DEVELOP A HAZARD INDEX USING MACHINE LEARNING APPROACH FOR THE HAZARD IDENTIFICATION OF CHEMICAL LOGISTIC WAREHOUSES

A Thesis

by

ZHUORAN ZHANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Qingsheng Wang
Committee Members,	Mahmoud El-Halwagi
	Farzan Sasangohar
Head of Department,	Arul Jayaraman

December 2019

Major Subject: Safety Engineering

ABSTRACT

With the rapid development of chemical process plants, the safe storage of hazardous chemicals become an essential topic. Several chemical warehouse incidents related to fire and explosion have been reported recently. Therefore, an accurate hazard identification method for the logistic warehouse is needed not only for the facility to develop a proper emergency response plan but also for the residents who live near the facility to have an effective hazard communication. Furthermore, the government can better allocate the resources for first responders to make fire protection strategies, and the stakeholders can lead to improved risk management.

The storage of hazardous chemicals in a warehouse is a complex problem. The potential hazards include flammability, reactivity, and interaction among different types of hazardous chemicals. Hazard index is a helpful tool to identify and quantify the hazard in a facility or a process unit. Various hazard indices are developed in history. However, the challenge for this research is to improve the current method with the novel technique to implement our purpose.

The first objective of this research is to develop a "Storage Hazard Factor" (SHF) to evaluate and rank the inherent hazards of chemicals stored in logistic warehouses. In the factor calculation, the inherent hazard of chemicals is determined by various parameters (*e.g.*, the NFPA rating, the flammability limit, and the protective action criteria values, *etc.*) and validated by the comparison with other indices. The current criteria for flammable hazard ratings are based on flash point, which is proved to be insufficient. Two

machine learning based methods will be used for the classification of liquid flammability considering aerosolization based on DIPPR 801 database. Subsequently, SHF and other warehouse safety penalty factors (*e.g.*, the quantity of the chemicals, the distance to the nearest fire department, *etc.*) are utilized to identify the Logistic Warehouse Hazard Index (LWHI) of the facilities. In the last chapter, LWHI is applied to an actual case from Houston Chronicle, and several statistical analyses are used to prove that the LWHI is helpful for hazard identification to emergency responders and hazard communication to the public.

DEDICATION

This thesis is dedicated to my parents, Wei Huang and Yuewei Zhang,

and my love, Rui Sun,

for all their motivation, encouragement, and support.

ACKNOWLEDGEMENTS

I would like to thank my committee chair and academic advisor, Dr. Qingsheng Wang, for his constant support and insightful advice on my graduate career. I also would like to express gratitude to my committee members, Dr. Mahmoud M. El-Halwagi and Dr. Farzan Sasangohar, for their patient guidance and innovative ideas for my research. Special thanks should be given to Dr. M. Sam Mannan, who was my previous committee chair and mentor. He gave me the chance to join the Mary Kay O'Connor Process Safety Center (MKOPSC), which is the best place for process safety study. My research cannot be successfully done without their help and support.

Moreover, I would like to thank my colleague Shuai Yuan, who is my mentor on the machine learning method, and created a novel classification method with me. I also would like to thank Mengxi Yu, as the team leader of 'Potential to Cause Harm to the Public (PCHP)' project, and Dr. James Holste, as the acting director of MKOPSC, and Mr. T. Michael O'Connor, as the research associate and co-founder of MKOPSC, for their help to let me collect the raw data used in my case study from our partners. Special thanks should be given to John from HazardHub, who helped me to collect an important factor from their HazardHub API.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, I wish to thank my parents for their support and encouragement throughout my studies. I also want to thank my girlfriend Rui Sun, for her love and support and wish Rui can successfully pass her Ph.D. final defense next year.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor Qingsheng Wang and Professor Mahmoud M. El-Halwagi of the Department of Chemical Engineering and Professor Farzan Sasangohar of the Department of Industrial and Systems Engineering. The previous thesis committee chair is Professor M. Sam Mannan of the Mary Kay O'Connor Process Safety Center before he passed away on September 11, 2018.

The classification methods depicted in Chapter 3 were conducted in part by Shuai Yuan of the Department of Chemical Engineering and were submitted to publish in 2019. Houston Chronicle collectted and MKOPSC provided the data analyzed for Chapter 5. All other work conducted for the thesis was completed by the student independently.

Funding Sources

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

NOMENCLATURE

AIChE	American Institute of Chemical Engineers
AIT	Autoignition Temperature
CAS	Chemical Abstracts Service
CCPS	Center for Chemical Process Safety
CGS	Centimetre–Gram–Second System
CSB	Chemical Safety and Hazard Investigation Board
DIPPR	Design Institute for Physical Properties
DOE	Department of Energy
DSC	Differential Scanning Calorimeter
EPA	Environmental Protection Agency
ESS	Error Sum of Squares
F&EI	Fire and Explosion Index
GHS	Globally Harmonized System
HC	Hierarchical Clustering
IFAL	Instantaneous Fractional Annual Loss
ITC	Intercontinental Terminals Company
KC	K-Mean Clustering
LC_{50}	Lethal Concentration of the chemical in air that kills 50%
LD_{50}	Lethal Dose causes the death of 50%
LFL	Lower Flammability Limit

LWHI	Logistic Warehouse Hazard Index
MF	Material Factor
MHI	Material Hazard Index
MIE	Minimum Ignition Energy
MKOPSC	Mary Kay O'Connor Process Safety Center
Mond FETI	Mond Fire, Explosion, and Toxicity Index
NFPA	National Fire Protection Association
NF	Degree of Flammability
NH	Degree of Health Hazard
NR	Degree of Reactivity
OSHA	Occupational Safety and Health Administration
PAC	Protective Action Criteria
PCA	Principal Component Analysis
PCHP	Potential to Cause Harm to the Public
PHA	Process Hazard Analysis
RRI	Reactivity Risk Index
RSST	Reactive System Screening Tool
SHF	Storage Hazard Factor
SMD	Sauter Mean Diameter
UFL	Upper Flammability Limit
WCSS	Within-Cluster Sum of Square
WFC	West Fertilizer Company

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Motivation

Since entering the 21st century, people enjoy the benefits of the rapid development of the industry. With the innovations of the new chemical process, process safety must also be up to date and accommodate the new chemical process.

Many industrial cities around the world are facing a dilemma between economic growth and population growth. With the blooming development of chemical process plants, the safe storage of hazardous chemicals becomes an essential topic. People should understand that some inherent properties of a chemical which makes it profitable to our society may be hazardous in the meantime. The researcher lived in Tianjin for five years, which is one of the economic centers in northern China. Tianjin has developed a sub-provincial district named Binhai New Area, which is near the largest port in northern China, where more than 1800 facilities are related to the storage of hazardous chemicals [1]. This thesis is focused on hazard identification for chemical logistics warehouses, which is inspired by the Tianjin explosion that happened on August 12, 2015 [2].

1.1.1. Tianjin explosion

On the night of August 12, 2015, a series of devastating explosions happened in the Binhai New Area near Tianjin Port. This explosion caused a total of 173 fatalities, including 110 firefighters, and about 800 non-fatal injuries. Total direct economic losses in this incident were over 1 billion dollars. The first two explosions were the most serious

ones. It involved the detonation of about 800 tonnes of ammonium nitrate and left with a crater of 60 meters in diameter. [2]

Based on the final government investigation report [3], there are more than 40 kinds of hazardous chemicals (a total of 3000 tonnes) stored onsite. Some of the substances (*e.g.*, sodium cyanide) are highly toxic *via* inhalation, ingestion, and skin absorption. Other chemicals (*i.e.*, potassium nitrate and ammonium nitrate) may be highly flammable or highly explosive. The explosion happened in a warehouse owned by Ruihai International Logistics, which is in charge of handling various hazardous materials within the Tianjin Port.

Based on the incident anatomy by Yan *et al.* [2], the direct causes of the two major fire and explosion are spontaneously burning nitrocellulose and the subsequent detonation of the ammonium nitrate. However, a catastrophic incident typically occurs because of a series of unexpected events among multiple failures. The government report classified this event as an 'accountability accident,' and Yan *et al.* also point out the root causes may include the lack of safety knowledge and insufficient hazard identification.

1.1.2. Similar incidents in the history

In addition to the Tianjin explosion, several similar incidents happened in the chemical warehouse that motivates this research. On April 17, 2013, a fire and explosion occurred at the West Fertilizer Company (WFC) in West, Texas. The violent detonation fatally injured 15 people, and the local hospitals treated more than 200 injured victims. After almost three years of investigation, the U.S. Chemical Safety and Hazard

Investigation Board (CSB) released the final investigation report, which reveals several details and concludes various possible causes [4]. Ammonium nitrate, which is principally used in manufacturing fertilizer, had been confirmed as the primary material that exploded. Babrauskas [5] pointed out that this incident is a failure of regulation, where the storage facilities may lack awareness for the potential hazard.

Another related incident happened on March 17, 2019, where a storage tank caught fire at the Intercontinental Terminals Company (ITC) Deer Park facility near Houston. ITC handles and stores the petrochemical product as well as the raw materials. Although this incident did not cause casualties, the releasing benzene, toluene, and xylene may cause harm to the public by contaminating the drinking water. For the safety of the residents, the City of Deer Park collaborated with Envirodyne Laboratories to monitor the presence of volatile organic carbons (VOCs) and semi-volatile organic carbons (SVOCs) [6].

Those three examples show that the logistic warehouse is a weak point in process safety — it usually stores numerous hazardous chemicals in a considerable amount.

1.2. Improper storage hazards

The hazards presented by storage depend on the inherent material characteristics and the storage condition. Both unawareness of the material inherent hazard and the incompatible storage situation may lead to chemical incidents like as mentioned in the previous section. Statistics show that chemical incidents caused by improper storage make up almost 25% of all chemical incidents [7, 8]. Thus, well-established hazard identification is essential for any laboratory or large-scale storage facility.

Material inherent hazard includes flammability, toxicity, reactivity, and interaction among different types of hazardous chemicals. Several characterization techniques have been invented to represent the relevant hazardous level of chemicals. Many popular classification techniques have been known over the years, including the National Fire Protection Association (NFPA) 704, and Globally Harmonized System (GHS). Details will be discussed in the next chapter.

Apart from the material inherent hazard, storage condition is another critical factor that requires extra attention. For liquids stored in a vessel or tank, a leak may happen on rare occasions, such as overpressured vessels due to overfilling. If the material is toxic, the leaking vapor cloud can potentially result in health problems. If the material is flammable, a fire or explosion could be a catastrophic consequence. When considering the interaction between chemicals, which is common in a logistic warehouse with various chemicals stored onsite, the segregation and separation of materials within the storage area [9] is the key to prevent and mitigate the possible incident. Also, the facility siting is a primary consideration for public safety. The population density near the facility and the distance to the nearest fire station will be considered as a penalty factor in this thesis.

1.3. Hazard identification

Hazard identification is a part of risk assessment. In other words, hazard identification should be the starting point of a full process of risk assessment. The scope of hazard identification is to identify hazards and risk factors that have the potential to cause harm. Hazard identification should have been done at the design phase of a project, and repeated multiple times before the project starts, during the routine inspections, and after the incident happens. A common way is to classify the hazard in the workplace as biological, chemical, ergonomic, physical, *etc.* [10] This thesis is mainly focused on inherent chemical hazards as well as considering key factors that may affect the public.

1.4. Research objectives

Based on the investigation reports of the storage facilities related to safety incidents, some hazardous chemicals are mentioned more than one time, such as ammonium nitrate. Former researchers in Mary Kay O'Connor Process Safety Center (MKOPSC) have investigated the thermal decomposition and runaway reaction characteristics of some hazardous chemicals [11-13] while few studies have explored the hazard identification applications.

The first objective of this research is to develop a "Storage Hazard Factor" (SHF) to evaluate and rank the inherent hazards of chemicals stored in logistic warehouses. In the factor calculation, the inherent hazard of chemicals is determined by various parameters (*e.g.*, the NFPA rating, the flammability limit, and the protective action criteria values, *etc.*) and validated by the comparison with other indices. Two machine learning

based methods will be used for the classification of flammability. Subsequently, SHF and other warehouse safety penalty factors (*e.g.*, the quantity of the chemicals, the distance to the nearest fire department, *etc.*) are utilized to identify the hazard index of the facilities.

The index can be used not only for the facility to develop a proper emergency response plan but also for the residents who live near the facility to have an effective hazard communication. Furthermore, the government can better allocate the resources for first responders to make fire protection strategies, and the stakeholders can achieve improved risk management for the facility.

2. LITERATURE REVIEW

2.1. Hazard versus Risk

Hazard and risk are not interchangeable concepts in many ways [14], not only the definition in the dictionary but also the interpretation by a chemical engineer. Specifically, it is common to use the definitions given by the Center for Chemical Process Safety (CCPS) from the American Institute of Chemical Engineers (AIChE) in chemical engineering study and research:

> ***Hazard****: An inherent chemical or physical characteristic that has the potential for causing damage to people, property, or the environment.*
>
> ***Risk****: A measure of human injury, environmental damage, or economic loss in terms of both the incident likelihood and the magnitude of the loss or injury.*

Briefly, the hazard is an inherent chemical characteristic that may cause potential harm, whereas risk is considering both consequences and the likelihood of an incident. In other words, any chemical represents some hazards due to their inherent properties. For example, propane is a highly flammable substance, and this inherent property cannot be interfered with by human action. However, this inherent hazard characteristic does not represent risk or imminent harm. Different mitigative methods can be applied to lower the likelihood of the potential consequence, which is fire and explosion in this case.

2.2. Hazard identification methods

As discussed in the last chapter, hazard identification should be the first step to develop a chemical process. In the past century, several methods have been invented to identify the hidden hazards in the workplace. This section will introduce some successful and widely used methods.

2.2.1. Process hazard checklists

A checklist might be the most simple tool of hazard identification. Even today, it is widely used all over the world. It is the fastest way to get a concept of the whole chemical process. However, the checklist cannot give more details about the potential hazard with the simple ‘yes’ or ‘no’ answers. It will be more effective if the answers include more details, but it may also become trivial at the same time. Therefore, the checklist is not a useful method for interpretation.

2.2.2. NFPA 704

NFPA 704 is a standard developed by the National Fire Protection Association (NFPA), which is also known as ‘hazard diamond,’ and it provides hazard information for degrees of reactivity, toxicity, and flammability. [15] The subtitle of NFPA 704 is ‘standard system for the identification of the hazards of materials for emergency response.’ As the full name refers, the NFPA 704 is designed to provide hazard information for emergency responders to take proper action for different emergency cases. Figure 1 shows an example of the NFPA hazard diamond.

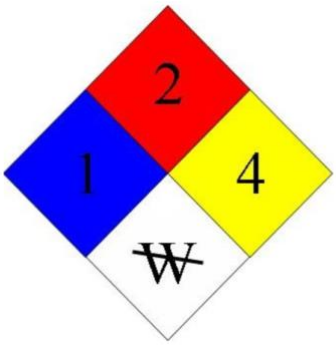


Figure 1. Example of NFPA hazard diamond [16]

As shown in Figure 1, the NFPA diamond [16] includes four divisions with different colors. Red, blue, and yellow divisions represent the degree of flammability (NF), health hazard (NH), and reactivity/instability (NR). Each colored division has a scale from 0 to 4, which 0 indicates a minimal hazard, and 4 indicates a severe hazard. The last white division is used for special hazards like W (unusual reactivity with water), OX (oxidizer), and SA (asphyxiant gas).

