PRACTICAL INHERENTLY SAFER DESIGN APPROACHES DURING EARLY PROCESS DESIGN STAGES AIMING FOR SUSTAINABILITY

A Dissertation

by

SUNHWA PARK

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Mahmoud El-Halwagi
Committee Members,	Stratos Pistikopoulos
	Qingsheng Wang
	Nancy Currie-Gregg
Head of Department,	Arul Jayaraman

May 2022

Major Subject: Chemical Engineering

ABSTRACT

In traditional industrial process design approaches, techno-economic criteria have been the primary objectives in the early process design stages. Safety is often considered only in the later design stage (e.g., detailed engineering stage). Such a traditional approach is that most of the design degrees of freedom, including technology and configuration issues, have already been determined when considering safety. Modifying a process is costly or unreliable at later stages. To solve this issue, there have been numerous attempts to consider process safety during the early design stages in safety engineers and researchers. In particular, special attention to adopting inherently safer design (ISD) has been made because ISD is deemed the most cost-effective risk reduction strategy at early design stages. However, it is still challenging to adopt ISD for process engineers at the early design stages because of the lack of guidance and insufficient information on upcoming process facilities.

To address this challenge, this dissertation consists of three peer-reviewed journal papers [Articles #1 - #3]. With respect to the progress of inherently safer design (in particular, during the early design stage) over the last three decades, Article #1 selects 73 inherent safety assessment tools, which can be utilized during the early design stages, and categorized into three groups: hazard-based inherent safety assessment tools (H-ISATs) for 22 tools, risk-based inherent safety assessment tools (R-ISATs) for 33 tools, and cost-optimal inherent safety assessment tools (CO-ISATs) for 18 tools. The goal of this article

is to enable process engineers to use all the available design degrees of freedom to mitigate risk early enough in the design process.

Article #2 analyzes 94 chemical process incidents investigated by the U.S. Chemical Safety and Hazard Investigation Board (CSB) reports. To analyze in a systematic approach, this article proposes 17 incident cause factors, 6 scenario factors, and 6 consequence factors to find out whether ISD would have helped to prevent these incidents.

Article #3 proposes hands-on predictive models of the flash point, the heat of combustion, lower flammability limit (LFL), and upper flammability limit (UFL). By incorporating the nonlinearity and transformation along with linearity of variables, this article constructed practical, reliable regression models thoroughly with readily available variables—the number of all atoms, molecular weights, and boiling points. The purpose is to enable a process engineer to quickly obtain hazardous properties of intended process materials.

DEDICATION

First and foremost, I would like to dedicate this dissertation to my family. This work would not have been possible without their unconditional love and support.

My husband, Hoseok Yoe, has sacrificed everything due to this journey and gave me strength whenever there were difficulties, even learning coding and data mining together. Thanks to you, I managed all required steps at Texas A&M University and built a happy family.

I feel thankful to my soon-to-be-born son, Henry, for always healthy movement, even though I could not get proper prenatal education during the pregnancy due to research and graduation preparations. I feel thankful to you for coming into my life.

Lastly, I would like to thank my parents, Chanheung Park, Mija Jung, and my parents-in-law, Jaeyoung Yeo and Haekyung Jung. Their love and support have always been more than I deserve.

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Dr. Mahmoud El-Halwagi. You are such a kindhearted mentor, great leader, and excellent researcher. I am very fortunate to have a great mentor like you during this challenging journey. You have provided more than enough support and guidance during my Ph.D. program. Without your support, my current achievements would not have been possible. Under your advice, I could grow intellectually and socially by performing various projects with other lab members.

I am also very grateful to my previous late advisor, Dr. M. Sam Mannan, for guiding me to this Ph.D. program and process safety research world. I will continue to promote safety as second nature in the industry by keeping in mind your saying, "Safety should not be a priority; priorities change. So, *Safety should be a core value, a second nature*".

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

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Mahmoud El-Halwagi [advisor], Professor Stratos Pistikopoulos, Professor Qingsheng Wang, and Professor Nancy Currie-Gregg. They have guided the overall direction of the entire dissertation. In Particular, Professor El-Halwagi made significant contributions as an advisor, a co-author, and a corresponding author of the journal articles. Professor Wang generously offered his expertise to facilitate research regarding chemical flammable properties.

Also, this work has been supported by two great mentors, a MKOPSC research Professor Hans Pasman and Dr. Dr. William J. Rogers. First, Professor Pasman has made valuable contributions as a mentor to this dissertation by collaborating on every journal article and guiding me in each step. His guidance has allowed me to learn to expand my horizon of process safety knowledge, the way of research, and leadership. Dr. William J. Rogers, as a mentor and co-author of an article, has guided me through the literature progress and provided his professional editing and proofreading support.

Dr. James P. Bailey of the Department of Industrial & System Engineering contributed much to analysis of data used in the development new predictive models of the flammable properties.

Funding Sources

This dissertation was funded by Mary Kay O'Connor Process Safety Center (MKOPSC).

NOMENCLATURE

2DGR	2-Dimensional Graphical Rating
ACO	An Ant Colony Optimization Algorithm
AEGL	Acute Exposure Guideline Level
AHI	Accident Hazard Index
AIT	Auto-Ignition Temperature
ALARP	As Low as Reasonably Practicable
ALOHA	Area Location of Hazardous Atmospheres
ANN	Artificial Neural Network
BLEVE	Boiling Liquid Expanding Vapor Explosion
CAS	Chemical Abstracts Service
CISI	Comprehensive Inherent Safety Index
CPI	Chemical Process industries
CRSI	Chemical Route Safety Index
CSB	the U.S. Chemical Safety and Hazard Investigation Board
DIPPR	Design Institute of Physical Properties Relationships
EHI	Environment Hazard Index
EISI	Enhanced Inherent Safety Index
EPA	the U.S. Environmental Protection Agency
ER	Emergency Responder
ERPG	Emergency Response Planning Guideline
ETA	Event Tree Analysis

ETHI	Environmental Toxicity Hazard Index
FEDI	Fire, Explosion and Damage Index
FMEA	Failure Mode and Effect Analysis
FTA	Fault Tree Analysis
GRAND	Graphical and Numerical Descriptive Technique for Inherent Safety Assessment
GNN	Graph Neural Networks
HAZOP	Hazard and Operability
H-ISAT	Hazard-based Inherent Safety Assessment Tool
HIRA	Hazard Identification and Ranking
HPS	Hazard Prevention Strategies
HQI	Health Quotient Index
I2SI	Integrated Inherent Safety Index
IBI	Inherent Benign-Ness Indicator
ICPRI	Inherent Chemical Process Route Index
IFCET	Inherent Fire Consequence Estimation Tool
IRA	Inherent Risk Assessment
iRET	integrated Risk Estimation Tool
IRDI	Inherent Risk of Design Index
IOHI	Inherent Occupational Health Index
ISAT	Inherent Safety Assessment Tool
ISAPEDS	Inherent Safety Assessment Technique for Preliminary Design Stage

ISBI	Inherent Safety Benefits Index
ISBL	Inside Battery Limits
ISD	Inherently Safer Design
i-SDT	inherently Safer Design Tool
ISI	Inherent Safety Index
ISIF	Inherent Safety Intervention Framework
ISIM	Inherent Safety Index Module
ISISTHE	Inherent Safety Index for Shell and Tube Heat Exchanger
IS-KPI	Inherent Safety Key Performance Indicators
ISPI	Inherent Safety Performance Indices
ISPP	Inherently Safer Process Piping
KPI	Key Performance Indicator
LFL	Lower Flammability Limit
LInX	Life Cycle Indexing System
LOOCV	Leave-One-Out-Cross Validation
MAE	Mean Absolute Error
MIE	Minimum Ignition Energy
MINLP	Mixture-Integer Non-Linear Program
MNL	Multiple Linear Regression
MSDS	Material Safety Data Sheet
MSE	Mean Square Error
NFPA	the US National Fire Protection Association

NLP	Nonlinear Programming
NPEI	Net Potential Economic Impact
NuDIST	Numerical Descriptive Inherent Safety Technique
OhHI	Occupational Health Hazard Index
OSBL	Outside Battery Limits
OSHA	the U.S. Occupational Safety and Health Administration
OSIHEN	Overall Safety Index for Heat Exchanger Network
PAC	Protective Action Criteria
PHA	Process Hazard Analysis
PIIS	Prototype Index for Inherent Safety
PFD	Process Flow Diagram
P&ID	Piping and Instrument Diagram
PLS	Partial Least Square
PRI	Process Route Index
PRHI	Process Route Healthiness Index
PSCI	Process Stream Characteristic Index
PSI	Process Stream Index
PSM	Process Safety Management
PSV	Process Safety Valve
QAISD	Qualitative Assessment for Inherently Safer Design
QRA	Quantitative Risk Analysis
QSPR	Quantitative Structure–Property Relationship

R-ISAT	Risk-based Inherent Safety Assessment Tool
RMSE	Root Mean Square Error
RRABD	Rapid Risk Analysis-based Design
SASWROIM	Safety and Sustainability Weighted Return on Investment Metric
SFBS	Sequential Floating Backward Selection
SFBS-MLR-L	SFBS-based MLR Models Incorporating Linearity
SFBS-MLR-NLI	SFBS-based MLR Models Incorporating Non-Linearity and Interaction
SFFS	Sequential Floating Forward Selection
SFFS-MLR-NLI	SFFS-based MLR Models Incorporating Non-Linearity and Interaction
SIL	Safety Integrity Level
SREST	Substance, Reactivity, Equipment and Safety-Technology
SVM	Support Vector Machine
SWeHI	Safety Weight Hazard Index
SWROIM	Sustainability Weighted Return on Investment Metric
TDI	Toxicity Damage Index
TEEL	Temporary Emergency Exposure Limit
TIM	Three-stage ISD Matrix
TORCAT	Toxic Release Consequence Analysis Tool
TRIRA	Toxic Release Inherent Risk Assessment
TRSI	Toxic Release Stream Index
TRRI	Toxic Release Route Index

UFL	Upper Flammability Limit
VBA	Visual Basic for Application
VCE	Vapor Cloud Explosion
VIF	Variance Inflation Factor
WAR	Waste Reduction

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Background

The development path of sustainable manufacturing has strived to balance the three pillars of sustainability: environment, society, and economy (El-Halwagi, 2017). In particular, the chemical process industry deals with hazardous materials and processes that can have a significant impact on the well-being of surrounding communities and natural resources. Therefore, we should extend a broader understanding of sustainability in the chemical process industry by including safety along with these three pillars.

Of various safety strategies, Inherently Safer Design (ISD) is one of the most effective risk reduction strategies for achieving sustainable chemical facilities. Instead of merely installing safety equipment or devices (CCPS, 2010), ISD is a proactive strategy for reducing the likelihood or the impact of incidents and achieving cost-optimal safety solutions in the chemical process industry.

Since early design stages (i.e., conceptual design and preliminary design) offer more flexibility to perform ISD, practitioners must consider ISD to apply it during these stages. If ISD is impractical to reduce the chance of potential incidents, it may be more challenging to achieve tolerable risk ranges with the two other safety strategies. In the final step of the risk management system, chemical companies must iterate this hierarchy strategy until tolerable risk ranges are obtained.

The following chapter (Chapter 2, Article #1) will provide a detailed background of ISD, including the necessities for adopting ISD during early design stages and relevant challenges.

1.2. Research question

Despite the shared insights among engineers in the chemical process industry that adopting ISD during the early design stages is a strategic tactic, practitioners often encounter problems implementing ISD due to the insufficient practical, reliable approaches. Therefore, more hands-on research efforts are necessary to appropriately enable process researchers or engineers to consider ISD during the early design stages.

As the first step, to come up with more reliable applicable ISD strategies, this dissertation is aimed at addressing the major research question:

- *How can we effectively measure inherently safer design levels during early design stage in which process information are insufficiently provided*?

1.3. Dissertation layout

This dissertation is composed of three articles to the following questions to answer the major question, as shown Figure 1.1. First, Chapter 2 (Article #1) lists the fundamental information of safety and ISD principles and reviews feasible safety assessment tools to measure ISD levels during the early design stages, along with other sustainability aspects. Then, this article categorizes the selected assessment tools based on three risk

management purposes: hazard, risk, and cost-optimal assessment. Chapter 2 also identifies the common variables that have been used as safety indicators among the selected assessment tools.

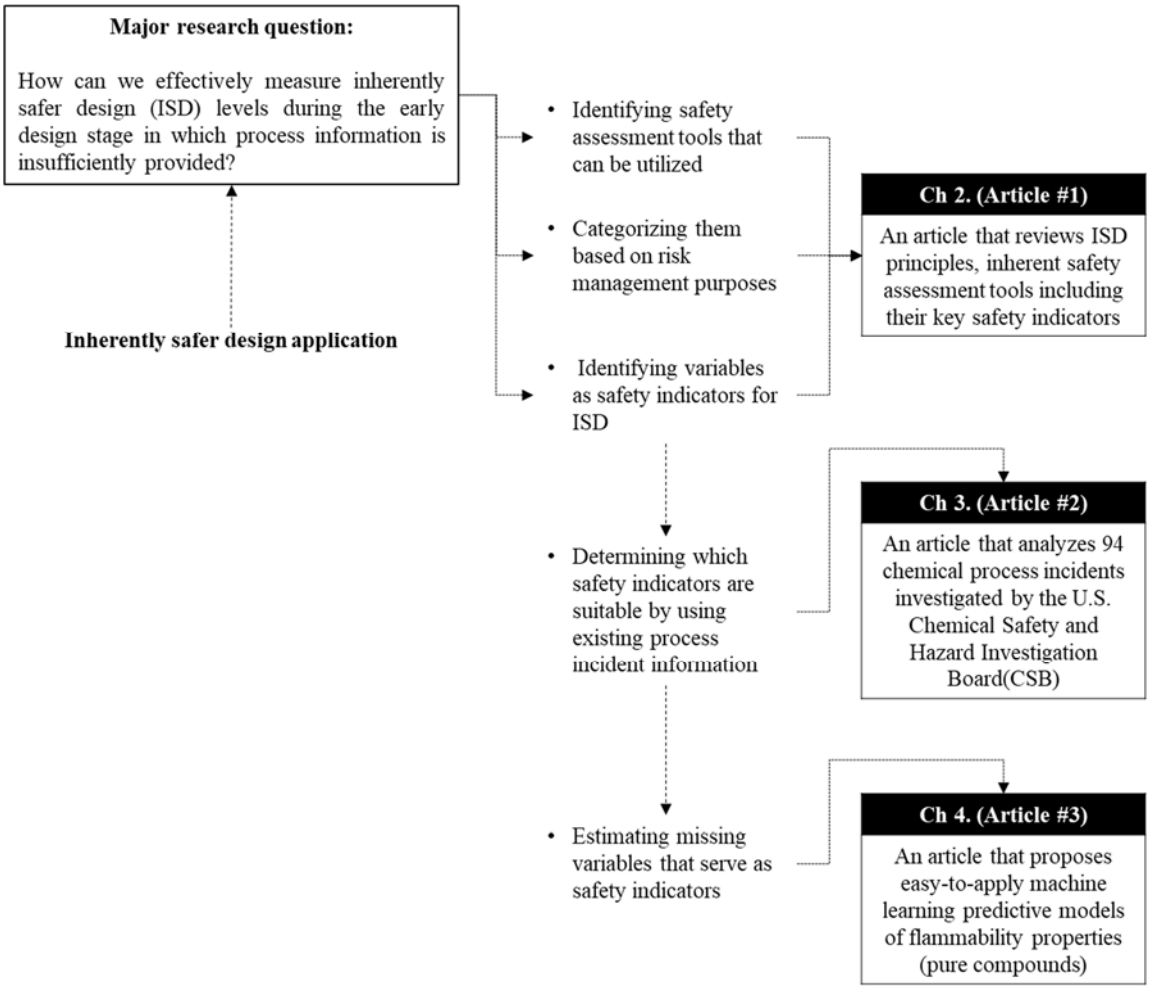


Figure 1.1 An overview of this dissertation layout

Based on the literature review in Chapter 2, Chapter 3 (Article #2) analyzes 94 chemical process incidents investigated by the U.S. Chemical Safety and Hazard

Investigation Board (CSB) to determine which safety indicators are suitable to apply to the previous process incidents.

Next, Chapter 4 (Articles #3) proposes more easy-to-apply machine learning predictive models for flammability properties than currently existing predictive models such as physical property models, group contribution models, and quantitative structure-property relationships (QSPR) models. The purpose of this article is to estimate missing safety indicators in order to enable practitioners to adopt in a practical way. In this article, the novel predictive models of flammability properties (i.e., flash point, the heat of combustion, lower flammability limit, and upper flammability limit of organic compounds) are developed via machine learning algorithms.

1.4. References of Chapter 1

CCPS, 2010. Inherently Safer Chemical Processes: A Life Cycle Approach. John Wiley & Sons, Incorporated).

El-Halwagi, M.M., 2017. Sustainable Design through Process Integration: Fundamentals and Applications to Industrial Pollution Prevention, Resource Conservation, and Profitability Enhancement. Butterworth-Heinemann.

2. INCORPORATING INHERENT SAFETY DURING THE CONCEPTUAL PROCESS DESIGN STAGE: A LITERATURE REVIEW[1]

This chapter reviews principal concepts, tools, and metrics for risk management and Inherently Safer Design (ISD) during the conceptual stage of process design. Even though there has been a profusion of papers regarding ISD, the targeted audience has typically been safety engineers, not process engineers. Thus, the goal of this chapter is to enable process engineers to use all the available design degrees of freedom to mitigate risk early enough in the design process. Mainly, this chapter analyzes ISD and inherent safety assessment tools (ISATs) from the perspective of inclusion in conceptual process design. This chapter also highlights the need to consider safety as a major component of process sustainability. In this chapter, 73 ISATs were selected, and these tools were categorized into three groups: hazard-based inherent safety assessment tools (H-ISATs) for 22 tools, risk-based inherent safety assessment tools (R-ISATs) for 33 tools, and cost-optimal inherent safety assessment tools (CO-ISATs) for 18 tools. This chapter also introduces an integrated framework for coordinating the conventional process design workflow with safety analysis at various levels of detail.

[1] Reprinted with permission from "Incorporating inherent safety during the conceptual process design stage: a literature review" by Park, S., Xu, S., Rogers, W., Pasman, H., & El-Halwagi, M. M. 2020, Journal of Loss Prevention in the Process Industries, 63, 104040, Copyright [2020] by Elsevier Ltd. All rights reserved.

2.1. Introduction

The design of industrial processes proceeds via several stages that start with high-level synthesis and screening of alternatives in the conceptual design stage. Afterward, the process continues with more analysis that yields the recommended design in the detailed design stage. Traditionally, techno-economic criteria have been the principal objectives in the early process design stage. Safety is usually considered in the detailed design stages. At this point, most of the design degrees of freedom including technological and configurational issues have already been determined. The objective of this chapter is to advocate for the consideration of inherent safety aspects in the conceptual design stage; specifically, the analysis focuses on how various inherent safety assessment tools (ISATs) can be utilized by process engineers. Previous review papers have shown overall views of incorporating inherent safety principles (e.g., grouping different types of methods or research groups) with the primary target audience being safety experts. This chapter is aimed at the process design community to show principal safety concepts, definitions, tools, and insights and to enable process engineers to apply inherent safety principles during the conceptual design stage.

2.1.1. Safety is a key component of in sustainability

The development path of sustainable manufacturing has strived to balance the three pillars of sustainability—environment, society, and economy (National Research Council, 2011). A broader understanding of sustainability in the chemical process industries (CPIs) should be extended to include safety along with the three aforementioned

pillars, as shown by Figure 2.1. The CPIs in particular deal with hazardous materials and processes that can have a major impact on the wellbeing of surrounding communities and natural resources.

Figure 2.1 Safety as a key pillar towards sustainability

2.1.2. Inherently safer design as a logical component of conceptual process design for better sustainability

Inherently Safer Design (ISD) is one of the most effective risk reduction strategies for achieving sustainable chemical facilities. The idea of "removing risks at the design stage" was originally developed in a pioneering study by a U.K. process engineer, Houston (1971), and was further developed in the ISD concept by Kletz (1985). Instead of merely installing safety equipment or devices, ISD is a proactive strategy for reducing the

likelihood or the impact of incidents and endeavoring to achieve cost-optimal safety solutions in the CPIs.

Without an appropriate ISD approach, chemical facilities will be at increased risk of experiencing devastating events. The lessons learned—from previous catastrophic chemical accidents— demonstrated the possible impacts on people, the environment, and the economy in our society (Eckerman, 2018). For example, the 1984 Bhopal disaster in India resulted in over 2,000–8,000 fatalities, the number depending on different sources (Eckerman, 2018; Lees, 2012); this disaster was triggered by methyl isocyanate, an extremely toxic intermediate, which was stored for further processing in unnecessary large amounts that far exceeded acceptable safety limits. Singh et al. (2010) also highlighted the importance of ISD by analyzing the 1998 Piper Alpha disaster, which resulted in 167 fatalities. Although systematic techniques have been developed to incorporate green chemistry in conceptual design (e.g., Julián-Durán and Laura, 2014; Crabtree and El-Halwagi, 1994), much less attention has been given to including inherently safer chemistry during the early stages of process design.

Natural threats are another reason why a chemical facility must apply ISD. Krausmann et al. (2011) concluded that natural events triggered 5% of industrial accidents involving process units or stored hazardous substances. In 2017, for instance, all chemical facilities in the U.S. Houston area expected flooding from Hurricane Harvey based on previous events involving excessive rainfall (DeRosa et al., 2019). However, the excessive rainfall from Hurricane Harvey was so unprecedented that the Arkema Crosby facility lost back-up power and failed to safely store organic peroxide products at low temperature.

Eventually, these organic peroxide products loaded in containers and for Harvey placed on trucks spontaneously combusted producing explosion effects, and people within a 1.5-mile radius were required to evacuate (Chemical Safety Board (CSB) 2018). Such an example—of unpredictable natural disasters—represents why process engineers must adopt ISD in the first place rather than attempting to control a hazard after an event it triggers occurs. By then it may be impossible to control. Specifically, Cozzani et al. (2010) referred to chemical accidents triggered by natural hazards (e.g., floods, earthquake, lightning, etc.) that lie at the interface of nature and technology as NaTech.

2.1.3. The direction of this chapter

This chapter comprises six sections. Section 2.2 explains the fundamental concepts of safety and ISD based on the developed concept of Kletz and Amyotte (2010), followed by an explanation of each process design stage. This section was prepared for both process engineers and safety engineers; a process engineer calls for basic safety knowledge, whereas a safety engineer desires to understand process design stages. Section 2.3 highlights a new review paper that mainly analyzes inherent safety assessment tools (ISATs) for the conceptual design stage via five previous review papers. Section 2.4 addresses the methods of reviewing literature to select ISATs in this chapter. Section 2.5 groups the selected ISATs in three categories and characterizes them. Finally, Section 2.6 concludes this chapter with recommendations and possible next steps.

2.2. Understanding inherently safer design (ISD)

2.2.1. Hazard vs. Risk

It is essential to make a clear distinction between hazard and risk. Table 2.1 shows key definitions. **Risk** is the result of a combination of consequence and likelihood, which can vary widely depending on the circumstances and conditions (Pasman 2015). On the other hand, a **hazard** is the intrinsic-damage potential (e.g., of a chemical or a process). As Kletz (1978) said: "what you don't have, can't leak," if chemical plants were designed without storage tanks, then storage tanks could not possibly leak. Similarly, if hazards could be eliminated, possible accidents could be completely avoided in advance.

Table 2.1 Definitions for Hazard vs. Risk

Term	Definition
Hazard [a]	A physical or chemical condition that has the potential for causing harm to something valued such as people, property, or the environment
Risk [a, b]	The combination of the expected Consequence and Likelihood of a single incident or a group of incidents, or in general the effect of uncertainty on objectives $\text{Risk} = f\ [\text{Consequence}, \text{Likelihood}]$
Consequence [b]	Potential loss that can be expressed in quantifiable units (e.g., monetary: $/event)
Likelihood [b]	The chance of occurrence of the incident (e.g., events/year)

[a] CCPS (2008), Guidelines for hazard evaluation procedures

[b] Pasman (2015), and the more general ISO (2009) definition

Traditionally, the CPIs have focused on reducing risk values with preventive measure — attempting to control a hazard by installing add-on safety equipment. Despite the chance of failure of protective equipment, it can still lower the risk ranking level, implying increased plant safety. However, the traditional process-design approach often fails to remove or sufficiently minimize the inherent hazard. To compensate later, it may require late design changes or costly preventative and more often protective measures. Therefore, in order to reduce or eliminate the inherent hazard, ISD is a much more proactive approach to reduce overall risk levels and improve plant safety. This approach is best implemented during the conceptual process design stage.

2.2.2. Four guidewords for ISD

The use of ISD is driven by with four guidewords — Intensification, Substitution, Attenuation, and Simplification (Kletz and Amyotte, 2010). ISD excludes/reduces hazards with these guidewords. Table 2.2 provides a brief overview of ISD principles. Because intensification is used to minimize the amount of hazardous chemicals, its application has been developed more actively in the CPIs (Tian et al., 2018) relative to other guidewords.

Table 2.2 Four guidewords for inherently safer design (Adopted from Kletz and Amyotte, 2010; CCPS, 2010; Abidin et al., 2016)

Guideword	Description

Intensification	Use smaller quantities of hazardous materials by reducing the size of relevant equipment (also called “Minimization”)
Substitution	Use less hazardous materials, reactions, and processes
Attenuation	Use more moderate process conditions through dilution, refrigeration, and process alternatives, but the total amount of chemicals might not be reduced (also called “Moderation”)
Simplification	Eliminate unnecessary complexity to decrease the likelihood of errors

2.2.3. Process risk management system incorporating ISD

For an improved risk management system, ISD should be applied in tandem with other safety strategies because ISD is not intended to be a stand-alone strategy. For example, Figure 2.2 shows the process risk management system incorporating the principles of ISD, domino effects, and layers of protection, based on the notions of Kletz and Amyotte (2010) and CCPS (2010). Because early design stages (i.e., conceptual design and preliminary design) offer more flexibility to perform ISD, practitioners must consider ISD to apply it during these stages. If ISD is impractical to reduce the chance of potential incidents, it may be more challenging to achieve tolerable risk ranges with the two other safety strategies. In the final step of the risk management system, chemical companies must iterate this hierarchy strategy until tolerable risk ranges are obtained.

The scope of ISD, as originally described by Kletz and Amyotte (2010), excluded the concept of domino effects. However, many researchers subsequent to Klets and Amyotte have contended that domino effects are an intrinsic characteristic to be

considered for ISD. For example, if chemical plants are inherently safer, corresponding domino effects are expected to decrease as well. For this reason, this chapter will not cover the secondary scope, domino effects.

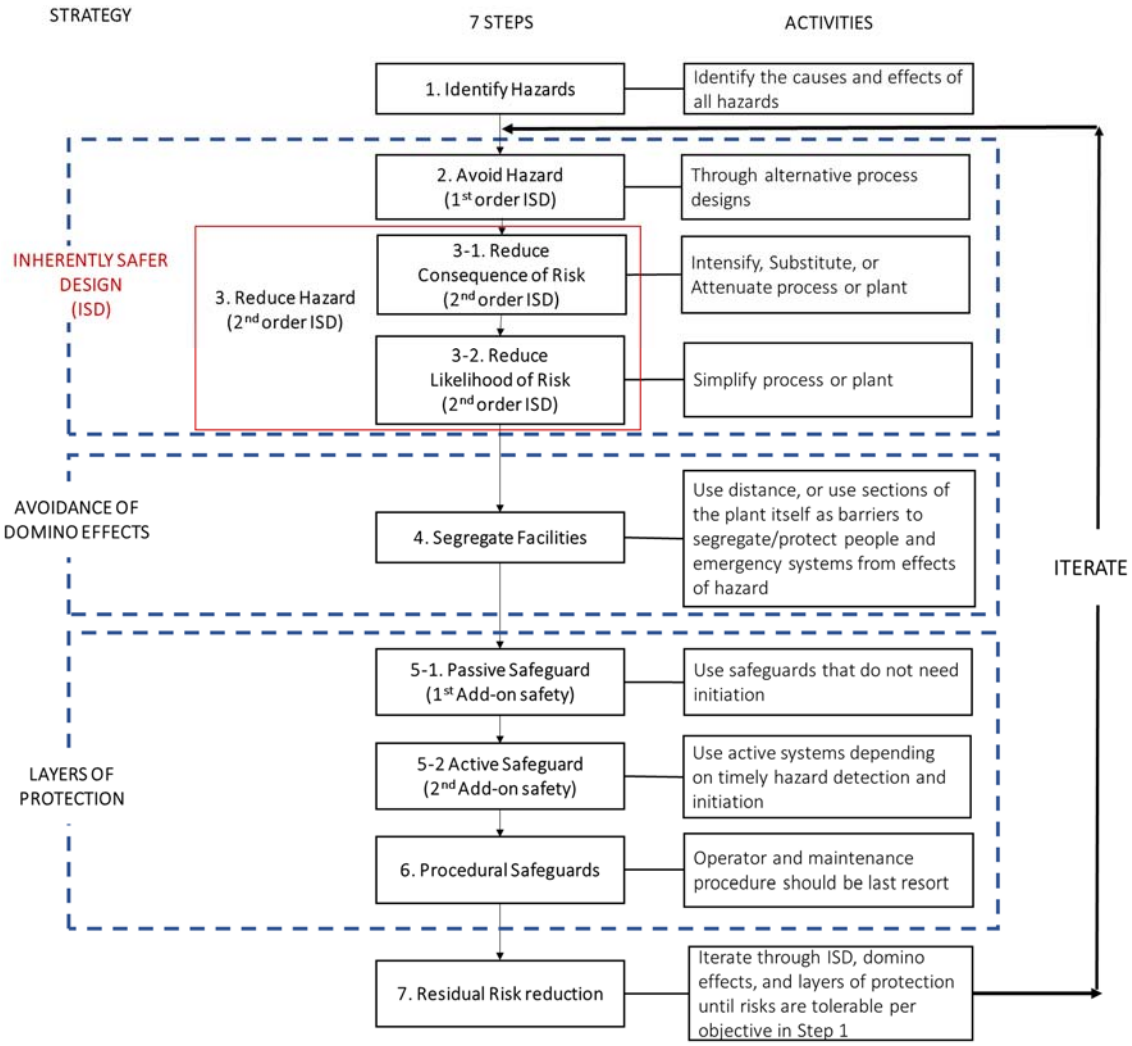


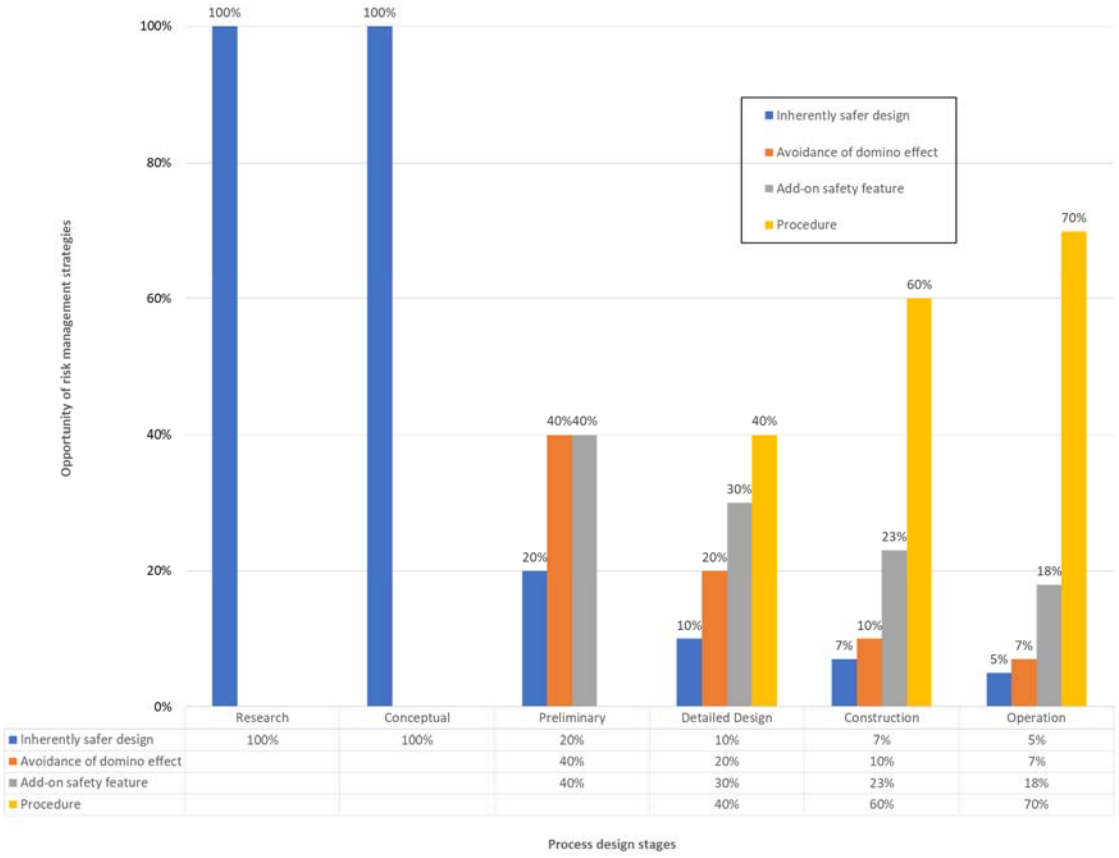
Figure 2.2 Process risk management system incorporating three strategies: inherently safer design, avoidance of domino effects, add-on safety and procedures (Modified from Kletz and Amyotte (2010) and CCPS (2010))

Figure 2.2 shows two approaches of ISD: (1) the 1st order ISD approach to eliminate hazards in a rigorous approach, and (2) the 2nd order ISD approach to reduce hazards (or risks) with its four guidewords (CCPS, 2010). Because the 1st order approach is not always applicable, the 2nd order approach should be the alternative. However, for both methods, eliminating or reducing hazards – the potentials – is more effective compared to reducing risks, as a risk is the result of a specific situation with uncertainty as mentioned in Section 2.2.1. Without sufficient information for the relevant situation, estimating and reducing risk is essentially impossible. However, if ISD compares two or multiple alternatives based on the same or similar conditions, researchers tend to prefer the 2nd order approach for risk reduction.

2.2.4. Importance of adopting ISD principles during the conceptual design stage

Process design proceeds through several stages starting with research and development followed by conceptual process synthesis and design where the basic chemical routes and primary functional steps are determined along with mass and energy balances and high-level techno-economic analysis. During conceptual design, the process engineers have substantial freedom in making design decisions. Next, detailed design involves thorough analysis, simulation, assessment, and optimization that will define the specifications necessary for construction then operation. When feasible during early design stages, the principal concepts of ISD are well aligned with sustainability. As can be seen from Figure 2.3, early design stages have much more flexibility to incorporate ISD features than the subsequent stages. Therefore, the opportunity to adopt ISD is ideal

at the research and conceptual design stages; such opportunity decreases, and the project cost increases if changes are made during the subsequent design stages. Once a conceptual design is completed, the other safety strategies should be applied along with ISD. However, in this case, the project cost would significantly increase to have the same risk level at the same reliability relative to if ISD was adopted during the conceptual design stage.



	Research	Conceptual	Preliminary	Detailed Design	Construction	Operation
Inherently safer design	100%	100%	20%	10%	7%	5%
Avoidance of domino effect			40%	20%	10%	7%
Add-on safety feature			40%	30%	23%	18%
Procedure				40%	60%	70%

Figure 2.3 Typical opportunities of risk management strategies during chemical plant projects

Nevertheless, implementing ISD may meet difficulty in the conceptual design stage for two main reasons: (1) there is insufficient time to incorporate ISD in a traditional approach, and (2) the limitation of available information (Table 2.3) to consider all possible hazardous scenarios. Once process designs are determined in basic or detailed engineering stages, process engineers develop the relevant documents. Traditionally, plant project engineers undertake safety analysis with piping and instrument diagrams (P&IDs) following basic engineering design with one of the hazard-identification tools (e.g., HAZard and OPerability (HAZOP), Failure Mode and Effect Analysis (FMEA) or What-if). After preliminary engineering design, it is quite difficult to change the main process or materials to increase ISD. Consequently, only opportunities for installing add-on safety features will increase instead, as shown in Figure 2.3 which shows typical values of opportunities for mitigating risk (expressed as percentages compared to 100% opportunities during research and conceptual design). Table 2.3 represents typically available information per each plant design stage (Towler and Sinnott, 2012). Although the available information is limited, it is crucial to apply the available information prudently to perform ISD effectively.

Table 2.3 Available information per each plant design stage (Modified from Towler and Sinnott, 2012)

Plant Design Stage	Available Information
Research concept	Chemistry, MSDS information
Conceptual design	Process flow diagram (PFD), Equipment list, Vessel designs, Reactor models, Simulation modeling
Preliminary design	P&IDs, Process control scheme, Metallurgy, Detailed mass and energy balance, Hydraulics, Offsites
Detailed design engineering	Mechanical designs, Instrument specs, Vendor details, Plot plans
Procurement, construction	Piping isometric, As built specs
Operation	Commissioning log, Operations log, Maintenance log

2.3. Necessity of this chapter

As stated in Section 2.2, the effect of ISD is maximized when ISD principles are applied during the conceptual design stages. Thus, process engineers must be consciously aware of ISD and apply it with robust safety evaluation tools. Without robust tools, engineers may fail to discover a potential hazard or risk, resulting in unsafe design.

By the early 1990s, multiple safety assessment tools had been developed such as HAZOP (Cameron et al., 2017), Dow (1994) Fire & Explosion Index (F&EI), Mond Index (Tyler, 1985), and Safety Weight Hazard Index (SWeHi) (Khan et al., 2001). Although these tools have contributed to the process safety field, they are inappropriate to use during the early design stages; however, they can be applied conditionally during or after the preliminary engineering stage. For example, a typical HAZOP study is performed with

P&IDs and plot plans that are made after finalizing an accurate process design scheme. As the most commonly used safety index, Dow F&EI has functioned to analyze safety for basic process design (Qi et al., 2019; Vázquez et al., 2017; Al-Mutairi et al., 2008; Suardin et al., 2007). However, this index tool also requires detailed information (e.g., equipment sizes, material inventories, flows from P&IDs, and plot plans), which may require many assumptions to obtain a total safety index value. Furthermore, Dow, (1994) examines hazard, flammability, and reactivity based on NFPA (1991). Because the NFPA standard is based on the standard process conditions—defined as a temperature of 20 °C and an absolute pressure of 1 atm—it does not enable the consideration of chemicals existing in different process conditions. The Mond Index (Tyler, 1985) also requires more detailed process design schemes (e.g., pipelines longer than 25 m are regarded as separate process units (Edwards and Lawrence, 1993)). Lastly, SWeHi requires information obtained during the detailed engineering stage such as plant layout, safety system, and barriers (Roy et al., 2016). Consequently, analyzing proper ISATs is necessary to maximize ISD principles during the early design stages.

Many researchers have focused on safety indices because of their quick and straightforward attributes (Abidin et al., 2018). A *safety index* is a tool that assesses the relative safety level among alternative processes using safety indicators (or parameters) that can differ depending on the index used. A safety indicator, typically represented as a numeric value, is the fundamental metric that represents the degree of safety of a process. The lower the value of a combination of indicators, the safer the process design is compared to alternatives.

2.3.1. Recent review papers relevant to inherent safety assessment tools (ISATs)

There have been several review papers discussing how ISATs have been developed. To provide coverage of the topic, five representative review papers focusing on various ISATS were selected. For each paper, the suggested tools were checked for potential adoption and utilization in the conceptual design stage.

Srinivasan and Natarajan (2012) examined the key developments of inherent safety with 111 articles published between 2000 and 2011. This article divided the research topics for inherent safety into eight sub-topics, as shown in Table 2.4.

As shown in Table 2.4, the authors considered ISD in the framework of risk management rather than in a strictly technical sense. Hence, this article expanded the scope of ISD beyond the inherently safer principles of Kletz and Amyotte (2010) by including the research of layout, add-on safety features, and human factors. Such expanded scope with several risk management features is beyond the scope of this paper.

Table 2.4 Eight sub-topics for inherently safety research (Adopted from Srinivasan and Natarajan, 2012)

Sub-topic of inherent safety	Note
Material	Defined as chemical's intrinsic characteristics
Chemistry	Defined as chemical reactions including runaway concerns
Unit operation	Mainly intensification such as microreactor systems

Sub-topic of inherent safety	Note
Flowsheet and layout	Including domino effects
Storage and transportation	Referred to as the inventory of hazardous materials stored
Process control and safety system	Including safety instrumented system (SIS), safety integrity level (SIL), and pressure safety valve (PSV)
Human factors and management systems	Significantly related to the likelihood of events
Assessment techniques	Shown as ISD indices

Roy et al. (2016) analyzed 25 representative safety indices for their viability at each stage of a lifecycle in a plant project. The analysis was performed based on the indicators required for executing the safety indices. In the authors' result, only 1 index among the 25 indices was adequate to be utilized during the conceptual design stage. However, this result can vary depending on the analysis approach on the safety indices. Although Roy and his co-authors regarded inventory as an indicator that can be obtained only with data after the conceptual design stage, it can be easily estimated during the conceptual design stage. For example, available flow rate data (e.g., kg/hr. or ton/hr.) obtained from a process simulation software can be exploited to estimate the potential chemical amount of an inventory (e.g., equipment or pipelines). With an emphasis on inventory, 9 out of 25 indices were selected to be adequate for usage during the conceptual design stage. The nine indices are:

- Mond Index (Lewis, 1979)

- PIIS Index (Edwards and Lawrence, 1993)
- ISI (Heikkilä, 1999)
- EHS Index (Koller et al., 1999)
- GreenPro-I (Khan et al., 2002)
- Life Cycle Index (Khan et al., 2004)
- Process Route Index (Leong and Shariff, 2009)
- Process Stream Index (Shariff et al., 2012)
- Risk-based Ranking of Hazardous Thermal Chemical (Busura et al., 2014).

It should be noted that the Mond Index should be excluded, given it was already deemed unsuitable, as mentioned in the beginning of Section 2.3. As an extension, this chapter will also analyze these 8 indices to determine their viability during the conceptual design stage.

Jafari et al. (2018) reviewed inherent process safety indicators through 62 peer-reviewed articles published between 1990 and 2017. Notably, this study analyzed two main aspects—ISD indicators and estimation approaches for ISD. As Figure 2.4 shows, the authors divided all indicators they found in the articles into six categories: (1) chemical and physical properties of materials, (2) process conditions, (3) reaction properties, (4) equipment, (5) types of activities/operations, and (6) consequences. Figure 2.4 presents the indicators the authors found in the 62 articles with black-color lining; possible indicators proposed by the authors, but not seen in the articles, are presented with red color lining. Furthermore, the authors divided the estimation approaches into six categories:

- Hybrid approach
- Equation-based approach
- Graphical approach
- Advanced mathematical approach
- Risk-based approach
- Relative ranking

Based on these six categories, the authors developed a table on how each indicator was estimated. However, the authors did not limit the concept of ISD as introduced by Kletz and Amyotte (2010) and included several articles associated with domino effects in their review—which is also out of the scope in this paper (see Table 2.2 in Section 2.2.4). For example, it may be challenging for process engineers to select appropriate indicators to use in order to evaluate the safety status. As an extension of the article, this chapter will pursue more practical comparisons among inherently safer indicators for improved practical applications.

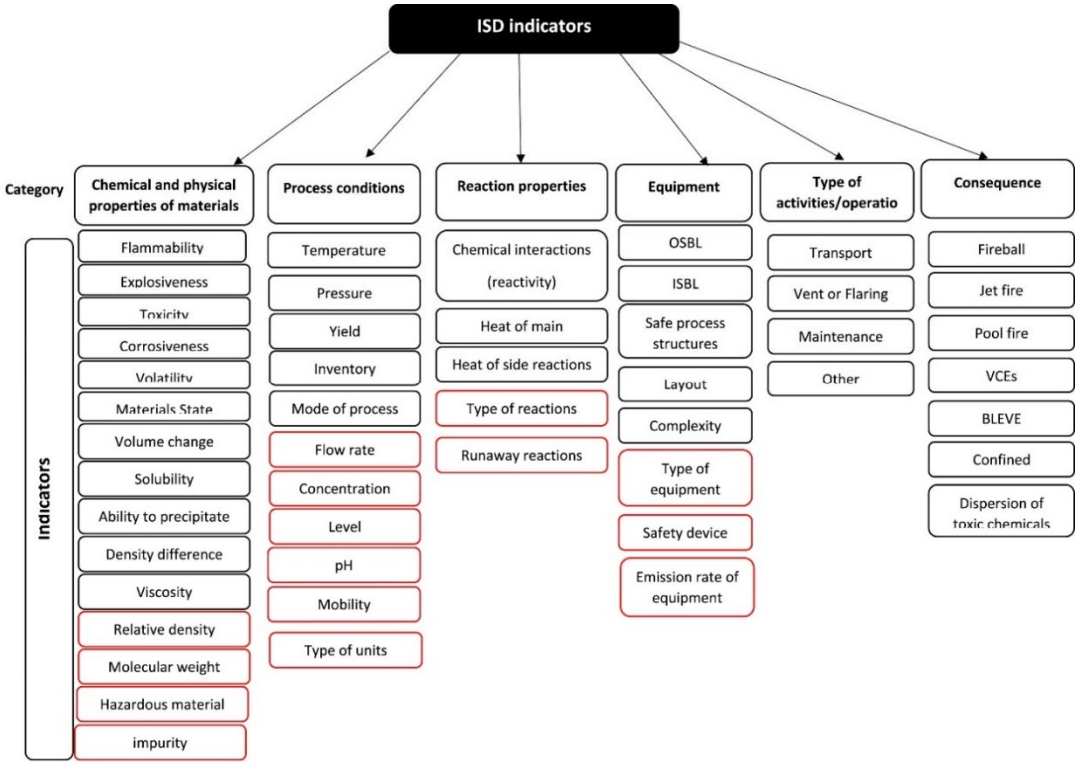


Figure 2.4 Summary of all indicators under six categories (Adopted from Jafari et al. (2018)[2]).

Tang et al. (2018) conducted a literature review of safety indicators commonly used in the offshore oil and gas industry. However, there was no specific mention about ISATs.

Athar et al. (2019) reviewed 93 ISATs extracted from 387 articles. The authors followed Kletz's ISD concept by analyzing ISATs that can be used during the early design stages. The authors found that a two-thirds majority of ISATs developed were hazard-

[2] Reprinted from Jafari, Mohammad Javad, Mohammadi, Heidar, Reniers, Genserik, Pouyakian, Mostafa, Nourai, Farshad, Ali Torabi, Seyed, Miandashti, Masoud Rafiee, 2018. 'Exploring inherent process safety indicators and approaches for their estimation: a systematic review. J. Loss Prev. Process. Ind. 52, 66–80, with permission from Elsevier

based tools rather than risk-based tools. The reviewed tools were tabulated to represent their Health, Safety, and Environment scope and output features in checkboxes under the following seven method categories:

- Index (57%)
- Consequence (13%)
- Graphical (7%)
- Numerical (5%)
- Computer-aid (15%)
- Optimization (3%)
- Experimental (1%)

The percentage in parentheses represents the respective proportion of total ISATs for each method. However, it is likely too difficult to divide ISATs based on the suggested categories due to many hybrid methods among the seven methods (e.g., indexing with computer-aid, indexing and graphical, or consequence and computer-aid).

2.4. Method of reviewing literature

As seen in the background information provided in the previous sections, this chapter aims to analyze the characteristics of inherent safety assessment tools (ISATs), which can be viable in the conceptual design stage. Various inherent safety tools were selected and reviewed. Afterward, the selected ISATs were categorized into risk management system terms, ‘hazard and risk’ (Table 2.1); These categories will help

practitioners smoothly adopt ISATs as the approach aligns closely with a conventional risk management system. Furthermore, this chapter will explore how economic, safety, and sustainability can be optimized and reconciled during the conceptual design stage.

In addition to the 291 articles from the previous review papers, further journal papers were retrieved based on the relevance to the proposed inherent safety tools or cost optimization with safety through the Scopus database or ACS website. Among those journal papers, ones that can be applied during the conceptual design stage were selected (see Table 2.3). The primary scopes of reviewing journal papers in this chapter are safety or cost optimization with safety. Inclusion or exclusion of specific ISATs was based on its relevance to conceptual design. For instance, a tool requires details that are not typically available during conceptual design, then it was excluded. The search for papers has been ended in the third week of March 2019.

2.5. Inherent safety assessment tools (ISATs) for the conceptual design stage

2.5.1. The division of three ISATs

This chapter analyzes ISATs based on the three categories shown in Figure 2.5: Hazard-based ISAT (H-ISAT), Risk-based ISAT (R-ISAT), and Cost-Optimal ISAT (CO-ISAT) (see Figure 2.6).

The primary purpose of each ISAT is denoted with bold letters in Figure 2.5. The purpose of H-ISATs has been to rank inherent hazard levels among alternatives, while R-ISATs have selected safer process design(s) via the measures of consequence and

likelihood. Finally, CO-ISATs have proposed optimal decision-making for sustainability with economic analysis and safety constraints.

Figure 2.5 shows the essential and supplementary steps for each ISAT distinguished with straight and dotted lines, respectively. For example, the consequence analysis in R-ISAT is an essential step in selecting an inherently safer process design, while the calculation of damage distance is a supplementary step. The connections between other ISATs are not requisite steps, so all connection lines are shown as dotted lines. The dotted lines between different ISATs represents multi-ISAT frameworks that have been proposed by researchers (e.g., ranking inherent levels in H-ISATs, followed by estimating inherent risk levels in R-ISATs).

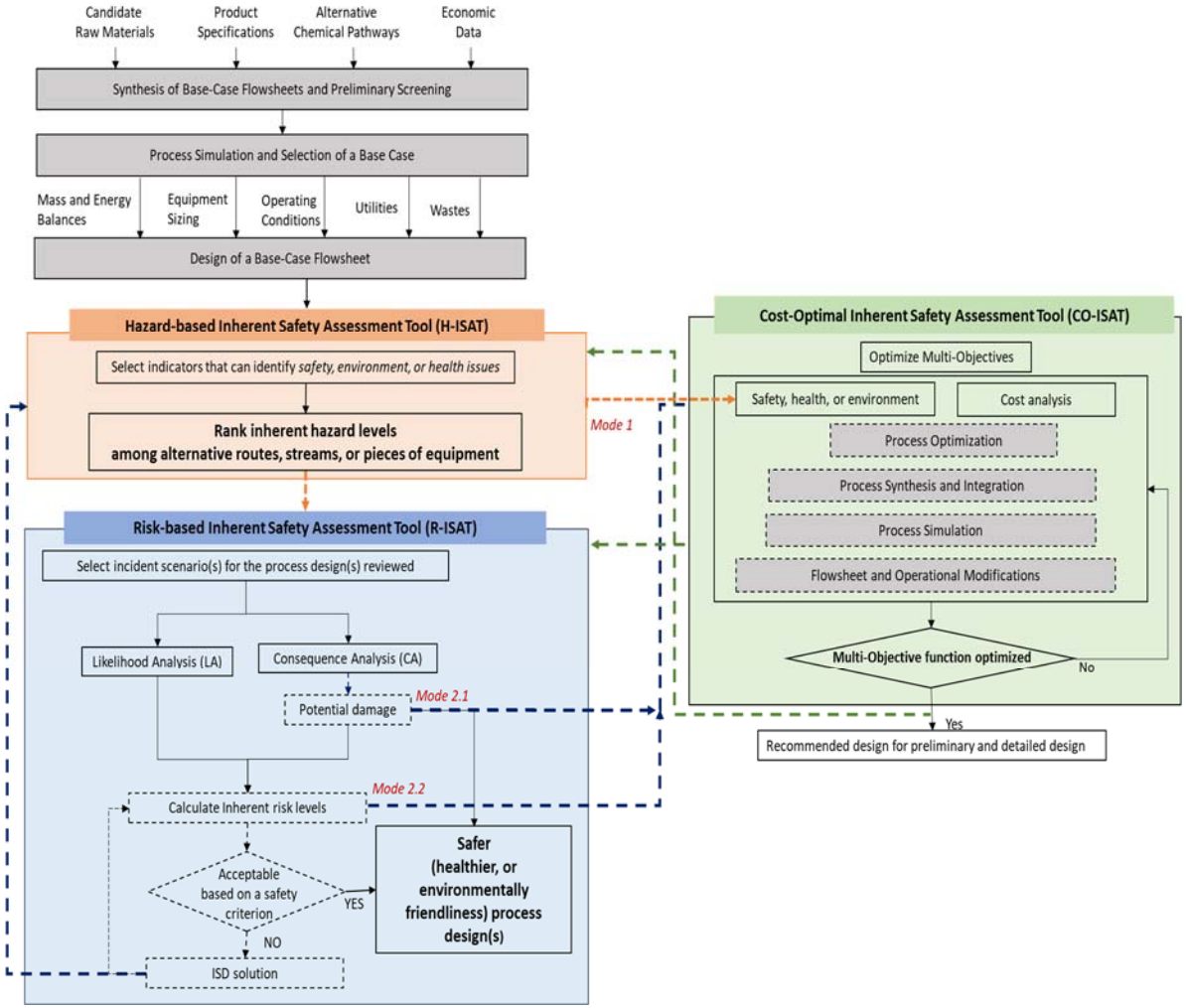


Figure 2.5 Schematic representation of three inherent safety assessment tools for the conceptual design stage

ISATs may be used independent of each other or in a specific sequence. The usage of the three ISATs can be described by the following use cases: (1) H-ISAT followed by CO-ISAT, (2) R-ISAT followed by a CO-ISAT, or (3) H-ISAT followed by R-ISAT, followed finally by CO-ISAT. The R-ISATs and CO-ISATs have more complex steps in comparison with H-ISAT. H-ISATs solely estimate inherent safety level via selected indicators on a comparative basis. In some cases, the procedure to determine H-ISATs is used as part of R-ISATs. However, R-ISATs select a safer process design based on two modes. The first mode, Mode 2.1, utilizes damage distance based on potential consequence. Meanwhile, the second mode, Mode 2.2, utilizes inherent risk level along with the measure of consequence and its likelihood. Moreover, CO-ISATs can have three types of safety constraints from (1) Mode 1: inherent hazard levels via H-ISATs, (2) Mode 2.1, or (3) Mode 2.2, which are ordered in increasing effort from (1) to (3). The choice will depend on the severity of the safety situation.

According to the above-mentioned points, the detailed analysis for each ISAT will be described in the following subsections.

2.5.2. Hazard-based inherent safety assessment tools (H-ISATs)

This study selected 22 H-ISATs that contains limited but viable information applicable during the conceptual design stage. Most of the selected H-ISATs were proposed for the safety scope except for two tools (IOHI in Table 2.5 #8, and graphical method for inherent occupational health assessment in Table 2.5 #11), which are for health. Table 2.5 lists the H-ISATs in chronological order to provide an overview with

five primary characteristics: (1) types of H-ISATs, (2) purpose of each tool, (3) computer aid, (4) types of chemical component, and (5) process design scopes compared for inherent safety levels.

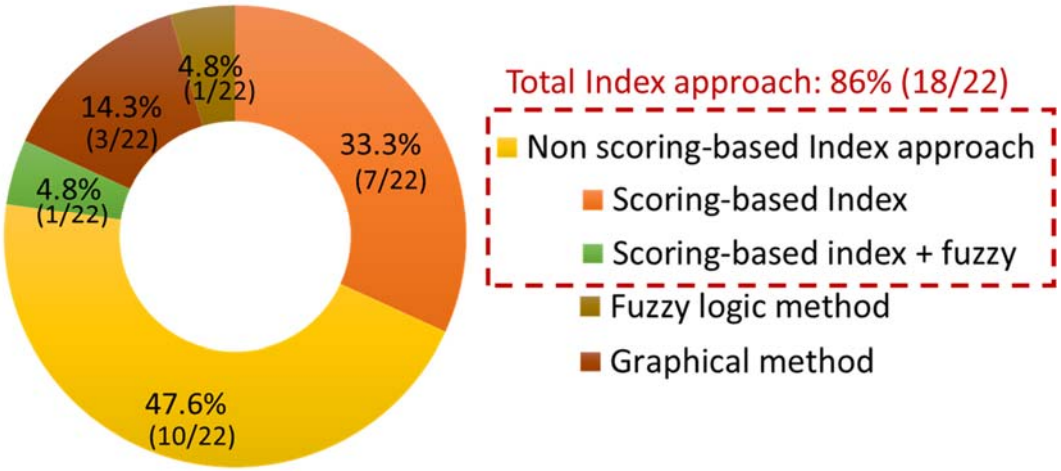


Figure 2.6 Distribution of different types among 22 selected H-ISATs

The selected 22 H-ISATs have been characterized into five different types of methods (Figure 2.6); index methods were the most commonly proposed methods followed by graphical methods, a fuzzy logic method, and a hybrid index plus fuzzy logic method.

Table 2.5 List of 22 Hazard-based Inherent Safety Assessment Tools (H-ISATs) selected

No.	Authors (Year)	H-ISAT		Remark	Computer aid	Chemical component		Process design scope compared for inherent safety levels			
		Name	Type [a]			Pure	Mixture	Route	Stream	Equip-ment	etc.
1	Edwards and Lawrence (1993); Lawrence (1996)	Prototype Index for Inherent Safety (**PIIS**)	S	Assessed the inherent safety of chemical process routes		√		√			
2	Heikkilä (1999)	Inherent Safety Index (**ISI**)	S	Proposed a practice that can be adopted quickly by the potential practitioners		√		√			
3	Palaniappan, Srinivasan, and Tan (2002a); (2002b)	**i-Safe**	S	As an extension of PIIS ISI, proposed an automated tool including five supplementary indices were proposed for rigorous ranking	√	√		√			
4	Gupta and Edwards (2003)	**Graphical method**	G	Avoided equalizing the same numerical value among different indicators		√		√			
5	Gentile, Rogers, and Mannan (2003);	**Fuzzy logic** based inherent safety **index**	F	Minimized uncertainty in scoring index with fuzzy	√	√				√ Reactor	√

No.	Authors (Year)	H-ISAT		Remark	Computer aid	Chemical component		Process design scope compared for inherent safety levels			
		Name	Type [a]			Pure	Mixture	Route	Stream	Equip-ment	etc.
	Gentile (2004)			logic methods, based on ISI.							
6	Leong and Shariff (2008)	Inherent Safety Index Module (**ISIM**)	S	Simplified process equipment that have unsafe level compared to others	√	√			√	√ Reactor	
7	Leong and Shariff (2009)	Process Route Index (**PRI**)	N	Estimated the outcome of an explosion in a non-scoring system	√		√	√			
8	Hassim and Hurme (2010a)	Inherent Occupational Health Index (**IOHI**)	N	Estimated inherent occupational health hazard using three different types of index calculations: additive-type, average-type, and worst case-type.		√		√			
9	Li et al. (2011)	Enhanced Inherent Safety Index (**EISI**)	S	Overcame ISI limitations—only considering maximum value for indicators and not	√	√		√		√	

No.	Authors (Year)	H-ISAT		Remark	Computer aid	Chemical component		Process design scope compared for inherent safety levels			
		Name	Type [a]			Pure	Mixture	Route	Stream	Equip-ment	etc.
				considering complexity							
10	Shariff, Leong, and Zaini (2012)	Process Stream Index (**PSI**)	N	Compared individual streams at one single process route	∨		∨		∨		
11	Hassim et al. (2013)	Graphical method for inherent occupational health assessment	G	Developed based on IOHI and experts' comments		∨		∨			
12	Gangadharan et al. (2013)	Comprehensive Inherent Safety Index (**CISI**)	S	Proposed a connection score among process steps as an extension of EISI	∨		∨	∨			
13	Zaini, Azmi, and Leong (2014)	Toxic Release Route Index (**TRRI**)	N	Compared toxicity among multiple process routes	∨	∨		∨			
14	Zaini, Azmi, and Leong (2014)	Toxic Release Stream Index (**TRSI**)	N	Compared toxicity among all process streams in a route after estimating TRRI	∨	∨			∨		

No.	Authors (Year)	H-ISAT		Remark	Computer aid	Chemical component		Process design scope compared for inherent safety levels			
		Name	Type [a]			Pure	Mixture	Route	Stream	Equip-ment	etc.
15	Ahmad, Hashim, and Hassim (2014)	Numerical Descriptive Inherent Safety Technique (**NuDIST**)	N	Overcame the disadvantages of index-based methods that evaluated inherently safer level with discrete safety indicators		∨		∨			
16	Ahmad et al. (2016)	**NuDIST**	N	Considered flammability limits for a mixture			∨	∨			
17	Ahmad, Hashim, and Hassim (2016)	Graphical and Numerical Descriptive Technique for Inherent Safety Assessment (**GRAND**) 2-Diemnsional Graphical Rating (**2DGR**)	G	Ranked, through visualization, inherent safety levels Identified the most hazardous chemical		∨		∨			
18	Pasha, Zaini, and Shariff (2017)	Inherent Safety Index for Shell and Tube Heat Exchanger (**ISISTHE**)	N	Calculated the inherent safety level of explosion for a Shell and Tube Heat Exchanger (STHE)	∨		∨			∨ (H/E)	

No.	Authors (Year)	H-ISAT		Remark	Computer aid	Chemical component		Process design scope compared for inherent safety levels			
		Name	Type [a]			Pure	Mixture	Route	Stream	Equip-ment	etc.
				through a similar approach of PSI							
19	Pasha, Zaini, and Shariff (2017)	Overall Safety Index for Heat Exchanger Network (**OSIHEN**)	N	Considered heat exchanger network	∨		∨	∨		∨ (H/E)	
20	Ahmad, Hashim, and Hassim (2017)	Inherent Safety Assessment Technique for Preliminary Design Stage (**ISAPED**)	N	Updated a tool as an extension of NuDIST and GRAND			∨			∨	
21	Song, Yoon, and Jang (2018)	Inherent Safety Performance Indices (**ISPI**)	S+F	Proposed a framework based on fuzzy logic methods with an index method	∨		∨	∨			
22	Athar et al. (2019)	Process Stream Characteristic Index (**PSCI**)	N	Assessed process piping as an extension of PSI	∨		∨		∨		

[a] S: Index type in a scoring system, N: Index type in a non-scoring system, G: graph approach, and F: Fuzzy logic approach, H/E: heat exchanger

59% of the tools in Table 2.5 adopted computer-aided methods to efficiently deal with large amounts of data. Prior to utilizing a computer simulator, it has been stated a challenge to obtain feasible process data, such as inventory values. In this manner, required overall yield was proposed to indirectly quantify the amount of chemicals in a process or supplement an inaccurate inventory value (e.g., PIIS, Table 2.5 #1). However, enabling practitioners to obtain process data through computer software reduces the time and effort required to collect indicator data in a process. For example, integrating data between a process simulator (e.g., Aspen Plus or HYSYS) and Visual Basic for Application (VBA) language can save significant time and streamline the safety analysis process.

Early H-ISATs calculated inherent safety levels with only pure chemicals, while several researchers have considered chemical mixtures since 2009.

