CHEMICAL ENGINEERING TRANSACTIONS
VOL. 78, 2020
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Jeng Shiun Lim, Nor Alafiza Yunus, Jiří Jaromír Klemeš

ISBN 978-88-95608-76-1; **ISSN** 2283-9216
DOI: 10.3303/CET2078056

Cetane Number Estimation of Pure Compound using Group Contribution Method

Shah Aznie Ariffin Kashinath, Haslenda Hashim, Azizul Azri Mustaffa, Nor Alafiza Yunus*

Process Systems Engineering Centre (PROSPECT), Research Institute for Sustainable Environment, School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
alafiza@utm.my

Water-in-diesel emulsion, also known as WIDE fuel, is one of the promising alternative fuels for diesel engines due to its positive impact on the performance and combustion characteristics of the engine, while at the same time reducing the emissions of NOx and particulate matter, without the need to modify the diesel engine. Cetane number is a key property that is considered in the formulation of diesel fuel. Cetane number is used to determine the ignition quality of fuel. One of the limitations in designing the diesel fuel formulation is the limited experimental data. The objective of this paper is to develop the cetane number property estimations for a pure compound. A property model was developed using the group contribution approach. In this approach, the molecular structure of the chemical was represented and divided into three levels, namely first-order (220 groups), second-order (130 groups), and third-order (74 groups). The cetane number with the group contribution occurrences of 271 chemicals were regressed using linear regression in MATLAB software to generate the contribution values of the three group levels. The regression step yielded contribution values of 43 groups for the first-order, 35 groups for the second-order, and 7 for the third-order. The coefficient of determination, R^2, for the cetane number property models was 0.9447, indicating that the proposed model had a good correlation and is reliable to use.

1. Introduction

The diesel engine is a type of internal combustion engine, which offers better fuel-to-power conversion efficiency. Many studies have explored the potential of diesel blends or water-in-diesel emulsion fuel as a promising fuel due to its positive impact on the performance, combustion characteristics, and emission factors of the engine. Cetane number (CN) is one of the most significant properties for determining the ignition quality of an engine. According to Eloisa et al. (2011), CN influences the engine start ability, emissions, and peak cylinder pressure and combustion noise. When the diesel fuel mixture or emulsion is formulated, all the chemicals involved during formulation will determine the physical properties of the fuel. Dhinesh and Annamalai (2018) found that by adding cerium oxide nanoparticles in emulsion fuel, the quality of the cetane number of the emulsion fuel was improved. Leng et al. (2018) reported that adding the surfactant to the emulsion improved the cold-flow properties and ignition delay (cetane number) of the diesel fuel. Physical and thermodynamic property prediction models for pure compounds are one of the vital prerequisites for performing tasks such as simulation and optimization and computer-aided molecular/mixture (product) design, especially when the experimental value for the property of the pure compound is not available or limited (Hukkerikar et al., 2012a). In the domain of property prediction models, a few researchers have implemented group-contribution (GC) methods to predict pure compound properties such as open cup flash point (Constantinou and Gani, 1994), melting point, boiling point (Marrero and Gani, 2001), viscosity and surface tension (Conte et al., 2008), lethal concentration, LC50 (Hukkerikar et al., 2012a), and heat of combustion (Yunus and Zahari, 2017). All these proposed models are generally suitable to obtain the needed property values since these methods provide the advantage of quick estimates without requiring substantial computational work. In GC methods, the property of a component is a function of structurally dependent parameters, which are determined as a function of the frequency of the groups that represent the molecules

Paper Received: 18/04/2019; Revised: 02/09/2019; Accepted: 14/09/2019
Please cite this article as: Ariffin Kashinath S.A., Hashim H., Mustaffa A.A., Yunus N.A., 2020, Cetane Number Estimation of Pure Compound using Group Contribution Method, Chemical Engineering Transactions, 78, 331-336 DOI:10.3303/CET2078056

and their contributions (Hukkerikar et al., 2012b). The GC method has proven able to provide a good prediction and only requires chemical structure as input (Yunus and Zahari, 2017). Due to its predictive capability, the GC method was considered for the CN estimation in this study. The objective of this paper is to propose a CN property model using the group contribution (GC) method. In the group-contribution approach, the molecular structure of the chemical is divided into a set of functional groups, where each group contributes to the value of the property (Conte et al., 2008).