Overall, the NFPA diamond is a useful rating system, which provides sufficient hazard information for emergency responders. However, the NFPA diamond only represents the inherent hazard of a chemical without taking into account the quantity, which is also an important factor for storage. Moreover, this standard is intended to be used by emergency responders, in which the applicability to the public is limited.

2.2.3. Quantitative methods

Quantitative methods use mathematical techniques to develop a precise and easy-to-understand hazard identification method. It requires a large base of input data, but the feedback is more reasonable than the simple checklist. An excellent example of quantitative methods is the hazard index, and it will be discussed in detail in the next section.

2.3. Hazard indices

The storage of hazardous chemicals in a warehouse is a complex problem. The potential hazards include flammability, reactivity, and interaction among different types of hazardous chemicals. Hazard index is a helpful tool to identify and quantify the hazard in a facility or a process unit. Various hazard indices were developed in history. Dow's Fire and Explosion Index is the most famous and widely used one, and others such as Mond Index, Dow's Chemical Exposure Index, IFAL Index, Weighted Average Risk Rating Index, *etc.* are developed or modified based on different scopes and purposes [17].

2.3.1. Dow's Fire and Explosion Index (F&EI)

The first edition of Dow's F&EI was issued in 1964 and used within Dow Chemical Company. [18] After the development over half a century, F&EI has been widely used in Dow and outside Dow and became the leading hazard index recognized by the chemical industries. The purpose of the F&EI system is to:

1. Quantify the hazards of potential fire, explosion, and reactive chemicals incidents in pertinent process unit;
2. Identify which process or equipment would create or escalate the incident by comparing the different F&EI for separate process unit;
3. Achieve effective risk communication to process management team.

The procedure for calculating the F&EI is described in the following flowchart (Figure 2). Moreover, the detailed steps should be:

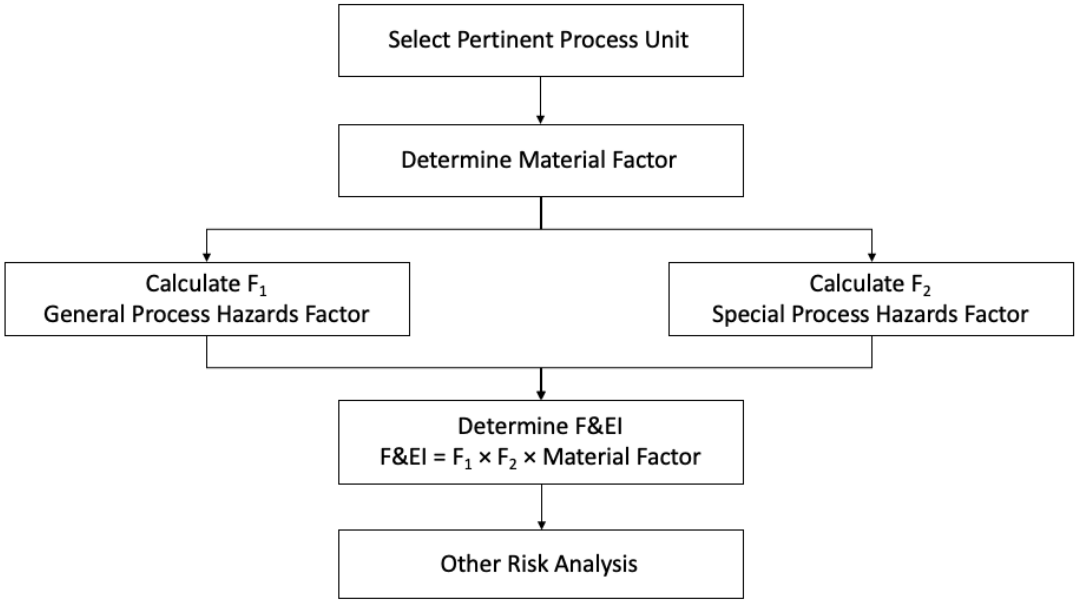


Figure 2. Procedure for calculating Dow's F&EI

1. The Pertinent Process Unit selection should consider the critical importance of the process, which would have the most significant impact on the magnitude of a potential fire or explosion.

2. The Material Factor (MF) should be determined for each process unit, and Dow Chemical Company develops its MF values.
3. The General Process Hazards Factor should be calculated by applying the appropriate penalties according to the F&EI Form. (Appendix A)
4. The Special Process Hazards Factor should be calculated by applying the appropriate penalties according to the F&EI Form. (Appendix A)
5. The F&EI should be calculated based on multiple Process Hazards Factors and MF.

2.3.2. MKOPSC's Potential to Cause Harm to the Public (PCHP)

In 2016, MKOPSC developed a hazard index called PCHP [19]. The objective of PCHP is to develop a methodology to assess chemical process facilities based on their potential to cause harm to the public. The PCHP index is a function of four factors: material hazards, quantities of chemicals, population densities, and accident history. Thus, the function is represented as follows in Equation 1.

$$PCHP\ Index = MHI * \prod F_i \qquad (Eqn. 1)$$

where material hazard index (*MHI*) represents the inherent hazard of the chemicals, and F_i represents the penalty value for i = quantity, population density, accident history.

2.3.3. Other hazard indices

The Mond fire, explosion, and toxicity index (Mond FETI) was developed as an extension to the Dow index [20] in 1985. As mentioned in the previous section, Dow's index considered mainly flammability and reactivity properties. However, Mond index also took toxicity into account and became a comprehensive safety index for process development.

The instantaneous fractional annual loss (IFAL) index was developed separately by the Insurance Technical Bureau for insurance assessment purposes. The main hazards accounted for this index are fires and explosions. However, in contrast to the Dow's F&EI and Mond FETI, the IFAL index is too complicated for manual calculation [17].

Other than mainly focusing on flammability properties, the Reactivity Risk Index (RRI) [21] was designed to recognize the hazards of reactive chemicals. Previous MKOPSC researchers measured chemical incompatibility data using Differential Scanning Calorimeter (DSC) and the Reactive System Screening Tool (RSST). The reactive risk can be expressed quantitatively by defining an RRI, as follows in Equation 2.

$$RRI = \tau \times \beta = \left(\frac{T_{process}}{T_{onset}}\right) \times \left(-\frac{\Delta H}{350}\right) \qquad (Eqn. 2)$$

where β is a thermodynamic quantity and is measured by the energy release potential of a substance, τ represents the probability of reaction occurrence.

2.4. Machine learning based method in process safety

Machine learning attracts much attention in recent years and has been applied in process safety in several aspects. Numerous works applied supervised learning to predict lower flammable limit (LFL), upper flammable limit (UFL), minimum ignition energy (MIE), and autoignition temperature [22-27]. Mage *et al.* utilized unsupervised learning to cluster the thermal stability of organic compounds into seven groups [28]. With the lack of study in liquid flammability considering aerosolization and the tendency of the machine learning approach, it is worthful to implement machine learning algorithms to liquid flammability rating.

2.4.1. K-Mean Clustering (KC)

The objective of KC [29] is to minimize the total within-cluster variation, which is shown in Equation 3. In the beginning, each of the observations is randomly assigned a number from 1 to K, which is called initial cluster assignments. The second step is to calculate the cluster centroid for each of the K clusters. The third step is to reassign each observation to the cluster whose centroid is the closest (defined using Euclidean distance). The second and third steps are repeated until the cluster assignments stop changing, which means the Equation 3 is satisfied.

$$\text{minimize}_{C_1,\ldots,C_k}\left\{\sum_{k=1}^{K}\frac{1}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}\left(x_{ij}-x_{i'j}\right)^2\right\} \qquad (Eqn.\,3)$$

where $C_1, \ldots, C_k$ represent cluster 1 to k, $|C_k|$ denotes the number of observations in the kth cluster, p is the number of features, and $\sum_{j=1}^{p}(x_{ij} - x_{i'j})^2$ is the Euclidean distance between two observations in the kth cluster.

Another graphical explanation (Figure 3) is easier to understand. In the following plot, six data points were chosen between ($0 < x < 6$) and ($0 < y < 4$). Firstly, each data point is assigned to a class number (red color refers to class I, and green color refers to class II), as shown in Figure 3(*b*). The number of data points in each cluster is decided by N (total number of data points) divided by K (number of clusters). In this example, N is six, K is two, and each cluster contains three data points. Then, the centroid of each class is calculated based on Euclidean distance between two observations in the kth cluster and is labeled as 'X' in each theme color in the plot. The next step is to calculate the distance from each data point to the centroid of each class. By comparing the result, each data point is reassigned to a class where the distance to the centroid is the closest, as shown in Figure 3(*c*). Finally, the local optimal result could be obtained after several iterations.

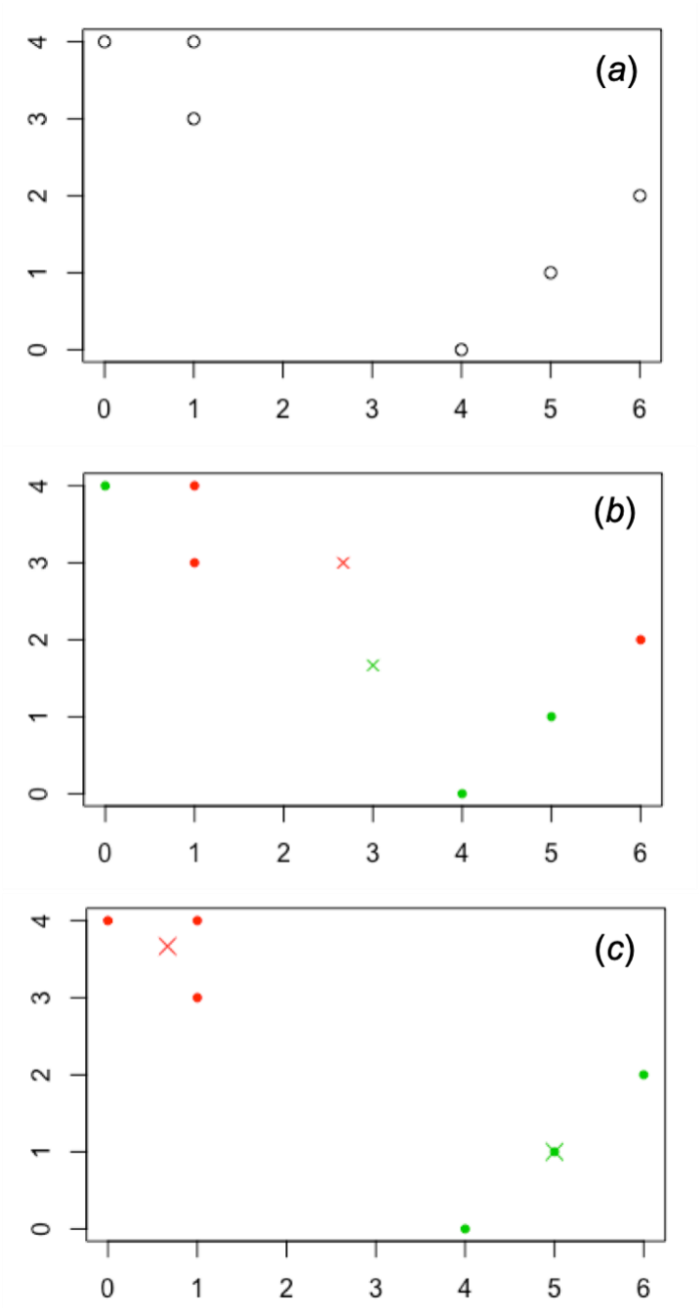


Figure 3. K-Mean clustering explanation: (*a*) Unclassified raw data; (*b*) Random classified data points and centroids; (*c*) Local optimal classification

2.4.2. Hierarchical Clustering (HC)

Hierarchical clustering [29] is an agglomerative clustering method that builds a dendrogram, starting from the leaves and combining clusters up to the trunk. In this method, it is necessary to specify the linkage in order to calculate the distance between two clusters. The common linkages are complete linkage, single linkage, average linkage, centroid linkage, and Ward linkage. The single linkage method calculates the distance between the closest elements in different clusters. Similarly, the complete method calculates the distance between the farthest clusters, and the average method finds the mean value for all pairwise distances. However, unlike the previous methods, Ward [30] suggested that for a large-scale (n > 100) dataset, the error sum of squares (ESS) should be calculated for each cluster, and the results should be compared by considering merging of the two clusters. This process should be repeated until the optimized solution is found. In this study, Ward linkage is applied. Equation 4 calculates the Euclidean distance between two clusters using Ward linkage.

$$\mathrm{dis(A,B)} = \sum_{i \in A \cup B} \|\vec{x}_i - \vec{m}_{A \cup B}\|^2 - \sum_{i \in A} \|\vec{x}_i - \vec{m}_A\|^2 - \sum_{i \in B} \|\vec{x}_i - \vec{m}_B\|^2 \qquad (Eqn.\,4)$$

where dis(A, B) is the Euclidean distance between cluster A and cluster B, and $\vec{m}_j$ is the center of cluster *j*.

Unlike the KC algorithm, the number of clusters is not required before conducting the clustering using the HC algorithm. As an example, Figure 4(*a*) shows the dendrogram

without truncation, and the dendrogram can be truncated with six leaves, as displayed in Figure 4(*b*).

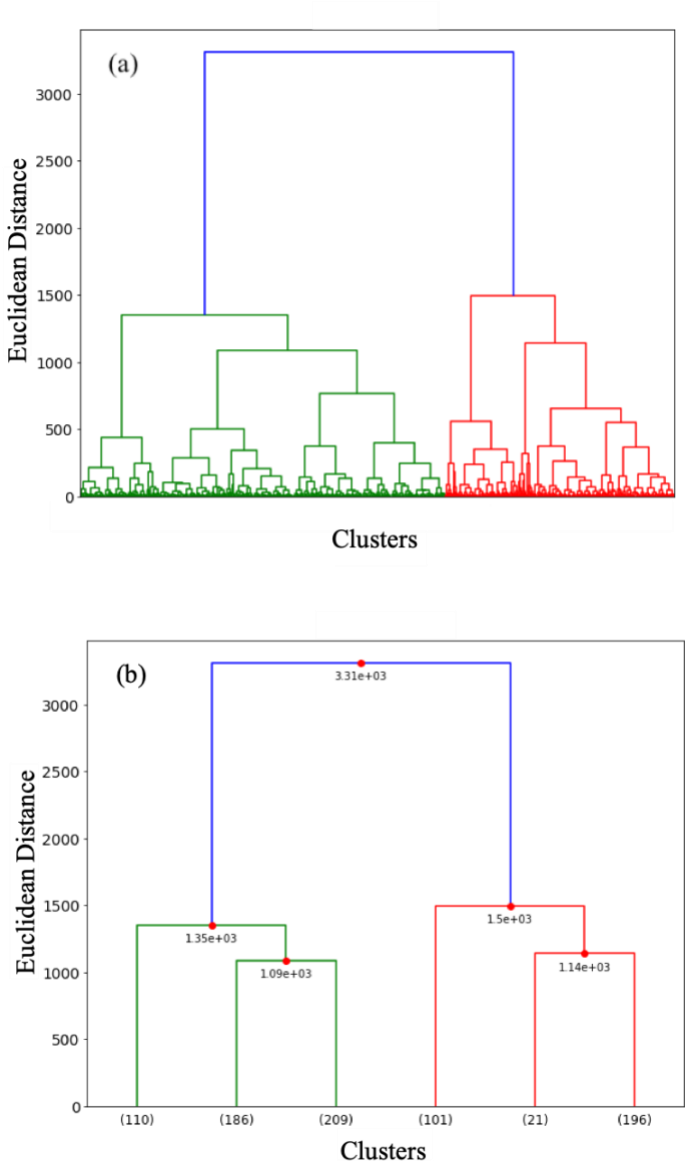


Figure 4. The dendrogram (a) without truncation, (b) with truncation.