It is worth defining different process scopes—process routes, steps, units, and streams—to compare H-ISATs, as authors typically developed their own methods with varying definitions. The following selected definitions will help clarify the H-ISATs mentioned in Table 2.5:

- Process route: an overall process compiled of a series of process steps—equipment and streams—converting raw materials to a required product and waste via intermediate materials (Lawrence, 1996)
- Process step: a partial process route that can be characterized based on its process functions (e.g., reaction steps and distillation trains), which typically includes multiple pieces of process equipment

- Process unit: defined, according to ISO (2006), as the "smallest element considered in the life cycle inventory analysis for which input and output data are quantified." The process unit was referred to as a process step by Palaniappan, Srinivasan, and Tan (2002a), whereas Li et al. (2011) and Gangadharan et al. (2013) referred to it as process equipment. To avoid confusion, this paper avoids usage of the "process unit" term
- Process stream: referred to as a process pipeline

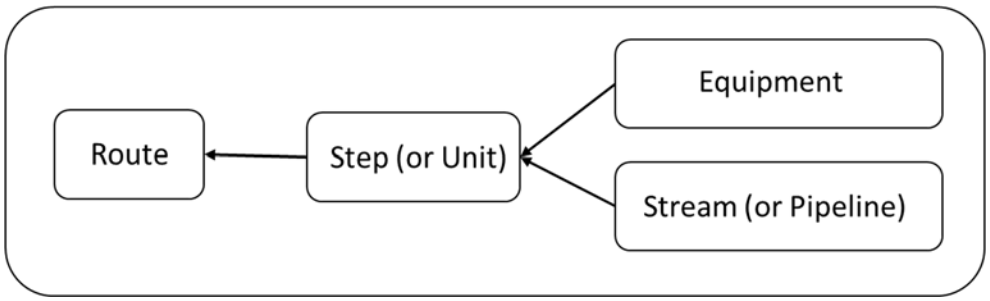


Figure 2.7 Process route configuration with process step (or unit), equipment, and stream

Table 2.5 also compared process scopes to estimate inherent safety levels (see Figure 2.7). Most of the initial H-ISATs compared multiple process routes to determine the safer process routes, while several later H-ISATs compared multiple streams or pieces of equipment within a process route to estimate safety levels. Figure 2.8 shows the process used in the case studies among 20 H-ISATs. Many authors adopted the same methyl Methacrylate process as it was used in the pioneer H-ISAT, PIIS, to have a reference to compare with their new method.

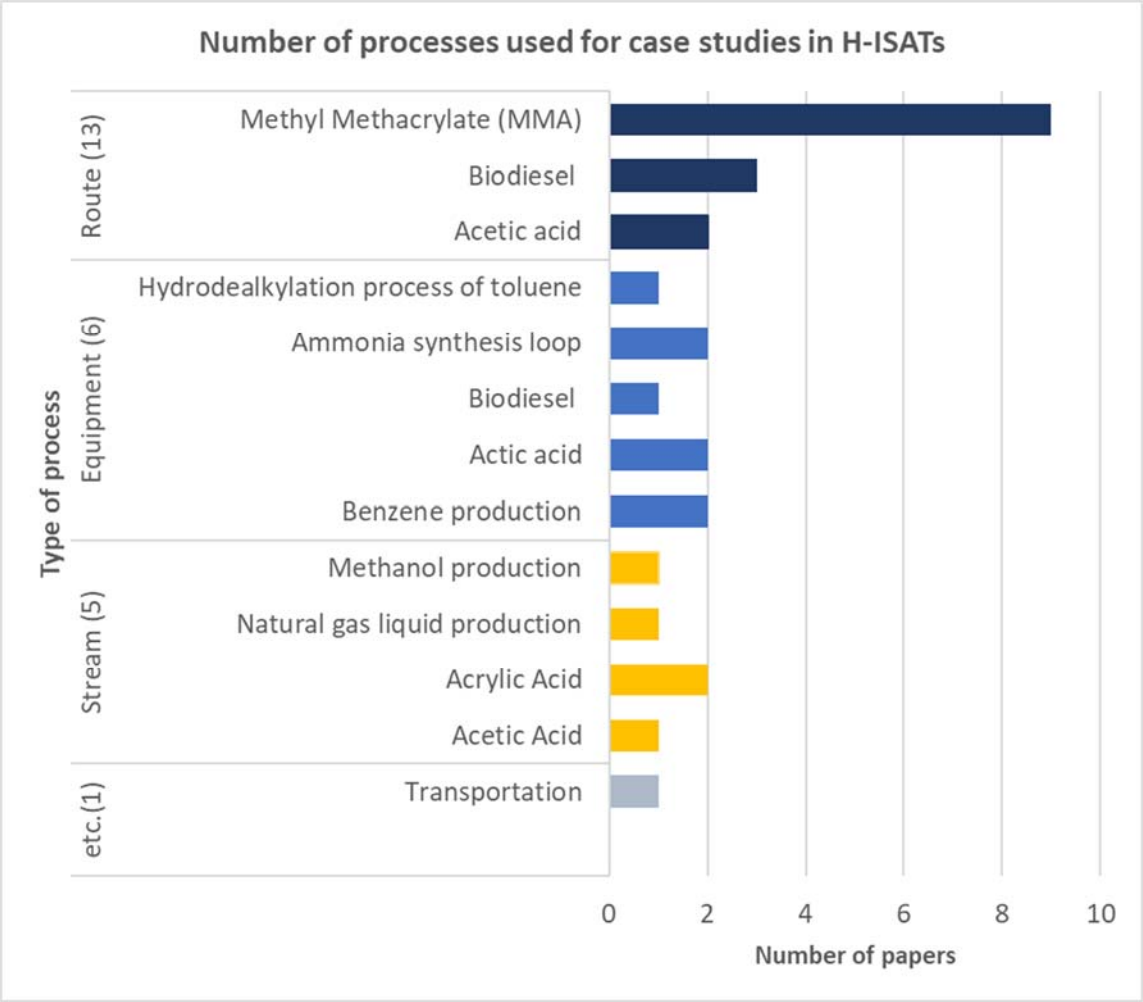


Figure 2.8 Histogram of process used in the case studies in 22 H-ISATs

All 18 indices in H-ISATs measure an inherent safety level with multiple safety indicators that can be classified into two categories: chemical indicators and process indicators. Chemical indicators account for the hazard presented by the chemicals in the process, while process indicators refer to the hazard presented by the operating conditions in the process. However, researchers have not clearly defined the differentiating factors between chemical indicators and process indicators. For instance, Edwards and Lawrence

(1993) regarded inventory as a chemical indicator, while Heikkilä (1999) regarded it as a process indicator. Further, heat of reaction was considered as a chemical indicator by Heikkilä (1999), while it was considered as a process indicator by Ahmad et al. (2014). This paper will consider whether an indicator is a process or chemical indicator based on the majority views among 18 indices. As a result, there were 13 chemical indicators and 7 process indicators as shown below:

For the 11 chemical indicators.

1) Flammability: the tendency of a material to combust in air (Heikkilä, 1999)
2) Explosiveness: ability to form an explosive mixture with air
3) Density: used to explain the principle of fluid mechanics that can impact during an incident
4) Lower heat of combustion: measure of energy
5) Toxicity: property that destroys life or causes injury, measured by the threshold limit (or exposure limit or R-phrase for the occupational health hazard in Hassim and Hurme (2010a))
6) Chemical interaction: consideration of unintended reaction among raw materials, intermediates, and byproducts (Palaniappan, Srinivasan, and Tan, 2002b)
7) Corrosiveness: associated with the reliability and integrity of a chemical plant, as corrosion increases the risk of leaks in a process by lessening the strength of an equipment material (Heikkilä, 1999)
8) Material phase (gas, liquid, and solid): related to the way chemicals will be handled and exposed (Hassim and Hurme, 2010a). For example, solid processes are likely

to be performed manually so that it has more changes to have higher exposure to workers

9) Reactivity: a measure of stability (Palaniappan, Srinivasan, and Tan, 2002a)
10) Heat of reaction (heat of main reaction and heat of side reaction): a measure of the hazard of the reaction and the possibility to control the reaction
11) Viscosity: a measure of the mass of the fluid flow regime inside process piping (Athar et al., 2019)
12) Volatility: related to inhalation and absorption in terms of occupational health hazard. It can be characterized by the vapor pressure or atmospheric boiling point in a liquid substance, or by a smaller dusk in a solid substance (Hassim and Hurme, 2010a)
13) Type of reaction: a measure of reactivity to supplement heat of reaction that cannot indicate all the reaction characteristic (Song et al., 2018)

For the 7 process indicators (Lawrence, 1996; Heikkilä, 1999).

1) Temperature: a measure of the thermal energy that cause a fire or an explosion (Srinivasan and Nhan, 2008) or of burns to humans (Hassim and Hurme, 2010a)
2) Pressure: a measure of the energy present that cause materials to escape from the process within a relatively short time

3) Inventory: the amount of process material that is closely related to reaction rate or conversion in a reactor. Inventory is the key indicator of the inherent safety principle, intensification
4) Type of equipment (ISBL and OSBL): a measure of the possibility that a piece of equipment is unsafe
5) Mode of process: three types of process modes—batch, semi-batch, and continuous process—can be related to workplace exposure (Hassim and Hurme, 2010a)
6) Yield: a measure of inventory or flow needed to obtain a required product
7) Structure: a description of inherent safety associated with the process configuration

Figure 2.9 shows the number of indicators used among 18 indices in H-ISATs. In terms of the number of indicators, the most used indicator was pressure, followed by temperature, explosiveness, and toxicity. On the other hand, some indicators (e.g., viscosity, type of equipment, and structure) were used only once.

Various indices in H-ISATs have been developed with either a scoring system or a non-scoring system. Indices in a scoring system grade the indicators depending on their scales (Appendix 2A), whereas indices in a non-scoring system utilize the actual indicator values in a process without any scoring (Appendix 2B). The two tables rigorously compare scores and equations of individual indices based on chemical and process indicators mentioned above.

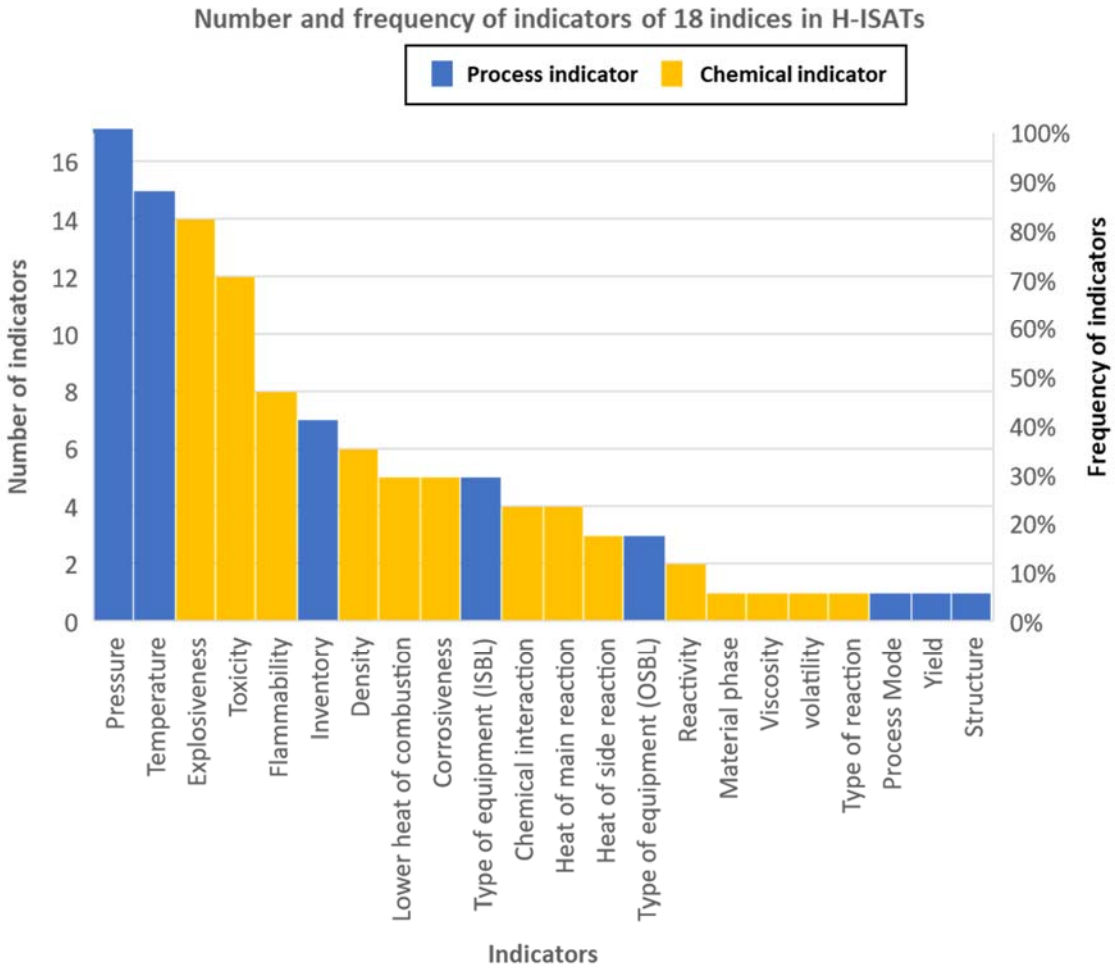
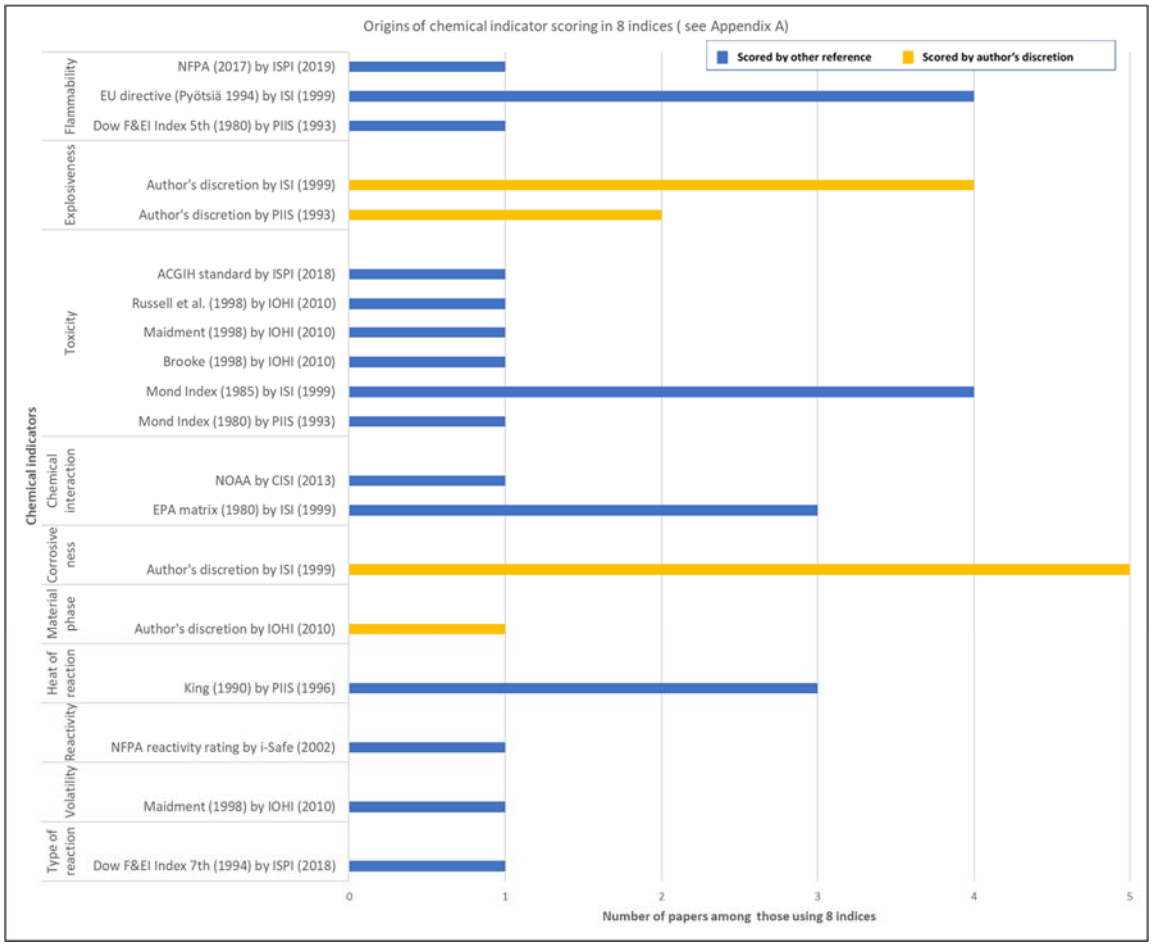


Figure 2.9 Histogram of indicators in H-ISATs

Appendix 2A compares 8 indices that score individual indicators via a sub-range to calculate total safety scores. Initially, indices in the scoring system served as a yardstick that measures inherent safety with possible indicators and their ranges. The range of each indicator must be reasonable to effectively compare the indicator among different process

routes, steps, units, or streams. Because the calculation approaches are highly dependent on the scoring technique, it is imperative to understand the differing guidelines of each index. Figure 2.10 represents the origins of the scoring systems among individual indicators. Preexisting scoring standards (e.g., Dow F&EI and Mond), adopted by a group of authors for one of the indices, are shown by blue colored bars, while authors' own discretional choices are shown by yellow-colored bars. The x-axis addresses the number of each scoring scale used among 7 indices. The analysis of the results in Figure 2.10 illustrates that most base standards used were proposed by PIIS and ISI, and these standards were reused in other indices. Three new methods were proposed as alternatives to scoring-based indices: a fuzzy logic method, non-scoring-based indices, and graphical methods. Because each indicator in the scoring-based indices results in subjective ranges and discontinuities at the range limits, the alternatives were proposed in an attempt to overcome these limitations. First, Gentile et al. (2003) attempted to minimize the uncertainty of scaled boundaries with a fuzzy logic method. Second, ten non-scoring-based indices (Appendix 2B) utilized actual chemical and process condition values for each indicator, instead of the value based on a scoring system. For example, operating pressure values were exploited in all ten indices in Appendix 2B. Even though Ahmad et al. (2014) distinguished their method from a scoring-based index method, NuDIST, which is based on statistical analysis, was classified in this paper into a non-scoring index. This is because the method is in accordance with our definition of a safety index in Section 2.3. Lastly, graphical methods were used in an attempt to visualize an inherent safety level instead of showing it with numbers. For example, one graphical method separately plotted

five key indicators (temperature, pressure, flammability, explosiveness, and toxicity) (Gupta and Edwards, 2003), while another 2-dimensional graphical rating, 2DGR, simultaneously plotted the frequency of the most hazardous chemicals with their process route and a total safety score for each route (Ahmad et al., 2016a).



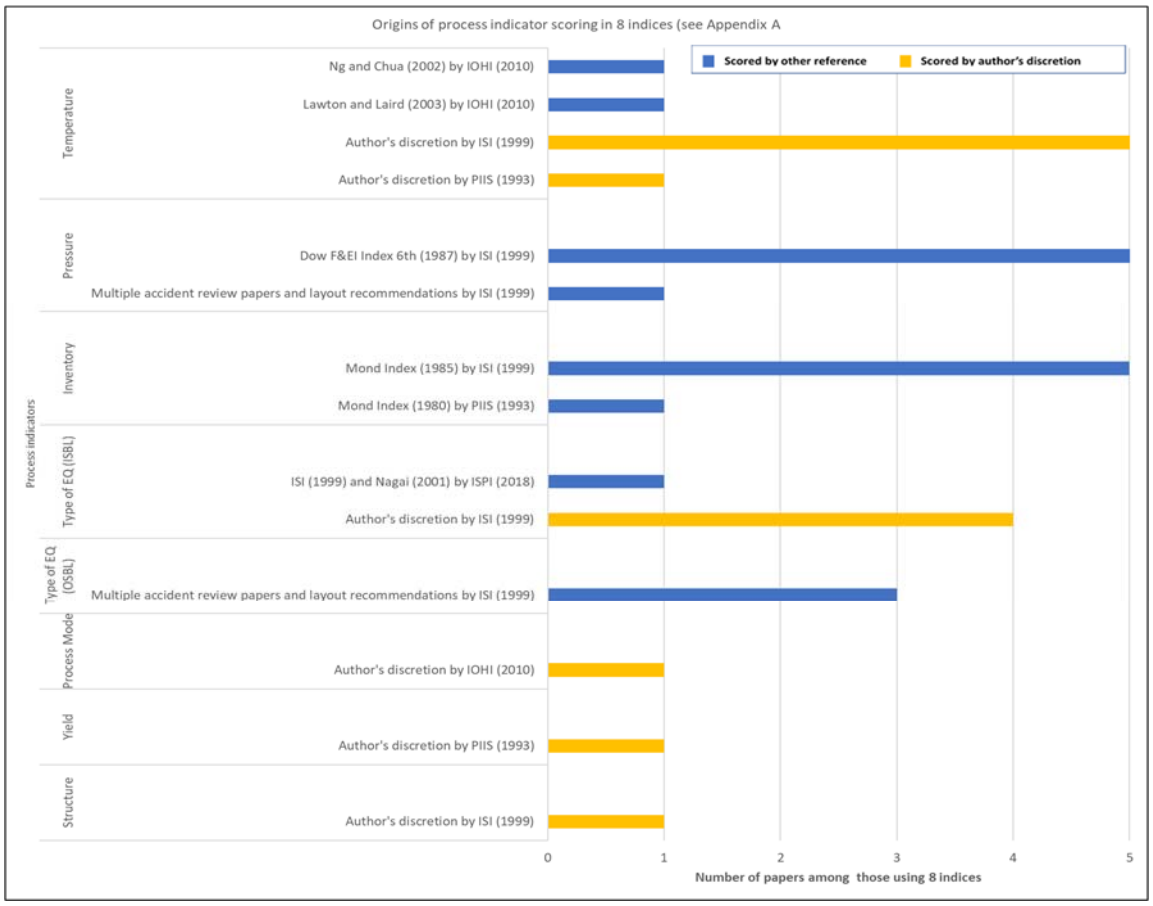


Figure 2.10 Papers with scoring scales of 8 indices in H-ISATs

2.5.3. Risk-based inherent safety assessment tools (R-ISATs)

The primary limitations of H-ISATs are the inability to illustrate potential incident scenarios and their damage impacts. H-ISATs aggregate multiple indicators for different types of hazards, which can be related to potential incidents. However, they do not cover the entirety of ISD.

To resolve these limitations, many researchers have developed R-ISATs, which can consider concrete incident scenarios through different approaches in the conceptual

design stage. Table 2.9 summarizes 33 selected R-ISATs. Although researchers have sought to propose R-ISATs as an easy-to-use method, multiple extra steps are unavoidable for R-ISATs in comparison with H-ISATs. For this reason, several frameworks—including expanded H-ISATs—also were proposed. The column "ISAT used" in Table 2.9 helps to understand the proposed framework as an extension step of H-ISATs.

The proposed R-ISATs can be grouped into three types of approaches: (1) qualitative-based, (2) index-based, and (3) quantitative-based approaches. The main difference among the three approaches is the way the incident scenarios are considered. Qualitative approaches depend on the practitioner's creative ideas and knowledge to obtain possible scenarios. Thereby, their possible scenarios (e.g., safer process or equipment) are enumerated without rigorous numeric values. In contrast, the outcome of quantitative approaches is numerically shown in the application of a quantitative risk analysis (QRA). During a QRA, a consequence analysis together with its likelihood is carried out to provide a measure of risk (AIChE, 2000). Furthermore, index-based approaches follow each author's scenarios or concerns through multiple sub-indices. A more detailed comparison is shown in Table 2.6.

Table 2.6 Summary of three different approaches in R-ISATs

	Qualitative-based approaches	**Index-based approaches**	**Quantitative-based approaches**
Basis of development	Brainstorming techniques (e.g., HAZOP or Checklist)	Author's subjective considerations	QRA

	Qualitative-based approaches	Index-based approaches	Quantitative-based approaches
Advantages	Potential to consider all hazardous scenarios	Relatively easy to perform	• Clear approach to select the inherent safety level or safer design among alternatives • More systematic approach
Disadvantages	Essential judgement	Limited scope for the consideration of possible scenarios	Potential to obtain enough information in the concept design stage

Similar to the conventional distinction between consequence and risk assessment tools, R-ISATs are divided into Mode 2.1, referred to as consequence assessment, and Mode 2.2, referred to as risk assessment (see Figure 2.5). Figure 2.11 provides the results obtained from the analysis of the 33 R-ISATs selected. Various index-based approaches have been utilized in Mode 2.1 R-ISATs, while Mode 2.2 is typically associated with quantitative-based approaches. This is due to how Mode 2.2 obtains inherent risk as a numerical value via possible consequence and its likelihood. Furthermore, it is noted that QAISD (Table 2.9 #14) and QI2SD (Table 2.9 #16) were grouped into Mode 2.2 as these tools qualitatively take into account the likelihood based on the ISD guideword—simplicity. Meanwhile, HPS (Table 2.9 #29), which does not include any likelihood aspect, was categorized into Mode 2.1.

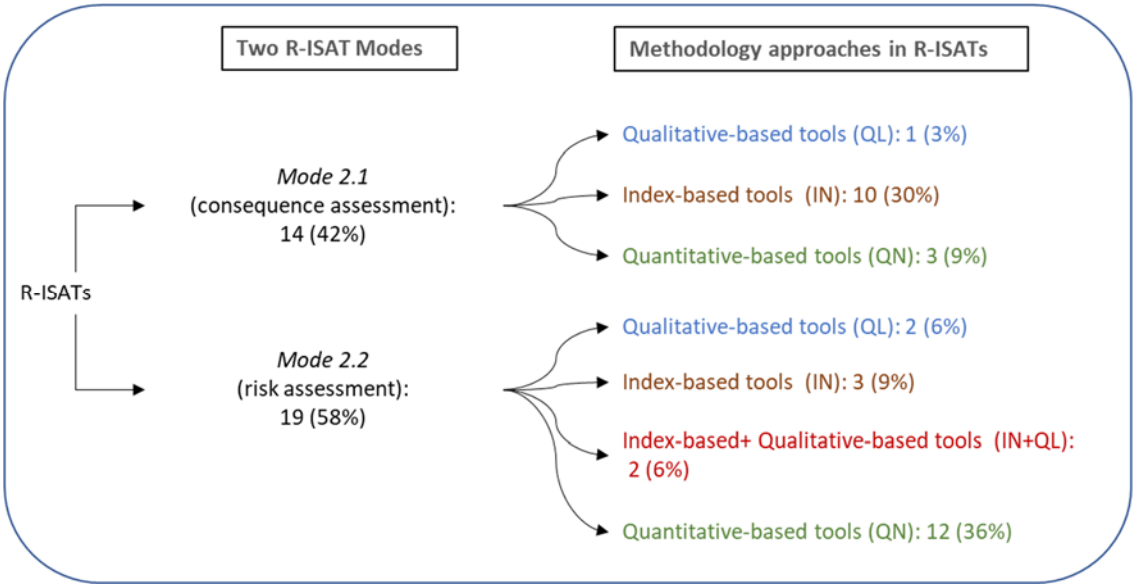
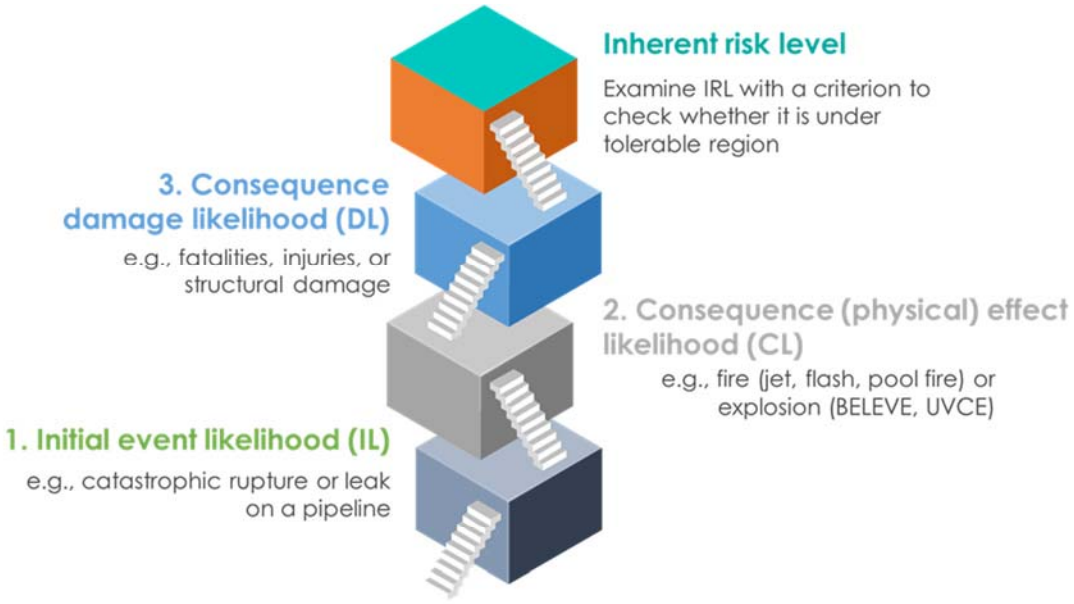


Figure 2.11 The division of R-ISATs into two modes and their approaches

R-ISATs in Mode 2.1 estimates the consequence impact in a selected scenario by considering damage distances, human life damages, and structure damages. Because the likelihood of possible incidents was not considered in Mode 2.1, the potential damage in a worst-cases scenario has been compared among the alternatives. Through this comparison, practitioners can determine which process design is safer.

Mode 2.2 R-ISATs provide the likelihood of possible incidents in detail, except for qualitative approaches. Three types of likelihood have been taken into account in a scenario to yield an inherent risk as shown in Figure 2.12: (1) initial event likelihood, (2) consequence likelihood, and (3) consequence damage likelihood. Fault tree analysis (FTA) is often utilized for the initial events, while event tree analysis (ETA) is employed for the outcome consequences due to the initial events. Finally, profit functions are for consequence damage impacts. Figure 2.12 shows that consequence damage likelihoods

are mostly utilized to estimate inherent risk, while initial event likelihoods are the least used.



Likelihood types	No suggested likelihood	IL	CL	DL
No.	1 (1/15, 7%)	5 (5/15, 33%)	7 (7/15, 47%)	9 (9/15, 60%)

Figure 2.12 Three types of likelihoods considered in the quantitative-based and index-based Mode 2.2 R-ISATs

Table 2.7 shows the breakdown of R-ISATs in terms of their scopes. Compared to health and environment, safety accounts for 67% of all R-ISATs (22/33) and has, therefore, been prioritized. However, in order to improve sustainability in chemical plants, multiple scopes in 18% R-ISATs have been considered simultaneously. All five studies,

including safety, health, and environment, employ index-based approaches to combine different scopes after utilizing each sub-index.

Table 2.7 The number of frequency of scopes considered in 33 R-ISATs

Scope considered	Safety	Health	Environment	Safety + Health	Health + Environment	Safety + Health +Environment
Number (Frequency)	22 (67%)	2 (6%)	3 (9%)	1 (3%)	0 (0%)	5 (15%)

The incident scenarios in R-ISATs can be classified into seven events: fire, explosion, dispersion of toxic chemicals, or the various combinations of them. Table 2.8 illustrates the number and frequency of the used scenarios in R-ISATs. In order to suggest a more realistic scenario, 46% of studies in the literature rigorously represented all three scenarios. As a pioneer, Lawrence (1996) introduced four-factor index without a concrete scenario to emphasize the importance of risk assessment in the early design stage

Table 2.8 The number and frequency of the 28 scenario types considered for safety scope in R-ISATs

Scenario type	No scenario	Fire	Explosion	Toxicity	Fire + Explosion	Fire + Toxicity	Explosion + Toxicity	Fire +Explosion + Toxicity
Number (Frequency)	1 (4%)	2 (7%)	3 (11%)	4 (14%)	4 (14%)	1 (4%)	0 (0%)	13 (46%)

Several researchers have employed optimization tools to suggest an effective ISD solution given multiple aspects to be considered effective. For instance, Patel et al. (2010) illustrated the process for obtaining 11 alternative solvents via the optimization-based tool, mixture integer non-linear program (MINLP), followed by consequence analysis and potential damage distance. Furthermore, Cabezas et al. (1999) optimized possible environmental effects.

Although the layout of a chemical plant is not determined in the conceptual design stage, several R-ISATs utilized ISD to propose sound alternatives. For example, in case a possible incident occurs, reasonable distances of equipment from an engineering building have been illustrated under the given safety criteria (Shariff et al., 2016).

In the early design stages, there is typically a striking lack of knowledge on lay-out, plant environment, and distance to residential areas, perhaps, even no idea where the plant, if built, will be located. In that case, what remains is in principle the consequence part of risk assessment — Mode 2.1 R-ISATs. For that in the first place, data are needed on the nature and quantity of the various inventories the plant will have. The reviewed Mode 2.1 R-ISATs all require knowledge and effort to calculate potential source terms and physical effects, that will impede designers to apply it.

The main advantage of Mode 2.2 is a well-defined acceptable or tolerable criterion that increase the potential for clear decision-making. In case that the inherent risk acceptability levels were not below the defined tolerable zone, it is required to find a better alternative. For example, Shariff and Zaini (2010) and Shariff and Wahab (2013) proposed one alternative that can reduce the potential consequence associated with a column and

storage vessel. Pasha et al., (2017) applied two inherently safer design principles-attenuation and simplification—for alternative designs on heat exchanger networks when the basic design did not meet the criterion. For the acceptable/tolerable criteria, "as Low as reasonably practicable" (ALARP) principle has been utilized in Mode 2.2 R-ISAT.

Even though R-ISATs have been developed to overcome the limitations of H-ISATs, they also have their own notable limitations. First, compared to H-ISATs, R-ISATs require additional effort—in terms of time, and safety knowledge (e.g., consequence modeling, ETA, FTA, or Probit functions). Second, scenarios considered with R-ISATs have inherent uncertainties when compared to conventional risk assessment tools. Conventional tools utilize detailed information as they are employed subsequently to detailed design. On the other hand, R-ISATs are ideally implemented during the conceptual design stage and, as such, can use only limited information available during that stage.

Table 2.9 List of 33 Risk-based Inherent Safety Assessment Tools (R-ISATs) selected

No.	Authors (Year)	R-ISAT		Remark	H-ISAT used	Mode		Scope			Likelihood		
		Name	Type [a]			2.1	2.2	S	H	E	IL	CL	DL
1	Lawrence (1996)	Four-factor index	IN	Update PIIS based on eight experts' opinions			∨	∨					
2	Cave and Edwards (1997)	Environment Hazard Index (**EHI**)	IN	Assessed the impact of hazardous chemicals on aquatic and terrestrial ecosystem		∨				∨			
3	Khan and Abbasi (1998a)	Rapid risk analysis-based design (**RRABD**)	QN	Proposed a tool to aid in decision-making with a scenario-based structure			∨	∨					∨
4	Khan and Abbasi (1998b)	Hazard Identification and Ranking (**HIRA**) Fire and Explosion Damage Index (**FEDI**) Toxicity Damage Index (**TDI**)	IN	Proposed HIRA including two indices: FEDI and TDI		∨		∨					

No.	Authors (Year)	R-ISAT		Remark	H-ISAT used	Mode		Scope			Likelihood		
		Name	Type [a]			2.1	2.2	S	H	E	IL	CL	DL
5	Koller, Fischer, and Hungerbühler (1999)	Assessment of environment-, health-, and safety	IN	Analyze source and starting points of possible problems		√		√	√	√			
6	Cabezas, Bare, and Mallick (1999)	WAste Reduction (**WAR**) algorithm	IN	utilized conservation equation and chemical environmental impact indices		√				√			
7	Gunasekera and Edwards (2003)	Atmospheric Hazard Index (**AHI**)	IN	Utilized fuzzy set for a scaling (weight scheme) to incorporate different 5 categories		√				√			
8	Shah, Fischer, and Hungerbühler (2003);Shah, Fischer, and Hungerbühler (2005)	Substance, Reactivity, Equipment and Safety-Technology (**SREST**) layer assessment tool	IN	Semi-quantitatively likelihood was proposed (e.g., Low, medium, high, moderate, critical high)			√	√	√			√	
9	Shariff et al. (2006)	Integrated Risk Estimation Tool (**iRET**)	QN	selected preliminary control room layout			√	√					√

No.	Authors (Year)	R-ISAT		Remark	H-ISAT used	Mode		Scope			Likelihood		
		Name	Type [a]			2.1	2.2	S	H	E	IL	CL	DL
10	Hassim and Edwards (2006)	Process Route Healthiness Index (**PRHI**)	IN	Quantified health hazards for workers		√			√				
11	Tugnoli, Cozzani, and Landucci (2007)	Key Performance Indicator (**KPI**)	QN	defined the potential accident scenarios associated to each process equipment through the damage distances			√	√			√		
12	Srinivasan and Nhan (2008)	Inherent Benign-ness Indicator (**IBI**)	IN	Quantified hazards, health and environmental impact of a chemical process route based on a statistical analysis		√		√	√	√			
13	Shariff and Leong (2009)	Inherent Risk Assessment (**IRA**)	QN	Integrated risk quantification tool with process design simulator			√	√			√	√	√
14	Hassim and Hurme (2010b)	Health Quotient Index (**HQI**)	IN	Quantified workers health risk from chronic exposure		√			√				
15	Patel, Ng, and Mannan (2010)	Design tool for inherently safer solvent process	QN	Determined inherently safer solvent incorporating an optimization tool		√		√					

No.	Authors (Year)	R-ISAT		Remark	H-ISAT used	Mode		Scope			Likelihood		
		Name	Type [a]			2.1	2.2	S	H	E	IL	CL	DL
16	Shariff and Zaini (2010)	TOxic Release Consequence Analysis Tool (**TORCAT**)	QN	Presented an evolution of iRET			∨	∨					∨
17	Rusli and Shariff (2010)	Qualitative Assessment for Inherently Safer Design (**QAISD**)	QL	Prepared work-flow diagram including possible ISD variables enables practitioners brainstorming			∨	∨					
18	Tugnoli et al. (2012)	Inherent Safety Key Performance Indicators (**IS-KPI**)	QN	As an extension of KPI, external threats were considered.		∨		∨			∨		
19	Rusli, Shariff, and Khan (2013)	Quantitative Index of Inherently Safer Design (**QI2SD**) framework Inherent Risk of Design Index (**IRDI**)	IN+QL	Identified risk based on two sub-indices, damage index (**DI**) and Likelihood index of Hazard Conflicts (**LIHC**) of IRDI			∨	∨					

No.	Authors (Year)	R-ISAT		Remark	H-ISAT used	Mode		Scope			Likelihood		
		Name	Type [a]			2.1	2.2	S	H	E	IL	CL	DL
20	Shariff and Wahab (2013)	Inherent Fire Consequence Estimation Tool (**IFCET**)	QN	demonstrated its potential to minimize the consequences of pool fire			∨	∨					∨
21	Shariff and Zaini (2013)	Toxic Release Inherent Risk Assessment (**TRIRA**)	QN	Demonstrated one of ISD principle, attenuation, after calculating risk level for a selected toxic release scenario			∨	∨			∨		
22	Rocha-Valadez et al. (2014)	Inherently safer sustained casing pressure testing for well integrity	QN	Predicted sustained casing pressure from early pressure buildup		∨		∨					
23	Ordouei, Elkamel, and Al-Sharrah (2014)	Risk index for used in conceptual design	IN	Used incident data for the severity and the likelihood			∨	∨				∨	∨
24	Abidin, Rusli, Shariff, et al. (2016)	Three-stage ISD Matrix (**TIM**)	QL	Identified the trade-off ISD modification toward the risk by combining a matrix and guide word techniques			∨	∨					

No.	Authors (Year)	R-ISAT		Remark	H-ISAT used	Mode		Scope			Likelihood		
		Name	Type [a]			2.1	2.2	S	H	E	IL	CL	DL
25	Abidin, Rusli, Buang, et al. (2016)	A framework including TIM and HIRA	IN+QL	Quantified the impact of ISD modification			∨	∨		∨			
26	Shariff, Wahab, and Rusli (2016)	**IFCET**	QN	Demonstrated its potential to minimize the consequences of BLEVE			∨	∨					∨
27	Jha, Pasha, and Zaini (2016)	Enhanced Inherent Safety Intervention Framework (**Enhanced ISIF**)	QN	Proposed a framework that can apply inherent safety principle by modifying the worst stream, selected via PRI and PSI	• PRI • PSI		∨	∨				∨	∨
28	Warnasooriya and Gunasekera (2017)	Inherent Chemical Process Route Index (**ICPRI**) Environmental Toxicity Hazard Index (**ETHI**) Occupational health Hazard Index (**OhHI**)	IN	Compared routes based on inherent safety, health, and environmental friendliness via weighting scheme	• PIIS	∨		∨	∨	∨			

No.	Authors (Year)	R-ISAT		Remark	H-ISAT used	Mode		Scope			Likelihood		
		Name	Type [a]			2.1	2.2	S	H	E	IL	CL	DL
		Chemical Route Safety Index (**CRSI**)											
29	Pasha, Zaini, and Shariff (2017)	**IRA**	QN	Assessed inherent risk for shell and tube heat exchanger network	• ISISTHE • OSIHEN		∨	∨			∨	∨	∨
30	Abidin et al. (2018)	Inherent Safety Benefits Index (**ISBI**)	IN	As an extension of TIM, quantitively identified the benefits of ISD options		∨		∨	∨	∨			
31	Eljack, Kazi, and Kazantzi (2019)	Inherently Safer Design Tool **(i-SDT)**	QN	Developed a semi-quantitative safety metric that can assess process safety based on historical incident data.			∨	∨				∨	
32	Athar et al. (2019)	Inherently Safer Process Piping (**ISPP**) technique	QN	Predicted the potential damage from a jet fire incident	PSCI		∨	∨				∨	
33	Ahmad et al. (2019)	Hazard Prevention Strategies (**HPS**)	QL	Proposed a qualitative thematic analysis based on the extracted data from accident databases		∨		∨					

S: Safety, H: Health, E: Environment, IL: Initial Likelihood, CL: Consequence Likelihood, DL: Consequence Damage Likelihood

2.5.4. Cost-optimal inherent safety assessment tools (CO-ISATs)

This section analyzes CO-ISATs to understand how cost-optimal inherent safety solutions are obtained.

The selected CO-ISATs were developed for the conceptual design stage in the same manner as previous ISATs in Sections 2.5.2 Hazard-based inherent safety assessment tools (H-ISATs), 2.5.3 Risk-based inherent safety assessment tools (R-ISATs). Table 2.12 chronologically lists the selected 18 CO-ISATs to provide an overview with four primary characteristics: (1) scopes, (2) ISAT used, (3) safety evaluation modes, and (4) decision-making procedures.

Reviewing used safety tools in CO-ISATs reveals whether the selected CO-ISATs were viable in the conceptual design stage. For example, initial CO-ISATs uses SWeHI (Table 2.12 #4) and Mond Index (Table 2.12 #5)—which may not be feasible in the conceptual design stage (see Section 2.3). That is, these CO-ISATs require additional information not available during the conceptual design stage as they were not developed based on H-ISATs or R-ISATs. However, these initial attempts at developing CO-ISATs (Khan and Amyotte, 2005; Tugnoli et al., 2008) influenced subsequent CO-ISATs. In contrast, more recent CO-ISATs have been developed based on H-ISATs (e.g., PRI and PSI) and R-ISATs (e.g., FEDI) to address these short-comings.

Because the simultaneous evaluation of safety and economy is fundamental in CO-ISATs, it is essential to comprehend the basis of each evaluation. In terms of safety evaluation, this study divides CO-ISATs into three modes depending on the approach to account for safety constraints (see Figure 2.5) — Mode 1, Mode 2.1, and Mode 2.2. In

Mode 1, safety is taken into account through inherent hazard levels by utilizing one H-ISAT (e.g., PSI or PRI in Table 2.5). In contrast, Mode 2.1 and Mode 2.2 consider safety based on the potential damages and inherent risk levels, respectively. Practitioners can utilize one of the three modes depending on the relevancy of each mode's characteristics. Table 2.10 summarizes the advantages and disadvantages of each mode of CO-ISATs. Furthermore, in terms of economy evaluation, Figure 2.13 represents 19 different types of costs considered. It is worth noting that approximately 33% of researchers accounted for incident costs (e.g., fatality, injury, environment damages, etc.) along with conventional cost evaluations (e.g., ROI, TAC, operating cost, etc.) as the conventional approach is not considered sufficient to estimate ISD effects.

Table 2.10 Summary of different safety evaluations in CO-ISATs

	Mode 1	**Mode 2.1**	**Mode 2.2**
Advantages	• Easy to perform • Straight-forward step and results	• Expected incident impacts • Potential to calculate incident impacts to monetary values directly.	• Expected incident impacts • Potential to calculate incident impacts to monetary values directly • Results can be associated with safety policy (e.g., ALARP)
Disadvantages	• Potential to fail possible incident scenarios	• Required further knowledge (e.g., consequence modeling) • Complicated calculation steps • Reliable incident database for incident scenarios	

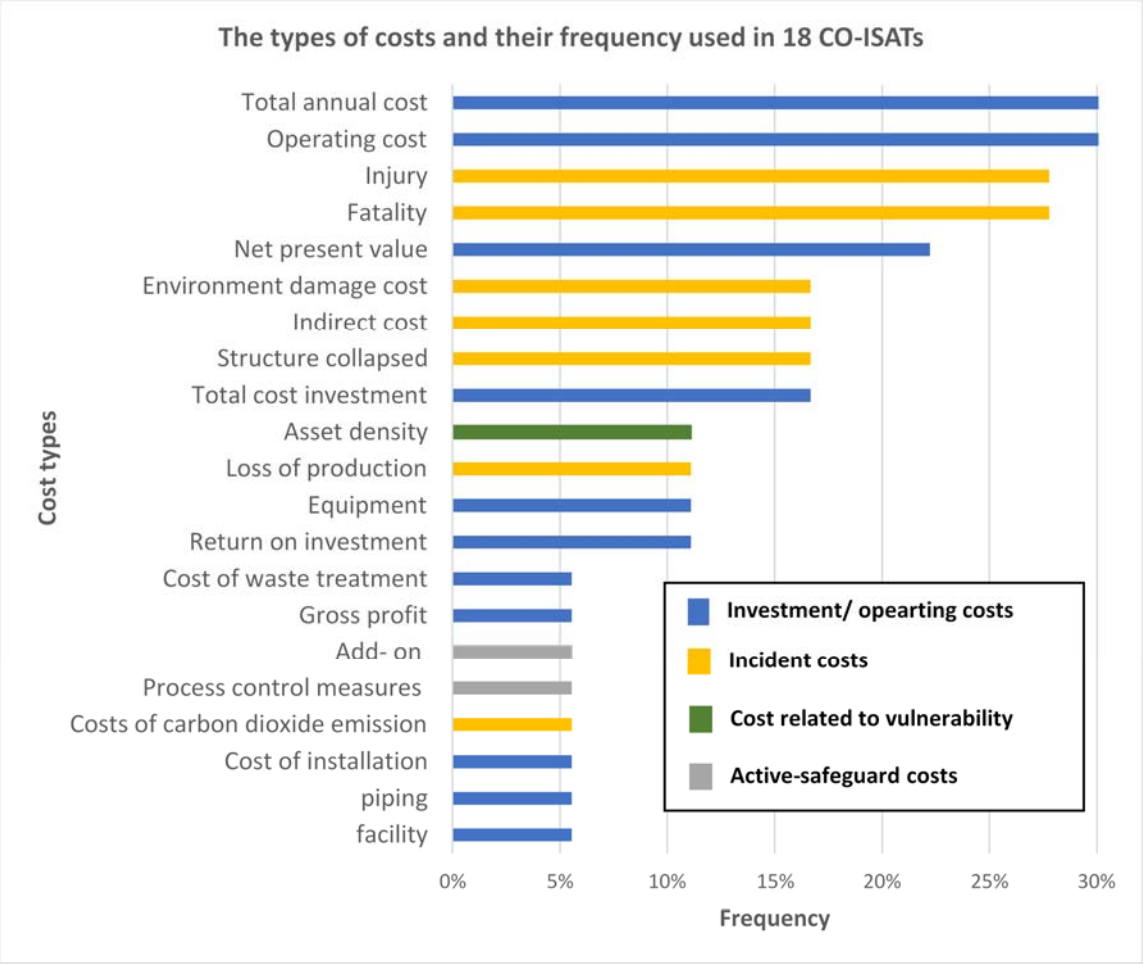


Figure 2.13 The 19 different types of cost considered in the 18 CO-ISATs

Initial attempts required additional information that are unlikely to be available in the conceptual design stage. Such information includes vulnerability factors of a specific area (Khan et al., 2004), or the cost of an add-on system (Khan and Amyotte, 2004). Indirect costs expected the losses arising from business interruptions, insurance, and penalties from the authorities, etc. The 12% (2/17) of CO-ISATs accounts for economic

aspect along with vulnerability, which is capable of being damaged in the surrounding area of the process plant.

Table 2.11 presents the scope considered in CO-ISATs between safety, health, and environment. Similar to R-ISATs, the most considered scope was safety at 44% (8/18) of CO-ISATs, followed by the five CO-ISATs in the combination of safety and environment. Lastly, four CO–ISATs were able to simultaneously consider the 3 scopes.

Table 2.11 The number and frequency of scopes considered in 18 CO-ISATs

Scope considered	S	H	E	S+H	S+E	H+E	S+ H+E
Number (Frequency)	8 (44%)	0 (0%)	1 (6%)	0 (0%)	5 (28%)	0 (0%)	4 (22%)

Note- S: Safety; H: Health; E: Environment

For CO-ISATs, the decision-making tools have been adopted to demonstrate cost factors along with safety. The used decision-making procedures in the 15 CO-ISATs are divided into the five main categories (Table 2.12): (1) multi-criteria decision-making procedure (MCD), (2) multi-objective formulation procedure (MOF), (3) sensitivity analysis, (4) index method, and (5) graphical method.

MCD by itself is *not* able to establish various solutions (or alternatives) in a single run. However, by utilizing a weighting scheme that accounts for the importance of individual attributes, MCD is capable of changing a multi-objective problem into a single-objective problem. For example, Analytical hierarchy process (AHP), which is a type of

MCD, finds an optimal solution by utilizing weight factors (Khan, Sadiq, and Veitch 2004; Khan, Sadiq, and Husain 2002).

Meanwhile, MOF finds multiple solutions in a single run. For example, Pareto front, which is a type of MOFs, presents an array of optimal solutions for multiple problems simultaneously. The optimal solution point is then selected using specific techniques such as Shannon's Entropy and Bellman-Zadeh's techniques (Eini et al. 2018).

Khan and Abbasi (1997) firstly proposed an index approach to evaluate inherent safety along with a cost consideration. The net inherent safety of a selected process is calculated with a formulation that combines values obtained from multiple sub-indices. Of the multiple sub-indices, an asset density index, which is associated with vulnerability of possible incidents, considers economic aspects.

A graphical approach was adopted to visually compare each alternative (e.g., energy, recycle ratio) (Thiruvenkataswamy et al. 2016; Ortiz-Espinoza et al. 2017).

Sensitivity analysis estimates how different values of an independent variable affect a dependent variable. This procedure was utilized with a standalone, MCD, MOF, or a graphical approach among the 15 CO-ISATs.

Around 60% (5/16) of CO-ISATs employ computer-aided software to use their tools effectively. Since Medina-Herrera et al. (2014) used Computer-aided Molecular Design (CAMD) and MATLAB, all subsequent researchers began utilizing computer-aided software including process simulators (e.g., Aspen HYSYS and Plus).

For reference, even though the optimization of storage vessel layouts is beyond the scope of this study, several researchers examined its application during the early design stages (de Lira-Flores, Gutiérrez-Antonio, and Vázquez-Román 2018; Bernechea and Viger 2013; Bernechea and Arnaldos 2014).

Table 2.12 shows that CO-ISATs may adopt more enhanced H-ISATs or R-ISATs. Some safety tools (e.g., SWeHI or Mond index) used in CO-ISATs may not be feasible in the conceptual design stage, as mentioned in Section 2.3. Some of the early ISATs (e.g., ISI proposed in 1999) were utilized in some cases. For better decision-making, researchers can reconcile a more plausible H-ISATs or R-ISATs when they perform CO-ISATs.

Table 2.12 List of 18 Cost-Optimal Inherent Safety Assessment Tools (CO-ISATs) selected

No.	Authors (Year)	Remark	ISAT (or safety tool) used	Scope			Modes used in CO-ISAT			Used decision-making procedure				
				S	H	E	Mode 1	Mode 2.1	Mode 2.2	MCD	MOF	IN	GA	SA
1	Khan and Abbasi (1997)	• New tool named as Accident Hazard Index (**AHI**)		∨		∨		∨				∨		
2	Young, Scharp, and Cabezas (2000)	• Weighting scheme	WAR			∨		∨		∨		∨		
3	Khan, Sadiq, and Husain (2002)	• New tool named as **GreenPro-I** • A holistic and integrated methodology for a cleaner and greener process		∨	∨	∨		∨		∨				∨
4	Khan, Sadiq, and Veitch (2004)	• New tool named as Life cycle indexing system (**LInX**)		∨	∨	∨		∨		∨				
5	Khan and Amyotte (2004);	• New tool named as Integrated Inherent Safety Index (**I2SI**)	SWeHI	∨		∨		∨		∨				

No.	Authors (Year)	Remark	ISAT (or safety tool) used	Scope			Modes used in CO-ISAT			Used decision-making procedure				
				S	H	E	Mode 1	Mode 2.1	Mode 2.2	MCD	MOF	IN	GA	SA
	(Khan and Amyotte 2005)	• Similar to the HAZOP study procedure												
6	Tugnoli, Santarelli, and Cozzani (2008)	• A specific quantitative assessment procedure with normalized indicators	Mond index	√	√	√			√	√				
7	Medina, Arnaldos, and Casal (2009)	• Calculated total cost based on the cost of equipment plus the cost of incidents		√					√		√			√
8	Medina-Herrera et al. (2014)	• Selected safer solvent including economic considerations		√				√			√			√
9	Medina-Herrera, Jiménez-Gutiérrez, and Mannan (2014)	• Selected safer distillation systems		√					√					√

No.	Authors (Year)	Remark	ISAT (or safety tool) used	Scope			Modes used in CO-ISAT			Used decision-making procedure				
				S	H	E	Mode 1	Mode 2.1	Mode 2.2	MCD	MOF	IN	GA	SA
10	Eini et al. (2015)	• Optimal decision-making for refrigerant cycles		∨					∨		∨			∨
11	Abidin, Rusli, Shariff, et al. (2016)	• As an extension task of TIM, cost was estimated	TIM	∨					∨				∨	
12	Eini, Shahhosseini, Delgarm, et al. (2016) and Eini, Shahhosseini, Javidi, et al. (2016)	• Considered thermo-economic analysis along with ISD		∨		∨			∨		∨			
13	Thiruvenkataswamy et al. (2016)	• Only considered main streams	FEDI	∨				∨					∨	
14	Ortiz-Espinoza, Jiménez-Gutiérrez,	• A procedure for the inclusion of inherent safety, economic factors,	• PRI • PSI	∨		∨	∨						∨	∨

No.	Authors (Year)	Remark	ISAT (or safety tool) used	Scope			Modes used in CO-ISAT			Used decision-making procedure				
				S	H	E	Mode 1	Mode 2.1	Mode 2.2	MCD	MOF	IN	GA	SA
	and El-Halwagi (2017)	and environment impact												
15	Martinez-Gomez et al. (2017)	• Proposed a multi-objective formulation approach		∨				∨						∨
16	Guillen-Cuevas et al. (2017)	• New tool named as Safety And Sustainability Weighted Return On Investment Metric (**SASWROIM**)	• FEDI • PRI	∨		∨	∨	∨		∨				∨
17	Eini et al. (2018)	• Optimal decision-making for a reactor network system		∨					∨		∨			∨
18	Teh et al. (2019)	• A systematic tool to conduct sustainability assessment via fuzzy optimization	• ISI • HQI • WAR	∨	∨	∨	∨			∨				

2.6. Concluding remarks and suggestions

The primary objective of this chapter was to establish a comprehensive perspective of ISATs that can enable process engineers to utilize the principles of inherent safety during the conceptual design stage. Therefore, the review focused on various ISATs that are potentially viable for including in conceptual design.

To start with, this review highlighted the fundamentals of safety in Section 2.2 and analyzed previous review papers for ISD in Section 2.3. Section 2.2 clarified and distinguished the various design stages. Because of the specific nature of conceptual design and the limited details available at that stage, it was found that most ISATs are not appropriate for immediate adoption by process engineers during conceptual design. Section 2.4 described the approach used in selecting and reviewing the literature. In Section 2.5, this review study explored viable ISATs that were developed for the conceptual design stage. Since safety was prioritized in this study, the selected tools were named ISATs. However, various tools for health, environment, and economics — the other elements of sustainable process design in CPI — also were taken into account in tandem with safety. Reorganizing the ISATs was imperative for deciding purpose and categorizing them accordingly. In this study, 73 ISATs were selected, and these tools were categorized into three groups: hazard-based inherent safety assessment tools (H-ISATs) for 22 tools, risk-based inherent safety assessment tools (R-ISATs) for 33 tools, and cost-optimal inherent safety assessment tools (CO-ISATs) for 18 tools. The goal of any of these ISATs is to discover a safer design and eventually, a more sustainable process design.

H-ISATs utilize multiple indicators available in the process design scheme, while

R-ISATs estimate potential damages or inherent risk. Furthermore, CO-ISATs account for multiple objectives simultaneously — safety and economics with health and/or environment as option. This chapter illustrated the necessary measures and goals for each ISAT (presented in Figure 2.5) and analyzed the detailed information provided in Section 2.5. The overall view of 73 ISATs would help practitioners determine the tool that would be optimal for them.

This review attempted to offer a comprehensive overview of ISATs. The review also identified that numerous inherent safety tools have been developed with the assumption that sufficient process details are available. This is not the case for the conceptual design stage during which many of the key design decision are still being made. For ISATs to be effectively and efficiently utilized by process engineers, safety researchers and engineers should endeavor to develop metrics and tools that are consistent with the nature of conceptual design and utilizable by process engineers. On the other hand, process engineers should endeavor to apply ISATs and to collaborate with safety experts on defining the needs and specifications of such ISATs and the quantification of safety features of Intensification, Substitution, Attenuation, and Simplification using appropriate level of details available during the conceptual design stage.

In addition to the aforementioned observations and recommendations, there are also needs for more studies in the following areas:

1. **ISATs for the offshore industry**: Of the 73 ISTATs, only Rocha-Valadez et al. (2014) (Table 2.9 #22) suggested an ISAT that could be used in the offshore

oil and gas industry. Therefore, further development for the specific case of the industry is required.

2. **ISATs utilizing a dynamic simulator to consider abnormal condition**: Up until now, all simulators used in ISATs have only been for steady-state conditions. However, studying abnormal conditions with a dynamic simulator would enable process designers to reduce the likelihood of potential adverse events, as many incidents occur during abnormal process conditions.
3. **ISATs adopting the integral process design step, process synthesis**: In general, process designers handle process design via process synthesis — whose aim is "to optimize the logical structure of chemical process, specifically the sequence of step, the choice of chemical employed and the source and destination of recycle stream" (Johns 2001; El-Halwagi 2017). In other words, rather than compromising feedstock and products, this approach seeks an optimal alternative from the given resources.
4. **ISATs for NaTech**: The occurrence of natural disasters have increased over the last few decades. Due to the intrinsic unpredictable nature of natural threats in the CPIs, resulting adverse incidents from natural threats are typically difficult to control and will add to the likelihood of an event. Therefore, ISD can proactively be applied for these cases.
5. **ISATs for supply chains**: There was a lack of consideration of ISD in the context of supply chains. Since numerous incidents have occurred in process supply chains, ISD would aid in creating an improved sustainable process

system. For example, the safety level and cost benefits can vary depending on feedstocks, transported chemicals, or chemical phases.