2. Methodology

There are five steps involved in the CN property estimation, as illustrated in Figure 1 below:

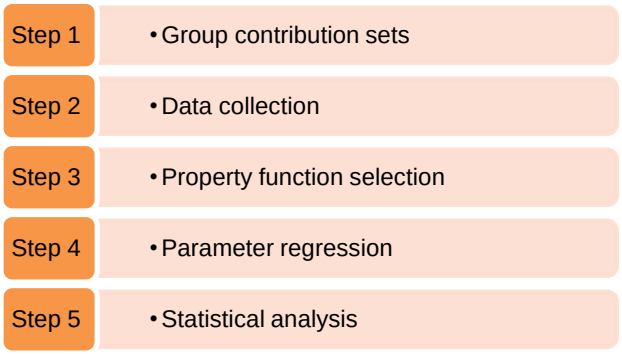

Figure 1: Methodology for cetane number (CN) property estimation

2.1 Group contribution sets

In this study, the functional group of the chemical was defined according to Marrero and Gani, (2001). This definition was employed to predict the variety of properties such as critical temperature, critical pressure, the Standard Gibbs energy, and the Standard Enthalpy of Vaporization covering more than 2,000 compounds. For the GC method, generally, the property estimation of chemicals is defined as three levels, namely first-order, second-order, and third-order groups. The basic (first) level uses the contributions of first-order groups that describe a wide variety of organic compounds. The higher (second) level provides additional structural information, which is not covered by the first-order groups, and thus provides corrections to the estimation at the first-level. The final level (third-order) provides the adjustment to the prediction made from the first and second level, where the contributions from the structure of complex molecules are calculated (Conte et al., 2008). More detailed information and distribution of each level are given in Figure 2.

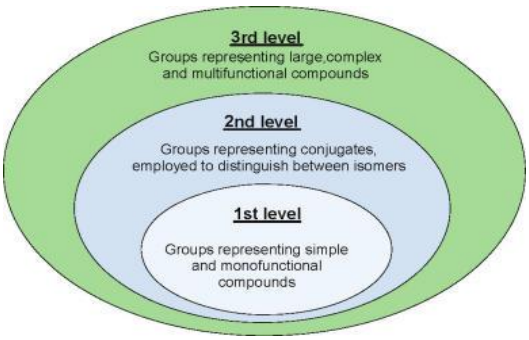

Figure 2: The description of the multilevel approach for the group contribution method (Conte et al., 2008).

Eq(1) below, as given by Marrero and Gani (2001), was used to represent the general form of the function *f(X)* of the target property *X* for the property estimation model:

$$f(X) = \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k \quad (1)$$

Where C_i, D_j and E_k are the contribution values for the first-order, second-order, and third-order groups with N_i, M_j, and O_k being the occurrences of each group.

2.2 Data collection

A CN dataset was collected as the initial database from Yanowitz et al., (2017) Compendium of Experimental Cetane Numbers, which contains 333 chemicals from the common families (alkanes, alkenes, ethers, esters, aldehydes, ketones, alcohols, and furan). The chemical structures of the compounds were identified and defined for the first-, second- and third-order levels according to Marrero and Gani (2001). Also, the occurrences of the chemicals were collected using ICAS software. The occurrences of the group contribution for all chemicals were used for the model regression in Step 4.

2.3 Property function selection

The property function was selected based on the CN trend observed in the data collected in Step 2. This function must show the best possible fit with the experimental data and should also provide good extrapolation capability. The collected experimental data on CN was plotted against the occurrences of the CH_2 group for various families of compounds to identify the best model for CN property. The resulting trend from the data collected shows that the property function is a linear function of the CN property function. Hence, the CN model is represented by Eq(2):

$$CN = CN_0 + \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k \quad (2)$$