2.4.3. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) [29] is a widely used method to reduce the dimension of features in a database. Before conducting PCA, the features should be normalized in the feature space, as shown in Equation 5.

$$\tilde{\phi}(x_i) = \phi(x_i) - \frac{1}{m}\sum_{k=1}^{m}\phi(x_k) \qquad (Eqn.5)$$

where $\phi(x_i)$ is the feature, and $\tilde{\phi}(x_i)$ is the normalized feature.

The covariance matrix is displayed in Equation 6, and the derivation of the projection of a new point onto the principal components is shown below.

$$\boldsymbol{C} = \frac{1}{p}\sum_{i=1}^{p}\tilde{\phi}(x_i)\tilde{\phi}(x_i)^T \qquad (Eqn.6)$$

The eigenvectors are:

$$\boldsymbol{C}\mathbf{v}_\mathrm{j} = \lambda_j \mathbf{v}_\mathrm{j}\,, \qquad j = 1, \ldots, N \qquad (Eqn.7)$$

Using the kernel function:

$$K(x_i, x_k) = \tilde{\phi}(x_i)\tilde{\phi}(x_i)^T \qquad (Eqn.8)$$

The eigenvectors can be written as a linear combination of normalized features:

$$\mathbf{v}_\mathrm{j} = \sum_{i=1}^{p} a_{ji}\tilde{\phi}(x_i) \qquad (Eqn.\,9)$$

Substituting Equation 6, 8 and 9 into Equation 7 results in:

$$\frac{1}{p}\sum_{i=1}^{p} \tilde{\phi}(x_i)\tilde{\phi}(x_i)^T \left(\sum_{l=1}^{p} a_{jl}\tilde{\phi}(x_l)\right) = \frac{1}{p}\sum_{i=1}^{p} \tilde{\phi}(x_i)\left(\sum_{l=1}^{p} a_{jl}K(x_i, x_l)\right) = \lambda_j \sum_{i=1}^{p} a_{jl}\tilde{\phi}(x_l)$$

$$(Eqn.\,10)$$

Multiplying both sides of Equation 10 by $\tilde{\phi}(x_k)^T$:

$$\frac{1}{p}\sum_{i=1}^{p} \tilde{\phi}(x_k)^T\tilde{\phi}(x_i)\left(\sum_{l=1}^{p} a_{jl}K(x_i, x_l)\right) = \lambda_j \sum_{i=1}^{p} a_{jl}\tilde{\phi}(x_k)^T\tilde{\phi}(x_l) \qquad (Eqn.\,11)$$

Simplifying:

$$\frac{1}{p}\sum_{i=1}^{p} K(x_k, x_i)\left(\sum_{l=1}^{p} a_{jl}K(x_i, x_l)\right) = \lambda_j \sum_{i=1}^{p} a_{jl}K(x_k, x_l) \qquad (Eqn.\,12)$$

The projection of a new point onto the principal components is:

$$\tilde{\phi}(\mathbf{x})^T\mathbf{v}_\mathrm{j} = \sum_{i=1}^{p} a_{ji}\tilde{\phi}(\mathbf{x})^T\tilde{\phi}(x_i) = \sum_{i=1}^{p} a_{ji}K(\mathbf{x}, x_i) \qquad (Eqn.\,13)$$

In this study, a radial basis function (RBF) kernel is utilized which provides a similarity measurement between two normalized feature vectors for use in the PCA, since RBF is a stable, isotropic, and infinitely smooth kernel. RBF kernel is widely used in classification and clustering studies. [31]

2.5. Aerosolization and flammability

The flammable and explosive hazards have been well studied. However, extra attention should be paid to the flammable and explosive hazards associated with aerosols, also known as mists or sprays, in a logistic warehouse and other storage facilities. In 1955, Eichhorn [32] reported that a liquid could not be ignited below its flash point, which means the flammable limit of the gas phase is clear. A clear UFL and LFL for vapor can be found in Figure 5, but the limit for aerosol (mist) is vague. Moreover, there are hundreds of liquids with the potential to aerosolize in the logistic warehouse and other storage facilities. [33, 34] Also, the consequences of an aerosol explosion have a so-called 'fuel-air bomb' effect. The unburnt droplets dispersed by the initial explosion will mix with air and cause a series of explosions. [35] Santon *et al*. conducted a survey and identified 33 incidents involve flammable aerosol, including lube oil, crude oil, kerosene, and naphtha. [36] Furthermore, the 2015 Buncefield fire and explosion is another evidence of detonation during an aerosol explosion. [37]

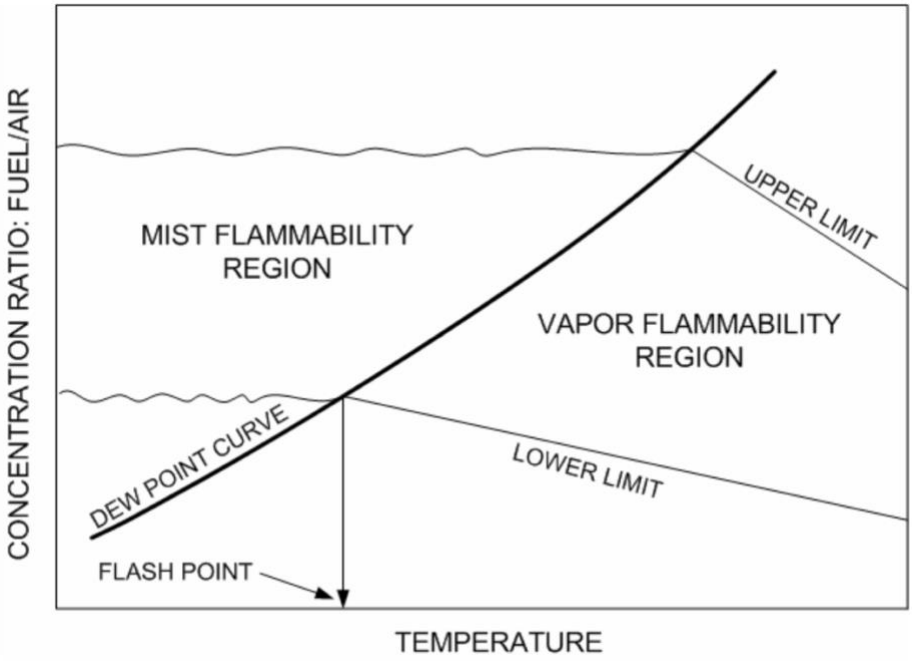


Figure 5. Schematic flammability diagram at constant pressure [32]

2.5.1. Probability of aerosolization

The probability of aerosolization can be determined as a function of viscosity and surface tension. Lefebvre *et al.* and Jasuja *et al*. reported that the Sauter Mean Diameter (SMD) of the droplet is proportional to viscosity and surface tension, as shown in Equation 14. [38, 39]

$$\mathrm{SMD} = \mathrm{A}\sigma^{0.25}\mu_L^{0.25}\rho_L^{0.125}d_0^{0.5}\rho_A^{-0.25}\Delta P_L^{-0.375} \qquad (Eqn.\,14)$$

where A is the coefficient, σ is the surface tension, μ_L is the liquid dynamic viscosity, ρ_L is the liquid density, d_0 is the orifice diameter, ρ_A is the air density, and ΔP_L is the pressure differential across the orifice.

Similarly, Elkotb *et al*. and Ingebo *et al*. also found the relationship between SMD and the liquid properties, as shown in Equation 15 and Equation 16. [40, 41] These three relationships support the methodology to estimate the tendency of a liquid to aerosolize based on viscosity and surface tension of the liquid.

$$\text{SMD} = 3.08 v_L^{0.385} (\sigma \rho_L)^{0.737} \rho_A^{0.06} \Delta P_L^{-0.135} \qquad (Eqn. 15)$$

$$\text{SMD} = 5.0 \left(\frac{\sigma \mu_L d_0^2}{\rho_A U_R^3 \rho_L} \right)^{0.25} \qquad (Eqn. 16)$$

where v_L is the kinematic viscosity of the liquid, σ is the surface tension, ρ_L is the liquid density, ρ_A is the air density, ΔP_L is the pressure differential across the orifice, μ_L is the liquid dynamic viscosity, d_0 is the orifice diameter, and U_R is the air velocity relative to liquid.

2.5.2. Physical properties related to aerosolization and flammability

2.5.2.1 Flash point

Flash point is the primary and most common physical property to measure the flammability. The flash point refers to the temperature where the liquid will generate enough vapor mixture with air and be ignited. However, the vapor mixture is inadequate to maintain constant combustion. Therefore, the ignition is temporary at its flash point. Generally, the flash point will increase with the increasing pressure.

Flash points may alter a few degrees based on different determination methods. The two commonly used methods are the open cup and closed cup. [42] Currently, the flash point is a commonly used criterion when classifying the degree of flammability.

In this thesis, the values of the flash point were obtained from the DIPPR 801 database with the unit of K.

2.5.2.2 Autoignition temperature

Autoignition temperature is another physical property related to flammability. Compared to the flash point determined from experimental data, which is a common standard for liquids and vapors, autoignition temperature can serve as a flammable parameter for solids.

Autoignition temperature is defined as a minimum temperature at which a flammable liquid or solid is capable of being ignited without any heat source from the surrounding environment. [43] The lower the autoignition temperature, the higher the probability of the substance can ignite itself.

In this thesis, the values of autoignition temperature were obtained from the DIPPR 801 database with the unit of K.

2.5.2.3 Flammable range

The flammable range represents a special range of compositions within which the fire or explosion will take place. To calculate this range, more knowledge about flammability limits is needed.

As shown in the fire triangle in Figure 6, three requirements are needed for a fire to occur. UFL is the upper flammable limit for a flammable mixture in air. Beyond this limit, the percent of oxygen is insufficient for a fire or explosion. In other words, the mixture is too rich for combustion. In contrast, LFL is the lower flammable limit. The mixture is also not combustible when there is not enough fuel; that is when the composition is lower than the LFL. Therefore, the mixture is combustible only within this flammable range.

In this thesis, the values of UFL/LFL were obtained from the DIPPR 801 database with no unit.

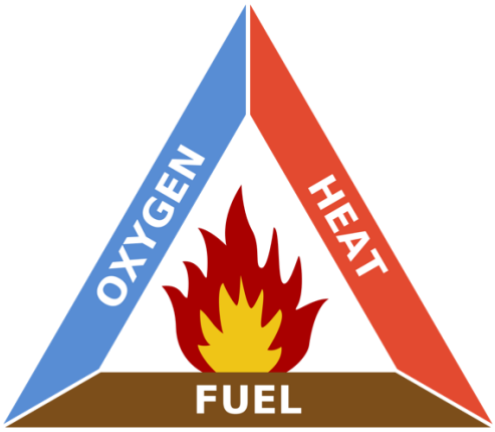


Figure 6. Fire triangle [42]

2.5.2.4 Viscosity

In the previous sections, the SMD is proportional to viscosity and surface tension. Viscosity reflects the resistance of a fluid to the flow and can be calculated by the ratio of the shearing stress to the velocity gradient in a fluid.

The SI unit of viscosity is the pascal second (Pa·s). However, this unit is rarely used in scientific and technical writing today. The centimeter–gram–second system of units (CGS) has defined poise (P) as dyne seconds per square centimeter (dyne·s/cm$_2$), which is named after the French physiologist Jean Poiseuille. [44] The relation between the SI unit and CGS unit is: 1 Pa·s = 10 P, 1 mPa·s = 1 cP.

In this thesis, the values of viscosity were obtained from the DIPPR 801 database with the unit of Pa·s and can be converted to cP for calculation.

2.5.2.5 Surface Tension

Surface tension is typically used to describe a phenomenon that happens on a liquid/gas interface. Unlike the liquid molecule inside the fluid, a molecule on the surface will not receive equal cohesive forces from the neighboring molecule. The fluid surface tends to shrink into the minimum surface area.

In this thesis, the values of surface tension were obtained from the DIPPR 801 database with the unit of mN/m.

3. METHODOLOGY

This chapter develops the methodology used in this thesis. Based on the literature review in Chapter 2, previous studies have developed several well-known indices in the past century. In general, every index is a function of the combination of inherent chemical properties and other factors that meets its purpose. In this chapter, we will develop the overall function for our index based on F&EI and then focus on the details about each parameter. For the degree of flammability, two machine learning based methods will be addressed.

3.1. Data collection

The Design Institute for Physical Properties (DIPPR) 801 is a project sponsored by AIChE, which provides more than 30 constant properties and nearly 50 thermophysical properties as well as molecular structures, hazard properties, and physical constants for more than 2000 compounds. This database is widely used in chemical property classification and prediction [22, 23]. After data cleaning, 823 organic compounds will be used in this research. The parameters used in this thesis are CAS No., flash point, autoignition temperature, UFL, LFL, viscosity, and surface tension.

3.2. Storage Hazard Factor (SHF)

Based on the literature reviews, various hazard indices are developed or modified based on different scopes and purposes. Considering that the index will be applied to the

chemical logistic warehouse, the overall index function can be represented as follows in Equation 17.

$$Logistic\ Warehouse\ Hazard\ Index\ (LWHI) = \sum F_i \times SHF \qquad (Eqn. 17)$$

where F_i represents different penalty factors such as quantity, population density, *SHF* represent the inherent hazard of the chemicals stored in the warehouse.

Based on the MKOPSC's PCHP project, the formula for calculating the SHF can be modified as follows in Equation 18.

$$SHF = 2^{Modified\ NF} + 2^{NR} + 2^{Modified\ NH} \qquad (Eqn. 18)$$

where *NR* represents the degree of reactivity, which will be determined by the original NFPA rating; *NH* represents the degree of the health hazard, which will be modified by PAC-3 value; *NF* represents the degree of flammability, which will be modified by two machine learning methods using DIPPR 801 database.

3.2.1. Modified NH

Protective Action Criteria (PAC) value is an exposure limit system, and this system is commonly used as the guideline for an emergency response to the concentration of the accidental release of the hazardous chemicals.

NH represents the degree of health hazard. The original NFPA rating criteria are based on LC_{50} and LD_{50}, which is more focused on the emergency conditions for the

working area. Since PAC-3 is the maximum airborne exposure resulting in the most severe consequence, which is life-threatening effects, PAC-3 is suitable for the modification of the NH value for our purpose.