2.7. Appendix 2A Summary of 8 indices using a scoring system in H-ISATs

Division	Indicator	Symbol	PIIS (1993)	ISI (1999)	i-Safe (2002)	ISIM (2008)
Chemical indicators	Flammability	I_{CF}	0–4	0–4	0–4	–
	Explosiveness	I_{CE}	1–10	0–4	0–4	0–4
	Density	I_{CD}	–	–	–	–
	Lower heat of combustion	I_{CH}	–	–	–	–
	Toxicity	I_{CT}	0–8	0–6	0–6	–
	Chemical interaction	I_{CI}	–	0–4	0–4	–
	Corrosiveness	I_{CC}	–	0–2	0–2	–
	Material phase	I_{CM}	–	–	–	–
	Reactivity	I_{CRE}	–	–	0–4	–
	Heat of reaction	I_{CR}	–	I_{CR1}: *main reaction* (0–4) I_{CR2}: *side reaction* (0–4)	I_{CR1}: *main reaction* (0–4) I_{CR2}: *side reaction* (0–4)	–
	Viscosity	I_{CV}	–	–	–	–
	Volatility	I_{CO}	–	–	–	–
	Type of reaction	I_{CK}	–	–	–	–
	Total chemical indicator	I_C	$I_C = (I_{CF} + I_{CE} + I_{CT})_{max}$ where I_C is the value of the chemical that has the highest total chemical score in a process step.	$I_C = (I_{CF} + I_{CE} + I_{CT})_{max}$ where I_{CF}, I_{CE}, and I_{CT} are the values of the chemical that has the highest total chemical score in a process step.	$I_C = \mathbf{max}\,(I_{CF} + I_{CE} + I_{CT} + I_{CI} + I_{CC} + I_{CRE} + I_{CR1} + I_{CR2})$ where individual indicators are the maximum values for one step in a process route for the worst-case.	–
Process indicators	Temperature	I_{PT}	0–10	0–4	0–4	0–4
	Pressure	I_{PP}	1–10	0–4	0–4	0–4
	Inventory	I_{PI}	1–10	0–5	0–5	–
	Type of equipment	I_{PE}	–	–	–	–
	ISBL	I_{PE1}	–	0–4	0–4	–
	OSBL	I_{PE2}	–	0–3	0–3	–
	Yield	I_{PY}	1–10	–	–	–
	Process modes	I_{PM}	–	–	–	–
	Structure	I_{PS}	–	0–5	–	–
	Total process indicator	I_P	$I_P = max(I_{PT} + I_{PP} + I_{PI} + I_{PY})$ where I_P is the sum of maximum individual indicators from the same process steps as I_c.	$I_P = max(I_{PT} + I_{PP} + I_{PI} + I_{PE} + I_{PS})$ where I_P is the sum of maximum individual indicators from the same process step as I_c.	$I_P = max(I_{PT} + I_{PP} + I_{PE1} + I_{PY})$ where I_P is the sum of maximum individual indicators from the same process step as I_c.	–
Total hazard score		I_{total}	$I_{Total} = \sum_{i}^{n}(I_C + I_P)$ where n is the number of process steps.	$I_{Total} = \sum_{i}^{n}(I_C + I_P)$ where n is the number of process steps.	$I_{Total} = \sum_{i}^{n}(I_C + I_P)$ where n is the number of process steps.	–

Division	Indicator	Symbol	IOHI (2010)	EISI (2011)	CISI (2013)
Chemical indicators	Flammability	I_{CF}	–	0–4	0–4
	Explosiveness	I_{CE}	–	0–4	0–4
	Density	I_{CD}	–	–	–
	Lower heat of combustion	I_{CH}	–	–	–
	Toxicity	I_{CT}	I_{CT1} based on OEL0–4 / I_{CT2} R-phrase: 0–5	0–6	0–6
	Chemical interaction	I_{CI}	–	0–4	0–4
	Corrosiveness	I_{CC}	0–2	0–2	0–2
	Material phase	I_{CM}	1–3	–	–
	Reactivity	I_{CRE}	–	–	–
	Heat of reaction	I_{CR}	–	–	–
	Viscosity	I_{CV}	–	–	–
	Volatility	I_{CO}	0–3	–	–
	Type of reaction	I_{CK}	–	–	–
	Total chemical indicator	$\boldsymbol{I_C}$	$\boldsymbol{I_C = \sum_i^n max(I_{CT1} + I_{CT2} + I_{CC} + I_{CM} + I_{CO})}$ where n is the number of steps	$\boldsymbol{I_C = \sum_i^n (I_{CFi} + I_{CEi} + I_{CTi} + I_{CIi} + I_{CCi}) \times (flow\ rate)_i}$ where n is the number of chemicals in a process route.	$\boldsymbol{I_C = \sum_i^n [(I_{CFi} + I_{CEi} + I_{CTi} + I_{CCi}) \times (flow\ rate)_i] + I_{CI} \times (total\ flow\ rate)}$ where n is the number of chemicals in a process route.
Process indicators	Temperature	I_{PT}	0–3	0–4	0–4
	Pressure	I_{PP}	0–3	0–4	0–4
	Inventory	I_{PI}	–	0–5	0–5
	Type of equipment	I_{PE}	–	–	–
	ISBL	I_{PE1}	–	0–4	0–4
	OSBL	I_{PE2}	–	0–3	–
	Yield	I_{PY}	–	–	–
	Process modes	I_{PM}	1–3	–	–
	Structure	I_{PS}	–	0–5	–
	Total process indicator	$\boldsymbol{I_P}$	$\boldsymbol{I_P = \sum_i^n (I_{PT} + I_{PP} + I_{PM})}$ where n is the number of steps	$\boldsymbol{I_P = \sum_i^n (I_{PTi} + I_{PPi} + I_{PIi} + I_{PE1i} + I_{PE2i} + I_{PSi}) \times No.of\ i}$ where n is the number of equipment types in a process route.	$\boldsymbol{I_P = \sum_i^n (I_{PTi} + I_{PPi} + I_{PIi} + I_{PEi}) \times connection\ score}$ where n is the number of equipment in a process route.
Total hazard score		$\boldsymbol{I_{total}}$	$\boldsymbol{I_{Total} = I_c + I_P}$	$\boldsymbol{I_{Total} = I_c + I_P}$	$\boldsymbol{I_{Total} = I_c + I_P}$

Division	Indicator	Symbol	ISPI (2018)
Chemical indicators	Flammability	I_{CF}	1–5
	Explosiveness	I_{CE}	1–10
	Density	I_{CD}	–
	Lower heat of combustion	I_{CH}	–
	Toxicity	I_{CT}	1–9
	Chemical interaction	I_{CI}	–
	Corrosiveness	I_{CC}	–
	Material phase	I_{CM}	–
	Reactivity	I_{CRE}	–
	Heat of reaction	I_{CR}	1–5
	Viscosity	I_{CV}	–
	Volatility	I_{CO}	–
	Type of reaction	I_{CK}	0.2–1
	Total chemical indicator	$\boldsymbol{I_C}$	$\boldsymbol{I_C = max(I_{CF} + I_{CE} + I_{CT} + I_{CR} \times I_{CK})}$ where I_{CF}, I_{CE}, and I_{CT} are the values of the chemical that has the highest total chemical score in a process step.
Process indicators	Temperature	I_{PT}	1–5
	Pressure	I_{PP}	1–5
	Inventory	I_{PI}	1–6
	Type of equipment	I_{PE}	–
	ISBL	I_{PE1}	0.1–1
	OSBL	I_{PE2}	–
	Process modes	I_{PM}	–
	Yield	I_{PY}	–
	Structure	I_{PS}	–
	Total process indicator	$\boldsymbol{I_P}$	$\boldsymbol{I_P = max\,(I_{PT} + I_{PP} + I_{PI}) \times I_{PE}}$ where I_P is the sum of maximum individual indicators from the same process step as I_c.
Total hazard score		$\boldsymbol{I_{total}}$	$\boldsymbol{I_{Total} = \sum_{i}^{n}(I_C + I_P) \times I_{S,max}}$ where n is the number of streams or equipment, and$\boldsymbol{I_S}$ is the supplement indicators of safeness, complexity, and operability.

2.8. Appendix 2B Summary of 10 indices using a non-scoring system in H-ISATs

Division	Indicator	Symbol	PRI (2009)	PSI (2012)	TRRI (2014)	TRSI (2014)
Chemical indicators	Flammability	I_{CF}	–	–	–	–
	Explosiveness	I_{CE}	$Average\ (\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%$	$\frac{(\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%\ of\ individual\ stream}{(\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%\ of\ all\ streams}$	–	–
	Density	I_{CD}	$Average\ fluid\ density\ [kg/m^3]$	$\frac{Density\ [kg/m^3]\ of\ individual\ stream}{Average\ density\ \ [kg/m^3]\ of\ all\ streams}$	$Average\ fluid\ density\ [kg/m^3]$	$\frac{Density\ value\ [kg/m^3]\ of\ individual\ stream}{Density\ value\ [kg/m^3]\ of\ all\ streams}$
	Lower heat of combustion	I_{CH}	$Average\ heating\ value\ [kJ/kg]$	$\frac{Heating\ value\ [kJ/kg]\ of\ individual\ stream}{Average\ heating\ \ [kJ/kg]\ of\ all\ streams}$	–	–
	Toxicity	I_{CT}	–	–	$Average\ toxicity\ level\ [ppm]$	$\frac{Toxicity\ level\ value\ [ppm]\ of\ individual\ stream}{Average\ toxicity\ level[ppm]\ of\ all\ streams}$
	Chemical interaction	I_{CI}	–	–	–	–
	Corrosiveness	I_{CC}	–	–	–	–
	Material phase	I_{CM}	–	–	–	–
	Reactivity	I_{CRE}	–	–	–	–
	Heat of reaction	I_{CR}	–	–	–	–
	Viscosity	I_{CV}	–	–	–	–
	Volatility	I_{CO}	–	–	–	–
	Type of reaction	I_{CK}	–	–	–	–
Process indicators	Temperature	I_{PT}	Used for the calculation of UFL_{mix} and LFL_{mix}	Used for the calculation of UFL_{mix} and LFL_{mix}	–	–
	Pressure	I_{PP}	$Average\ pressure[bar]$	$\frac{Pressure\ \ [bar]\ of\ individual\ stream}{Average\ pressure\ \ [bar] of\ all\ streams}$	$Average\ pressure[bar]$	$\frac{Pressure\ \ [bar]\ of\ individual\ stream}{Average\ pressure\ \ [bar]\ of\ all\ streams}$
	Inventory	I_{PI}	–	–	–	–
	Type of equipment	I_{PE}	–	–	–	–
	ISBL	I_{PE1}	–	–	–	–
	OSBL	I_{PE2}	–	–	–	–
	Process modes	I_{PM}	–	–	–	–
	Yield	I_{PY}	–	–	–	–
	Structure	I_{PS}	–	–	–	–
Total hazard score		$\boldsymbol{I_{total}}$	$\boldsymbol{I_{Total} = \frac{I_{CE} \times I_{CD} \times I_{CH} \times I_{PP}}{10^8}}$	$\boldsymbol{I_{total} = A_0 \times (I_{CE} \times I_{CD} \times I_{CH} \times I_{PP})}$ where A_0 is a magnifying factor. $\boldsymbol{A_0 = 10}$ was used for the case study.	$\boldsymbol{I_{Total} = A_0 \times (I_{CD} \times I_{CT} \times I_{PP})}$ where A_0 is a magnifying factor	$\boldsymbol{I_{total} = A_0 \times (I_{CD} \times I_{CH} \times I_{PP})}$ where A_0 is a magnifying factor.

Division	Indicator	Symbol	TRSI (2014)	NuDIST (2014;2016)	ISAPED (2017)
Chemical indicators	Flammability	I_{CF}	–	$100 \times \left(1 - \left(\frac{1}{1+3.03e^{-0.02x}}\right)\right)$ where, x *is a flash point*	$100 \times \left(1 - \left(\frac{1}{1+3.03e^{-0.02x}}\right)\right)$ where, x is a flash point
	Explosiveness	I_{CE}	–	$100 \times \left(\frac{1}{1+1096.63e^{-0.14x}}\right)$ *where*, x *is* $(\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%$	$100 \times \left(\frac{1}{1+1096.63e^{-0.14x}}\right)$ where x *is* $(\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%$
	Density	I_{CD}	$\frac{Density\ value\ [kg/m^3]\ of\ individual\ stream}{Density\ value\ [kg/m^3]\ of\ all\ streams}$	–	–
	Lower heat of combustion	I_{CH}	–	–	–
	Toxicity	I_{CT}	$\frac{Toxicity\ level\ value\ [ppm]\ of\ individual\ stream}{Average\ toxicity\ level[ppm]\ of all\ streams}$	$100 \times \left(1 - \left(\frac{1}{1+403.4288e^{-0.012x}}\right)\right)$ *wherer is* TLV − STEL (short − term exposure limit)	$100 \times \left(1 - \left(\frac{1}{1+403.4288e^{-0.012x}}\right)\right)$ *wherer is* TLV − STEL (short − term exposure limit)
	Chemical interaction	I_{CI}	–	–	–
	Corrosiveness	I_{CC}	–	–	–
	Material phase	I_{CM}	–	–	–
	Reactivity	I_{CRE}	–	$100 \times \left(\frac{1}{1+207.43e^{-2.8x}}\right)$ where x is the NFPA reactivity rating	–
	Heat of reaction	I_{CR}	–	$100 \times \left(\frac{1}{1+601.85e^{-0.016x}}\right)$ *where x is the heat of endotermic reaction* $100 \times \left(\frac{1}{1+403.43e^{0.006x}}\right)$ *where x is the heat of exothermic reaction*	–
	Viscosity	I_{CV}	–	–	–
	Volatility	I_{CO}	–	–	–
	Type of reaction	I_{CK}	–	–	–
Process indicators	Temperature	I_{PT}	–	$100 \times \left(\frac{1}{1+403.43e^{-0.012x}}\right)$ *where x is the temperature above* 25℃ $100 \times \left(1 - \left(\frac{1}{1+0.0025e^{-0.012x}}\right)\right)$ where x is the temperature below 25℃	$100 \times \left(\frac{1}{1+403.43e^{-0.012x}}\right)$ *where x is the temperature above* 25℃ $100 \times \left(1 - \left(\frac{1}{1+0.0025e^{-0.012x}}\right)\right)$ where x is the temperature below 25℃
	Pressure	I_{PP}	$\frac{Pressure\ \ [bar]\ of\ individual\ stream}{Average\ pressure\ \ [bar]\ of\ all\ streams}$	$100 \times \left(\frac{1}{1+148.41e^{-0.2x}}\right)$ *where* x is the operating pressure [bar]	$100 \times \left(\frac{1}{1+148.41e^{-0.2x}}\right)$ where x is the operating pressure [bar]
	Inventory	I_{PI}	–	$100 \times \left(1 - \left(\frac{1}{1+1339.43e^{-0.12x}}\right)\right)$ where x is the percentage yield.	–
	Type of equipment	I_{PE}	–	–	–
	ISBL	I_{PE1}	–	–	–
	OSBL	I_{PE2}	–	–	–
	Process modes	I_{PM}	–	–	–
	Yield	I_{PY}	–	–	–
	Structure	I_{PS}	–	–	–
Total hazard score		$\boldsymbol{I_{total}}$	$\boldsymbol{I_{total} = A_0 \times (I_{CD} \times I_{CH} \times I_{PP})}$ where A_0 is a magnifying factor.	$\boldsymbol{I_{Total} = I_{CF} + I_{CE} + I_{CT} + I_{CRE} + I_{CR} + I_{PT} + I_{PP}}$ where each indicator uses the maximum value.	$\boldsymbol{I_{Total} = I_{CF} + I_{CE} + I_{CT} + I_{PT} + I_{PP}}$ where each indicator uses the maximum value.

Division	Indicator	Symbol	ISISTHE (2017)	OSIHEN (2017)	PSCI (2019)
Chemical indicators	Flammability	I_{CF}	–	–	–
	Explosiveness	I_{CE}	$\frac{(\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%\ of\ a\ specific\ STHE}{(\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%\ of\ all\ heat\ exchangers}$	$\frac{(\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%\ of\ a\ specific\ STHE}{(\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%\ of\ all\ heat\ exchangers}$	$\frac{(\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%\ of\ individual\ stream}{(\mathrm{UFL}_{mix} - \mathrm{LFL}_{mix})\%\ of\ all\ streams}$
	Density	I_{CD}	–	–	$\frac{Density\ [kg/m^3]\ of\ individual\ stream}{Average\ density\ [kg/m^3]\ of\ all\ streams}$
	Lower heat of combustion	I_{CH}	$\frac{Heating\ value\ [kJ/kg]\ of\ a\ specific\ STHE}{Average\ heating\ [kJ/kg]\ of\ all\ heat\ exchangers}$	$\frac{Heating\ value\ [kJ/kg]\ of\ a\ specifc\ STHE}{Average\ heating\ [kJ/kg]\ of\ all\ heat\ exchangers}$	$\frac{Heating\ value\ [kJ/kg]\ of\ individual\ stream}{Average\ heating\ [kJ/kg]\ of\ all\ streams}$
	Toxicity	I_{CT}	–	–	–
	Chemical interaction	I_{CI}	–	–	–
	Corrosiveness	I_{CC}	–	–	–
	Material phase	I_{CM}	–	–	–
	Reactivity	I_{CRE}	–	–	–
	Heat of reaction	I_{CR}	–	–	–
	Viscosity	I_{CV}	–	–	$\frac{Viscosity\ of\ individual\ stream}{Average\ viscosity\ of\ all\ streams}$
	Volatility	I_{CO}	–	–	–
	Type of reaction	I_{CK}	–	–	–
Process indicators	Temperature	I_{PT}	$\frac{CMTD\ [C^o]\ of\ a\ specific\ STHE}{Average\ CMTD\ [C^o]of\ all\ heat\ exchangers}$ where CMTD is corrected mean temperature difference	$\frac{CMTD\ [C^o]\ of\ a\ specific\ STHE}{Average\ CMTD\ [C^o]of\ all\ heat\ exchangers}$ where CMTD is corrected mean temperature difference	$\frac{Temperature\ [C^o]\ of\ individual\ stream}{Average\ temperature\ [C^o]of\ all\ streams}$
	Pressure	I_{PP}	$\frac{Pressure\ [kPa]of\ a\ specific\ STHE}{Average\ pressure\ [kPa]of\ all\ heat\ exchangers}$	$\frac{Pressure\ [kPa]of\ a\ specific\ STHE}{Average\ pressure\ [kPa]of\ all\ heat\ exchangers}$	$\frac{Pressure\ [bar]\ of\ individual\ stream}{Average\ pressure\ [bar]of\ all\ streams}$
	Inventory	I_{PI}	–	–	–
	Type of equipment	I_{PE}	–	–	–
	ISBL	I_{PE1}	–	–	–
	OSBL	I_{PE2}	–	–	–
	Process modes	I_{PM}	–	–	–
	Yield	I_{PY}	–	–	–
	Structure	I_{PS}	–	–	–
Total hazard score		$\boldsymbol{I_{total}}$	$\boldsymbol{I_{total} = I_{CE} \times I_{CH} \times I_{CT} \times I_{PP}}$	$\boldsymbol{I_{total} = \sum_{i}^{n} (I_{CE} \times I_{CH} \times I_{CT} \times I_{PP})_i}$ where i is a specific STHE and n is the number of STHE	$\boldsymbol{I_{total} = A_0 \times (I_{CE} \times I_{CD} \times I_{CH} \times I_{CV} \times I_{PT} \times I_{PP})}$ where A_0 is a magnifying factor.

2.9. References of Chapter 2

Abidin, Mardhati Zainal, Rusli, Risza, Buang, Azizul, Mohd Shariff, Azmi, Khan, Faisal Irshad, 2016a. 'Resolving inherent safety conflict using quantitative and qualitative technique. J. Loss Prev. Process. Ind. 44, 95–111.

Abidin, Mardhati Zainal, Rusli, Risza, Khan, Faisal, Mohd Shariff, Azmi, 2018. 'Development of inherent safety benefits index to analyse the impact of inherent safety implementation. Process Saf. Environ. Prot. 117, 454–472.

Abidin, Mardhati Zainal, Rusli, Risza, Mohd Shariff, Azmi, Khan, Faisal Irshad, 2016b. 'Three-stage ISD matrix (TIM) tool to review the impact of inherently safer design implementation. Process Saf. Environ. Prot. 99, 30–42.

Ahmad, Syaza I., Hashim, Haslenda, Hassim, Mimi H., 2014. 'Numerical descriptive inherent safety technique (NuDIST) for inherent safety assessment in petrochemical industry'. Process Saf. Environ. Prot. 92, 379–389.

Ahmad, Syaza I., Hashim, Haslenda, Hassim, Mimi H., 2016a. A graphical method for assessing inherent safety during research and development phase of process design. J. Loss Prev. Process. Ind. 42, 59–69.

Ahmad, S.I., Hashim, H., Hassim, M.H., Muis, Z.A., 2016. 'Inherent safety assessment of biodiesel production: flammability parameter. Procedia Eng. 148, 1177–1183.

Ahmad, Syaza I., Hashim, Haslenda, Hassim, Mimi H., Rashid, Roslina, 2019. 'Development of hazard prevention strategies for inherent safety assessment during early stage of process design. Process Saf. Environ. Prot. 121, 271–280.

Ahmad, Syaza Izyanni, Hashim, H., Hassim, M.H., 2017. Inherent safety assessment technique for preliminary design stage. Chem. Eng. Trans. 56, 1345–1350.

AIChE, C.C.P.S., 2000. 'Guidelines for Chemical Process Quantitative Risk Analysis (New York, NY).

Al-Mutairi, Eid, M., Suardin, Jaffee A., Sam Mannan, M., El-Halwagi, Mahmoud M., 2008. 'An optimization approach to the integration of inherently safer design and process scheduling. J. Loss Prev. Process. Ind. 21, 543–549.

Athar, M., Mohd Shariff, A., Buang, A., 2019. 'A review of inherent assessment for sustainable process design. J. Clean. Prod. 233, 242–263.

Athar, Muhammad, Mohd Shariff, Azmi, Buang, Azizul, Shuaib Shaikh, Muhammad, Tan, Lian See, 2019. 'Inherent safety for sustainable process design of process piping at the preliminary design stage. J. Clean. Prod. 209, 1307–1318.

Bernechea, E.J., Arnaldos, J., 2014. Optimizing the design of storage facilities through the application of ISD and QRA. Process Saf. Environ. Prot. 92, 598–615.

Bernechea, E.J., Viger, J.A., 2013. Design optimization of hazardous substance storage facilities to minimize project risk. Saf. Sci. 51, 49–62.

Busura, Saheed, Khan, Faisal, Kelly, Hawboldt, Iliyas, AbdulJelil, 2014. 'Quantitative risk-based ranking of chemicals considering hazardous thermal reactions. J. Chem. Health Saf. 21, 27–38.

Cabezas, Heriberto, Bare, Jane C., Subir, K Mallick, 1999. 'Pollution prevention with chemical process simulators: the generalized waste reduction (WAR) algorithm—full version. Comput. Chem. Eng. 23, 623–634.

Cameron, Ian, Mannan, Sam, , Erzsébet Németh, Park, Sunhwa, Pasman, Hans, Rogers, William, Benjamin, Seligmann, 2017. 'Process hazard analysis, hazard identification and scenario definition: are the conventional tools sufficient, or should and can we do much better? Process Saf. Environ. Prot. 110, 53–70.

Cave, S.R., Edwards, D.W., 1997. 'Chemical process route selection based on assessment of inherent environmental hazard. Comput. Chem. Eng. 21, S965–S970.

CCPS, 2008. Guidelines for Hazard Evaluation Procedures. John Wiley & Sons, Incorporated.

CCPS, 2010. Inherently Safer Chemical Processes: A Life Cycle Approach. John Wiley & Sons, Incorporated).

Chemical Safety Board (CSB), 2018. Organic peroxide decomposition, release, and fire at Arkema Crosby following Hurricane Harvey flooding.

Cozzani, Valerio, Campedel, Michela, Renni, Elisabetta, Krausmann, Elisabeth, 2010. 'Industrial accidents triggered by flood events: analysis of past accidents. J. Hazard Mater. 175, 501–509.

Crabtree, E.W., El-Halwagi, M.M., 1994. Synthesis of environmentally acceptable reactions. In: AIChE Symposium Series., vol. 90. American Institute of Chemical Engineers, New York, NY. No. 303, 1971-c2002.

de Lira-Flores, J.A., Claudia Gutiérrez-Antonio, and Richart Vázquez-Román, R., 2018. A MILP approach for optimal storage vessels layout based on the quantitative risk analysis methodology. Process Saf. Environ. Prot. 120, 1–13.

DeRosa, Sean E., Kimura, Yosuke, Stadtherr, Mark A., Gary, McGaughey, McDonald-Buller, Elena, Allen, David T., 2019. 'Network modeling of the US petrochemical industry under raw material and Hurricane Harvey disruptions. Ind. Eng. Chem. Res.58, 12801–12815.

Dow, 1994. Dow fire and explosion index hazard classification guide. American Institute of Chemical Engineers, New York.

Eckerman, Ingrid, 2018. 'Bhopal Gas Catastrophe 1984: Causes and Consequences.

Edwards, David W., Lawrence, Duncan, 1993. 'Assessing the inherent safety of chemical process routes: is there a relation between plant costs and inherent safety? Process Saf. Environ. Prot. 71, 252–258.

Eini, Saeed, Abdolhamidzadeh, Bahman, Reniers, Genserik, Rashtchian, Davood, 2015. 'Optimization procedure to select an inherently safer design scheme. Process Saf. Environ. Prot. 93, 89–98.

Eini, Saeed, Javidi, Majid, Shahhosseini, Hamid Reza, Rashtchian, Davood, 2018. 'Inherently safer design of a reactor network system: a case study. J. Loss Prev. Process. Ind. 51, 112–124.

Eini, Saeed, Shahhosseini, Hamidreza, Delgarm, Navid, Lee, Moonyong, Bahadori, Alireza, 2016b. 'Multi-objective optimization of a cascade refrigeration system: exergetic, economic, environmental, and inherent safety analysis. Appl.Therm. Eng. 107, 804–817.

Eini, Saeed, Shahhosseini, Hamid Reza, Javidi, Majid, Sharifzadeh, Mahdi, Rashtchian, Davood, 2016a. 'Inherently safe and economically optimal design using multi-objective optimization: the case of a refrigeration cycle. Process Saf. Environ. Prot. 104, 254–267.

El-Halwagi, M.M., 2017. Sustainable Design through Process Integration: Fundamentals and Applications to Industrial Pollution Prevention, Resource Conservation, and Profitability Enhancement. Butterworth-Heinemann.

Eljack, Fadwa, Kazi, Monzure-Khoda, Kazantzi, Vasiliki, 2019. 'Inherently safer design tool (i-SDT): a property-based risk quantification metric for inherently safer design during the early stage of process synthesis. J. Loss Prev. Process. Ind. 57, 280–290.

Gangadharan, Preeti, Singh, Ravinder, Cheng, Fangqin, Helen, H Lou, 2013. 'Novel methodology for inherent safety assessment in the process design stage. Ind. Eng. Chem. Res. 52, 5921–5933.

Gentile, M., Rogers, W.J., Mannan, M.S., 2003. 'Development of a fuzzy logic-based inherent safety index. Process Saf. Environ. Prot. 81, 444–456.

Gentile, Michela, 2004. 'Development of a Hierarchical Fuzzy Model for the Evaluation of Inherent Safety. Texas A&M University.

Guillen-Cuevas, Karen, Ortiz-Espinoza, Andrea P., Ozinan, Ecem Ozinan, Arturo Jiménez-Gutiérrez, Arturo, Nikolaos, K Kazantzis, El-Halwagi, Mahmoud M., 2017. 'Incorporation of safety and sustainability in conceptual design via a return on investment metric. ACS Sustain. Chem. Eng. 6, 1411–1416.

Gunasekera, M.Y., Edwards, D.W., 2003. Estimating the environmental impact of catastrophic chemical releases to the atmosphere: an index method for ranking alternative chemical process routes. Process Saf. Environ. Prot. 81, 463–474.

Gupta, J.P., Edwards, David W., 2003. 'A simple graphical method for measuring inherent safety. J. Hazard Mater. 104, 15–30.

Hassim, M.H., Edwards, D.W., 2006. 'Development of a methodology for assessing inherent occupational health hazards. Process Saf. Environ. Prot. 84, 378–390.

Hassim, M.H., Hurme, M., 2010a. 'Inherent occupational health assessment during process research and development stage. J. Loss Prev. Process. Ind. 23, 127–138.

Hassim, Mimi H., Hurme, Markku, 2010b. 'Inherent occupational health assessment during preliminary design stage. J. Loss Prev. Process. Ind. 23, 476–482.

Hassim, M.H., Hurme, M., Edwards, D.W., Aziz, N.N., Rahim, F.L., 2013. 'Simple graphical method for inherent occupational health assessment. Process Saf. Environ. Prot. 91, 438–451.

Heikkil€a, Anna-Mari, 1999. Inherent Safety in Process Plant Design: an Index-Based Approach. VTT Technical Research Centre of Finland.

Houston, D.E.L., 1971. 'New Approaches to the Safety Problem', Major Loss Prevention in the Process Industries, p. 210.

ISO, ISO14040, 2006. 'Environmental management — life cycle assessment — principles and framework. Accessed June 5. https://www.iso.org/obp/ui/#iso:std:iso:14040: ed-2:v1:en:term:3.34.

ISO, ISO31000, 2009. Risk Management–Principles and Guidelines. International Organization for Standardization, Geneva, Switzerland.

Jafari, Mohammad Javad, Mohammadi, Heidar, Reniers, Genserik, Pouyakian, Mostafa, Nourai, Farshad, Ali Torabi, Seyed, Miandashti, Masoud Rafiee, 2018. 'Exploring

inherent process safety indicators and approaches for their estimation: a systematic review. J. Loss Prev. Process. Ind. 52, 66–80.

Jha, Varsha, Pasha, Mohsin, Zaini, Dzulkarnain, 2016. Enhanced inherent safety intervention framework. Procedia Eng. 148, 1051–1057.

Johns, W.R., 2001. Process synthesis: poised for a wider role. Chem. Eng. Prog. 97, 59–65.

Julián-Durán, Laura, M., et al., 2014. Techno-economic assessment and environmental impact of shale gas alternatives to methanol. ACS Sustain. Chem. Eng. 2.10, 2338–2344.

Khan, Faisal I., Abbasi, S.A., 1998a. Inherently safer design based on rapid risk analysis. J. Loss Prev. Process. Ind. 11, 361-72.

Khan, Faisal I., Abbasi, S.A., 1998b. 'Multivariate hazard identification and ranking system. Process Saf. Prog. 17, 157–170.

Khan, Faisal I., Amyotte, Paul R., 2004. 'Integrated inherent safety index (I2SI): a tool for inherent safety evaluation. Process Saf. Prog. 23, 136–148.

Khan, Faisal I., Amyotte, Paul R., 2005. 'I2SI: a comprehensive quantitative tool for inherent safety and cost evaluation. J. Loss Prev. Process. Ind. 18, 310–326.

Khan, Faisal I., Husain, T., Abbasi, Sakineh A., 2001. 'Safety weighted hazard index (SWeHI): a new, user-friendly tool for swift yet comprehensive hazard identification

and safety evaluation in chemical process industrie. Process Saf. Environ. Prot. 79, 65–80.

Khan, Faisal I., Sadiq, Rehan, Husain, Tahir, 2002. 'GreenPro-I: a risk-based life cycle assessment and decision-making methodology for process plant design. Environ. Model. Softw 17, 669–692.

Khan, Faisal I., Sadiq, Rehan, Veitch, B., 2004. 'Life cycle iNdeX (LInX): a new indexing procedure for process and product design and decision-making. J. Clean. Prod. 12, 59–76.

Khan, F.I., Abbasi, Sakineh A., 1997. 'Accident hazard index: a multi-attribute method for process industry hazard rating. Process Saf. Environ. Prot. 75, 217–224.

Kletz, Trevor A., 1978. What you don't have, can't leak. Chem. Ind. 6, 287–292.

Kletz, Trevor A., 1985. 'Inherently safer plants. Plant Oper. Prog. 4, 164–167.

Kletz, T.A., Amyotte, P., 2010. Process Plants: A Handbook for Inherently Safer Design. CRC Press.

Koller, Guntram, Fischer, Ulrich, Hungerbühler, Konrad, 1999. 'Assessment of environment-, health-and safety aspects of fine chemical processes during early design phases. Comput. Chem. Eng. 23, S63–S66.

Krausmann, Elisabeth, Cozzani, Valerio, Salzano, Ernesto, Renni, Elisabetta, 2011. Industrial accidents triggered by natural hazards: an emerging risk issue. Nat. Hazards Earth Syst. Sci. 11, 921–929.

Lawrence, Duncan, 1996. Quantifying Inherent Safety of Chemical Process Routes. © Duncan Lawrence.

Lees, Frank, 2012. Lees' Loss Prevention in the Process Industries: Hazard Identification, Assessment and Control. Butterworth-Heinemann.

Leong, C.T., Shariff, A.M., 2008. 'Inherent safety index module (ISIM) to assess inherent safety level during preliminary design stage. Process Saf. Environ. Prot. 86, 113–119.

Leong, C.T., Shariff, A.M., 2009. 'Process route index (PRI) to assess level of explosiveness for inherent safety quantification. J. Loss Prev. Process. Ind. 22, 216–221.

Lewis, D.J., 1979. The Mond fire, explosion, and toxicity index—a development of the dow index. In: Proceedings of the AIChE on Loss Prevention Symposium, New York.

Li, Xiang, Zanwar, Anand, Jayswal, Abhishek, Helen, H Lou, Huang, Yinlun, 2011. Incorporating exergy analysis and inherent safety analysis for sustainability assessment of biofuels. Ind. Eng. Chem. Res. 50, 2981–2993.

Martinez-Gomez, Juan, Nápoles-Rivera, Fabricio, Ponce-Ortega, José María, El-Halwagi, Mahmoud M., 2017. 'Optimization of the production of syngas from shale gas with economic and safety considerations. Appl. Therm. Eng. 110, 678–685.

Medina-Herrera, Nancy, Grossmann, Ignacio E., Sam Mannan, M., Jiménez-Gutiérrez, Arturo, 2014. An approach for solvent selection in extractive distillation systems including safety considerations. Ind. Eng. Chem. Res. 53, 12023–12031.

Medina-Herrera, Nancy, Jiménez-Gutiérrez, Arturo, Sam Mannan, M., 2014.'Development of inherently safer distillation systems. J. Loss Prev. Process. Ind. 29, 225–239.

Medina, Héctor Arnaldos, Josep, Casal, Joaquim, 2009. 'Process design optimization and risk analysis. J. Loss Prev. Process. Ind. 22, 566–573.

National Research Council, 2011. Sustainability and the US EPA. National Academies Press.

NFPA, 1991. Fire hazard properties OF flammable liquids, gases, and volatile solids. Ordouei, Mohammad Hossein, Ali, Elkamel, Al-Sharrah, Ghanima, 2014. New simple indices for risk assessment and hazards reduction at the conceptual design stage of a chemical process. Chem. Eng. Sci. 119, 218–229.

Ortiz-Espinoza, Andrea, P., Jiménez-Gutiérrez, Arturo, El-Halwagi, Mahmoud M., 2017. 'Including inherent safety in the design of chemical processes. Ind. Eng. Chem. Res. 56, 14507–14517.

Palaniappan, Chidambaram, Srinivasan, Rajagopalan, Tan, Reginald, 2002a. Expert system for the design of inherently safer processes. 1. Route selection stage. Ind. Eng. Chem. Res. 41, 6698–6710.

Palaniappan, Chidambaram, Srinivasan, Rajagopalan, Tan, Reginald B., 2002b. Expert system for the design of inherently safer processes. 2. Flowsheet development stage. Ind. Eng. Chem. Res. 41, 6711–6722.

Pasha, M., Zaini, D., Shariff, A.M., 2017. 'Inherently safer design for heat exchanger network. J. Loss Prev. Process. Ind. 48, 55–70.

Pasman, Hans J., 2015. Risk Analysis and Control for Industrial Processes-Gas, Oil and Chemicals: A System Perspective for Assessing and Avoiding Low-Probability, High-Consequence Events. Butterworth-Heinemann.

Patel, Suhani J., Ng, Dedy, Sam Mannan, M., 2010. 'Inherently safer design of solvent processes at the conceptual stage: practical application for substitution. J. Loss Prev. Process. Ind. 23, 483–491.

Qi, Meng, Zhu, Jiayong, Xu, Jiaming, Zhao, Dongfeng, Liu, Yi, 2019. Investigation of Inherently Safer Design through Process Intensification: Novel Safety Assessment Methodology and Case Study in C3-Alkyne Hydrogenation Distillation Process. Industrial & Engineering Chemistry Research.

Rocha-Valadez, Tony, Mentzer, Ray A, Rashid Hasan, A., Sam Mannan, M., 2014. 'Inherently safer sustained casing pressure testing for well integrity evaluation. J. Loss Prev. Process. Ind. 29, 209–215.

Roy, Nitin, Eljack, Fadwa, Jiménez-Gutiérrez, Arturo, Zhang, Bin, Thiruvenkataswamy, Preetha, El-Halwagi, Mahmoud, Sam Mannan, M., 2016. A review of safety indices for process design. Curr. Opin. Chem. Eng. 14, 42–48.

Rusli, R., Shariff, A.M., 2010. 'Qualitative assessment for inherently safer design (QAISD) at preliminary design stage. J. Loss Prev. Process. Ind. 23, 157–165.

Rusli, R., Shariff, A Mohd, Khan, F.I., 2013. Evaluating hazard conflicts using inherently safer design concept. Saf. Sci. 53, 61–72.

Shah, S., Fischer, Ulrich, Hungerbühler, Konrad, 2003. A hierarchical approach for the evaluation of chemical process aspects from the perspective of inherent safety. Process Saf. Environ. Prot. 81, 430–443.

Shah, Shailesh, Fischer, Ulrich, Hungerbühler, Konrad, 2005. 'Assessment of chemical process hazards in early design stages. J. Loss Prev. Process. Ind. 18, 335–352.

Shariff, Azmi Mohd, Leong, Chan T., 2009. 'Inherent risk assessment—a new concept to evaluate risk in preliminary design stage. Process Saf. Environ. Prot. 87, 371–376.

Shariff, Azmi Mohd, Leong, Chan T., Zaini, Dzulkarnain, 2012. 'Using process stream index (PSI) to assess inherent safety level during preliminary design stage. Saf. Sci. 50, 1098–1103.

Shariff, Azmi Mohd, Rusli, Risza, Leong, Chan T., Radhakrishnan, V.R., Buang, Azizul, 2006. 'Inherent safety tool for explosion consequences study. J. Loss Prev. Process. Ind. 19, 409–418.

Shariff, Azmi Mohd, Abdul Wahab, Nordiana, Rusli, Risza, 2016. 'Assessing the hazards from a BLEVE and minimizing its impacts using the inherent safety concept. J. Loss Prev. Process. Ind. 41, 303–314.

Shariff, A.M., Wahab, N.A., 2013. 'Inherent fire consequence estimation tool (IFCET) for preliminary design of process plant. Fire Saf. J. 59, 47–54.

Shariff, Azmi Mohd, Zaini, Dzulkarnain, 2010. 'Toxic release consequence analysis tool (TORCAT) for inherently safer design plant. J. Hazard Mater. 182, 394–402.

Shariff, A.M., Zaini, D., 2013. 'Inherent risk assessment methodology in preliminary design stage: a case study for toxic release. J. Loss Prev. Process. Ind. 26, 605–613.

Singh, Binder, Paul, Jukes, Ben Poblete, Wittkower, Bob, 2010. '20 Years on lessons learned from Piper Alpha. The evolution of concurrent and inherently safe design. J. Loss Prev. Process. Ind. 23, 936–953.

Song, Daesung, Yoon, En Sup, Jang, Namjin, 2018. 'A framework and method for the assessment of inherent safety to enhance sustainability in conceptual chemical process design. J. Loss Prev. Process. Ind. 54, 10–17.

Srinivasan, Rajagopalan, Natarajan, Sathish, 2012. 'Developments in inherent safety: a review of the progress during 2001–2011 and opportunities ahead. Process Saf. Environ. Prot. 90, 389–403.

Srinivasan, R., Nhan, N.T., 2008. 'A statistical approach for evaluating inherent benignness of chemical process routes in early design stages. Process Saf. Environ. Prot. 86, 163–174.

Suardin, J., Mannan, M.S., El-Halwagi, M., 2007. The integration of Dow's fire and explosion index (F&EI) into process design and optimization to achieve inherently safer design. J. Loss Prev. Process. Ind. 20 (1), 79–90.

Tang, Kuok Ho Daniel, Md Dawal, Siti Zawiah, Udoncy Olugu, Ezutah, 2018. A review of the offshore oil and gas safety indices. Saf. Sci. 109, 344–352.

Teh, Shin Yee, Chua, Kian Boon, Hooi Hong, Boon, Ling, Alex JW., Andiappan, Viknesh, Dominic, CY Foo, Hassim, Mimi H., Denny, KS Ng, 2019. A hybrid multi-objective optimization framework for preliminary process design based on health. Saf. Environ. Impact', Process. 7, 200.

Thiruvenkataswamy, Preetha, Eljack, Fadwa T., Roy, Nitin, Sam Mannan, M., El-Halwagi, Mahmoud M., 2016. 'Safety and techno-economic analysis of ethylene

technologies. J. Loss Prev. Process. Ind. 39, 74–84.

Tian, Yuhe, Demirel, Salih Emre, Hasan, MM Faruque, Pistikopoulos, Efstratios N., 2018. An Overview of Process Systems Engineering Approaches for Process Intensification: State of the Art', Chemical Engineering and Processing-Process Intensification.

Towler, G., Sinnott, R., 2012. Chemical Engineering Design: Principles, Practice and Economics of Plant and Process Design. Elsevier.

Tugnoli, Alessandro, Cozzani, Valerio, Landucci, Gabriele, 2007. 'A consequence-based approach to the quantitative assessment of inherent safety. AIChE J. 53, 3171–3182.

Tugnoli, Alessandro, Landucci, Gabriele, Salzano, Ernesto, Cozzani, Valerio, 2012. 'Supporting the selection of process and plant design options by Inherent Safety KPIs. J. Loss Prev. Process. Ind. 25, 830–842.

Tugnoli, Alessandro, Santarelli, Francesco, Cozzani, Valerio, 2008. An approach to quantitative sustainability assessment in the early stages of process design. Environ.Sci. Technol. 42, 4555–4562.

Tyler, B.J., 1985. 'Using the Mond Index to measure inherent hazards. Plant Oper. Prog. 4, 172–175.

Vázquez, Daniel, Quirante, Natalia, Ruiz-Femenia, Rubén, Fernández, María J., Salcedo-Díaz, Raquel, Gómez-Rico, M Francisca, Caballero, José A., 2017. 'Systematic

methods for inherently safer process design: comparison among inherent safety indexes by dimensionality reduction. In: Computer Aided Chemical Engineering. Elsevier.

Warnasooriya, Sureshinie, Gunasekera, Manisha Yasanthi, 2017. Assessing inherent environmental, health and safety hazards in chemical process route selection. Process Saf. Environ. Prot. 105, 224–236.

Young, Douglas, Scharp, Richard, Cabezas, Heriberto, 2000. The waste reduction (WAR) algorithm: environmental impacts, energy consumption, and engineering economics. Waste Manag. 20, 605–615.

Zaini, Dzulkarnain, Azmi, Mohd Shariff, Leong, Chan T., 2014. Three-tier inherent safety quantification (3-TISQ) for toxic release at preliminary design stage. In: Applied Mechanics and Materials. Trans Tech Publ, pp. 426–430.

3. WHAT CAN THE TROVE OF THE INCIDENT INVESTIGATIONS TEACH US? A DETAILED ANALYSIS OF INFORMATION CHARACTERISTICS AMONG CHEMICAL PROCESS INCIDENTS INVESTIGATED BY THE CSB[3]

Understanding the commonalities among previous chemical process incidents can help mitigate recurring incidents in the chemical process industry and will be useful background knowledge for process designers intending to foster inherent safety. The U.S. Chemical Safety and Hazard Investigation Board (CSB) reports provide detailed and vital incident information that can be used to identify possible commonalities. This chapter aims to develop a systematic approach for extracting data from the CSB reports with the objective of establishing these commonalities. Data were extracted based on three categories: attributed incident causes, scenarios, and consequences. Seventeen causal factors were classified as chemical indicators or process indicators. Twelve chemical indicators are associated with the hazards of the chemicals involved in the incidents, whereas five process indicators account for the hazards presented by process conditions at the time of the incident. Seven scenario factors represent incident sequences, equipment types, operating modes, process units, domino effects, detonation likelihood for explosion incidents, and population densities. Finally, three consequence factors were selected based on types of chemical incidents, casualties, population densities, and economic losses. Data

[3] Reprinted with permission from "What can the trove of CSB incident investigations teach us? A detailed analysis of information characteristics among chemical process incidents investigated by the CSB" by Park, S., Mendez, E., Bailey, J. P., Rogers, W., Pasman, H. J., & El-Halwagi, M. M. 2021, Journal of Loss Prevention in the Process Industries, 69, 104389, Copyright [2021] by Elsevier Ltd. All rights reserved.

from 87 CSB reports covering 94 incidents were extracted and analyzed according to the proposed approach. Based on these findings, this chapter proposes guidelines for future collection of information to provide valuable resources for prediction and risk reduction of future incidents.

3.1. Introduction

Past chemical process incidents can hold insights into systematic failure of certain safety policies, methodological approaches, or hardware in the process industry. Extracting the commonalities in previous incidents may enable us to proactively reduce the likelihood of similar incidents in the process industry from occurring. The U.S. Chemical Safety and Hazard Investigation Board (CSB) investigation reports are a noteworthy resource due to their detailed incident information. Established by the Clean Air Act Amendments in 1990, the CSB is responsible for determining the root and contributing causes of an incident, issuing safety recommendations, studying chemical safety issues, and evaluating the effectiveness of other government agencies involved in chemical safety. Since 1998, the CSB has investigated chemical incidents that primarily occurred at land-based U.S. industrial facilities, and it has produced detailed reports for each investigated incident. Because of each committee's differences in developing a CSB report, there have been variabilities in the way information is provided in these reports. Such variabilities hamper the establishment of consistent predictive and preventive analyses. However, most CSB reports adhere to specific safety standards or guidelines and have consequently provided a base level of consistent information. Along with incident

causes and consequences, the reports covered whether Process Safety Management (PSM) was applied according to the 1992 Occupational Safety and Health Administration PSM Rule (OSHA, 1992) in plant, whether the Process Hazard Analysis (PHA) contained in the PSM was performed, and recommendations related to PSM were followed. A company must apply the PSM standard if it occupies a flammable chemical category (Category 1) or if it meets a threshold working quantity for 137 highly toxic and reactive chemicals specified by the Occupational Safety and Health Administration, OSHA (2020a). It should be emphasized that a number of CSB reports also included other detailed information such as incident scenarios and process conditions.

Several researchers have attempted to consolidate and categorized the CSB reports but at a high abstraction level. Most of them analyzed CSB reports based on PSM, PHA, and recommendations made to each incident facility. For example, Kaszniak (2010) analyzed 21 CSB reports to examine the problems related to PHA, one of the PSM elements. Amyotte et al. (2011) analyzed 63 CSB reports and bulletins that mainly concerned inherent safety principles and the hierarchy of controls along with PSM elements. Baybutt (2016) reviewed 64 CSB reports to identify the top contributing factors that were mainly related to PSM elements.

Nevertheless, an unanswered question remains whether the CSB reports can also sufficiently provide lower-level information such as chemical types involved and process conditions. In particular, since such lower-level information has been deemed key factor in seeking inherently safer design (ISD), the information would help further develop the ISD strategy; ISD aims to proactively eliminate hazards or risks and is one of the most

effective risk reduction strategies (Park et al., 2020). To check the presence of lower-level information in CSB reports, CSB reports are analyzed in more detail.

In this chapter, Section 3.2 describes the objective in detail. In Section 3.3, the methodology used is described, and in Section 3.4, the results are presented. Finally, conclusions are summarized, and future work is indicated in Section 3.5.

3.2. Objective and approach of this chapter

As mentioned, this chapter analyzes CSB reports at a more granular level of detail than previously published. The main purpose is to check whether useful information can be obtained in the form of commonalities in incident contributing factors to take ISD into account. On the basis of the former analysis, this study also checks the data availability of the chemical properties associated with the chemicals in the CSB reports; the data is obtained from the publicly available databases. Up to July 20, 2020, the CSB has investigated 94 incidents that occurred in the U.S., mainly at onshore facilities. The full list of those incidents is given in the Appendix 3A.

To date, there has been no publicly available systematic study to identify the commonality of process incident characterizing factors that can be extracted from the given incident scenario and other information in the CSB reports. The reason for this lack of effort is likely associated with the difficulty in utilizing information from inconsistently formatted CSB reports. The lack of consistency poses a significant challenge in obtaining and comparing process conditions that, if available, were often referred to in different units. In addition, information related to incident scenarios were not methodically

described. Hence, additional effort is required to categorize the scenario characteristics described in the CSB reports.

3.3. Method

This section outlines our method of collecting incident data on multiple characterizing factors in CSB reports and additional external sources in order to shed more light on understanding these chemical process incidents (furthermore briefly referred to as 'chemical incidents'). The conceptual flow diagram for this method is illustrated in Figure 3.1. As mentioned in Section 3.1, the chemical incident data obtained from CSB reports have no standard format; however, various incident data can be consistently collected based on three groups of factors: (1) factors related to incident causes, (2) factors related to incident scenarios, and (3) factors related to incident consequences. Generally, a hazardous chemical can trigger a chemical incident after leaking/being released from some process equipment. Hence, the knowledge associated with chemical properties at a specific process condition is essential to understanding the causes of chemical incidents. The data for scenario factors is also reviewed to understand the connection between incident causes and consequences. In the last step, this method evaluates data for consequence factors from CSB reports and external sources.

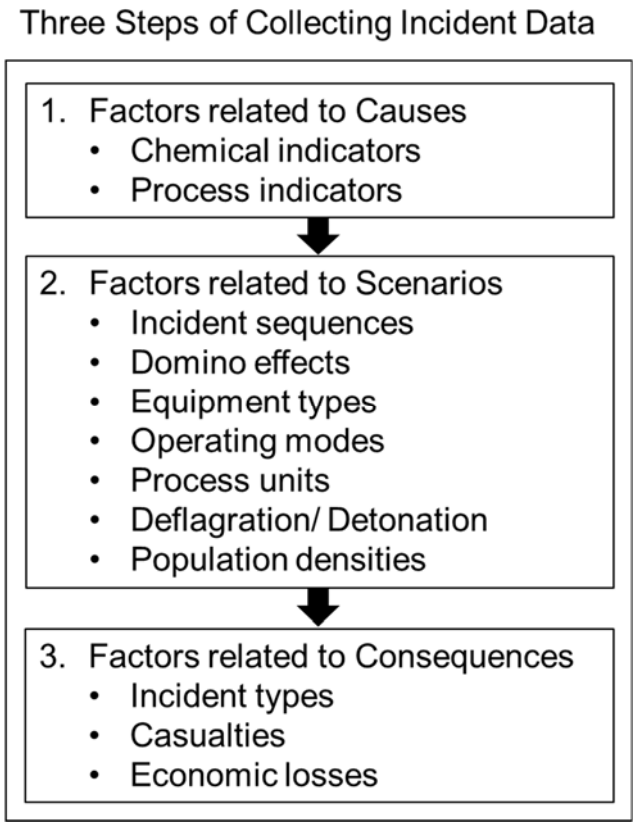


Figure 3.1 Flow diagram of the proposed method.

3.3.1. Step 1: factors related to chemical incident causes

Causal factors were divided into two groups – chemical indicators and process indicators. Chemical indicators are associated with the hazards of the chemicals involved in incidents, whereas process indicators are associated with process conditions when an incident occurs (Park et al., 2020). As shown in Table 3.1, 17 causal indicators (twelve chemical and five process indicators) were selected in this study because they have been often used in previous studies as safety indicators (Park et al., 2020).

Table 3.1 List of selected 17 causal factors in two groups: chemical and process indicators

12 Types of Chemical Indicators	5 Types of Process Indicators
• Chemical involved • Molecular weight • Flashpoint • Autoignition temperature • Boiling temperature • Flammability limit: Upper Flammability Limit (UFL) - Lower Flammability Limit (LFL) • Vapor specific gravity • Toxicity— life-threatening health effect level of Protective Action Criteria for chemicals (PAC-3) [a] • NFPA 704 standard: Reactive ratings • NFPA 704 standard: Special properties • Heat of combustion • Dust relevancy	• Quantity of chemical (inventory) • Incident pressure • Incident temperature • Relevancy of vacuum • Equipment type

[a] Toxic chemical exposure causing acute effects requires prompt action (Handbook, 2016). For the 60-minute health effect criteria, PAC-3 is preferred because it provides sufficient data, compared to AEGL-3, ERPG-3, and TEEL-3 concentration; for definitions, see https://edms.energy.gov/pac/TeelDef).

3.3.1.1. Collecting data associated with chemical indicators

The 12 chemical property values were collected from various references. First, data of chemical involved (type) and dust relevancy was gathered based on the CSB

reports' descriptions. Except for the special case CSB reports remarked, data of molecular weight, boiling point, flashpoint, autoignition temperature, the heat of combustion, upper flammability limit (UFL), and lower flammability limit (LFL) was obtained from Yaw's Chemical Properties Handbook (1999), DIPPR (2019), NIST (2020), or numerous Material Safety Data Sheets (MSDSs). The data of toxicity criteria, PAC-3, was collected from the U.S. Department of Energy (2020). Lastly, standard reactive rating and special properties were gathered from NFPA 704 (2017).

For chemical mixtures, the following guideline was created to prioritize what chemical property (i.e., chemical indicators) data should be collected.

1. If the CSB reports assumed there was a principal chemical that was the main contributor to the incident, then select that chemical and add its chemical properties.
2. If the CSB reports did not specify a principal chemical but included the chemical composition of a mixture, then calculate the mixture's UFL and LFL via the approximate Equations (1) and (2), respectively.
3. If not all the flammability limits or other properties of the mixture constituents were known, then select the severest of the constituents' present — the widest range of explosive limits, the heaviest molecular weight, lowest autoignition temperature, lowest boiling temperature, highest vapor specific gravity, lowest toxicity value, highest reactive rating in NFPA 704 standard, and highest heat of combustion.

$$UFL_{mix} = \frac{1}{\sum_{i=1}^{n} \frac{y_i}{UFL_i}}$$ **Equation 3.1**

where

UFL_i is the upper flammability limit (in volume %) of chemical i in fuel and air (vapor mixture),

y_i is the mole fraction of chemical i, and

n is the number of combustible chemical species in a vapor mixture (Crowl & Louvar, 2001).

Similarly,

$$LFL_{mix} = \frac{1}{\sum_{i=1}^{n} \frac{y_i}{LFL_i}}$$ **Equation 3.2**

where

LFL_i is the lower flammability limit (in volume %) of chemical i in fuel and air (vapor mixture).

Nevertheless, uncertainties of all selected property values implicitly remained. These values were determined by methods differing in detail, while the mixture relations were also approximations. Moreover, not all CSB reports did provide the specific chemical compositions involved.

3.3.1.2. Collecting data associated with process indicators

This section reviews the availability of information regarding the amounts of the chemicals involved in chemical incidents.

Because different types of units (i.e., mass, volume, mass flowrate, and volumetric units) were used in the CSB reports to characterize quantities, we converted all quantities into a mass unit (Figure 3.2). Of various units representing chemical quantities, a mass unit was preferred because mass units are comparable regardless of process conditions. To convert volume units, standard densities were employed; the densities at 25 °C if available (otherwise, either 15 °C or 20 °C) were used. A fixed chemical release time of 30 min was adopted for mass flowrate units if there was no estimated release time in the CSB reports. Regarding volumetric flowrate units, a 30-min release time and density at standard conditions were used together.

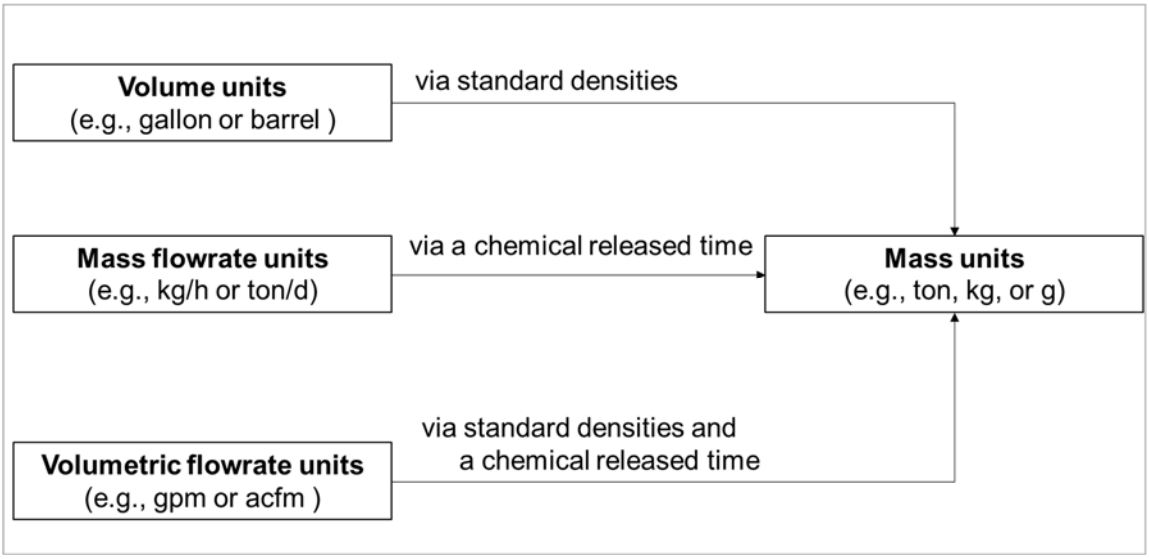


Figure 3.2 Consolidating the units for chemical quantities into mass units

If a chemical involved was not remarked in the CSB reports, it was impossible to convert a given quantity expressed in volume or volumetric flowrate units because the

appropriate chemical density was not selected. Meanwhile, when the quantity was not reported, an alternative was proposed to estimate the value in this study. For instance, in the case of an explosion, it was possible to utilize the equivalent TNT mass method to estimate the quantity of chemical involved (Equation 3.3). For estimating the quantity, the blast strength of an explosion was measured via a function of distance from the farthest damage point to the chemical source. The overpressure for calculating the TNT method can be assessed through an incident aftermath photo in a CSB report.

$$m_{TNT} = \frac{\eta m \Delta H_c}{E_{TNT}} \qquad \textbf{Equation 3.3}$$

where

m_{TNT} is the equivalent mass of TNT (mass),

η is the empirical explosion efficiency (unitless number varying between 1% and 10%). The value of 1% was adopted to consider maximum chemical amount),

ΔH_c is the energy of explosion of the flammable gas (energy/mass), and

E_{TNT} is the energy of explosion of TNT (Crowl & Louvar, 2001).

Regarding process pressure and temperature values, process conditions at the time of each incident were considered instead of normal operating conditions. This is because some incidents occur outside of normal operating conditions (e.g., while shutting down or maintenance). Hence, the pressure and temperature of an incident can differ from values

at normal operating conditions. When the CSB reports merely indicated that an incident occurred at the atmospheric conditions but without a specific statement, the pressure and temperature at the time of the incident were chosen as 1 atm and 25 °C, respectively.

In the case that CSB reports did not state a specific value for vacuum pressure, we separately checked the information on processes that were operated under the vacuum conditions.

In agreement with the CCPS (1998) categories, ten types of equipment were considered to systematically cover all types of equipment in the chemical industry. Appendix 3B in Appendix lists the ten types and the detail subtypes for each equipment type.

3.3.2. Step 2: factors related to chemical incident scenarios

For incident scenario factors, the following seven factors were assessed: successive incident types, equipment types, operating modes, process units, domino effects, detonation relevancy of explosion incidents, and population densities (as indicated earlier in Figure 3.1)

An incident, e.g., an explosion, may be followed by a fire or a toxic release from the same affected part of the plant. Therefore, successive incidents — an initial incident and its following incidents — were explored. Here is not meant so-called *domino effect events*, which will be addressed further below. Successive incidents describe a series of phenomena appearing during an incident, unlike traditional chemical incident types, which are often assigned as one type of incident and will be explained further in Section 3.3.3.

The successive incident types were determined along with the incident descriptions in the CSB reports. Three successive event types were identified based on the specified successive incident types (Table 3.2). For example, if, according to a CSB report, an explosion and a fire occurred, such as is the case when a release from a pressurized vessel formed a vapor cloud and a rain-out pool, subsequently, the vapor cloud found an ignition source and exploded, and by flash-back ignited the pool. In the preceding scenario, there was a series of VCE and a pool fire. Such successive incident types were selected in that order. When CSB reports did not specify the incident type, a fire and an explosion were labeled 'Unspecified fire' and 'Unspecified explosion', respectively, as illustrated in Table 3.2.

Table 3.2 14 specified incident types for successive incidents

	Specified incident types
Fire	• Not specified fire • Flash fire • Fireball • Jet fire • Pool fire
Explosion	• Not specified explosion • Physical explosion [a] • Vapor Cloud Explosion (VCE) • Boiling Liquid Expanding Vapor Explosion (BLEVE) • Dust explosion
Chemical release	• Toxic release • Liquid chemical release
Asphyxiation	Asphyxiation
Blowout	Blowout

[a] According to CCPS (2000), a physical explosion is defined as the incident associated with a catastrophic rupture of a pressurized gas-filled vessel. The pressurization need not but can be caused by an internal gas explosion.

Along with successive incidents, domino effects were considered. Most CSB investigators have not employed the technical term 'domino effect', although their descriptions of incidents occasionally imply a domino effect. Two definitions of a domino effect were considered to consistently determine it. First, in Khan and Abbasi's (1998) words: "situations when a fire, explosion, missile, and/or toxic load generated by an incident in one unit in the industry causes secondary and high order incidents in other units". Second, Antonioni et al. (2009) defined it more strictly as "the secondary effect is stronger than the primary one", which was also called escalation. Since determining a domino effect based on the latter definition was challenging due to the lack of information obtained from the CSB reports, the former definition was adopted in this study. For clarification, the "other units" in the former definition were regarded as "other equipment" in a process-related plant. Besides, the incident that led to projectile debris potentially causing secondary incidents was not categorized as a domino effect.

As mentioned in Section 3.1, most researchers have considered only one single main equipment type from which a chemical leaked or was released during the chemical incident (see Section 3.3.3). However, the relevant equipment types from which the released material originated was examined along with the main equipment type. For example, when an incident occurred due to hydrocarbon leakage from a pipeline that was connected to a process vessel, the process vessel was designated as the relevant equipment in such an incident.

The incidents of the CSB reports were categorized into 10 types of operating modes:

- Routine operation
- Startup after a shutdown
- Startup of new construction/equipment
- While shutting down
- Maintenance operation after a shutdown
- Maintenance operation without a shutdown
- During special test and trial
- Loading or Unloading operation
- Other operation modes
- Operating modes not reported

Because the CSB reports have not covered merely process-related incidents, but also storage and other incidents, process units were distinguished based on three sub-factors: (1) Inside Battery Limit (ISBL) unit in a chemical process facility, (2) Outside Battery Limit (OSBL) unit in a chemical process facility, and (3) Not Applicable (N/A), which refers to a facility that is not process-related.

Besides, detonation was investigated based on the severity of the damage caused in relation to the exploded amount, the occurrence or likelihood of detonation, and being the most destructive among explosion events. In a number of cases, e.g., with solids or liquids able to detonate the occurrence was obvious. However, it should be noted that

verifying whether a detonation occurred is rather complicated and uncertain. In particular, attempting to distinguish the explosion type in VCE would be necessary to properly understand the correlation between incident type and its consequences, because also VCE detonations are capable of producing much more damage than deflagrations (Chamberlain et al., 2019).[4] In case of gas explosions in confined situation, e.g., pipelines, a transition from deflagration to detonation is well possible and this study looked whether this occurred.

Finally, the residential population densities surrounding chemical incident facilities were collected because these may be relevant to consequence impacts. The data was obtained from two data sources: the EPA (https://echo.epa.gov/) and the Missouri Census Data Center (http://mcdc.missouri.edu/applications/caps.html). The unit of data was persons per square mile.

3.3.3. Step 3: factors related to chemical incident consequences

Chemical incident consequences were explored through six factors — types of chemical incidents, casualties (injuries and fatalities), and economic losses (OSHA and EPA penalties); due to lack of information, concrete economic losses and environmental damage remain outside consideration.

Chemical incidents were categorized based on the following nine types:

[4] It should be noted that verifying whether a VCE detonation occurred has a long time been rather uncertain and a matter of controversy, but in recent years clear detonation markers for VCE have been identified by Oran et al. (2020). The authors distinguished detonation impact from typical explosions, such as VCE deflagration, BLEVE, or pressure burst.

- Liquid chemical release without fire or explosion incident
- Fire incident
- Explosion incident
- Toxic release incident
- Fire + Explosion incident
- Fire + Toxic dispersion incident
- Explosion + Toxic dispersion incident
- Fire + Explosion + Toxic dispersion
- Others

The above incident types include successive incidents (Table 3.2) and domino effects together.

For casualties, the CSB reports described four different measures: fatality, injury, medical attention, and hospitalization.

Using direct financial damages in the aftermaths of chemical incidents would be preferred way to represent economic losses. However, it is challenging to obtain such explicit damage values. Alternatively, OSHA and the EPA penalties were considered in this study as possible indicators of economic losses. Once an incident occurs, the OSHA penalties are imposed due to the violation of the OSHA guidance associated with onsite impacts, whereas the EPA penalties are for offsite impacts. Since the EPA penalties require the budget to comply with environmental requirements for our community, the penalties imply to a certain extent economic offsite damage. In contrast, the penalties of

OSHA do not require the budget of such required compliance. Hence, using the OSHA penalties as an economic indicator would grossly understate the actual losses. Nevertheless, this study analyzed the OSHA penalties to check whether the values could be utilized to investigate any relations with other incident factors, such as a correlation between casualties (e.g., fatalities or injuries) and the OSHA penalties. Many OSHA penalties differed between its initial and actual penalties, so we examined both OSHA penalties. Although labeling penalties of OSHA and the EPA would be more precise than the term, "economic losses", this factor was named as economic losses in order to leave the door open for future works.