2.4 Parameter regression

The Levenberg-Marquardt method was selected for the regression step to minimize the sum of squares of the differences between the experimental and estimated values of the CN property, as per the method outlined in Conte et al., (2008) for the parameter regression of surface tension and viscosity. The regression step was done using MATLAB. For Eq(2), CN is the cetane number of the chemical and CN_o is the universal constant for the model. The contribution values of the contribution groups C_i, D_j, and E_k was determined in three steps. The first step involved determining the universal constant, CN_o, and the contribution value of the first-order groups. The results of the universal constant and the group contribution value of the first-order group were used to determine the contribution value of the second-order group, D_j. The final step was the regression of the contribution value for the third-order group, E_k, using the results of the first and second steps. The final results of the regression step, which consists of the universal constant and the contribution values of the three levels, were analyzed to identify the outliers in the experimental data. The identified outliers were removed and the GC model parameters regressed again. Following that, 35 chemicals were identified as the outliers, as these did not follow the average trend. Meanwhile, 298 chemicals that fulfilled the trend were used for parameter regression. The training and testing data were divided randomly, with 271 chemicals as the training set and 26 chemicals as the testing set. The testing step is known as the validation step to verify the capability of the model.

2.5 Statistical analysis

The statistical analyses employed in this study are the Standard Deviation (SD), the Relative Deviation (RD), the Average Absolute Error (AAE), and the Average Relative Error (ARE), as defined by Eq(3) to Eq(6). All these equations are commonly used to verify the property model developed using the GC-based method (Marrero and Gani, 2001).

$$SD = \sqrt{\sum \frac{(\theta_i^{est} - \theta_i^{exp})^2}{N}} \quad (3)$$

$$RD = \frac{|\theta_i^{est} - \theta_i^{exp}|}{\theta_i^{exp}} 100 \quad (4)$$

$$AAE = \sum \frac{|\theta_i^{est} - \theta_i^{exp}|}{N} \quad (5)$$

$$ARE = \frac{\sum RD_i}{N} \quad (6)$$

Where N is the number of experimental data and θ_i^{est} and θ_i^{exp} are the predicted cetane number and experimental cetane number. The results of the statistical analysis are shown in Section 3 below, where a good prediction model is demonstrated with an R^2 value close to unity.

3. Results and discussion

The results of the contribution values are shown in Tables 1, 2, and 3 for the first-order, second-order, and third-order groups. From these tables, it can be concluded that the regression of the experimental data generated (43, 35, and 7) group contribution values for the first, second, and third-order groups. It is important to highlight that all the groups listed in Tables 1 to 3 followed the functional groups presented in Marrero and Gani, (2001) for the first-, second- and third-order levels. The results in Tables 1 to 3 were used to predict the CN of the chemicals in the training dataset. The plots of the estimated values of CN were then compared to the experimental training and testing data, as shown in Figure 3. The model with the three-level groups predicted the CN accurately, with an R^2 value equal to 0.9447 and 0.926 for the training set and the testing set. The CN_o value based on Eq(2) is 16.04.

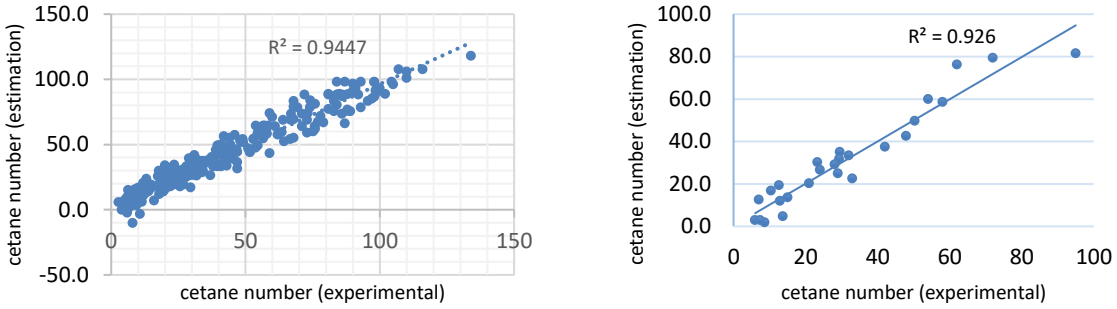

Figure 3: Estimated CN versus experimental CN for (a) training data and (b) testing data