The Department of Energy 's (DOE) current PAC dataset is Revision 29, which is published in May 2016 [45]. It provides chemical exposure limit values for 3146 chemicals. Table 1 provides the criteria used to modify the NH value.

Table 1. Modified NH determination guide

Min PAC-3 value (mg/m3)	**Max PAC-3 value (mg/m3)**	**Modified NH**
0	100	4
101	1000	3
1001	10000	2
10001	100000	1
100001	… or NA	0

3.2.2. Modified NF

As mentioned in Chapter 2, the most widely used chemical classification method is NFPA 704, GHS, and OSHA (29 CFR 1910.106). However, both of these criteria (Table 2) are based on flash points only [42]. Evidence shows that liquid can be ignited below its flash point if it is in some particular condition, such as aerosol form [36, 37]. In this research, flash point, autoignition temperature, surface tension, and viscosity are selected to modify the classification using KC and HC algorithm with PCA.

Table 2. Current standards for liquid flammability rating and classification

Standard	Flammability rating and classification	Criteria
NFPA 704	0	Materials will not burn in air when exposed to a temperature of 1500°F for a period of 5 minutes
	1	Flash point at or above 200°F
	2	Flash point between 100 and 200 °F
	3	Flash point between 73 and 100°F
	4	Flash point below 73°F
GHS classification and labeling of chemicals	1	Flash point < 23°C and boiling point ≤ 35°C
	2	Flash point < 23°C and boiling point > 35°C
	3	Flash point ≥ 23°C and ≤ 60°C
	4	Flash point > 60°C and ≤ 93°C

Table 2. Continued

Standard	Flammability rating and classification	Criteria
OSHA (29 CFR 1910.106)	1	Flash point < 73.4°F and boiling point ≤ 95°F
	2	Flash point < 73.4°F and boiling point > 95°F
	3	Flash point ≥ 73.4°F and ≤ 140°F. When a category 3 liquid with a flash point ≥ 100°F is heated for use to within 30°F of its flash point, it shall be handled in accordance with the requirements for a Category 3 liquid with a flash point < 100°F.
	4	Flash point > 140°F and ≤ 199.4°F. When a category 4 liquid is heated for use to within 30°F of its flash point, it shall be handled in accordance with the requirements for a Category 3 liquid with a flash point < 100°F.
	5	When a liquid with a flash point > 199.4°F is heated for use to within 30°F of its flash point, it shall be handled in accordance with the requirements for a Category 4 flammable liquid.

In this study, the KC and HC algorithm is implemented through the Python package, Scikit-Learn [46]. The number of clusters is determined by the elbow method, which plots the within-cluster sum of square (WCSS) with respect to the number of

clusters [47]. Figure 7 shows the example of the elbow plot when implementing the KC algorithm on liquid flammability clustering based on flash point and autoignition temperature. The number of clusters is five in this thesis.

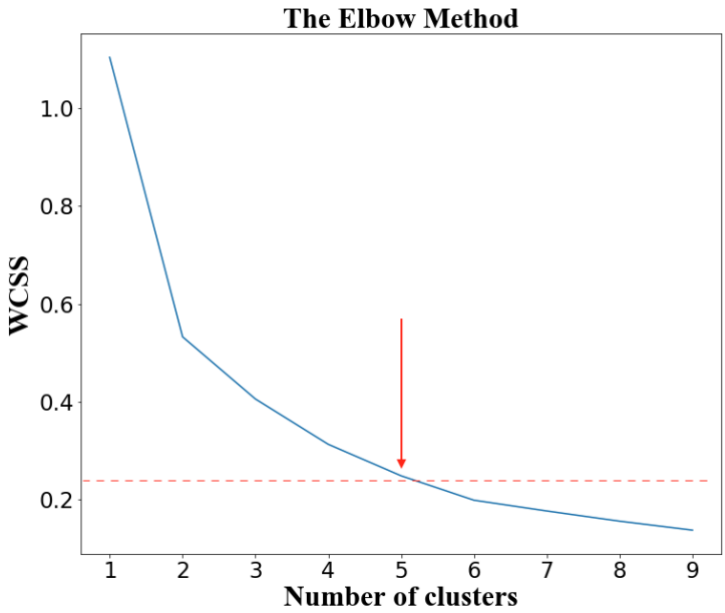


Figure 7. Within-cluster sum of square (WCSS) and the number of clusters

This modification method is reliant on the availability of the data. Despite the lack of data, the original NFPA rating with simple update (if UFL – LFL > 10%, then NF + 1 with a maximum of 4) can be used for SHF calculations.

3.3. Penalty factors

The other important part of Equation 17 is $\sum F_i$, which represents different penalty factors. In this study, quantity, population density, and distance to the nearest fire station are selected to be the penalty factors. The determination guides for each factor are described in this section.

3.3.1. Quantity

Quantity is an important factor that should be considered first when designing a hazard index. Besides the inherent hazard of a hazardous chemical, the amount of chemicals stored in the facility also reveals the level of hazard.

For example, facility A has a chemical X stored onsite with a high SHF and a small amount. However, facility B has a chemical Y stored onsite with a moderate SHF, but the quantity is enormous. When considering the hazard index, we should balance the inherent hazard and quantity. In this simple case, facility B is also dangerous, although chemical Y is inherently safer than chemical X.

The following table shows the determination guide of the quantity penalty value (Table 3).

Table 3. Penalty value of quantity determination guide

Original code	Min (Pounds)	Max (Pounds)	Penalty value
1	0	99	1.2
2	100	499	1.4
3	500	999	1.4
4	1,000	4,999	1.6
5	5,000	9,999	1.6
6	10,000	24,999	1.8
7	25,000	49,999	1.8
8	50,000	74,999	1.8
9	75,000	99,999	1.8
10	100,000	499,999	2
11	500,000	999,999	2
12	1,000,000	9,999,999	2
13	10,000,000	…	2

3.3.2. Population density

Besides the inherent hazard of a chemical and the quantity stored in the facility, another important factor is the safety impact to the public. Given the coordinate of a facility, we defined the population in a radius of two miles near the facility that can be used to represent the population density factor in Equation 17.

Population density information is retrieved on LandView 6.0 [48], a geographic information system software. The following table shows the determination guide of population density penalty value (Table 4).

Table 4. Penalty value of population density determination guide (in a radius of two miles near the facility)

Min	Max	Penalty value
10	100	1.2
101	1000	1.4
1001	10000	1.6
10001	100000	1.8
100001	…	2

3.3.3. Distance to the nearest fire station

The last factor we choose in this study is 'distance to the nearest fire station.' In the previous sections, we considered the inherent hazard, quantity, and the potential impact to the public. Furthermore, last but not least, we choose a factor that can reflect the mitigation process, which is an essential point for a storage facility.

Distance to the nearest fire station (FS) is retrieved from HazardHub, a provider of property-level hazard risk database [49]. The following table shows the determination guide of distance to the FS penalty value (Table 5).

Table 5. Penalty value of distance to FS determination guide

Min (Miles)	Max (Miles)	Penalty value
0	1	1.2
1.01	2	1.4
2.01	3	1.6
3.01	4	1.8
4.01	…	2

3.3.4. Accident history

The accident history of a facility is another indicator of the performance of chemical safety. This information can reflect part of the safety management of the facility and how well it will perform in the potential chemical incidents. Since the accident history is not available so far, this thesis does not consider this factor in the LWHI calculations.

4. RESULTS AND DISCUSSION

This chapter provides the results obtained with the application of the proposed methodology. The analysis is based on calculations of database DIPPR 801 from AIChE. This chapter discussed mainly the calculation of SHF and the modification of parameters. The first section discussed a novel classification method based on the machine learning approach. In the following sections, some statistical techniques are used to test the behavior of SHF mathematically. The calculation of LWHI contains other factors that are related to real data, and it will be addressed and discussed in the next chapter.

4.1. NF modification

4.1.1. Database visualization

Before conducting the liquid flammability rating with the inclusion of aerosolization, we would like to investigate the distribution of observations for each liquid property, and the scatter plots of each pair of liquid properties. Appendix B shows the aggregated scatter plots, distribution plots, and heatmap of liquid properties. The diagonal of Appendix B shows the distribution of observations for each liquid property. For example, Figure 8 from Appendix B shows that the distributions of flash point and surface tension are normally distributed. However, the distributions of autoignition temperature and viscosity are right-skewed.

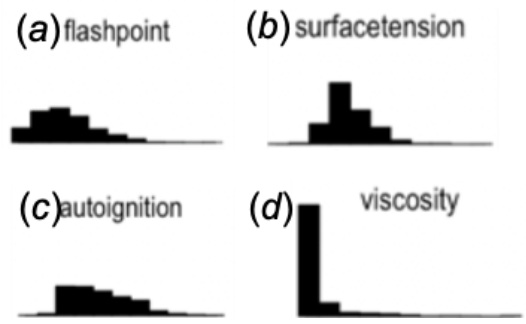


Figure 8. Distributions of liquid properties: (*a*) flash points; (b) surface tension; (c) autoignition temperature; (d) viscosity

The left part of Appendix B is the scatter plot of each pair of liquid properties. For example, a positive slope is plotted for the relationship between flash point and molecular weight, shown in the 5th plot from the top in the first column on the left part of Appendix B, and the magnified plot in Figure 9.

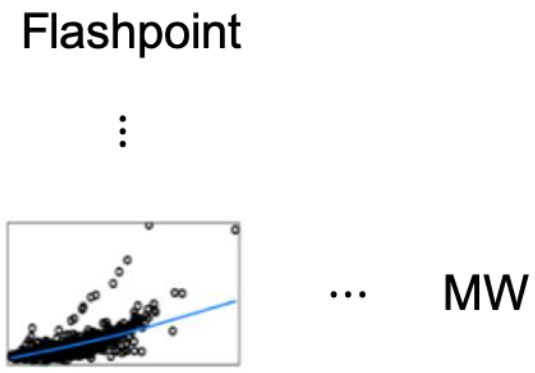


Figure 9. Scatter plot of flash point and molecular weight

The right part of Appendix B shows the statistical correlation between each pair of liquid properties, including Pearson coefficient (ρ), Kendal coefficient (τ), Spearman coefficient (r), and the P-value for Pearson coefficient (p). For example, the statistical correlation between flash point and vapor pressure is found in the upper rightmost location. The Pearson coefficient between flash point and vapor pressure is -0.99, which means an entirely negative correlation.

4.1.2. KC and HC algorithm

As discussed before, the number of clusters determined by the elbow method is five in this thesis. Thus, the 823 organic compounds from DIPPR 801 were split into five groups and rated from 0 to 4 as in the NFPA rating. The KC clustering is based on flash point and autoignition temperature, which is different from the NFPA rating. The compounds in the group with a rating of 4 are the compounds with the highest flammability. On the other hand, compounds with a rating of 0 have the lowest flammability. Figure 10 shows the data distribution.

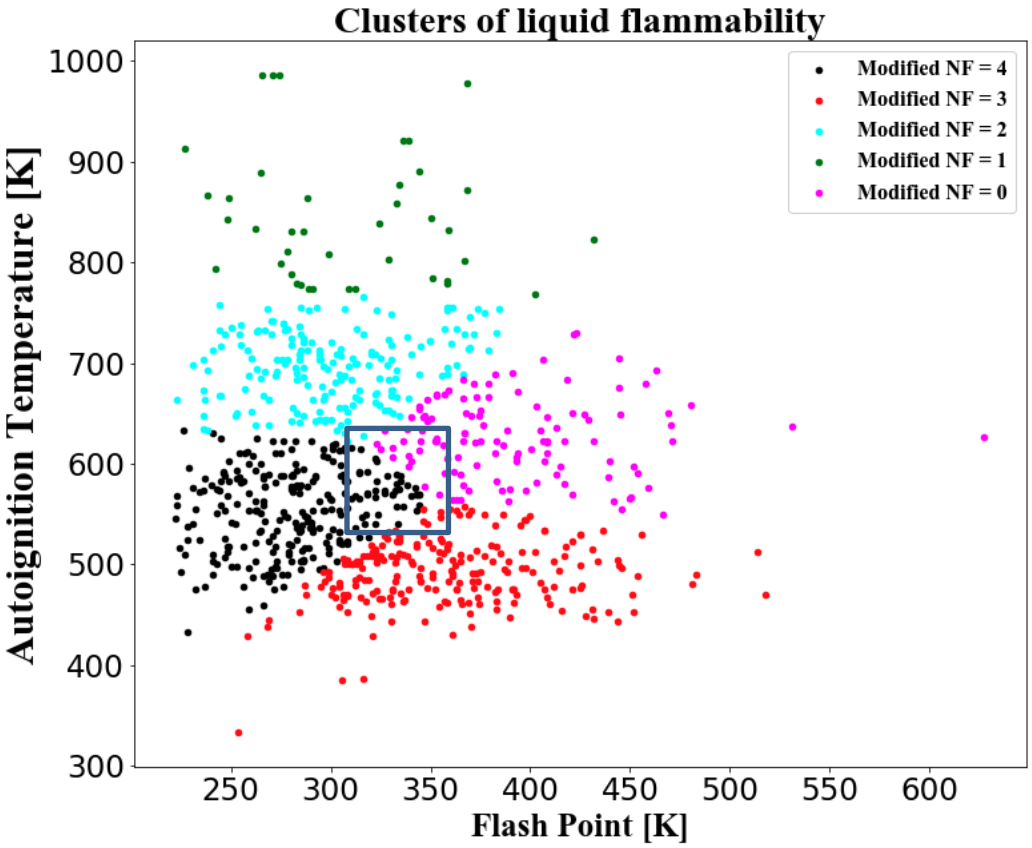


Figure 10. Clusters of liquid flammability using KC algorithm

In Figure 10, some data points labeled black have a medium flash point, and medium autoignition temperature comparing to the neighbor points labeled red and cyan. Those points either have a high flash point and low autoignition temperature, or have a low flash point and high autoignition temperature. However, the black label means NF = 4, which is higher than the red (NF = 3) and cyan (NF = 2) label. Similar results and doubts show at the boundary of different clusters in the circled area.

Similarly, Figure 11 displays the dendrogram of clustering through the HC algorithm. Also, 823 organic compounds from DIPPR 801 are split into five groups and rated from 0 to 4 as the same criteria with the HC algorithm.