3.4. Results and discussion

As explained in Section 3.3, we examined available CSB reports to obtain useful information that may explain chemical incidents based on twenty-seven factors. Such factors were categorized into three groups: cause, scenario, and consequence factors. This section shows the data availability for each factor step-by-step and briefly introduces attributes of the obtained data.

3.4.1. Brief high-level investigation

Before presenting the lower-level information of the CSB reports in the following subsections, we would like to share high-level information aligned with previous

researchers' contributions mentioned in the Introduction. A brief overview is given regarding the incident facility sites and PHA performance alongside PSM coverage.

The locations of incidents investigated by the CSB over the last two decades are presented in a U.S. map (Fig. 3.3). Most investigated incidents occurred in the states of Texas, Louisiana, West Virginia, California, and Ohio. Figure 3.3 distinguishes between incidents of PSM standard covered facilities and other incidents through different colors. Red dots represent the incidents that met the PSM standard (38.3% of the CSB incidents), blue dots represent the incidents that did not meet the PSM standard (41.5%), and orange dots represent where there was no clear information regarding the PSM standards (20.1%).

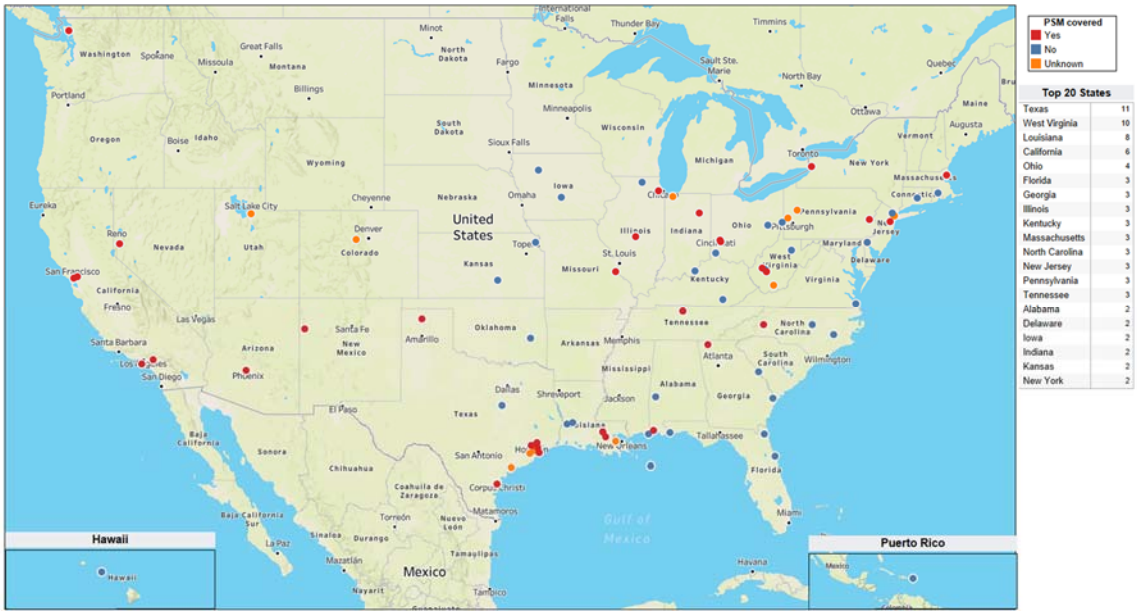


Figure 3.3 The U.S. map of 94 incidents investigated by the CSB

Although most of the facilities that included highly flammable or toxic chemicals are covered by the PSM standard, facilities covered by the standard are not always more hazardous. This is because there are often exceptions for the PSM coverages. For example, some facilities have insufficient information about chemical mixtures to determine if the PSM standard covers them (e.g., #16 and #67 in Appendix 3A). According to OSHA guidelines, oil or gas well-operating facilities that include highly flammable hydrocarbons are not covered by the PSM standard (e.g., # 80 and #92 in Appendix 3A). Furthermore, in many states public sectors do not observe the OSHA guidelines (e.g., #37 in Appendix 3A).

Many CSB investigation teams examined whether the chemical facilities performed a PHA prior to an incident. Based on the CSB reports, the correlation between the PSM standard coverage and PHA performance was analyzed, as shown in Figure 3.4. This figure illustrates PHA performance in terms of the number of incidents at facilities covered by the PSM standard. As we intuitively grasped, the facilities covered by the PSM standard had actively more often performed a PHA compared to the uncovered facilities. The PHA performance ratio was 77.8% (28 in 36) and 46.3% (19 in 40) for the covered and uncovered facilities, respectively. However, of the facilities covered by the PSM, a considerable percentage, 61.6% (22 in 36), inappropriately performed PHA.

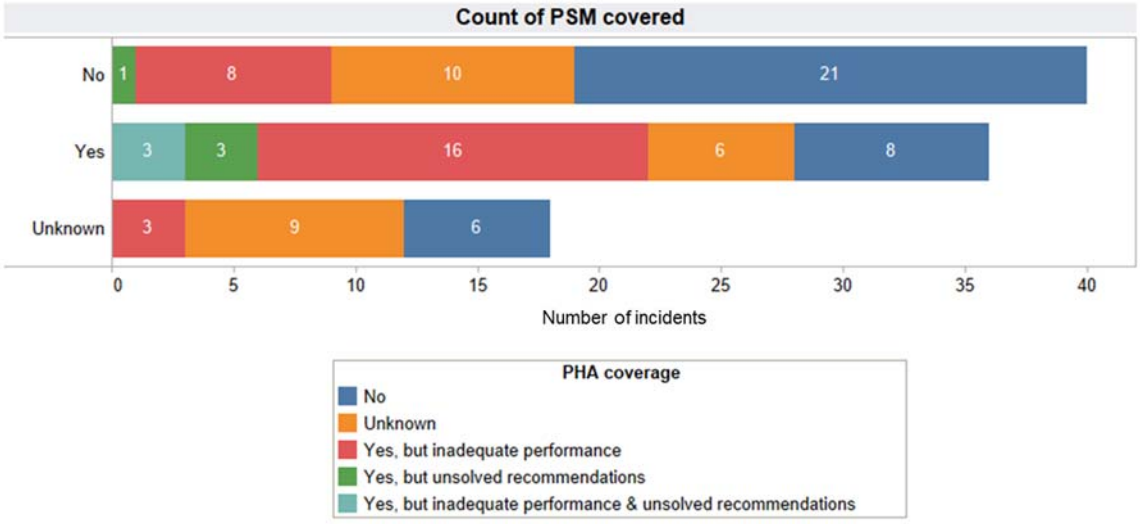


Figure 3.4 Number of chemical incidents in the CSB reports based on PSM standard coverage and PHA performance

3.4.2. Availability of data related to causal factors and their attributes

The 84 chemicals that contributed to the 94 chemical incidents were identified (Table 3.3). Of the 94 incidents in the CSB reports, 50.0% (47 in 94) involved a single chemical and the other half two or more chemicals. Because several CSB reports could not apparently provide the main chemical(s) involved in the chemical incident, such chemicals were labeled with the same term as the CSB reports. For example, when the CSB reports referred to the main chemical merely as 'hydrocarbons' without a detailed explanation of actual chemical types, the chemical was labeled as 'hydrocarbons. In addition, when any chemical was not determined in the CSB reports, it was categorized as 'unknown'. Although such labels, hydrocarbon and unknown, were listed as a reference in Table 3.3, their chemical properties (i.e., chemical indicators) were not further investigated in this study.

Table 3.3 Count of chemicals involved in chemical incidents investigated in the CSB reports

Chemical involved	Count	Chemical involved	Count	Chemical involved	Count	Chemical involved	Count
Hydrocarbons	6	Acetylene	1	Gasoline	1	PETN (Pentaerythritol, tetranitrate)	1
Chlorine	4	Acrylic monomer	1	Gilsonite	1	Phenolic resin dust	1
Methane	4	Alkylates (Isooctane)	1	Hydrochloric acid	1	Phosgene	1
Methanol	4	Allyl alcohol	1	Hydrogen fluoride	1	Propylene glycol phenyl	1
Propane	4	Allyl chloride	1	Hydrogen sulfide	1	Raffene oil	1
Sulfuric acid	4	Alpha-pinene	1	hydroxylamine	1	RDX (hexahydro-1,3,5-trinitro,1,3,5-triazocine)	1
Unknown chemical	4	Aluminum dust	1	iso-propyl alcohol	1	Silver nitrate	1
Ammonia	3	Ammonium nitrate	1	Maltodextrin	1	Sodium hydrosulfide (NaHS)	1
Hydrogen	3	Amodel (polymer)	1	Methomyl	1	Sodium hydroxide	1
(Light) naphtha	2	Antimony pentachloride	1	Methyl ethyl ketone (MEK)	1	Sucrose	1
Heptane	2	Benzene	1	Methyl mercaptan	1	Tetrahydrofuran (THF)	1
Iron dust	2	Benzoyl peroxide (BPO)	1	Methylcyclopenta-dienyl manganese tricarbonyl	1	Titanium (power)	1
Isobutane	2	Calcium carbide	1	Mineral seal oil	1	TNT (2,4,6-trinitrotoluene)	1
Naphtha	2	Caramel color liquid	1	Mineral spirits	1	Urea ammonium nitrate solution	1
Nitrogen	2	Carbon black	1	Mono-nitrotoluene	1	Vinyl chloride monomer (VCM)	1
Oleum (fuming sulfuric acid)	2	Carbon monoxide	1	Nitric acid	1	Vinyl fluoride	1
Polyethylene (dust, wax)	2	Chloromethane	1	Nitric oxide	1	VM&P naphtha	1
Propylene	2	Cyclohexane methanol	1	Nitrous oxide	1	Zinc	1
Sodium hypochlorite	2	Ethyl acetate	1	n-Propyl alcohol	1	Zinc oxide	1
2-Ethyl-1-hexylamine	1	Ethylene oxide (EO)	1	ortho-Nitro chlorobenzene	1	Zirconium	1

Chemical involved	Count	Chemical involved	Count	Chemical involved	Count	Chemical involved	Count
4-Methylcyclo-hexanemethanol	1	Gas oil	1	Peroxydicarbonate	1	Monomethylamine (MMA)	1

Grand Total 121

The data availability of 12 chemical indicators was analyzed (Table 3.4). Based on the CSB descriptions, the chemicals involved in 90 incidents were determined after they were matched with appropriate CAS numbers. The chemicals that could not be matched with any CAS number were deemed missing data; there were such chemicals (i.e., unknown or hydrocarbons) in four incidents. After that, the data of 10 chemical indicators in Table 3.4, except for dust relevancy, were collected from various external databases and references. Two categorical variables — reactive rating and special properties — were selected based on NFPA 704 standard (2017). Of the remaining 8 continuous variables, the values of boiling point, toxic measure PAC-3, and molecular weight were available in most cases from external databases, such as Yaw's Chemical Properties Handbook (1999), DIPPR (2019), or MSDSs. For the 90 incidents at which main chemicals were known, the data of boiling points and molecular weights and PAC-3 was available in most of the incidents. In contrast, the data of autoignition temperature and flammability limit was less available. The selection of dust relevancy was performed exclusively based on the CSB descriptions.

Table 3.4 Available information for each chemical indicator in the CSB report

No.	Chemical indicator	Count	Percentage of 94 incidents
1	Chemical involved	90 [a]	95.7 %
2	Molecular weight	86	91.5 %
3	Flashpoint	63	67.0 %
4	Autoignition temperature	63	67.0 %
5	Boiling temperature	88	93.6 %
6	Flammability limit	54	57.4 %
7	Vapor specific gravity	77	81.9 %
8	PAC-3	85	90.4 %
9	Heat of combustion	57	60.6 %
10	Reactive ratings in NFPA 704 standard	90	95.7 %
11	Special properties in NFPA 704 standard [b]	90 [c]	95.7 %
12	Dust relevancy	10 (Yes) + 84 (no)	100 %

[a] The value was calculated except for chemicals that cannot be matched with any chemical CAS number (i.e., unknown chemicals).
[b] There are four types of special properties considered: oxidizer, a substance reacting with water, simple asphyxiant gas, and strong corrosive acid or base.
[c] The number of incidents was counted when at least one special property was related to the incident.

Similarly, the availability of process indicators in the CSB reports was analyzed (Table 3.5). Process indicators were expressed in various, often not SI units but probably in units used in the facility investigated. To obtain chemical quantities in a consistent unit, we converted their numerical values into a SI mass unit, tonne; the types and frequencies

of the chemical quantity units used in the CSB reports were counted in Table 3.6 for the sake of statistics. As shown in Table 3.5, among the three numerical process indicators (chemical quantity, incident pressure, and incident temperature), the data of process pressure at an incident (here called incident pressure, not to confuse with the term used to indicate an incoming shock wave pressure) was more available or predictable in the CSB reports, compared to the others. The two categorical process indicators (relevancy of vacuum and equipment type) were analyzed in detail in Table 3.7 and Figure 3.5(a), respectively. Table 3.7 displays the pressure units and their used frequencies concerning vacuum conditions. As shown in this table, various units of process pressures were employed. Atm, psig, psi, bara, and mmHg were the most frequently used in that order, and 12 incidents were associated with vacuum processes. Finally, the types of main equipment related to chemical releases were analyzed; 31.9% of the incidents were related to vessels, followed by piping and piping components (26.6%, 25 in 94) and reactors (10.6%, 10 in 96).

Table 3.5 Available information for each process indicator in the CSB reports

No.	Process indicator [unit]	Count	Percentage of 94 incidents
1	Quantity of chemical involved [tonne]	71	75.50%
2	Incident pressure [bara]	78	83.00%
3	Incident temperature [°C]	71	75.50%
4	Relevancy of vacuum [unitless]	12(Yes) 81 (No) 1 (Not informed)	100.00%

5	Equipment type [unitless]	See Fig. 3.6(a)	98.90%

Table 3.6 Used units to represent quantities of chemicals in the CSB reports

	Unit	Count	Percentage of 94 incidents
Flowrate	acfm	1	1.1 %
	gpm	2	2.1 %
	kg/h	1	1.1 %
	lb/h	1	1.1 %
	st/d	1	1.1 %
Subtotal (flowrate)		**6**	**6.4 %**
Volume	barrel	1	1.1 %
	ft^3	1	1.1%
	gal	30	32.0 %
Subtotal (volume)		**32**	**34.0 %**
Mass	g	2	2.1 %
	lbm	28	29.8 %
	tonne	4	4.3 %
Subtotal (Mass)		**31**	**46.3 %**
N/A (missing data)		22	23.4%
Grand Total		94	100.0 %

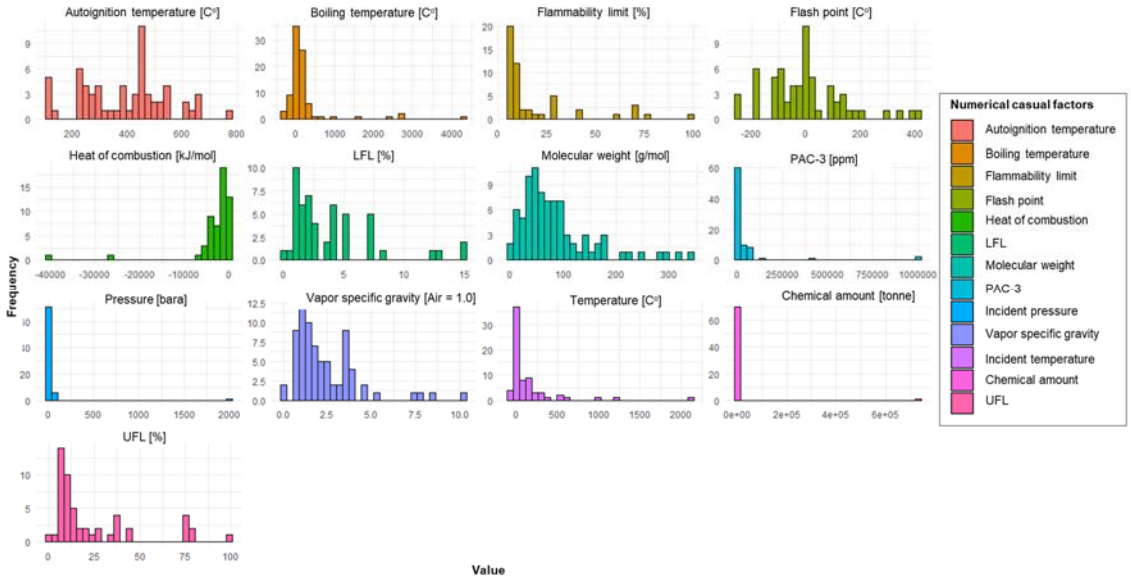
acfm, actual cubic feet per minute; gpm, gallon per minute; kg/hr, kilogram per hour; lb/hr, pound mass per hour; st/d, standard tonne per day; ft^3, cubic feet; gal, gallon; oz, fluid ounce; g, gram; lbm, pound-mass

Table 3.7 Used pressure units in the CSB reports concerning vacuum conditions

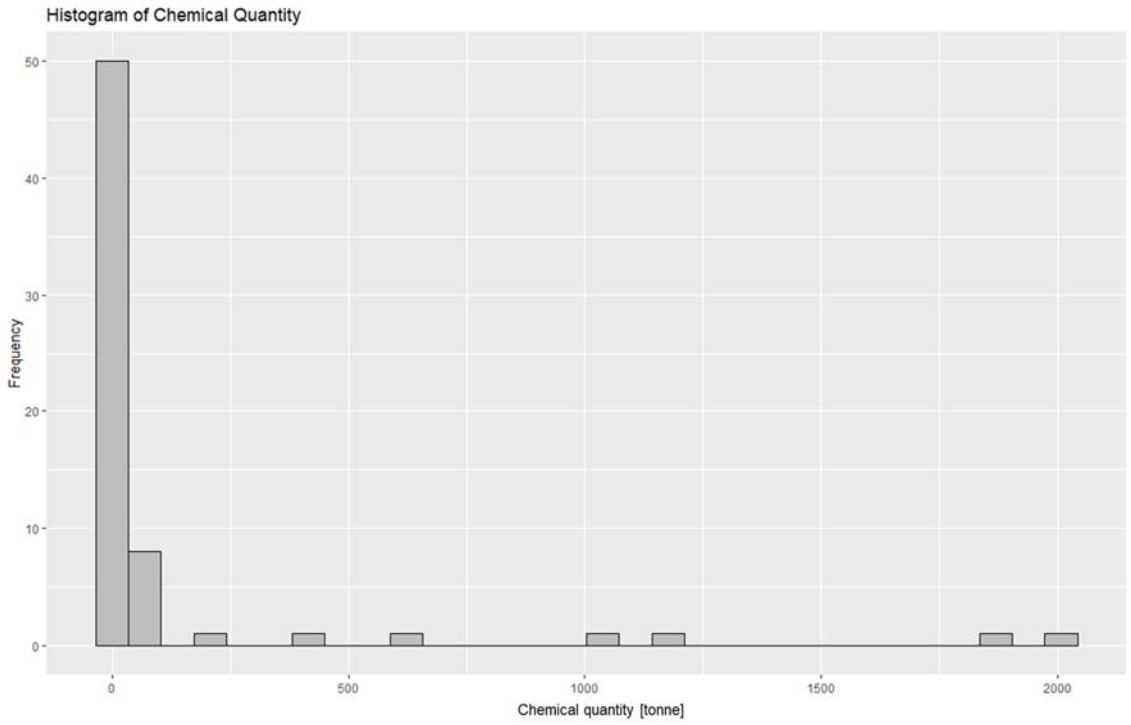
Related to vacuum	atm	bara	mmHg	psi	psig	N/A	Grand Total
No	31	3	0	6	26	15	81
Yes	4	2	1	0	4	1	12

N/I (no information)	1	0	0	0	0	0	1
Grand Total	36 38.3%	5 5.3%	1 1.1%	6 6.4%	30 31.9%	16 17.0%	94 100.0%

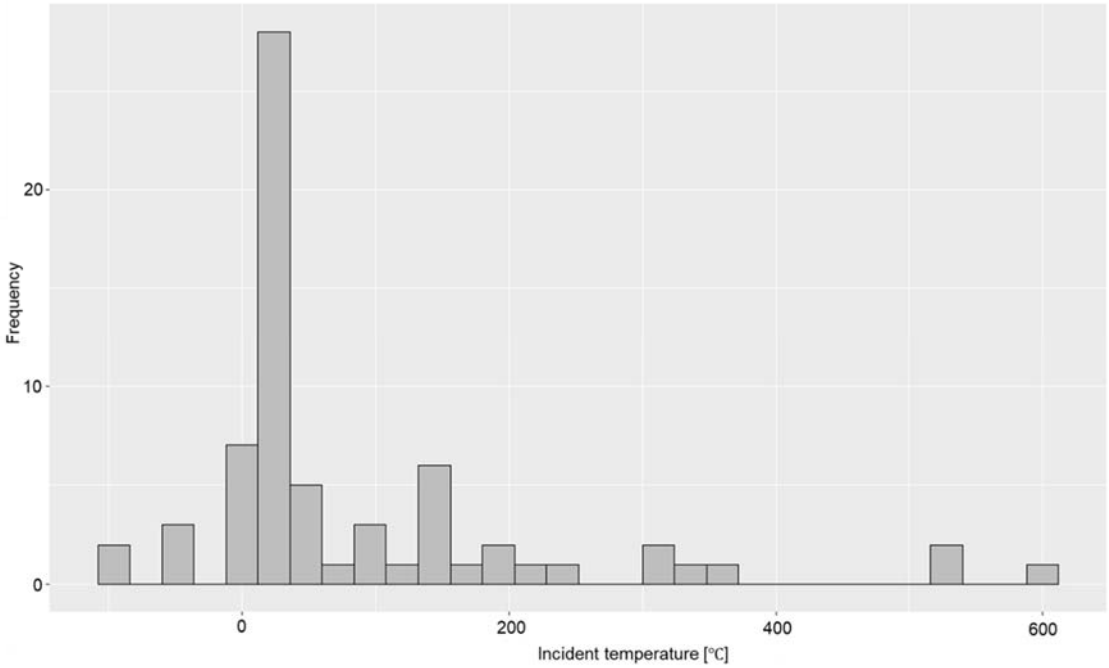
The collected numerical chemical and process indicators were analyzed further through their histograms to understand each attribute. Figure 3.5(a) displays the 13 numerical causal indicators: chemical quantity, incident pressure, incident temperature, flash point, boiling point, molecular weight, vapor specific gravity, autoignition temperature, LFL, ULF, flammability limit, and PAC-3. As illustrated in Figure 3.5(a), the shapes of the three process indicators' distributions are extremely right-skewed. Hence, excluding several outliers would be helpful to grasp the indicators' generic attributes. Thereby, after excluding outliers, further analysis of process indicators was performed. For example, many chemical quantities of the chemical incidents were under 2500 tonnes, except for six outliers of chemical quantities (#10, #48, #53, #69, #74, and #80 in Appendix 3A). After excluding these outliers, the histogram of chemical quantities indicates that most process chemical incidents occurred while chemicals less than 100 tonnes were leaking (See Figure 3.5(b)). Incident temperature and pressure are also redisplayed after excluding their outliers in Figure 3.5(c) and (d), respectively; in this study, the threshold of incident temperature's outliers was above 1000 ℃ (#52, #68, and #77 in Appendix 3A), and the threshold of the pressure's outlier was above 250 bara (#69 in Appendix 3A). These figures show that the temperature range of most incidents is from 25 to 150 ℃, and the pressure range from 1 to 10 bara.



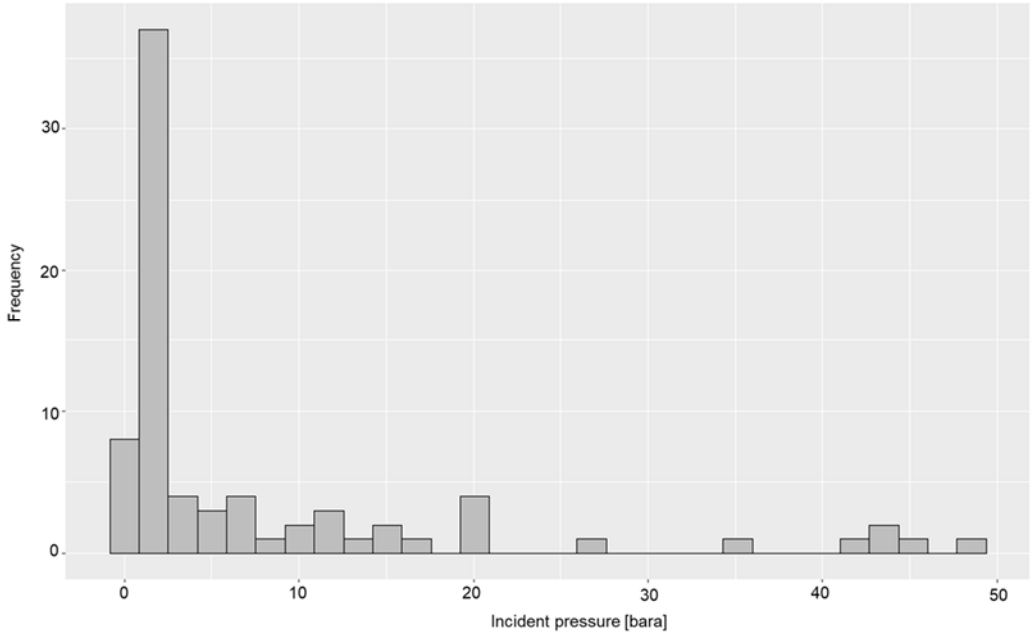
(a) Histograms of all numerical causal factors including outliers



(b) Histogram of chemical quantity without 6 outliers



(c) Histogram of incident temperature without 3 outliers



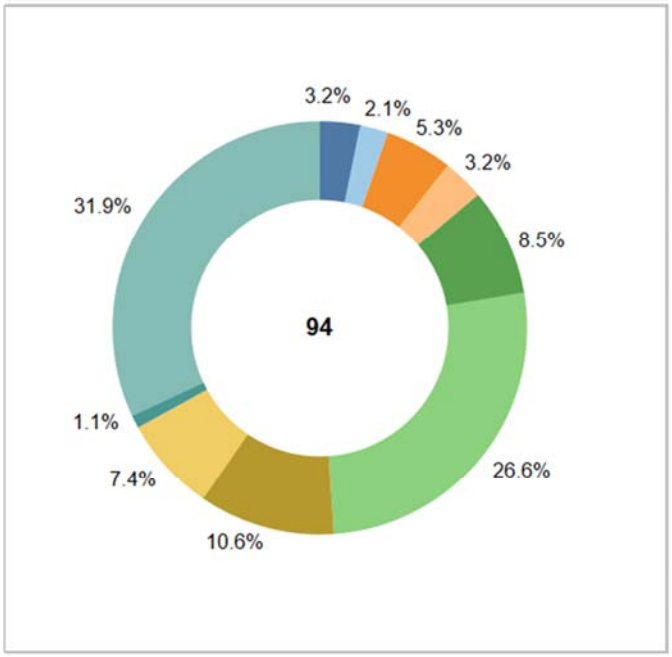
(d) Histogram of incident pressure without 1 outlier

Figure 3.5 The histograms of numerical indicators of the CSB reports

3.4.3. Attributes in collected scenario factors

This subsection shows the attributes of the collected seven scenario factors: relevant equipment types, incident sequences, domino effects, detonations/deflagration, operating modes, process units, and population densities. These scenario factors were considered either to comprehend causal and consequence factors in detail or to find causal relationships between them by adding supplementary reasons. Therefore, the results of several scenario factors can be explained along with the aforementioned causal factors (Section 3.4.2) or the following consequence factors (Section 3.4.3).

The gathered information on relevant equipment types was investigated along with the main equipment type (Figure 3.6). The frequencies of main equipment types in the chemical incidents show the point where chemicals often leaked. Meanwhile, the frequencies of relevant equipment types enable us to understand the flow of the chemicals involved in the incidents. For example, the two most frequently reported types of main equipment are ‘vessel’ (31.9%, 30 in 94) and ‘piping and piping components’ (26.6%, 25 in 94), as shown in Figure 3.6(a). As shown in Figure 3.6(b), most chemicals leaked at ‘piping and piping components fed from other equipment, such as mass transfer equipment types and vessels, at the time of the incidents. That is, the consequences of the incidents became severe because of the chemical supply from other equipment. In contrast, most chemicals that leaked from vessels did not flow from other equipment. This is because the amount of the chemicals in vessels was sufficient to make the event a severe incident.



(a) Distribution of main equipment (EQ) related to chemical leakages in the CSB reports

Main EQ	1st relevant EQ	2nd relevant EQ	Count	
Fired EQ			3	3.2%
Fluid transfer EQ			1	1.1%
	Heat transfer EQ	Others	1	1.1%
Heat transfer EQ			3	3.2%
	Mass transfer EQ		1	1.1%
	Piping and piping components		1	1.1%
Mass transfer EQ			2	2.1%
	Piping and piping components		1	1.1%
Others			7	7.4%
	Fluid transfer EQ	Piping and piping components	1	1.1%
Piping and piping components			10	10.6%
	Fired EQ		1	1.1%
	Fluid transfer EQ	Vessel	1	1.1%
	Heat transfer EQ		2	2.1%
	Mass transfer EQ	Solid-fluid separators	1	1.1%
			3	3.2%
	Others		1	1.1%
	Reactor		2	2.1%
	Solids handling EQ		1	1.1%
	Vessel		2	2.1%
	Unknown		1	1.1%
Reactor			9	9.6%
	Piping and piping components		1	1.1%
Solids handling EQ			6	6.4%
	Fluid transfer EQ		1	1.1%
Vessel			17	18.1%
	Fluid transfer EQ		2	2.1%
	Mass transfer EQ	Others	1	1.1%
		Fluid transfer EQ	1	1.1%
	Piping and piping components	Vessel	1	1.1%
			8	8.5%
Unknown			1	1.1%
Grand Total				**100.0%**

(b) Scenarios of chemical leakages from equipment (EQ)

Figure 3.6 Analysis of equipment related to chemical leakages concerning relevant equipment (scenarios)

The information on incident sequence provides the common mechanisms and the frequencies of successive incident events. The most frequent first incidents were physical explosions (26.6%, 25 in 94), followed by toxic releases (20.2%, 19 in 94), VCEs (13.8%, 13 in 94), flash fires (9.6%, 9 in 94), unspecified fires (7.4%, 7 in 94), and dust explosions (6.4%, 6 in 94). Figure 3.7 shows the frequency of successive incident events; it indicates that successive incidents rarely occur when the initial incident is a toxic release or asphyxiant event. In contrast, various successive incidents occur more frequently when fire or explosions are the initial incidents.

Aside from incident sequences, the domino effect was also analyzed using the guidelines described in Section 3.3. Of the 94 incidents, it was clear that 68% of the incidents were not associated with domino effects. However, deciding whether domino effects occurred at the rest of the incidents was challenging because of the lack of information in the CSB reports.

Of 64 incidents where explosions occurred, a deflagration to detonation transition was likely to have occurred in three incidents (#1, #21, and #79 in Appendix 3A).

1st incident	2nd incident	3rd incident	Count
Fireball			2
	Physical explosions	Unspecified fire	1
	Unspecified explosion		1
Flash fire			4
	Physical explosions		1
	Pool fire		2
	Unspecified explosion		2
Unspecified fire			2
	BLEVE		1
	Physical explosions		2
	Unspecified explosion		2
Blowout	Unspecified fire		1
		Lliquid chemical release	1
Dust explosion	Fireball		3
	Flash fire		1
	Unspecified fire		2
BLEVE	Unspecified fire		1
Physical explosion			6
	Fireball		7
	Flash fire		1
	Unspecified fire		10
		Unspecified explosion	1
VCE			2
	Flash fire		1
	Pool fire		2
	Unspecified fire		8
Unspecified explosion	Fireball		1
	Unspecified fire		2
		Lliquid chemical release	1
Toxic release			19
Asphyxiation			2
Liquid chemical release			2

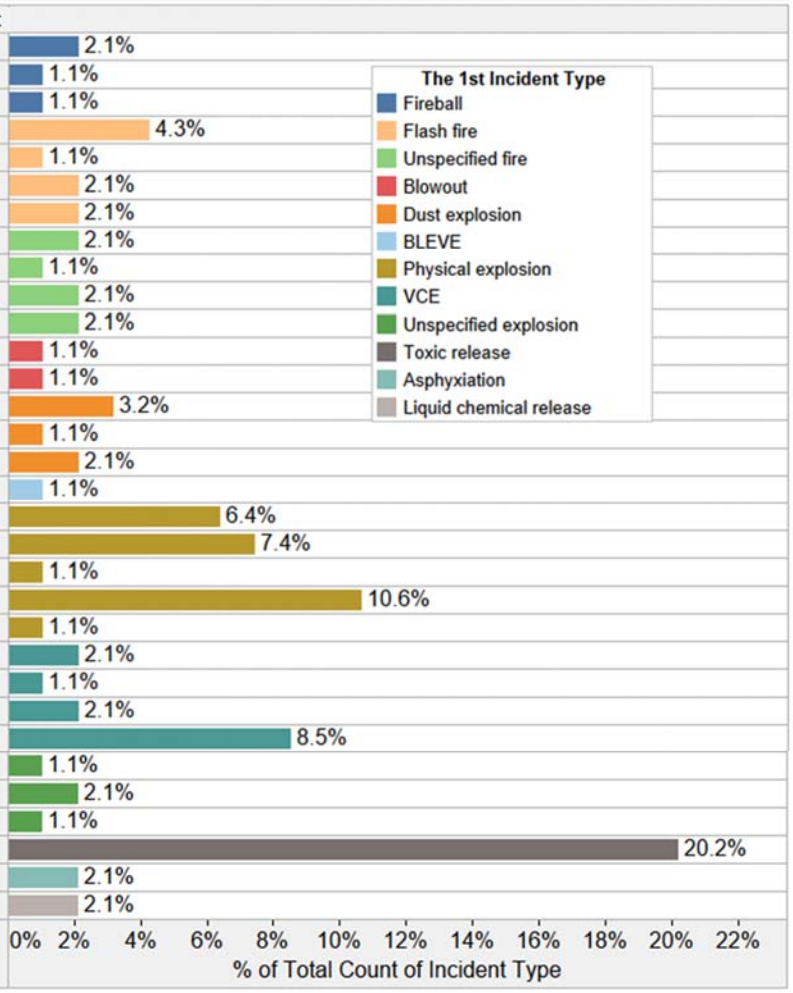


Figure 3.7 The count of the successive incident scenarios in the CSB reports

The process operation modes during the incidents were then analyzed (Table 3.8). This showed that 39.4% of incidents occurred during a routine operating mode, and 37.2% were associated with a shutdown (#2, #4, #5, and #6 in Table 3.8). Besides, a quarter of incidents occurred during maintenance (# 5 and #6 in Table 3.8).

Table 3.8 Process operation modes related to chemical incidents in the CSB report

No.	Process operating mode	Count	Percentage of 94 incidents
1	Routine operation	37	39.4 %
2	Startup after a shutdown	7	7.4 %
3	Startup of new construction or equipment	6	5.3 %
4	While shutting down	4	4.3 %
5	Maintenance operation after a shutdown	8	8.5 %
6	Maintenance operation without a shutdown	16	17.0 %
7	During special test or trial	1	1.1 %
8	Loading or unloading operation	6	6.4 %
9	Other operation modes	9	9.6 %
10	Operating mode not reported	0	0 %
	Grand total	94	100 %

The examination of process units that are ISBL, OSBL, or other cases showed that over 85% of the incidents occurred at chemical process facilities (either ISBL or OSBL). 57 Chemical incidents in the CSB reports occurred in ISBL units, 28 incidents in OSBL

units, and 9 incidents in neither of these units. This indicates that the incidents described in the CSB reports are more associated with the main process units, ISBL.

Regarding the residential population densities, most of the data were extracted from the EPA except for 6 incident facilities (#3, #5, #16, #40, #55, #67 in Appendix 3A), which were not available in the EPA data. The data of these 6 facilities were obtained from the Missouri Census Data Center based in 2010 instead. Based on the collected data, the average density per incident type (consequence factor) was illustrated in various ranges of residential population densities from 5 to 2,790 persons per square mile, as shown in Figure 3.8. In particular, the average population density surrounding the facilities associated with toxic dispersion incidents were less dense compared to fire + explosion incident cases. This may confirm the expectation that facilities handling toxic materials are set up in less densely populated areas.

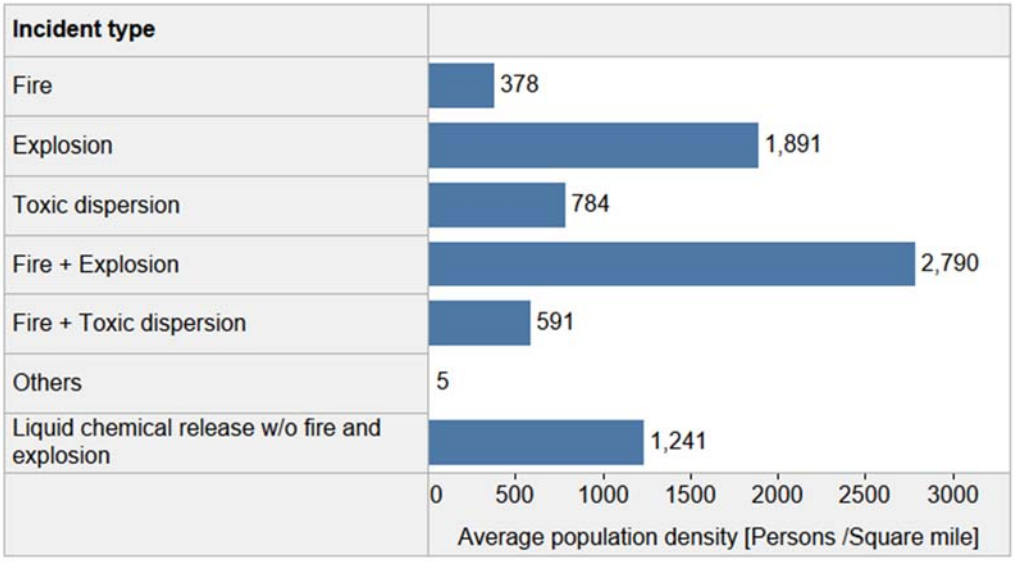


Figure 3.8 Average residential population densities per each incident type

3.4.4. Attributes in collected consequence factors

This subsection shows the information of the analysis of three consequence factors: chemical incident types, casualties, and economic losses (penalties).

All 94 incidents were categorized into 8 chemical incident types (Table 3.9). The most frequently investigated incident type was fire + explosion incident (58.5%), followed by toxic dispersion (20.2%), fire (7.4%), and explosion (7.4%). Two asphyxiant incidents corresponded to the incident type, 'others'. To specify incidents further, the data of incident types can be integrated with the data of incident sequences, described in Section 3.4.3.

Table 3.9 List of incident types in the CSB reports

Incident Type	**Count**	**Percentage of 94 incidents**
Liquid chemical release without fire or explosion	2	2.1 %
Toxic dispersion	19	20.2 %
Fire	7	7.4 %
Explosion	7	7.4 %
Fire + Explosion	55	58.5 %
Fire + Toxic dispersion	2	2.1 %
Explosion + Toxic dispersion	2	2.1 %
Others	2	2.1 %
Grand Total	**94**	**100 %**

Although the CSB reports often showed the number of individuals needing medical attention and hospitalization, not all CSB reports recorded such injury types in

detail. Alternatively, without distinction between different injury types, the number of injuries and fatalities was counted as reported by the CSB. For example, several CSB reports (e.g., # 31, #75, and #76 in Appendix 3A) referred to casualties as medical attention (treatment) and residents' hospitalization. In such cases, hospitalization was assigned to injury, and medical attention was not included in this study.

The incident fatalities and injuries were analyzed together in various approaches. First, the sum of injuries in the 94 incidents was 1,174, whereas the sum of fatalities was 185. That is, the overall ratio of the sum of fatalities to the sum of injuries is around 6.3 in the 94 CSB incidents. Subsequently, the spread of fatalities and injuries in ratio per incident was illustrated by means of a scatter plot, as illustrated in Figure 3.9. The color in the figure shows incident type details, and the marks are labeled by the incident numbers presented in Appendix 3A. Purple dots in this figure indicate that there were significant casualties in several fire + explosion incidents. In addition, the different numbers aggregated per incident types were analyzed (Figure 3.10). To understand the chemical incident casualties in detail, injuries and fatalities were further divided into onsite people, emergency responders (ER), and offsite people. Figure 3.10(a) represents the average number of all injuries, onsite, ER, and offsite, from left to right. Similarly, Figure 3.10(b) shows the average number of all fatalities, onsite, ER, and offsite. These figures indicate that the average fatalities were higher than its corresponding injuries at fire incidents, unlike the other incident types. Besides, explosion incidents (explosion or fire + explosion) generated more emergency responders' casualties. Meanwhile, toxic dispersion and liquid

chemical release incidents significantly impacted surrounding communities because considerable people from offsite got injured from such incidents.

After comparing the average originally and actually imposed OSHA penalties depending on incident types, we concluded that PSM coverages affected the increased amount of actual OSHA penalties. When PSM covered process facilities of CSB incidents, the penalties on the facilities between original and actual penalties were almost the same. In contrast, actual penalties of the facilities that were not covered by PSM were often reduced. For this reason, the original OSHA penalties were utilized to compare with EPA penalties based on incident consequences without PSM coverage.

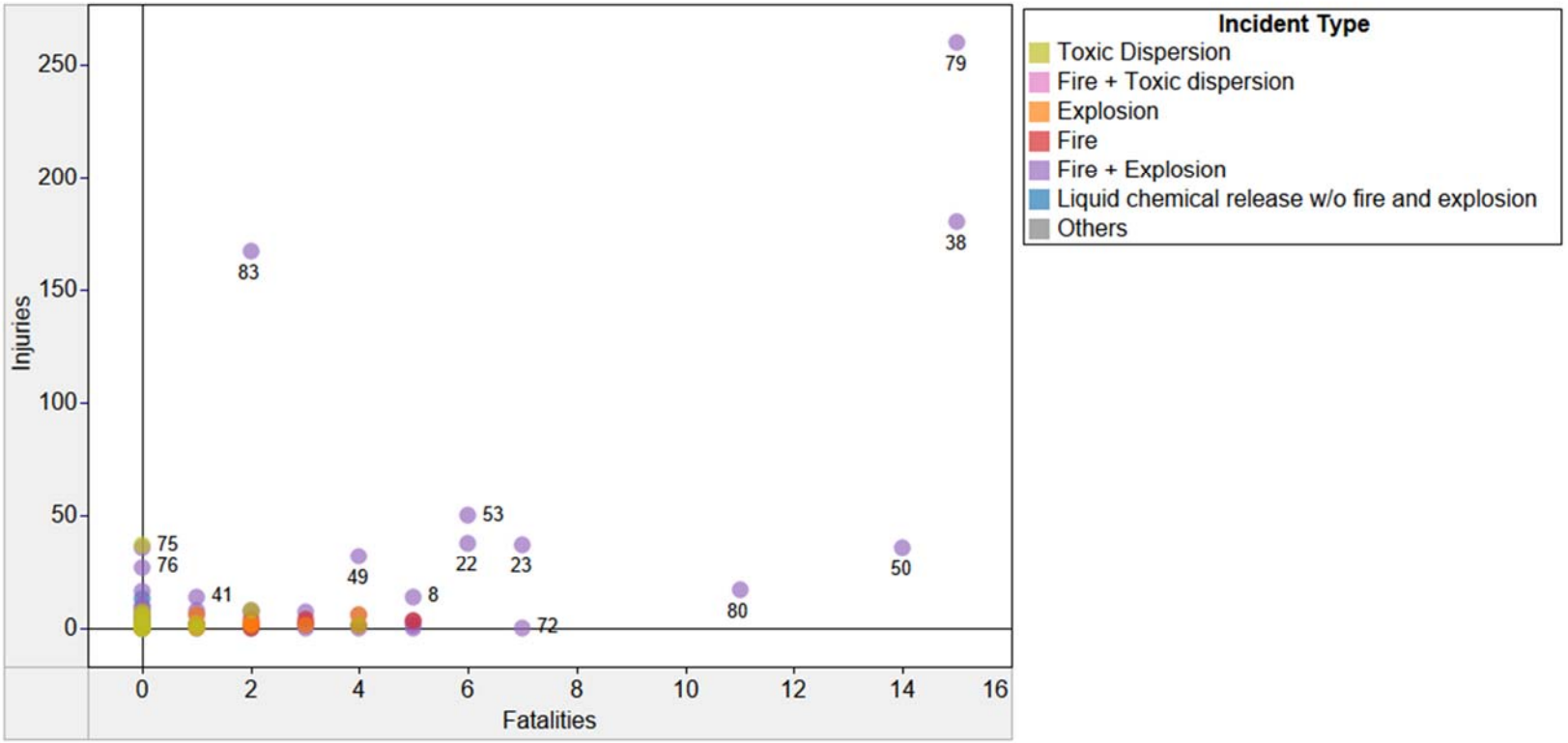
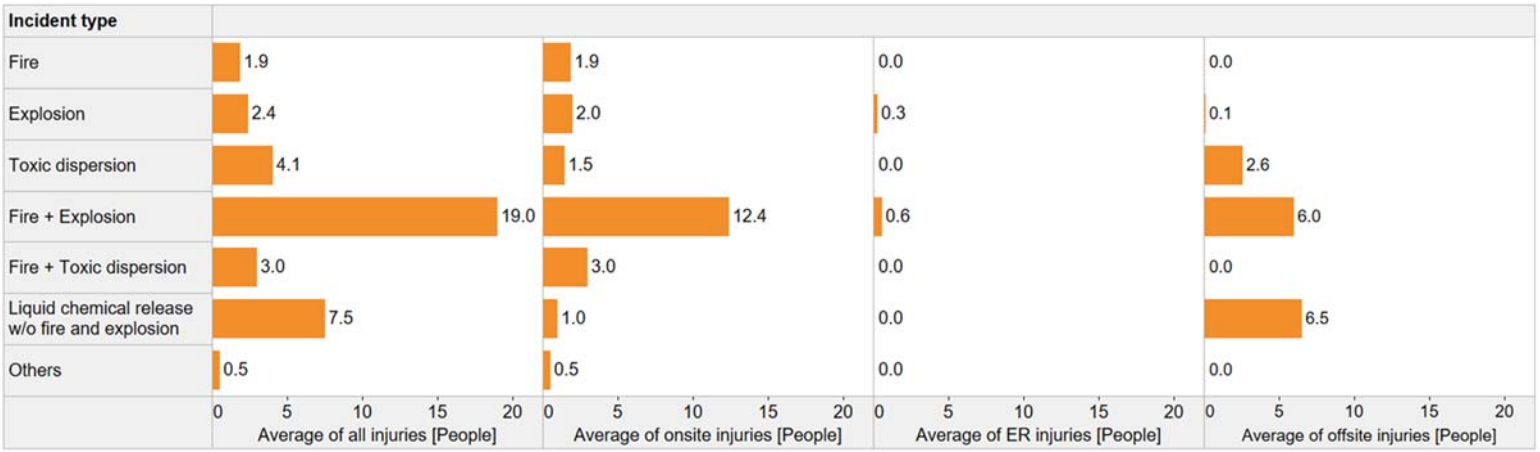
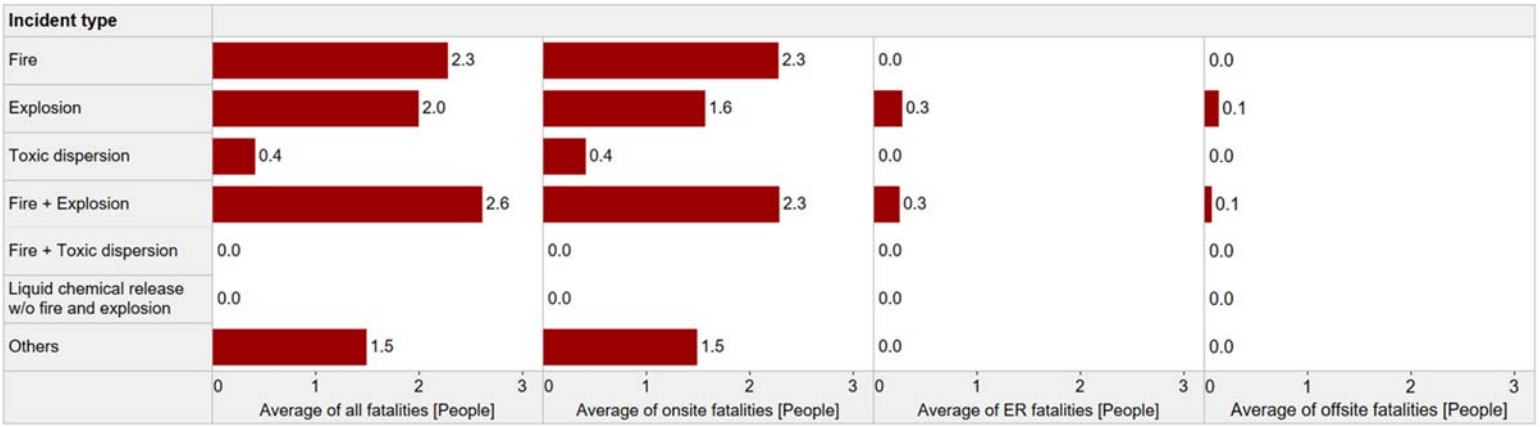


Figure 3.9 Scatter plot of chemical incident injuries versus fatalities from the 94 CSB incidents



(a) Average number of different injuries



(b) Average number of different fatalities

Figure 3.10 Average number of fatalities and injuries for each incident type

For the penalties of OSHA and EPA, the number of times being imposed and average fine amounts depending on incident type were examined (Table 3.10 and Figure 3.11). Table 3.10 shows the counts of OSHA- and EPA penalties that were broken down by incident types. OSHA imposed penalties in 92.5% (86 in 93) of the incidents, whereas the EPA imposed penalties in 34.4% (32 in 93). Since a public sector facility was not relevant to these regulatory penalties, one incident (#37 in Appendix 3A) was excluded from this analysis. The average of the 86 original OSHA penalties was $1.03 MM, and the average of the 32 EPA penalties was $177.90 MM. The two topmost EPA penalties (#80 and #38 in Appendix 3A) account for 93.5% and 3.1% of the total EPA penalties, respectively. Meanwhile, the two topmost OSHA original penalties (#38 and #58 in Appendix 3A) account for 57.6% and 19.0% of the total original OSHA penalties. Hence, both penalty distributions are extremely right-skewed by such very severe incidents. Therefore, analyzing both average penalty values of each incident type were performed after excluding three incidents (#38, #80, and #58 in Appendix 3A). Figure 3.11 shows the results of this analysis. It indicates that the average of EPA penalties was much higher than the average of OSHA penalties. This is because the EPA domain covers sites and surroundings, while the OSHA only covers the sites proper. Nevertheless, fire incidents had a tendency of comparatively low impact on surrounding, compared to the other incidents.

Table 3.10 Count of incident penalties imposed by OSHA and EPA

	Liquid chemical release without fire or explosion	Toxic dispersion	Fire	Explosion	Fire + Explosion	Fire + Toxic dispersion	Others	Grand Total
Count of OSHA penalties	2	16	6	7	51	2	2	**86**
Count of EPA penalties	2	7	1	1	20	1	0	**32**

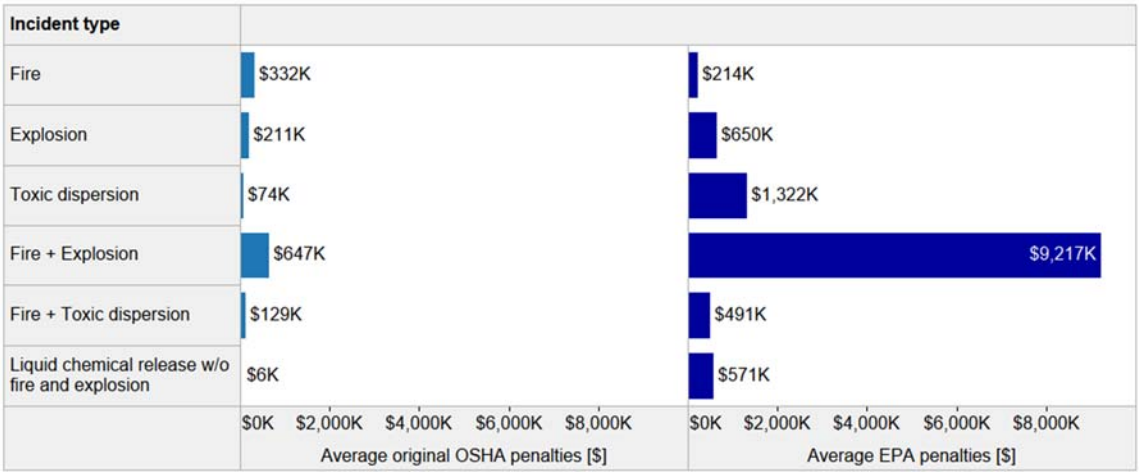


Figure 3.11 Average original OSHA and EPA penalties for each incident type[5]

[5] It excludes three incidents (Deep Horizon explosion, BP Texas City, and DuPont Corporation Toxic Chemical Releases) that account for the topmost penalties for either OSHA or EPA penalties.

3.4.5. Analysis of causal and consequence factors together

Causal and consequence factors can be utilized together to grasp the collected data in detail. This subsection, as an example, shows the data availability of low-level information associated with the process and chemical indicators (causal factors) per incident type (consequence factor). In particular, missing and observed data patterns would be worth examining to unfold the patterns of marked process conditions in the CSB reports per incident type. These data patterns also enable us to identify the level of the associated chemical characteristics and their properties' availability from publicly available databases.

Two representative incident types were used to investigate data availability patterns: (1) combustion incidents and (2) toxic dispersion incidents. For combustion, a total of 69 incidents (55 fire + explosion, 7 fire, 7 explosion incidents) in the CSB reports were selected. For toxic dispersion, a total of 21 incidents (19 toxic dispersion and 2 fire + toxic dispersion incidents) were selected; the 2 fire + toxic dispersion incidents were categorized as toxic dispersion incidents here because the CSB investigated them due to the severe impact from toxic dispersions.

Figure 3.12 shows the missing data patterns of 13 numeric chemical and process indicators of the 69 combustion incidents. The yellow color represents the missing data, and the blue color represents the observed data part. Figure 3.12(a) presents indicators in the sequence of high missing ratios in the dataset from left to right. Each value of LFL, flammability limit, and combustion heat was missing in more than 40% of the 69 combustion incidents. The CSB teams tended to pay less attention to chemical quantity

estimate than to other process indicators (incident temperature and incident pressure. Figure 3.12(b) shows the amount of missing and observed indicators in certain combinations of the indicators. The very bottom line of Figure 3.12(b) indicates that the numeric indicators at around 36% of the combustion incidents were fully observed. Also, it suggests that the data will be more available when the missing data can be estimated.

Similarly, Figure 3.13 shows the missing data patterns of 13 numerical indicators of 21 toxic dispersion incidents. Figure 3.13(a) displays the 13 indicators in the sequence of high missing ratios from left to right. The values of six combustion-related properties (flash point, autoignition temperature, flammability limit, LFL, UFL, and the heat of combustion) were not accessible in more than half of the incidents, while the toxic exposure guideline, PAC-3, was accessible for all the incident cases. That is, it obviously indicates that the chemicals involved in toxic dispersion incidents are less flammable but highly toxic. Furthermore, the CSB teams had a tendency to establish chemical quantities for toxic dispersion incidents rather than other process conditions (incident temperature and pressures).

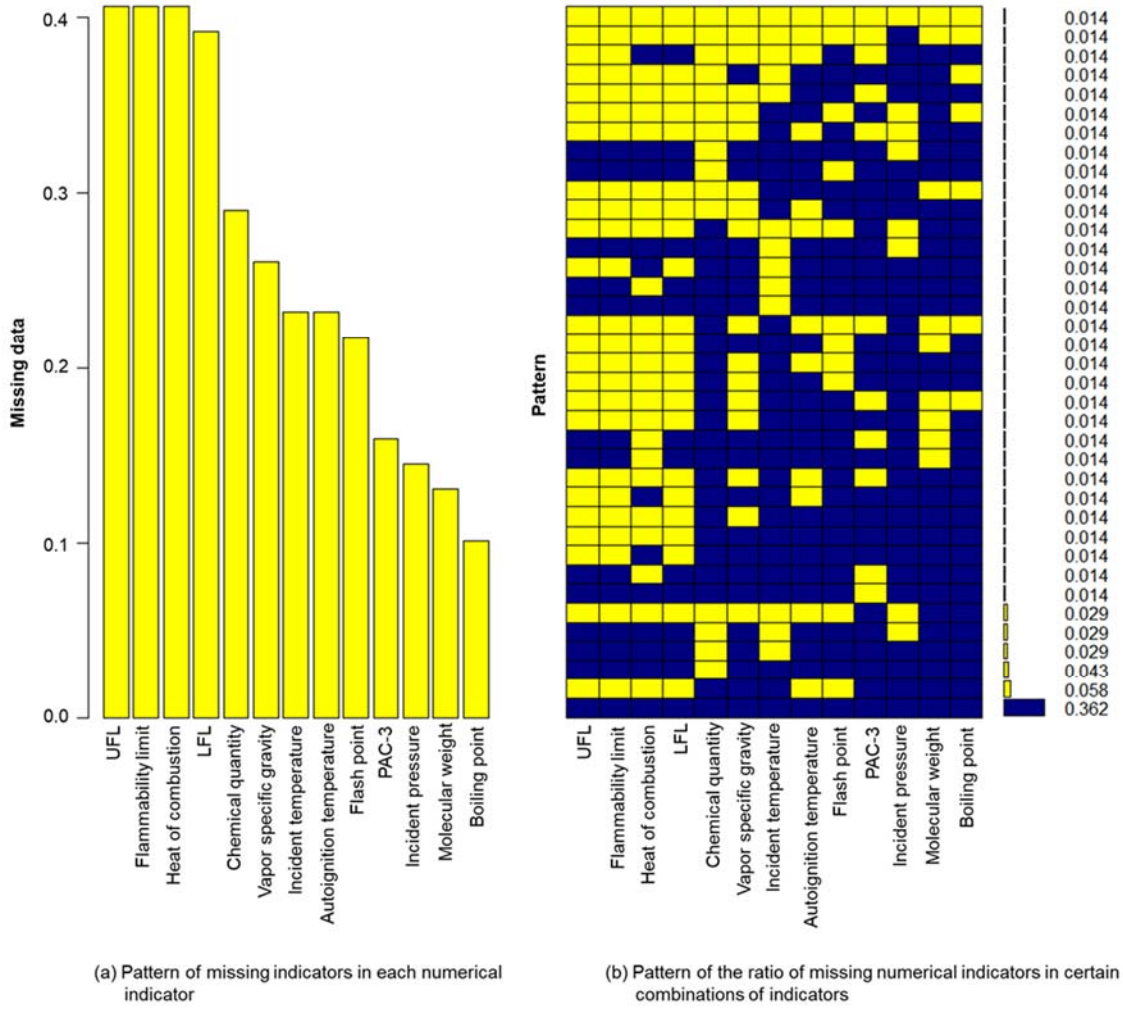


Figure 3.12 Data availability of numerical process and chemical indicators in 69 fire and/or explosion incidents

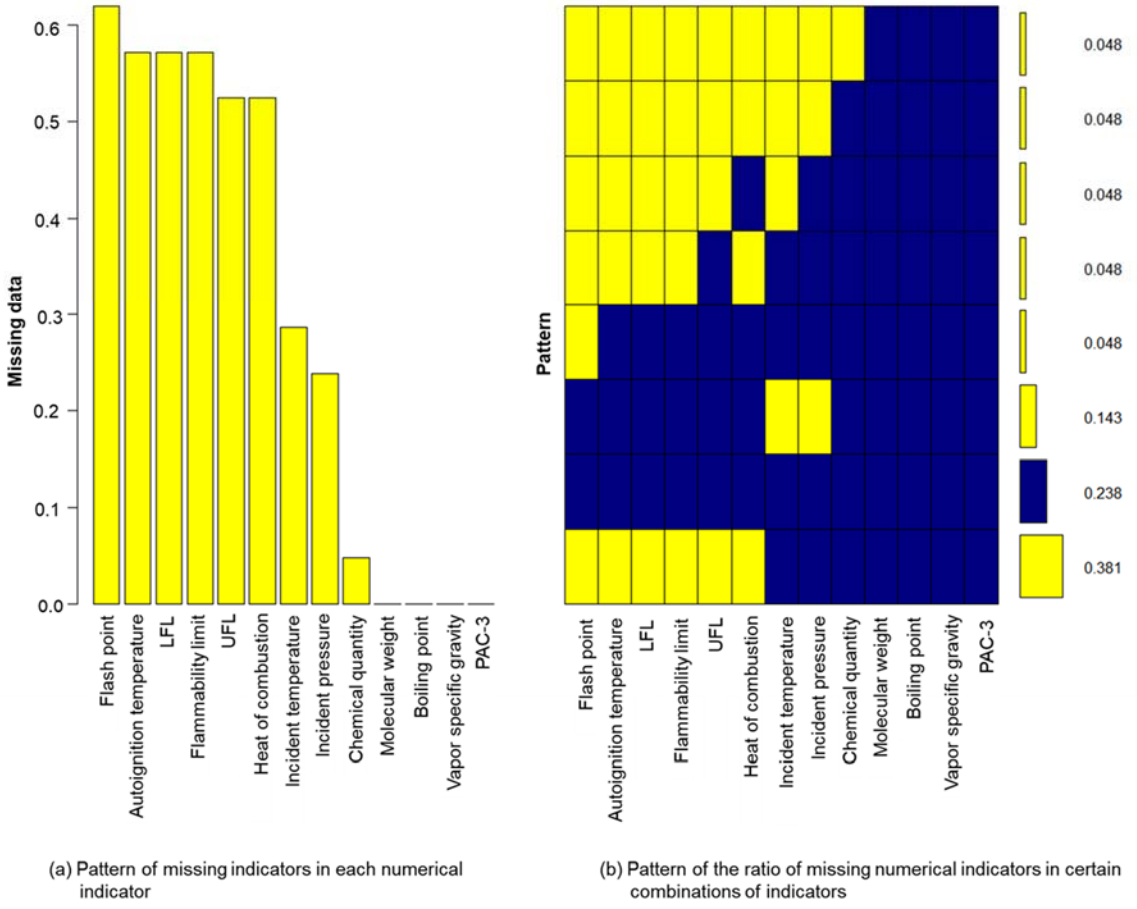


Figure 3.13 Data availability of numerical process and chemical indicators in 21 toxic dispersion incidents

Figures 3.12 and 3.13 indicate that the available information about chemical compounds' combustion-related properties was often limited. Although chemical compounds' hazardous characteristics are provided most often for combustible or toxic chemicals, such information was unavailable for incombustible chemicals. For this reason, in the case of chemical incidents due to incombustible chemicals, there is much more missing hazardous properties' data.

