SUSTAINABILITY AND SAFETY STUDY OF TANK TO PROPELLER PROCESS

A Dissertation

by

CHENXI JI

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
Co-Chair of Committee,	Qingsheng Wang
Committee Members,	M. M. Faruque Hasan
	Li Zeng
Head of Department,	Arul Jayaraman

December 2021

Major Subject: Chemical Engineering

ABSTRACT

Many public concerns have been brought to the increasingly intense greenhouse effects. The International Maritime Organization (IMO) has ambitious strategies to limit the air pollutant emissions from the merchant ships in a long run, especially for carbon, sulfur, methane and nitrogen oxides.

To achieve IMO 2050 decarbonization objectives, more than one solution are required for maritime energy transition, from electric batteries for onboard activities to a variety of "green fuels" as well as safe and sustainable process design of onboard carbon capture, utilization, and storage (CCUS). Our work is focusing on screening promising marine fuels and providing safer and more sustainable carbon capture systems for maritime industry from the perspective of process safety and process systems engineering.

This work can be divided into four major parts: Tank to propeller (TTP) sustainability study focuses on providing solutions on marine fuel consumption and TTP exhaust gas emission control, and a bottom-up emission inventory model was developed by analyzing and optimizing multiple parameters; Then an onboard carbon capture system called TTP post-combustion carbon capture (TTPPCC) system was proposed by integrating ship engine process modeling with chemical absorption/desorption process modeling techniques, this work covers a thorough sustainability evaluation based on

emission reduction efficiency, energy penalty, and carbon cyclic capacity among two single aqueous amines, MEA and diisopropanolamine (DIPA), and one blended amine with a promoter, methyldiethanolamine (MDEA) with piperazine (PZ); The first TTP safety study aims at identifying the contributors influencing liquid aerosol flammability and solving their data deficiencies by developing quantitative structure−property relationship (QSPR) models, 1215 liquid chemicals and 14 predictors have been input to train the developed machine learning models via k-fold cross validation with the consideration of principal component analysis; The second TTP process safety study makes contributions on exploring inherently safer marine fuels by offering a liquid combustion risk criterion for ship compression ignition engines, two unsupervised machine learning clustering models were developed by considering liquid flammability flame propagation and aerosol formulation characteristics.

DEDICATION

This dissertation is dedicated to my beloved

parents, parents-in-law,

and

my brilliant and supportive wife, Chaoyao, and

my energetic and sweet little boy, Yibu.

ACKNOWLEDGEMENTS

On the very outset of this work, I would like to extend my heartfelt obligation towards all the personages who have walked through this journey together. Life is a continuous process, and I have been always grateful to meet and to work with so many great individuals. Without their active guidance, help and encouragement, I would not have made headway in this program.

First of all, I would like to express my deepest appreciation to my advisor, Dr. Mahmoud M. El-Halwagi for guiding me during these past three years. Besides the top-level attitude to the research, Dr. El-Halwagi influenced me a lot on how to be an excellent student-caring instructor, and I will never forget his great support during the most difficult period in my Ph.D. study.

I would also like to thank my co-chair Dr. Qingsheng Wang for inspiring me to achieve high level performance on our collaborative projects. Profound appreciation goes to my committee members, Dr. M. M. Faruque Hasan and Dr. Li Zeng for their precious comments and suggestions on my research projects and the scientific thinking during the doctoral program. The life philosophy discussion with Dr. Zeng will always be in my mind.

Moreover, I am grateful for Dr. M. Sam Mannan, my first advisor for supervising me for more than two years and leading me to this fabulous platform, the Mary Kay O'Connor

Process Safety Center in our chemical engineering department. He will always be memorized by all the staff and students in our center.

Thanks also go to my friends, colleagues, the department faculty, and staff for making my time at Texas A&M University a great experience. It is my great honor to have you during these four years.

Finally, to my family, their love and selfless support help me chase my life goals all the time. Special thanks to the warmest part in my heart, my wife Nicole, and my little boy Evin, for their encouragement, peace, and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Mahmoud El-Halwagi, Professor Qingsheng Wang, and Professor M. M. Faruque Hasan of the Department of Chemical Engineering and Dr. Li Zeng of the Department of Industrial and Systems Engineering.

All work conducted for the dissertation was completed by the student independently.

Funding Sources

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

NOMENCLATURE

AB	Auxiliary Boiler
AE	Auxiliary Engine
AIS	Automatic Identification System
AIT	Auto Iginition Temperature
BCA	Black Carbon Aerosol
CAPEX	Capital Expenditure
CIE	Compression Ignition Engine
C_p	Specific Heat Capacity
DWT	Deadweight Ton
Da	Damkohler number
$D_{3,2}$, SMD	Sauter mean diameter (μm)
FP	Flash Point
FR	Flammability Range
GPR	Gaussian process regression
HFO	Heavy Fuel Oil
HPDF	High Pressure Injection Dual Fuel
IMO	International Maritime Organization
LDV	Liquid dynamic viscosity
LICRI	Liquid In-cylinder Combustion Risk Index
LNG	Liquefied Natural Gas