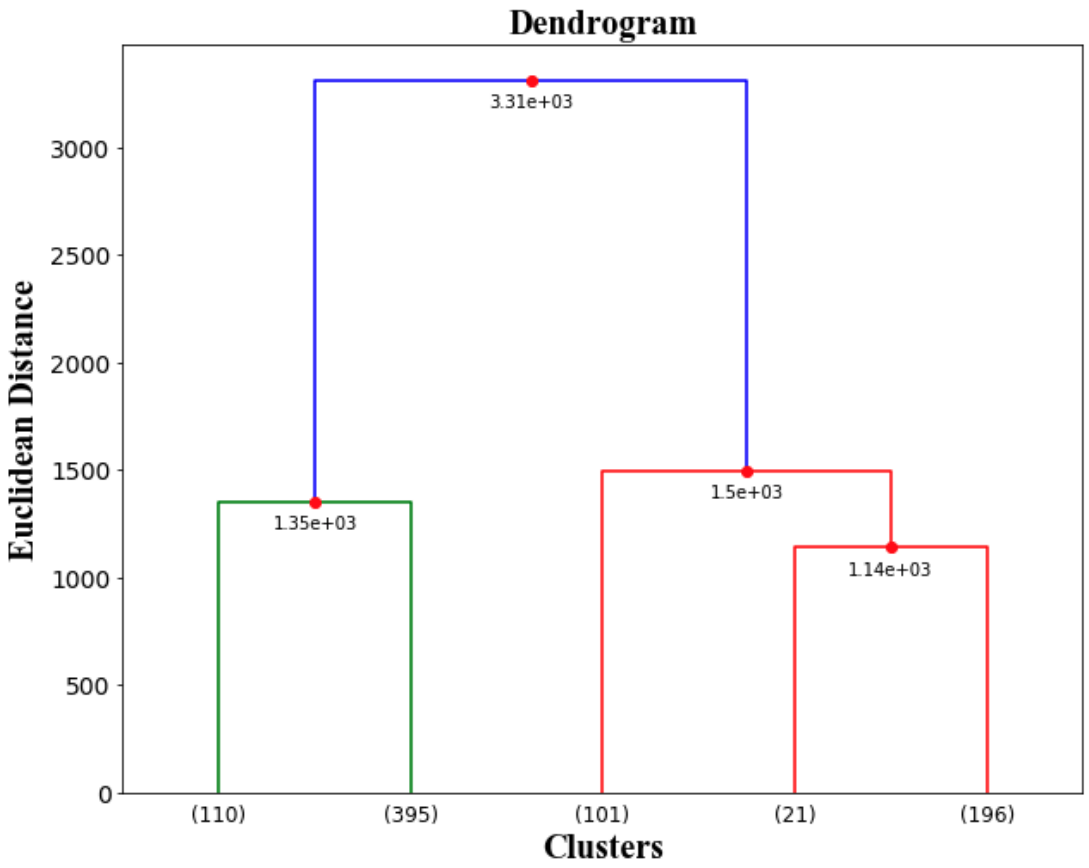


Figure 11. Truncated dendrogram of liquid flammability using HC algorithm

The agglomerative clustering results will be assigned to each data point. Figure 12 shows the visualized plot in Cartesian coordinates.

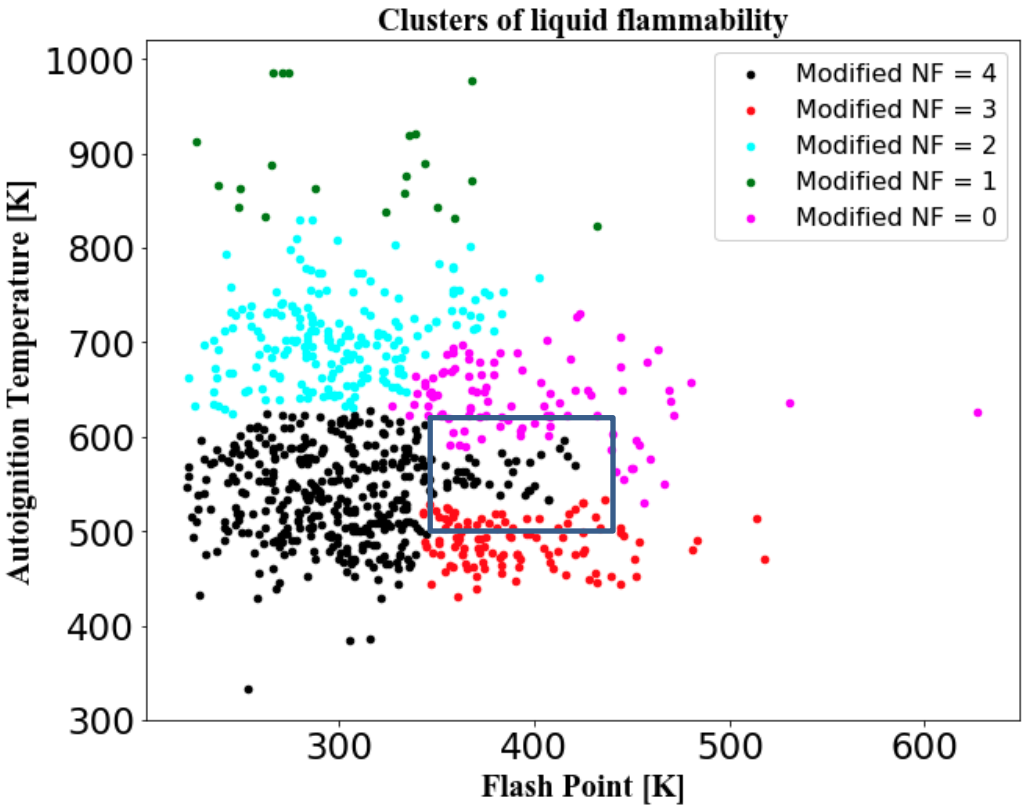


Figure 12. Clusters of liquid flammability using HC algorithm

In Figure 12, the results located in the controversial boundary between the black (NF = 4) and the red (NF = 3) regions are more reasonable. However, this time, a misclassification may happen in the circled area. With a similar flash point, the black labeled data points have the medium autoignition temperature comparing to the red and magenta labeled data points. However, these data points were classified as NF = 4, which is the most hazardous material among all. On the other hand, some points with the lower

autoignition temperature were classified as NF = 3, which is less dangerous than black labeled points.

Comparing the KC and HC algorithms, there are 653 out of 823 compounds with the same rating for liquid flammability in both algorithms. Table 6 shows the liquids with significantly different ratings between the two algorithms. Those liquids in Table 6 require more attention when conducting a risk assessment with inherent flammability.

Table 6. Liquids with significant different ratings between KC and HC algorithm

Substance name	**Flammability rating (KC)**	**Flammability rating (HC)**
o-ethylaniline	2	0
hexylene glycol	0	4
cetyl methacrylate	3	0
3-methyl-1-pentene	4	2
1-dodecanol	3	4
4-methyl-1-octanol	0	4

As a result, the KC algorithm has a more reasonable rating for the clustering of liquid flammability since the circled area is smaller in Figure 10 compared to Figure 12. Another reason is that the misclassification in the KC algorithm is more likely to happen on the boundary of two clusters, whereas the misclassification in the HC algorithm is more likely to happen in an area. Both results with KC and HC algorithm are considering the

flash point and autoignition temperature in two dimensions. Therefore, the results are highly interpretable since the X and Y axis both have physical meaning.

However, to consider liquid aerosolization probability at the same time, it is necessary to reduce the features for visualization and more straightforward calculation. The PCA method will be applied in the next section.

4.1.3. PCA with KC and HC algorithm

The main purpose of NF modification is to consider aerosolization. In the previous chapter, we conclude that viscosity and surface tension can be used as two indicators of aerosolization. To reduce the flash point, autoignition temperature, viscosity, and surface tension into two principal components (PCA1 and PCA2), the RBF kernel function was applied when reducing four features. Another advantage is that PCA does not need to specify the weight of contributions of liquid aerosolization and flammability. Figure 13 shows the clustering results by the KC and HC algorithm based on PCA1 and PCA2. Figure 14 shows the dendrogram of the HC algorithm. Besides the advantages of PCA, one thing that needs to keep in mind is that both X and Y axes in Figure 13 have no physical meaning. This is the main disadvantage of the PCA method.

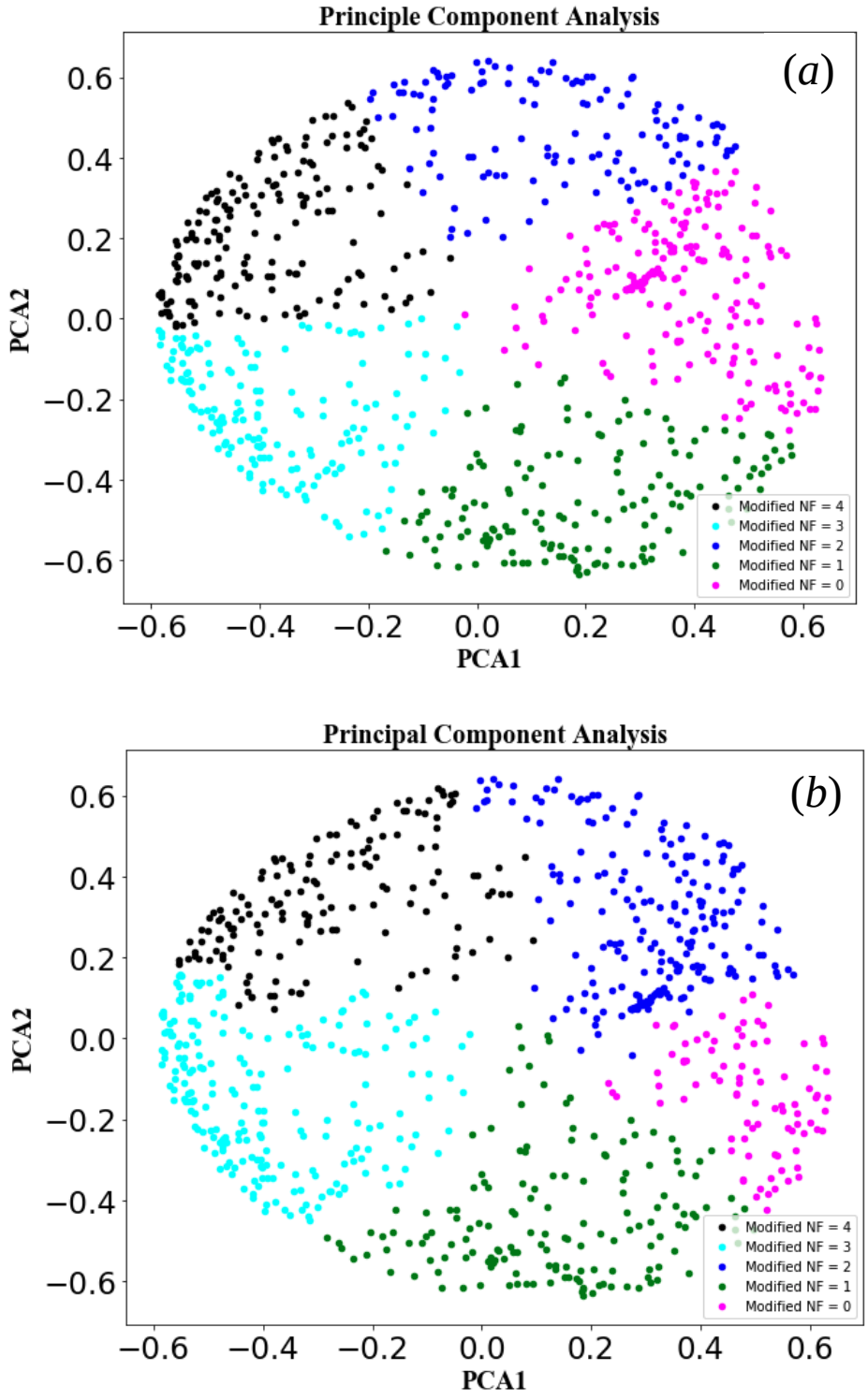


Figure 13. Principal component (PCA1 and PCA2) clusters using (a) KC algorithm (b) HC algorithm.

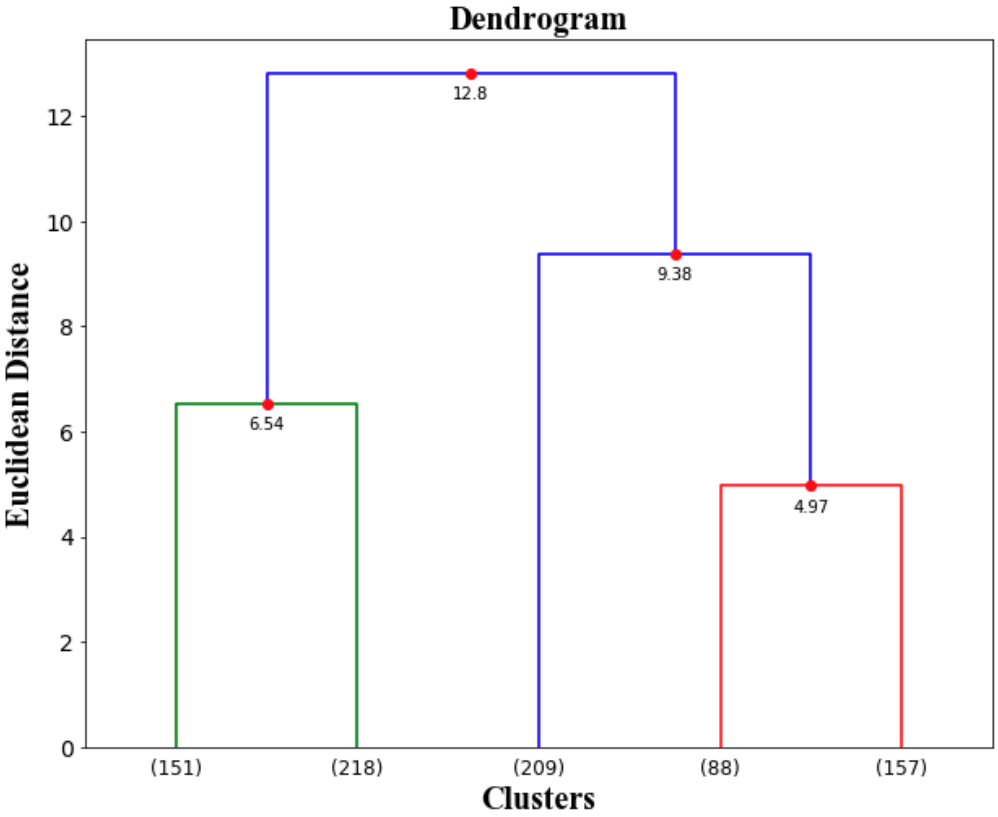


Figure 14. Truncated dendrogram of liquid flammability using HC algorithm with PCA

4.2. Sample SHF calculations

After updating each parameter, three SHF sample calculations of different chemicals are shown below.

4.2.1. Diisobutyl ketone

Table 7 is the sample hazard review for diisobutyl ketone from DIPPR 801, DOE's PAC, and NFPA database.

Table 7. Hazard review: Diisobutyl ketone

Parameter	**Data**
CAS No.	108-83-8
NFPA NR	0
NFPA NH	1
PAC-3	12000 mg/m3
NFPA NF	2
Flash point	322.15 K
Autoignition temperature	669.15 K
LFL/UFL	0.8% / 6.2%
Viscosity	0.896 cP at 25 °C
Surface tension	23.9 mN/m at 25 °C

Based on the data above, the SHF should be:

1. The original NFPA NR rating is 0 and kept the same in our calculation.
2. The original NFPA NH rating is 1, and the PAC-3 value is 12000 mg/m3. Based on Table 1 in the last chapter, the modified NH value is still 1.
3. The original NFPA NF rating is 2, the flash point is 322.15 K, autoignition temperature is 669.15 K, LFL / UFL are 0.8% / 6.2%, viscosity is 0.896 cP, and surface tension is 23.9 mN/m. The results of the machine learning method with both KC and HC algorithms are both 2. So the NF remains unchanged.
4. For this chemical, the original NFPA rating does not change.
5. $SHF = 2^2 + 2^0 + 2^1 = 7$

4.2.2. 2-Butoxyethanol

Table 8 is the sample hazard review for 2-Butoxyethanol from DIPPR 801, DOE's PAC, and NFPA database.