3.5. Conclusions and future work

This chapter's primary purpose was to investigate the potential of using the incidents investigated by CSB reports for enhancing an inherently safer design (ISD) strategy. Chemical indicators (chemical hazardous properties) and process indicators (process conditions) are deemed essential metric factors for performing ISD. Several indicators, in particular, have widely been adopted in the index-based inherent safety assessment tools. The selections of these indicators were often made based on theoretical approaches and less on heuristic ones. However, understanding the indicators' characteristics from previous chemical process incidents in heuristic approach would significantly help select key indicators to enhance the ISD strategy's reliability. Since the characteristics of chemical and process indicators are only associated with chemical incident causes, the knowledge about such indicators would be more powerful along with the characteristics of other risk assessment factors, such as incident scenarios and consequence, to properly grasp previous incidents. The posterior characteristics of causal factors (chemical and process indicators) can then be utilized in the ISD strategy. Up until now, there has not been a U.S. database that concisely represents such detailed information associated with previous chemical process incidents. Therefore, the currently available 87 CSB reports (94 incidents) were analyzed in this chapter because they provide much more detailed information than other U.S. incident databases. Depending on the quality of low-level information of the CSB reports, it may either reinforce a currently existing inherent safety assessment tool or propose a more reliable one.

The key points of this chapter are the following:

- Based on the purpose and background, we established a method to collect possible incident causal, scenario, or consequence factors in the CSB reports. A guideline was introduced to collect the data for the 27 sub-factors that consisted of 17 causal factors, 7 scenario factors, and 3 consequence factors. The causal factors were further divided into 12 chemical process indicators and 5 chemical indicators.
- The data availability of causal factors was analyzed after identifying chemicals involved in 90 incidents out of the 94 incidents. Based on the CAS numbers matched with the identified chemicals, the data of chemical indicators was collected. Two categorical chemical indicators, reactive rating and special notes of NFPA 704 standards, were fully examined. Then, the 8 numeric chemical indicators' available data was collected from external databases. The collected numeric chemical indicators showed that boiling point, molecular weight, and PAC-3 data were easily available, whereas autoignition temperature and flammability limit were less available. Meanwhile, the rest of causal factor data were examined merely according to the CSB reports' descriptions. In particular, the data of three numerical process indicators, the quantity of chemicals involved, incident temperature, and incident pressures, was established (or could be estimated) from 75% to 83% of the 94 incidents. The histograms of each numeric causal factor were displayed afterward.
- The analysis of 7 scenario factors were presented: successive incidents, relevant equipment type, domino effects, detonation/deflagration, operating modes, process units, and population densities. By integrating successive incidents information,

the detailed mechanisms of chemical incidents were investigated. While determining domino effects through the CSB reports' descriptions was problematic, determining the existence of detonation transition was performed as it appeared from the damage pattern to be rather obvious. In addition, most chemical incidents occurred during routine operation, shutdown, or maintenance operation modes. The residential population densities were displayed for each incident type.

- The analysis of consequence factors showed that the most frequently investigated incident types were fire + explosion and toxic dispersion incidents. In addition, it indicated that incident types were relevant to the number and types of casualties. Regarding two regulatory penalties, those by EPA and OSHA, it was observed that although the cost of OSHA penalties was generally much less than the EPA ones, the frequency of OSHA penalties was higher.
- This chapter examined the numeric causal factors' availability for the two incident types: combustion and toxic dispersion incidents. The result of this analysis revealed that the CSB tended to state the quantity of chemicals involved in toxic dispersion incidents rather than in case of combustion incidents. In contrast, the other process indicators, incident pressure and temperature, were remarked more often for the combustion incidents.
- The data extracted from the CSB reports and currently existing chemical databases can be utilized for ISD. However, the small amount of sample size may bias the analysis of chemical process incidents. Since the available data about the causal

factors were limited due to missing data, further attempts may be required to estimate currently unavailable values of indicators. In particular, the missing data of numerical chemical properties can be estimated by using currently existing group contribution models, Quantitative Structure–Property Relationship (QSPR) models, or new hands-on models. This would enable us to utilize more currently existing data of the CSB reports.

Although this chapter was initiated to study the potential of low-level information extracted from the CSB report to utilize in ISD strategy, such information can also be used in other safety risk analysis approaches. As Section 3.4.5 illustrated, the collected data (causal, scenario, and consequence factors) may be utilized to analyze various incident aspects. As the applications showed in Section 3.4, integrating various aspects (causal, scenario, and consequence factors) to describe chemical incidents enables us to analyze the factors that caused patterns of incidents. In particular, Industry 4.0 era can benefit from the significant potential of such a detailed analysis.

Even though the CSB reports' information was inspected multiple times, the results of this chapter contain uncertainty as it was necessary to estimate unstated parameters. Information of the CSB reports may be incorrectly marked as we manually collected it. Besides, process conditions may be inappropriately determined. For example, when an incident occurred at the atmospheric conditions, if there was no specific CSB report statement, the atmospheric conditions were consistently selected as 25℃ and 1atm. That is, these conditions may differ from the actual incident condition. In addition, we often

faced difficulty in determining the process conditions at the time of the incident; when the incident occurred, because of temporary external reasons, such as hot work.

Although the CSB investigators spelled out causes, scenarios, and consequences for each incident, the reports had two major limitations for providing sufficient lower-level information. First, the reports for each incident were written by various independent CSB investigator teams, which led to a lack of consistency in formats and units. Also, we cannot guarantee that the incident information is representative of all incidents happening in the U.S. because the CSB selected these incidents depending on incident severity or public attention. That is, the incidents in CSB reports tend to be catastrophic or high-profile incidents, rather than representing a typical average chemical incident.

Given the absence of a standardized incident database, perhaps it is much more challenging for researchers to collect data instead of analyzing the data. From this perspective, the format inconsistency among CSB reports suggests that standardized formats in such incident reports published by regulatory agencies could improve a general sharing of incident causes and recommendations. Traditionally, each incident was often analyzed for a standalone lesson. For this reason, the consistency among incident reports has not been the main issue under consideration. As information technology, however, has rapidly developed, the chemical process safety community can easily obtain broader insights among similar incidents beyond one single incident. For example, natural language processing can be utilized to automatically retrieve data of interest (e.g., commonalities and patterns) in a consistent format document, without tremendous human efforts. Machine learning or deep learning techniques can also be adopted to numerically

analyze the collected data. Using the resultant patterns, a safety assessment technique can be proposed; although various safety assessment tools that combine multiple chemical or process indicators have been proposed to measure safety levels, the indicators were selected in a rather subjective or theoretical way. Hence, the risk-based insights learned from a sufficient incident dataset can inversely be utilized to develop a better safety assessment, where safety is the complement of risk.

3.6. Appendix 3A 94 incidents investigated by the CSB

No.[1]	Incident	Incident Date	PSM covered	PHA performance
1*	Sierra Chemical Co. High Explosives Accident	1/7/1998	Yes	Yes
2*	Union Carbide Corp. Nitrogen Asphyxiation Incident	3/27/1998	Unknown	No
3*	Herrig Brothers Farm Propane Tank Explosion	4/9/1998	No	Yes
4	Morton International Inc. Runaway Chemical Reaction	4/8/1998	No	Yes
5	Sonat Exploration Co. Catastrophic Vessel Overpressurization	3/4/1998	No	No
6	Tosco Avon Refinery Petroleum Naphtha Fire	2/23/1999	Yes	No
7	Bethlehem Steel Corporation Gas Condensate Fire	2/2/2001	Unknown **	No
8	Concept Sciences Hydroxylamine Explosion	2/19/1999	Yes	Yes

No.[1]	Incident	Incident Date	PSM covered	PHA performance
9	BP Amoco Thermal Decomposition Incident	3/13/2001	No	Yes
10	Motiva Enterprises Sulfuric Acid Tank Explosion	7/17/2001	No	Yes
11	Georgia-Pacific Corp. Hydrogen Sulfide Poisoning	1/16/2002	No	No
12	Third Coast Industries Petroleum Products Facility Fire	5/1/2002	Unknown **	Unknown
13	DPC Enterprises Festus Chlorine Release	8/14/2002	Yes	Unknown
14	BLSR Operating Ltd. Vapor Cloud Fire	1/13/2003	No ***	Unknown
15	Environmental Enterprises Hydrogen Sulfide Release	12/11/2002	No	Unknown
16*	Kaltech Industries Waste Mixing Explosion	4/25/2002	Unknown	No
17	First Chemical Corp. Reactive Chemical Explosion	10/13/2002	No	Yes **
18	Catalyst Systems Inc. Reactive Chemical Explosion	1/2/2003	No	No
19	D.D. Williamson & Co. Catastrophic Vessel Failure	4/11/2003	No	No
20	Technic Inc. Ventilation System Explosion	2/7/2003	No	No
21	Isotec/Sigma Aldrich Nitric Oxide Explosion	9/21/2003	Yes	Yes
22	West Pharmaceutical Services Dust Explosion and Fire	1/29/2003	No	No

No.[1]	Incident	Incident Date	PSM covered	PHA performance
23	CTA Acoustics Dust Explosion and Fire	2/20/2020	No	Unknown **
24	Honeywell Chemical Incidents (Chlorine)	7/20/2003	Yes	Yes
25	Honeywell Chemical Incidents (Antimony pentachloride)	7/29/2003	No	No
26	Honeywell Chemical Incidents (Hydrogen fluoride)	8/13/2003	Yes	Yes **
27	Hayes Lemmerz Dust Explosions and Fire	10/29/2003	No	No
28	Giant Industries Refinery Explosions and Fire	4/8/2004	Yes	No **
29	Acetylene Service Company Gas Explosion	1/25/2005	Yes	Yes
30	Sterigenics Ethylene Oxide Explosion	8/19/2004	Yes	Yes
31	MFG Chemical Inc. Toxic Gas Release	4/12/2004	Yes	No
32	Marcus Oil and Chemical Tank Explosion	12/3/2004	No	No
33	Formosa Plastics Propylene Explosion	10/6/2005	U **	Yes
34*	Valero Refinery Asphyxiation Incident	11/5/2005	Unknown	No
35	DPC Enterprises Glendale Chlorine Release	11/17/2003	Yes	Yes
36	Formosa Plastics Vinyl Chloride Explosion	4/23/2004	Yes	Yes
37	Bethune Point Wastewater Plant Explosion [2]	1/11/2006	No	No

No.[1]	Incident	Incident Date	PSM covered	PHA performance
38	BP Texas City America Refinery Explosion	3/23/2005	Yes	Yes
39	Universal Form Clamp Co. Explosion and Fire	6/14/2006	Yes	No
40	Partridge Raleigh Oilfield Explosion and Fire	6/5/2006	No ***	No
41	Synthron Chemical Explosion and Vapor Cloud Explosion	1/31/2006	Yes	No
42	EQ Hazardous Waste Plant Explosions and Fire	10/5/2006	No	Unknown
43	CAI / Arnel Chemical Plant Explosion	11/22/2006	Yes	No
44	Barton Solvents Explosions and Fire	7/17/2007	No	Unknown
45	Valero Refinery Propane Fire	2/16/2007	Yes	Yes
46	Barton Solvents Flammable Liquid Explosion and Fire	10/29/2007	No ***	Unknown
47*	Little General Store Propane Explosion	1/30/2007	Unknown	No
48	Allied Terminals Fertilizer Tank Collapse	11/11/2008	No ***	Unknown **
49	T2 Laboratories Inc. Reactive Chemical Explosion	12/19/2007	No	No
50	Imperial Sugar Company Dust Explosion and Fire	2/7/2008	No	No **
51	INDSPEC Chemical Corporation Oleum Release	10/11/2008	Yes ***	Yes
52	CITGO Refinery Hydrofluoric Acid Release and Fire	7/19/2009	Yes	Yes

No.[1]	Incident	Incident Date	PSM covered	PHA performance
53	Kleen Energy Natural Gas Explosion	2/7/2010	No	No **
54	Veolia Environmental Services Flammable Vapor Explosion and Fire	5/4/2009	Yes	No
55	Xcel Energy Company Hydroelectric Tunnel Fire	10/2/2007	Unknown **	Yes
56	Bayer CropScience Pesticide Waste Tank Explosion	8/28/2008	Yes	Yes
57	Goodyear Heat Exchanger Rupture	6/11/2008	Yes	Unknown
58	DuPont Corporation Toxic Chemical Releases (Phosgene)	1/23/2010	Yes	Yes
59	DuPont Corporation Toxic Chemical Releases (Methyl chloride)	1/22/2010	Unknown	Unknown
60	DuPont Corporation Toxic Chemical Releases (Oleum)	1/23/2010	Unknown	Unknown
61*	DuPont Corporation Toxic Chemical Releases (Methanol)	9/21/2010	Unknown	Unknown
62*	DuPont Corporation Toxic Chemical Releases (Monomethylamine)	12/3/2010	Unknown	Unknown
63*	Hoeganaes Corporation Fatal Flash Fires (Iron dust)	1/31/2011	No	Unknown
64*	Hoeganaes Corporation Fatal Flash Fires (Iron dust)	3/29/2011	No	Unknown
65	Hoeganaes Corporation Fatal	5/27/2011	Yes **	Unknown

No.[1]	Incident	Incident Date	PSM covered	PHA performance
	Flash Fires (Hydrogen)			
66	E. I. DuPont De Nemours Co. Fatal Hot Work Explosion	11/9/2010	Yes	Yes
67*	Donaldson Enterprises, Inc. Fatal Fireworks Disassembly Explosion and Fire	04/08/2011	No	No
68	Carbide Industries Fire and Explosion	3/21/2011	No	Unknown
69	NDK Crystal Inc. High-pressure vessel rupture	12/7/2009	No	Unknown
70	Silver Eagle Refinery and Catastrophic Pipe Explosion	11/4/2009	Unknown **	Unknown
71*	Silver Eagle Refinery Flash Fire and Explosion	1/12/2009	Unknown	Unknown
72	Tesoro Refinery Fatal Explosion and Fire	4/2/2010	Yes	Yes
73	AL Solutions Dust Explosion	12/9/2010	No	Yes **
74	US Ink Fire	10/9/2012	Unknown	No **
75	Millard Refrigerated Services Ammonia Release	8/23/2010	Yes **	Unknown
76*	Chevron Refinery Fire	8/6/2012	Yes [3]	Yes
77	Horsehead Holding Company Fatal Explosion and Fire	7/22/2010	Unknown **	Unknown
78	Caribbean Petroleum Refining Tank Explosion and Fire	10/23/2009	No	No
79*	West Fertilizer Explosion and Fire	4/17/2013	No	No
80*	Macondo Blowout and Explosion	4/20/2010	No	Yes

No.[1]	Incident	Incident Date	PSM covered	PHA performance
81*	Tesoro Martinez Sulfuric Acid Spill	2/12/2014	Yes	Yes
82*	Tesoro Martinez Sulfuric Acid Spill	3/10/2014	Yes	No
83*	Williams Olefins Plant Explosion and Fire	6/13/2013	Yes	Yes
84*	AirGas Facility Fatal Explosion	8/28/2016	No	No
85*	ExxonMobil Refinery Explosion	2/18/2015	Yes [3]	Yes
86*	Freedom Industries Chemical Release	1/9/2014	No	No
87*	ExxonMobil Refinery Chemical Release and Fire	11/22/2016	Yes	Yes
88*	MGPI Processing, Inc. Toxic Chemical Release	10/20/2016	No	Yes
89	Packaging Corporation of America Hot Work Explosion	2/8/2017	No	No
90*	Arkema Inc. Chemical Plant Fire	8/29/2017	Yes	Yes
91*	Enterprise Pascagoula Gas Plant Explosion and Fire	6/27/2016	Yes	Unknown
92*	Pryor Trust Fatal Gas Well Blowout and Fire	1/22/2018	No	No
93*	DuPont La Porte Facility Toxic Chemical Release	11/15/2014	Yes	Yes
94*	Midland Resource Recovery Explosion	5/24/2017	No	No

* Incidents newly added compared to the analysis of Baybutt (2016)
** Discrepant outcomes compared to the analysis of Baybutt (2016)
*** There were specific notifications for the PSM standard coverages in the CSB reports, but it was determined based on the OSHA guideline (Occupational Safety and Health Administration OSHA, 2020)

[1] Chronological order based on publishing incident reports
[2] As a public sector, the incident site was not covered by the OSHA standard (OSHA, 2020b).
[3] California's process safety management (PSM) regulations

3.7. Appendix 3B Used 10 equipment categories with their subcategories (CCPS, 1998)

Equipment types	Sub-categories of each equipment type
Vessel	• In-process vessel (surge drums, flash drums, separators, accumulators, etc.) • Pressurized tank (spheres, bullets) • Atmospheric, fixed rook storage tank (cone/done roof) • Atmospheric, floating roof storage tanks
Reactor	• Batch reactors • Semi-batch reactors • Continuous flow stirred tank reactors (CSTR) • Plug flow tubular reactor (PFR) • Packed-bed reactor (continuous) • Packed-tube reactors (continuous) • Fluid-bed reactors
Mass transfer equipment	• Absorption • Adsorption • Extraction • Distillation • Scrubbing • Stripping • Washing
Heat transfer equipment	• Shell and tube exchange • Air-cooled exchange • Direct contact changers • Other types including helical, spiral, plate and frame, and carbon block exchange
Dryer	• Spray dryers • Tray dryers • Fluid bed dryers • Conveying (flash, mechanical, and pneumatic) dryers • Rotary dryers
Fluid transfer equipment	• Blowers • Pumps • Compressor
Solid-fluid separators	• Centrifuges • Filters • Dust collectors • Cyclones • Electrostatic precipitators

Equipment types	Sub-categories of each equipment type
Solids handling and process equipment	• Mechanical conveyors • Pneumatic conveying system • Comminution equipment (mills, grinder, crushers) • Sieving (screening) equipment • Power blenders (mixers) • Solid feeders (rotary valves, screw feeders, etc.) • Spray granulators and coaters
Fired equipment	• Process furnaces • Boilers • Thermal incinerators • Catalytic incinerators
Piping and piping components	• Piping (metallic, nonmetallic, lined, jacketed, double-walled) • Components (flanges, expansion joints, gaskets, bolts, etc.)

3.8. References of Chapter 3

Amyotte, P.R., MacDonald, D.K., Khan, F.I., 2011. An analysis of CSB investigation reports concerning the hierarchy of controls. Process Saf. Prog. 30 (3), 261–265.

Antonioni, G., Spadoni, G., Cozzani, V., 2009. Application of domino effect quantitative risk assessment to an extended industrial area. J. Loss Prev. Process. Ind. 22, 614–624.

Baybutt, P., 2016. Insights into process safety incidents from an analysis of CSB investigations. J. Loss Prev. Process. Ind. 43, 537–548.

CCPS, 1998. Guidelines for Design Solutions for Process Equipment Failures. American Institute of Chemical Engineers, New York.

CCPS, 2000. Guidelines for Chemical Process Quantitative Risk Analysis. American Institute of Chemical Engineers, New York.

Chamberlain, G., Oran, E.S., Pekalski, A., 2019. An analysis of severe vapour cloud explosions and detonations in the process industries. Chemical Engineering Transactions 77, 853–858.

Crowl, D.A., Louvar, J.F., 2001. Chemical Process Safety: Fundamentals with Applications. Pearson Education.

DIPPR, 2019. DIPPR Project 801 - Full Version. https://app.knovel.com/hotlink/toc/id:kpDIPPRPF7/dippr-project-801-full/dippr-project-801-full.

Handbook, D., 2016. Temporary Emergency Exposure Limits for Chemicals: Methods and Practice. https://www.standards.doe.gov/standards-documents/1000/1046-Bhdbk-2016/@@images/file.

Kaszniak, M., 2010. Oversights and omissions in process hazard analyses: lessons learned from CSB investigations. Process Saf. Prog. 29 (3), 264–269.

Khan, F.I., Abbasi, S., 1998. Models for domino effect analysis in chemical process industries. Process Saf. Prog. 17 (2), 107–123.

NFPA 704, 2017. Standard System for the Identification of the Hazards of Materials for Emergency Response. National Fire Protection Association.

NIST, 2020. National Institute of Standards and Technology Chemistry WebBook. https://webbook.nist.gov/chemistry.

Oran, E.S., Chamberlain, G., Pekalski, A., 2020. Mechanisms and occurrence of detonations in vapor cloud explosions. Prog. Energy Combust. Sci. 77, 100804. https://doi.org/10.1016/j.pecs.2019.100804.

OSHA, 1992. OSHA Standard 29 CFR 1910, Occupational Health and Safety Administration. U.S. Department of Labor. https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=9760.

OSHA, 2020a. Process safety management. In: OSHA 3132 2000 (Reprinted). Occupational Safety and Health Administration, U.S. Department of Labor.

OSHA, 2020b. Process safety management. 'State plans. Accessed Nov. 23, 2020. https ://www.osha.gov/stateplans.

Park, S., Xu, S., Rogers, W., Pasman, H., El-Halwagi, M.M., 2020. Incorporating inherent safety during the conceptual process design stage: a literature review. J. Loss Prev. Process. Ind. 63, 104040.

U.S. Department of Energy, 2020. Tables in PDF and Excel Format (Rev. 29A). https:// edms.energy.gov/pac/TeelDocs.

Yaws, C.L., 1999. Chemical Properties Handbook: McGraw-Hill.

4. FAST, EASY-TO-USE, MACHINE LEARNING-DEVELOPED MODELS OF PREDICTION OF FLASH POINT, HEAT OF COMBUSTION, AND LOWER AND UPPER FLAMMABILITY LIMITS FOR INHERENTLY SAFER DESIGN [6]

This chapter proposes easy-to-apply machine learning-developed models, which predict four flammability properties of pure organic compounds: the flash point, heat of combustion, lower flammability limit (LFL), and upper flammability limit (UFL). These flammability properties pose a strong impact on the inherently safer design of industrial processes. Similar to quantitative structure-property relationship (QSPR) or group contribution models, machine learning algorithms are utilized in this study to establish predictive models. Compared to previous models, this chapter uses readily available variables (i.e., the numbers of atomic elements, molecular weights, and normal boiling points) as default variables without the analysis of detailed molecular structures or the in-depth knowledge of chemistry. This chapter consists of two steps: Step (1) building multiple linear regression (MLR) models by incorporating default input variables and Step (2) building MLR models by incorporating interaction and transformed variables to improve the predictions from the models in Step 1. In Step 1, an optimal subset of predictors is identified by con- structing an MLR model via the sequential floating backward selection (SFBS) algorithm. As a result of Step 1, the two constructed models

[6] Reprinted with permission from "Fast, easy-to-use, machine learning-developed models of prediction of flash point, heat of combustion, and lower and upper flammability limits for inherently safer design" by Park, S., Bailey, J. P., Pasman, H. J., Wang, Q., & El-Halwagi, M. M. (2021) Computers & Chemical Engineering, 155, 107524, Copyright [2021] by Elsevier Ltd. All rights reserved.

of the flash point and heat of combustion are found to be adequate, while the predictability of LFL and UFL are insufficient. In Step 2, MLR models incorporating nonlinearity and interaction terms are constructed via the sequential floating forward selection (SFFS) algorithm by selecting the optimal subset of default variables. The results show that all the constructed models in Step 2 are adequate as predictive models; the mean absolute errors (MAEs) of the flash point, heat of combustion, LFL, and UFL are 7.31 (5.67 via the SFBS) [K], 60.6 (61.87 via the SFBS) [kJ/mol], 0.21 (0.19 via the SFBS) [vol.%], and 2.44 (2.33 via the SFBS) [vol.%], respectively. Compared to previous models, the approved models in this study provide highly competitive performance with enhanced simplicity and interpretability.

4.1. Introduction

Combustion incidents are among the most commonly occur- ring incident types in the process industry. The general prerequisite condition for combustion incidents is that the released chemical must be flammable (CCPS, 1999). The released chemical may then cause a combustion incident when the surrounding temperature reaches or surpasses the chemical's autoignition or flash point temperature within its flammable range (also called explosive range) (Albahri, 2015). The minimum ignition energy (MIE) and the heat of combustion are additional flammability properties associated with the combustion incidents (Wang et al., 2017). Some flammability properties are often addressed together to grasp combustion incidents properly (Park et al., 2020). This was done within the context of choice making in inherently safer conceptual design, in which an easy-to-use

method for any flammable that would be sufficiently accurate, would support a designer in decision making on alternatives.

Since flammable chemicals such a fuels, solvents, and raw materials are indispensable in the process industry, finding sensible strategies for handling flammability has been an ongoing topic of investigation. Of several flammability properties, particular attributes are adopted as indicators of susceptibility to ignition in academic fields, safety standards, and regulatory authorities. For example, the U.S. National Fire Protection Association (NFPA) and the U.S. Occupational Safety and Health Administration (OSHA) have employed the flash point as a criterion for enforcing the safety handling of flammable chemicals. The analysis of Park et al. (2020), which is a review paper for safety indices, shows that many studies have often adopted four properties—flash point, heat of combustion, and flammable range —as chemical predictor variables in inherent safety indices.

Table 4.1 List of selected five chemical properties associated with the hazards of chemicals

Hazardous chemical property	Definition
Flash point	The lowest temperature at which an ignition source causes the vapors of a substance to ignite in atmospheric pressure (ASTM, 2016).
Heat of combustion (Also called enthalpy of combustion)	The net increase in heat content when a substance undergoes complete oxidation in its standard state at ambient conditions (Yaws, 1999).

Hazardous chemical property	Definition
Lower flammability limit (LFL)	The minimum concentration of a combustible substance that is capable of propagating a flame in a homogeneous mixture of the combustible and a gaseous oxidizer under the specified conditions (ASTM, 2004).
Upper flammability limit (UFL)	The maximum concentration of a combustible substance that is capable of propagating a flame in a homogenous mixture of the combustible and a gaseous oxidizer under the specified condition of test (ASTM, 2004).
Flammable range (Also, called explosive range)	The difference between UFL and LFL (i.e., UFL-LFL vol.%]).

Flammability properties for specific chemicals are not always available from the experimental data. This is because numerous chemicals are newly synthesized in the chemical process industry and carrying out experiments for toxic, explosive, or radioactive compounds is extremely difficult and costly (High & Danner, 1987; Pan et al., 2010).

Alternatively, studies have been geared toward predicting chemical flammability properties in a practical manner. These studies can be categorized into three models (Gharagheizi et al., 2008): (1) a physical property model, (2) a group contribution model, and (3) a quantitative structure-property relationship (QSPR) model, as illustrated in Table 4.2. Predictive models were actively proposed in physical property model approaches between the late 1950s and the mid-1990s. Then, more complicated and accurate approaches, which were group contribution models and QSPR models, were actively proposed from 2000 onward.

Table 4.2 Comparison of the three model types of predicting flammability properties

	Physical property model	**Group contribution model**	**QSPR model**
Predictor variables	• Other physical properties: boiling points, vapor pressure, specific density, stoichiometric concentration of fuel in air, different flammability properties with a target flammability property, etc.	• Molecular structures	• Molecular descriptors (by considering molecular structures)
Performance of feature selection technique	• No	• No	• Yes, most cases
Advantages	• Straightforward to perform • Easy to interpret the proposed models	• Estimation without considerable computational resources like QSPR models	• Flexible models • High accuracy via numerous input parameters, molecular descriptors
Disadvantage	• High dependency from other physical properties • Limited applicability scope of chemical compounds	• Required knowledge about various molecular structures. • Difficulty in decomposing structures for several	• Multiple steps to obtain molecular descriptor values using licensed commercial software programs

	Physical property model	Group contribution model	QSPR model
	• Chance of underfitting because of few input parameters	concurrent functional groups in a considered molecule • Numerous predictor variables (Tens of predictors) • Limited applicability scope of chemical compounds	• Often required the skill of coding to perform machine learning algorithms • Hard to intuitively interpret the proposed models.
Exemplary studies (flash point)	Mack et al. (1923), Butler et al. (1956), Affens (1966), Affens and McLaren (1972), Ishiuchi (1976), Fujii and Hermann (1982), Kanury (1983), Patil (1988), Satyanarayana and Kakati (1991), Hshieh (1997), Catoire and Naudet (2004), Jiao, Ji, et al. (2020)	Suzuki et al. (1991), Albahri (2003), Pan et al. (2007), Gharagheizi et al. (2008), J. Rowley et al. (2010), Lazzús (2010), Carroll et al. (2010a), Carroll et al. (2010b), Carroll, Godinho, et al. (2011), Carroll, Lin, et al. (2011), J. R. Rowley et al. (2011), Keshavarz and Ghanbarzadeh (2011), Jia et al. (2012), Mathieu and Alaime (2014), Albahri (2015), Frutiger et al. (2016), Keshavarz et al. (2016), Alibakshi (2018)	Tetteh et al. (1999), Katritzky et al. (2001), Catoire and Naudet (2004), Katritzky et al. (2007), Gharagheizi and Alamdari (2008), Patel et al. (2009), Pan et al. (2010), Bagheri, Bagheri, et al. (2012), Bagheri, Borhani, et al. (2012), Torabian and Sobati (2019)

	Physical property model	**Group contribution model**	**QSPR model**
Exemplary studies (heat of combustion)	Hshieh (1999), Britton (2002),	-	Duchowicz et al. (2007), Gharagheizi (2008b),
Exemplary studies (LFL)	Suzuki (1994), Suzuki and Ishida (1995), Hshieh (1999), Britton (2002), Jiao, Ji, et al. (2020)	Seaton (1991), Frutiger et al. (2016),	Gharagheizi (2008a), Pan et al. (2010), Bagheri, Rajabi, et al. (2012), Chen et al. (2017), Wang et al. (2018), Jiao, Yuan, et al. (2020)
Exemplary studies (UFL)	Suzuki and Ishida (1995), Jiao, Ji, et al. (2020)	Seaton (1991), Frutiger et al. (2016),	Pan et al. (2009b), Pan et al. (2010), Wang et al. (2019)

All of the three aforementioned approaches find an empirical correlation between a target flammability property and predictor variables, based on the assumption that structurally similar compounds may have similar flammable activities. Table 4.2 summarizes the characteristics of each approach to developing a predictive model. As its name implies, physical property models use other physical properties (e.g., boiling point or vapor pressure) as predictor variables to predict a desirable flammability property among structurally similar compounds. In a first step, group contribution and QSPR models initially describe molecular structures either by functional groups or 3D-descriptors with chemistry knowledge or specialized software programs to obtain default variables. In particular, QSPR models use molecular descriptors, which are encoded molecular structure characteristics, after identifying geometrically optimized molecular structures via specialized software programs[7] (Jiao et al., 2019). These descriptors numerically express various chemical structural properties such as constitutional, topological, geometrical, thermodynamic, quantum chemical, and charge-related characteristics. Since thousands of high dimensional descriptors are produced to describe numerous compounds' characteristics, more informative descriptors for the prediction of a target property are selected through multiple feature selection procedures. In the last step of all models, the selected variables are correlated to target properties by regression algorithms. Thus, the type of predictor variables are the main differences among the three

[7] Most QSPR models obtain descriptors via specialized software program. However, recently, graph neural networks (GNNs) have been attracting attention in the prediction of molecular properties without the need to manually select molecular descriptors, as the GNNs learn properties from a molecular graph representation in an end-to-end training (Coley et al., 2017).

different types of predictive models. Moreover, physical property and group contribution models are constructed using determined predictor variables without a feature selection procedure, whereas most QSPR models have a feature selection step to identify informative variables. Hence, in comparison to the other two, the predictors of QSPR models are systematically selected compared to the other models with a feature selection procedure.

To present the big picture of the published predictive models, this study divided the previous models into three groups based on predictor variables. However, there are often concurrent predictor variables among the three groups. For example, the numbers of specific atoms or boiling points, which are typical predictor variables in physical property models, are also often adopted as predictor variables in the other model types. For instance, the group contribution models proposed by Carroll et al. (2010a) and Godinho et al. (2012) use the number of carbon and boiling points along with selected group contribution terms to predict the flash point. The group contribution model proposed by Keshavarz and Ghanbarzadeh (2011) also counts the number of carbons and hydrogens of the functional groups to distinguish between saturated and unsaturated hydrocarbons. The QSPR model, which predicts flash points proposed by Katritzky et al. (2001), contains the normal boiling point as a predictor variable along with selected descriptors. Hence, since predictive models have been enhanced gradually from physical property models to group contribution or QSPR models, these predictor variables of physical property models have been retained in the other models. Even though the number of specific atoms and the boiling point have played important roles as predictors, to our best knowledge, there has

been no attempt to systematically check which atomic elements and boiling points are more informative to predict a target property, such as through a feature selection technique.

The predictive models have been enhanced by focusing on two primary points: (1) providing competitive predictability and (2) including more generic organic compounds. Physical property models and QSPR models have expanded the scope of organic compounds from hydrocarbons, which are the simplest organic compounds only containing carbon and hydrogen, to more generic organic compounds, which contain other atomic elements along with carbon and hydrogen. For example, initial predictive models (Butler et al.'s physical property model (1956) and Pan et al.'s QSPR model (2010)) are only limited to pure hydrocarbons. Then, for covering more generic organic compounds, physical property models have been extending more organic compounds by developing multiple models based on atomic elements or functional groups. Also, for better predictability, complex equation forms that are expressed by polynomial, logarithmic or nonlinear terms have been proposed in physical property models. Meanwhile, group contribution and QSPR models have been developed by means of the rapid development of information technology. Using various machine learning algorithms such as multiple linear regression (MLR), artificial neural network (ANN), support vector machine (SVM), k-nearest-neighbors (KNN), and random forest (RF), the predictability of these models has increased (Alibakshi, 2018; Jiao, Hu, et al., 2020; Katritzky et al., 2007; Keshavarz et al., 2016; Lazzús, 2010; Patel et al., 2009; Wang et al., 2019; Yuan et al., 2019). The promising results in the previous QSPR studies demonstrate the feasibility of predicting

flammability properties based on molecular structures. The recent development of QSPR models provides accurate and reliable predictions of target properties (Liu & Liu, 2010). Recently, several QSPR models were proposed to estimate the flammability property of mixture compounds beyond the pure organic compounds (Jiao, Yuan, et al., 2020; Torabian & Sobati, 2019; Wang et al., 2018; Wang et al., 2019).

Although the proposed QSPR models have resulted in high accuracy predictions, there are two main limitations when design and risk analysis practitioners attempt to use them. First, these models may be feasible to only a limited group of researchers. Since descriptors work as predictor variables in QSPR models, the information of descriptors is indispensable for establishing or utilizing a QSPR model (Borhani et al., 2016). To obtain the information, utilizing a couple of specialized chemical packages is essential. Thus, the generic first step of QSPR that prepares descriptors causes practitioners to struggle when predicting a target flammability property. In the process industry (or even other academic research groups), practitioners hardly access such computer programs and information. Even after obtaining the descriptors, there is another big challenge for practitioners for utilizing QSPR models. Often, published QSPR models account for a certain range of a target property or the machine learning algorithms creating a model via complicated black box model without a concrete equation form from an MLR algorithm. Therefore, to adopt a QSPR model, practitioners must be proficient in coding to perform the identical machine learning algorithms. Although a variety of predictive models have been proposed in academic fields, adopting these models is still difficult for practitioners due to the demand

for either in-depth chemistry or computer science knowledge in associated with the models.

The above was substantiated when we analyzed the chemical process incidents investigated by the US Chemical Safety and Hazard Investigation Board (CSB) (Park et al., 2021), in our quest for more straightforward judgment on whether a change in conceptual design improves inherent safety. In terms of the data availability of chemical properties associated with the 69 combustion incidents: more than 40% of the incident data was missing for the LFL, UFL, and heat of combustion, and more than 20% of the data was missing for the flash point. Estimating this missing flammability data could give a better understanding of the possible lack of the inherently safer design among previous incidents' characteristics. However, due to suitable predictors' unenviabilities, adopting various QSPR models for each flammability property seemed difficult. This difficulty of missing data then raised demand regarding practical predictive models for these flammability properties. As Pistikopoulos et al. (2021) emphasized the importance of data application within scientific chemical engineering, a more practical and simpler acquisition method for flammability properties data was considered, which led us to the current approach incorporating machine learning techniques.

In the remainder of this chapter, Section 4.2 provides the study's objective and contributions. Section 4.3 describes the data and methods used. The results and discussion are presented in Section 4.4. Finally, concluding remarks are addressed in Section 4.5.

4.2. Objective and contributions of this study

This study aims to propose practical, reliable predictive models for four flammability properties of pure organic compounds: flash points, heat of combustion, LFL, and UFL. In order to establish practical models, predictor variables were selected among easily accessible variables (i.e., numbers of atomic elements per molecule, molecular weight, and boiling point). Apart from the variables' accessibility, their intrinsic characteristics may be associated with molecular structures such as descriptors, which are encoded molecular structure properties in QSPR studies. Compared with the limited number of default predictor variables in this study, hundreds of descriptors are produced in QSPR studies to describe various chemical compounds' characteristics. Given that there are often high correlations among the descriptors, this study intuitively creates transformed variables from the default input variables and their interaction terms. This study assumes that the new variables might be more informative in predicting a target flammability property than the original default variables. In other words, in this study, more compounded or transformed relationships (e.g., polynomial or logarithm) are considered when a simple relationship is insufficient for accurately predicting flammability properties. These strategies, in turn, lead to the establishment of MLR models incorporating consideration of nonlinear and interaction terms. MLR yields a simple equation, so practitioners are easily able to apply the results. Of hundreds of input variables (e.g., 119 atom types, boiling points, molecular weight, their nonlinear forms, and their numerous interaction terms), more informative ones are selected as predictor

variables through feature selection algorithms. Finally, the accuracy of the proposed MLR model is evaluated through machine learning validation techniques.

Put simply, the novel contributions of this current study are:

1. To use easily accessible physical properties, the normal boiling point, molecular weight per mole, and *all* atomic elements as default input variables, unlike currently published predictive models.
2. To selected informative atomic elements in *a systematic manner via feature selections*.
3. To prepared new input variable transformed from the default input variables or from *interaction terms* between individual input variables.
4. To establish statistical MLR models, which are tractable models, *incorporating consideration of nonlinear and interaction predictors* by using the prepared transformed input variables.
5. To interpret *how likely primary atomic elements and their interactions* contribute to each flammability property's characteristics.

4.3. Materials and methods

A generic schematic in this study is depicted in Figure 4.1. This study consists of two steps: Step (1) building MLR models for the four flammability properties by incorporating default input variables and Step (2) building MLR models incorporating interaction and transformed predictor variables as needed to improve the predictions from the models in

Step 1. Specifically, nonlinear and interaction terms allow for more flexible, but not necessarily more interpretable, models. In both steps, the MLR can take the form of Equation 4.1:

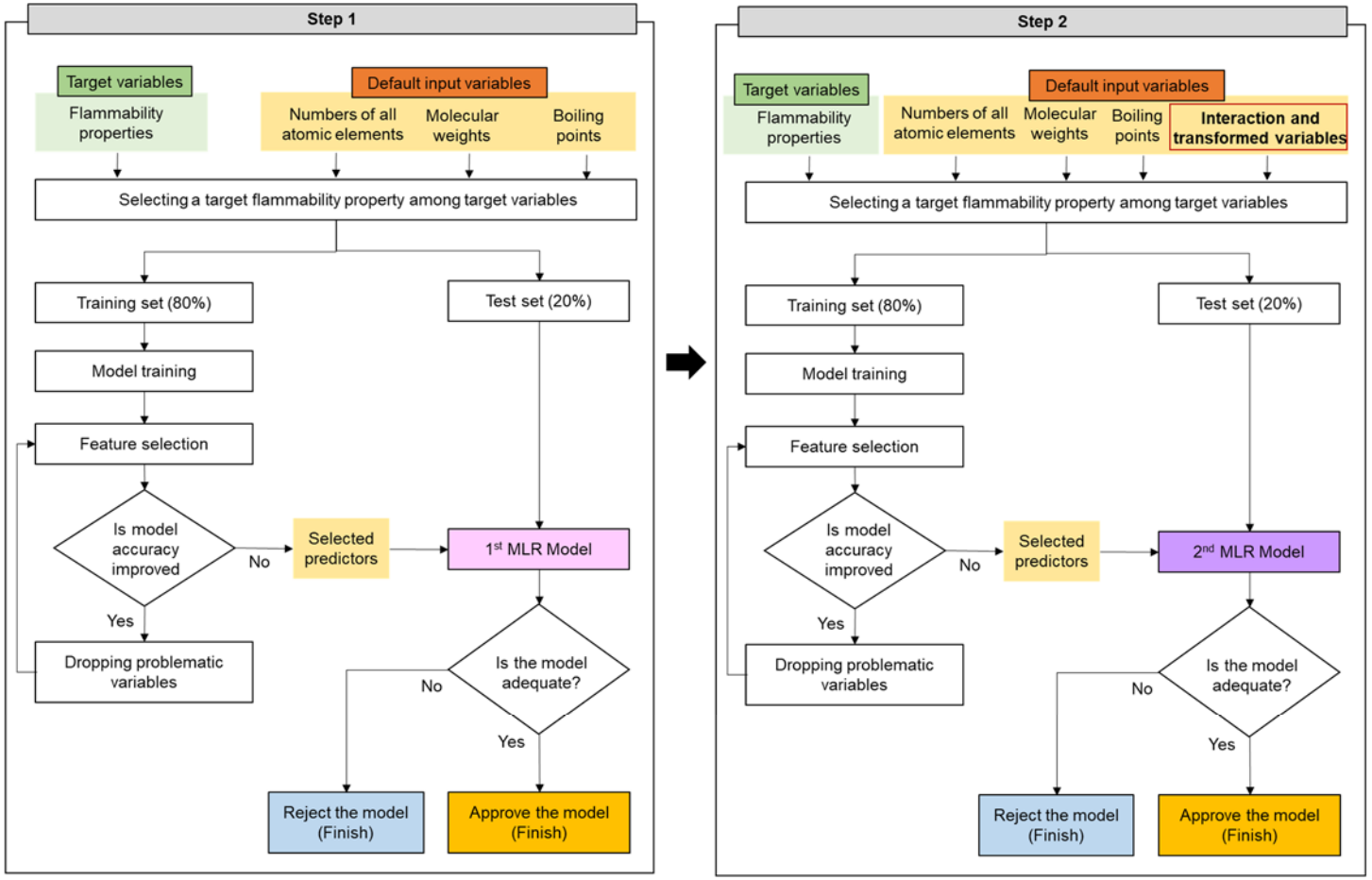
$$\boldsymbol{y \approx \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_i x_i + \cdots + \hat{\beta}_n x_n} \qquad \textbf{Equation 4.1}$$

where

y represents an observed flammability property of interest,

x_i is the i^{th} predictor variable, and

$\hat{\beta}_i$ is the i^{th} regression coefficient that correlates between y and x .



(a) Step 1: to establish an MLR model incorporating default input variables which are solely linear variables

(b) Step 2: to establish an MLR model incorporating consideration of nonlinear and interaction input variables along with the default input variables

Figure 4.1 Flow diagram of the proposed method

4.3.1. Comprehensive overview of dataset

The American Institute of Chemical Engineer's (AIChE's) Design Institute for Physical Properties Relationships (DIPPR, 2019) database was used in this study to establish accurate statistical models. The DIPPR data base contains temperature dependent and independent properties of more than two thousand industrially important compounds. As our study relates to fire and explosion propensity in view of industrial conceptual process design involving mainly organic compounds, we selected those from the data base. Dependent on the type of property, this set comprises between 1,700 and 1,850 compounds of various types with different functional groups. For this reason, many QSPR models have been constructed based on this data. The DIPPR data reference included all of the information required for this study, such as molecular formula, molecular weight, boiling point, flash point, heat of combustion (net standard enthalpy of combustion at 298 K), LFL, and UFL. All necessary information was compiled together based on the chemical abstracts service (CAS) numbers or substance names. Organic compounds were then selected. The combined available data of a target flammability property and boiling point is essential for establishing a predictive model (The organic compounds whose boiling points and each target property are exclusively available were analyzed). Consequently, a slightly inconsistent dataset was investigated for a different target flammability variable, depending on how it matched with the boiling point. The numbers of the dataset for boiling point, heat of combustion, LFL, and UFL were 1741, 1850, 1733, and 1711, respectively.

Depending on the hallmark represented by given data, different tactics should be explored. Therefore, understanding the dataset was the first task of this study to obtain an

idea of strength (measured by correlation), form (e.g., linear, quadratic, etc.), or direction (positive or negative) of any relationship that exists between default variables.

Figure 2 displays that the data distributions of normal boiling points and flash points were roughly normal (or Gaussian) distributions, i.e., a bell shape (Marshall & Jonker, 2010). However, the other variables' data followed non-normal distribution. For example, the number of molecular weight values, the number of carbon atoms per molecule, and data of LFL and UFL were positively skewed. In contrast, the data distribution of heat of combustion was moderately negatively skewed.

The pairwise correlations were analyzed to capture an insight into the potential of boiling points, molecular weight, and carbon elements as predictors (Figure 4.3). Since most variables in the study were not normally distributed as illustrated in Figure 4.2, the Spearman rank correlation that may be appropriately applied to asymmetric marginal distributions was adopted (Maindonald & Braun, 2006). Spearman rank correlations were calculated using Equation 4.2. This unitless coefficient is a standardized version of the covariance that is between -1 and 1(Rhys, 2020). The higher (in magnitude) the correlation, the more associated two variables are. For example, the values of -1 and 1 indicate that the correlation between a pair of variables is perfectly negative or positive, respectively. A correlation of zero implies that there is no relationship at all. For instance, the correlation value, 0.98, between boiling point and flash point indicates that they are highly related. In contrast, the correlation between the number of carbon atoms and heat of combustion is highly negative because their correlation value is -0.97. UFL and LFL are also highly negatively correlated with the number of carbons. Figure 4.3 indicates that

the four flammability properties (UFL, LFL, heat of combustion, and flash point) are highly correlated with any of default possible predictor variables (boiling point, the number of carbons, and molecular weight).

$$\boldsymbol{Spearman\ rank\ correlation = 1 - \frac{6\sum d_i^2}{n(n^2-1)}} \qquad \textbf{Equation 4.2}$$

where

n is the number of observations,

d_i is the difference between the ranks of corresponding variables, and

Σ denotes a sum of the multiple terms.

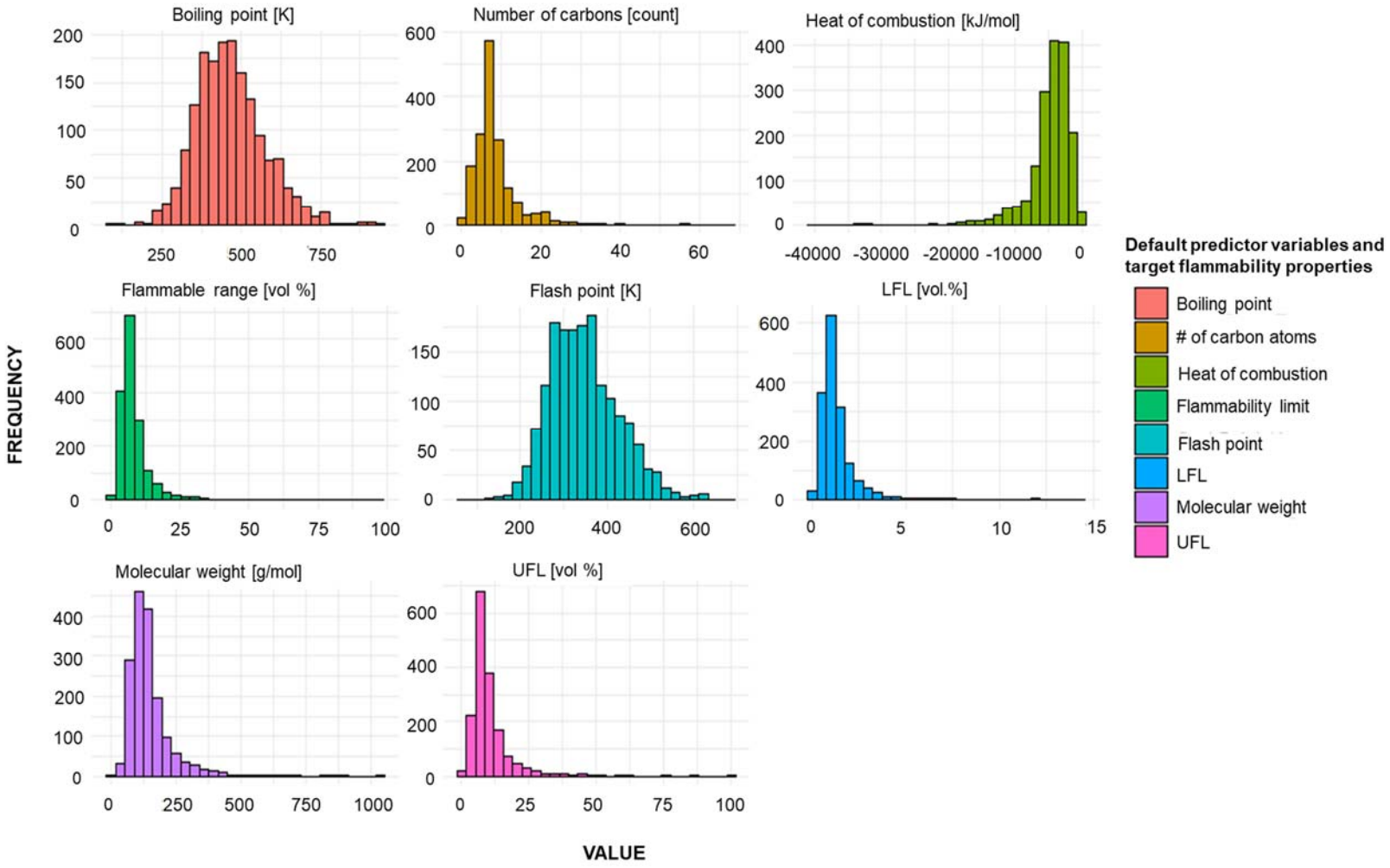


Figure 4.2 Eight histograms of variables in the collected data (*n*=1,738)

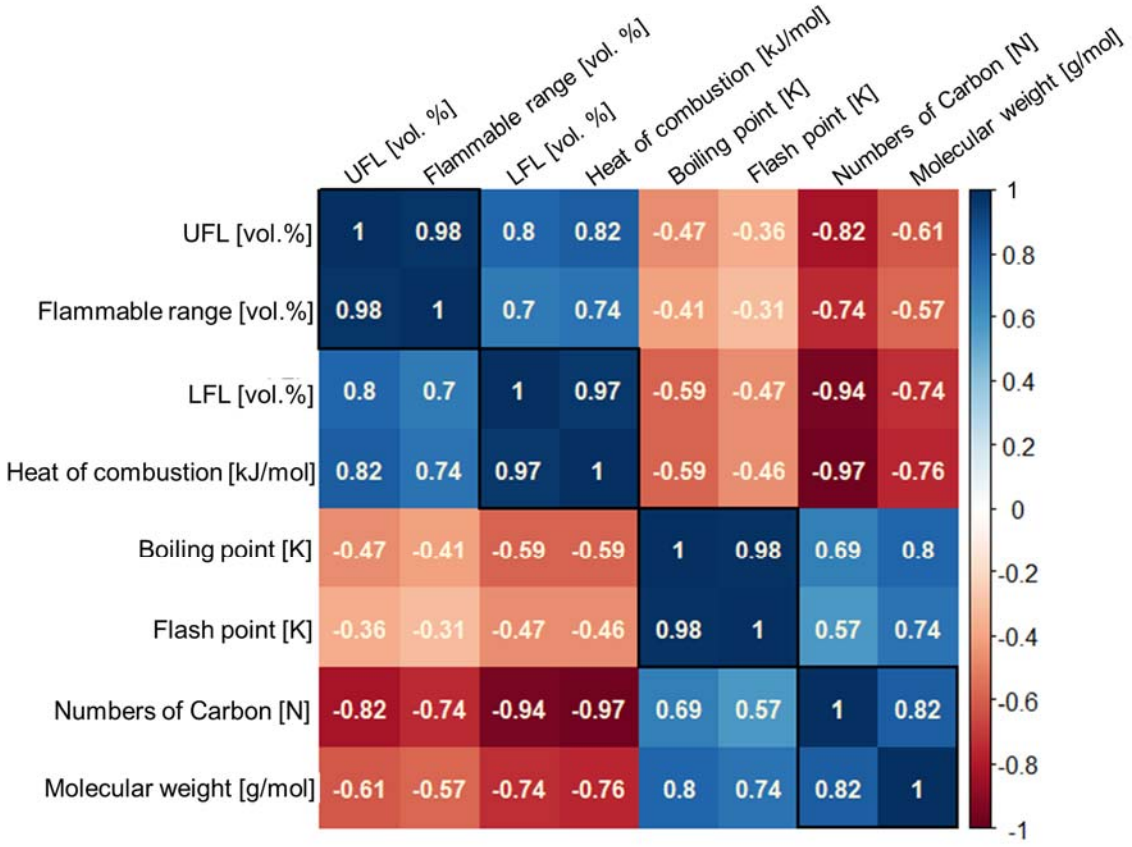


Figure 4.3 Spearman rank correlation plot of eight chemical properties in the entire data set (*n*=1,738)

4.3.2. Preprocessing of dataset

We prepared a dataset for input variables, which are possible predictor variables, based on the compiled data from the DIPPR (2019). In both Steps 1 and 2, the prepared input variables were (1) the number of all atomic elements per each molecule, (2) normal boiling point, and (3) molecular weight. First, individual atoms within every organic compound were counted. After enumerating all atom types of the periodic table of elements, the numbers of used atoms were counted in each molecular formula. Even

though various previous predictive models utilized the numbers of specific atoms (e.g., carbon, halogen groups) as predictors, this study thoroughly examined all 119 atoms as possible predictors to ensure that there was no missing information on the atomic elements. Second, the normal boiling point was selected. Since its intrinsic characteristics, such as volatility, which is correlated to flammability properties, boiling point has been exploited as a predictor variable in previous predictive models. In a similar manner, the normal boiling point was selected as an input variable in this study. Finally, molecular weight was chosen as another input variable. This choice was intuitively made through a couple of predictors of physical property models and QSPR models. For instance, instead of specific gravity in physical property models, or a gravity index in QSPR models, molecular weight may replace their properties to take into account the weight-related characteristics of organic compounds.

To refine the initial model, 122 input variables were additionally constructed by transforming or interacting between default input variables in Step 2. First, the transformed variables were created via quadratic, square root, or logarithmic forms of input variables. Such transformations have been performed to establish a better predictive model in other published methods to capture nonlinear relationships. Since generating transformed variables of all atomic elements is computationally expensive, 10 atomic elements (i.e., C, H, O, F, Cl, Br, I, N, Si, S), which were selected at least once as predictors through feature selections in Step 1, were transformed. Such feature selections were conducted within a training set (80%) of all data (Further explanation of the training set will be provided in Section 4.3.4). Then, their transformation forms were appended as

input variables, along with the transformation forms of molecular weight and boiling point. Afterward, using the transformed and default input variables, the interaction terms were further created. It should be noted that the interaction terms among predictor variables were firstly proposed in this study to account for their interaction effects (or synergy effects) in a macro scale. Since chemical-physical properties, including flammability properties, were often interpreted through either a group bond or a chemical bond, we intuitively devised interaction terms to consider such bonds. For example, the interaction terms between molecular weight and atomic elements may capture a group bond's behavior, which is an integration of two molecular groups. This assumption was made because molecular weight might refer to a bulk-related factor. Similarly, the interactions between different atoms may also account for chemical bonds between them. This assumption was also made because simply the interaction between nonidentical atoms might be associated with a polarity-related factor. For example, the interaction terms between the number of carbons and hydrogens may capture the characteristics of the alkyl group's bond types (e.g., single, double, and triple bonds). The ratio of electronegativities between different atoms may contribute to chemical potential due to the imbalanced charged atoms. For this reason, the fraction (ratio) terms were also constructed as an interaction term type. Lastly, the interaction terms between boiling point and the other input variables (atomic element and molecular weight) were also considered in case the former interaction terms were not enough to capture the characteristics of a group or chemical bonds. Finally, a total of 243 variables for each model were employed

as input variables in Step 2. All the prepared variables are provided in the Supplementary Materials.

In case the created model's performance was insufficient (e.g., $R^2 < 0.6$), some distinguishable outliers were discarded for additionally developing a more reliable model.

4.3.3. Selecting a target variable

In the sequence of flash point, heat of combustion, LFL, and UFL, predictive models were proposed. In Step 1, predictive models for each target variable were established only using the default input variables. Then, in Step 2, the transformed and interaction variables were added into input variables, along with the default predictor variables to improve the accuracy of the model.

4.3.4. Developing and validating a model

The organic compounds were randomly split into a training (80%) and a test set (20%). The training set and test set were utilized for model development and model validation, respectively. This idea of partitioning the data into a training and test set is a standard approach in machine learning to get more accurate predictions of a model's accuracy (Hastie et al., 2009).

The set of variables was trimmed in order to improve both model accuracy and interpretability by using feature selection algorithms. In this study, two feature selection algorithms were used: (1) sequential floating forward selection (SFFS) and (2) sequential floating backward selection (SFBS). These feature selection algorithms find out a small,

optimal subset of input variables to accurately predict a target property via MLR (Rhys, 2020). The SFFS starts from the empty set and iteratively adds the most significant variables. Significance is measured by estimating the mean square error (MSE) (Equation 4.6) of the result MLR using the selected variables via cross validation techniques. In contrast, the SFBS starts with all input variables and successively excludes an insignificant variable in each step. Afterward, a series of conditional inclusions are made from excluded variables if an improvement can be reinforced to the preceding sets (Pudil et al., 1994). These two feature selection algorithms were implemented through the R package, *mlr* (Bischl et al., 2020), based on the training set. Also, every permutation of a predictor variable in these algorithms was cross-validated (i.e.,10-fold cross-validation) to calculate internal performance within the training set. After performing the two algorithms, the established model that best predicted a target property was retained.

Multicollinearity was then examined for an established model using the subset of predictors extracted from the feature selection algorithms in Step 1. Multicollinearity refers to the situation collinearity exists among three or more variables, so it results in standard errors that are highly inflated (Fred Nwanganga, 2020; James et al., 2017). Hence, multicollinearity can generate a problem in linear regression models in terms of their reliability. This study used variance inflation factor (VIF) for the diagnosis of multicollinearity using Equation 4.3.

$$\boldsymbol{VIF}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}\right) = \frac{\mathbf{1}}{\mathbf{1}-\boldsymbol{R}^{\mathbf{2}}_{\boldsymbol{X_i}|\boldsymbol{X_{-i}}}} \qquad \textbf{Equation 4.3}$$

where $R^2_{X_i|X_{-i}}$ is the R^2 from a regression of X_i on the remaining explanatory variables.

The value of VIF is between 1 and infinity. The larger the value is, the larger multicollinearity exists. In this study, the VIF value of 10 was used as a threshold (Alin, 2010). When VIF was greater than 10, we dropped pragmatic variables that presented multicollinearity in the proposed model from the feature selection algorithms in Step 1.

Nevertheless, the high VIFs caused by the inclusion of square or interaction terms of other variables were ignored because the p-value for these variables is not affected by the multicollinearity (Allison, 2012). For this reason, the analysis of VIF in Step 2 with interaction terms was not performed.

Next, the predictive capability of the developed model was evaluated using three statistical metrics: the coefficient of determination, denoted as R^2, root mean square error (RMSE), and mean absolute error (MAE or also called average absolute error (AAE)). The equations of these metrics are expressed as follows:

$$\boldsymbol{R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad \textbf{Equation 4.4}$$

where

y_i is the observed value of a target variable, $\hat{y}$ is a predicted value,

$\bar{y}$ is the mean of the observed value in the dataset, and

n is the number of observed chemical compounds.

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad \textbf{Equation 4.5}$$

$$\boldsymbol{MAE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \qquad \textbf{Equation 4.6}$$

The internal predictive capability of the established MLR model was evaluated via leave-one-out cross-validation RMSE ($RMSE_{Loo}$) and R-squared (Q^2_{LOO}) on each training set, which is calculated using Equations 4.7 and 4.8.

$$\boldsymbol{RMSE_{Loo}} = \sqrt{\frac{1}{l}\sum_{i=1}^{l=training\ set}(y_i - \hat{y}_{ic})^2} \qquad \textbf{Equation 4.7}$$

$$\boldsymbol{Q^2_{Loo}} = 1 - \frac{\sum_{i=1}^{l=training\ set}(y_i - \hat{y}_{ic})^2}{\sum_{i=1}^{l=training\ set}(y_i - \bar{y}_{train})^2} \qquad \textbf{Equation 4.8}$$

where

y_i and $\bar{y}_{train}$ are the observed target variable and mean value in the training data set, respectively.

$\hat{y}_{ic}$ is the predictive variable of ith object estimated by a model without using the ith objective (Bagheri, Rajabi, et al., 2012).

In this validation, the model is repeatedly fit l times using $l - 1$ observed values in the training set.

The external predictive capability of the MLR models on the external test set was calculated by Q^2_{ext}, using Equation 4.9.

$$\boldsymbol{Q^2_{ext} = 1 - \frac{\sum_{i=1}^{m=test}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m=test}(y_i - \bar{y}_{train})^2}} \qquad \textbf{Equation 4.9}$$

$R^2, RMSE, and\ MAE$ were computed using both the training and test set, respectively. The $RMSE_{Loo}$ and Q^2_{Loo} of the training set and the Q^2_{ext} of the test set were employed to investigate the model's overfitting by comparing with the RMSE and R^2_{train} of the training set and R^2_{test} of the test set.

4.3.5. Assessing the adequacy of the created model

Finally, the adequacy of the created MLR model was assessed by checking the goodness of fit through (1) a scatter plot between the predicted vs. observed points of a target flammability property, and (2) a residual plot against the predicted points. In the scatter plot between predicted vs. observed points, the diagonal line represents the perfect correlation of observed and predicted flammability values. Therefore, we checked the vicinity to the diagonal line. Meanwhile, the residual refers to the different between the actual of y, y_i and the predicted value of $\hat{y}_i$(Sheather, 2009). In the residual plot, a discernible pattern is identified in case there is a systematic error. Therefore, when we recognized a discernible pattern in the plot, the model was determined as an inadequate

model to use. Based on these two observations, adequate models were selected in this study.

4.3.6. The computer specification used

This computation procedure was performed on a 1.90GHz Intel XPS 13 9370 (Dell Inc.) with 16GB Ram under window 10 Pro. The results from the SFFS feature selection were obtained within 30 minutes and those from the SFBS feature selection were obtained in around 3-7 hours.

4.4. Results and discussions

This section presents more reliable, applicable models among the various constructed models via the SFFS and SFBS feature selection algorithms. The models created through these two algorithms had relatively different strengths and weaknesses. In terms of accuracy, the performances of models constructed by the SFBS were superior through various predictors, except for heat of combustion. In terms of applicability or generalization, the SFFS algorithm-based models were superior because they had fewer predictors and showed a similar performance between training and test sets. In Step 1 (Figure 4.1), since less than ten predictor variables were selected on the basis of both feature selection algorithms, the SFBS-based MLR models incorporating linearity (SFBS-MLR-L) were determined by taking reliability into account. In contrast, many predictors were often selected through the SFBS in Step 2. Such models incorporating many predictors could make it complicated for process designers to quickly estimate a target

property. For example, the number of generated predictor variables of flash point, heat of combustion, LFL, and UFL prediction models via SFBS in Step 2 were 46, 34, 59, and 57, respectively. For this reason, more easily applicable models, SFFS-based MLR models incorporating nonlinearity and interaction (SFFS-MLR-NLI), were presented in detail in this section with the relevant interpretations. To check whether this method can achieve a more accurate model in this study, the SFBS-based MLR models incorporating nonlinearity and interaction (SFBS-MLR-NLI) of flash point, heat of combustion, LFL, and UFL are presented in Appendices 4A-4D, respectively.