Table 1: Contribution values of the first-order group

No.	Group	C_i	No.	Group	C_i
1	CH_3	3.49	23	CHCO	-11.296
2	CH_2	4.879	24	CHO	20.867
3	CH	-3.7	25	CH_3COO	-5.795
4	C	-7.296	26	CH_2COO	-12.851
5	CH_2=CH	-1.174	27	aC-COO	-4.005
6	CH=CH	-8.088	28	COO	-6.28
7	CH_2=C	-2.623	29	CH_3O	15.253
8	CH=C	-3.646	30	CH-O	-2.696
9	aCH	-2.76	31	aC-O	-0.385
10	aC	0.09	32	OCH_2CH_2OH	-0.832
11	aC	-3.848	33	OCH_2CHOH	-4.023
12	aC	-16.722	34	CH_2cyc	1.557
13	aC-CH3	0.49	35	CHcyc	-1.354
14	aC-CH_2	-3.601	36	Ccyc	-5.537
15	aC-CH	-11.132	37	CH=CHcyc	-2.253
16	aC-C	-23	38	CH=Ccyc	-4.766
17	aC-CH=CH_2	-22.999	39	CH_2=Ccyc	1.047
18	OH	-13.769	40	O	1
19	aC-OH	0.596	41	CO	-7.022
20	COOH	-21.342	42	-O-	26.861
21	CH_3CO	-5.334	43	Ccyc=C	-3.811
22	CH_2CO	-8.401			

Table 2: Contribution values of the second-order group

No.	Group	D_j	No.	Group	D_j
1	$(CH_3)_2CH$	2.010	19	aC-CH_n-O- (n in 1..2)	-2.804
2	$(CH_3)_3C$	-3.767	20	aC-$CH(CH_3)_2$	4.235
3	$CH(CH_3)CH(CH_3)$	-7.785	21	(CH_n=C)cyc-CH_3 (n in 0..2)	2.141
4	$CH(CH_3)C(CH_3)_2$	-7.836	22	(CH_n=C)cyc-CH_2 (n in 0..2)	-4.525
5	CH_n=CH_m-CH_p=CH_k (k,m,n,p in 0..2)	4.047	23	CHcyc-CH_3	0.025
6	CH_3-CH_m=CH_n (m,n in 0..2)	-0.199	24	CHcyc-CH_2	0.384
7	CH_2-CH_m=CH_n (m,n in 0..2)	-0.595	25	CHcyc-CH	-2.602
8	CH_p-CH_m=CH_n (m,n in 0..2; p in 0..1)	0.252	26	CHcyc-C	1.716
9	CHCHO or CCHO	-21.095	27	CHcyc-C=CH_n (n in 1..2)	-1.943
10	CH_3COCH_2	-1.461	28	CHcyc-OH	2.654
11	CHCOOH or CCOOH	12.637	29	Ccyc-CH_3	0.043
12	CH_3COOCH or CH_3COOC	8.058	30	AROMRINGs1s2	0.226
13	CHOH	-1.507	31	AROMRINGs1s3	0.618
14	COH	5.850	32	AROMRINGs1s4	-5.665
15	COO-CH_n-CH_m-OOC (n, m in 1..2)	-3.885	33	AROMRINGs1s2s3	0.867
16	OOC-CH_m-CH_m-COO (n, m in 1..2)	10.422	34	AROMRINGs1s2s4	0.717
17	CH_m-O-CH_n=CH_p (m,n,p in 0..3)	-0.036	35	AROMRINGs1s3s5	-2.293
18	CH_n=CH_m-COO-CH_p (m,n,p in 0..3)	-2.349			

Table 3: Contribution values of the third-order group

No .	Group	E_k
1	aC-CHncyc (fused rings) (n in 0..1)	4.569
2	CH multiring	0.183
3	C multiring	-2.779
4	aC-CH_n-O-CH_m-aC (different rings) (n,m in 0..2)	5.608
5	AROM.FUSED[2]	-9.330
6	AROM.FUSED[2]s1	8.782
7	AROM.FUSED[2]s2	9.603

Table 4 shows the statistical results of the regression procedure for Step 1 to Step 3. The statistical parameter values (R^2, AAE, ARE, and SD) are given for all three levels of the compounds in the dataset containing 271 chemicals. The result better predicted the property if the regression step considered all three group contribution values to predict CN, as proven by the improvement in the R^2 values and the reduced standard deviation.

