Grouping of Retention Index on Gas Chromatography using Cluster Analysis

I Husna¹, A Rusyana⁵, Muslem¹, G M Idroes³, R Suhendra⁴and R Idroes¹,²,⁺

¹ Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Kopelma Darussalam, Banda Aceh 23111, Indonesia
² Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Kopelma Darussalam, Banda Aceh 23111, Indonesia
³ Department of Chemical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Kopelma Darussalam, Banda Aceh 23111, Indonesia
⁴ Department of Informatics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Kopelma Darussalam, Banda Aceh 23111, Indonesia
⁵ Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Kopelma Darussalam, Banda Aceh 23111, Indonesia

Email: rinaldi.idroes@unsyiah.ac.id

Abstract. Retention index data consisting of 146 compounds commonly found in fuels are used to analyze the level of a psycho-chemical property similarity with cluster analysis methods. Cluster analysis is applied based on the temperature and column dependencies. The results show that the retention index was divided into four groups. The first group consists of benzene and cycloalkanes. The retention index of the compound shows high-temperature dependence and dependence on small column polarity. The second group consists of aromatic compounds and compounds that contain N, S, O. The retention index of these compounds shows dependence on temperature and polarity of large columns. The third group consists of aromatic compounds and hydrocarbons. The retention index of hydrocarbon compounds shows temperature dependence smaller than N, S, O. compounds and also the dependence on small column polarity. And the fourth group consists of only two compounds namely n-Hentriacontane and chrysene. The retention index of the two compounds shows a significant temperature dependence.

1. Introduction
Retention index was first introduced by Erwin Kovats in 1958 which became known as the Kovats retention index. The Kovats retention index uses the n-alkane homologous series as the standard series to change the non-fixed retention time into a system that is relatively stable against the change of analysis conditions [1]. The non-fixed time behavior makes this parameter just use as a separation marker [2] of a compound that is subsequently identified by mass spectrometry [3]. On another hand, the stable value of the retention index makes it possible to obtain psycho-chemical [4] and thermodynamic [1] information of a compound.

¹⁺Corresponding author: rinaldi.idroes@unsyiah.ac.id
Besides stability, the retention index also has the lowest systematic error than other retention parameters such as retention time, adjusted retention time, capacity factor, etc [5]. These conditions make it possible to have a relationship between psycho-chemical and thermodynamic properties and retention index. To obtain this relationship, the statistical method is required to analyze the retention index data from the various compounds and classified into the group.

Idroes et al., (2018) have grouped flavor and fragrance retention indexes obtained from some column using the biplot method. The results show that retention indexes are divided into three groups. Each group consists of the similarity of the psycho-chemical property of the compounds. The first group consists of n-alkanes with lower carbon numbers and lower retention indexes. The second group consists of n-alkanes with higher carbon numbers and terpenoid compounds with higher Kow log values. The third group consists of alcohol, aldehydes, esters, and carboxylic acid compounds [6].

In this study, the retention index data of pure compounds that are commonly found in fuels derived from bricks and petroleum are analyzed for similarity levels based on groups formed using cluster methods. The psycho-chemical properties from the compounds in each group are identified and compared to other groups.

2. Literature Review
Cluster analysis is a statistical method for grouping or combining n objects into k groups [7]. Cluster analysis is one of the multivariate techniques for grouping objects into suitable groups based on their resemblance. Objects formed in the same group will be more similar than objects between groups. The similarity between the objects is measured by the Euclidian distance. The equation from the Euclidian distance is:

\[ d(x,y) = [(x_1-y_1)^2 + (x_2-y_2)^2 + \ldots + (x_p-y_p)^2]^{1/2} \]

Where: \( d(x,y) \) is the distance between object i to object j, \( x_p \) is the middle value in the group i and \( y_p \) is the middle value in the group j [8]. Cluster analysis aims to determine the diversity of the object under study and to find out the similarities between objects in a group formed. The greater the similarity in the group, the better the grouping will be.

Group analysis is divided into two, namely hierarchical and non-hierarchical. There are two approaches in the analysis of hierarchical groups, one of which is the agglomerative approach. Agglomerative cluster analysis method grouping objects starting from n objects into one group and it is also called a down to top. To measure the dissimilarity between groups formed can be using the average linkage. Average Linkage measures or looks for similarities between groups by looking at the average distance of observation data. The formula for Average Linkage is as follows:

\[ d_{G,A}(G,H) = \frac{1}{N_G N_H} \sum_{i \in G} \sum_{j \in H} d_{ij} \]

Analysis of hierarchical groups initially consisted of separate groups so that there were as many groups as observational data. Then similar data will be entered into the same group, repeated until formed into a group that contains all observational data [9]. The results of the cluster analysis are dendrogram based on distances between objects. The hierarchical cluster analysis method requires measurement of inequality (distance) for each data to be grouped to facilitate the development of the dendrogram.