4.4.1. Results of flash point models

The number of matched datasets among the flash point, normal boiling point, and molecular weight in the compiled dataset is 1,741.

First, by performing the SFBS-MLR-L procedure on the training set, the optimum subset of eight predictors was achieved in Step 1. The constructed model for predicting the flash point is presented as the following:

$$\boldsymbol{T_f\,[K] \approx f(the\ numbers\ of\ C, O, N, F, Cl, Si, S\ and\ T_b)}$$

$$\boldsymbol{= 2.7735 - 1.8443\,nC + 2.7454nO\ + 2.3241nN + 2.9889nF}$$

$$\boldsymbol{+ 6.2254nCl - 3.2484nSi - 7.9198nS + 0.7665\,T_b}$$

Equation 4.10

where nC, nO, nN, nF, nCl, nSi and nS are the numbers of carbon, oxygen, nitrogen, fluorine, chlorine, silicon, and sulfur atoms, respectively,

T_f and T_b is the flash point [K] and normal boiling point [K] of an organic compound of interest.

Given that the boiling point values (i.e., 81.7 - 1,089K) are much higher than the rest of the predictors, the most informative predictor would be the normal boiling point. Hence, the flash point increases linearly, particularly with the normal boiling point of pure organic compounds. The selected seven atoms then assist in improving the fit of the model. Of the seven atoms, three atoms contribute to reducing flash point values having negative coefficients. Conversely, the remaining four atoms lead to increased flash point values with positive coefficients. Although the magnitude of every coefficient is not in proportion to its corresponding electronegativities, we could interpret that the atoms having weak electronegativities (i.e., carbon, silicon, and sulfur) facilitate the volatility of organic compounds, so they function to decrease the flash point values. In stark contrast, the atom elements with stronger electronegativities (i.e., oxygen, nitrogen, fluorine, and chlorine) than the former three elements prevent organic compounds from volatilizing by leading to higher flash point values. However, note that this interpretation is merely valid when the boiling point exists in the model. These selected predictors' coefficients are statistically significant (p-value < 0.05), and multicollinearity is less likely to exist (VIT<10), as illustrated in Table 4.3.

Table 4.3 Selected predictors and their coefficients of the flash point SFBS-MLR-L model (Equation 4.10)

No.	Variables	Coefficient	Std. error	p-value	VIF
0	Intercept	2.7735E+00	2.5210E+00	0.2715	-
1	nC	-1.8443E+00	1.1823E-01	< 2E-16	3.002
2	nCl	6.2254E+00	8.6501E-01	1.01E-12	1.090
3	nF	2.9889E+00	7.1382E-01	3.00E-05	1.054
4	nN	2.3241E+00	9.0694E-01	0.0105	1.242
5	nO	2.7454E+00	3.4598E-01	4.31E-15	1.618
6	nS	-7.9198E+00	1.7457E+00	6.21E-06	1.064
7	nSi	-3.2484E+00	7.2282E-01	7.57E-06	1.137
8	T_b	7.6654E-01	7.4290E-03	< 2E-16	3.615

Second, by performing the SFFS-MLR-NLI procedure in the training set, the optimum subset of fifteen predictors was achieved in Step 2. The constructed model for predicting the flash point is presented as the following:

$$\boldsymbol{T_f\ [K]}$$

$$\boldsymbol{\approx f(the\ numbers\ of\ C, H, O, N, Cl, Br, I, Si, S, T_b, MW, interaction\ terms)}$$

$$\boldsymbol{= 259.0 - 7.532nCl - 17.04nSi + 49.33\sqrt{T_b} - 84.17(log_{10}nC)}$$

$$\boldsymbol{+ 39.47\log_{10}MW - 366.7\log_{10}(T_b) - 4.569\exp(nBr) - 27.75(nI)^2}$$

$$\boldsymbol{+ T_b(0.024nSi - 0.1217nS) + 5.733(nC \times nS)}$$

$$\boldsymbol{+ nCl(2.083nC - 1.384nH + 5.421nO - 11.17nN)}$$

Equation 4.11

where $nC, nH, nO, nN, nl, nBr, nI, nSi$ and nS are the numbers of carbon, hydrogen, oxygen, nitrogen, chlorine, bromine, iodine, silicon, and sulfur atoms, respectively. MW is the molecular weight [g/mol] of an organic compound of interest.

The selected predictors' coefficients are statistically significant (p-value < 0.05), as illustrated in Table 4.4. This model was constructed by combining two linear, six nonlinearity, and eight interaction terms of the eight predictors. Compared to Equation 4.10 in Step 1, Equation 4.11 newly contains two halogen atoms — bromine, and iodine — as predictors by excluding fluorine. The contribution of chlorine atoms in this SFFS-MNR-NLI model is expressed in various terms, which are its standalone terms and multiple interaction terms with carbon, hydrogen, oxygen, and nitrogen. Moreover, the normal boiling point is used to accurately fit with the model in the two transformed terms and interaction terms with silicon and sulfur atoms.

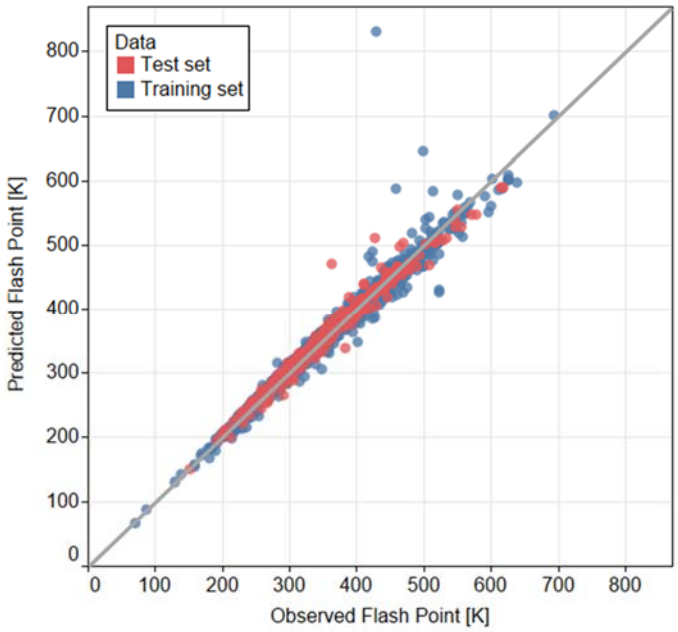
Table 4.4 Selected predictors and their coefficients of the flash point SFFS-MLR-NLI model (Equation 4.11)

No	Variables	Coefficient	Std. error	p-value
0	Intercept	2.5900E+02	5.44E+01	2.16E-06
1	nCl	-7.5320E+00	1.91E+00	8.21E-05
2	nSi	-1.7040E+01	4.49E+00	0.000156
3	$\sqrt{T_b}$	4.9330E+01	1.28E+00	< 2E-16
4	$\log_{10} C$	-8.4170E+01	3.25E+00	< 2E-16
5	$\log_{10} MW$	3.9470E+01	5.01E+00	6.98E-15
6	$\log_{10}(T_b)$	-3.6670E+02	2.99E+01	< 2E-16
7	$exp\ (nBr)$	-4.5690E+00	9.62E-01	2.23E-06
8	$(nI)^2$	-2.7750E+01	7.08E+00	9.35E-05
9	$T_b \times nSi$	2.4040E-02	7.68E-03	0.001796
10	$T_b \times nS$	-1.2170E-01	8.62E-03	< 2E-16
11	$nC \times nCl$	2.0830E+00	3.64E-01	1.32E-08
12	$nC \times nS$	5.7330E+00	5.68E-01	< 2E-16
13	$nH \times nS$	-1.3840E+00	2.49E-01	3.16E-08
14	$nO \times nCl$	5.4210E+00	1.09E+00	8.09E-07
15	$nN \times nCl$	-1.1170E+01	3.44E+00	0.001193

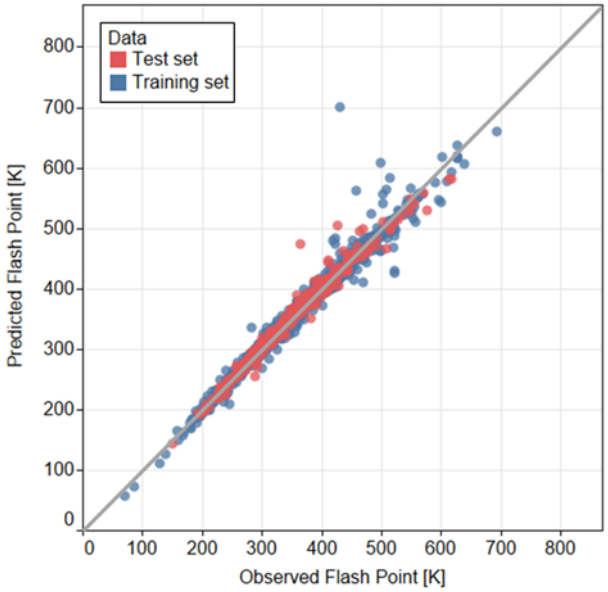
Subsequently, the two created models (Equations 4.10 and 4.11) were validated by predicting the flash points in the training and test set. As a result, the predicted flash points for 1,741 organic compounds in the dataset were computed and presented in the Supplementary Materials. Table 5 displays the main resultant statistical metrics. In both models, the generated consistent and high values of $R^2_{training}, Q^2_{Loo}, R^2_{test}, and\ Q^2_{ext}$ indicate that they are stable predictive models both internally and externally. Two plots of the predicted flash point values versus the observed ones in the two models are shown in Figure 4.4. As illustrated in this figure, these two model predictions for all organic compounds in the dataset are very accurate, with the average absolute difference between the predicted value and the observed value with 0.22% and 0.17%, respectively (Moreno et al., 2013). Although the accuracy of the SFFS-MLR-NLI (Step 2) appears to be slightly better, there is no statistical difference between the two models' performance.

Table 4.5 Predictabilities of the two created models for predicting flash points

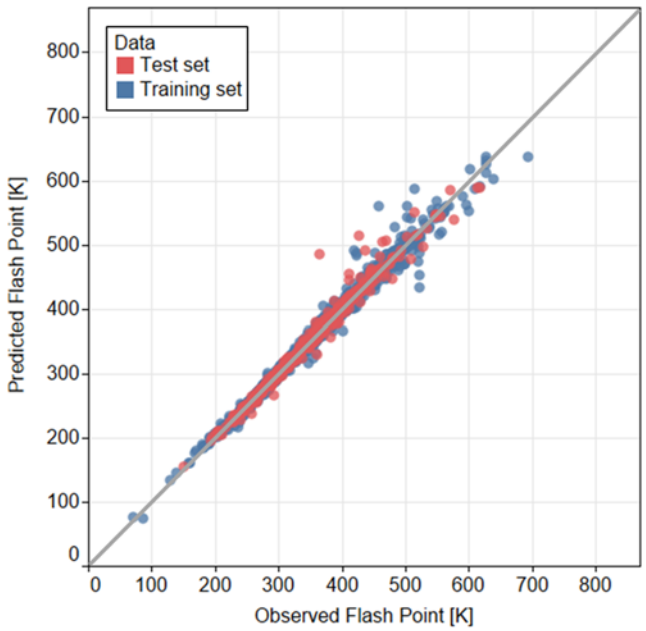
	SFBS-MLR-L model in Step 1			SFFS-MLR-NLI in Step 2		
Statistical metric	Training set	Test set	Entire dataset	Training set	Test set	Entire dataset
R^2	0.959	0.976	0.963	0.972	0.977	0.973
RMSE	16.37	12.50	15.68	13.56	12.21	13.30
$RMSE_{Loo}$	17.33	-	-	14.88	-	-
Q^2_{Loo}	0.956	-	-	0.966	-	-
Q^2_{ext}	-	0.976	-	-	0.977	-
MAE	8.27	8.07	8.23	7.31	7.32	7.31
Std. error	16.43	12.50	15.69	13.64	12.22	13.31
Number of compounds	1,394	347	1,741	1,394	347	1,741



(a) Predicted (Equation 4.10) vs. observed flash points in Step 1 via the SFBS



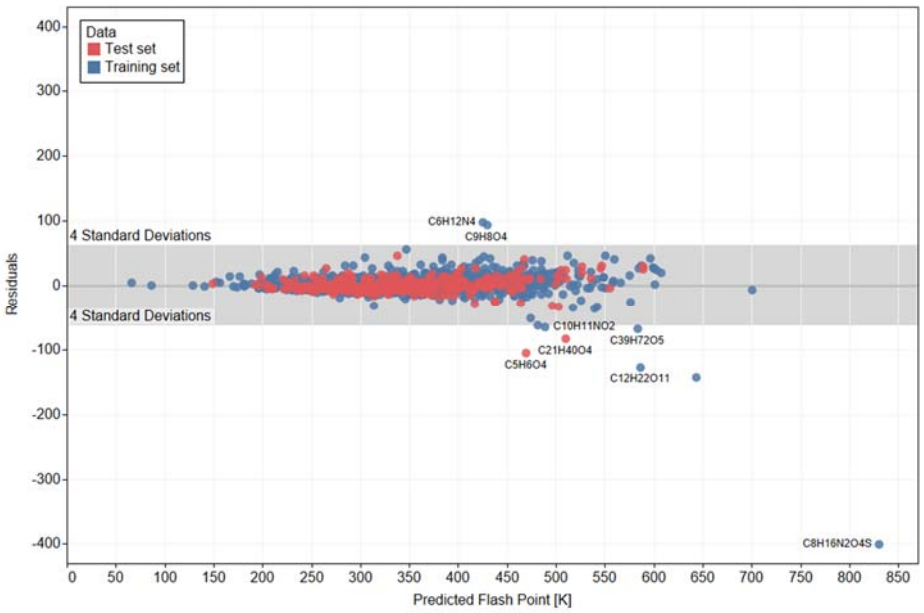
(b) Predicted (Equation 4.11) vs. observed flash points in Step 2 via the SFFS



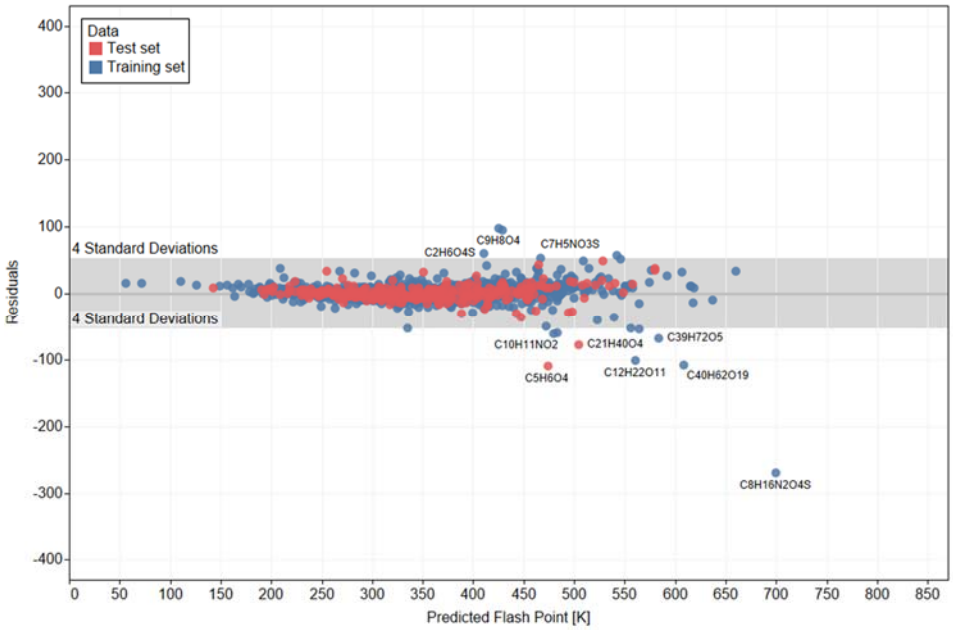
(c)* Predicted (Equation 4A1) vs. observed flash points in Step 2 via the SFBS

Figure 4.4 Correlations between the precited and observed flash point values for both training and test sets. The diagonal line represents perfect correlation of observed and predicted flash point values. *The detailed SFBS-MLR_NLI model is additionally provided in Appendix A to check whether this method can achieve a more accurate model in this study. Graphs (b) and (c) show that the predicted flash points via the SFBS-MLR-NLI model are closer to the diagonal line than the SFFS-MLR-NLI model.

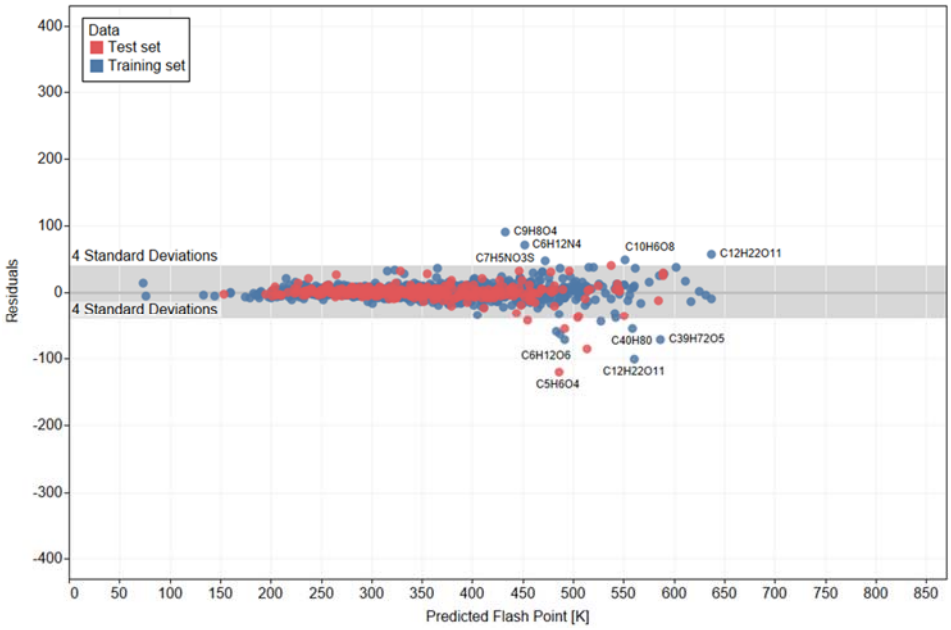
To determine whether any pattern was missed in the proposed MLR model, residual plots were examined. Given the two models (Equations 4.10 and 4.11), we plotted the residuals against the fitted values (Figure 4.5). Since both plots, Figure 4.5 (a) and (b), are randomly distributed on both sides of the zero line, we can conclude that there is no systematic error in developing the proposed models. Hence, these two models are valid for predicting the flash points of pure organic compounds.



(a) Residuals (Step 1 for flash point via the SFBS) vs predicted values (Equation 10)



(b) Residuals (Step 2 for flash point via the SFFS) vs predicted values (Equation 11)



*(c) Residuals (Step 2 for flash point via the SFBS) vs predicted values (Equation 4A1 in Appendix 4A)

Figure 4.5 Plots of residuals versus the predicted flash point values by the created models. *The detailed model of Graph (c) is additionally provided in Appendix 4A to check whether this method can achieve a more accurate model in this study.

Figure 4.5 also indicates outliers for each model. The outliers (>4s) are methenamine ($C_6H_{12}N_4$), 1-ethyl-3-methylimidazolium ethyl sulfate ($C_8H_{16}N_2O_4S$), 2-(acetyloxy)benzoic acid ($C_9H_8O_4$), acetoacetanilide ($C_{10}H_{11}NO_2$), sucrose ($C_{12}H_{22}O_{11}$), glycerol dioleate ($C_{39}H_{72}O_5$), sucrose acetate isobutyrate ($C_{40}H_{62}O_{19}$), citraconic acid ($C_5H_6O_4$), and glycerol 1-monooleate ($C_{21}H_{40}O_4$) in Step 1. Although these nine outliers in Step 1 still remain as outliers in Step 2, the residuals among the outliers decreased from 131.8 K to 93.9K. Of the nine outliers, eight chemical compounds contain more than two oxygen atoms in their molecular formulas, along with carbon and hydrogen atoms. These

outliers' common characteristics imply that more variables may be required to capture the interaction terms of oxygen with other atoms.

The SFBS-based model in Step 2 is provided in Appendix 4A. As shown in its statistical results in Appendix 4A, its predicted vs. observed values in Figures 4.4 (c), and residual plot in Figure 4.5 (c), the SFBS-MLR-NLI model is more accurate than the SFFS-MLR- NLI model and is adequate as a predictive model of the flash point. However, the model is less applicable than the SFFS-MLR-NLI model because of 46 predictor terms.

Table 4.6 summarizes the performance of the adequate three predictive models. This table also includes four flash point models constructed either by group contribution or by QSPR with the same data source, DIPPR. Although the same data source was utilized in all models, it could not be confirmed whether the dataset of the flash point models in this study covered the entire dataset the previous models used. Nevertheless, statistical metrics such as $higher\ R^2$, $lower\ MAE[K]$, $lower\ RME[K]$, and applicability to a wider variety of chemical compounds for flash point values indicate that the proposed flash point predictive models in this study seem to come out as being slightly better than the previous models.

Table 4.6 Comparison with the pressure model and previous models (the flash point of organic compounds)

Model	$R^2_{dataset}$ ($R^2_{training}$/ R^2_{test})	MAE $[K]$ (MAE_{train}/ MAE_{test})	$RMSE_{data}$ $[K]$ ($RMSE_{train}$/ $RMSE_{test}$)	Std. error	Dataset Size *	Property range (Applicable range) [K]	Note
Gharagheizi and Alamdari (2008)	0.967 (0.967/0.971)	10.2 (- / -)	-	12.7 (12.69/12.02)	1,030 (824/206)	-	GA-MLR QSPR model
Gharagheizi et al. (2008)	0.976 (0.977/0.966)	8.10 (7.90/9.94)	11.21 (10.97/13.24)	11.21 (10.99/13.14)	1,378 (1,321/330)	185.0-585.1	MLR Group contribution model
Bagheri, Bagheri, et al. (2012)	0.861 (0.863/0.860)	21.1 (21.09/21.39)	28.0 (28.07/27.89)	-	1,651 (1,321/330)	85.0-597.0	GA-MLR QSPR model
Bagheri, Bagheri, et al. (2012)	0.885 (0.880/0.905)	19.8 (19.5/18.33)	25.0 (26.00/23.42)	-	1,651 (1,321/330)	85.0-597.0	SVR QSPR model
Step 1 (Equation 4.10)	0.963 (0.959/0.976)	8.23 (8.27/8.07)	15.68 (16.37/12.50)	15.69 (16.63 /12.50)	1,741 (1,394/347)	71.0 -694.0	SFBS-MLR-L (Step 1 in this study)
Step 2 (Equation 4.11)	0.973 (0.972/0.977)	7.31 (7.31/7.32)	13.30 (13.56/12.21)	13.31 (13.64/12.22)	1,741 (1,394/347)	71.0 -694.0	SFFS-MLR-NLI (Step 2 in this study)
Step 2 (Equation A1 in Appendix 4A)	0.984 (0.986/0.976)	5.67 (5.49/6.41)	10.22 (9.59/12.43)	10.22 (9.76/12.45)	1,741 (1,394/347)	71.0 -694.0	SFBS-MLR-NLI (Step 2 in this study)

*All data reference is DIPPR 801

4.4.2. Results of heat of combustion models

The number of matched datasets among the heat of combustion, normal boiling point, and molecular weight in the compiled dataset is 1,850.

The number of matched datasets among the heat of combustion, normal boiling point, and molecular weight in the compiled dataset is 1,850.

$$\Delta H_C^o \left[\frac{\boldsymbol{kJ}}{\boldsymbol{mol}}\right] \approx (\boldsymbol{the\ numbers\ of\ C, H, O, N, F, Cl, Si, S, P, Pb}\ )$$
$$= \boldsymbol{-30.514 - 425.831nC - 90.766nH + 169.7306nO}$$
$$\boldsymbol{- 106.9996nN + 224.3168nF + 62.7552nCl - 683.1319nSi}$$
$$\boldsymbol{- 295.9456nS - 419.4349nP - 647.0202nPb}$$

Equation 4.12

where $nC, nH, nO, nN, nF, nCl, nSi, nS, nP, and\ nPb$ refer to the number of carbons, hydrogen, oxygen, fluorine, chlorine, silicon, sulfur, phosphorus, and lead atoms, respectively.

It should be noted that the model composes merely these ten atomic elements without molecular weight and boiling point terms. The contributions of N, F, Cl, Si, S, P, and Pb were derived from 205, 77, 124, 37, 90, 5, and 1 data point in 1,482 training dataset. The selected predictors' coefficients are statistically significant (p-value < 0.05), and multicollinearity is less likely to exist (VIF<10), as illustrated in Table 4.7.

The heat of combustion, ΔH_C^o, is negative when heat is released into the surroundings (i.e., exothermic reaction), and positive when heat is absorbed from the surroundings (e.g., endothermic reaction). In this manner, as indicated in Equation 4.12, the atomic elements with negative coefficients — C, H, N, Si, S, P and Pb — contribute to releasing heat that accompanying the combustion reaction. On the other hand, the three atomic elements with positive coefficients — O, F, and Cl — contribute to absorbing heat from the surroundings because they are likely to have positive reaction enthalpies, which are not common phenomena in combustion reactions. Representative organic compounds that absorb heat accompanying the reaction are freon gases such as chlorodifluoromethane ($CClF_3$) and carbon tetrafluoride (CF_4) in the current dataset.

Table 4.7 Selected predictors and their coefficients of the heat of combustion SFBS-MLR-L model (Equation 4.12)

No.	Variables	Coefficient	Std. error	p-value	VIF
0	Intercept	-3.0514E+01	6.8866E+00	1.01E-05	-
1	C	-4.2583E+02	1.4451E+00	< 2E-16	6.730
2	Cl	6.2755E+01	5.6633E+00	< 2E-16	1.093
3	F	2.2432E+02	2.0586E+00	< 2E-16	1.118
4	H	-9.0766E+01	7.7990E-01	< 2E-16	7.187
5	N	-1.0700E+02	5.5747E+00	< 2E-16	1.044
6	O	1.6973E+02	2.2941E+00	< 2E-16	1.147
7	P	-4.1943E+02	5.9844E+01	3.65E-12	1.013
8	Pb	-6.4702E+02	1.3291E+02	1.25E-06	1.002
9	S	-2.9595E+02	1.3105E+01	< 2E-16	1.029
10	Si	-6.8313E+02	7.0212E+00	< 2E-16	1.191

Second, by performing the SFFS-MLR-NLI procedure in the training set, the optimum subset of fifteen predictors was achieved in Step 2. The constructed model for predicting the heat of combustion is presented as the following:

$$\Delta H_C^o \left[\frac{kJ}{mol}\right]$$

$$\approx f(\textbf{the numbers of } C, H, O, N, F, Cl, Br, I, Si, S, P, T_b, MW, \textbf{interaction terms})$$

$$= -849.5 - 403.9nC - 83.83nH + 203.4nO + 282.6nF + 162.9nCl$$

$$+ 252.2nBr - 588nSi - 224.4nS - 319.2nP + 1.109T_b - 39.32\sqrt{MW}$$

$$+ 426.3log_{10}MW + 0.3186exp(nN) + 209.3exp(nI)$$

$$- MW(0.018T_b + 0.2111nN) + 2.909(nC \times nO) + nH(6.78nF - 1.289nO)$$

$$- 29.01(nO \times nN) - \frac{nN}{nC}\left(168 - 105.7\frac{nN}{nC}\right) + 33710\frac{1}{(T_b \times nC)}$$

Equation 4.13

The selected predictors' coefficients are statistically significant (p-value < 0.05), as illustrated in Table 4.8. This model was constructed by combining ten linear, four transformed, and nine interaction terms of the twenty-three predictors. Compared to Equation 4.12 in Step 1, Equation 4.13 contains a new atomic element (Iodine, noted as I) and nitrogen (noted N) in their logarithmic terms, and has one atomic element (lead, Pb) removed. Furthermore, boiling point, molecular weight, and their interaction terms with element atoms are newly included in Equation 4.13. new atomic elements — N and I — are included via their logarithmic terms. Once the boiling point physical property is

included, the interpretation might conflict with the relationship between the individual atomic element and the target property. However, we could infer some insights of the interaction terms of Equation 4.13. For example, the ratio term of nitrogen and carbon, $-\frac{nN}{nC}\left(168-105.7\frac{nN}{nC}\right)$, could trade off the effect. These terms produce positive values whenever the ratio is higher than 1.6. Therefore, the bond between nitrogen and carbon atoms positively affects absorbing heat from the surroundings. Besides, hydrogen bonds may reinforce exothermic or endothermic heat magnitude. For instance, the interaction between hydrogen and fluorine atoms with a positive coefficient value reinforces the endothermic heat magnitude of the standalone endothermic heat from fluorine atom. In contrast, the interaction between hydrogen and oxygen with a negative value reinforces the exothermic heat magnitude of the standalone exothermic heat due to oxygen. Hence, such interaction terms between different atoms aid in a better well-fit model.

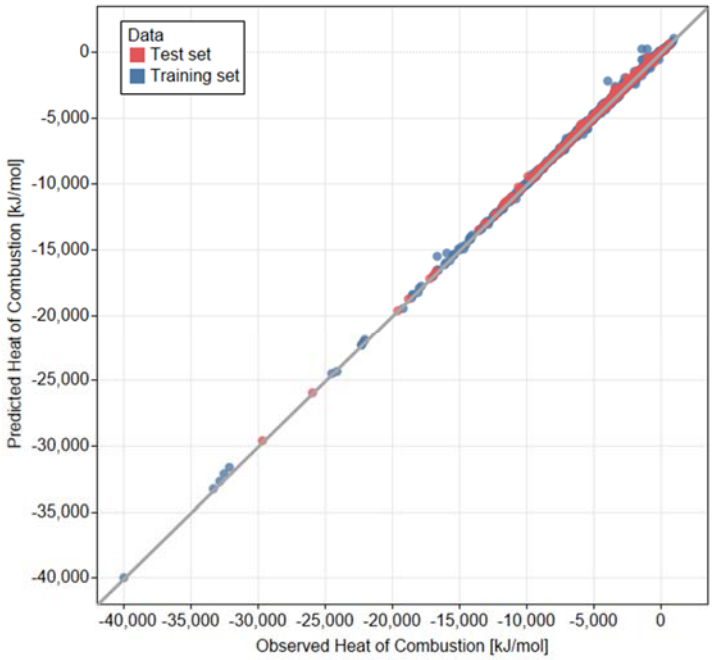
Table 4.8 Selected predictors and their coefficients of the heat of combustion SFFS-MLR-Model (Equation 4.13)

	Variables	Coefficient	Std. error	p-value
0	Intercept	-8.4950E+02	2.1380E+02	7.42E-05
1	nC	-4.0390E+02	6.2680E+00	< 2E-16
2	nH	-8.3830E+01	9.5420E-01	< 2E-16
3	nO	2.0340E+02	8.9500E+00	< 2E-16
4	nF	2.8260E+02	8.1580E+00	< 2E-16
5	nCl	1.6290E+02	1.7760E+01	< 2E-16
6	nBr	2.5220E+02	4.0050E+01	3.98E-10
7	nS	-2.2440E+02	1.9390E+01	< 2E-16
8	nSi	-5.8800E+02	1.4800E+01	< 2E-16
9	nP	-3.1920E+02	5.3690E+01	3.45E-09
10	T_b	1.1090E+00	9.1510E-02	< 2E-16
11	$\sqrt{MW}$	-8.9320E+01	2.1590E+01	3.70E-05

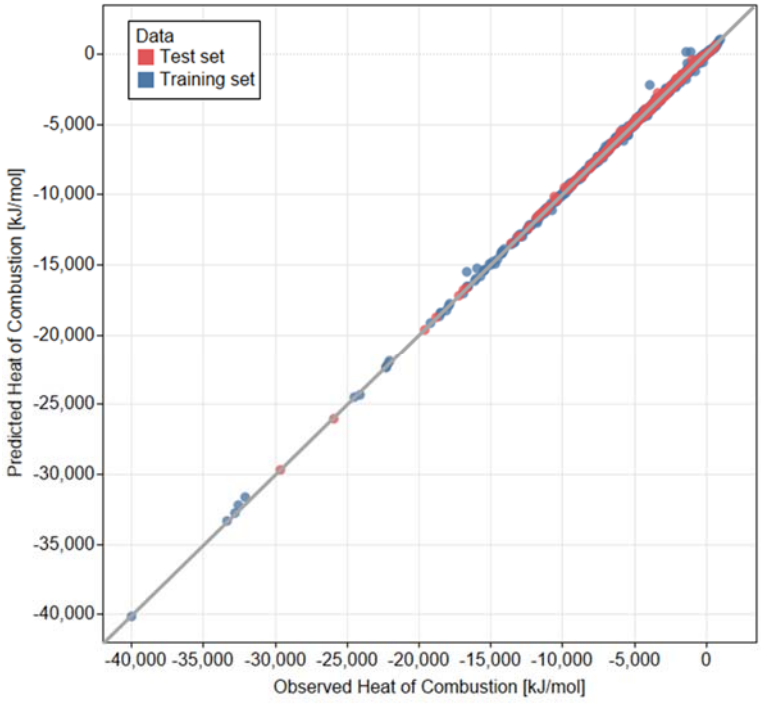
	Variables	Coefficient	Std. error	p-value
12	$\log_{10} MW$	4.2630E+02	1.9060E+02	0.025451
13	$exp\ (nN)$	3.1860E-01	6.5470E-02	1.26E-06
14	$exp\ (nI)$	2.0930E+02	4.1850E+01	6.38E-07
15	$MW \times T_b$	-1.8320E-03	4.1250E-04	9.58E-06
16	$MW \times nN$	-2.1110E-01	6.7450E-02	0.001785
17	$nC \times nO$	2.9090E+00	5.9270E-01	1.02E-06
18	$nH \times nO$	-1.2890E+00	3.3470E-01	0.000123
19	$nH \times nF$	6.7800E+00	1.9100E+00	0.000399
20	$nO \times nN$	-2.9010E+01	2.5330E+00	< 2E-16
21	$\frac{nN}{nC}$	-1.6800E+02	5.3350E+01	0.001667
22	$\frac{(nN)^2}{(nC)^2}$	1.0570E+02	3.6160E+01	0.003517
23	$\frac{1}{(T_b \times nC)}$	3.3710E+04	5.9630E+03	1.88E-08

Table 4.9 Predictabilities of the two created models for predicting the heat of combustion

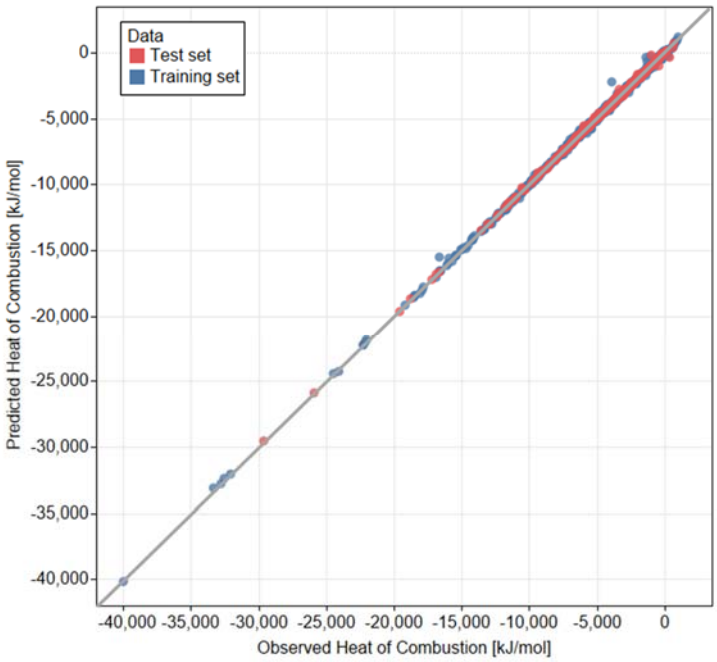
	SFBS-MLR-L model in Step 1			SFFS-MLR-NLI in Step 2		
Statistical metric	Training set	Test set	Overall dataset	Training set	Test set	Overall dataset
R^2	0.9987	0.9992	0.9988	0.9991	0.9994	0.9991
RMSE [kJ/mol]	132.24	102.45	126.87	112.29	86.71	107.68
$RMSE_{Loo}$ [kJ/mol]	135.81	-	-	118.73	-	-
Q^2_{Loo}	0.9986	-	-	0.9989	-	-
Q^2_{ext}	-	0.9992	-	-	0.9994	-
MAE [kJ/mol]	73.36	65.77	71.85	61.86	55.53	60.60
Std. error	132.7	102.44	126.90	113.20	86.53	107.70
Number of compounds	1,482	368	1,850	1,482	368	1,850



(a) Predicted (Equation 4.12) vs. observed heats of combustion in Step 1 via the SFBS



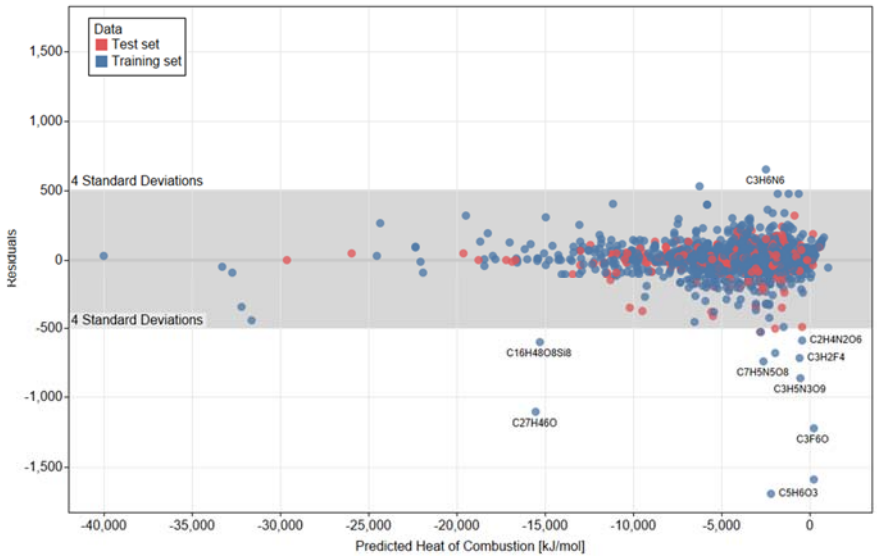
(b) Predicted (Equation 4.13) vs. observed heats of combustion in Step 2 via the SFFS



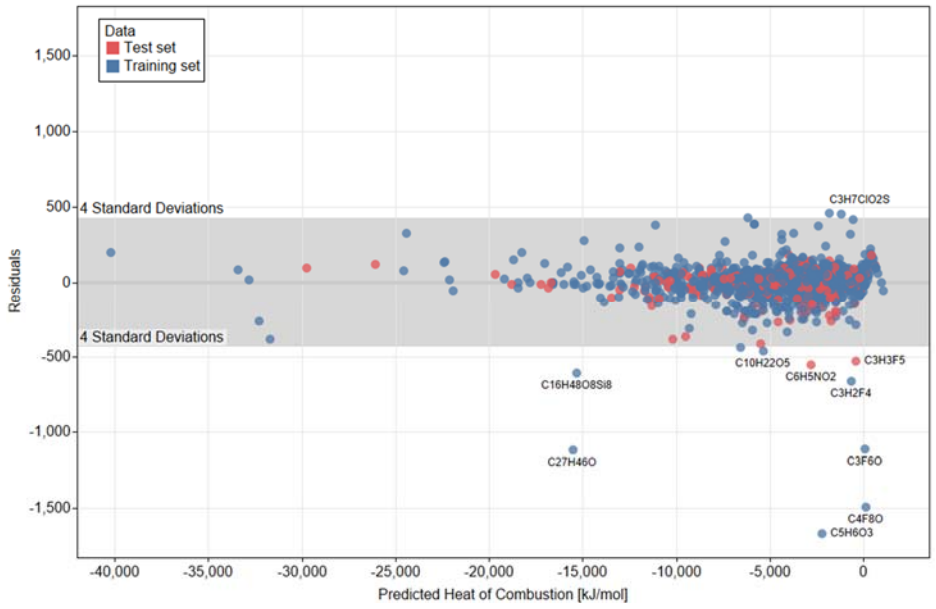
(c)* Predicted (Equation 4B1) vs. observed heats of combustion in Step 2 via the SFBS

Figure 4.6 Correlations between the predicted and observed heat of combustion values for both training and test sets. The diagonal line represents perfect correlation of observed and predicted flash point values. *The detailed SFBS-MLR_NLI model is additionally provided in Appendix 4B to check whether this method can achieve a more accurate model in this study. Graphs (a)- (c) indicate that all of the predicted heats of combustion via the three models are closer to the diagonal line**.**

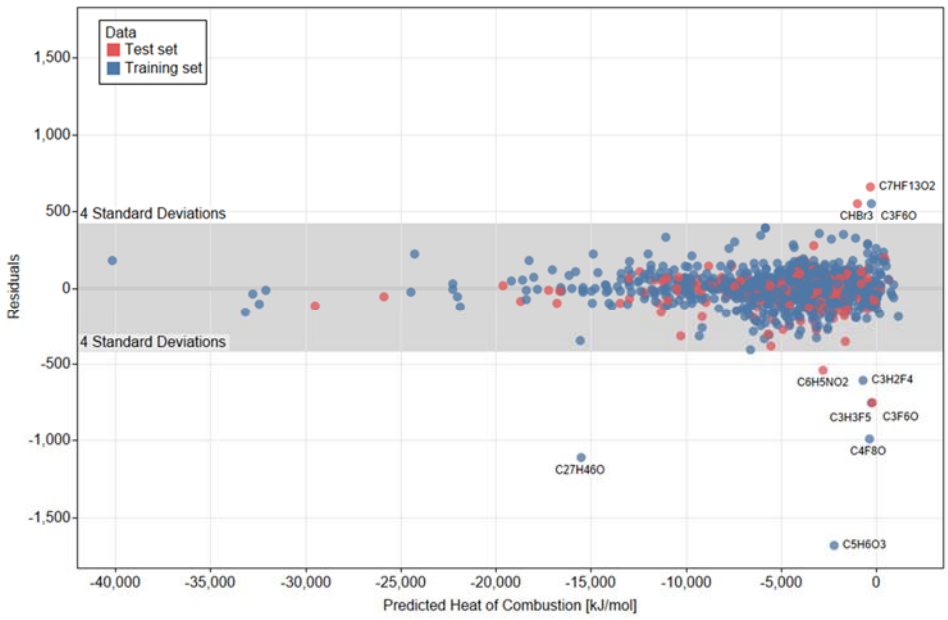
To determine whether any pattern was missed in the proposed MLR model, residual plots were examined. Given the two models (Equations 4.12 and 4.13), we plotted the residuals against the fitted values (Figure 4.7). Since both plots are randomly distributed on both sides of the zero line, we can conclude that there is no systematic error in developing the proposed model. Hence, these two models are valid for predicting the heats of combustion of pure organic compounds.



(a) Residuals (Step 1 for heat of combustion via the SFBS) vs predicted values (Equation 4.12)



(b) Residuals (Step 2 for heat of combustion via the SFFS) vs predicted values (Equation 4.13)



(c)* Residuals (Step 2 for heat of combustion via the SFBS) vs predicted values (Equation 4B1)

Figure 4.7 Plots of residuals versus the predicted heat of combustion values by the two created model. The detailed model of Graph (c) is additionally provided in Appendix 4B to check whether this method can achieve a more accurate model in this study, but it indicates that the SFBS-MLR-NLI model is overfitting with more outliers from the test set than the SFFS-MLR-NLI model**.**

Moreover, Figure 4.7 indicates outliers for each model. The outliers (>4s) are trifluoromethyl trifluoro vinyl ether (C_3F_6O), pentafluoroethyl trifluoro vinyl ether (C_4F_8O), (S)-4-ethenyl-1,3-dioxolan-2-one ($C_5H_6O_3$), 2,3,3,3-tetrafluoropropene ($C_3H_2F_4$), pentafluoroethyl trifluorovinyl ether (C_4F_8O), (S)-4-ethenyl-1,3-dioxolan-2-one ($C_5H_6O_3$) and hexadecamethylcyclooctasiloxane ($C_{16}H_{48}O_8Si_8$) in both models. Like flash point predictive models, these outliers mostly contain more than an oxygen atom in their molecular formulas along with carbon and hydrogen atoms.

Although the residuals among the outliers (>4s) slightly decreased from 838.3 kJ/mol. in Step 1 to 797.1 kJ/mol. in Step 2, there is no outstanding difference between the two steps. This result indicates that there are almost no interaction effects for the characteristics regarding heat of combustion property.

The SFBS-based model in Step 2 is provided in Appendix 4B. As shown in its statistical results in Appendix 4B, its predicted vs. observed values in Figure 4.6 (c), and residual plot in Figure 4.7 (c), the SFBS-MLR-NLI model is adequate as a predictive model. However, this model has slightly poorer performance than the SFFS-MLR-NLI model

summarizes the performance of the three adequate predictive models. This table also includes additional models constructed by QSPR with the same data source, DIPPR. Although the same data source was utilized in all models, it no confirmation could be obtained whether the dataset of the heat of combustion models in this study covered the entire dataset the QSPR model used. Nevertheless, statistical metrics such as $higher\ R^2$, $lower\ MAE[K]$, $lower\ RMSE[K]$, lower standard error, and applicability to a wider variety of chemical compounds for the heat of combustion values indicate that the proposed heat of combustion predictive models in this study seem to come out as being slightly better than the previous model.

Table 4.10 Comparison with the present model and previous models (the heat of combustion of organic compounds)

Model	$R^2_{dataset}$ ($R^2_{training}$/ R^2_{test})	MAE [kJ/mol]	RMSE [kJ/mol]	Std. error	Dataset Size *	Property range (Applicable range) [kJ/mol]	Note
Pan et al. (2011)	0.995 (0.995/0.996)	- (122.21/104.14)	- (210.27/163.23)	-	1,650 (1,322/328)	-22,100 – 954.8	ACO[1]-PLS[2] QSPR model
Equation 4.12	0.999 (0.999/0.999)	71.85 (73.36/65.77)	126.87 (132.24/102.45)	126.9 (132.7/102.44)	1,850 (1,482/368)	-39,985.3 – 982.7	SFBS-MLR-L (Step 1 in this study)
Equation 4.13	0.999 (0.999/0.999)	60.60 (61.86/55.53)	107.68 (112.29/86.71)	107.70 (113.2/86.53)	1,850 (1,482/368)	-39,985.3 – 982.7	SFFS-MLR-NLI (Step 2 in this study)
Equation 4B1 (Appendix 4B)	0.999 (0.999/0.999)	61.87 (61.68/62.63)	105.28 (105.50/104.35)	105.29 (105.50/104.35)	1,850 (1,482/368)	-39,985.3 – 982.7	SFBS-MLR-NLI (Step 2 in this study)

*All data reference is DIPPR 801. After constructing the heat of combustion models, one distinguishable outlier, dehydroabietylamine ($C_{20}H_{31}N$), in all models was discarded. The outlier's input variables used for calculation were $nC = 20, nH = 31, nN = 1, and\ MW = 285.467[\frac{g}{mol}]$. This erroneous value could probably be attributed to an experimental error or typo in the DIPPR database. The observed heat of combustion value of dehydroabietylamine was -1,550 kJ/mol. from the DIPPR database, whereas its corresponding predicted values were -11,409 kJ/mol. (in Step 1 via SFBS), -11,312 kJ/mol. (in Step 2 via SFFS), and -10,847 kJ/mol. (in Step 2 via SFBS). Given that another flammability property database, Yaw's Chemical Properties Handbook (1999), provided its value as -11,440 kJ/mol that was much closer to this study's predicted values, the heat of combustion value of dehydroabietylamine in the DIPPR database might incorrectly be addressed. Moreover, the compared study, Pan et al. (2011) also does not include this chemical for its model. Therefore, this study recreated the models as shown in Section 4.4.2 and Appendix 4B after the outlier to provide more reliable predictive models.

[1] ACO—ant colony optimization algorithm

[2] PLS—partial least square

4.4.3. Results of LFL models

The number of matched datasets among the LFLs, boiling points, and molecular weights in the compiled dataset is 1,733.

Initially, the LFL SFBS-MLR-L model was constructed similar to the flash point and the heat of combustion SFBS-MLR-L models. However, the performance of this model was inadequate after analyzing its performance (a detailed explanation of why this model, Equation 4C1, was rejected is given in the section). The rejected model is provided in Appendix 4C1 to prevent practitioners applying the unreliable model.

In a second step, by performing the SFFS-MLR-NLI procedure in the training set, the optimum subset of nine predictors was created in Step 2. The created model for predicting the heat of combustion is presented as the following:

$$\begin{aligned} \boldsymbol{LFL}\ [\boldsymbol{vol.}\%] &\approx \boldsymbol{f}(\boldsymbol{Interactions\ of\ C, H, O, F, Cl, Si, S}) \\ &= \mathbf{-0.1014} + \boldsymbol{nCl}(\mathbf{0.3103}\boldsymbol{nO} - \mathbf{0.0613}\boldsymbol{nH}) \\ &+ \frac{\mathbf{1}}{\boldsymbol{nC}}(\mathbf{10.3579} + \mathbf{1.6783}\boldsymbol{nF} + \mathbf{2.0032}\boldsymbol{nCl}) \\ &+ \frac{\mathbf{1}}{(\boldsymbol{nC})^{\mathbf{2}}}(\mathbf{-1.1007}\boldsymbol{nH} + \mathbf{0.496}(\boldsymbol{nO})^{\mathbf{2}} - \mathbf{4.3578}\boldsymbol{nS} - \mathbf{6.896}\boldsymbol{nSi}) \end{aligned}$$

Equation 4.14

The selected predictors' coefficients are statistically significant (p-value < 0.05), as illustrated in Table 4.11. This model was constructed by combining interaction terms between atomic elements (Equation 4.14). Thus, the created model implies that the LFL

values of organic compounds are determined by diverse molecular interactions rather than a single atom's characteristics. These interaction predictors correspond with the descriptors of QSPR models, which are associated with chemical interactions (e.g., Me — mean atomic Sanderson electronegativity, and GATS1v — information related to the atomic Van der Walls volumes of a molecule) (Bagheri, Rajabi, et al., 2012; Pan et al., 2009a).

Furthermore, there are two significant findings through the selected fraction terms (antoehr interaction type). First, the sign of the fraction terms may be associated with the selected atom's electronegativities. Second, these fractions may describe a wide range of LFL among lighter organic compounds (the number of C<2). Regarding the sign of the fraction, for example, $\left(\frac{nO}{nC}\right)^2, \frac{nF}{nC}, and \frac{nCl}{nC}$ terms have the plus sign on their coefficients in Equation 4.14. They commonly consider more electronegative atoms than carbon on their numerators. On the other hand, $\frac{nH}{(nC)^2}, \frac{nS}{(nC)^2}, and\ \frac{nSi}{(nC)^2}$ terms on their coefficients, and when taken into account, the atoms on the numerators are less electronegative than carbon. Then, the fraction terms with the plus sign make increased LFL values, whereas the terms with the minus sign make reduced LFL values.

However, the question remains how certain organic compounds have relatively higher LFL values than the other compounds, in particular among lighter compounds (C<2). For example, the LFL values of chlorofluoromethane (CH_2ClF) and dichloromethane (CH_2Cl_2) are 14.4 and 14.0 vol.%, which are the highest LFL values in the dataset. On the other hand, those of carbon disulfide (CS_2) and methyl mercaptan

(CH_4S) are 1.3 and 3.9 vol.%, which are much lower than the former chemicals (DIPPR, 2019). Such a phenomena can be explained by addressing an *organic atmosphere of counter charge*. According to Huckel and Debye (1923), the atmosphere occurs when there is a slight imbalance of charge in ionic solutions because there is a slight excess of cations near any anion and a slight excess of anion near any cation (Atkins & De Paula, 2017). Since each ion is oppositely charged, the energy of the solution is lower than the ideal solution, which is uniformly charged, and therefore its chemical potential is lower than in an ideal solution. Although this study's organic compounds are not for ion solutions but of high-pressure vapors, which are the precondition of combustions, the high-pressure vapors could show a similar behavior by producing a slight imbalance of charge. That is, these compounds with such imbalanced charges may have lower chemical potential than other vapors.

To estimate the ratios between the electronegativities of other atoms and the electronegativity of carbon atom in a round way, the ratios between the numbers of other atoms and the number of carbons were prepared. As a result, when organic compounds had fewer carbons, the ratios between the numbers of other atoms and the number of carbons tended to increase. Hence, we could conclude that the atmospheres of counter charge among the organic compounds with fewer carbons are reinforced because these compounds' charges are easily imbalanced at high pressure. Therefore, when compared to other vapor compounds, these organics could have a distinct behavior with a wider range of LFL values. When the charges of organic compounds are more imbalanced due to more electronegative atoms and fewer carbons, more molecules (i.e., higher LFL % vol.)

could be required to accompany a combustion reaction due to their lower chemical potentials. For example, some low carbon halogenated organic compounds (e.g., halon 1211 CF2BrCl) are capable of inhibiting flames or extinguishing materials rather than generating flames, although they are banned because of their ozone-hold enlarging properties.

Nevertheless, the magnitudes of coefficients in Equation 4.14 are not directly proportional to the electronegativities of individual atoms. This result presents that the constructed model takes into account other causes such as chemical's steric effects along with the interaction terms

Table 4.11 Selected predictors and their coefficients of the LFL SFFS-MLR-NLI model (Equation 4.14)

No.	Variables	Coefficient	Std. error	p-value
0	Intercept	-1.0143E-01	2.3130E-02	1.25E-05
1	$nH \times nCl$	-6.1280E-02	6.2100E-03	< 2E-16
2	$nO \times nCl$	3.1025E-01	3.6580E-02	< 2e-16
3	$\frac{1}{nC}$	1.0358E+01	2.0865E-01	< 2E-16
4	$\frac{nH}{(nC)^2}$	-1.1007E+00	6.4580E-02	< 2E-16
5	$\frac{(nO)^2}{(nC)^2}$	4.9599E-01	3.9590E-02	< 2E-16
6	$\frac{nF}{nC}$	1.6783E+00	7.4730E-02	< 2E-16
7	$\frac{nCl}{nC}$	2.0032E+00	1.0776E-01	< 2E-16
8	$\frac{nS}{(nC)^2}$	-4.3578E+00	1.8918E-01	< 2E-16
9	$\frac{nSi}{(nC)^2}$	-6.8960E+00	4.0987E-01	< 2E-16

Subsequently, the created model (Equation 4.14) was validated by predicting the LFL values in the training and test set. As a result, the predicted LFL values for 1,733 organic compounds in the dataset were computed and presented in the Supplementary Materials. The main resultant statistical metrics are presented in Table 4.12, the correlations between the predicted and observed LFL values in Figure 4.8, and corresponding residual plots are shown in Figure 4.9.

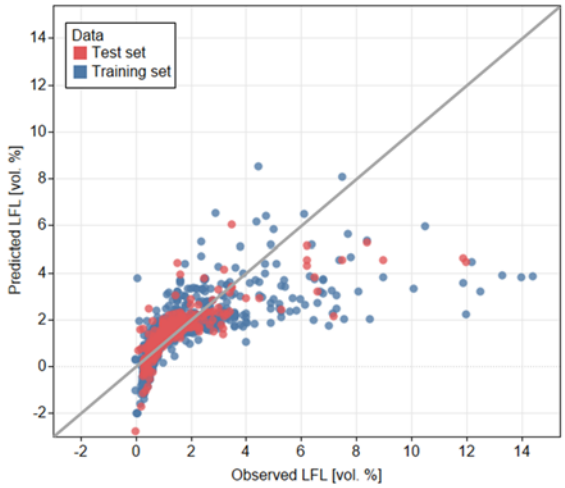
As illustrated in Table 4C (a), the low R^2 (< 0.5) value indicates that the selected predictor variables in Step 1 are not further explaining the variation of LFL values, regardless of the variable significance (Table 4C1). Figure 4.8(a) also indicates that the values that are obtained from Equation 4C1 in Step 1 failed to predict the observed LFLs, because most LFLs points are not close to the regressed diagonal line. Finally, the residual plot of the model, which has a non-flat pattern between -3 and 2 predicted LFLs [vol.%], evidently reveals that essential predictor variables are missing in Equation 4C1 (Figure 4.9(a)). Therefore, we concluded that Equation 4.14 incorporating simple linearity is invalid in predicting LFL values.

In contrast, the SFFS-MLR-NLI model of Step 2 (Equation 4.14) generated consistent and high values of $R^2_{training},\ Q^2_{Loo}, R^2_{test}, and\ Q^2_{ext}$. This result indicates that this model is a stable predictive model internally and externally. This result clearly shows that the predictive performance significantly increases using various atomic interaction terms. The plot of the predicted LFL values versus the observed ones in the model shows that most LFL points are close to the regressed diagonal line (Figure 4.8(b)), even though there is still room to improve. To determine whether any pattern was missed in the MLR

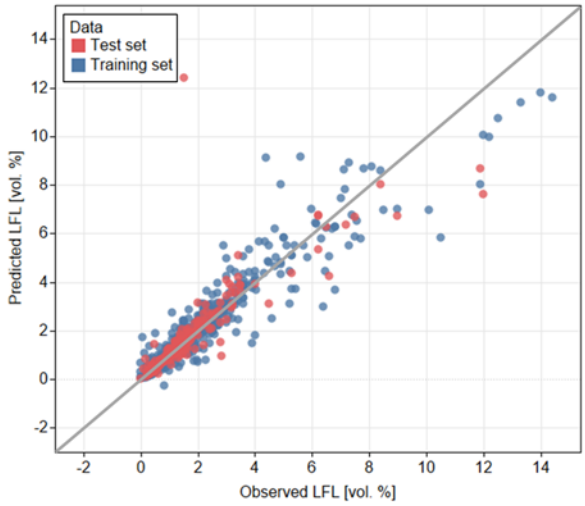
model, residual plots were examined, as shown in Figure 4.9(b)). Since the plot is randomly distributed on both sides of the zero line, we could conclude that there is no systematic error in developing the proposed SFFS-MLR-NLI model. Therefore, this model is valid for predicting the LFL values of pure organic compounds

Table 4.12 Predictabilities of the SFFS-MLR-NLI model for predicting LFL in Step 2 (Equation 4.14)

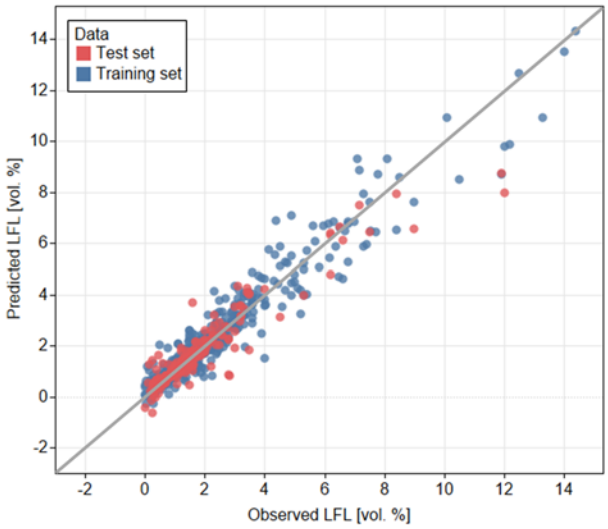
Statistical metric	Training set	Test set	Overall dataset
R^2	0.898	0.741	0.868
RMSE [vol.%]	0.469	0.741	0.535
$RMSE_{Loo}$ [vol.%]	0.506	-	-
Q^2_{Loo}	0.882	-	-
Q^2_{ext}	-	0.741	-
MAE [vol.%]	0.202	0.229	0.207
Std. error	0.471	0.742	0.535
Number of compounds	1,388	345	1,733



(a)* Predicted (Equation 4C1) vs. observed LFLs in Step 1 via the SFBS (**the model rejected**)

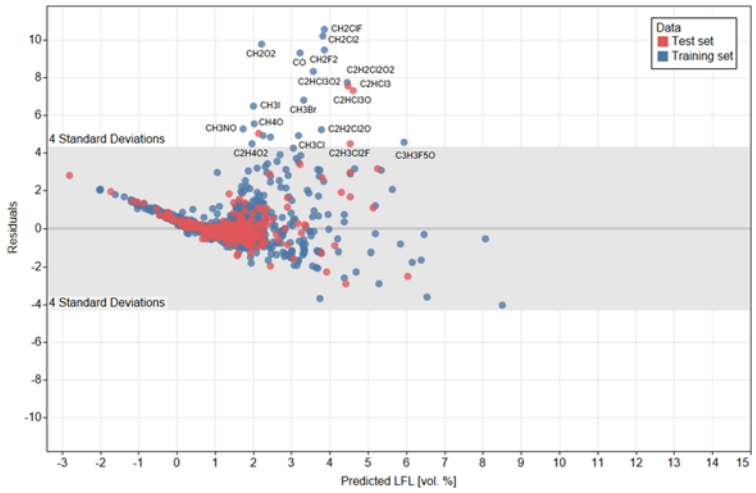


(b) Predicted (Equation 4.14) vs. observed LFLs in Step 2 via the SFFS

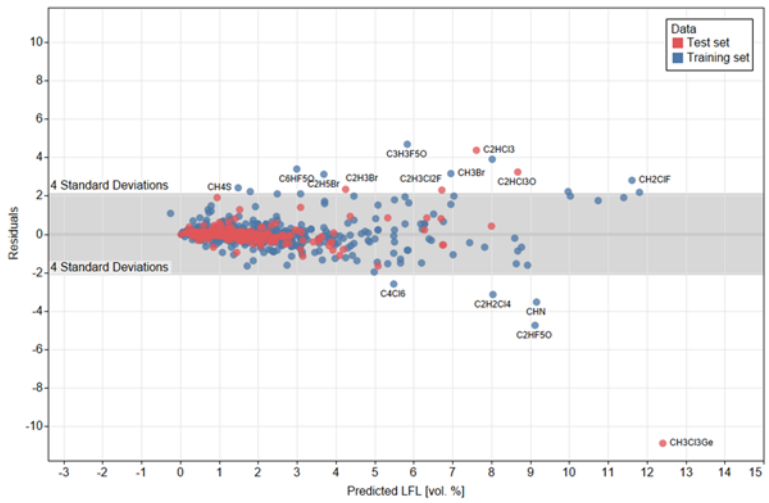


(c)* Predicted (Equation 4C2) vs. observed LFLs in Step 2 via the SFBS

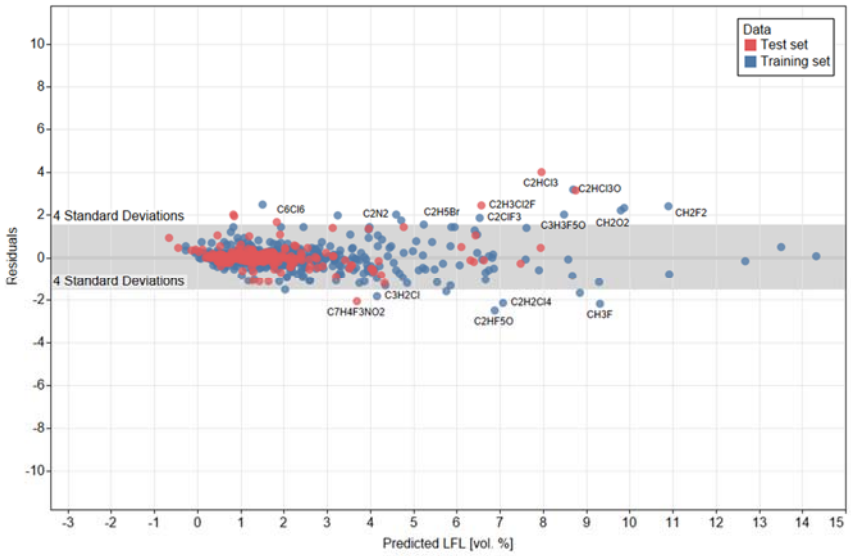
Figure 4.8 Correlation between the predicted and observed LFLs for both training and test sets. The diagonal line represents perfect correlation of observed and predicted LFLs. *The detailed models of Graphs (a) and (c) are additionally provided in Appendix 4C to check whether this method can achieve a more accurate model in this study**.**



(a)* Residuals (Step 1 for LFL via the SFBS) vs predicted values (Equation 4C1) (**the model rejected**)



(b) Residuals (Step 2 for LFL via the SFFS) vs predicted values (Equation 4.14)



(c)* Residuals (Step 2 for LFL via the SFBS) vs predicted values (Equation 4C2)

Figure 4.9 Plots of residuals versus the observed LFL values for the constructed models. *The detail of the Graphs (a) and (c) model is additionally provided in Appendices 4C to check whether this method can achieve a more accurate model in this study. This graph shows how the predictabilities gradually increased by adding more flexible terms. The most notable outlier, Methyl germanium trichloride (CH_3Cl_3Ge) in the SFBS-MLR-L in Step 1 and the SFFS-MLR-L in Step 2, is not the outlier any more in the SFBS-MLR-NLI in Step 2**.**

Figure 4.9 also indicates outliers of the created LFL models. All outliers (>4s) in Step 1, 14 outliers contain less than or equal to two carbon atoms in their molecular formulas. The remarkable outliers such as chlorofluoromethane (CH_2ClF) and dichloromethane (CH_2Cl_2) indicate that the simple MLR model (Eq. 14) fails to capture the behavior of these lighter halogenated organic compounds' LFL values. Although several outliers in Step 1 remain as outliers in Step 2, the average residuals among the outliers considerably decreased from 6.8 vol.% to 3.5 vol.%. Presumably, this outlying

behavior has to ascribed to halogen radicals' relatively strong influence on flame kinetics leading to some degree of suppression.