LPDF	Low Pressure Injection Dual Fuel
LPG	Liquefied Petroleum Gas
LVP	Liquid Vapor Pressure
MAE	Mean Absolute Error
ME	Main Engine
MEPC	The Marine Environment Protection Committee
MGO	Marine Gas Oil
MSE	mean squared error
OPEX	Operation Expenditure
PCA	principle component analysis
QSPR	Quantitative Structure–Property Relationships
RMSE	Root Means Squared Error
R_{vessel}	Explosion vessel diameter (m)
SFC	Specific Fuel Consumption
SSD	Slow Speed Diesel
ST	Surface Tension
S_l	Laminar flame speed (m/s)
S_T	Turbulent flame speed (m/s)
TEU	Twenty Foot Equivalent Unit
TTP	Tank to Propeller
VLSFO	Very Low Sulphur Fuel Oil

Greek symbols

α	Thermal diffusivity
ρ	Liquid Density
δ_L	Flame thickness (m)
$\dot{\omega}$	Reaction rate
μ'	Turbulence intensity

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION*

The oil and gas sector have been always considered as a large capital and technology intensive sector, while the process industries focus on the infrastructures and manufacturing technologies which transform feedstocks to value-added products via the means of chemical and physical (El-Halwagi, 2017). Many process techniques have been widely adopted to the oil and gas sector, leading to great economic boost. In another aspect, the world's economy relies heavily as well on maritime transportation since more than 90% of the global trade are carried by marine transportation (International Maritime Organization, 2019), which is commonly deemed as the midstream of the oil and gas sector. As illustrated in Figure 1.1, the value chain of the marine fuel oil comprises the exploration and extraction of feedstock, refining the raw material, products transportation to bunkering station, and refueling to the ships. This whole life cycle of marine fuel oil is often referred as the well to propeller (WTP) or well to wake process, including the well to tank (WTT) process, which covers the upstream and midstream part of the value chain, and the tank to propeller (TTP) process that defines the fuel functional stage in the various types of ships (Bengtsson et al., 2011).

* Reprinted with permission from "Ji, Chenxi, and Mahmoud M. El-Halwagi. "A data-driven study of IMO compliant fuel emissions with consideration of black carbon aerosols." *Ocean Engineering* 218 (2020): 108241. Ji, Chenxi, Shuai Yuan, Zeren Jiao, Mitchell Huffman, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Predicting flammability-leading properties for liquid aerosol safety via machine learning." *Process Safety and Environmental Protection* 148 (2021): 1357-1366. Ji, Chenxi, Zeren Jiao, Shuai Yuan, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Development of novel combustion risk index for flammable liquids based on unsupervised clustering algorithms." *Journal of Loss Prevention in the Process Industries* 70 (2021): 104422." Ji, Chenxi, Shuai Yuan, Mitchell Huffman, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Post-combustion carbon capture for tank to propeller via process modeling and simulation." *Journal of CO2 Utilization* 51 (2021): 101655.

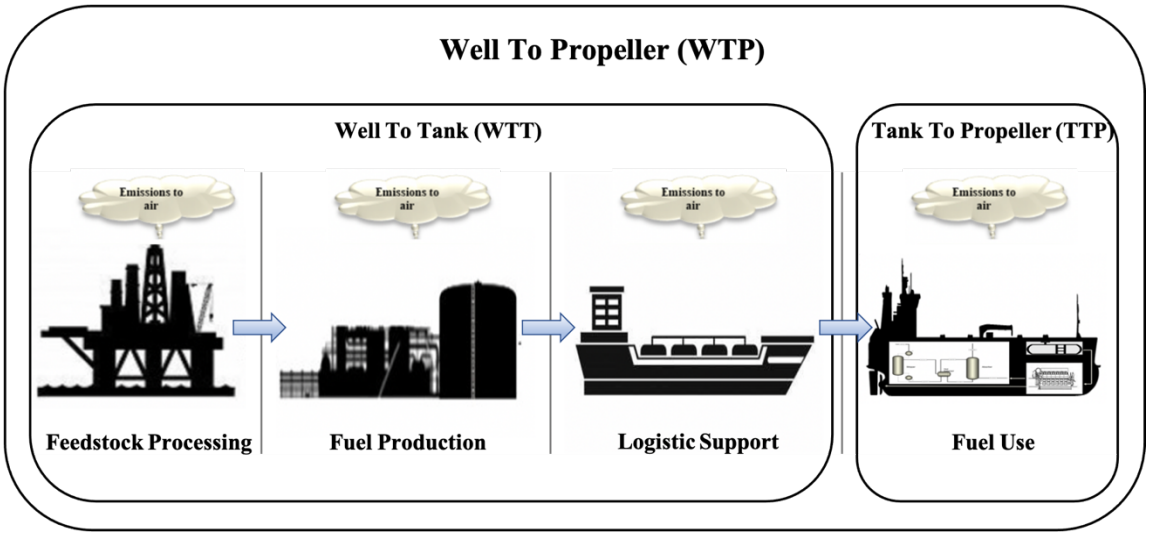


Figure 1.1 Definition of WTP, WTT and TTP

In the recent decade, many concerns have been brought on environmental and safety infrastructures and economic criteria of the well to propeller process. In the shipping industry, the Marine Environment Protection Committee (MEPC) of the International Maritime Organization (IMO) has issued a long-term plan to phase out the major ship-resulted air polluters, such as SO_x, NO_x, and CO_2. Carbon emissions are set to be reduced by at least 40% by 2030 and 70% by 2050 compared to the benchmark emissions of 2008 (MEPC 72, 2018).