Table 8. Hazard review: 2-Butoxyethanol

Parameter	**Data**
CAS No.	111-76-2
NFPA NR	0
NFPA NH	2
PAC-3	3400 mg/m3

Table 8. Continued

Parameter	**Data**
NFPA NF	2
Flash point	334.15 K
Autoignition temperature	511.15 K
LFL/UFL	1.1% / 12.7%
Viscosity	2.9 cP at 25 °C
Surface tension	26.1 mN/m at 25 °C

Based on the data above, the SHF should be:

1. The original NFPA NR rating is 0 and kept the same in our calculation.
2. The original NFPA NH rating is 2, and the PAC-3 value is 3400 mg/m3. Based on Table 1 in the last chapter, the modified NH value is still 2.
3. The original NFPA NF rating is 2, the flash point is 334.15 K, autoignition temperature is 511.15 K, LFL / UFL are 1.1% / 12.7%, viscosity is 2.9 cP, and surface tension is 26.1 mN/m. The result of the machine learning method using KC algorithm is 3, but the result becomes 4 when using HC algorithm. As we discussed in the previous section, the result with KC algorithm is more reasonable. So the modified NF value is 3.
4. $SHF = 2^3 + 2^0 + 2^2 = 13$

4.2.3. N, N-Dimethylcyclohexylamine

Table 9 is the sample hazard review for N, N-Dimethylcyclohexylamine from DIPPR 801, DOE's PAC, and NFPA database.

Table 9. Hazard review: N, N-Dimethylcyclohexylamine

Parameter	**Data**
CAS No.	98-94-2
NFPA NR	0
NFPA NH	3
PAC-3	66 mg/m3
NFPA NF	2
Flash point	312.15 K
Autoignition temperature	NA
LFL/UFL	3.6% / 19%
Viscosity	NA
Surface tension	NA

Based on the data above, the SHF should be:

1. The original NFPA NR rating is 0 and kept the same in our calculation.
2. The original NFPA NH rating is 3, and the PAC-3 value is 66 mg/m3. Based on Table 1 in the last chapter, the modified NH value is 4.

3. The original NFPA NF rating is 2, the flash point is 312.15 K, LFL / UFL are 3.6% / 19%, but the autoignition temperature, viscosity, and surface tension data are not available for DIPPR 801 database. The machine learning methods with both KC and HC algorithms can not be applied to this chemical. In alternative, the flammable rage is 19% - 3.6% = 15.4% > 10%. So the modified NF is 2 + 1 = 3.
4. For this chemical, both NH and NF change with our method. Especially for NH, the PAC-3 value of 66 mg/m3 is a potential hazard to public.
5. $SHF = 2^3 + 2^0 + 2^4 = 25$

Those three examples show SHF calculation under different conditions. With the method mentioned above, there are 170 SHFs for different chemicals that can be calculated from the original database and used for the case study in the next chapter. The full list of 170 chemicals SHF is in Appendix C.

4.3. Statistical analyses

In this section, statistical techniques are used to test the behavior of SHF mathematically. Some significant statistical results are selected to prove that the SHF method is applicable to the hazard index. The following three subsections will describe the descriptive statistics, the histogram of 170 chemicals, and the percentile analysis. The purpose of this section is to validate the SHF for a broader application.

4.3.1. Descriptive statistics

With the proposed SHF methodology, 170 different chemicals were calculated to obtain the SHF value. Table 10 shows the a summary of the descriptive statistics for this sample.

Table 10. Summary of descriptive statistics for SHF

Statistical parameter	**Value**
Mean	13.75
Median	12
Mode	11
Standard Deviation (SD)	5.90
Standard Error of Mean (SEM)	0.45
Range	45
Minimum	3
Maximum	48
Counts	170

From the statistics above, the mean value of this dataset is 13.75, the median is 12, and the mode is 11. Compared with the range from 3 to 48, the results indicate that the distribution of the dataset is right-skewed since mode < mean. However, the mean and

median are close to each other, which indicates the dataset is nearly normal distribution shaped. The visualization of a histogram will further verify this conclusion.

Moreover, the standard deviation of this dataset is 5.90, which is not small comparing to the total range. But the standard error is only 0.45, which means the sample statistics can precisely approximate the population.

4.3.2. Generation of frequency histogram

As mentioned before, Figure 15 is the histogram of 170 SHF values for different chemicals.

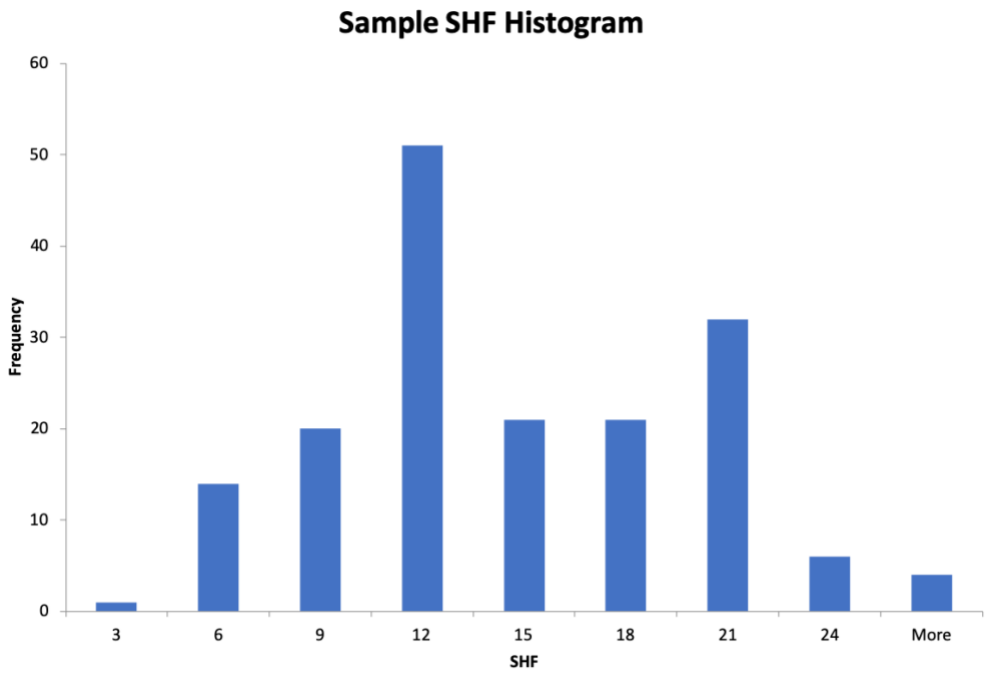


Figure 15. Sample SHF histogram

This histogram shows that the majority of chemicals in this sample have a SHF value around 12, which is $2^2 + 2^2 + 2^2$, indicating most chemicals have the intermediate

hazard level. The second peak of SHF value is located in the range of 18 to 21. Those chemicals may need extra attention to storage.

4.3.3. Percentile analysis

For this dataset, the percentile analysis results can be used to determine what percentage of data lies within a particular value of the SHF. This is another way to show the distribution of the dataset. The percentile analysis is conducted and summarized in Table 11.

Table 11. Percentile analysis of the sample SHF

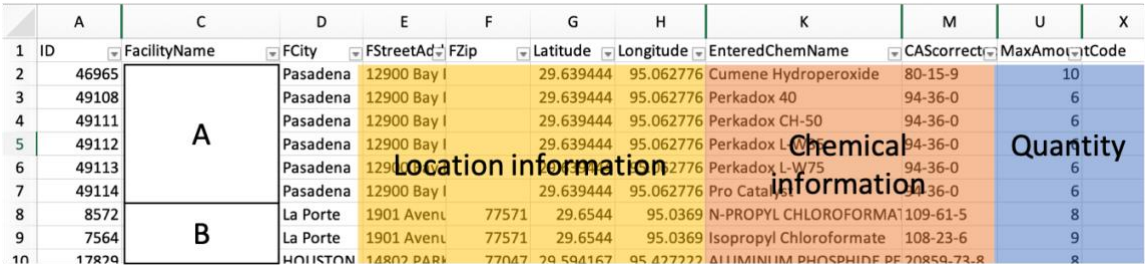
Percentiles	**SHF range**
25%	[3, 11]
50%	[3, 13]
75%	[3, 19]
90%	[3, 21]
100%	[3, 48]

The above analysis indicates that the majority of the data (over 50%) falls in the range of 3 to 13, which is the same as the result we got from the histogram. Both histogram and percentile analysis suggest that the majority of chemicals used for SHF have the value in the range of 3 to 13.

5. CASE STUDY

5.1. Data collection

Houston Chronicle has published a series of articles [50], aiming at exploring fatal mistakes that could have the most significant consequences and probes that put the citizen in jeopardy. Houston Chronicle has collected 2581 facilities and over 18000 chemical records in the greater Houston area. The raw data is in EPA Tier II standard and shared with MKOPSC. After data cleaning, at least 33 warehouses that have more than 400 records and over 170 kinds of hazardous chemicals will be used in this research. The raw database (Figure 16, and all company names is blacked out for privacy) includes company information, location information, chemical information, and storage quantity.



	A	C	D	E	F	G	H	K	M	U	X
1	ID	FacilityName	FCity	FStreetAd	FZip	Latitude	Longitude	EnteredChemName	CAScorrect	MaxAmo	tCode
2	46965		Pasadena	12900 Bay I		29.639444	95.062776	Cumene Hydroperoxide	80-15-9	10	
3	49108		Pasadena	12900 Bay I		29.639444	95.062776	Perkadox 40	94-36-0	6	
4	49111		Pasadena	12900 Bay I		29.639444	95.062776	Perkadox CH-50	94-36-0	6	
5	49112		Pasadena	12900 Bay I		29.639444	95.062776	Perkadox L-W75	94-36-0	6	
6	49113		Pasadena	12900 Bay I		29.639444	95.062776	Perkadox L-W75	94-36-0	6	
7	49114		Pasadena	12900 Bay I		29.639444	95.062776	Pro Catalyst	94-36-0	6	
8	8572		La Porte	1901 Avenu	77571	29.6544	95.0369	N-PROPYL CHLOROFORMA	109-61-5	8	
9	7564		La Porte	1901 Avenu	77571	29.6544	95.0369	Isopropyl Chloroformate	108-23-6	9	
10	17829		HOUSTON	14802 PARK	77047	29.594167	95.427222	ALUMINUM PHOSPHIDE PE	20859-73-8	8	

Figure 16. Example of raw data from the Houston Chronicle

5.2. Sample LWHI calculation

With the chemical information, SHF value can be obtained by using the method discussed above, and the storage quantity information can be converted into units in pounds. Furthermore, population density and distance to the nearest fire station can be extracted

from the location information. The 2-butoxyethanol example from the last chapter will be expanded with other information to help calculate the LWHI. (Table 12)

Table 12. Tier II information for 2-butoxyethanol in facility #33

Parameter	**Data**
CAS No.	111-76-2
SHF	13
Quantity	4 (original code)
Population density	4154 (in a radius of two miles)
Distance to FS	1.62 miles

Based on the data above, the LWHI should be:

1. SHF for 2-butoxyethanol is 13 based on the calculation from the last chapter.
2. For facility #33, the quantity indicator of 2-butoxyethanol is 4 (1000 to 5000 pounds range), and the penalty value is 1.6 based on Table 3.
3. For facility #33, the population density in a radius of two miles is 4154, and the penalty value is also 1.6 based on Table 4.
4. For facility #33, the distance to the nearest fire station is 1.62 miles, and the penalty value is 1.4 based on Table 5.
5. Therefore, the LWHI for 2-butoxyethanol in facility #33 is:

   $13 \times 1.6 \times 1.6 \times 1.4 = 46.592$.

5.3. Statistical analyses

Following the previously mentioned methodology, the LWHI is calculated for each facility. Since a facility can have multiple chemicals with variable quantities, the LWHI is first calculated for each record. Figure 17 shows the distribution of LWHI for each record.

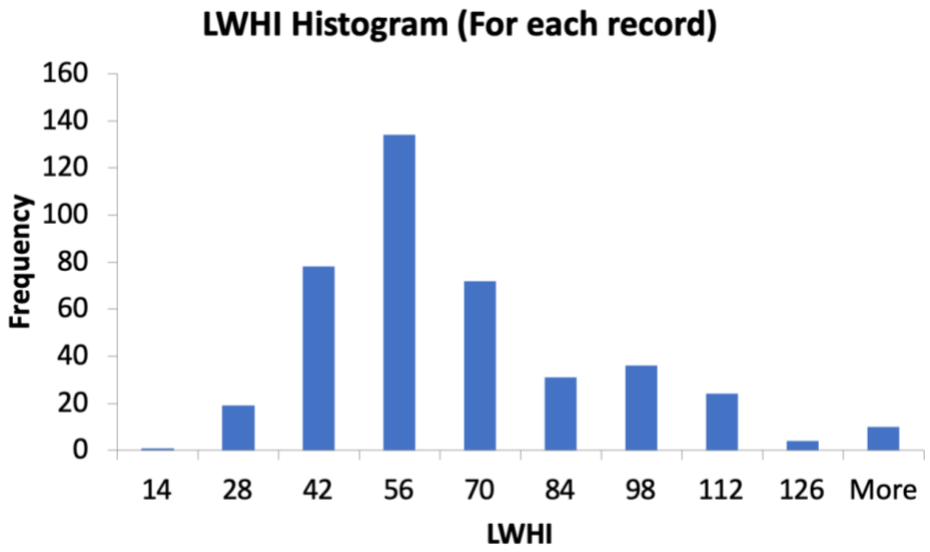


Figure 17. LWHI histogram (for each record)

In the figure above, the dataset of LWHI for each record is near to normal distribution, suggesting that the majority of the chemicals have a LHWI for the intermediate level. The findings in SHF distribution can prove this similar conclusion. With this histogram, the manager can pay more attention to the chemicals which have a higher value in LWHI.

If we average out all records for each facility, another histogram (Figure 18) shows a similar trend. Most facilities have the medium level LWHI, and regulatory authorities

can focus more on those facilities that have higher LWHI values. For example, the government can make more effective hazard communication with the surrounding public, and the manager can arrange more resources (*i.e.*, fire brigades, specialty clinics, *etc.*) near the facility when developing the emergency response plan.