The SFBS-based model in Step 2 is provided in Appendix 4C. As shown in its statistical results in Appendix 4C, its predicted vs. observed values in Figure 4.8(c), and residual plot in Figure 4.9(c), the SFBS-MLR-NLI model is adequate as a predictive model. In addition, the model has higher accuracy than the other proposed models. However, the model is less applicable than the SFFS-MLR-NLI model because of 59 predictor terms.

Table 4.13 summarizes the performance of the two adequate predictive models. This table also includes two models constructed by QSPR with the same data source, DIPPR. Although the same data source was utilized in all models, it could not be confirmed whether the dataset of LFL models in this study covered the entire dataset the QSPR models used. Nevertheless, statistical metrics such as $higher\ R^2,\ lower\ MAE[K]$, $lower\ RMSE[K]$, lower standard error, and applicability to a wider variety of chemical compounds for the LFL values indicate that the proposed LFL predictive models seem to come out as being slightly better than the previous models.

It is clear from the findings that the SFBS-MLR-NLI model presented here can perform competitively to QSPR MLR models in predicting LFL values. That is, rather numerous interaction terms are required to capture the characteristics of LFL values. Meanwhile, the crisp model, SFBS-MNR-NLI model, enables us to feasibly interpret the relationship between atomic elements and LFL by providing reliable, accurate performance. On the other hand, the model simply incorporating linearities among default

variables fails to predict LFL because the characteristics of LFL may be strongly associated with molecular interactions.

Table 4.13 Comparison with the current model and previous models (the LFL of pure organic compounds)

Model	$R^2_{dataset}$ ($R^2_{training}$/ R^2_{test})	MAE [vol. %]	RMSE [vol.%]	Std. error	Dataset Size* (training/test)	Property range (Applicable range) [vol. %]	Note
Bagheri, Rajabi, et al. (2012)	0.906 (0.906/0.908)	0.1867 (0.191/0.162)	0.335 (0.339/0.318)	-	1,615 (1,292/323)	0.1-12.0 vol.%	QSPR model (MLR)
Bagheri, Rajabi, et al. (2012)	0.929 (0.930/0.926)	0.153 (0.158/0.135)	0.287 (0.290/0.278)	-	1,615 (1,292/323)	0.1-12.0 vol.%	QSPR model (ANN)
Equation 4.14	0.868 (0.898/0.741)	0.207 (0.202/0.229)	0.535 (0.469/0.741)	0.535 (0.47/0.74)	1,733 (1,388/345)	0.0002 – 14.4 vol.%	SFFS-MLR- NLI (Step 2 in this study)
Equation 4C2 (Appendix 4C2)	0.930 (0.939/0.894)	0.188 (0.188/0.218)	0.389 (0.365/0.473)	0.389 (0.37/0.47)	1,733 (1,388/345)	0.0002 – 14.4 vol.%	SFBS-MLR- NLI (Step 2 in this study)

*All data reference is DIPPR 801

4.4.4. Results of UFL models

The number of matched datasets among the UFLs, boiling points, and molecular weights in the compiled dataset is 1,711.

Initially, the UFL SFBS-MLR-L model was constructed similar to the flash point and the heat of combustion SFBS-MLR-L models. However, the performance of this model was inadequate after analyzing its performance (a detailed explanation of why this model, Equation 4D1, was rejected is given in the section). The rejected model is provided in Appendix 4D1 to prevent practitioners applying the unreliable model.

In a second step, by performing the SFFS-MLR-NLI procedure in the training set, the optimum subset of nineteen predictors was created in Step 2. The created model for predicting the UFL is presented as the following.

$$\begin{aligned}
&\boldsymbol{UFL}\ [\boldsymbol{vol.}\%] \\
&\approx \boldsymbol{f(nC, nH, nN, nO, nF, nCl, nBr, nI, nSi, nS, MW, interaction\ terms)} \\
&= \mathbf{7.7890} + \mathbf{0.4438}\boldsymbol{nH} - \mathbf{2.8850}\sqrt{\boldsymbol{nC}} + \mathbf{0.0247}\boldsymbol{exp}(\boldsymbol{nN}) + \mathbf{0.0184}\boldsymbol{exp}(\boldsymbol{nCl}) \\
&+ \mathbf{0.9688}\boldsymbol{exp}(\boldsymbol{nBr}) + \mathbf{4.1030}\boldsymbol{exp}(\boldsymbol{nI}) - \mathbf{0.0023}(\boldsymbol{nH})^2 - \mathbf{2.7690}\ (\boldsymbol{Si})^2 \\
&+ \mathbf{1.702}(\boldsymbol{nS})^2 + \mathbf{0.0063}(\boldsymbol{MW} \times \boldsymbol{nF}) + \mathbf{0.2575}(\boldsymbol{nH} \times \boldsymbol{nSi}) \\
&+ \boldsymbol{nO}(\mathbf{1.0010}\boldsymbol{nSi} - \mathbf{1.789}\boldsymbol{nS}) \\
&+ \frac{\mathbf{1}}{\boldsymbol{nC}}\left(\mathbf{73.04} - \mathbf{29.07}\frac{\mathbf{1}}{\boldsymbol{nC}} - \mathbf{4.49}\boldsymbol{nH} + \mathbf{7.458}\boldsymbol{nO} + \mathbf{45.51}\boldsymbol{nSi} - \mathbf{0.2899}\boldsymbol{MW}\right)
\end{aligned}$$

Equation 4.15

The selected predictors' coefficients are statistically significant (p-value < 0.05), as illustrated in Table 4.14. Compared to the performance of Equation 4D1 in Step 1, various transformed variables (e.g., squared, logarithmic, exponential, and quadratic), interaction terms between atoms, and molecular weights are used as predictors in Equation 4.15. The created model implies that the UFL values of organic compounds are determined by various atomic interactions and bulk-related factors (e.g., weights). These selected predictors correspond with the descriptors of QSPR models, which are associated with atomic interactions and bulk-related factor (e.g., BELm1 accounts for weighted by atomic masses, BEHp1 and BELp1 account for the weighted by atomic polarizabilities, and ATS2v considers atomic Van der Waals volumes (Pan et al., 2009b))

Table 4.14 Selected predictors and their coefficients of the UFL SFFS-MLR-NLI model (Equation 4.15)

	Variables	Coefficient	Std. error	p-value
0	Intercept	7.7890E+00	3.18E+00	1.4350E-02
1	nH	4.4380E-01	7.03E-02	3.6900E-10
2	$\sqrt{nC}$	-2.8850E+00	8.41E-01	6.1800E-04
3	$exp\ (nN)$	2.4740E-02	8.24E-03	2.7400E-03
4	$exp\ (nCl)$	1.8430E-02	9.11E-03	4.3222E-02
5	$exp\ (nBr)$	9.6880E-01	4.07E-01	1.7359E-02
6	$exp\ (nI)$	4.1030E+00	1.46E+00	4.9380E-03
7	$(nH)^2$	-2.3220E-03	4.67E-04	7.4400E-07
8	$(nSi)^2$	-2.7690E+00	6.26E-01	1.0500E-05
9	$(nS)^2$	1.7020E+00	3.66E-01	3.7000E-06
10	$MW \times nF$	6.3250E-03	1.43E-03	1.0600E-05
11	$nH \times nSi$	2.5750E-01	7.78E-02	9.6500E-04
12	$nO \times nSi$	1.0010E+00	3.36E-01	2.9640E-03
13	$nO \times nS$	-1.7890E+00	7.40E-01	1.5715E-02
14	$\frac{1}{nC}$	7.3040E+01	6.27E+00	< 2E-16

	Variables	Coefficient	Std. error	p-value
15	$\frac{1}{(nC)^2}$	-2.9070E+01	4.91E+00	4.1300E-09
16	$\frac{nH}{nC}$	-4.4900E+00	4.79E-01	< 2E-16
17	$\frac{nO}{nC}$	7.4580E+00	6.11E-01	< 2E-16
18	$\frac{nSi}{nC}$	4.5510E+01	3.09E+00	< 2E-16
19	$\frac{MW}{nC}$	-2.8990E-01	2.61E-02	< 2E-16

Similar to the LFL SFFS-MLR-NLI model in Step 2, Equation 4.15 can be interpreted using the differences between carbon and other atoms' electronegativity. Like the LFL characteristic, lighter organic compounds (the number of C<2) tend to have wider ranges of UFLs than the other compounds. However, the UFL range of these lighter compounds is even wider than the LFL one among the same compounds. Of the lighter compounds, for example, the LFL range is from 1.3 to 14.4 vol.%, whereas the UFL range is from 6.35 to 100.0 vol.%. Such phenomena may be explained through the unusual exothermic decomposition reaction of several low carbon compounds in the absence of air. When the released heat from decomposing a compound is larger than the observed heat for the formation of its products, once external energy is added, the remained heat could trigger a self-sustaining propagating reaction zone. In other words, a deflagration, and in some cases, even a detonation, can propagate without the presence of air. Furthermore, the radical reaction mechanism at low temperatures is complex with many partially oxygenated intermediates as this reaction leads to slowly decreasing combustion heats with decreasing air concentration. These complexities then could cause a relatively

wide range of experimental results and challenges when estimating accurate UFL values. Pan et al. (2009b) also noticed the different explosive manner of the UFL.

Subsequently, the created model (Equation 4.15) was validated by predicting the UFL values in the training and test set. As a result, the predicted UFL values for 1,711 organic compounds in the dataset were computed and presented in the Supplementary Materials. The main resultant statistical metrics are presented in Table 4.15, the correlations between the predicted and observed UFL values in Figure 4.10, and corresponding residual plots are shown in Figure 4.11.

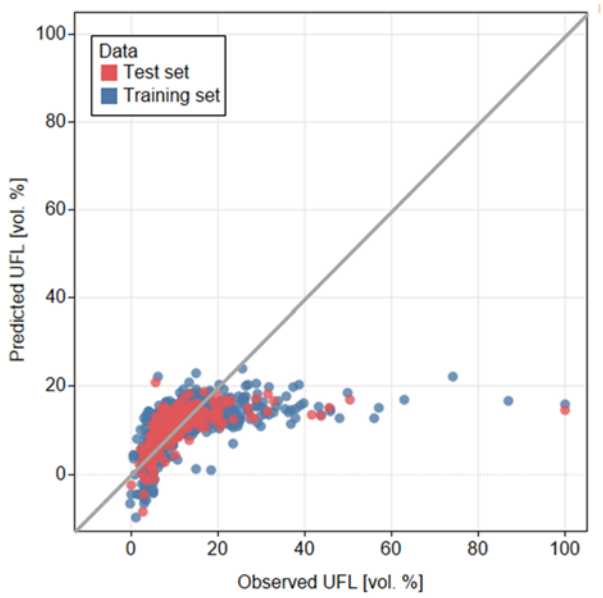
As illustrated in Table 4.15(a), the low R^2 (< 0.5) value indicates that the selected predictor variables in Step 1 are not explaining much in the variation of UFL values, regardless of the variable significance (Table D1). Figure 4.10(a) also indicates that the values that are obtained from Equation 4D1 in Step 1 failed to predict the observed UFLs, because many UFLs points are not close to the regressed diagonal line. Finally, the residual plot of the model, which has a non-flat pattern between -10 and 15 predicted UFLs [vol.%], evidently reveals that essential predictor variables are missing in Equation 4D1 (Figure 4.11(a)). Therefore, we concluded that Equation 4D1 incorporating simple linearity is invalid to predict UFL values.

In contrast, the SFFS-MLR-NLI model of Step 2 (Equation 4.15) generated consistent and high values of $R^2_{training}$, $Q^2_{Loo}, R^2_{test}, and\ Q^2_{ext}$ (> 0.5) than Equation 4D1. As can be seen from Table 4.15(b), the $R^2_{dataset}$ and Q^2_{Loo}, of Eq. 17 is 0.580 and 0.516, respectively, both of which were low for a good model. However, the similar predictive errors (RMSE and MAE) between training and test sets suggests that the

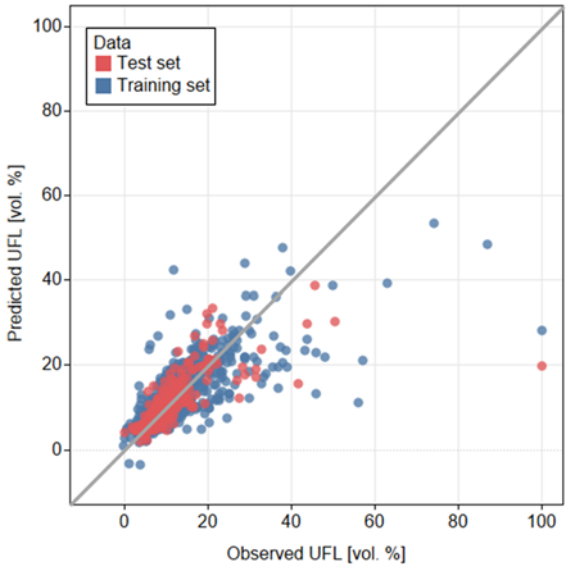
constructed model has both predictive ability and generalization performance. The plot of the predicted UFL values versus the observed ones in the model shows that most UFL points are close to the regressed diagonal line (Figure 4.10(b)), even though there is still room to improve. To determine whether any pattern was missed in the MLR model, residual plots were examined, as shown in Figure 4.11(b)). Since the plot is randomly distributed on both sides of the zero line, we could conclude that there is no systematic error in developing the proposed SFF-MLR-NLI model. Therefore, this model is valid for predicting the UFL values of pure organic compounds.

Table 4.15 Predictability of SFFS-MLR-NLI model for predicting UFL in Step 2 (Equation 4.15)

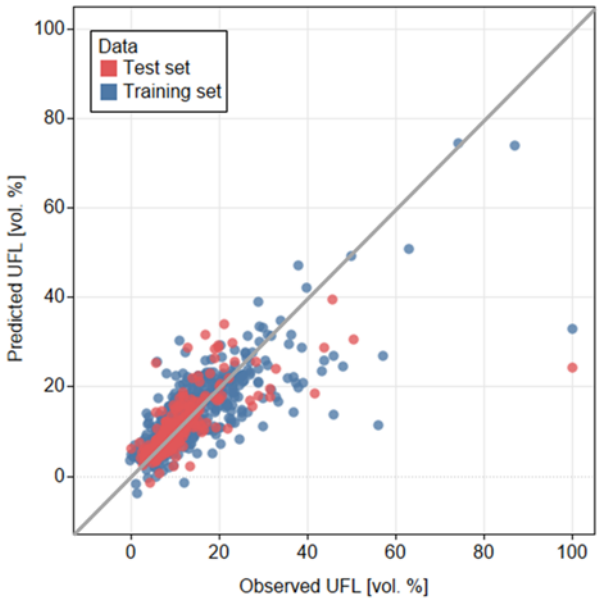
Statistical metric	Training set	Test set	Overall dataset
R^2	0.602	0.501	0.580
RMSE [vol.%]	4.833	5.726	5.023
$RMSE_{Loo}$	5.159	-	-
Q^2_{Loo}	0.549	-	-
Q^2_{ext}	-	0.501	-
MAE	2.439	2.463	2.444
Std. error	4.869	5.733	5.025
Number of compounds	1,370	341	1,711



(a)* Predicted (Equation 4D1) vs. observed UFLs in Step 1 via the SFBS (**this model rejected**)

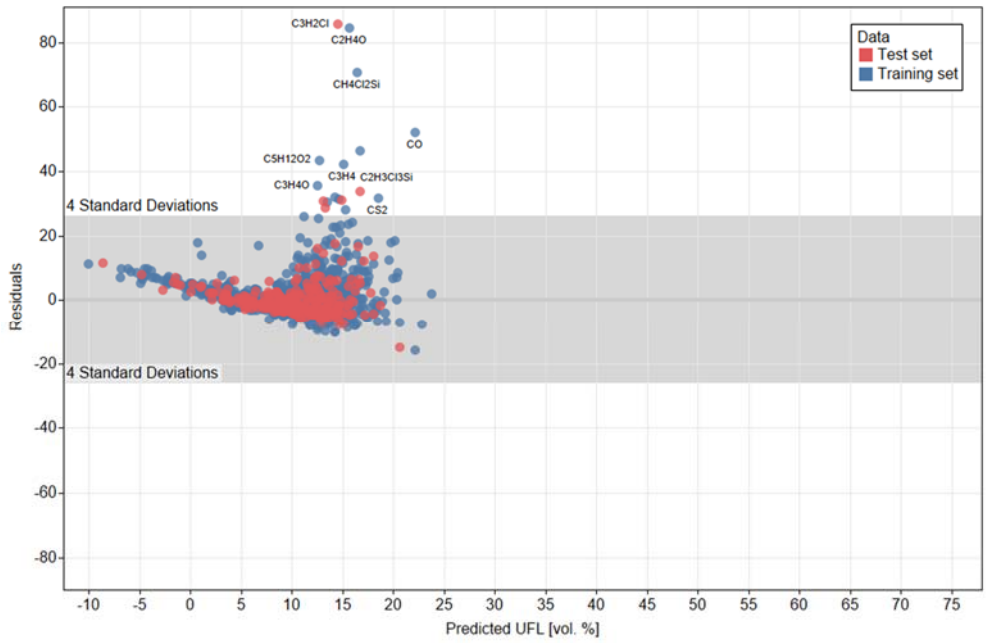


(b) Predicted (Equation 4.15) vs. observed UFLs in Step 2 via the SFFS

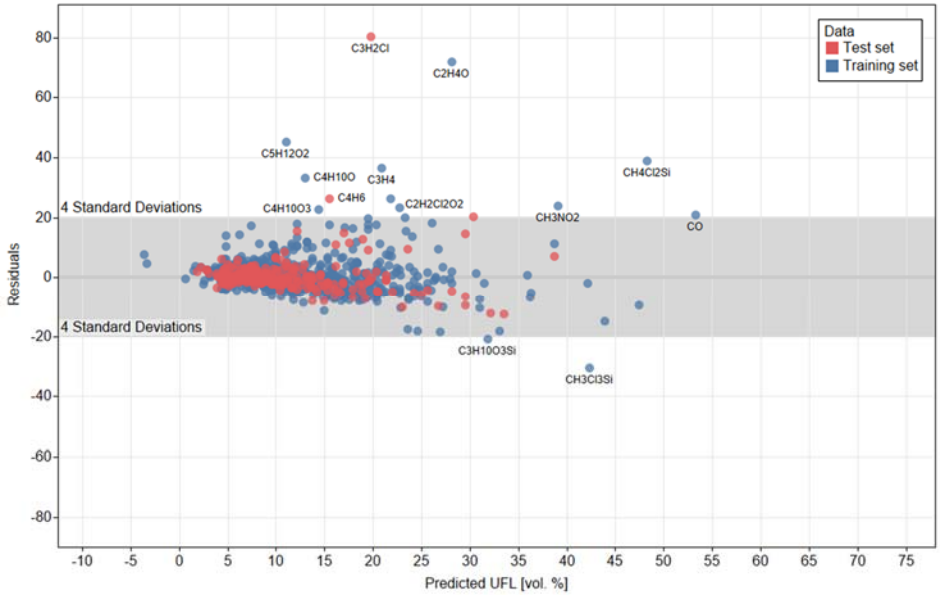


(c)* Predicted (Equation 4D2) vs. observed UFLs in Step 2 via the SFBS

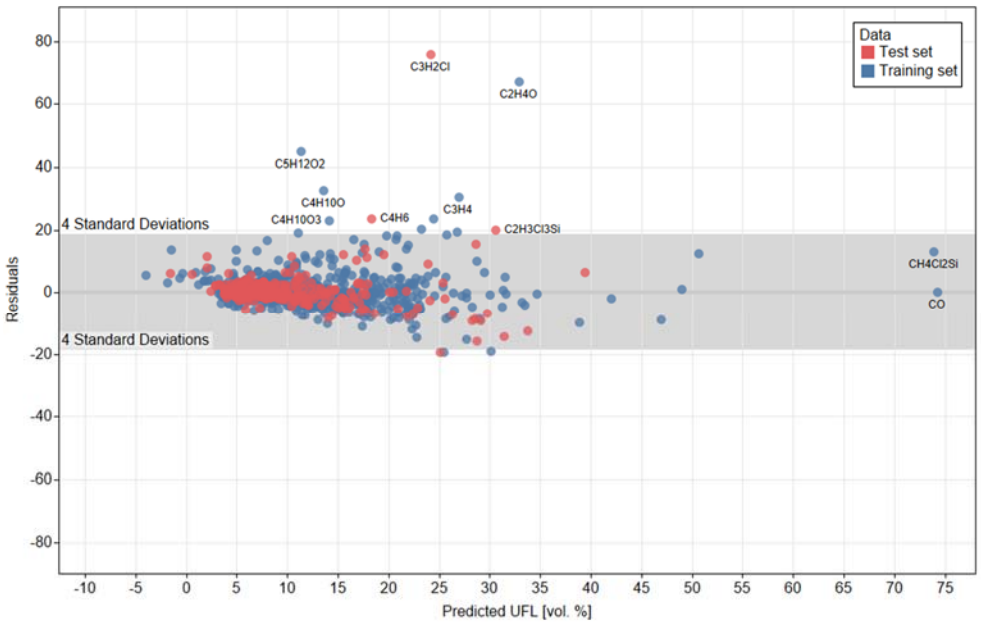
Figure 4.10 Correlations between the predicted and observed UFLs for both training and test sets. The diagonal line represents perfect correlation of observed and predicted UFLs. *The detail of the Graphs (a) and (c) model are additionally provided in Appendix 4D to check whether this method can achieve a more accurate model in this study.



(a)* Residuals (Step 1 for UFL via the SFBS) vs predicted values (Equation 4D1) (**this model rejected**)



(b) Residuals (Step 2 for UFL vis the SFFS) vs predicted values (Equation 4.15)



(c)* Residuals (Step 2 for UFL vis the SFBS) vs predicted values (Equation 4D2)

Figure 4.11 Plots of residuals versus the observed UFL values for the constructed models. *The details of the Graphs (a) and (c) models are additionally provided in Appendix 4D to check whether this method can achieve a more accurate model in this study. This graph shows how the predictabilities gradually increased by adding more flexible terms. The two notable outliers, carbon monoxide (CO) and methyl dichlorosilane (CH_4Cl_2Si) in the SFBS-MLR-L in Step 1 and the SFFS-MLR-L in Step 2, are not the outliers any more in the SFBS-MLR-NLI in Step 2**.**

Figure 4.11 also indicates outliers for the two created UFL models. All outliers (>4s) in Step 1, 17 outliers contain less or equal to four carbon atoms in their molecular formulas. Although many outliers in Step 1 still remain as outliers in Step 2, the average residuals among the outliers considerably decreased from 43.3 vol. % 35.5 vol.%. Moreover, several compounds that are outliers in Step 2 are not outliers any more in Step 2 such as carbon disulfide (CS_2) and dichloroacetic acid ($C_2H_2Cl_2O_2$). Hence, the

predictability increases by capturing the behaviors of atomic interaction among lighter compounds in Equation 4.15, compared to Equation 4D1.

The UFL SFBS-based model in Step 2 is also provided in Appendix 4D. As shown in its statistical results in Appendix 4D, its predicted vs. observed values in Figure 4.10(c), and residual plot in Figure 4.11(c), the SFBS-MLR-NLI model is adequate as a predictive model. In addition, the model has higher accuracy than the other proposed models. However, the model is less applicable than the SFFS-MLR-NLI model because of 57 predictor terms.

Although the proposed UFL SFFS-MLR-NLI model is adequate, the model had rather low predictability: the value of R^2 was 0.58. In order to construct a more generic QSPR model, Pan et al. (2009b) excluded 9 outliers by narrowing down the UFL range from 2.7 – 100.0 vol.% to 2.7 – 50.0 vol.%. Similar to the study of Pan et al. (2009b), we additionally established a more generic UFL model in Step 2 after excluding the SFBS-MLR-NLI model's 15 outliers (dichloroacetic acid, ethylene oxide, ethyleneimine, methylacetylene, propargyl alcohol, trimethylchlorosilane, trimethoxysilane, diethyl ether, diethylene glycol, 1-ethoxy-2-propanol, nitrobenzene, trichlorovinylsilane, 1-chloro-2-propyne, dimethylacetylene, and sucrose acetate isobutyrate). The range of the UFL values was changed from 0.0475 – 100.0 vol.% to 0.0475 – 87.0 vol.%.

By performing the SFFS-MLR-NLI procedure in the training set without the fifteen outliers, the optimum subset of fifteen predictors was selected. The created model for predicting the UFL is presented as the following.

$$\boldsymbol{UFL}\,[\boldsymbol{vol.}\%]$$

$$\approx \boldsymbol{f(nC, nH, nO, nN, nF, nCl, nBr, nSi, nS, MW, interaction\ terms)}$$

$$= \mathbf{3.636} - \mathbf{0.0686}\boldsymbol{nH} + \mathbf{0.00001}(\boldsymbol{MW})^{\mathbf{2}} + \mathbf{0.2266}(\boldsymbol{nO})^{\mathbf{2}} + \mathbf{0.1740}\ (\boldsymbol{nN})^{\mathbf{2}}$$

$$+ \mathbf{0.8261}(\boldsymbol{nS})^{\mathbf{2}} - \mathbf{0.0118}(\boldsymbol{MW} \times \boldsymbol{nO}) + \mathbf{0.1799}(\boldsymbol{nC} \times \boldsymbol{nF}) + \mathbf{0.0746}(\boldsymbol{nH} \times \boldsymbol{nO})$$

$$+ \frac{\mathbf{1}}{\boldsymbol{nC}}(\mathbf{46.62} + \mathbf{5.476}\boldsymbol{nO} + \mathbf{38.47}\boldsymbol{nSi})$$

$$+ \frac{\mathbf{1}}{(\boldsymbol{nC})^{\mathbf{2}}}(-\mathbf{5.749}\boldsymbol{nH} - \mathbf{2.051}(\boldsymbol{nCl})^{\mathbf{2}} - \mathbf{7.024}\boldsymbol{nBr} - \mathbf{0.0009}(\boldsymbol{MW})^{\mathbf{2}})$$

Equation 4.16

The selected predictors' coefficients are statistically significant (p-value < 0.05), as illustrated in Table 4.16. Compared to the performance of Equation 4.15, The new model was established without iodine atom. Also, the same atomic elements involved in Equation 4.15 were contained in this model with rather different transformed terms. However, similar to Equation 4.15, this model also composes various atomic interactions and bulk-related factors (weights).

Table 4.16 Selected predictors and their coefficients of the UFL SFFS-MLR-NLI model (Equation 4.16)

	Variables	Coefficient	Std. error	p-value
0	Intercept	3.6360E+00	3.5110E-01	< 2E-16
1	nH	-6.8570E-02	1.9000E-02	0.000316
2	$(MW)^2$	1.4150E-05	5.3130E-06	0.007798
3	$(nO)^2$	2.2660E-01	3.7230E-02	1.44E-09
4	$(nN)^2$	1.7400E-01	5.1830E-02	0.000806
5	$(nS)^2$	8.2610E-01	2.5880E-01	0.001439
6	$MW \times nO$	-1.1800E-02	1.4620E-03	1.35E-15

	Variables	Coefficient	Std. error	p-value
7	$nC \times nF$	1.7990E-01	4.3850E-02	4.26E-05
8	$nH \times nO$	7.4630E-02	1.0610E-02	2.96E-12
9	$\frac{1}{nC}$	4.6620E+01	1.7020E+00	< 2E-16
10	$\frac{nH}{(nC)^2}$	-5.7490E+00	4.8460E-01	< 2E-16
11	$\frac{nO}{nC}$	5.4760E+00	4.9130E-01	< 2E-16
12	$\frac{(nCl)^2}{(nC)^2}$	-2.0510E+00	4.5990E-01	8.72E-06
13	$\frac{nBr}{(nC)^2}$	-7.0240E+00	3.4880E+00	0.044192
14	$\frac{nSi}{nC}$	3.8470E+01	1.9300E+00	< 2E-16
15	$\frac{(MW)^2}{(nC)^2}$	-9.4040E-04	1.6990E-04	3.63E-08

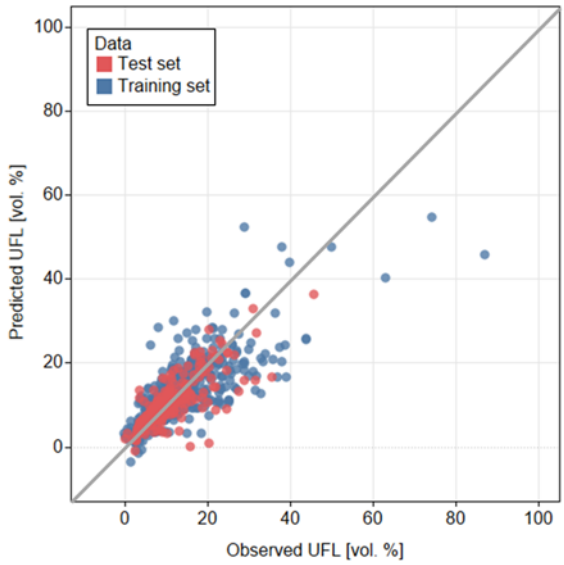
The SFFS-MLR-NLI model (Equation 4.16) without the 15 outliers generated consistent and high values of $R^2_{training}$, $Q^2_{Loo}, R^2_{test}, and\ Q^2_{ext}$ (> 0.5) than Equation 4.15. As can be seen from Table 4.17, the $R^2_{dataset}$ and Q^2_{Loo}, of Equation 4.16 is 0.677 and 0.626, respectively. This result indicates that this model is now a stable model internally and externally. This result clearly shows that the predictive performance significantly increases after excluding the 15 outliers. The plot of the predicted UFL values versus the observed ones in the model shows that most UFL points are close to the regressed diagonal line (Figure 4.12(a)), even though there is still room to improve. To determine whether any pattern was missed in the MLR model, residual plots were examined, as shown in Figure 4.13(a)). Since the plot is randomly distributed on both sides of the zero line, we could conclude that there is no systematic error in developing the

proposed SFFS-MLR-NLI model. Therefore, this model is valid for predicting the generic UFL values of pure organic compounds.

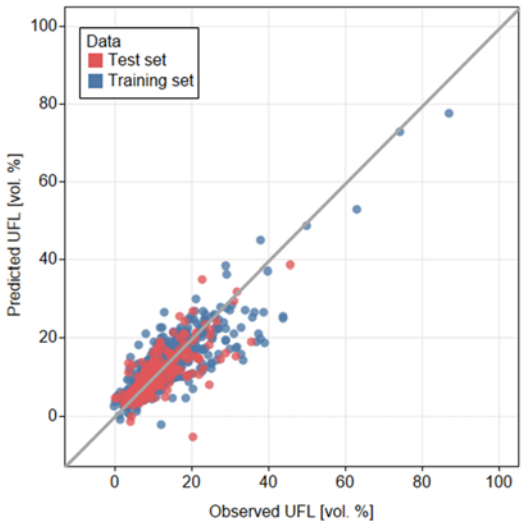
Table 4.17 Predictabilities of the two constructed models for predicting UFL

The resultant statistical parameters of the SFFS-MLR-L model without 15 outliers in Step 2

Statistical metric	Training set	Test set	Overall dataset
R^2	0.684	0.646	0.677
RMSE [vol.%]	3.380	3.155	3.336
$RMSE_{Loo}$	4.107	-	-
Q^2_{Loo}	0.626	-	-
Q^2_{ext}	-	0.646	-
MAE	2.104	2.178	2.119
Std. error	3.791	3.594	3.737
Number of compounds	1,359	337	1,696

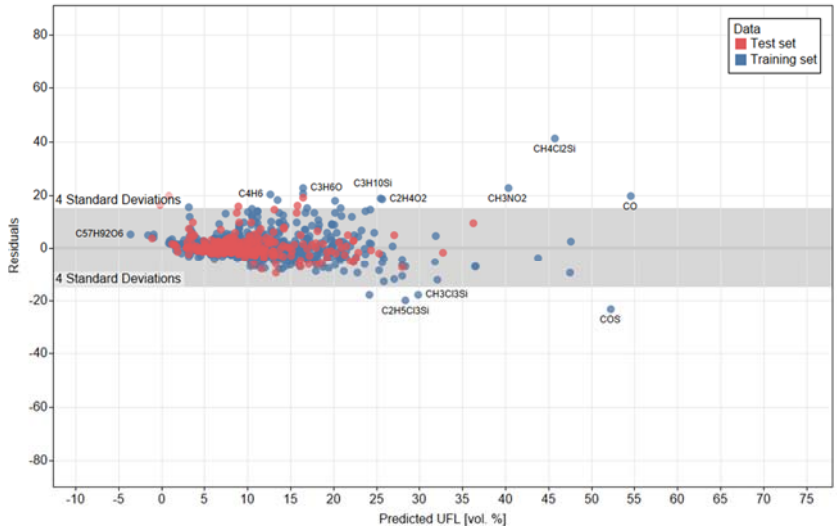


(a) Predicted (Equation 4.16) vs. observed UFLs without 15 outliers in Step 2 via the SFFS

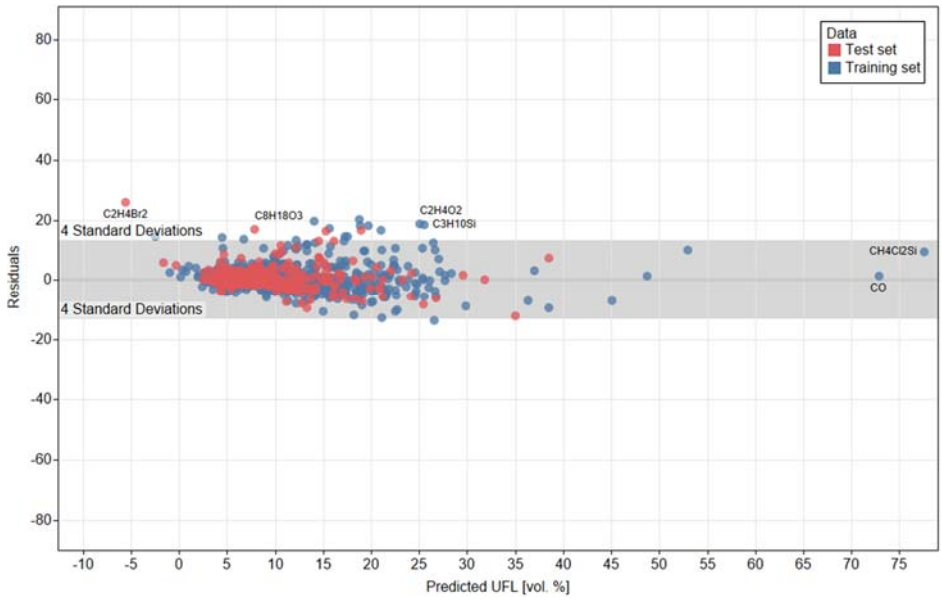


(b)* Predicted (Equation 4D3) vs. observed UFLs without 15 outliers in Step 2 via the SFBS

Figure 4.12 Correlations between the predicted and observed UFLs for both training and test sets. The diagonal line represents perfect correlation of observed and predicted UFLs. *The detail of the Graph (b) model is additionally provided in Appendix 4D to check whether this method can achieve a more accurate model in this study.



(a) Residuals (Step 2 for UFL via the SFFS without 15 outliers) vs predicted values (Equation 4.16)



(b)* Residuals (Step 2 for UFL vis the SFBS with 15 outliers) vs predicted values (Equation 4D3)

Figure 4.13 Plots of residuals versus the observed UFL values for the constructed models. *The detail of the Graph (b) model is additionally provided in Appendix 4D3 to check whether this method can achieve a more accurate model in this study. This graph shows how the predictabilities gradually increased by adding more flexible terms.

Table 4.18 summarizes the performance of the four adequate predictive models. This table also includes two models constructed by QSPR with the same data source, DIPPR. Although the same data source was utilized in all models, it could not be confirmed whether the dataset of UFL models in this study covered the entire dataset the QSPR models used. Nevertheless, statistical metrics such as $higher\ R^2, lower\ MAE[K]$, $lower\ RMSE[K]$, lower standard error, and applicability to a wider variety of chemical compounds for the UFL values indicate that the proposed UFL predictive models in this study seem to come out as being slightly better than the previous models.

Table 4.18 Comparison with the current model and previous model (the UFL of pure organic compounds)

Model	$R^2_{dataset}$ ($R^2_{training}$/ R^2_{test})	MAE [vol.%]	RMSE [vol.%]	Std. error	Dataset Size*	Property range (Applicable range) [vol. %]	Note
Pan et al. (2009b)[1]	- (0.570/0.523)	- (2.80/2.93)	- (6.39/7.46)	-	588 (470/118)	2.7 – 100.0 vol.%	QSPR model (GA-MLR)
Pan et al. (2009b)	- (0.758/0.751)	- (1.85/1.75)	- (2.70/2.77)	-	579 [2] (465/114)	2.7 – 50.0 vol.%	QSPR model (GA-MLR)
Equation 4.15	0.580 (0.602/0.501)	2.444 (2.44/2.46)	5.023 (4.83/5.73)	5.025 (4.87/5.73)	1,711 (1,370/341)	0.0475 – 100.0 vol.%	SFFS-MLR- NLI (Step 1 in this study)
Equation 4D2 (Appendix 4D2)	0.639 (0.679/0.495)	2.329 (2.25/2.64)	4.657 (4.34/5.76)	4.658 (4.44/5.76)	1,711 (1,370/341)	0.0475 – 100.0 vol.%	SFBS-MLR- NLI (Step 2 in this study)
Equation 4.16	0.677 (0.684/0.646)	2.119 (2.10/2.18)	3.336 (3.38/3.16)	3.737 (3.79/3.59)	1,696 [3] (1,359/337)	0.0475 – 87.0 vol.%	SFFS-MLR- NLI (Step 2 in this study)
Equation 4D3 (Appendix 4D3)	0.748 (0.774/0.615)	1.991 (1.92/2.28)	3.303 (3.18/3.75)	3.303 (3.25/3.74)	1,696 [3] (1,359/337)	0.0475 – 87.0 vol.%	SFBS-MLR- NLI (Step 2 in this study)

*All data reference is DIPPR 801

[1] Of several QSPR models, this model has the most similar property ranges for the UFL values from the DIPPR 801 reference.

[2] The 9 outliers with more than 50 vol.% UFL values were excluded.

[3] The 15 outliers, which were far from other data points (>4s) of the SFBS-MLR-NLI model, were excluded.

4.5. Conclusions

It is often challenging for practitioners to estimate flammability properties with currently existing predictive models. Many practitioners' main challenge is the demand for in-depth chemical or computer science knowledge associated with the predictive models. Therefore, this study aimed to propose new reliable predictive models that practitioners can easily adopt.

Based on the data obtained from the DIPPR (2019), this study established MLR models that predict the flash point, heat of combustion, LFL, and UFL values of pure organic compounds. To utilize readily available predictor variables, the numbers of all atoms from their molecular formulas, molecular weights, and boiling points were adopted as 121 default input variables (possible predictors). Of various machine learning algorithms, the MLR algorithm was selected because the algorithms' models would enable process engineers to adopt them quickly. However, the MLR method was generally limited to linear fitting. To overcome this MLR limitation, atomic interactions and transformation terms were additionally prepared by transforming the default input variables. This study consisted of two steps to clearly compare the effects of atomic interactions and transformation terms.

- In Step 1, an optimal subset from the 121 variables, which are exclusively linear terms, was selected with the SFBS feature selection algorithm. Based on the selected predictors, the SFBS-MLR-L models were created for predicting each flammability property. After analyzing the four created models, the flash points and the heats of combustion of organic compounds could be estimated

with the SFBS-MLR-L, with the values of R^2 as 0.976 and 0.999 for test sets, respectively. In contrast to the former models, the LFL and UFL SFBS-MLR-L models were not applicable, with the value of R^2as 0.552 and 0.253 for test sets, respectively. These observations implied that the combinations of simple linear terms were reasonably sufficient at predicting the flash points and the heats of combustion but were not applicable for predicting the LFL and UFL values.

- In Step 2, the default variables in Step 1 were transformed to incorporate the nonlinear and interaction terms together. With the SFFS feature selection, an optimal subset of predictors was selected in Step 2. Similar to Step 1, based on the selected predictors, the SFFS-MLR-NLI models were constructed for each flammability property. Although the models were constructed by the SFBS in Step 2, the SFBS-MLR-NLI models generally showed better accuracies than the SFFS-MLR-NLI models except for heat of combustion prediction. Since the SFBS-MLR-NLI models included numerous predictors, they were not appealing for applications. Therefore, for simplicity and generalization, the SFFS-MLR-NLI models were primarily presented in Step 2. All of the four SFFS-MLR-NLI models were valid, with the values of R^2 as 0.977 (for the flash point), 0.999 (for the heat of combustion), 0.741 (for the LFL), and 0.501 (for the UFL) for each test set. All of the constructed models in Step 2 were adequate as predictive models of the four flammability properties. Given that the proposed models' accuracies had a considerably similar accuracy as the

QSPR models, we concluded that the prepared variables of Step 2 successfully described the various molecular structures to capture the characteristics of flammability properties like the QSPR approaches.

Through the SFBS in Step 1, the SFFS in Step 2, and the SFBS in Step 2, the constructed models were gradually enhanced. Since the flash point SFBS-MLR-L models were already remarkable, there was no significant increase in these models' predictabilities in Step 2. The heat of combustion's SFFS-MLR-NLI model proved exceptionally more accurate and generalizable than its corresponding SFBS-MLR-NLI model due to an extraordinary outlier in the training set. This result indicated that the linear terms with several interaction terms were enough to predict the heat of combustion, and the SFBS-MLR-NLI overfitted the data. Meanwhile, the LFL and UFL models showed substantial improvement in Step 2. Hence, this result indicated that the values of LFL and UFL were determined by various atomic interactions and bulk-related factors, such as the Van der Waals force and steric effects. The heats of combustion of organic compounds were mainly associated with atomic types.

Since the created models in this study are provided with equation forms, they can be utilized as parts of the mathematical formulations in optimization along with other process aspects. For example, the LFL and UFL models can be adopted as part of the formula nonlinear programming (NLP) or mixed-integer nonlinear programming (MINLP) optimization models, while the flash point and heat of combustion with the

linear programming formulations. Through these approaches, flammability properties would proactively be considered during the early design stages.

Despite the ease and the acceptable performance of the predictive models in this study, these models still have two limitations. First, several proposed models utilize another physical property, boiling point, as a predictor variable. However, the boiling point is the most applicable property compared to the target flammability properties, so this limitation may not be a severe barrier. Second, numerous atomic interaction terms are selected, especially when using the SFBS in Step 2, but these interactions are not often statistically significant. This result points to the fact that atomic interactions strongly affect flammability properties, but the interactions between atoms are rather weak and unstable. When an atomic element, for example nitrogen, is selected from the SFBS feature selection, multiple variables are selected as predictors to express the nitrogen atom's effect, such as $nN, (nN)^2, MW \times nN, T_b \times nN, nN \times nH, nN \times nO, nN \times nS, \frac{nN}{nC}, and \left(\frac{nN}{nC}\right)^2$ in predicting the UFL. These multiple terms clearly show that nitrogen affects the UFL values with various interactions (intermolecular and intramolecular forces) with other atoms along with itself. For this reason, numerous predictors should be utilized together to describe interaction effects, in particular via the SFBS in Step 2. Such a circumstance clearly demonstrates the complex relationship between atomic interactions and a flammability property. However, it makes the proposed model via the SFBS in Step 2 seem rather complicated. For this reason, we primarily presented and interpreted the models constructed via the SFFS in Step 2, which were better for generalization internally and externally with fewer predictors. Since this is the first study to clearly incorporate

atomic interaction terms, we could revise more effective input variables that can describe the complicated interaction terms with few predictors.

In conclusion, as all the results discussed above showed, the proposed regression models in this study possess some apparent advantages. First, the number of predictive variables used was readily available. Second, the proposed predictive models are both practical and sufficiently accurate to practitioners. Third, the proposed MLR models will be favored for their ease of interpretability. Therefore, the chemical engineering community could benefit from these results, which would give practitioners the possibility to quickly estimate missing data of the flash point, heat of combustion, LFL, and UFL of organic compounds.

4.6. Supplementary materials

Supplementary material associated with this study can be found, in the online version, at doi: 10.1016/j.compchemeng.2021. 107524.

4.7. Appendix 4A Flash point SFBS-MLR-NLI model in Step 2

$$T_f\ [K] \approx f(the\ numbers\ of\ C, H, O, N, Si, S, F, Cl, T_b, MW, interaction\ terms\ )$$

$$= 1511 + 9.2170nN - 17.49nSi - 0.2634MW + 6.827T_b$$
$$- 173.8\sqrt{T_b} - 118.9(log_{10}nC) - 0.0020MW^2 - 0.0022T_b^2$$
$$+ 0.0416(nH)^2 + 17.09(nN)^2 - 3.4710(nCl)^2 + 7.2740(nS)^2$$
$$+ MW(0.0009T_b + 0.0425nC + 0.0259nF + 0.1194nCl - 0.7915nN)$$
$$+ T_b(-0.0077nC + 0.0018nH + 0.006nO - 0.0661nN - 0.0801nS)$$
$$+ nC(-0.1942nH + 12.26nN + 3.891nS)$$
$$+ nH(-1.681nCl + 0.4372nSi - 0.8864nS)$$
$$+ nO(13.04nN + 2.121nSi + 5.679nS) + nN(19.49nCl + 23.43nSi)$$
$$+ \frac{1}{(nC)^2}\left(525.5 - 184.9\frac{1}{nC}\right) + \frac{nO}{nC}\left(-15.65 + 38.75\frac{1}{nC} - 9.779\frac{nO}{nC}\right)$$
$$+ 5.29\frac{nF}{nC} + \frac{nSi}{(nC)^2}\ (-127.7 + 147.0nSi)$$
$$+ \frac{1}{(T_b \times nC)}\left(-98{,}550 + 3{,}276{,}000\frac{1}{(T_b \times nC)}\right) - 0.4932\frac{T_b}{(nC)^2}$$
$$+ \frac{MW}{nC}\left(0.771 - 0.7427\frac{1}{nC}\right)$$

Equation 4A1

where $nCl, nSi, nC, nBr, nI, nSi, nS, nH$, nO and nN are the numbers of chlorine, silicon, carbon, bromine, iodine, silicone, sulfur, hydrogen, oxygen, and nitrogen. T_b is the normal boiling point [K], and MW is molecular weight [g/mol] of the chemical compound of interest.

Table 4A1 Selected predictors and their coefficients of the flash point SFBS-MLR-NLI model (Equation 4A1)

No	Variables	Coefficient	Std. error	p-value
0	Intercept	1.5110E+03	1.3400E+02	< 2e-16
1	nN	9.2170E+00	3.6690E+00	0.012103
2	nSi	-1.7490E+01	2.7720E+00	3.80E-10
3	MW	-2.6340E-01	9.3060E-02	0.004723
4	T_b	6.8270E+00	4.3620E-01	< 2E-16
5	$\sqrt{T_b}$	-1.7380E+02	1.4020E+01	< 2E-16
6	$\log_{10} C$	-1.1890E+02	1.6400E+01	6.94E-13
7	MW^2	-2.0200E-03	1.6570E-04	< 2E-16
8	T_b^2	-2.1910E-03	1.3660E-04	< 2E-16
9	$(nH)^2$	4.1630E-02	9.3360E-03	8.92E-06
10	$(nN)^2$	1.7090E+01	1.8710E+00	< 2E-16
11	$(nCl)^2$	-3.4710E+00	9.9820E-01	0.000523
12	$(nS)^2$	7.2740E+00	2.5300E+00	0.004097
13	$MW \times T_b$	8.9950E-04	1.9770E-04	5.87E-06
14	$MW \times nC$	4.2480E-02	4.2710E-03	< 2E-16
15	$MW \times nF$	2.5940E-02	5.0490E-03	3.19E-07
16	$MW \times nCl$	1.1940E-01	2.2130E-02	8.06E-08
17	$MW \times nN$	-7.9150E-01	1.1090E-01	1.54E-12
18	$T_b \times nC$	-7.6950E-03	1.6340E-03	2.75E-06
19	$T_b \times nH$	1.8330E-03	3.3710E-04	6.39E-08
20	$T_b \times nO$	6.0160E-03	1.3680E-03	1.19E-05
21	$T_b \times nN$	-6.6080E-02	1.0280E-02	1.79E-10
22	$T_b \times nS$	-8.0120E-02	1.6050E-02	6.80E-07
23	$nC \times nH$	-1.9420E-01	3.3890E-02	1.24E-08
24	$nC \times nN$	1.2260E+01	1.5390E+00	3.41E-15
25	$nC \times nS$	3.8910E+00	7.0680E-01	4.40E-08
26	$nH \times nCl$	-1.6810E+00	2.3920E-01	3.35E-12
27	$nH \times nSi$	4.3720E-01	8.6860E-02	5.48E-07
28	$nH \times nS$	-8.8640E-01	2.4400E-01	0.000291
29	$nO \times nN$	1.3040E+01	1.8100E+00	9.68E-13
30	$nO \times nSi$	2.1210E+00	5.6690E-01	0.000192
31	$nO \times nS$	5.6790E+00	1.8380E+00	0.002048
32	$nN \times nCl$	1.9490E+01	4.5740E+00	2.17E-05

No	Variables	Coefficient	Std. error	p-value
33	$nN \times nSi$	2.3430E+01	5.8760E+00	7.05E-05
34	$\frac{1}{(nC)^2}$	5.2550E+02	9.9600E+01	1.53E-07
35	$\frac{1}{(nC)^3}$	-1.8490E+02	4.7070E+01	9.02E-05
36	$\frac{nO}{nC}$	-1.5650E+01	5.3440E+00	0.003468
37	$\frac{nO}{(nC)^2}$	3.8750E+01	7.7380E+00	6.24E-07
38	$\frac{(nO)^2}{(nC)^2}$	-9.7790E+00	2.5710E+00	0.000149
39	$\frac{nF}{nC}$	5.2900E+00	1.8390E+00	0.004077
40	$\frac{nSi}{(nC)^2}$	-1.2770E+02	5.4340E+01	0.018967
41	$\frac{(nSi)^2}{(nC)^2}$	1.4700E+02	5.6680E+01	0.009596
42	$\frac{1}{(T_b \times nC)}$	-9.8550E+04	1.4600E+04	2.17E-11
43	$\frac{1}{(T_b \times nC)^2}$	3.2760E+06	6.3310E+05	2.64E-07
44	$\frac{T_b}{(nC)^2}$	-4.9320E-01	9.1300E-02	7.78E-08
45	$\frac{MW}{nC}$	7.7100E-01	2.1810E-01	0.00042
46	$\frac{MW}{(nC)^2}$	-7.4270E-01	1.9510E-01	0.000147

Table 4A2 Predictable capability of the developed flash point model in Step 2

Statistical metric	Training set	Test set	Entire dataset
R^2	0.986	0.976	0.984
$RMSE$	9.59	12.43	10.22
$RMSE_{Loo}$	12.03	-	-
Q^2_{Loo}	0.976	-	-
Q^2_{ext}	-	0.976	-
MAE	5.49	6.41	5.67
Std. error	9.76	12.45	10.22
$Number\ of\ compounds$	1,394	347	1,741

4.8. Appendix 4B Heat of combustion SFBS-MLR-NLI model in Step 2

By performing the SFBS-MLR-NLI procedure in the training set, the optimum subset of 64 predictors was selected in Step 2. The constructed model for predicting the heat of combustion is presented as the following:

$$\Delta H_C^o \left[\frac{kJ}{mol}\right]$$

$$\approx f(the\ numbers\ of\ C, H, O, N, F, Cl, Si, S, P, Pb, T_b, MW, interaction\ terms)$$

$$= -7{,}627 - 67.58nH + 617.9nO + 328.1nN + 525.2nF + 1{,}401nCl + 3{,}382nBr$$

$$+ 187.6nSi - 532.5nP - 30.52MW - 87.9\sqrt{MW} + 4{,}068\ exp(nI) + 3{,}535\ exp(nS)$$

$$- 3.094(nC)^2 - 1.341(nH)^2 - 36.3(nO)^2 + 32.14(nF)^2 - 56.34(nCl)^2$$

$$- 5{,}266(nS)^2$$

$$+ MW(0.0056T_b + 1.17nH + 2.265nO - 1.561nF + 0.8412nCl + 1.892nBr$$

$$+ 18.67nI) - T_b(0.0422nH - 0.8717nF + 2.783nBr + 9.356nI)$$

$$- nC(12.25nH + 26.91nO + 27.3nCl - 227.4nI + 16.79nS)$$

$$- nH(21.17nO + 13.04nF + 46.96nCl + 87.35nBr + 163.6nI + 16.18nN$$

$$+ 30.22nSi + 34.58nS)$$

$$- nO(85.9nF + 124.3nCl + 68.49nN + 88.83nSi + 110.8nS) - 135.5(nN \times nCl)$$

$$- \frac{1}{nC}\left(1{,}558 - 980\frac{1}{nC} + 1{,}198nCl + 2{,}617nBr + 3{,}999nI + 995.7nS + 861.0ni\right)$$

$$+ \frac{nH}{nC}\left(127 - 154.6\frac{1}{nC}\right) - \frac{nO}{(nC)^2}(255.7 + 109.7nO) - \frac{1}{(nC)^2}(388nF + 396.2nN)$$

$$+ \frac{36{,}140}{(T_b \times nC)} + 0.003\frac{(T_b)^2}{(nC)^2} + \frac{MW}{nC}\left(37.89 - 0.0341\frac{MW}{nC}\right) \qquad \text{Equation 4B1}$$

Table 4B1. Selected predictors and their coefficients of the heat of combustion SFBS-MLR-NLI model (Equation 4B1)

No.	Variables	Coefficient	Std. error	p-value
0	Intercept	-7.6270E+03	5.9340E+02	< 2E-16
1	nH	-6.7580E+01	6.6440E+00	< 2E-16
2	nO	6.1790E+02	1.1670E+01	< 2E-16
3	nN	3.2810E+02	1.6550E+01	< 2E-16
4	nF	5.2520E+02	2.8800E+01	< 2E-16
5	nCl	1.4010E+03	5.7460E+01	< 2E-16
6	nBr	3.3820E+03	1.8770E+02	< 2E-16
7	nSi	1.8760E+02	3.0050E+01	5.7000E-10
8	nP	-5.3250E+02	7.9050E+01	2.3600E-11
9	MW	-3.0520E+01	1.3630E+00	< 2E-16
10	$\sqrt{MW}$	-8.7900E+01	2.6170E+01	8.0300E-04
11	exp (nI)	4.0680E+03	5.0060E+02	9.5500E-16
12	exp (nS)	3.5350E+03	2.8280E+02	< 2E-16
13	$(nC)^2$	-3.0940E+00	4.7720E-01	1.2200E-10
14	$(nH)^2$	-1.3410E+00	1.1870E-01	< 2E-16
15	$(nO)^2$	-3.6300E+01	9.8340E+00	2.3200E-04
16	$(nF)^2$	3.2140E+01	4.5210E+00	1.8500E-12
17	$(nCl)^2$	-5.6340E+01	1.2490E+01	7.0400E-06
18	$(nS)^2$	-5.2660E+03	4.2860E+02	< 2E-16
19	$MW \times T_b$	5.6390E-03	1.0530E-03	1.0100E-07
20	$MW \times nH$	1.1700E+00	4.6470E-02	< 2E-16
21	$MW \times nO$	2.2650E+00	6.1420E-01	2.3500E-04
22	$MW \times nF$	-1.5610E+00	2.0580E-01	5.8100E-14
23	$MW \times nCl$	8.4120E-01	3.5010E-01	1.6416E-02
24	$MW \times nBr$	1.8920E+00	5.5110E-01	6.1300E-04
25	$T_b \times nH$	-4.2150E-02	1.1100E-02	1.5300E-04
26	$T_b \times nF$	8.7170E-01	1.0770E-01	1.2500E-15
27	$T_b \times nBr$	-2.7830E+00	5.7490E-01	1.4400E-06
28	$T_b \times nI$	-9.3560E+00	2.6970E+00	5.3900E-04
29	$nC \times nH$	-1.2250E+01	8.4260E-01	< 2E-16
30	$nC \times nO$	-2.6910E+01	7.5020E+00	3.4600E-04
31	$nC \times nCl$	-2.7300E+01	6.7180E+00	5.0900E-05
32	$nC \times nI$	2.2740E+02	8.3660E+01	6.6370E-03
33	$C \times nS$	-1.6790E+01	5.4720E+00	2.2010E-03
34	$nH \times nO$	-2.1170E+01	1.0500E+00	< 2E-16
35	$nH \times nF$	-1.3040E+01	2.6040E+00	6.1800E-07
36	$nH \times nCl$	-4.6960E+01	3.1450E+00	< 2E-16
37	$nH \times nBr$	-8.7350E+01	4.9930E+00	< 2E-16

No.	Variables	Coefficient	Std. error	p-value
38	$nH \times nI$	-1.6360E+02	1.9000E+01	< 2E-16
39	$nH \times nN$	-1.6180E+01	1.0440E+00	< 2E-16
40	$nH \times nSi$	-3.0220E+01	1.6200E+00	< 2E-16
41	$nH \times nS$	-3.4580E+01	3.3510E+00	< 2E-16
42	$nO \times nF$	-8.5900E+01	1.3270E+01	1.3100E-10
43	$nO \times nCl$	-1.2430E+02	2.3830E+01	2.1000E-07
44	$nO \times nN$	-6.8490E+01	8.9180E+00	2.9400E-14
45	$nO \times nSi$	-8.8830E+01	1.8420E+01	1.5700E-06
46	$nO \times nS$	-1.1080E+02	2.2550E+01	1.0100E-06
47	$nN \times nCl$	-1.3550E+02	2.9040E+01	3.3700E-06
48	$\frac{1}{nC}$	-1.5580E+03	2.7380E+02	1.5400E-08
49	$\frac{1}{(nC)^2}$	9.8090E+02	1.7340E+02	1.8700E-08
50	$\frac{nH}{nC}$	1.2760E+02	2.8200E+01	6.5800E-06
51	$\frac{nH}{(nC)^2}$	-1.5460E+02	3.4600E+01	8.5600E-06
52	$\frac{nO}{(nC)^2}$	-2.5570E+02	6.0630E+01	2.6300E-05
53	$\frac{(nO)^2}{(nC)^2}$	-1.0970E+02	1.5850E+01	6.8600E-12
54	$\frac{nF}{(nC)^2}$	-3.8800E+02	3.3180E+01	< 2E-16
55	$\frac{nCl}{nC}$	-1.1980E+03	9.4690E+01	< 2E-16
56	$\frac{nBr}{nC}$	-2.6170E+03	1.6380E+02	< 2E-16
57	$\frac{nI}{nC}$	-3.9990E+03	3.9530E+02	< 2E-16
58	$\frac{nS}{nC}$	-9.9570E+02	1.0510E+02	< 2E-16
59	$\frac{nSi}{nC}$	-8.6100E+02	9.8410E+01	< 2E-16
60	$\frac{nN}{(nC)^2}$	-3.9620E+02	7.4940E+01	1.4400E-07
61	$\frac{1}{(T_b \times nC)}$	3.6140E+04	1.5910E+04	2.3222E-02
62	$\frac{(T_b)^2}{(nC)^2}$	2.9630E-03	8.8680E-04	8.5500E-04

No.	Variables	Coefficient	Std. error	p-value
63	$\frac{MW}{nC}$	3.7890E+01	2.9190E+00	< 2E-16
64	$\frac{(MW)^2}{(nC)^2}$	-3.4050E-02	7.8200E-03	1.4300E-05

Table 4B2. The resultant statistical parameters of the heat of combustion SFBS-MLR-NLI model in Step 2

Statistical metric	Training set	Test set	Overall dataset
R^2	0.9992	0.9992	0.9991
RMSE [kJ/mol]	105.50	104.35	105.28
$RMSE_{Loo}$[kJ/mol]	126.36	-	-
Q^2_{Loo}	0.9989	-	-
Q^2_{ext}	-	0.9991	-
MAE [kJ/mol]	61.68	62.63	61.87
Std. error	107.9	104.16	105.29
Number of compounds	1,482	368	1,850

4.9. Appendix 4C1 LFL SFBS-MLR-L model in Step 1 (rejected)

$$LFL\ [vol.\%] \approx f(C, O, N, F, Cl, Br, Si, T_b)$$

$$= 3.4833 - 0.0453nC + 0.2273nO + 0.2622nN + 0.7860nF$$

$$+ 0.9797nCl + 1.0533nBr - 0.2144nSi - 0.0046T_b$$

Equation 4C1 (the model rejected)