3. Data and Methods
3.1. Data collection
This research takes retention index data from Lai’s 1995 literary article [10]. The sample used was 146 pure compounds that are generally found in liquid fuels derived from coal and petroleum. The columns used are two capillary columns with different stationary phases of polarity. First is the Rtx-50
column with a 50% phenyl-50% methyl polysiloxane stationary phase and the second is a DB-5 column with a 5% phenyl-95% methyl polysiloxane stationary phase. The column temperature is programmed linearly from levels 40 to 310 at a heating rate of 2, 4, 6 °C min⁻¹.

In this research, the n-alkane series is used as a standard to calculate the retention index of several compounds because it has a robust retention index.

3.2. Data analysis
Cluster analysis is performed using XLSTAT software. After the data is entered into the software, the analysis is carried out using the cluster analysis.

4. Result and discussion
The retention index is included in the secondary retention parameter because it is generated through mathematical calculations using primary retention parameter data which is dead time and retention time. The retention index is a parameter that is often used as a standard for gas chromatography using the n-alkane series as an internal standard under isothermal conditions. The retention index is also defined as 100 times nₖ (carbon number) [11]. In this study, the retention index data is used to see the level of similarity based on groups formed using the cluster analysis. The compounds analyzed are common compounds found in fuels. Alkanes, cycloalkanes, and aromatics are the three main components that are often found in liquid fuels.

![Dendrogram using Average Linkage (Between Groups)](image)

Figure 1. Dendrogram of cluster analysis.
Table 1. Compounds in their cluster.

| Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|-----------|-----------|-----------|-----------|
| 1-Ethyl-3-Methylbenzene | 1,3,5-Trimethylbenzene | cis-Octahydro-1H-Indene | sec-Butylbenzene |
| 1,2,4-Trimethylbenzene | n-Butylocyclohexane | 1-Decane | tert-Butylocyclohexane |
| n-Decane | Isopropylbenzene | 2-Methylnladen | 1-Methylnladen |
| n-Pentylcyclohexane | cis-Decalin | n-Undecane | 1-Undecane |
| trans-Decalin | n-Butylbenzene | Indan | Phenyl Ethyl Ether |
| 1,3-Dimethylnbenzene | 1,2-Dimethylnbenzene | n-Nonane | 1-Nonane |
| n-Propylcyclohexane | Cyclohexene | Benzene | n-Heptane |
| 1-Heptane | Cyclohexene | Methylocyclohexane | Toluene |
| Pyridine | cis-1,2-Dimethylocyclohexane | Dimethylocyclohexane | Dimethylocyclohexane |
| cis-1,3-Dimethylocyclohexane | n-Octane | 1-Octane | 1-Methylocyclohexane |
| n-Hexene | n-Pentane | Hexylbenzene | 1,2,3,4-Tetramethylbenzene |
| Cluster 2 | Cluster 3 | Cluster 4 |
| 1,3,5-Triethylbenzene | n-Tridecane | 1-Tridecane |
| Bicyclohexyl | n-Dodecane | 1-Dodecane |
| 1,4-Dipropylbenzene | Tetralin | n-Tetradecane |
| 5,6,7,8-Tetrahydro-3Methylquinoline | 2-Methylnaphthalene | Cyclohexylbeneze |
| Quinoline | Naphthalene | Benzothiophene |
| 1,2-Dimethylnaphthalene | 1,3-Dimethylnaphthalene | 2,7-Dimethylnaphthalene |
| 1-Ethylnaphthalene | 2-Ethylnaphthalene | 2,6-Di-tert-Butylphenol |
| 3-Methylbiphenyl | 1,8-Dimethylnaphthalene | 5,6,7,8-Tetrahydro-1-Naphthol |
| 1-Hexadecene | n-Octylbenzene | n-Pentadecane |
| 4,4’-Dimethylbiphenyl | Fluoren | 2-Naphthol |
| 1,2,3,4,5,6,7,8Octahydroacridine | Dibenzy1 Ether | n-Decylocyclohexane |
| n-Octadecane | 1-Octadecane | n-Nonadecane |
| Cluster 3 | Cluster 4 |
| Phenanthrene | Anthracene | 1-Phenylnaphthalene | n-Eicosane |
| n-Henicosane | n-docosane | n-Tricosane | Pyrine |
| p-Terphenyl | Fluoranthene | n-Tetracosane | n-Pentacosane |
| n-Hexacosane | | | |
| n-Hentriacontane | Chrysene | | |
The results obtained from the cluster analysis in the form of a dendrogram showing grouping based on similarities in column polarity and the heating rate between the compounds studied. From the results of the analysis of these compounds are grouped into four main groups, see Figure 1 and Table 1. The first group consists of benzene and cycloalkanes. The retention index of the compound shows high-temperature dependence and dependence on small column polarity. The second group consists of aromatic compounds and compounds that contain N, S, O such as Quinoline, Indanol, and Benzothiophene. The retention index of the compound shows greater temperature dependence and is very dependent on the polarity of the column used. The third group consists of aromatic compounds and hydrocarbons. The retention index of hydrocarbon compounds shows a smaller temperature dependence than N, S, O, and dependence on small column polarity for hydrocarbon compounds, whereas N, S, O compounds are very dependent on column polarity. The fourth group consists of only two compounds namely n-Hentriacontane and Chrysene. These compounds show significant temperature dependence. The retention index of all the compounds studied was reduced by decreasing column polarity (from Rtx-50 to DB-5).