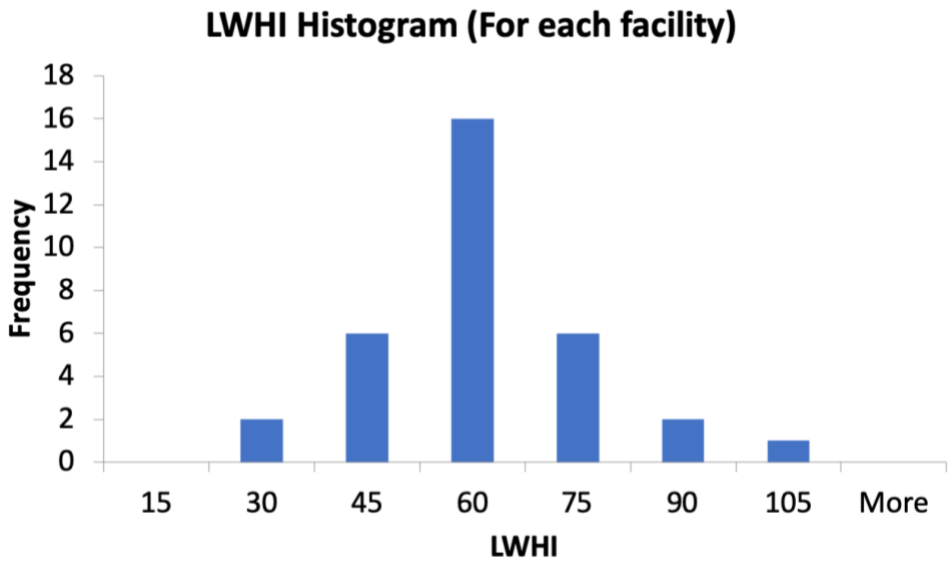


Figure 18. LWHI histogram (for each facility)

6. CONCLUSIONS AND FUTURE WORK

In this thesis, a hazard index for the hazard identification of chemical logistic warehouses was created and named LWHI. This index aims to to calculate the potential hazards in a logistic facility numerically. These results can be utilized by the manager or the emergency responder to develop their hazard chemicals management plan.

To reach the goal mentioned above, the SHF was introduced to the index. First, two machine learning based methods for liquid flammability rating with the consideration of aerosolization have been proposed. The first method applies KC and HC algorithms in machine learning to chemical classification for their high interpretability. The 823 organic compounds in DIPPR 801 are clustered into five groups based on their flash point and autoignition temperature. The five groups regarding liquid flammability are then rated from 0 to 4, where 4 is the most hazardous rating. With the analysis mentioned in Chapter 4, the KC algorithm has a more reasonable rating on liquid flammability clustering.

The second method presented uses PCA to reduce the four features (*i.e.*, flash point, autoignition temperature, viscosity, and surface tension) into two principal components (PCA1 and PCA2). The advantage of the PCA rating method is that the weight of the contribution of the four features is automatically considered. Admittedly, the lack of interpretability is a disadvantage of the PCA method as the principal components do not have physical significance but only statistical significance. However, compared with traditional flammability classification methods, which only rely on flash point and boiling point, the two proposed methods have shown a statistical correlation

with liquid flammability. Additionally, one distinct disadvantage of traditional flammability classification methods is that the threshold values are determined by human, which invariably has a bias, while machine learning based methods partly eliminate this bias. Also, the boundary of traditional flammability classification methods is linear, while the boundary of the proposed machine learning based methods can be nonlinear to eliminate some misclassification caused by the linear boundary.

After the modified classification methods and the SHF were developed, LWHI can be calculated with the proposed equation. In Chapter 5, we applied real data from the Houston Chronicle to test and verify LWHI. The results showed a high level of reliability, and the distribution of LWHI is a right-skewed normal distribution. With this reliable result, the LWHI can serve as a simple and effective hazard identification method that can be included in the overall PHA (Process Hazard Analysis) process of the facility.

In the future, this study has several scopes for improvement. The machine learning methods still have room for improvement, although it fits the scope of this thesis. For example, 'how to assign the initial clustering numbers?' is a popular question in a statistical study. In order to make the clustering result more persuasive and applicable to more chemicals, more statistical questions need to be answered.

NR values, which is another factor that may affect process safety, were not modified when calculating SHF in this study. Some incident reports mentioned that [2, 4] the interaction between different chemicals due to the storage space limitation might be a root cause for fire and explosion. Moreover, if the chemical reacts with water, like

ammonium nitrate [11], it also affects the design of an emergency response plan. Possible ways to modify NR may include:

1. Consider the special hazard;
2. Consider the interactions between different chemicals;
3. Consider the scale effect of the storage facility, which means different chemicals store together may amplify its hazards.

Other than SHF modification, penalty value selection could be another aspect to improve. The LWHI can expand to other scopes with different factors, such as the incident history of the facility, *etc.* By incorporating different aspects of the chemical facility, the LWHI can be further enhanced and become a more comprehensive index.

REFERENCES

[1] 天津市 2017 年国民经济和社会发展统计公报 / Statistical Communiqué of Tianjin on the 2017 National Economic and Social Development in, Statistical Bureau of Tianjin, 2017.

[2] G. Fu, J. Wang, M. Yan, Anatomy of Tianjin Port fire and explosion: Process and causes, Process Safety Progress, 35 (2016) 216-220.

[3] 天津港“8·12”瑞海公司危险品仓库特别重大火灾爆炸事故调查报告 / 8.12 Tianjin Explosion: Official Final Investigation Report in, State Administration of Work Safety, 2016.

[4] Final investigation report: West Fertilizer plant fire and explosion, in, US Chemical Safety and Hazard Investigation Board, 2016.

[5] V. Babrauskas, The West, Texas, ammonium nitrate explosion: A failure of regulation, Journal of Fire Sciences, 35 (2017) 396-414.

[6] D.P.E. Service, ITC Fire Updates, in, City of Deer Park website, 2019.

[7] Hazard Investigation: Improving Reactive Hazard Management, in, U.S. Chemical Safety and Hazard Investigation Board, 2001.

[8] Recommendations for Addressing Recurring Chemical Incidents at the U.S., in, Department of Energy, 2005.

[9] S. Mannan, Chapter 16 - Storage and Transport, in: Lees' Process Safety Essentials : Hazard Identification, Assessment and Control, Elsevier Science & Technology, Oxford, UNITED STATES, 2013.

[10] OSH Answers Fact Sheets, in, Canadian Centre for Occupational Health & Safety, 2019.

[11] Z. Han, S. Sachdeva, M.I. Papadaki, M.S. Mannan, Ammonium nitrate thermal decomposition with additives, Journal of Loss Prevention in the Process Industries, 35 (2015) 307-315.

[12] C.-P. Lin, J.-S. Li, J.-M. Tseng, M.S. Mannan, Thermal runaway reaction for highly exothermic material in safe storage temperature, Journal of Loss Prevention in the Process Industries, 40 (2016) 259-265.

[13] Q. Wang, D. Ng, M.S. Mannan, Study on the Reaction Mechanism and Kinetics of the Thermal Decomposition of Nitroethane, Industrial & Engineering Chemistry Research, 48 (2009) 8745-8751.

[14] M.S. Mannan, Director's Corner, in, Centerline, Vol. 19, No. 3, 2015.

[15] N.F.P. Association, NFPA 704: Standard System for the Identification of the Hazards of Materials for Emergency Response, in, 2017.

[16] G.L. Head, B.C. Wagner, III, The NFPA 704 diamond, Professional Safety, 40 (1995) 20.

[17] S. Mannan, Chapter 6 - Hazard Identification, in: Lees' Process Safety Essentials : Hazard Identification, Assessment and Control, Elsevier Science & Technology, Oxford, UNITED STATES, 2013.

[18] Dow's fire & explosion index : hazard classification guide., 7th ed. ed., American Institute of Chemical Engineers, 1994.

[19] Ranking of Chemical Facilities Based on the Potential to Cause Harm to the Public, in, Mary Kay O'Connor Process Safety Center, 2016.

[20] B.J. Tyler, Using the mond index to measure inherent hazards, Plant/Operations Progress, 4 (1985) 172-175.

[21] S.R. Saraf, W.J. Rogers, M.S. Mannan, Using screening test data to recognize reactive chemical hazards, Journal of Hazardous Materials, 104 (2003) 255-267.

[22] B. Wang, H. Yi, K. Xu, Q. Wang, Prediction of the self-accelerating decomposition temperature of organic peroxides using QSPR models, Journal of Thermal Analysis & Calorimetry, 128 (2017) 399-406.

[23] B. Wang, L. Zhou, K. Xu, Q. Wang, Prediction of Minimum Ignition Energy from Molecular Structure Using Quantitative Structure–Property Relationship (QSPR) Models, Industrial & Engineering Chemistry Research, 56 (2017) 47-51.

[24] Y. Pan, J. Jiang, R. Wang, H. Cao, Advantages of support vector machine in QSPR studies for predicting auto-ignition temperatures of organic compounds, Chemometrics and Intelligent Laboratory Systems, 92 (2008) 169-178.

[25] S. Yuan, Z. Jiao, N. Quddus, J.S. Kwon, II, C.V. Mashuga, Developing Quantitative Structure–Property Relationship Models To Predict the Upper Flammability Limit Using Machine Learning, Industrial & Engineering Chemistry Research, 58 (2019) 3531-3537.

[26] Y. Pan, J. Jiang, R. Wang, H. Cao, Y. Cui, A novel QSPR model for prediction of lower flammability limits of organic compounds based on support vector machine, Journal of Hazardous Materials, 168 (2009) 962-969.

[27] P. He, Y. Pan, J.-c. Jiang, Prediction of the self-accelerating decomposition temperature of organic peroxides based on support vector machine, Procedia Engineering, 211 (2018) 215-225.

[28] L. Mage, N. Baati, A. Nanchen, F. Stoessel, T. Meyer, A systematic approach for thermal stability predictions of chemicals and their risk assessment: Pattern recognition and compounds classification based on thermal decomposition curves, Process Safety and Environmental Protection, 110 (2017) 43-52.

[29] G. James, D. Witten, T. Hastie, R. Tibshirani, An introduction to statistical learning : with applications in R. Corrected edition. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Corrected edition. ed., Springer, 2014.

[30] J.H. Ward, Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 58 (1963) 236-244.

[31] B. Scholkopf, S. Kah-Kay, C.J.C. Burges, F. Girosi, P. Niyogi, T. Poggio, V. Vapnik, Comparing support vector machines with Gaussian kernels to radial basis function classifiers, IEEE Transactions on Signal Processing, 45 (1997) 2758-2765.

[32] Eichhorn, Careful! Mist can explode, Petroleum Refiner, (1955) 194-196.

[33] S. Gant, R. Bettis, R. Santon, I. Buckland, P. Bowen, P. Kay, Generation of flammable mists from high flashpoint fluids: Literature review, Institution of Chemical Engineers Symposium Series, (2012) 327-338.

[34] P. Kannan, T. Flechas, E. Mendez, L. Angarita, P. Chaudhari, Y. Hong, M.S. Mannan, A web-based collection and analysis of process safety incidents, Journal of Loss Prevention in the Process Industries, 44 (2016) 171-192.

[35] S. Yuan, C. Ji, A. Monhollen, J.S. Kwon, II, C. Mashuga, Experimental and thermodynamic study of aerosol explosions in a 36 L apparatus, Fuel, 245 (2019) 467-477.

[36] R.C. Santon, Mist fires and explosions - An incident survey, Institution of Chemical Engineers Symposium Series, (2009) 370-374.

[37] M.S. Mannan, The Buncefield explosion and fire–lessons learned, Process Safety Progress, 30 (2011) 138-142.

[38] A.H. Lefebvre, Flame radiation in gas turbine combustion chambers, International Journal of Heat and Mass Transfer, 27 (1984) 1493-1510.

[39] A.K. Jasuja, Atomization of Crude and Residual Fuel Oils, Journal of Engineering for Gas Turbines and Power, 101 (1979) 250-258.

[40] M.M. Elkotb, Fuel atomization for spray modelling, Progress in Energy and Combustion Science, 8 (1982) 61-91.

[41] R.D. Ingebo, H.H. Foster, Drop-size Distribution for Cross Current Breakup of Liquid Jets in Airstreams, in, 1957.

[42] D.A. Crowl, J.F. Louvar, Chemical process safety : fundamentals with applications. 3rd ed. Daniel A. Crowl, Joseph F. Louvar, 3rd ed. ed., Prentice Hall, 2011.

[43] C.E. Baukal, The Slipcover for The John Zink Hamworthy Combustion Handbook, Boca Raton, 2013.

[44] A.D. McNaught, A. Wilkinson, Compendium of chemical terminology : IUPAC recommendations. 2nd ed. compiled by Alan D. McNaught and Andrew Wilkinson, 2nd ed. ed., Blackwell Science, 1997.

[45] D.o. Energy, Protective Action Criteria for Chemicals - Including AEGLs, ERPGs, & TEELs, Revision 29, in, 2016.

[46] F. Pedregosa, Ga, #235, l. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, #201, d. Duchesnay, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12 (2011) 2825-2830.

[47] T. Kodinariya, P.R. Makwana, Review on Determining of Cluster in K-means Clustering, International Journal of Advance Research in Computer Science and Management Studies, 1 (2013) 90-95.

[48] LandView 6 on DVD : a viewer for EPA, Census Bureau, USGS data and maps, U.S. Dept. of Commerce, Economics and Statistics Administration, U.S. Census Bureau, 2003.

[49] John, HazardHub API, in: HazardHub (Ed.), 2019.

[50] M. Collette, M. Dempsey, Chemical Breakdown, in, Houston Chronicle, 2016.

APPENDIX A

DOW'S FIRE & EXPLOSION INDEX FORM

DOW

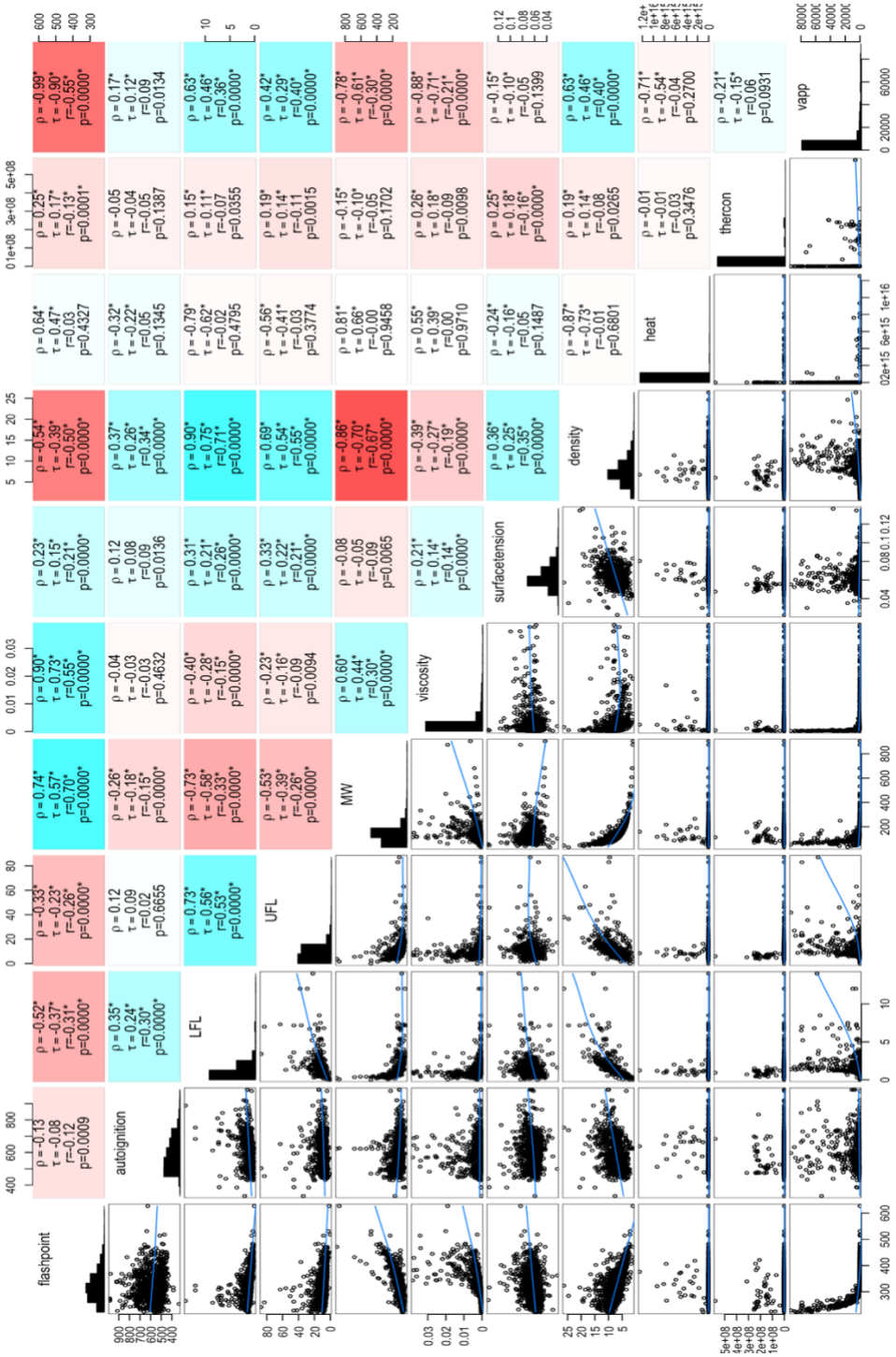
FIRE & EXPLOSION INDEX

AREA / COUNTRY	DIVISION	LOCATION	DATE
SITE	MANUFACTURING UNIT	PROCESS UNIT	
PREPARED BY:	APPROVED BY: (Superintendent)	BUILDING	
REVIEWED BY: (Management)	REVIEWED BY: (Technology Center)	REVIEWED BY: (Safety & Loss Prevention)	
MATERIALS IN PROCESS UNIT			
STATE OF OPERATION ___ DESIGN ___ START UP ___ NORMAL OPERATION ___ SHUTDOWN	BASIC MATERIAL(S) FOR MATERIAL FACTOR		
MATERIAL FACTOR (See Table 1 or Appendices A or B) Note requirements when unit temperature over 140 °F (60 °C)			