Table 4C1. Selected predictors and their coefficients of the LFL SFBS-MLR-L model (Equation 4C1, the model rejected)

No.	Variables	Coefficient	Std. error	p-value	VIF
0	Intercept	3.4833E+00	1.6627E-01	< 2E-16	-
1	nC	-4.5279E-02	8.0008E-03	1.85E-08	2.863
2	nF	7.8599E-01	5.3792E-02	< 2E-16	1.038
3	nCl	9.7969E-01	5.7562E-02	< 2E-16	1.077
4	nBr	1.0533E+00	2.2458E-01	3.00E-06	1.014
5	nN	2.6219E-01	5.5235E-02	2.28E-06	1.241
6	nO	2.2732E-01	2.2331E-02	< 2E-16	1.639
7	nSi	-2.1439E-01	4.8017E-02	8.66E-06	1.135
8	T_b	-4.5567E-03	4.9040E-04	< 2E-16	3.537

Table 4C2. Predictable capability of the developed LFL SFBS-MLR-L model in Step 1 (Equation 4C1, the model rejected)

Statistical metric	Training set	Test set	Overall dataset
R^2	0.441	0.552	0.463
RMSE [vol.%]	1.101	0.974	1.077
$RMSE_{Loo}$ [vol.%]	1.120	-	-
Q^2_{Loo}	0.422	-	-
Q^2_{ext}	-	0.552	-
MAE [vol.%]	0.586	0.534	0.57
Std. error	1.104	0.975	1.077
Number of compounds	1,388	345	1,733

4.10. Appendix 4C2 LFL SFBS-MLR-NLI model in Step 2

$$
\begin{aligned}
&LFL\ [vol.\%] \\
&\approx f(the\ numbers\ of\ H, C, F, Br, O, N, Si, Si, Cl, I, P, MW, T_b, interaction\ terms,) \\
&= 187.7 + 0.1148nH - 00809nC - 4.878nF - 0.0695MW - 0.1267T_b \\
&+ 1.272\sqrt{MW} + 10.92\sqrt{T_b} - 145.7Log_{10}(T_b) - 0.0026exp(nF) + 16.62exp(nBr) \\
&- 0.0006MW^2 - 0.1444(nC)^2 - 0.0009(nH)^2 + 0.1652(nO)^2 + 0.1829(nN)^2 \\
&+ 0.528(nF)^2 - 19.81(nBr)^2 - 0.223(nSi)^2 + 0.5779(nS)^2 \\
&+ MW(0.00004T_b + 0.0199nC + 0.0014nH + 0.0107nF + 0.0205nCl + 0.0726nI \\
&- 0.0043nN + 0.0214nSi) + T_b(-0.0152nBr + 0.0009nSi) \\
&- nC(0.024nH + 0.0758nO + 0.3785nCl + 1.48nI + 0.4329nSi + 0.1463nS) \\
&+ nH(0.1718nF - 0.0928nCl) \\
&+ nO(0.6683nF + 0.6741nCl + 0.3221nN + 0.2726nSi + 0.6509nS) + 30.46\frac{1}{nC} \\
&+ \frac{nH}{nC}\left(-1.516 + 0.1474\frac{nH}{nC}\right) + \frac{nF}{nC}\left(9.43 - 2.426\frac{nF}{nC}\right) + 1.459\frac{nCl}{nC} - 5.51\frac{nS}{(nC)^2} \\
&- 8.601\frac{nSi}{(nC)^2} - 2.1\frac{nN}{(nC)^2} + \frac{nP}{nC}\left(24.06 - 535.36\frac{nP}{nC}\right) \\
&+ \frac{1}{(T_b \times nC)}\left(-6455 + 262{,}500\frac{1}{(T_b \times nC)}\right) - 0.02\frac{T_b}{(nC)^2} + \frac{55.5}{(MW \times nC)} \\
&+ \frac{MW}{(nC)^2}(0.1108 - 0.0006MW)
\end{aligned}
$$

Equation 4C4

Table 4C3. Selected predictors and their coefficients of the LFL SFBS-MLR-NLI model (Equation 4C2)

No.	Variables	Coefficient	Std. error	p-value
0	Intercept	1.8770E+02	2.5950E+01	7.89E-13
1	nH	1.1480E-01	1.2330E-02	< 2E-16
2	nC	-8.0920E-02	3.2910E-02	0.014078
3	nF	-4.8780E+00	3.5400E-01	< 2E-16
4	MW	-6.9530E-02	8.3570E-03	< 2E-16
5	T_b	-1.2670E-01	1.9560E-02	1.32E-10
6	$\sqrt{MW}$	1.2720E+00	1.3170E-01	< 2E-16
7	$\sqrt{T_b}$	1.0920E+01	1.5830E+00	7.94E-12
8	$\log_{10} T_b$	-1.4570E+02	1.9410E+01	1.10E-13
9	$exp\ (nF)$	-2.6390E-03	1.8690E-04	< 2E-16
10	$exp\ (nBr)$	1.6620E+01	2.7080E+00	1.10E-09
11	MW^2	-6.3140E-04	7.3760E-05	< 2E-16
12	$(nC)^2$	-1.4440E-01	1.7060E-02	< 2E-16
13	$(nH)^2$	-9.3220E-04	4.4160E-04	0.03495
14	$(nO)^2$	1.6520E-01	1.8850E-02	< 2E-16
15	$(nN)^2$	1.8290E-01	2.4410E-02	1.21E-13
16	$(nF)^2$	5.2800E-01	3.8880E-02	< 2E-16
17	$(nBr)^2$	-1.9810E+01	3.7470E+00	1.44E-07
18	$(nSi)^2$	-2.2300E-01	7.9440E-02	0.005079
19	$(nS)^2$	5.7790E-01	9.1260E-02	3.30E-10
20	$MW \times T_b$	3.7820E-05	6.8790E-06	4.63E-08
21	$MW \times nC$	1.9940E-02	2.2480E-03	< 2E-16
22	$MW \times nH$	1.4210E-03	1.7560E-04	1.31E-15
23	$MW \times nF$	1.0730E-02	1.2460E-03	< 2E-16
24	$MW \times nCl$	2.0530E-02	2.5690E-03	2.87E-15
25	$MW \times nI$	7.2580E-02	1.0280E-02	2.65E-12
26	$MW \times nN$	-4.3130E-03	7.7900E-04	3.71E-08
27	$MW \times nSi$	2.1440E-02	3.4290E-03	5.40E-10
28	$T_b \times nBr$	-1.5170E-02	2.6470E-03	1.23E-08
29	$T_b \times nSi$	9.3670E-04	2.7600E-04	0.000709
30	$nC \times nH$	-2.4030E-02	2.7390E-03	< 2E-16
31	$nC \times nO$	-7.5820E-02	1.1840E-02	2.11E-10
32	$nC \times nCl$	-3.7850E-01	5.0300E-02	9.77E-14
33	$nC \times nI$	-1.4800E+00	2.7890E-01	1.31E-07
34	$nC \times nSi$	-4.3290E-01	6.6410E-02	1.01E-10
35	$nC \times nS$	-1.4630E-01	2.4930E-02	5.52E-09
36	$nH \times nF$	1.7180E-01	1.9170E-02	< 2E-16

No.	Variables	Coefficient	Std. error	p-value
37	$nH \times nCl$	-9.2810E-02	1.1100E-02	< 2E-16
38	$nO \times nF$	6.6830E-01	5.4340E-02	< 2E-16
39	$nO \times nCl$	6.7410E-01	5.3660E-02	< 2E-16
40	$nO \times nN$	3.2210E-01	4.2230E-02	4.59E-14
41	$nO \times nSi$	2.7150E-01	4.9670E-02	5.48E-08
42	$nO \times nS$	6.5090E-01	9.2140E-02	2.60E-12
43	$\frac{1}{nC}$	3.0460E+01	1.8710E+00	< 2E-16
44	$\frac{nH}{nC}$	-1.5160E+00	1.4970E-01	< 2E-16
45	$\frac{(nH)^2}{(nC)^2}$	1.4740E-01	3.1780E-02	3.85E-06
46	$\frac{nF}{nC}$	9.4300E+00	6.2850E-01	< 2E-16
47	$\frac{(nF)^2}{(nC)^2}$	-2.4260E+00	1.8840E-01	< 2E-16
48	$\frac{nCl}{nC}$	1.4590E+00	1.6540E-01	< 2E-16
49	$\frac{nS}{(nC)^2}$	-5.5100E+00	2.3380E-01	< 2E-16
50	$\frac{nSi}{(nC)^2}$	-8.6010E+00	4.9210E-01	< 2E-16
51	$\frac{nN}{(nC)^2}$	-2.1000E+00	2.4700E-01	< 2E-16
52	$\frac{nP}{nC}$	2.4060E+01	4.0590E+00	3.93E-09
53	$\frac{(nP)^2}{(nC)^2}$	-5.5360E+01	1.1970E+01	4.12E-06
54	$\frac{1}{(T_b \times nC)}$	-6.4550E+03	6.2370E+02	< 2E-16
55	$\frac{1}{(T_b \times nC)^2}$	2.6250E+05	2.4810E+04	< 2E-16
56	$\frac{T_b}{(nC)^2}$	-2.0050E-02	3.0000E-03	3.39E-11
57	$\frac{1}{(MW \times nC)}$	5.5500E+01	2.2680E+01	0.014537
58	$\frac{MW}{(nC)^2}$	1.1080E-01	1.9420E-02	1.44E-08
59	$\frac{(MW)^2}{(nC)^2}$	-5.6770E-04	1.0080E-04	2.14E-08

Table 4C4. Predictable capability of the developed LFL SFBS-MLR-NLI model in Step 2

Statistical metric	Training set	Test set	Overall dataset
R^2	0.939	0.894	0.930
RMSE [kJ/mol]	0.365	0.473	0.389
$RMSE_{Loo}$[kJ/mol]	0.566	-	-
Q^2_{Loo}	0.859	-	-
Q^2_{ext}	-	0.894	-
MAE [kJ/mol]	0.188	0.218	0.188
Std. error	0.373	0.472	0.389
Number of compounds	1,388	345	1,733

4.11. Appendix 4D1 UFL SFBS-MLR-NLI model in Step 1 (rejected)

$$UFL\ [vol.\%] \approx f(nH, nO, nN, nF, nCl, nS, T_b)$$

$$= 23.0538 - 0.115nH + 1.6175nO + 2.0730nN + 0.9481nF$$

$$+ 1.6571nCl + 2.5315nS - 0.0301\ T_b$$

Equation 4D1 (the model rejected)

Table 4D1. Selected predictors and their coefficients of the UFL SFBS-MLR-L model (Equation 4D1, the model rejected)

No.	Variables	Coefficient	Std. error	p-value	VIF
0	Intercept	2.3423E+01	9.0410E-01	< 2E-16	-
1	nH	-1.1343E-01	2.1116E-02	1.43E-08	1.755
2	nN	2.0397E+00	3.3660E-01	1.81E-10	1.146
3	nO	1.5283E+00	1.3716E-01	< 2E-16	1.386
4	nS	2.4082E+00	7.3911E-01	0.000281	1.044
5	nF	9.1756E-01	3.1154E-01	0.001629	1.045
6	nCl	1.8287E+00	3.7281E-01	4.67E-06	1.078
7	T_b	-3.0734E-02	2.4110E-03	< 2E-16	2.119

Table 4D2. Predictable capability of the developed UFL SFBS-MLR-L model in Step 1 (Equation 4D1, the model rejected)

Statistical metric	Training set	Test set	Overall dataset
R^2	0.284	0.253	0.287
RMSE [vol.%]	6.628	7.002	6.546
$RMSE_{Loo}$	6.484	-	-
Q^2_{Loo}	0.284	-	-
Q^2_{ext}	-	0.253	-
MAE	3.726	3.650	3.711
Std. error	6.446	7.012	6.548
Number of compounds	1,370	341	1,711

4.12. Appendix 4D2 UFL SFBS-MLR-NLI model in Step 2

$$
\begin{aligned}
&UFL\ [vol.\%] \\
&\approx f(the\ numbers\ of\ H, C, F, Br, O, N, Si, Si, Cl, I, P, MW, T_b, interaction\ terms,) \\
&= -2{,}937.0 - 3.298nO - 5.838nF - 12.86nBr + 4.663T_b - 19.38\sqrt{nC} \\
&+ 5.874\sqrt{MW} - 276.9\sqrt{T_b} + 2{,}643(\log_{10} T_b) - 0.0391\exp(nN) - 15.41\exp(nBr) \\
&- 126.2\exp(nS) + 0.0033MW^2 - 0.0008T_b^2 + 0.465(nC)^2 + 0.0109(nH)^2 \\
&- 0.4797(nO)^2 - 0.7709(nF)^2 - 5.316(nSi)^2 + 192.5(nS)^2 \\
&- MW(0.0788nC + 0.0177nH + 0.013nO + 0.0992nCl + 0.0359nN + 0.3497nS) \\
&+ T_b(-0.382nI + 0.0856nS) + nC(0.2222nH + 16.67nI + 2.065nS) \\
&+ nH(0.2925nO + 1.194nBr + 0.4011nN + 0.5054nSi + 0.9694nS) \\
&- nO(1.763nF + 1.514nSi) + \frac{1}{(nC)^2}\left(-292.9 + \frac{92.02}{nC}\right) - 4.012\frac{nH}{nC} + 17.35\frac{nO}{(nC)^2} \\
&+ \frac{1}{nC}(7.835nF + 128.6nI) + \frac{nBr}{(nC)^2}(-87.24 + 116.3nBr) + 10.62\frac{(nS)^2}{(nC)^2} \\
&+ 93.1\frac{nSi}{(nC)^2} + \frac{nN}{nC}\left(-24.08 + \frac{20.73}{nC} + 11.51\frac{nN}{nC}\right) + \frac{nP}{nC}\left(-154.9 + 376.5\frac{nP}{nC}\right) \\
&+ 1{,}274{,}000\frac{1}{(T_b \times nC)^2} + \frac{1}{(MW \times nC)}\left(6{,}204 - 56{,}420\frac{1}{(MW \times nC)}\right) \\
&+ \frac{MW}{(nC)^2}(2.607 - 0.0161MW)
\end{aligned}
$$

Equation 4D2

Table 4D3. Selected predictors and their coefficients of the UFL SFBS-MLR-NLI model (Equation 4D2)

	Variables	Coefficient	Std. error	p-value
0	Intercept	-2.9370E+03	5.9260E+02	8.11E-07
1	nO	-3.2980E+00	5.2960E-01	6.39E-10
2	nF	-5.8380E+00	1.2440E+00	2.97E-06
3	nBr	-1.2860E+01	5.2160E+00	0.013815
4	T_b	4.6630E+00	1.0760E+00	1.57E-05
5	$\sqrt{nC}$	-1.9380E+01	1.7610E+00	< 2E-16
6	$\sqrt{MW}$	5.8740E+00	5.1440E-01	< 2E-16
7	$\sqrt{T_b}$	-2.7690E+02	5.9750E+01	3.93E-06
8	$\log_{10} T_b$	2.6430E+03	5.3470E+02	8.64E-07
9	$exp\ (nN)$	-3.9050E-02	1.7790E-02	0.028348
10	$exp\ (nBr)$	-1.5410E+01	2.7640E+00	2.98E-08
11	$exp\ (nS)$	-1.2620E+02	2.2430E+01	2.26E-08
12	MW^2	3.2790E-03	2.9400E-04	< 2E-16
13	T_b^2	-7.6890E-04	2.0260E-04	0.000154
14	$(nC)^2$	4.6500E-01	5.3000E-02	< 2E-16
15	$(nH)^2$	1.0920E-02	5.0490E-03	0.0307
16	$(nO)^2$	-4.7970E-01	9.8360E-02	1.21E-06
17	$(nF)^2$	-7.7090E-01	1.4590E-01	1.47E-07
18	$(nSi)^2$	-5.3160E+00	6.3560E-01	< 2E-16
19	$(nS)^2$	1.9250E+02	3.3390E+01	1.02E-08
20	$MW \times nC$	-7.8750E-02	6.6570E-03	< 2E-16
21	$MW \times nH$	-1.7730E-02	2.5670E-03	7.67E-12
22	$MW \times nO$	-1.3010E-02	6.2280E-03	0.036935
23	$MW \times nCl$	-9.9190E-02	9.9430E-03	< 2E-16
24	$MW \times nN$	-3.5920E-02	5.9400E-03	1.92E-09
25	$MW \times nS$	-3.4970E-01	4.7750E-02	4.19E-13
26	$T_b \times nI$	-3.8200E-01	5.1260E-02	1.67E-13
27	$T_b \times nS$	8.5600E-02	1.9610E-02	1.37E-05
28	$nC \times nH$	2.2220E-01	3.2210E-02	8.19E-12
29	$nC \times nI$	1.6670E+01	3.7650E+00	1.04E-05
30	$nC \times nS$	2.0650E+00	4.9900E-01	3.72E-05
31	$nH \times nO$	2.9250E-01	4.1620E-02	3.35E-12
32	$nH \times nBr$	1.1940E+00	2.9430E-01	5.22E-05
33	$nH \times nN$	4.0110E-01	5.8190E-02	8.48E-12
34	$nH \times nSi$	5.0540E-01	9.3840E-02	8.52E-08
35	$nH \times nS$	9.6940E-01	1.6400E-01	4.33E-09
36	$nO \times nF$	-1.7630E+00	4.6350E-01	0.000149

	Variables	Coefficient	Std. error	p-value
37	$nO \times nSi$	-1.5140E+00	3.9350E-01	0.000124
38	$\frac{1}{(nC)^2}$	-2.9290E+02	3.6700E+01	3.10E-15
39	$\frac{1}{(nC)^3}$	9.2020E+01	1.7850E+01	2.90E-07
40	$\frac{nH}{nC}$	-4.0120E+00	3.6710E-01	< 2E-16
41	$\frac{nO}{(nC)^2}$	1.7350E+01	1.7960E+00	< 2E-16
42	$\frac{nF}{nC}$	7.8350E+00	2.2930E+00	0.000653
43	$\frac{nBr}{(nC)^2}$	-8.7240E+01	2.3730E+01	0.000246
44	$\frac{(nBr)^2}{(nC)^2}$	1.1630E+02	2.1700E+01	9.99E-08
45	$\frac{nI}{nC}$	1.2860E+02	1.8990E+01	1.95E-11
46	$\frac{(nS)^2}{(nC)^2}$	1.0620E+01	1.7980E+00	4.35E-09
47	$\frac{nSi}{(nC)^2}$	9.3100E+01	7.7940E+00	< 2E-16
48	$\frac{nN}{nC}$	-2.4080E+01	4.0190E+00	2.69E-09
49	$\frac{nN}{(nC)^2}$	2.0730E+01	5.3740E+00	0.00012
50	$\frac{(nN)^2}{(nC)^2}$	1.1510E+01	2.9240E+00	8.68E-05
51	$\frac{nP}{nC}$	-1.5490E+02	4.0140E+01	0.00012
52	$\frac{(nP)^2}{(nC)^2}$	3.7650E+02	1.3220E+02	0.004457
53	$\frac{1}{(T_b \times nC)^2}$	1.2740E+06	1.8810E+05	1.91E-11
54	$\frac{1}{(MW \times nC)}$	6.2040E+03	5.0000E+02	< 2E-16
55	$\frac{1}{(MW \times nC)^2}$	-5.6420E+04	4.5010E+03	< 2E-16
56	$\frac{MW}{(nC)^2}$	2.6070E+00	3.2980E-01	5.61E-15

	Variables	Coefficient	Std. error	p-value
57	$\frac{(MW)^2}{(nC)^2}$	-1.61E-02	1.54E-03	< 2E-16

Table 4D4. Predictable capability of the developed UFL SFBS-MLR-NLI model in Step 2

Statistical metric	Training set	Test set	Overall dataset
R^2	0.679	0.495	0.639
RMSE [kJ/mol]	4.340	5.757	4.657
$RMSE_{Loo}$[kJ/mol]	5.184	-	-
Q^2_{Loo}	0.554	-	-
Q^2_{ext}	-	0.495	-
MAE [kJ/mol]	2.251	2.644	2.329
Std. error	4.435	5.763	4.658
Number of compounds	1,370	341	1,711

4.13. Appendix 4D3 UFL SFBS-MLR-NLI model without 15 outliers

By performing the SFBS-MLR-NLI procedure in the training set without the fifteen outliers, the optimum subset of fifteen predictors was selected. The created model for predicting the UFL is presented as the following:

$$
\begin{aligned}
&UFL\ [vol.\%] \\
&\approx f(the\ numbers\ of\ H, C, F, Br, O, N, Si, Si, Cl, I, P, MW, T_b, interaction\ terms,) \\
&= -1{,}718 - 2.687nO - 3.066nF - 18.79nBr + 2.829T_b - 17.76\sqrt{nC} + 5.055\sqrt{MW} \\
&- 168.3\sqrt{T_b} + 2{,}596(\log_{10} T_b) - 0.0318\exp(nN) - 4.301\exp(nBr) \\
&- 115.4\exp(nS) + 0.0021MW^2 - 0.0005T_b^2 + 0.4112(nC)^2 - 0.4455(nO)^2 \\
&- 0.4927(nF)^2 - 5.855(nSi)^2 + 177.5(nS)^2 \\
&- MW(0.0616nC + 0.0119nH + 0.0587nCl + 0.0241nN + 0.3474nS) \\
&+ nC(0.1809nH + 16.01nI + 2.38nS) \\
&+ nH(0.2171nO + 0.801nBr + 0.3446nN + 0.6256nSi + 0.8274nS) \\
&- 1.071(nO \times nF) + \frac{1}{(nC)^2}\left(-205.3 + \frac{64.39}{nC}\right) - 3.618\frac{nH}{nC} + 16.1\frac{nO}{(nC)^2} \\
&+ \frac{1}{nC}(4.214nF + 135.0nI - 26.93nP) \\
&+ \frac{1}{(nC)^2}(24.07(nBr)^2 + 10.0(nS)^2 + 101.4nSi) + 93.1\frac{nSi}{(nC)^2} \\
&+ \frac{nN}{nC}\left(-18.71 + \frac{17.23}{nC} + 12.47\frac{nN}{nC}\right) + 848{,}500\frac{1}{(T_b \times nC)^2} \\
&+ \frac{1}{(MW \times nC)}\left(3{,}980 - 35{,}950\frac{1}{(MW \times nC)}\right) + \frac{MW}{(nC)^2}(2.091 - 0.0147MW)
\end{aligned}
$$

Equation 4D3

Table 4D5. Selected predictors and their coefficients of the UFL SFBS-MLR-NLI model without fifteen outliers (Equation 4D3)

No.	Variables	Coefficient	Std. error	p-value
0	Intercept	-1.7180E+03	4.4980E+02	1.40E-04
1	nO	-2.6870E+00	3.4540E-01	1.48E-14
2	nF	-3.0660E+00	8.2110E-01	0.000196
3	nBr	-1.8790E+01	3.5410E+00	1.31E-07
4	T_b	2.8290E+00	7.9950E-01	0.000417
5	$\sqrt{nC}$	-1.7760E+01	1.1500E+00	< 2E-16
6	$\sqrt{MW}$	5.0550E+00	3.3670E-01	< 2E-16
7	$\sqrt{T_b}$	-1.6830E+02	4.4640E+01	0.000171
8	$\log_{10} T_b$	1.5960E+03	4.0250E+02	7.74E-05
9	$exp\ (nN)$	-3.1800E-02	1.3020E-02	0.014769
10	$exp\ (nBr)$	-4.3010E+00	1.2350E+00	0.000513
11	$exp\ (nS)$	-1.1540E+02	1.5620E+01	2.71E-13
12	MW^2	2.1420E-03	1.5680E-04	< 2E-16
13	T_b^2	-4.5360E-04	1.4960E-04	0.002475
14	$(nC)^2$	4.1120E-01	2.8740E-02	< 2E-16
15	$(nO)^2$	-4.4550E-01	4.9110E-02	< 2E-16
16	$(nF)^2$	-4.9270E-01	1.0070E-01	1.12E-06
17	$(nSi)^2$	-5.8550E+00	4.3170E-01	< 2E-16
18	$(nS)^2$	1.7750E+02	2.3310E+01	4.97E-14
19	$MW \times nC$	-6.1580E-02	3.9890E-03	< 2E-16
20	$MW \times nH$	-1.1890E-02	1.4920E-03	3.41E-15
21	$MW \times nCl$	-5.8680E-02	4.9940E-03	< 2E-16
22	$MW \times nN$	-2.4140E-02	4.3130E-03	2.64E-08
23	$MW \times nS$	-3.4740E-01	3.5410E-02	< 2E-16
24	$T_b \times nI$	-3.2460E-01	3.8340E-02	< 2E-16
25	$T_b \times nS$	8.3190E-02	1.2710E-02	8.37E-11
26	$nC \times nH$	1.8090E-01	2.1930E-02	3.91E-16
27	$nC \times nI$	1.6010E+01	2.7880E+00	1.16E-08
28	$nC \times nS$	2.3800E+00	3.7910E-01	4.67E-10
29	$nH \times nO$	2.1710E-01	2.2800E-02	< 2E-16
30	$nH \times nBr$	8.0100E-01	1.4910E-01	9.15E-08
31	$nH \times nN$	3.4460E-01	4.2680E-02	1.52E-15
32	$nH \times nSi$	6.2560E-01	6.1880E-02	< 2E-16
33	$nH \times nS$	8.2740E-01	1.1510E-01	1.09E-12
34	$nO \times nF$	-1.0710E+00	3.6540E-01	0.003445
35	$\frac{1}{(nC)^2}$	-2.0530E+02	3.3580E+01	1.29E-09

No.	Variables	Coefficient	Std. error	p-value
36	$\frac{1}{(nC)^3}$	6.4390E+01	1.5400E+01	3.09E-05
37	$\frac{nH}{nC}$	-3.6180E+00	2.7110E-01	< 2E-16
38	$\frac{nO}{(nC)^2}$	1.6100E+01	1.3750E+00	< 2E-16
39	$\frac{nF}{nC}$	4.2140E+00	1.5950E+00	0.008332
40	$\frac{(nBr)^2}{(nC)^2}$	2.4070E+01	8.5480E+00	0.004938
41	$\frac{nI}{nC}$	1.3500E+02	1.4250E+01	< 2E-16
42	$\frac{(nS)^2}{(nC)^2}$	1.0000E+01	1.3100E+00	4.48E-14
43	$\frac{nSi}{(nC)^2}$	1.0140E+02	5.8120E+00	< 2E-16
44	$\frac{nN}{nC}$	-1.8710E+01	3.0110E+00	6.99E-10
45	$\frac{nN}{(nC)^2}$	1.7230E+01	4.7260E+00	0.000277
46	$\frac{(nN)^2}{(nC)^2}$	1.2470E+01	3.9190E+00	0.0015
47	$\frac{nP}{nC}$	-2.6930E+01	8.7450E+00	0.00212
48	$\frac{1}{(T_b \times nC)^2}$	8.4850E+05	1.4800E+05	1.23E-08
49	$\frac{1}{(MW \times nC)}$	3.9800E+03	5.0560E+02	7.34E-15
50	$\frac{1}{(MW \times nC)^2}$	-3.5950E+04	4.2660E+03	< 2E-16
51	$\frac{MW}{(nC)^2}$	2.0910E+00	2.7530E-01	5.84E-14
52	$\frac{(MW)^2}{(nC)^2}$	-1.4720E-02	1.1610E-03	< 2E-16

Table 4D6. Predictable capability of the developed UFL SFBS-MLR-NLI model without 15 outliers in Step 2

Statistical metric	Training set	Test set	Overall dataset
R^2	0.774	0.615	0.748
RMSE [kJ/mol]	3.182	3.751	3.303
$RMSE_{Loo}$[kJ/mol]	4.242	-	-
Q^2_{Loo}	0.603	-	-
Q^2_{ext}	-	0.615	-
MAE [kJ/mol]	1.919	2.279	1.991
Std. error	3.246	3.742	3.303
Number of compounds	1,370	341	1,711

4.14. References of Chapter 4

Affens, W. A. (1966). Flammability Properties of Hydrocarbon Fuels. Interrelations of Flammability Properties of n-Alkanes in Air. Journal of Chemical and Engineering Data, 11(2), 197-202. https://doi.org/10.1021/je60029a022.

Affens, W. A., & McLaren, G. W. (1972). Flammability properties of hydrocarbon solutions in air. Journal of Chemical and Engineering Data, 17(4), 482-488. https://doi.org/10.1021/je60055a040.

Albahri, T. A. (2003). Flammability characteristics of pure hydrocarbons. Chemical Engineering Science, 58(16), 3629-3641. https://doi.org/10.1016/S0009-2509(03)00251-3.

Albahri, T. A. (2015). MNLR and ANN structural group contribution methods for predicting the flash point temperature of pure compounds in the transportation fuels range. Process Safety and Environmental Protection, 93, 182-191. https://doi.org/10.1016/j.psep.2014.03.005.

Alibakshi, A. (2018). Strategies to develop robust neural network models: Prediction of flash point as a case study. Analytica chimica acta, 1026, 69-76. https://doi.org/10.1016/j.aca.2018.05.015.

Alin, A. (2010). Multicollinearity. Wiley Interdisciplinary Reviews: Computational Statistics, 2(3), 370-374. https://doi.org/10.1002/wics.84.

Allison, P. (2012). When can you safely ignore multicollinearity. Statistical horizons, 5(1), 1-2.

ASTM. (2004). Standard Test Method forConcentration Limits of Flammability of Chemicals (Vaporsand Gases). In E681-04. West Conshohocken, PA: ASTM International.

ASTM. (2016). Standard Test Method forFlash Point by Tag Closed Cup Tester. In D56-16a. West Conshohocken,PA: ASTM International.

Atkins, P., & De Paula, J. (2017). Elements of physical chemistry (7th ed.): Oxford University Press, USA.

Bagheri, M., Bagheri, M., Heidari, F., & Fazeli, A. (2012). Nonlinear molecular based modeling of the flash point for application in inherently safer design. Journal of Loss Prevention in the Process Industries, 25(1), 40-51. https://doi.org/10.1016/j.jlp.2011.06.025.

Bagheri, M., Borhani, T. N. G., & Zahedi, G. (2012). Estimation of flash point and autoignition temperature of organic sulfur chemicals. Energy Conversion and Management, 58, 185-196. https://doi.org/10.1016/j.enconman.2012.01.014.

Bagheri, M., Rajabi, M., Mirbagheri, M., & Amin, M. (2012). BPSO-MLR and ANFIS based modeling of lower flammability limit. Journal of Loss Prevention in the Process Industries, 25(2), 373-382. https://doi.org/10.1016/j.jlp.2011.10.005.

Bischl, B., Lang, M., Kotthoff, L., Schratz, P., Schiffner, J., Richter, J., Jones Z., Casalicchio, G., Gallo, M. (2020). Package 'mlr'. Retrieved from https://mlr.mlr-org.com/

Borhani, T. N. G., Saniedanesh, M., Bagheri, M., & Lim, J. S. (2016). QSPR prediction of the hydroxyl radical rate constant of water contaminants. Water research, 98, 344-353. https://doi.org/10.1016/j.watres.2016.04.038.

Britton, L. G. (2002). Using heats of oxidation to evaluate flammability hazards. Process Safety Progress, 21(1), 31-54. https://doi.org/10.1002/prs.680210108.

Butler, R., Cooke, G., Lukk, G., & Jameson, B. (1956). Prediction of flash points of middle distillates. Industrial & Engineering Chemistry, 48(4), 808-812.

Carroll, F. A., Godinho, J. M., & Quina, F. H. (2011). Development of a simple method to predict boiling points and flash points of acyclic alkenes. Industrial & Engineering Chemistry Research, 50(24), 14221-14225. https://doi.org/10.1021/ie201241e.

Carroll, F. A., Lin, C.-Y., & Quina, F. H. (2010a). Calculating flash point numbers from molecular structure: an improved method for predicting the flash points of acyclic alkanes. Energy & fuels, 24(1), 392-395. https://doi.org/10.1021/ef900883u.

Carroll, F. A., Lin, C.-Y., & Quina, F. H. (2010b). Improved prediction of hydrocarbon flash points from boiling point data. Energy & fuels, 24(9), 4854-4856.https://doi.org/10.1021/ef1005836.

Carroll, F. A., Lin, C.-Y., & Quina, F. H. (2011). Simple method to evaluate and to predict flash points of organic compounds. Industrial & Engineering Chemistry Research, 50(8), 4796-4800. https://doi.org/10.1021/ie1021283.

Catoire, L., & Naudet, V. (2004). A unique equation to estimate flash points of selected pure liquids application to the correction of probably erroneous flash point values. Journal of Physical and Chemical Reference Data, 33(4), 1083-1111. https://doi.org/10.1063/1.1835321.

CCPS, Center for Chemical Process Safety. (1999). Guidelines for consequence analysis of chemical releases (Vol. 1): Wiley-AIChE.

Chen, C.-C., Lai, C.-P., & Guo, Y.-C. (2017). A novel model for predicting lower flammability limits using quantitative structure activity relationship approach. Journal of Loss Prevention in the Process Industries, 49, 240-247. https://doi.org/10.1016/j.jlp.2017.07.007.

Coley, C. W., Barzilay, R., Green, W. H., Jaakkola, T. S., & Jensen, K. F. (2017). Convolutional embedding of attributed molecular graphs for physical property prediction. Journal of chemical information and modeling, 57(8), 1757-1772. https://doi.org/10.1021/acs.jcim.6b00601.

DIPPR, Design Institute for Physical Properties Relationships. (2019). DIPPR Project 801 - Full Version. Retrieved from:

https://app.knovel.com/hotlink/toc/id:kpDIPPRPF7/dippr-project-801-full/dippr-project-801-full

Duchowicz, P. R., Garro, J. C., Andrada, M. F., Castro, E. A., & Fernández, F. M. (2007). QSPR modeling of heats of combustion for carboxylic acids. QSAR & Combinatorial Science, 26(5), 647-652. https://doi.org/10.1002/qsar.200630073.

Fred Nwanganga, M. C. (2020). Practical Machine Learning in R. Indianapolis, Indiana: John Wiley & Sons, Inc.,.

Frutiger, J., Marcarie, C., Abildskov, J., & Sin, G. (2016). Group-contribution based property estimation and uncertainty analysis for flammability-related properties. Journal of hazardous materials, 318, 783-793. https://doi.org/10.1016/j.jhazmat.2016.06.018.

Fujii, A., & Hermann, E. R. (1982). Correlation between flash points and vapor pressures of organic compounds. Journal of Safety Research, 13(4), 163-175. https://doi.org/10.1016/0022-4375(82)90032-9.

Gharagheizi, F. (2008a). Quantitative structure− property relationship for prediction of the lower flammability limit of pure compounds. Energy & fuels, 22(5), 3037-3039. https://doi.org/10.1021/ef800375b.

Gharagheizi, F. (2008b). A simple equation for prediction of net heat of combustion of pure chemicals. Chemometrics and Intelligent Laboratory Systems, 91(2), 177-180. https://doi.org/10.1016/j.chemolab.2007.11.003.

Gharagheizi, F., & Alamdari, R. F. (2008). Prediction of flash point temperature of pure components using a quantitative structure–property relationship model. QSAR & Combinatorial Science, 27(6), 679-683. https://doi.org/10.1002/qsar.200730110.

Gharagheizi, F., Alamdari, R. F., & Angaji, M. T. (2008). A new neural network− group contribution method for estimation of flash point temperature of pure components. Energy & fuels, 22(3), 1628-1635. https://doi.org/10.1021/ef700753t.

Godinho, J. M., Carroll, F. A., & Quina, F. H. (2012). A simple method to evaluate, correlate and predict boiling and flash points of alkynes. Journal of the Brazilian Chemical Society, 23(10), 1895-1899. http://dx.doi.org/10.1590/S0103-50532012005000064.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction: Springer Science & Business Media.

High, M. S., & Danner, R. P. (1987). Prediction of upper flammability limit by a group contribution method. Industrial & Engineering Chemistry Research, 26(7), 1395-1399. https://doi.org/10.1021/ie00067a021.

Hshieh, F. Y. (1997). Correlation of closed-cup flash points with normal boiling points for silicone and general organic compounds. Fire and materials, 21(6), 277-282. https://doi.org/10.1002/(SICI)1099-1018(199711/12)21:6<277::AID-FAM617>3.0.CO;2-3.

Hshieh, F. Y. (1999). Predicting heats of combustion and lower flammability limits of organosilicon compounds. Fire and materials, 23(2), 79-89. https://doi.org/10.1002/(SICI)1099-1018(199903/04)23:2<79::AID-FAM673>3.0.CO;2-F.

Huckel, E., & Debye, P. (1923). Zur theorie der elektrolyte. i. gefrierpunktserniedrigung und verwandte erscheinungen. Phys. Z, 24, 185-206.

Ishiuchi, Y. (1976). Prediction of flash points of flammable liquids. Anzen Kogaku, 15, 382-386.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). An Introduction to Statistical Learning with Applications in R. In: Springer, New York, NY.

Jia, Q., Wang, Q., Ma, P., Xia, S., Yan, F., & Tang, H. (2012). Prediction of the flash point temperature of organic compounds with the positional distributive contribution method. Journal of Chemical & Engineering Data, 57(12), 3357-3367. https://doi.org/10.1021/je301070f.

Jiao, Z., Hu, P., Xu, H., & Wang, Q. (2020). Machine Learning and Deep Learning in Chemical Health and Safety: A Systematic Review of Techniques and Applications. ACS Chemical Health & Safety, 27(6), 316-334. https://doi.org/10.1021/acs.chas.0c00075.

Jiao, Z., Escobar-Hernandez, H. U., Parker, T., & Wang, Q. (2019). Review of recent developments of quantitative structure-property relationship models on fire and

explosion-related properties. Process Safety and Environmental Protection, 129, 280-290. https://doi.org/10.1016/j.psep.2019.06.027. https://doi.org/10.1016/j.jlp.2020.104226

Jiao, Z., Ji, C., Yuan, S., Zhang, Z., & Wang, Q. (2020). Development of machine learning based prediction models for hazardous properties of chemical mixtures. Journal of Loss Prevention in the Process Industries, 67, 104226.

Jiao, Z., Yuan, S., Zhang, Z., & Wang, Q. (2020). Machine learning prediction of hydrocarbon mixture lower flammability limits using quantitative structure-property relationship models. Process Safety Progress, 39(2), e12103. https://doi.org/10.1002/prs.12103.

Kanury, A. M. (1983). A relationship between the flash point, boiling point and the lean limit of flammability of liquid fuels. Combustion science and technology, 31(5-6), 297-302. https://doi.org/10.1080/00102208308923647.

Katritzky, A. R., Petrukhin, R., Jain, R., & Karelson, M. (2001). QSPR analysis of flash points. Journal of chemical information and computer sciences, 41(6), 1521-1530. https://doi.org/10.1021/ci010043e.

Katritzky, A. R., Stoyanova-Slavova, I. B., Dobchev, D. A., & Karelson, M. (2007). QSPR modeling of flash points: An update. Journal of Molecular Graphics and Modelling, 26(2), 529-536. https://doi.org/10.1016/j.jmgm.2007.03.006.

Keshavarz, M. H., & Ghanbarzadeh, M. (2011). Simple method for reliable predicting flash points of unsaturated hydrocarbons. Journal of hazardous materials, 193, 335-341. https://doi.org/10.1016/j.jhazmat.2011.07.044.

Keshavarz, M. H., Jafari, M., Kamalvand, M., Karami, A., Keshavarz, Z., Zamani, A., & Rajaee, S. (2016). A simple and reliable method for prediction of flash point of alcohols based on their elemental composition and structural parameters. Process Safety and Environmental Protection, 102, 1-8. https://doi.org/10.1016/j.psep.2016.01.018.

Lazzús, J. A. (2010). Prediction of flash point temperature of organic compounds using a hybrid method of group contribution+ neural network+ particle swarm optimization. Chinese Journal of Chemical Engineering, 18(5), 817-823. https://doi.org/10.1016/S1004-9541(09)60133-6.

Liu, X., & Liu, Z. (2010). Research progress on flash point prediction. Journal of Chemical & Engineering Data, 55(9), 2943-2950. https://doi.org/10.1021/je1003143.

Mack, E., Boord, C., & Barham, H. (1923). Calculation of flash points for pure organic substances. Industrial & Engineering Chemistry, 15(9), 963-965.

Maindonald, J., & Braun, J. (2006). Data analysis and graphics using R: an example-based approach (Vol. 10): Cambridge University Press.

Marshall, G., & Jonker, L. (2010). An introduction to descriptive statistics: A review and practical guide. Radiography, 16(4), e1-e7. https://doi.org/10.1016/j.radi.2010.01.001.

Mathieu, D., & Alaime, T. (2014). Insight into the contribution of individual functional groups to the flash point of organic compounds. Journal of hazardous materials, 267, 169-174. https://doi.org/10.1016/j.jhazmat.2013.12.047.

Moreno, J. J. M., Pol, A. P., Abad, A. S., & Blasco, B. C. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. Psicothema, 25(4), 500-506.

Pan, Y., Jiang, J., Ding, X., Wang, R., & Jiang, J. (2010). Prediction of flammability characteristics of pure hydrocarbons from molecular structures. AIChE journal, 56(3), 690-701. https://doi.org/10.1002/aic.12007.

Pan, Y., Jiang, J., Wang, R., Cao, H., & Cui, Y. (2009a). A novel QSPR model for prediction of lower flammability limits of organic compounds based on support vector machine. Journal of hazardous materials, 168(2-3), 962-969. https://doi.org/10.1016/j.jhazmat.2009.02.122.

Pan, Y., Jiang, J., Wang, R., Cao, H., & Cui, Y. (2009b). Prediction of the upper flammability limits of organic compounds from molecular structures. Industrial & Engineering Chemistry Research, 48(10), 5064-5069. https://doi.org/10.1021/ie900193r.

Pan, Y., Jiang, J., Wang, R., & Jiang, J. (2011). Predicting the net heat of combustion of organic compounds from molecular structures based on ant colony optimization. Journal of Loss Prevention in the Process Industries, 24(1), 85-89. https://doi.org/10.1016/j.jlp.2010.11.001.

Pan, Y., Jiang, J., & Wang, Z. (2007). Quantitative structure–property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network. Journal of hazardous materials, 147(1-2), 424-430. https://doi.org/10.1016/j.jhazmat.2007.01.025.

Park, S., Mendez, E., Bailey, J. P., Rogers, W., Pasman, H. J., & El-Halwagi, M. M. (2021). What can the trove of CSB incident investigations teach us? A detailed analysis of information characteristics among chemical process incidents investigated by the CSB. Journal of Loss Prevention in the Process Industries, 69, 104389. https://doi.org/10.1016/j.jlp.2021.104389.

Park, S., Xu, S., Rogers, W., Pasman, H., & El-Halwagi, M. M. (2020). Incorporating inherent safety during the conceptual process design stage: A literature review. Journal of Loss Prevention in the Process Industries, 63, 104040. https://doi.org/10.1016/j.jlp.2019.104040.

Patel, S. J., Ng, D., & Mannan, M. S. (2009). QSPR flash point prediction of solvents using topological indices for application in computer aided molecular design.

Industrial & Engineering Chemistry Research, 48(15), 7378-7387. https://doi.org/10.1021/ie9000794.

Patil, G. (1988). Estimation of flash point. Fire and materials, 12(3), 127-131. https://doi.org/10.1002/fam.810120307.

Pistikopoulos, E., Barbosa-Povoa, A., Lee, J. H., Misener, R., Mitsos, A., Reklaitis, G., Venkatasubramanian, V., You, F., Gani, R. (2021). Process Systems Engineering–The Generation Next? Computers & Chemical Engineering, 107252. https://doi.org/10.1016/j.compchemeng.2021.107252.

Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. Pattern recognition letters, 15(11), 1119-1125.

Rhys, H. I. (2020). Machine Learning with R, the tidyverse, and mlr: Manning Publications.

Rowley, J., Rowley, R., & Wilding, W. (2010). Estimation of the flash point of pure organic chemicals from structural contributions. Process Safety Progress, 29(4), 353-358. https://doi.org/10.1002/prs.10401.

Rowley, J. R., Rowley, R. L., & Wilding, W. V. (2011). Prediction of pure-component flash points for organic compounds. Fire and materials, 35(6), 343-351.

Satyanarayana, K., & Kakati, M. (1991). Note: Correlation of flash points. Fire and materials, 15(2), 97-100. https://doi.org/10.1002/fam.810150208.

Seaton, W. H. (1991). Group contribution method for predicting the lower and the upper flammable limits of vapors in air. Journal of hazardous materials, 27(2), 169-185. https://doi.org/10.1016/0304-3894(91)80028-M.

Sheather, S. (2009). A modern approach to regression with R: Springer Science & Business Media.

Suzuki, T. (1994). Note: Empirical relationship between lower flammability limits and standard enthalpies of combustion of organic compounds. Fire and materials, 18(5), 333-336. https://doi.org/10.1002/fam.810180509.

Suzuki, T., & Ishida, M. (1995). Neural network techniques applied to predict flammability limits of organic compounds. Fire and materials, 19(4), 179-189. https://doi.org/10.1002/fam.810190404.

Suzuki, T., Ohtaguchi, K., & Koide, K. (1991). A method for estimating flash points of organic compounds from molecular structures. Journal of chemical engineering of Japan, 24(2), 258-261. https://doi.org/10.1252/jcej.24.258.

Tetteh, J., Suzuki, T., Metcalfe, E., & Howells, S. (1999). Quantitative structure− property relationships for the estimation of boiling point and flash point using a radial basis function neural network. Journal of chemical information and computer sciences, 39(3), 491-507. https://doi.org/10.1021/ci980026y.

Torabian, E., & Sobati, M. A. (2019). New structure-based models for the prediction of flash point of multi-component organic mixtures. Thermochimica Acta, 672, 162-172. https://doi.org/10.1016/j.tca.2018.11.012.

Wang, B., Park, H., Xu, K., & Wang, Q. (2018). Prediction of lower flammability limits of blended gases based on quantitative structure–property relationship. Journal of Thermal Analysis and Calorimetry, 132(2), 1125-1130. https://doi.org/10.1007/s10973-017-6941-9.

Wang, B., Xu, K., & Wang, Q. (2019). Prediction of upper flammability limits for fuel mixtures using quantitative structure–property relationship models. Chemical Engineering Communications, 206(2), 247-253. https://doi.org/10.1080/00986445.2018.1483350.

Wang, B., Zhou, L., Xu, K., & Wang, Q. (2017). Prediction of minimum ignition energy from molecular structure using quantitative structure–property relationship (QSPR) models. Industrial & Engineering Chemistry Research, 56(1), 47-51. https://doi.org/10.1021/acs.iecr.6b04347.

Yaws, C. L. (1999). Chemical properties handbook: McGraw-Hill.

Yuan, S., Jiao, Z., Quddus, N., Kwon, J. S.-I., & Mashuga, C. V. (2019). Developing quantitative structure–property relationship models to predict the upper flammability limit using machine learning. Industrial & Engineering Chemistry Research, 58(8), 3531-3537. https://doi.org/10.1021/acs.iecr.8b05938.

5. CONCLUSIONS

This final chapter begins by presenting a summary of key findings and contributions of these findings. Next, it discusses several limitations of this dissertation and possible future works based on the current research progress given in this dissertation.

5.1. Summary of key findings and contributions

This dissertation showed a multifaceted approach to address the key research question "how can we effectively measure inherently safer design (ISD) levels during the early design stages in which process information are insufficiently provided?". This dissertation sought to answer the following specific research questions to approach the key research questions:

(1) Which safety assessment tools can be utilized for ISD measurement during early design stages

(2) How can we categorize such safety assessment tools for the risk management system?

(3) Which variables can serve as safety indicators among the selected safety assessment tools?

(4) How can we utilize safety indicators for ISD measurement via previous process incidents? and

(5) How can we estimate missing variables that serve as safety indicators?

For the first three questions, a literature review (Chapter 2) established a comprehensive perspective of inherent safety assessment tools (ISATs) developed over the last 30 years by a host of researchers. Even though this review prioritized safety, the other elements of sustainable process design — health, environment, and economics — were also taken into consideration. This review selected 73 viable ISATs that could be adopted with only process information of the early design stages. Afterward, to help practitioners to understand the purposes of these ISATs regarding safety in a risk management system, the ISATs were reorganized into three groups: hazard-based ISAT (H-ISAT), risk-based ISAT (R-ISAT), and cost-optimal ISAT (CO-ISAT) for 22 tools, 33 tools, and 18 tools, respectively. Based on the overall view provided in this review, practitioners could determine the tool that would be optimal for themselves (e.g., the preferences of chemical companies or engineers for safety considerations).

To address the fourth question, "how can we utilize safety indicators for ISD measurement via previous process incidents", this dissertation (Chapter 3) analyzed the U.S. Chemical Safety and Hazard Investigation Board (CSB) reports at a more granular level of detail than previously published articles. Since the CSB reports often seemed to provide detailed process incident information and dealt with high-profile U.S. process incidents, it was a noteworthy resource to understand remarkable process incidents and to see whether inherently safer design features had been present or not. A guideline was established to collect the data after proposing 17 causal factors (12 types of chemical indicators and 5 types of process indicators), 7 scenario factors (incident sequences, domino effects, equipment types, operating modes, process units, deflagration/detonation,

population densities), and 3 consequence factors (incident types, casualties, and economic losses). The analysis results indicated that data available was too limited to draw firm conclusions on possible risk reduction by ISD features. Even numerous data on causal factors that were mostly adopted as safety indicators among existing ISATs was missing. Therefore, this dissertation concluded that further efforts were required to estimate currently unavailable values of indicators to identify commonalities among more incident cases. This research approach was the first to study the potential of low-level information extracted from the CSB reports to utilize in ISD strategy. Also, this dissertation raised the problem of providing inconsistent formats among the CSB reports, which prevented using the CSB information effectively.

Lastly, because in Chapter 3 (Article #2) there were many hazardous material property data that could serve as safety indicators for flammable chemical characteristics for determining causal factors, this dissertation attempted in Chapter 4 (Article #3) to effectively estimate these missing variable data. Compared to the previous predictive models — physical property model, group contribution model, and quantitative structure-property relationship (QSPR) model, the newly proposed models in this dissertation were not only easy-to-use, but also provided highly competitive performance to predict the flash point, heat of combustion, LFL, and UFL of pure organic compounds. This dissertation used readily available variables — the numbers of atomic elements, molecular weights, and normal boiling points — or their combinations as predictors to propose easy-to-use models. Then, multiple linear regression (MLR) models were built based on the selected predictors via feature selections for the flash point, heat of combustion, LFL, and UFL.

Machine learning algorithms were adopted to set proper predictors among numerous default variables and built MLR models. Chapter 4 showed that the accuracy of the created models was gradually enhanced by adding the interaction terms of default variables. This result indicated that the linear terms of default variables were sufficient to predict the flash point and the heat of combustion. Meanwhile, the LFL and UFL models' performance significantly increased by adding the interaction terms of the default variables. Since the created MLR models in this dissertation were provided with equation forms, they can be utilized as parts of the mathematical formulations in optimization along with other process aspects.

Overall, this dissertation would provide practitioners in the chemical engineering community an overview of inherent safety assessment tools (Chapter 2), an analysis of previous chemical process incidents for trying to provide better understanding of common incident mechanisms, and possibly, information available on inherent safety features through various safety indicators (Chapter 3). Also, it could offer the possibility to quickly estimate missing data of the flammability properties of organic compounds through easy-to-handle correlation equations (Chapter 4).

It should be noted that the research results of this dissertation for better sustainability in the chemical process industry will continue to be circulated through peer-reviewed journals and peer-reviewed conference presentations to developers of ISD applications and users of information technology in chemical engineering.

5.2. Limitations

With regard to the studied conducted for this dissertation, some limitations need to be acknowledged.

First, the scope of this dissertation for ISATs was limited to the steady-state process conditions of fixed process sites. As stated in Chapter 2, there was only one ISAT that could be utilized in the offshore oil and gas industry, and currently proposed ISATs have only been for steady-state process conditions (for normal process conditions) without dynamic conditions (for abnormal process conditions). The currently occurring limitation of this dissertation was the same as those of previous journal papers and research regarding ISD. In this regard, it is recommended that future studies examine various chemical process industries (e.g., offshore industry, supply chain, and green energy industry) including their dynamic process conditions.

Second, the generalizability of the findings for U.S. chemical process incidents was limited because of the lack of information. Even though this dissertation (Chapter 3) analyzed the process incidents based upon data from the CSB reports that offered detailed information, the study suffered from insufficient sample size and availability of possible safety indicators. To overcome this issue, this dissertation attempted to propose hands-on predictive models for flammable chemical properties in the following chapter (Chapter 4).

The principal limitation of this dissertation was failed to determine which ISAT would be appropriate to apply based on actual data among existing ones or to propose a more reliable ISAT. Further research needs to deal with more incident and safety indicator information to achieve this major objective.

5.3. Future works

To resolve the limitations in Section 5.2 and answer the major research question of this dissertation, further studies on the current topic are therefore recommended, as shown in Figure 5.2.

First, the article (Future work #1) that proposes easy-to-apply machine learning predictive models of flammability properties for chemical mixture components is suggested. The proposed models in Chapter 4 (Article #3) were promising for quickly predicting flammable chemical properties of pure compounds. These results raise intriguing questions regarding the extent of chemical mixture compounds' predictive models. Since most chemicals used in the chemical process industry are mixtures, establishing hands-on predictive models of chemical mixtures would help to use accurate chemical indicators that reflect more realistic incident scenarios.

Second, an article (Future work #2) that identifies the key safety indicators of ISD is recommended. This article can be developed properly based on the sufficient incident database and chemical property database. Hence, once the missing flammable chemical properties are correctly estimated via Chapter 4 (Article #3) and future #1, we may ascertain which variables are more effectively predictable as the key safety indicators of ISD based on the CSB reports.

Finally, based on all the findings of this dissertation, future work #1 and #2, an article (Future work #3) that checks the performances of existing ISATs or proposes a more reliable ISAT is suggested. The ISAT, well-performed among existing ISATs or newly proposed, should be evaluated based on the data-driven approach with previous

incident data to prove its reliability, as illustrated in Figure 5.1. This article ultimately can offer the answer for the major research question of this dissertation, “How can we effectively measure inherently safer design (ISD) levels during the early design stage in which process information is insufficiently provided?”.

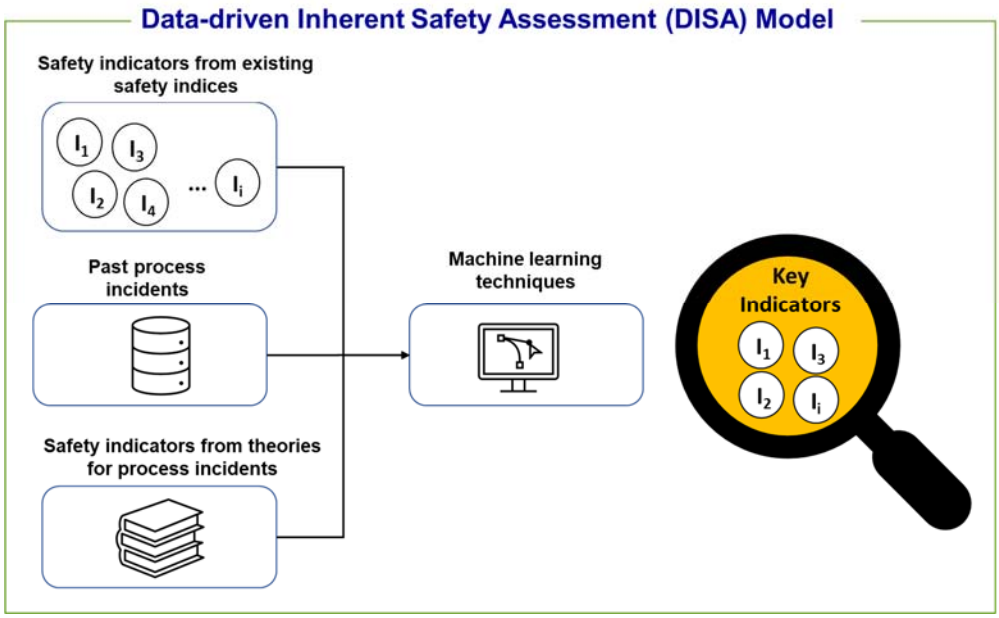


Figure 5.1 Proposed future work for identifying key safety indicators for ISD

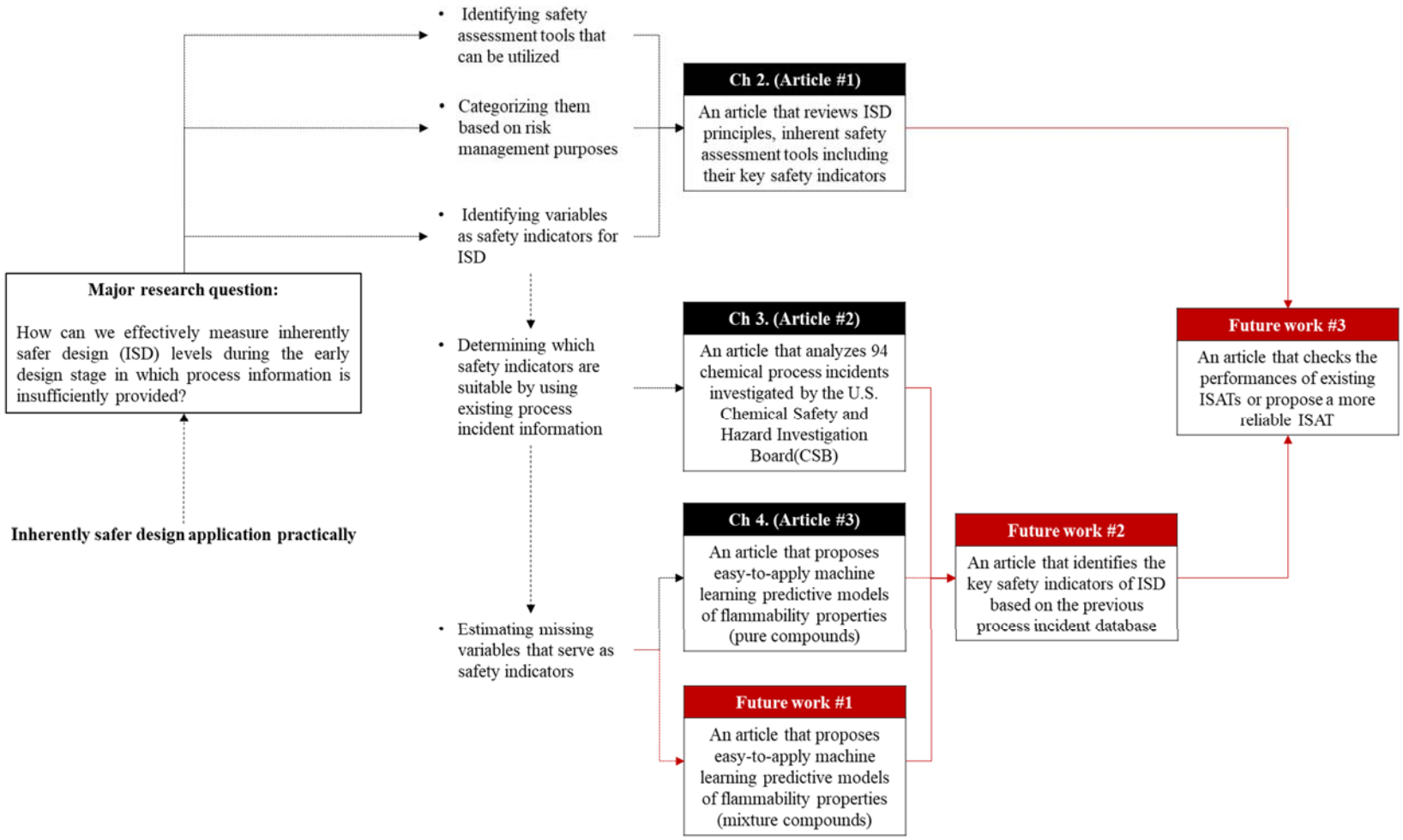


Figure 5.2 The relationship of this dissertation and possible future works

APPENDIX A COPYRIGHT PERMISSIONS

Copyright permission for Chapter 2 (Article #1)

Incorporating inherent safety during the conceptual process design stage: A literature review

Author: Sunhwa Park,Sheng Xu,William Rogers,Hans Pasman,Mahmoud M. El-Halwagi

Publication: Journal of Loss Prevention in the Process Industries

Publisher: Elsevier

Date: January 2020

Journal Author Rights

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW

Copyright permission for Chapter 2 (Figure 2.4)

Exploring inherent process safety indicators and approaches for their estimation: A systematic review

Author:
Mohammad Javad Jafari,Heidar Mohammadi,Genserik Reniers,Mostafa Pouyakian,Farshad Nourai,Seyed Ali Torabi,Masoud Rafiee Miandashti

Publication: Journal of Loss Prevention in the Process Industries

Publisher: Elsevier

Date: March 2018

Order Completed

Thank you for your order.

This Agreement between Texas A&M -- Sunhwa Park ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

Printable Details

License Number	5239520964863
License date	Jan 31, 2022

Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Loss Prevention in the Process Industries
Licensed Content Title	Exploring inherent process safety indicators and approaches for their estimation: A systematic review
Licensed Content Author	Mohammad Javad Jafari,Heidar Mohammadi,Genserik Reniers,Mostafa Pouyakian,Farshad Nourai,Seyed Ali Torabi,Masoud Rafiee Miandashti
Licensed Content Date	Mar 1, 2018
Licensed Content Volume	52
Licensed Content Issue	n/a
Licensed Content Pages	15

Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

About Your Work

Title	PRACTICAL INHERENTLY SAFER DESIGN APPROACHES DURING EARLY PROCESS DESIGN STAGES AIMING FOR SUSTAINABILITY
Institution name	Texas A&M
Expected presentation date	May 2022

Additional Data

Order reference number	1
Portions	Appendix 3

Requestor Location

Requestor Location	Texas A&M 1100 Hensel Dr. COLLEGE STATION, TX 77840 United States Attn: Texas A&M

Tax Details

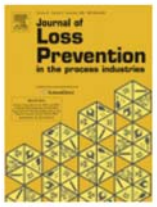
Publisher Tax ID	98-0397604

Price

Total	0.00 USD

Total: 0.00 USD

Copyright permission for Chapter 3 (Article #2)



What can the trove of CSB incident investigations teach us? A detailed analysis of information characteristics among chemical process incidents investigated by the CSB

Author: Sunhwa Park,Edna Mendez,James P. Bailey,William Rogers,Hans J. Pasman,Mahmoud M. El-Halwagi

Publication: Journal of Loss Prevention in the Process Industries

Publisher: Elsevier

Date: March 2021

Journal Author Rights

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW

Copyright permission for Chapter 4 (Article #3)

Fast, easy-to-use, machine learning-developed models of prediction of flash point, heat of combustion, and lower and upper flammability limits for inherently safer design

Author: Sunhwa Park,James P. Bailey,Hans J. Pasman,Qingsheng Wang,Mahmoud M. El-Halwagi

Publication: Computers & Chemical Engineering

Publisher: Elsevier

Date: December 2021

Journal Author Rights

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW