methylheptane	C_8H_{18}	37.0			26.8	27.0		35.0	35	
3,3-Dimethylhexane	C_8H_{18}				55.5			83.4		
3,4-Dimethylhexane	C_8H_{18}	20.0			75.5	71.0		81.7	73	
4-Methylheptane	C_8H_{18}				26.7			39.0		
n-Octane	C_8H_{18}	63.8	58.2	65.0	0.0			0.0		
2-Methyloctane	C_9H_{20}	39.0			7.0			11.0		
2,2-Dimethyl-3-ethylpentane	C_9H_{20}				112.1			99.5		
2,2-Dimethylheptane	C_9H_{20}	49.0			50.3		52.0	60.5	54	
2,2,3,3-Tetramethylhexane	C_9H_{20}				112.8			92.4		
2,2,3,3-Tetramethylpentane	C_9H_{20}				116.8			95.0		
2,2,5-Trimethylhexane	C_9H_{20}	24.0								
2,4-Dimethyl-3-ethylpentane	C_9H_{20}				105.3			96.6		
3,3-Diethylpentane	C_9H_{20}	17.0			84.0	83.0		91.6		97
3,3,5-Trimethylheptane	C_9H_{20}				86.5			88.7		
n-Nonane	C_9H_{20}	73.0	60.9	74.0						
2,2-Dimethyloctane	$C_{10}H_{22}$	59.0	55.0		49.0			46.0		
2,6-Dimethyloctane	$C_{10}H_{22}$	51.7								
n-Decane	$C_{10}H_{22}$	76.9	65.0	83.0						
n-Undecane	$C_{11}H_{24}$	81.1	79.0	83.0						
2,2,4,5,6-Pentamethylheptane	$C_{12}H_{26}$	9.0								
2,2,4,6,6-Pentamethylheptane	$C_{12}H_{26}$	9.0								
2,7-Dimethyl-4,5-diethyloctane	$C_{12}H_{26}$	39.0								
3-Ethyldecane	$C_{12}H_{26}$	47.0								
4,5-Diethyloctane	$C_{12}H_{26}$	20.0	18.0		0.0			25.0		
n-Dodecane	$C_{12}H_{26}$	87.6	72.9	87.6						
2,5-Dimethylundecane	$C_{13}H_{28}$	58.0								
4-Propyldecane	$C_{13}H_{28}$	39.5								
5-Butylnonane	$C_{13}H_{28}$	53.0								
7,8-Dimethyltetradecane	$C_{13}H_{28}$	40.5								
n-Tridecane	$C_{13}H_{28}$	90.0	88.0	91.0						
n-Tetradecane	$C_{14}H_{30}$	96.1	85.1	96.1						

Compound Name	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
2,7-Dimethyl-4,5-diethyloctane	$C_{14}H_{30}$	39.0								
2,6,10-Trimethyldodecane	$C_{15}H_{32}$	58.0	58.0	59.1						
n-Pentadecane	$C_{15}H_{32}$	98.0	95.0	98.0						
2,2,4,4,6,8,8-Heptamethylnonane	$C_{16}H_{34}$	15.0	14.2	15.1						
5-Butyldodecane	$C_{16}H_{34}$	45.0								
7,8-Dimethyltetradecane	$C_{16}H_{34}$	40.0								
n-Hexadecane (Cetane)	$C_{16}H_{34}$	100.0	92.0	105.0						
7-Butyltridecane	$C_{17}H_{36}$	70.0								
n-Hetadecane	$C_{17}H_{36}$	105.0								
2-Methylheptadecane	$C_{18}H_{38}$	67.0								
5,6-Dibutyldecane	$C_{18}H_{38}$	29.8								
7,8-Diethyltetradecane	$C_{18}H_{38}$	67.0								
8-Propylpentadecane	$C_{18}H_{38}$	48.0								
9-Methylheptadecane	$C_{18}H_{38}$	30.0								
9,10-Dimethylocatdecane	$C_{18}H_{38}$	59.5								
n-Octadecane	$C_{18}H_{38}$	108.3								
2-Methyloctadecane	$C_{19}H_{40}$	104.4								
n-Nonadecane	$C_{19}H_{40}$	110.0								
9,10-Dimethyloctadecane	$C_{20}H_{42}$	60.0								
n-Eicosane	$C_{20}H_{42}$	112.0								
7-Hexylpentadecane	$C_{21}H_{44}$	83.0								
8-Hexylpentadecane	$C_{21}H_{44}$	83.0								
2,9-Dimethyl-5,6-diisopentyldecane	$C_{22}H_{46}$	48.0								
9-Heptylpentadecane	$C_{22}H_{46}$	88.0								
10,13-Dimethyldocosane	$C_{24}H_{50}$	56.0								
9-Heptylheptadecane	$C_{24}H_{50}$	88.0								
9,10-Dipropyloctadecane	$C_{24}H_{50}$	47.3								
9,10-Dipropyloctadecane	$C_{24}H_{50}$	47.0								

Table B-2. Cetane and Octane Number Data for Olefins.

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Ethene	C_2H_4				97.3		100.0	75.6		
Propene	C_3H_6				101.8	100.0		84.9		
1-Butene	C_4H_8				98.8			79.9		80
2-Butene	C_4H_8				101.6			99.9	83	
2-Methylpropene	C_4H_8				106.3			90.3		
Cyclopentene	C_5H_8	20			140.0					
1-Pentene	C_5H_{10}	20			87.9		92.0	77.1		
2-Methyl-1-butene	C_5H_{10}	20			98.3			81.9		
2-Methyl-2-Butene	C_5H_{10}				97.3			84.7		
2-Pentene	C_5H_{10}				87.8		98.0	87.8		
3-Methyl-1-butene	C_5H_{10}				97.5			97.5		
Cyclohexene	C_6H_{10}	23.8								
1-Hexene	C_6H_{12}	27	25.8	27.3	76.4		80.0	63.4		
2-Ethyl-1-butene	C_6H_{12}				99.3			84.3		
2-Hexene	C_6H_{12}				92.7	89.0		80.8		
2-Methyl-1-pentene	C_6H_{12}				94.2			81.5		
2-Methyl-2-pentene	C_6H_{12}	23.8			95.7			80.9		
2,3-Dimethyl-1-butene	C_6H_{12}				101.3			82.8		
2,3-Dimethyl-2-butene	C_6H_{12}				97.4			80.5		
3-Hexene	C_6H_{12}		.		94.0	84.0		80.1		
3-Methyl-1-pentene	C_6H_{12}				96.0			81.2		
3-Methyl-2-pentene	C_6H_{12}				97.2			83.0		
3,3-Dimethyl-1-butene	C_6H_{12}				111.7			93.3		
4-Methyl-2-pentene	C_6H_{12}				98.9			81.0		
4-Methyl-1-cyclohexene	C_6H_{12}	28.0								
1-Heptene	C_7H_{14}	32			54.5			50.7		
2-Ethyl-3-methyl-1-butene	C_7H_{14}				97.0			82.0		
2-Methyl-1-hexene	C_7H_{14}				90.7			78.8		
2-Methyl-2-hexene	C_7H_{14}				92.0			79.8		
2-Methyl-3-hexene	C_7H_{14}				97.9			82.0		
2,3-Dimethyl-1-pentene	C_7H_{14}				99.3			84.2		
2,3-Dimethyl-2-pentene	C_7H_{14}				97.5			80.0		
2,3,3-Trimethyl-1-butene	C_7H_{14}				105.3			90.5		

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
2,4-Dimethyl-1-pentene	C_7H_{14}				99.2			84.6		
2,4-Dimethyl-2-pentene	C_7H_{14}				100.0			84.6		
3-Ethyl-1-pentene	C_7H_{14}				95.6			81.6		
3-Ethyl-2-pentene	C_7H_{14}				93.7			80.6		
3-Heptene	C_7H_{14}				90.0			79.3		
3-Methyl-1-hexene	C_7H_{14}				82.2			71.5		
3-Methyl-2-hexene	C_7H_{14}				92.0			79.8		
3-Methyl-3-hexene	C_7H_{14}				96.2			81.4		
3,3-Dimethyl-1-pentene	C_7H_{14}				103.5			86.1		
3,4-Dimethyl-1-pentene	C_7H_{14}				98.9			80.9		
3,4-Dimethyl-2-pentene	C_7H_{14}				96.0			82.2		
4-Methyl-1-hexene	C_7H_{14}				86.4			74.0		
4-Methyl-2-hexene	C_7H_{14}				97.6			83.0		
4,4-Dimethyl-1-pentene	C_7H_{14}				104.4			85.4		
4,4-Dimethyl-2-pentene	C_7H_{14}				105.3			90.6		
5-Methyl-1-hexene	C_7H_{14}				75.5			65.0		
5-Methyl-2-hexene	C_7H_{14}				94.3			81.2		
cis-2-Heptene	C_7H_{14}	44.0			73.4			68.8		
4-Vinyl-1-Cyclohexene	C_8H_{12}	32.0								
Vinyl cyclohexane	C_8H_{14}	38.0								
1-Octene	C_8H_{16}	41	40.0	41.0	28.7		39.0	34.7		
2-Methyl-1-heptene	C_8H_{16}				70.2			66.3		
2-Methyl-2-heptene	C_8H_{16}				79.8			73.1		
2-Methyl-3-heptene	C_8H_{16}				94.6			80.6		
2-Octene	C_8H_{16}	43.0								
2,2-Dimethyl-3-Hexene	C_8H_{16}				106.0			88.5		
2,2,4-Trimethyl-1-pentene	C_8H_{16}	10.0	-3.0	11.0						
2,3-Dimethyl-1-Hexene	C_8H_{16}				96.3			83.6		
2,3-Dimethyl-2-Hexene	C_8H_{16}				93.1			79.3		
2,3,3-Trimethyl-1-pentene	C_8H_{16}				106.0			85.7		
2,3,4-Trimethyl-2-pentene	C_8H_{16}				96.6			80.9		
2,4,4-Trimethyl-1-pentene	C_8H_{16}				106.0			86.5		
2,4,4-Trimethyl-2-pentene	C_8H_{16}				102.0	89.0		85.6		
2,4,4-Trimethyl-2-pentene	C_8H_{16}				103.5			86.2		
2,5-Dimethyl-2-Hexene	C_8H_{16}				95.2			82.2		

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
2,5-Dimethyl-3-Hexene	C_8H_{16}				101.9			85.4		
3-Ethyl-2-methyl-1-pentene	C_8H_{16}				99.5			85.3		
3-Ethyl-2-methyl-2-pentene	C_8H_{16}				95.6			82.0		
3,4,4-Trimethyl-2-pentene	C_8H_{16}				103.0			86.1		
6-Methyl-1-heptene	C_8H_{16}				63.8			62.6		
6-Methyl-2-heptene	C_8H_{16}				71.3			65.5		
6-Methyl-3-heptene	C_8H_{16}				91.3			82.0		
cis-3-Octene	C_8H_{16}	38.1								
trans-3-Octene	C_8H_{16}	34.0								
1-Nonene	C_9H_{18}	51.0								
2-Methyl-1-octene	C_9H_{18}							83.5		
2,6-Dimethylheptene	C_9H_{18}	51.0						74.8		
3-Methyl-1-octene	C_9H_{18}									
alpha-Pinene	$C_{10}H_{16}$	25.0								
beta-Pinene	$C_{10}H_{16}$	22.0								
1,9-Decadiene	$C_{10}H_{18}$	41.0								
1-Decene	$C_{10}H_{20}$	49.1								
1-Undecene	$C_{11}H_{22}$	65.0								
4,4-Diethyl-1-heptene	$C_{11}H_{22}$				79.8			74.6		
1-Dodecene	$C_{12}H_{24}$	56.8								
1-Tetradecene	$C_{14}H_{28}$	82.7								
7-Butyltridecene	$C_{15}H_{24}$	36.0								
Bisabolene	$C_{15}H_{24}$	32.0								
2,6,7-Trimethyl-2,6-tridecadiene	$C_{16}H_{30}$	24.0								
1-Hexadecene	$C_{16}H_{32}$	86.0								
2,2,6,6,8,8-Hexamethyl-4-methylene-nonane	$C_{16}H_{32}$	5.0								
4-Butyl-4-dodecene	$C_{16}H_{32}$	45.0								
5-Butyl-4-dodecene	$C_{16}H_{32}$	45.0								
3,12-Diethyl-3,11-tetradecadiene	$C_{18}H_{34}$	26.0								
1-Octadecene	$C_{18}H_{36}$	90.0								
7,10-Dimethyl-8-hexadecene	$C_{18}H_{36}$	43.0								
8-Propyl-8-Pentadecene	$C_{18}H_{36}$	45.0								
9-Methyl-9-heptadecene	$C_{18}H_{36}$	66.0								
7-Hexyl-7-pentadecene	$C_{21}H_{42}$	47.0								
10,13-Dimethyl-11-doeicosese	$C_{24}H_{48}$	56.0								

Table B-3. Cetane and Octane Number Data for Alkynes.

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Acetylene	C_2H_2				80.0					
2-Butyne	C_4H_6				85.9			70.2		
2-Pentyne	C_5H_8				98.0					
1-Heptyne	C_7H_{12}	22.0			84.0					
3-Heptyne	C_7H_{12}				40.0					
1-Octyne	C_8H_{14}				50.5			51.5		

Table B-4. Cetane and Octane Number Data for Naphthenes.

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Cyclopentane	C_5H_{10}	6.1			101.6	100.0		84.9		
1,1,2-Trimethylcyclopropane	C_6H_{12}	17.0			109.0			89.0		
Cyclohexane	C_6H_{12}	16.9	13.2	18.0	84.0			77.6		
Methylcyclopentane	C_6H_{12}	17.2			89.3	81.5		81.0		
1,1-Dimethylcyclopentane	C_7H_{14}				92.3			89.3		
1,3-Diemthylcyclopentane	C_7H_{14}	27.0			79.9			72.9		
Cycloheptane	C_7H_{14}				38.9			40.8		
Ethylcyclopentane	C_7H_{14}				67.2	62		61.2		
Methylcyclohexane	C_7H_{14}	20.0	20.0	24.4	73.8			73.8		
1-Ethyl-3-methylcyclopentane	C_8H_{16}	21.9			57.6			59.8		
1,1-Dimethylcyclohexane	C_8H_{16}	21.0			87.3			85.9		
1,1,3-Trimethylcyclopentane	C_8H_{16}	16.0			81.7			83.5		
1,2-Dimethylcyclohexane	C_8H_{16}	23.0			80.9		86.0	78.7		
1,2,4-Trimethylcyclopentane	C_8H_{16}	19.0			89.2			79.5		
1,3-Dimethylcyclohexane	C_8H_{16}	23.0			69.3		77.0	67.6		
1,4-Dimethylcyclohexane	C_8H_{16}				67.7		74.0	64.2		
Cyclooctane	C_8H_{16}	22.3			71.0			51.2		
Ethylcyclohexane	C_8H_{16}	35.8			46.5			40.8		
Isopropylcyclopentane	C_8H_{16}				81.1			76.2		
n-Propylcyclopentane	C_8H_{16}	37.0			31.2			28.1		

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
1-Ethyl-1-methylcyclohexane	C_9H_{18}				68.7			76.7		
1-Ethyl-2-methylcyclohexane	C_9H_{18}				74.0					
1-Ethyl-3-methylcyclohexane	C_9H_{18}				58.0					
1-Ethyl-4-methylcyclohexane	C_9H_{18}				54.0					
1,1,2-Trimethylcyclohexane	C_9H_{18}	29.0			95.7		96.0	87.7		91
1,1,2,4-Tetramethylcyclopentane	C_9H_{18}	34.2			96.2			88.0		
1,1,3-Trimethylcyclohexane	C_9H_{18}	29.3			81.3			82.6		
1,2,3-Trimethylcyclohexane	C_9H_{18}	29.8			84.8			81.8		
1,2,4-Trimethylcyclohexane	C_9H_{18}	30.5			72.9			74.3		
1,3,5-Trimethylcyclohexane	C_9H_{18}	30.5			63.8			63.3		
Ethylcycloheptane	C_9H_{18}				28.0			30.0		
iso-Butylcyclopentane	C_9H_{18}	47.8			81.1			76.2		
iso-Propylcyclohexane	C_9H_{18}	41.1			62.8			61.1		
n-Propylcyclohexane	C_9H_{18}	52.0			17.8			14.0		
cis-Decalin	$C_{10}H_{18}$	39.4	39.4	41.6						
Decalin	$C_{10}H_{18}$	48.0	42.1	48.0						
trans-Decalin	$C_{10}H_{18}$	32.0	31.8	32.0						
1-Methyl-2-propylcyclohexane	$C_{10}H_{20}$	58.0			29.9		49.0	38.8		
1-Methyl-3-propylcyclohexane	$C_{10}H_{20}$				39.0					
1-Methyl-4-propylcyclohexane	$C_{10}H_{20}$				34.0					
iso-Butylcyclohexane	$C_{10}H_{20}$	50.0			33.7			28.9		
n-Butylcyclohexane	$C_{10}H_{20}$	46.5	46.5	48.0						
sec-Butylcyclohexane	$C_{10}H_{20}$	35.1			51.0			55.2		
tert-Butylcyclohexane	$C_{10}H_{20}$	29.9			98.5			89.2		
1-Methyl-2-butylcyclohexane	$C_{11}H_{22}$				39.0					
1-Methyl-3-butylcyclohexane	$C_{11}H_{22}$				34.0					
1-Methyl-4-butylcyclohexane	$C_{11}H_{22}$				28.0					
Bicyclohexyl	$C_{12}H_{20}$	51.0	47.4	53.0						
3-Cyclohexylhexane	$C_{12}H_{24}$	36.0								
n-Propyldecalin	$C_{13}H_{24}$	35.0								
Perhydrophenanthrene	$C_{14}H_{24}$	38.8								
n-Butyldecalin	$C_{14}H_{26}$	31.0								
sec-Butyldecalin	$C_{14}H_{26}$	34.0								
tert-Butyldecalin	$C_{14}H_{26}$	24.0								
1,3,5-Triisopropylcyclohexane	$C_{15}H_{30}$	25.3								

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
2-Methyl-3-cyclohexylnonane	$C_{15}H_{30}$	63.0	56.0	70.0						
n-Octyldecalin	$C_{18}H_{36}$	31.0								
1-Methyl-3-dodecylcyclohexane	$C_{19}H_{38}$	70.0								
2-Cyclohexyltetradecane	$C_{20}H_{40}$	57.0								
2-Methyl-2-cyclohexyanylpentadecane	$C_{22}H_{44}$	45.0								
1,2,4-Trimethyl-5-hexadecylcyclohexane	$C_{25}H_{50}$	42.0								
5-Cyclohexyleicosane	$C_{26}H_{52}$	66.0								

Table B-5. Cetane and Octane Number Data for Aromatics.

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Benzene	C_6H_6	0.0	-10.0	15.0	101.0	98.0		93.0		
Toluene	C_7H_8	-5.0	-5.0	3.0	121.0	114.0	124.0	103.0		107.0
Styrene	C_8H_8				103.0			100.2		
Ethylbenzene	C_8H_{10}	4.0			108.3			97.9		
m-Xylene	C_8H_{10}	-1.0			145.0			102.8		
o-Xylene	C_8H_{10}	8.3			120.0			100.0		
p-Xylene	C_8H_{10}	-4.0			146.0			101.2		
Indene	C_9H_8				102.3			100.7		
2-propenylbenzene	C_9H_{10}				102.1			100.1		
2,3-Dihydroindene	C_9H_{10}				100.3			89.8		
cis-1-propenylbenzene	C_9H_{10}				100.5			91.7		
trans-1-propenylbenzene	C_9H_{10}				100.4			92.1		
1,2,3-Trimethylbenzene	C_9H_{12}	7.6			118.0	100.5		101.1		
1,2,4-Trimethylbenzene	C_9H_{12}	8.9			148.0	101.4		110.6		
1,3,5-Trimethylbenzene	C_9H_{12}	8.0			170.0	106.0		100.6		
Isopropylbenzene	C_9H_{12}	7.0			67.3		102.1	60.5		99.3
m-Ethyltoluene	C_9H_{12}				101.8			100.0		
n-Propylbenzene	C_9H_{12}	16.0			101.5			98.7		
o-Ethyltoluene	C_9H_{12}				100.2			92.1		
p-Ethyltoluene	C_9H_{12}							97.0		

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Naphtalene	$C_{10}H_8$	22.0								
Tetralin	$C_{10}H_{12}$	13.0	8.9	21.3	96.5			81.9		
1-Methyl-2-isopropylbenzene	$C_{10}H_{14}$				100.6			95.2		
1-Methyl-2-propylbenzene	$C_{10}H_{14}$				100.3			92.2		
1-Methyl-3-propylbenzene	$C_{10}H_{14}$				101.8			100.0		
1-Methyl-4-isopropylbenzene	$C_{10}H_{14}$	4.0			101.4			60.5		97.7
1-Methyl-4-propylbenzene	$C_{10}H_{14}$	2.0						100.2		
1,2-Dimethyl-3-Ethylbenzene	$C_{10}H_{14}$				100.4			91.9		
1,2,3,4-Tetramethylbenzene	$C_{10}H_{14}$	17.0			100.5			100.0		
1,2,4,5-Tetramethylbenzene	$C_{10}H_{14}$	1.0								
1,3-Diethylbenzene	$C_{10}H_{14}$	5.0			103.0			97.0		
1,3-Dimethyl-4-Ethylbenzene	$C_{10}H_{14}$				100.6			95.9		
1,3-Dimethyl-5-Ethylbenzene	$C_{10}H_{14}$				102.7			100.2		
1,4-Diethylbenzene	$C_{10}H_{14}$				100.6			95.2		
1,4-Dimethyl-2-Ethylbenzene	$C_{10}H_{14}$				100.6			96.0		
Isobutylbenzene	$C_{10}H_{14}$				101.6			98.0		
n-Butylbenzene	$C_{10}H_{14}$	12.0			100.4			94.5		
sec-Butylbenzene	$C_{10}H_{14}$	6.0			100.7			95.7		
tert-Butylbenzene	$C_{10}H_{14}$	-1.0			103.0			100.8		
1-Methylnaphthalene	$C_{11}H_{10}$	0.0	-4.0	0.0						
2-Methylnaphthalene	$C_{11}H_{10}$	6.0								
Biphenyl	$C_{12}H_{10}$	21.0	12.0	21.0						
2,6-Dimethylnaphthalene	$C_{12}H_{12}$	-7.0								
n-Pentylbenzene	$C_{12}H_{16}$	9.0	8.0	18.0						
m-Diisopropylbenzene	$C_{12}H_{18}$	-7.0	-12.0	-3.0						
n-Hexylbenzene	$C_{12}H_{18}$	26.0								
Dipheylmethane	$C_{13}H_{12}$	11.0								
n-Propyltetralin	$C_{13}H_{18}$	8.0								
n-Heptylbenzene	$C_{13}H_{20}$	35.0	34.0	35.0						
1,2-Diphenylethane	$C_{14}H_{14}$	1.0								
1-Butylnaphthalene	$C_{14}H_{16}$	6.0								
2-(tert-Butyl)-naphthalene	$C_{14}H_{16}$	3.0								
cis-n-Bityltetralin	$C_{14}H_{20}$	18.0								
sec-Butyltetralin	$C_{14}H_{20}$	7.0								
tert-Butyltetralin	$C_{14}H_{20}$	17.0								

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
trans-n-Butyltetralin	$C_{14}H_{20}$	14.0								
2-Phenyloctane	$C_{14}H_{22}$	33.0								
n-Octylbenzene	$C_{14}H_{22}$	32.0	32.0	43.0						
1,3,5-Triisopropylbenzene	$C_{15}H_{24}$	2.8								
n-nonylbenzene	$C_{15}H_{24}$	50.0								
n-Octylxylene	$C_{15}H_{24}$	20.0								
2-Methyl-2-(beta-naphthyl)hexane	$C_{17}H_{22}$	10.0								
2-Phenyl-2-undecene	$C_{15}H_{26}$	23.0								
2-Phenylundecane	$C_{15}H_{28}$	51.0								
2-Octylnaphthalene	$C_{18}H_{24}$	18.0								
4-Methyl-4-(2-naphthyl)heptane	$C_{18}H_{24}$	9.0								
n-Octyltetralin	$C_{18}H_{28}$	18.0								
4-Phenyldodecane	$C_{18}H_{30}$	42.0								
n-Dodecylbenzene	$C_{18}H_{30}$	51.0								
7-Phenyltridecane	$C_{19}H_{32}$	41.0								
3,6-Dimethyl-3-(beta-naphthyl)octane	$C_{20}H_{28}$	18.0								
5-Methyl-5-(beta-naphthyl)nonane	$C_{20}H_{28}$	12.0								
2-Phenyltetradecane	$C_{20}H_{34}$	49.0								
n-Tetradecylbenzene	$C_{20}H_{34}$	72.0								
2-Methyl-2-(beta-naphthyl)decane	$C_{21}H_{30}$	18.0								
3-Ethyl-3-(beta-naphthyl)nonane	$C_{21}H_{30}$	13.0								
2-Methyl-2-phenylpentadecane	$C_{22}H_{38}$	39.0								
2-Methyl-4-isobutyl-4phenylundecane	$C_{22}H_{38}$	18.0								
2-Methyl-2-phenylheptadecane	$C_{24}H_{42}$	39.0								
5-Butyl-5-phenyltetradecane	$C_{24}H_{42}$	58.0								
1,2,4-Trimethyl-5-hexadecylbenzene	$C_{25}H_{44}$	42.0								
5-Phenyleicosane	$C_{26}H_{46}$	39.0								

Table B-6. Cetane and Octane Number Data for Oxygenates Other Than Esters.

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Alcohols										
Methanol	CH_4O	5.0	2.0	5.0	122.0		130.0	93.0		
Ethanol	C_2H_6O	12.0	2.0	12.0	111.0	107.4	121.0	96	90.0	97.0
isoPropanol	C_3H_8O				117.0	112.5	106.0	95.0		99.0
n-Propanol	C_3H_8O	12.0			102.6		104.0	98.0	89.0	
2-Butanol	$C_4H_{10}O$	8.5			108.0	105.0		91.0	87.0	93.0
Isobutanol	$C_4H_{10}O$	8.5			105.1	105.0		90.0		
n-Butanol	$C_4H_{10}O$	12.0	3.7	17.0	96.0		98.0	78.0	85.0	90.0
tert-Butanol	$C_4H_{10}O$	5.6			107.0	105.0		94.0	89.0	104.0
1-Pentanol	$C_5H_{12}O$	18.2	18.2	20.0	78.0			74.0		
2-Methyl-2-butanol	$C_5H_{12}O$				97.0					
3-Methyl-1-butanol	$C_5H_{12}O$				113.0			88.0		
Isopentanol	$C_5H_{12}O$	18.4			98.8	94.0		84.0		
1-Hexanol	$C_6H_{14}O$	23.3			69.3			78.0		
1-Heptanol	$C_7H_{16}O$	29.5								
1-Octanol	$C_8H_{18}O$	39.1	33.7	39.1						
2-Ethyl-1-hexanol	$C_8H_{18}O$	23.5								
3-Octanol	$C_8H_{18}O$	25.1								
1-Nonaol	$C_9H_{20}O$	46.2								
2-Nonanol	$C_9H_{20}O$	39.6								
beta-Citromellol	$C_{10}H_{20}O$	25.6								
1-Decanol	$C_{10}H_{22}O$	50.3								
3,7-Dimethyl-1-octanol	$C_{10}H_{22}O$	29.3								
1-Undecanol	$C_{11}H_{24}O$	53.2								
1-Dodecanol	$C_{12}H_{26}O$	63.6								
1-Tetradecanol	$C_{14}H_{30}O$	80.8	51.0	80.8						
Palmitoleyl Alcohol	$C_{16}H_{32}O$	46.0								
1-Hexadecanol	$C_{16}H_{34}O$	68.0								
Linolenyl Alcohol	$C_{18}H_{32}O$	41.0								
Linoleyl Alcohol	$C_{18}H_{34}O$	44.0								
Oleyl Alcohol	$C_{18}H_{36}O$	51.0								
1-Octadecanol	$C_{18}H_{38}O$	81.0								

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Ethers										
Dimethyl Ether	C_2H_6O	78.0	55.0	78.0				78		
2-Methoxyethanol	$C_3H_8O_2$	13.0								
Dimethoxymethane	$C_3H_8O_2$	62.9	29.0	55.0						
1-Methoxy-2-Propanol	$C_4H_{10}O$	19.0								
Diethylether	$C_4H_{10}O$	140.0	140.0	160.0						
1-Methoxy-2-propanol	$C_4H_{10}O_2$	57.2								
1,2-Dimethoxyethane	$C_4H_{10}O_2$	94.0	90.0	98.0						
Methyl-tertbutyl ether	$C_5H_{12}O$				118.0	115.2		101.0	97.2	
Poly(Oxyymethylene) dimethyl ethers	$C_5H_{12}O_3$	63.0								
2-Ethoxyethyl acetate	$C_6H_{12}O_3$	40.0								
Ethyl-tert-butyl Ether	$C_6H_{14}O$	86.4			118.0			102.0		
tert-Amyl methyl ether	$C_6H_{14}O$				109.0		115.0	99.0	98	
1,1-Diethoxyethane	$C_6H_{14}O_2$	40.0								
2-Butoxyethanol	$C_6H_{14}O_2$	59.6								
2-Methoxyethyl ether	$C_6H_{14}O_3$	170.0	112.0	170.0						
Poly(oxymethylene)dimethyl ethers	$C_6H_{14}O_3$	90.0								
2,4,7,9-Tetraoxadecane	$C_6H_{14}O_4$	64.5								
Hexymethyl ether	$C_7H_{16}O$	99.8								
1-Butoxy-2-propanol	$C_7H_{16}O_2$	36.1								
Dipropylene glycol monomethyl ether	$C_7H_{16}O_3$	43.9	43.9	52.0						
Triethylene glycol monomethyl ether	$C_7H_{16}O_4$	80.7								
Dibutyl ether	$C_8H_{18}O$	95.5	91.0	100.0						
Diisobutyl ether	$C_8H_{18}O$	59.7								
Dimethoxyhexane	$C_8H_{18}O_2$	88.0								
Diethoxybutane	$C_8H_{18}O_2$	97.0								
2-Ethoxyethyl ether	$C_8H_{18}O_3$	151.0	113.0	151.0						
Triethylene glycol dimethyl ether	$C_8H_{18}O_4$	120.0								
Dibutoxymethane	$C_9H_{20}O_2$	74.0								
Rose oxide	$C_{10}H_8O$	30.0								
DIisoamyl ether	$C_{10}H_{22}O$	96.3								
Dipentyl ether	$C_{10}H_{22}O$	135.0	111.0	135.0						
Tripropylene glycol monoethyl ether	$C_{10}H_{22}O_4$	63.0	63.0	81.3						

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Furans and Cyclic Ethers										
Furan	C_4H_4O	7.0								
2-3-Dihydrofuran	C_4H_6O	20.0								
2,5-Dihydrofuran	C_4H_6O	15.6								
Tetrahydrofuran	C_4H_6O	21.9								
2-Methylfuran	C_5H_6O	8.9			103.0			86.0		
2-Furfuryl Alcohol	$C_5H_6O_2$	10.8								
2-Methyltetrahydrofuran	$C_5H_{10}O$	22.0			86.0			73.0		
2-Tetrahydrofurfuryl Alcohol	$C_5H_{10}O_2$	17.9								
2,5-Dimethylfuran	C_6H_6O	10.9			119.0					
2-Ethylfuran	C_6H_8O	10.2								
2-Ethyltetrahydrofuran	$C_6H_{12}O$	28.1								
2-Butylfuran	$C_8H_{12}O$	13.1								
2-Butyltetrahydrofuran	$C_8H_{16}O$	45.5								
Aldehydes and Ketones										
Acetone	C_3H_6O				110.0					
Butanal	C_4H_8O	41.1								
Isobutyraldehyde	C_4H_8O	21.1								
Methyl Ethyl Ketone	C_4H_8O				118.0					
Pentanal	$C_5H_{10}O$	62.2								
3-Pentanone	$C_5H_{10}O$	19.5								
Cyclohexanone	$C_6H_{10}O$	10.0		10.4				100		
4-Methyl-3-penten-2-one	$C_6H_{10}O$							99		
Hexanal	$C_6H_{12}O$	75.2								
4-Methyl-2-penanone	$C_6H_{12}O$	12.6								
Cycloheptanone	$C_7H_{12}O$	22.5								
2-Heptanone	$C_7H_{14}O$	30.0								
3-cyclohexene-1-carboxaldehyde	$C_8H_{12}O$	28.1								
Octanal	$C_8H_{16}O$	80.5		102.5				91		
2-Octanone	$C_8H_{16}O$	36.6								
3-Octanone	$C_8H_{16}O$	36.0						80		
2-Nonanobe	$C_9H_{18}O$	46.1								
3,3,5-Trimethylcyclohexanone	$C_{10}H_{18}O$	11.9								
Menthone	$C_{10}H_{18}O$	20.6								

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Carboxylic Acids										
Decanoic Acid	$C_{10}H_{20}O_2$	48								
Linolenic Acid	$C_{18}H_{30}O_2$	29								
cis,cis-9,12-Octadecadienoic Acid	$C_{18}H_{32}O_2$	31								
cis-9-Octadecenoic Acid	$C_{18}H_{34}O_2$	46								
Octadecanoic Acid	$C_{18}H_{36}O_2$	62								

Table B-7. Cetane and Octane Number Data for Esters.

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Ester of Carboxylic Acids										
Methyl Butanoate	$C_5H_{10}O_2$	6.4	6.0	6.4						
Methyl Pentanoate	$C_6H_{12}O_2$	13.3			105.0			105.0		
Methyl Sorbate	$C7H_{10}O_2$	6.0								
Ethyl Levulinate	$C_7H_{12}O_3$	<5			110.0			102.0		
Ethyl Pentanoate	$C_7H_{14}O_2$	18.6								
Methyl Capronate	$C_7H_{14}O_2$	18.0	18.0	23.9						
Butyl Butanoate	$C_8H_{16}O_2$	17.8								
Methyl Heptanoate	$C_8H_{16}O_2$	34.2								
n-Hexyl Acetate	$C_8H_{16}O_2$	33.8								
Propyl Pentanoate	$C_8H_{16}O_2$	20.7								
Butyl Levulinate	$C_9H_{16}O_2$	14.4			98.0			96.0		
Butyl Pentanoate	$C_9H_{18}O_2$	23.5								
Methyl Octanoate	$C_9H_{18}O_2$	33.6	33.6	34.0						
Pentyl Pentanoate	$C_{10}H_{20}O_2$	28.8	27.6	28.8						
Methyl-9-decenoate	$C_{11}H_{20}O_2$	38.3								
Methyl Decanoate	$C_{11}H_{22}O_2$	47.9	47.9	54.1						
Butyl Octanoate	$C_{12}H_{14}O_2$	39.6								
Decyl Acetate	$C_{12}H_{14}O_2$	62.0								
Ethyl Decanoate	$C_{12}H_{14}O_2$	51.2	51.2	60.0						
Isopropyle Decanoate	$C_{13}H_{26}O_2$	46.6								
Methyl Laurate	$C_{13}H_{26}O_2$	61.2	54.0	70.0						
Octyl Valerate	$C_{13}H_{26}O_2$	49.0								

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Propyl Decanoate	$C_{13}H_{26}O_2$	52.9	52.9	64.0						
Butyl Decanoate	$C_{13}H_{28}O_2$	54.6		63.0						
Dodecyl Acetate	$C_{13}H_{28}O_2$	77.0								
Ethyl Laurate	$C_{13}H_{28}O_2$	73.0								
Decyl Valerate	$C_{15}H_{30}O_2$	61.0								
Methyl Myristate	$C_{15}H_{30}O_2$	73.5	66.2	73.5						
Propyl Laurate	$C_{15}H_{30}O_2$	71.0								
Butyl Laurate	$C_{16}H_{32}O_2$	73.0								
Ethyl Myristate	$C_{16}H_{32}O_2$	66.9	66.9	72.0						
Hexyl Caprate	$C_{16}H_{32}O_2$	64.0								
Tetradecyl Acetate	$C_{16}H_{32}O_2$	81.0								
Methyl cis-2-Hexadecenate	$C_{17}H_{32}O_2$	51.0								
Dodecyl Valerate	$C_{17}H_{34}O_2$	67.0								
Methyl Hexadecanate	$C_{17}H_{34}O_2$	85.9	74.3	91.0						
Propyl Myristate	$C_{17}H_{34}O_2$	71.0								
Butyl Myristate	$C_{18}H_{36}O_2$	69.4	69.4	73.0						
Ethyl Hexadecanate	$C_{18}H_{36}O_2$	93.1	80.0	93.1						
Hexadecyl Acetate	$C_{18}H_{36}O_2$	86.0								
Hexyl Laurate	$C_{18}H_{36}O_2$	74.0								
Methyl alpha-Linolenate	$C_{19}H_{32}O_2$	22.7								
Methyl gamma-Linolenate	$C_{19}H_{32}O_2$	29.2								
Methyl Linolenate	$C_{19}H_{32}O_2$	45.9	23.0	45.9						
Methyl cis,cis-9,12-Octadecadienate	$C_{19}H_{34}O_2$	38.2	38.0	43.9						
Methyl Linolelaidate	$C_{19}H_{34}O_2$	43.0								
Methyl cis-9-Octadecenate	$C_{19}H_{36}O_2$	59.3	55.0	80.0						
Methyl Elaidate	$C_{19}H_{36}O_2$	57.2								
Methyl Petroselinate	$C_{19}H_{36}O_2$	58.6								
Methyl Ricnoleate	$C_{19}H_{36}O_2$	37.4								
Isopropyl Hexadecanate	$C_{19}H_{38}O_2$	92.6								
Methyl Octadecanate	$C_{19}H_{38}O_2$	86.9	75.6	101.0						
Propyl Hexadecanate	$C_{19}H_{38}O_2$	85.0	83.0	85.0						
Tetradecyl Valerate	$C_{19}H_{38}O_2$	68.0								
Methyl Asclepate	$C_{19}H_{38}O_3$	53.9								
Ethyl Linolenate	$C_{20}H_{34}O_2$	27.0								
Ethyl cis,cis-9,12-Octadecadienate	$C_{20}H_{36}O_2$	39.6	37.0	44.4						

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Ethyl cis-9-Octadecenate	$C_{20}H_{38}O_2$	67.8	54.0	72.0						
2-Butyl Hexadecanate	$C_{20}H_{40}O_2$	84.8								
Butyl Hexadecanate	$C_{20}H_{40}O_2$	91.9	87.0	92.0						
Decyl Caprate	$C_{20}H_{40}O_2$	81.0								
Ethyl Octadecanate	$C_{20}H_{40}O_2$	97.7	76.8	98.0						
Hexyl Myristate	$C_{20}H_{40}O_2$	72.0								
Isobutyl Hexadecanate	$C_{20}H_{40}O_2$	83.6								
Octadecyl Acetate	$C_{20}H_{40}O_2$	90.0								
Octyl Laurate	$C_{20}H_{40}O_2$	84.0								
Methyl 5(Z) 8(Z) 11(Z) 14(Z) Eicosatetraenoate	$C_{21}H_{34}O_2$	29.6								
Propyl Linolenate	$C_{21}H_{36}O_2$	27.0								
Propyl cis,cis-9,12-Octadecadienate	$C_{21}H_{38}O_2$	44.0	41.0	44.0						
Isopropyl cis-9-Octadecenate	$C_{21}H_{40}O_2$	86.6								
Methyl Gondoate	$C_{21}H_{40}O_2$	73.2								
Propyl cis-9-Octadecenate	$C_{21}H_{40}O_2$	58.8								
Propyl cis-9-Octadecenate	$C_{21}H_{40}O_2$	58.8	56.0	72.0						
Hexadecyl Valerate	$C_{21}H_{42}O_2$	70.0								
Isopropyl Octadecanate	$C_{21}H_{42}O_2$	96.5								
Methyl Arachidate	$C_{21}H_{42}O_2$	100.0								
Propyl Octadecanate	$C_{21}H_{42}O_2$	90.9								
Butyl Linolenate	$C_{22}H_{38}O_2$	29.0								
Butyl cis,cis-9,12-Octadecadienate	$C_{22}H_{40}O_2$	53.5								
2-Butyl cis-9-Octadecenate	$C_{22}H_{42}O_2$	71.9								
Butyl cis-9-Octadecenate	$C_{22}H_{42}O_2$	61.6	60.0	102.0						
Isobutyl cis-9-Octadecenate	$C_{22}H_{42}O_2$	59.6								
2-Butyl Octadecanate	$C_{22}H_{44}O_2$	97.5								
Butyl Octadecanate	$C_{22}H_{44}O_2$	92.5								
Decyl Laurate	$C_{22}H_{44}O_2$	84.0								
Hexyl Hexanoate	$C_{22}H_{44}O_2$	87.0								
Isobutyl Octadecanate	$C_{22}H_{44}O_2$	99.3								
Octyl Myristate	$C_{22}H_{44}O_2$	71.0								
Methyl 4(Z) 7(Z) 10(Z) 13(Z0 16(Z) 19(Z) Docosahexaenoate	$C_{23}H_{34}O_2$	24.4								
Methyl Erucate	$C_{23}H_{44}O_2$	74.2								

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Hexyl cis-9-Octadecenate	$C_{23}H_{46}O_2$	102.0								
2-Ethylhexyl Hexadecanate	$C_{24}H_{48}O_2$	98.2	98.0	107.0						
Decyl Myristate	$C_{24}H_{48}O_2$	72.0								
Dodecyl Laurate	$C_{24}H_{48}O_2$	85.0								
2-Ethylhexyl cis-9-Octadecenate	$C_{26}H_{50}O_2$	88.2								
Octyl cis-9-Octadecenate	$C_{26}H_{50}O_2$	131.0								
2-Ethylhexyl Octadecanate	$C_{26}H_{52}O_2$	115.5								
Decyl Hexadecanoate	$C_{26}H_{52}O_2$	91.0								
Dodecyl Myristate	$C_{26}H_{52}O_2$	74.0								
Decyl cis-9-Octadecenate	$C_{28}H_{54}O_2$	134.0								
Hexadecyl Laurate	$C_{28}H_{56}O_2$	88.0								
Esters of Dicarboxylic Acids										
Dimethyl Malonate	$C_5H_8O_4$	15.0								
Diethyl Oxalate	$C_6H_{10}O_4$	21.0								
Diethyl Succinate	$C_8H_{14}O_4$	21.0								
Dimethyl Adipate	$C_8H_{14}O_4$	5.0								
Glycerol Tracetate	$C_9H_{14}O_4$	<5								
Dimethyl Phthalate	$C_{10}H_{10}O_4$	19.0								
Diethyl Adipate	$C_{10}H_{18}O_4$	15.0								
Dibutyl Malonate	$C_{11}H_{20}O_4$	21.0								
Dimethyl Azelate	$C_{11}H_{20}O_4$	24.0								
Dibutyl Butanedioate	$C_{12}H_{14}O_4$	21.0								
Dibutyl Fumarate	$C_{12}H_{20}O_4$	23.0								
Dibutyl Maleate	$C_{12}H_{20}O_4$	29.0	28.0							
Dibutyl Succinate	$C_{12}H_{22}O_4$	13.0								
Diethyl Azelate	$C_{13}H_{24}O_4$	47.0								
Dibutyl Adipate	$C_{14}H_{26}O_4$	81.0								
Diethyl Sebacate	$C_{14}H_{26}O_4$	47.0								
Dibutyl Phthalate	$C_{16}H_{22}O_4$	38.0								
Dibutyl Azalate	$C_{17}H_{32}O_4$	83.0								
Dihexyl Phthalate	$C_{20}H_{30}O_4$	48.0								
Dihexal Azalate	$C_{21}H_{40}O_4$	99.0								
Dioctyl Adipate	$C_{22}H_{42}O_4$	89.0								
Dioxtyl Sebacate	$C_{26}H_{50}O_4$	70.0								

Compound Dame	Formula	Cetane Number			Research Octane Number			Motor Octane Number		
		Nominal	Minimum	Maximum	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Triglycerides										
Tribyrin	$C_{15}H_{26}O_6$	-5.0								
Trilaurin	$C_{39}H_{74}O_6$	100.0								
Trimyristin	$C_{45}H_{86}O_6$	100.0								
Tripalmitin	$C_{51}H_{98}O_6$	89.0								
Trilonlenin	$C_{57}H_{92}O_6$	23.0								
Trilinolein	$C_{57}H_{98}O_6$	32.0								
Triolein	$C_{57}H_{104}O_6$	45.0								
Tristearin	$C_{57}H_{110}O_6$	85.0								

APPENDIX C
PARAMETERS FOR THE ARTIFICIAL NEURAL NETWORK

C.1 INTRODUCTION

This appendix contains tabulations of the parameters for the ANN representation of CN and ONs. Figure B.1 is a graphical representation of neural network structure. The input parameters are the number of each type of functional group comprising the molecule. Input parameters for this neural network are not scaled, which differs from the usual practice. Input nodes are labeled with the appropriate functional group.

This ANN for cetane and octane number contains eight hidden nodes. Four hidden nodes are connected to the output node for CN. These nodes are labeled CN-1 through CN-4. The remaining four hidden nodes are connected to the output nodes for RON and MON. These nodes are labeled RON-1, RON-2, MON-1, and MON-2. The ANN has three output nodes representing CN, RON, and MON. Output values are defined on the interval [0,1], and they must be scaled to obtain the final values of CN, RON, and MON.

C.2 PARAMETERS FOR HIDDEN NODE INPUTS

The input to a hidden node is a linear combination of the number of functional groups in the molecule.

$$x_k = G_{o,k} + \sum_{j \in groups} G_{j,k} \cdot n_j \ , \qquad \text{(C-1)}$$

where x_k is the input to hidden node k, $G_{o,k}$ is a constant for hidden node k, $G_{j,k}$ is the contribution of group j to hidden node k, and n_j is the number of j groups in the molecule. Values of $G_{o,k}$ and $G_{j,k}$ are given in Table C-1.

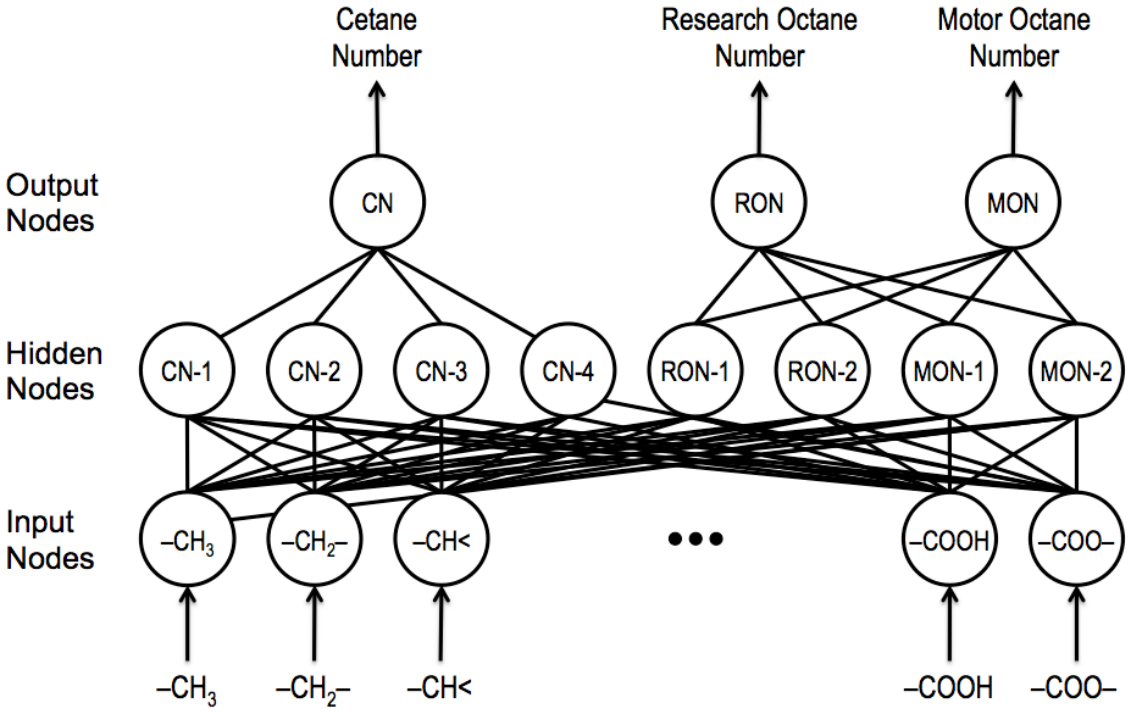

Fig. C-1. Structure of the artificial neural network representing cetane and octane numbers.

Table C-1. Group contribution parameters for hidden nodes ($G_{o,k}$ and $G_{j,k}$).

Group	Hidden Node Label							
	CN-1	CN-2	CN-3	CN-4	RON-1	RON-2	MON-1	MON-2
Constant	0.013	1.372	1.607	4.583	81.133	-0.051	4.467	0.134
Paraffins								
–CH_3	-0.011	-0.687	-0.031	-0.620	-29.490	-1.565	-0.837	8.709
–CH_2–	0.054	1.146	0.058	-0.806	-8.844	-0.839	-0.647	-4.699
–CH< (C# = 2)	-0.152	5.323	-0.386	-0.149	32.432	0.312	0.583	-0.284
–CH< (C# = 3)	-0.221	4.483	-0.374	0.972	23.796	0.001	0.965	1.893
–CH< (C# ≥ 4)	-0.853	-1.738	-1.400	-0.412	1.808	0.000	0.989	5.263
>C< (C# = 2)	0.118	5.753	0.787	0.160	64.174	-4.161	2.586	-9.351
>C< (C# ≥ 3)	0.184	-1.593	1.032	-0.160	47.887	-3.997	3.441	6.040
Olefins / Alkynes								
=CH_2	-0.130	2.666	-0.201	-0.069	0.133	-3.970	-1.047	-13.767
=CH– (C# = 2)	-0.033	-0.013	0.037	-0.105	3.987	-4.492	0.711	-20.790
=CH– (C# = 3)	-0.006	2.918	0.254	-0.918	0.281	-3.021	-0.339	-13.711
=CH– (C# ≥ 4)	-0.135	0.000	-0.149	-0.147	0.015	-1.522	1.249	3.284
>C= (C# = 2)	-1.272	6.285	-2.089	-1.668	-17.971	-1.371	3.042	-27.211
>C= (C# ≥ 3)	-0.673	4.277	-0.800	0.108	11.631	0.071	1.257	-24.226
≡CH	-0.136	0.000	0.000	0.000	-50.914	-5.785	0.794	-0.001
≡C–	-0.136	0.000	0.000	0.000	2.085	0.419	-0.238	-0.043
Naphthenes								
–CH_2–	-0.237	0.017	-0.497	-1.361	-11.993	-0.627	-0.509	0.120
–CH<	0.193	1.540	0.601	0.040	4.251	-6.175	0.376	-9.293
–CH< (ortho)	-1.317	1.656	-2.280	-0.166	18.516	0.747	0.862	-8.680
>C<	-0.300	1.546	-0.539	-0.775	37.083	-1.006	1.800	-3.705
=CH–	0.211	0.610	0.407	-0.033	0.000	-0.063	4.515	0.001
>C=	-1.219	5.213	-1.461	0.058	0.000	0.000	0.000	0.000
Aromatics								
=CH–	-0.236	0.719	-0.259	-0.840	-19.935	-1.116	0.791	0.331
=C< (ring/ring)	-2.165	-0.147	-1.571	0.120	0.000	-0.695	-0.036	-0.264
=C<	0.415	-0.749	0.175	4.568	0.404	4.497	7.226	0.100
=C< (ortho)	-1.407	1.653	-3.857	2.142	0.402	-4.770	-1.109	-0.047
=C< (meta)	0.406	0.596	2.047	0.786	0.403	-0.284	0.551	0.000
=C< (para)	-0.321	1.318	-0.152	0.225	0.403	-1.059	2.492	-0.023

Group	Hidden Node Label							
	CN-1	CN-2	CN-3	CN-4	RON-1	RON-2	MON-1	MON-2
Oxygenates								
–OH	-0.494	2.640	-0.566	0.011	-23.999	-1.237	1.134	-17.370
=O (aldehyde)	0.711	-0.039	5.804	0.089	8.281	0.983	1.508	45.838
=O (ketone)	-1.394	-3.374	-3.878	1.969	8.277	0.982	1.503	45.841
–O– (1^{st})	0.969	3.270	1.793	1.399	-9.654	-8.132	-0.281	6.334
–O– (2^{nd}+)	0.265	-1.664	0.166	7.562	0.000	0.761	-0.372	0.016
–O– (ring)	0.026	0.707	-0.063	0.249	0.126	8.073	-2.335	-0.074
–O– (aromatic)	-0.329	8.043	-0.481	-1.099	-0.010	-3.510	-1.137	0.000
–COOH	-0.550	0.077	-0.554	-5.863	0.000	0.000	0.000	0.000
–COO– (1^{st})	-0.435	3.414	-0.801	3.022	0.596	2.303	9.909	17.222
–COO– (2^{nd})	-0.024	5.173	-0.048	1.345	0.000	0.000	0.000	0.000
–COO– (3^{rd})	-0.736	-0.059	-0.296	1.247	0.000	0.000	0.000	0.000

In this study, some functional groups account for position on the carbon chain or ring. Other functional groups depend on the number of functional groups in the molecule. The symbol C# refers to carbon number and it designated the distance the group is from a terminal carbon or oxygen atom on the longest chain containing the group. C# = 1 corresponds to a terminal carbon atom.

Example 1 – 2,4,8-trimethylnonane
The methyl group on the eighth carbon atom is located adjacent to a terminal carbon atom, so it is attached to a C# = 2 –CH< group. The groups in this molecule are

5 $–CH_3$
2 –CH< (C# = 2)
1 –CH< (C# ≥ 4)
4 $–CH_2–$

Example 2 – 2,7-dimethyl-1-octanol
Because oxygen in the alcohol group is counted when determine C#, the methylene group adjacent to the alcohol group has C# = 2. The methyl group attached to the second carbon atom in the IUPAC notation has C# = 3 according to the definitions of the groups in Table B.1. The seventh carbon atom in the octane chain is adjacent to a terminal carbon atom, so C# = 2. The groups in this molecule are

3 $–CH_3$
1 –CH< (C# = 2)
1 –CH< (C# = 3)
5 $–CH_2–$
1 –OH

For ringed compounds with multiple side groups, one side group is picked as a reference position. Choice of the reference position is not arbitrary, and it does not correspond to the IUPAC nomenclature. The group for the ring group for the reference position is the same as for a monosubstituted ring. The carbon

atoms with other side groups attached are designated *ortho*-, *meta*-, or *para*- depending on position relative to the reference position. The groups for aromatic compounds includes a group designated =C< (ring/ring). This group corresponds to the carbon atom at the juncture of two aromatic rings found in compounds such as a naphthalene or anthracene. This group does not apply to aromatic carbon atoms connect by an aliphatic bond such as the bond the joins the two phenyl groups in a biphenyl molecule.

Example 3 – 1,2,4-trimethylbenzene
The carbon atom in the second position according to IUPAC notation is the reference position because it minimizes the distances from the other side group to the reference position. The group for a monosubstituted aromatic compound is used for the reference carbon atom. The first carbon according to IUPAC notation is in the *otho* position relative to the reference position, and the fourth carbon is in the *meta* position relative to the reference position. The groups in this molecule are

3 $-CH_3$
1 =C< (aromatic)
1 =C< (ortho) (aromatic)
1 =C< (meta) (aromatic)
3 =CH– (aromatic)

Example 4 – 1,2,4-trimethylcyclohexane
As in Example 3, the carbon atom in the second position according to IUPAC notation is the reference position because it minimizes the distances from the other side group to the reference position. 1he carbon atom in the first position is in the *ortho* position relative to the reference position. No special group are defined for carbon atoms in the *meta* and *para* positions, so they are represent by the same group that is used for the reference position or a monosubstituted ring. The groups in this molecule are

3 $-CH_3$
2 –CH< (ring)
1 –CH< (ortho) (ring)
3 $-CH_2-$ (ring)

Example 5 – naphthalene
The naphthalene molecule of to integrated aromatic rings containing two =C< (ring/ring) groups. The groups in this molecule are

8 =CH– (aromatic)
2 =C< (ring/ring) (aromatic)

Example 6 – biphenyl
For this group contribution method, biphenyl should be treated as a benzene molecule with a phenyl side group. The carbon atoms that are joined in the biphenyl molecule should be treated as an aromatic carbon with a side group. The groups in this molecule are

10 =CH– (aromatic)
2 =C< (aromatic)

Two groups are used to represent an aldehyde or a ketone in this group contribution method. For an aldehyde, the two groups are =CH– and =O (aldehyde). For ketones, the two groups are =C< and =O (ketone). The functional groups for ethers and esters depend distinguish between molecules with a single group an molecules with multiple groups. For an ether containing only one -O- group, the -O- (1^{st}) group in Table B.1 is used. For an ether containing more than on -O- group, the -O- (1^{st}) is used for one of

the -O- groups and the -O- (2^{nd}) is used for all additional groups. The same method is used to represent fuels containing multiple ester groups.

Example 7 – 2,6-dimethyl-4-heptanone (diisobutyl ketone)
In this molecule, the methyl groups are located adjacent to the terminal carbon atoms, so the carbon atoms to which they are attached have C# = 2. The ketone group is the four carbon atom counting from either terminal carbon, so the carbon group for the ketone is =C< (C# ≥ 3). The groups in this molecule are

4 –CH_3
2 –CH< (C# = 2)
2 –CH_2–
1 =C< (C# ≥ 3)
1 =O (ketone)

Example 8 – diethyl oxalate
When a molecule contains more than one ester group, each ester group is represent by a different function group in this group contribution correlation. Diethyl oxalate contains two ester groups, which are represented by –COO– (1^{st}) and –COO– (2^{nd}). The groups in this molecule are

2 –CH_3
2 –CH_2–
1 –COO– (1^{st})
1 –COO– (2^{nd})

C.3 PARAMETERS FOR OUTPUT NODE INPUTS

The parameters for the output nodes represent the connections between the hidden nodes and the output nodes. The input to an output node is a linear combination of outputs from the hidden nodes.

$$y_k = H_{o,k} + \sum_{j \in hidden\,nodes} H_{j,k} \cdot O_j \ , \qquad \text{(C-2)}$$

where y_k is the input to output node k, $H_{o,k}$ is a constant for output node k, $H_{j,k}$ is a parameter for the output from hidden node j to output node k, O_j is the output from hidden node j. O_j is defined on the interval [0,1]. The parameters for determining the input to the output nodes are given in Table C-2.

Table C-2. Parameters for output node inputs.

Hidden Node Label	Output Node Label		
	CN	RON	MON
Constant	-4.454	-2.152	-2.035
CN-1	4.553	0.000	0.000
CN-2	3.430	0.000	0.000
CN-3	-2.314	0.000	0.000
CN-4	-0.507	0.000	0.000
RON-1	0.000	0.153	0.076
RON-2	0.000	4.375	0.875
MON-1	0.000	2.295	2.029
MON-2	0.000	0.025	0.152

C.4. SCALING FACTORS FOR OUTPUT PARAMETERS

This ANN uses the logistic function for the nodes. The range of the logistic function is defined on the interval (0,1), so the output variables must be scaled to obtain cetane and octane numbers. Scaling is based on a simple linear transformation of variables. The predicted values of cetane and octane number are given by the following equation.

$$Z_i = LB_i + \left(UB_i - LB_i\right) \cdot z_i \quad , \tag{C-3}$$

where Z_i for property i, LB_i is the lower bound for i, UB_i is the upper bound for i, and z_i is the output from output node i. Property i is either CN, RON, or MON. Values for the lower and upper bounds are given in Table C-3.

Table C-3. Upper and lower bounds for output variables.

Parameter	Output Node Label		
	CN	RON	MON
Lower Bound	-50	-50	-50
Upper Bound	250	225	225

APPENDIX D
PARAMETERS FOR A CONVENTIONAL GROUP CONTRIBUTION METHODS

D.1. INTRODUCTION

To evaluate the advantages of using an ANN-based group contribution method, conventional group contribution correlations were developed for CN and ONs. In a conventional group contribution method, structure is characterized by the sum of the group contributions.

$$x_k = \sum_{j \in groups} G_{j,k} \cdot n_j \ , \tag{D-1}$$

where x_k is the sum of the group contribution factors for property k, $G_{j,k}$ is the contribution of group j to the input to property k, and n_j is the number of j groups in the molecule. Property k is CN, RON, or MON. Property k is assumed to be a function of x_k. In this study, two functions were considered. The first function was a polynomial, and the second was a bounded function.

D.2. POLYNOMIAL FUNCTION

The function relating x_k to property k is assumed to be a fourth-order polynomial.

$$Z_k = a_k + b_k \cdot x_k + c_k \cdot x_k^2 + d_k \cdot x_k^3 + e_k \cdot x_k^4 \ , \tag{D-2}$$

where Z_k is property k and a_k, b_k, c_k, d_k, and e_k are empirical constants. The functional groups given in Table C-1 of Appendix C were also used for the conventional group contribution correlation. The parameters $G_{j,k}$ in Eq. (D-1) and the constants in Eq. (D-2) were determined from a least-squares fit to the data in Appendix B. The calculations were implemented in Excel, and the Solver add-in with the multistart option was used to minimize the sum-of-squared errors. The resulting correlation is very similar to Albahri's group contribution method.[1] The group contribution factors are given in Table D-1 and the constants for the polynomial equation are given in Table D-2. Errors in this correlation are summarized in Table D-3.

Table D-1. Group factors for conventional a group contribution correlation with a polynomial function.

Group	Property			Group	Property		
	CN	RON	MON		CN	RON	MON
Paraffins							
–CH_3	0.529	1.566	0.272	–CH< (C# ≥ 4)	-2.598	-1.036	-0.361
–CH_2–	0.637	1.092	0.180	>C< (C# = 2)	-1.558	-4.365	-0.803
–CH< (C# = 2)	0.080	-1.257	-0.217	>C< (C# ≥ 3)	-2.731	-4.789	-1.149
–CH< (C# = 3)	-1.522	-1.598	-0.302				
Olefins / Alkynes							
=CH_2	-0.107	1.579	0.340	>C= (C# = 2)	-2.057	-3.278	-0.601
=CH– (C# = 2)	-0.086	-0.965	-0.168	>C= (C# ≥ 3)	-4.068	-6.832	-0.144
=CH– (C# = 3)	0.413	0.433	0.164	≡CH	-0.353	-1.122	-0.227
=CH– (C# ≥ 4)	-0.787	-1.923	-0.360	≡C–	-0.353	0.244	0.188

Group	Property			Group	Property		
	CN	RON	MON		CN	RON	MON
Naphthenes							
–CH_2–	0.358	0.846	0.163	>C<	-0.975	-2.670	-0.637
–CH<	-0.443	-0.456	-0.074	=CH–	-0.237	-0.672	-0.062
–CH< (ortho)	-0.673	-1.264	-0.226	>C=	-0.099	-0.354	-0.346
Aromatics							
=CH–	-0.113	0.045	-0.002	=C< (ortho)	-0.103	-0.290	-0.217
=C< (ring/ring)	0.047	0.000	0.213	=C< (meta)	-2.206	-0.366	-0.073
=C<	-2.866	-1.715	-0.360	=C< (para)	-2.308	-0.103	-0.264
Oxygenates							
–OH	-2.229	-1.089	-0.191	–O– (aromatic)	-3.373	0.006	0.045
=O (aldehyde)	1.730	-0.001	-0.379	–COOH	-2.576	0.000	0.045
=O (ketone)	2.109	-0.001	-0.359	–COO– (1^{st})	-2.222	-7.319	-0.684
–O– (1^{st})	3.419	-1.322	-0.001	–COO– (2^{nd})	-0.773	0.000	0.044
–O– (2^{nd}+)	2.138	-1.219	-0.001	–COO– (3^{rd})	-9.999	0.000	0.044
–O– (ring)	0.499	-0.082	0.006				

Table D-2. Constants for Eq. (D-2).