1. General Process Hazards	Penalty Factor Range	Penalty Factor Used(1)
Base Factor	1.00	1.00
A. Exothermic Chemical Reactions	0.30 to 1.25	
B. Endothermic Processes	0.20 to 0.40	
C. Material Handling and Transfer	0.25 to 1.05	
D. Enclosed or Indoor Process Units	0.25 to 0.90	
E. Access	0.20 to 0.35	
F. Drainage and Spill Control _____ gal or cu.m.	0.25 to 0.50	
General Process Hazards Factor (F_1)		
2. Special Process Hazards		
Base Factor	1.00	1.00
A. Toxic Material(s)	0.20 to 0.80	
B. Sub-Atmospheric Pressure (< 500 mm Hg)	0.50	
C. Operation In or Near Flammable Range ___ Inerted ___ Not Inerted		
1. Tank Farms Storage Flammable Liquids	0.50	
2. Process Upset or Purge Failure	0.30	
3. Always in Flammable Range	0.80	
D. Dust Explosion (See Table 3)	0.25 to 2.00	
E. Pressure (See Figure 2) Operating Pressure _____ psig or kPa gauge Relief Setting _____ psig or kPa gauge		
F. Low Temperature	0.20 to 0.30	
G. Quantity of Flammable/Unstable Material: Quantity _____ lb or kg H_C = _____ BTU/lb or kcal/kg		
1. Liquids or Gases in Process (See Figure 3)		
2. Liquids or Gases in Storage (See Figure 4)		
3. Combustible Solids in Storage, Dust in Process (See Figure 5)		
H. Corrosion and Erosion	0.10 to 0.75	
I. Leakage – Joints and Packing	0.10 to 1.50	
J. Use of Fired Equipment (See Figure 6)		
K. Hot Oil Heat Exchange System (See Table 5)	0.15 to 1.15	
L. Rotating Equipment	0.50	
Special Process Hazards Factor (F_2)		
Process Unit Hazards Factor ($F_1 \times F_2$) = F_3		
Fire and Explosion Index ($F_3 \times MF$ = F&EI)		

APPENDIX B

SCATTER PLOTS, DISTRIBUTION PLOTS, AND HEATMAP OF LIQUID PROPERTIES



APPENDIX C

FULL LIST OF 170 SHF VALUES

Chemical Name	**CAS**	**NH**	**NF**	**NR**	**SHF**
Gel	1302-78-9	4	1	0	19
Barite-barium sulfate	7727-43-7	2	0	0	6
Diesel	68334-30-5	2	2	0	9
Calcium chloride water	10043-52-4	3	0	1	11
Toluene 2,6 diisocyanate	91-08-7	4	1	0	19
Silica talc non asbesto	14807-96-6	4	0	0	18
Diethanolamine	111-42-2	3	2	0	13
Phosphoric acid	7664-38-2	3	0	0	10
Aminoethyl	140-31-8	3	3	0	17
Ethyl benzene	100-41-4	2	3	0	13
Cristobalite	14464-46-1	4	0	0	18
3.3,-dimethylnethylenedi (cyclohexylamine)	6864-37-5	4	1	0	19
Sodium hydroxide	1310-73-2	4	0	1	19
Potassium hydroxide	1310-58-3	4	0	1	19
Phenylenediamine 1,3	108-45-2	4	1	0	19
Zinc	7440-66-6	3	3	2	20
Potassium nitrate	7757-79-1	3	0	3	17
N,N-dimethylcyclohexylamine	98-94-2	4	3	0	25

Hydroquinone	123-31-9	4	1	0	19
Phenol	108-95-2	3	2	0	13
Cyclohexylamine	108-91-8	3	3	0	17
Tributyl tetradecyl phosphoniumchlordie	81741-28-8	2	3	0	13
Fatty acid polyamides	68410-23-1	2	1	0	7
Mineral spirits	8052-41-3	0	2	0	6
Sodium slat of trolyltriazole	64665-57-2	3	1	0	11
Phosphonic acid	6419-19-8	3	0	0	10
Carbon black	1333-86-4	3	0	0	10
Bisphenol epoxy resin	28064-14-4	3	0	0	10
Magnesium oxide	1309-48-4	3	0	1	11
Benzyl alcohol	100-51-6	3	2	0	13
1,4-butanediol digycidyl ethyer	2425-79-8	3	1	0	11
Tris-2,4,6-dimethylaminomethyl	90-72-2	3	1	0	11
Polyamine/ethyleneamines	112-24-3	3	1	0	11
Trimer of hexamethylene diisocyanatehexane 1,6-diisocyanato - homopolymer	28182-81-2	3	1	0	11
Sodium sulfate	7757-82-6	3	1	1	12
Tetraethylenepentamine	112-57-2	3	1	0	11
Xylene	1330-20-7	1	3	0	11
Methyl ethyl ketone	78-93-3	1	3	0	11

N-butyl acetate	123-86-4	1	3	0	11
Methyl propyl ketone	107-87-9	1	3	0	11
Morpholine	110-91-8	1	4	0	19
Battery acid	7664-93-9	3	0	2	13
Anhydrous ammonia	7664-41-7	3	2	0	13
Lead	7439-92-1	3	0	0	10
Acetic acid 60%	64-19-7	3	3	0	17
Benzyl chloride	100-44-7	3	2	1	14
Ammonium persulfate	7727-54-0	4	0	1	19
Maleic anhydride	108-31-6	4	2	1	22
Frac sand	14808-60-7	4	0	0	18
Sodium metasilicate	6834-92-0	3	0	0	10
Dichloroethyl ether	111-44-4	2	3	1	14
Sodium acid pyrophosphate	7758-16-9	3	0	0	10
Ammonium chloride	12125-02-9	3	0	0	10
Ammonium bifluoride	1341-49-7	3	0	0	10
Citric acid anhydrous	77-92-9	3	1	0	11
Fumaric acid	110-17-8	2	1	0	7
Nonylphenol ethoxylate 15 mole (sulfonic 150)	9016-45-9	3	1	0	11
Lubricating oil base stock automatic transmission fluid	64742-54-7	2	1	0	7

Ethylene glycol	107-21-1	3	2	0	13
Calcium hypochlorite, dry	7778-54-3	3	0	1	11
Ammonium hydroxide	1336-21-6	2	0	0	6
Cuprous oxide	1317-39-1	4	0	0	18
White spirit	64742-88-7	4	2	0	21
1-methoxy-2-propanol	107-98-2	2	4	0	21
Aromatic 100 fluid	64742-95-6	2	2	0	9
Butyl alcohol	71-36-3	1	3	0	11
Napthalene	91-20-3	2	2	0	9
Ethylene glycol monbutyl ether	111-76-2	2	3	0	13
2-methyl-2-butene	513-35-9	4	4	0	33
Polymeric diphenylmethane	9016-87-9	4	0	0	18
1,1-dioxide therahydrothiophen	126-33-0	4	1	0	19
4,4-diphenylmethane	101-68-8	4	1	0	19
Methylene bis (4-cyclohexylisocyanate)	5124-30-1	4	1	0	19
2, 6-xyphenol	576-26-1	4	2	0	21
Aniline	62-53-3	4	2	0	21
Methyl isobutyl ketone	108-10-1	1	3	1	12
Ehtyl 3-ethyoxypropionate	763-69-9	3	2	0	13
Trans 1,2 dichloroethylene	156-60-5	2	3	2	16
P-xylene	106-42-3	2	3	0	13

Styrene monomer	100-42-5	2	3	2	16
Isophorone diisocyanate	4098-71-9	4	1	1	20
N propylbenzene	103-65-1	1	2	0	7
M-cresol	108-39-4	2	2	0	9
2-propenal,3-phenyl	104-55-2	3	2	0	13
Isooctane	540-84-1	1	3	0	11
1,1,1,3,3-pentafluoropropane	460-73-1	1	0	0	4
Diisobutyl ketone, 2,6-dimethyl-4-hepton	108-83-8	1	2	0	7
Propylene glycol monomethyl ether aceta	108-65-6	1	3	0	11
1,2,4-trimethylbenzene	95-63-6	2	2	0	9
N-decane	124-18-5	2	2	0	9
Propylene carbonate	108-32-7	3	2	0	13
Phenol, 2,4-dimethyl	105-67-9	3	1	0	11
N-octane	111-65-9	1	3	0	11
Methylcyclohexane	108-87-2	1	3	0	11
N-amyl acetate	628-63-7	1	2	0	7
Fiberglass	65997-17-3	3	0	0	10
Vinyl acetate	108-88-3	1	3	0	11
Bebezene,diethenyl-,polymer with ethe	69011-20-7	2	0	0	6
O-cresol	95-48-7	2	2	0	9
P-cresol	106-44-5	2	2	0	9

Heptane	142-82-5	1	3	0	11
Kerosene	8008-20-6	2	2	0	9
Ethanolamine	141-43-5	2	3	0	13
Chromium oxide	1308-38-9	4	0	0	18
Butyl acrylate	141-32-2	2	3	0	13
Iron oxide	1309-37-1	2	0	0	6
Methyl acrylate	96-33-3	2	4	2	24
Methanol	67-56-1	2	4	0	21
Manganese	7439-96-5	2	0	2	9
Tetrasodium pyrophosphate	7722-88-5	3	0	0	10
Polyethylene	9002-88-4	3	0	0	10
Iron oxide	1317-61-9	2	0	0	6
Monosodium phosphate anhydrous	7558-80-7	2	0	0	6
Copper	7440-50-8	4	3	0	25
Hydrochloric acid	7647-01-0	3	0	2	13
Paraquat dichloride	1910-42-5	4	0	0	18
Silica, amorphous hydrated	7631-86-9	4	0	0	18
Methane, difluoro	75-10-5	1	4	1	20
Activated carbon	7440-44-0	4	0	0	18
Methomyl	16752-77-5	4	1	0	19
Calcium oxide	1305-78-8	3	0	2	13

Sodium oxide	1313-59-3	4	0	0	18
Oxamyl	23135-22-0	4	1	0	19
Polytetrafluoroethylene	9002-84-0	4	0	0	18
Chlorothalonil	1897-45-6	4	1	0	19
Benzenesulfonic acid, 4-methyl-,methyl e	80-48-8	4	1	0	19
Alumina	1344-28-1	4	0	0	18
Titanium dioxide	13463-67-7	4	0	0	18
Silica gel, amorphous	7699-41-4	4	0	0	18
Acrylic acid	79-10-7	3	3	2	20
Coal tar pitch	65996-93-2	3	3	0	17
Rosin	8050-09-7	4	1	0	19
Polyoxy propylene diamine	9046-10-0	4	1	0	19
1,1,2 tetrafluorothethane	811-97-2	0	0	0	3
Chlorodifluoromethane	75-45-6	2	0	0	6
Copper hydroxide	20427-59-2	3	1	0	11
Dodecylbenzene sulfonate	27176-87-0	3	1	0	11
Magnesite	546-93-0	2	0	0	6
Vinylidene fluoride hexafluropropene	9011-17-0	2	0	0	6
Propylene glycol	57-55-6	2	1	0	7
Heptan-2-one	110-43-0	1	2	0	7
Ethylene glycol	2807-30-9	2	3	0	13

2,4-d acetic acid	94-75-7	3	0	0	10
4,4'-methylene-bis (2-chloroaniline)	101-14-4	3	1	0	11
4-nonylphenol,branched	84852-15-3	3	1	0	11
Zinc oxide	1314-13-2	2	0	0	6
Dimethyl acetamide	127-19-5	2	2	0	9
Cyclohexanone	108-94-1	1	3	0	11
Ethanol	64-17-5	1	4	0	19
Isopropyl alcohol	67-63-0	1	4	0	19
Calcium carbonate	1317-65-3	2	0	0	6
Nonylphenol	25154-52-3	3	1	0	11
Mancozeb	8018-01-7	3	1	0	11
Isobutyl alcohol	78-83-1	1	3	0	11
Ammonium acetate	631-61-8	3	1	0	11
Ammonium sulfate	7783-20-2	3	0	0	10
Momoammonium	7722-76-1	3	0	1	11
Aluminum	7429-90-5	3	3	3	24
Paraformalydehyde * 400412	50-00-0	4	4	4	48
Sodium * 400148	7440-23-5	3	3	3	24
Acetone * 400411126645, 34471	67-64-1	1	4	2	22
Methyl methacrylate * monomer4000096/dammma 25	80-62-6	2	4	2	24

Benzyldimethylamine/400191* 3602000	103-83-3	4	2	0	21
Ethylenediamine * - 400149 - 1W	107-15-3	4	2	0	21
Flammable Liquid, n.o.s.* Kerosene, petroleum, 0-xylene, 400090/000963,000964	95-47-6	2	3	0	13
Pyridine * 400191-1851000	110-86-1	1	4	0	19
Tetrachloroethylene * 400149- 1S	127-18-4	2	0	0	6
Vanadium pentoxide * 400146-40716,40717,40714	1314-62-1	3	0	0	10
Diesel fuel	68476-30-2	2	2	0	9
Sodium carbonate (soda ash dense)	497-19-8	3	0	1	11