Table 4: Comparison of the statistical analysis of training data for the contribution of all group levels

Statistical analysis	1st order	1st and 2nd order	1st, 2nd and 3rd order
R^2	0.9348	0.9445	0.9447
Average absolute error, AAE	5.930	5.4206	5.39
Average relative error, ARE	23.10	20.83	20.48
Standard deviation, SD	7.482	6.904	6.8933

To prove its capability, the model was tested using a testing dataset. The contribution values from Table 1 to Table 3 were used to predict the CN of the chemicals in the testing dataset. A comparison between the experimental CN and the estimated CN for 26 compounds is shown in Figure 3b. The R^2 value of the predicted CN for the testing dataset was 0.926. Therefore, the predictive performance of the model is acceptable, at least for these compounds. Table 5 shows an example of the calculation of the CN value of 1-decanol using the developed property model. For reference, the experimental value of CN for 1-decanol is 50.3. By using this model, the prediction value of 1-decanol returned a value of 49.7. Therefore, the model developed using the group contribution method can accurately predict the cetane number (CN) of the compound.

Table 5: Prediction of the cetane number of 1-decanol

Property	Value	
Formula Molecule	$C_{10}H_{22}O$	
Molecular structure		
Universal constant, CN_0	16.04	
First-order group	*Group Contribution (C_i)*	Occurrences (N_i)
CH_3	3.490	1
CH_2	4.878	9
OH	-13.769	1
Using Eq(2):	$CN = CN_0 + \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k$	
	CN = 16.04 + 1(3.49) + 9(4.878) + 1(-13.769) = 49.7	

4. Conclusion

A cetane number (CN) property model was developed in this study and the performance of the model tested. The CN experimental data was compared to the estimated CN data to identify the prediction accuracy of the model. As a validation step, a set of testing data containing 26 chemicals was used to verify the prediction accuracy of the model. The R^2 value of the training data and the testing data was 0.9447 and 0.926. The model developed using the group-contribution approach is therefore reliable and can thus serve as an alternative method to estimate the CN of nonionic surfactants.

Acknowledgements

The authors would like to thank to the Ministry of Education Malaysia and Universiti Teknologi Malaysia (UTM) for providing the research fund for this study (Q.J130000.2651.16J47 and Q.J130000.2446.04G20).

Reference

Constantinou L., Gani R., 1994, New group-contribution method for the estimation of properties of pure compounds, AIChE Journal, 10, 1697-1710.

Conte E., Marstinho A., Matos H.A., Gani R., 2008, Combined group-contribution and atom connectivity index-based methods for estimation of surface tension and viscosity, Industrial & Engineering Chemistry Research, 47(20), 7940–7954.

Dhinesh B., Annamalai M., 2018, A study on performance, combustion and emission behaviour of diesel engine powered by novel nano nerium oleander biofuel, Journal of Cleaner Production, 196, 74–83.

Eloisa T.J., Marta S.J., Andreja G., Irenca L., Dorado M.P., Breda K., 2011, Physical and chemical properties of ethanol-diesel fuel blends, Fuel, 90, 795 – 802.

Hukkerikar A.S., Kalakul S., Sarup B., Young D.M., Sin G., Gani R., 2012a, Estimation of environment-related properties of chemicals for design of sustainable processes: Development of group-contribution$^+$ (GC^+) property models and uncertainty analysis, Journal of Chemical Information and Modelling, 52(11), 2823-2839.

Hukkerikar A.S., Sarup B., Ten Kate A., Abildskov J., Sin G., Gani R., 2012b, Group contribution$^+$ (GC^+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis, Fluid Phase Equilibria, 321, 25-43.

ICAS v18.0, 2014, CAPEC, Technical University of Denmark, Lyngby, Denmark.

Leng L., Chen J., Leng S., Li W., Huang H., Li H., Zhou W., 2018, Surfactant assisted upgrading fuel properties of waste cooking oil biodiesel, Journal of Cleaner Production, 210, 1376-1384.

Marrero J., Gani R., 2001, Group contribution-based estimation of pure component properties, Fluid Phase Equilibria, 183-184,183-208.

US DoE, 2017, Compendium of Experimental Cetane Number Data, US Department of Energy <https://www.nrel.gov/docs/fy17osti/67585.pdf > accessed 20.01.2019.

Yunus N.A., Zahari N.N.N.N.M., 2017, Prediction of standard heat of combustion using two-step regression, Chemical Engineering Transactions, 56,1063-1068.