Hidden Node Label	Output Node Label		
	CN	RON	MON
a_k	17.0493	105.286	93.354
b_k	6.3931	3.289	11.871
c_k	0.4384	-0.713	-14.666
d_k	-0.044	-0.311	-36.844
e_k	0.0006	0.024	12.125

Table D-3. Standard deviation of errors for the polynomial group contribution correlation.

Type of Compound	Property		
	CN	RON	MON
Paraffins	14.3	9.4	8.3
Olefins	13.7	11.6	12.1
Alkynes	N/A	11.2	1.2
Naphthenes	13.7	10.1	8.0
Aromatics	10.6	6.3	5.5
Oxygenates	13.1	6.9	8.3
Overall	**13.2**	**10.2**	**9.2**

,

D.3. BOUNDED FUNCTION

Polynomial functions are unbounded, so correlations based on polynomial fits can yield inaccurate and unrealistic results when extrapolated. In an attempt to obtain a correlation that could be extrapolated with greater confidence, a bounded function was used to relate the sum of group contributions to CN and ONs. In this alternative formulation, the sum of group contributions includes a constant.

$$x_k = G_{o,k} + \sum_{j \in groups} G_{j,k} \cdot n_j \quad , \tag{D-3}$$

where x_k is the sum of the group contribution factors for property k, $G_{o,k}$ is a constant for property k, $G_{j,k}$ is the contribution of group j to the input to property k, and n_j is the number of j groups in the molecule. A nested logistic function was used to relate the sum of group contributions to cetane and octane numbers.

$$Z_k = LB_k + \frac{\left(UB_k - LB_k\right)}{1 + \exp\left(-\alpha_k - \dfrac{\beta_k}{1+\exp\left(-x_k\right)}\right)} \quad , \tag{D-4}$$

where Z_k is property k, LB_k is the lower bound on property k, UB_k is the upper bound on property k, and α_k and β_k are constants. LB_k, UB_k, α_k, and β_k are empirical constants determined form a regression analysis. This function was based on the simple ANN shown in Fig. D-1. The group contribution factors are given in Table D-4 and the constants for the Eq. (D-4) are given in Table D-5. Errors in this correlation are summarized in Table D-6.

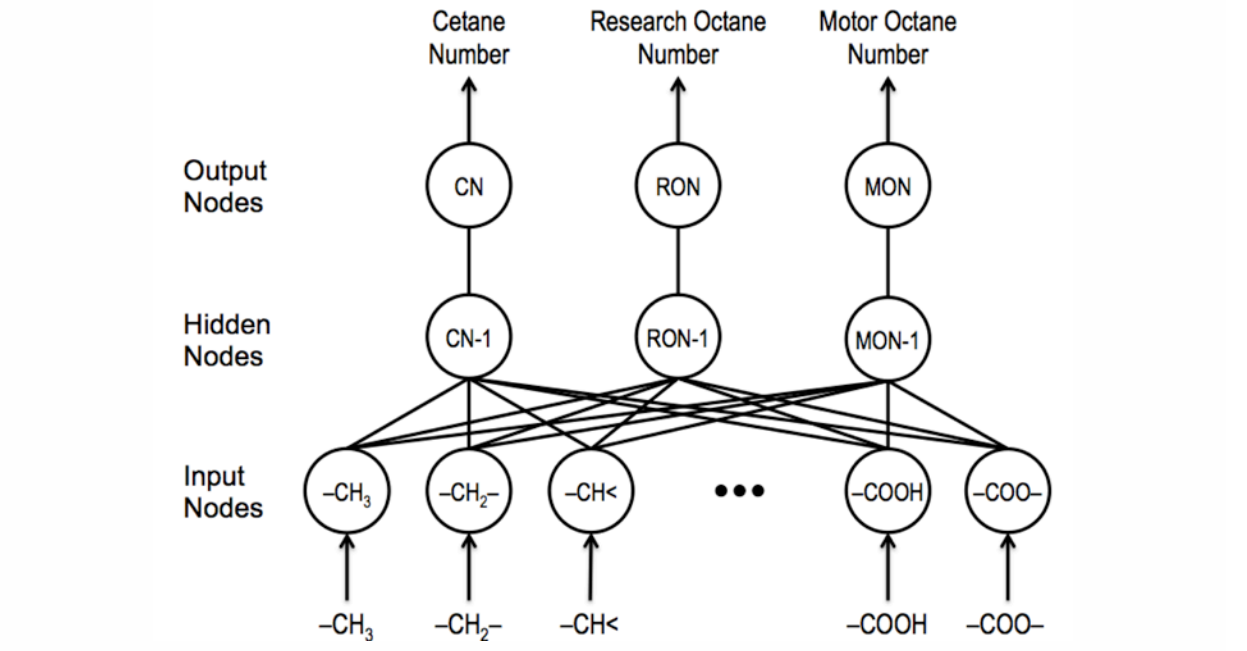

Fig. D-1. Simple ANN used to generate the bound function for a group contribution correlation represented by Eq. (D-3)

Table D-4. Group factors for conventional a group contribution correlation with a bounded function.

Group	Property			Group	Property		
	CN	RON	MON		CN	RON	MON
Constant, $G_{o,k}$	-0.045	0.019	-0.001				
Paraffins							
$-CH_3$	-0.012	-0.086	-0.054	–CH< (C# ≥ 4)	0.067	0.075	0.080
$-CH_2-$	-0.021	-0.063	-0.041	>C< (C# = 2)	0.033	0.258	0.169
–CH< (C# = 2)	-0.011	0.075	0.047	>C< (C# ≥ 3)	0.079	0.291	0.241
–CH< (C# = 3)	0.029	0.097	0.066				
Olefins / Alkynes							
$=CH_2$	-0.004	-0.060	-0.061	>C= (C# = 2)	0.066	0.109	0.125
=CH– (C# = 2)	-0.001	0.043	0.028	>C= (C# ≥ 3)	0.112	0.076	0.029
=CH– (C# = 3)	-0.023	0.000	-0.023	≡CH	-4.285	0.062	0.061
=CH– (C# ≥ 4)	0.028	0.095	0.069	≡C–	0.100	-0.001	-0.046
Naphthenes							
$-CH_2-$	-0.014	-0.030	-0.034	>C<	0.011	0.100	0.135
–CH<	0.007	0.018	0.012	=CH–	0.002	0.588	0.239
–CH< (ortho)	0.020	0.048	0.048	>C=	0.090	0.001	-0.167
Aromatics							
=CH–	0.000	-0.013	0.069	=C< (ortho)	0.005	-0.015	0.086
=C< (ring/ring)	0.015	0.001	-0.313	=C< (meta)	0.931	0.019	0.243
=C<	0.087	0.170	-0.236	=C< (para)	0.707	-0.017	0.117
Oxygenates							
–OH	0.087	0.079	0.053	–O– (aromatic)	0.042	0.016	-0.112
=O (aldehyde)	-0.073	0.887	0.147	–COOH	0.080	3.594	0.004
=O (ketone)	-0.068	-0.123	0.146	–COO– (1^{st})	0.050	2.948	1.795
–O– (1^{st})	-0.164	1.121	0.020	–COO– (2^{nd})	0.028	3.594	1.609
–O– (2^{nd}+)	-0.028	0.232	0.005	–COO– (3^{rd})	0.407	3.594	1.609
–O– (ring)	-0.013	-0.021	-0.108				

Table D-5. Constants for Eq. (D-4).

Hidden Node Label	Output Node Label		
	CN	RON	MON
LB_k	0.0	-30.0	-80.0
UB_k	100.0	110.0	100.0
α_k	-18.771	24.954	21.947
β_k	40.850	-57.007	-52.183

Table D-6. Standard deviation of errors for the bound function group contribution correlation.

Type of Compound	Output Node Label		
	CN	RON	MON
Paraffins	13.5	9.5	7.6
Olefins	10.5	6.4	8.0
Alkynes	N/A	14.9	0.0
Naphthenes	13.5	10.5	7.7
Aromatics	10.1	5.3	4.6
Oxygenates	13.8	16.0	5.4
Overall	13.5	10.3	7.2

REFERENCES

1. T. Albahri, "Structural Group Contribution Method for Predicting the Octane Number of Pure Hydrocarbons and Their Mixtures," *Industrial and Engineering Chemistry Research*, 42: 657 – 662 (2003).