5. Conclusion

Based on the analysis of the retention index data of 146 pure compounds commonly found in liquid fuels can be grouped into four groups. The first group consists of benzene and cycloalkanes. The second group consists of aromatic compounds and compounds that contain N, S, O. The third group consists of aromatic compounds and hydrocarbons. And the fourth group consists of only two compounds namely n-Hentriacontane and Chrysene. Compounds in each group have similarities between one compound with other compounds.

Reference

[1] Didaoui, L., Touabet, A., Meklati, B. Y., Alia, E., Ezzouar, B. P. B., Lewin-kretzschmar, U., & Engewald, W., 1999, Comparison of Retention Indices of Some Explosives and Related Compounds Calculated by Different Mathematical Methods in Reversed-Phase Liquid Chromatography, *Journal of HRC-High Resolution Chromatography*, 11, 613–618

[2] Earlia, N., Rahmad, Amin, M., Prakoeswa, C. R. S., Khairan., and Idroes, R., The Potential Effect of Fatty Acids from Pliek U on Epidermal Fatty Acid Binding Protein: Chromatography and Bioinformatic Studies, *Sains Malaysia*, 48, 1019–1024.

[3] Pratiwi, S. U. T., Lagendijk, E. L., Weert, S. V., Idroes, R., Hertiari, T., and Hondel, C. V. D., 2015, Effect of Cinnamomum burmannii Nees ex Bl. and Massoia aromatica Becc. Essential Oils on Planktonic Growth and Biofilm formation of Pseudomonas aeruginosa and Staphylococcus aureus In Vitro, International Journal of Applied Research in Natural Products Vol. 8 (2), pp. 1-13.

[4] Nowotnik, D.P. and Narra, R. K., 1993. A Comparison of methods for the determination of dead time in a reversed-phase high-performance liquid chromatography system used for the measurement of lipophilicity, J. of Liquid Chromatography 16(18), 3919 – 3932.

[5] Idroes, R., 2011, Dead Time Determination and Its Influence to Retention Index in GC System Using Polar Capillary Column *Maj. Farm. Indones* 3 257-264

[6] Japnur, A. F., Idroes R., Suhendra, R., Rusyana, A., 2018, Kovats Retention Index Analysis of Flavor and Fragrance Compound using Biplot Statistical Method in Gas Chromatography Systems, *AIC Conf*

[7] Afifi, A. A., & Clark, V 1999 *Computer-Aided Multivariate Analysis* (3rd ed.). New York: Chapman & Hall/ CRC.
[8] Johnson, R. A., & Wichern, D. W 2007 *Applied Multivariate Statistical Analysis* (6th ed.). New Jersey: Pearson

[9] Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th ed.). United Kingdom: John Wiley & Sons, Ltd.

[10] Lai, Wei-Chuan, Chunshan Song 1995 Temperature-Programmed Retention Indices for G.C. And G.C.-M.S. Analysis of Coal- And Petroleum-Derived Liquid Fuels. *Fuel*. 10 1436-1451

[11] Idroes, R., 2005, Determination of Absolute Index System in High Performance Liquid Chromatography (RP-HPLC) *Malaysian J. Anal. Sci.* 3 224-232.