Figure 1.2 shows grams CO_2 equivalent (CO_2, CH_4 and N_2O considered) per megajoules of energy contained in heavy fuel oil (HFO) (g CO_2eq/MJ) of the WTT (Figure 1.2(a)) and TTP (Figure 1.2(b)) from our collected data (Argonne National Laboratory, n.d.; Baldi et al., 2013; Baresic, D.; Smith, T.; Raucci, C.; Rehmatulla, N.; Narula, K.; Rojon, 2019; Bengtsson et al., 2012; El-Houjeiri et al., 2019, 2013; Gilbert et al., 2018; Manganaro and Lawal, 2016; Verbeek, Ruud; Kadijk, Gerrit;Mensch, Pim van; Wulffers, Chris; Beemt, Bas van den; Fraga, 2011). The median value of the TTP CO_2 equivalent emissions is 78 grams per MJ while that value of WTP is around 10 grams per MJ. Around 90% global warming potential

(GWP) of the whole WTP process may attribute to the tank to propeller process. Therefore, TTP process is essential for the global warming effects of the whole shipping fuel value chain.

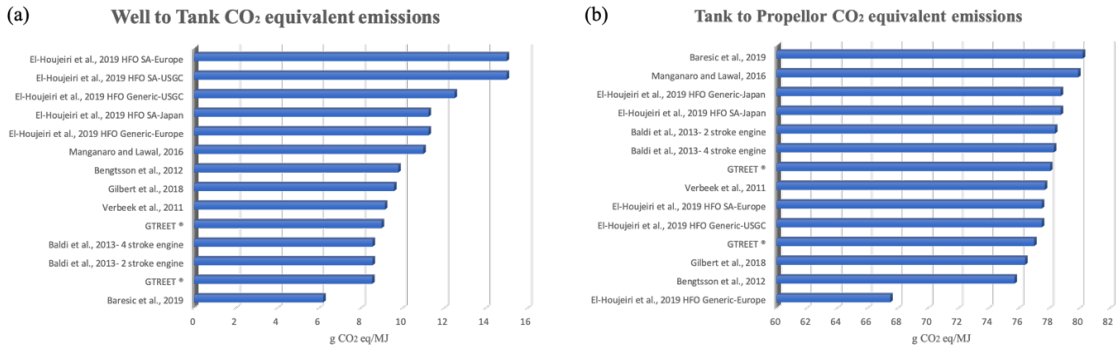


Figure 1.2 (a) CO_2 equivalent emission for well to tank and (b) tank to propeller processes

To achieve IMO 2050 decarbonization objectives, more than one solution are required for maritime energy transition, from electric batteries for onboard activities to a variety of "green fuels" as well as safe and sustainable process design of onboard carbon capture, utilization, and storage (CCUS). Our work is focusing on screening promising marine fuels and providing safer and more sustainable carbon capture systems for maritime industry from the perspective of process safety and process systems enegineering.

In general, this work is comprised of four major parts: Tank to propeller sustainability study focuses on providing solutions on marine fuel consumption and TTP exhaust gas emission control; Then an onboard carbon capture system called TTP post-combustion carbon capture (TTPPCC) system was proposed by integrating ship engine process modeling with chemical absorption/desorption process modeling techniques; The first TTP safety study aims at identifying the contributors influencing liquid aerosol flammability and solving their data deficiencies by developing quantitative structure−property relationship (QSPR) models; The

second TTP process safety study makes contributions on exploring inherently safer marine fuels by offering a liquid combustion risk criterion for ship compression ignition engines.

1.1. TTP fuel sustainability analysis

As a result of the extensive maritime transportation activities, there is a substantial impact on sustainable development of marine transportation. The International Maritime Organization (IMO) has been working on reducing greenhouse gases (GHG) emissions and other pollutants such as sulfur oxides, nitrogen oxidizes. The latest Marine Environment Protection Committee (MEPC) regulation, Annex VI of the International convention for the Prevention of Pollution from Ship (MARPOL), called "2020 Sulfur Cap", is set to globally reduce the fuel sulfur content on board from 3.5% mass to 0.5% mass outside the emission control areas (ECA). Furthermore, the IMO's carbon footprint reduction targets of 2050 has brought many challenges and opportunities for each stage of the shipping fuel supply chain. Towards complying with the MARPOL convention, the shipping companies and oil refineries have made various improvements.

The current standard for shipping fuel is ISO 8217: 2017 and its latest revision, which was published on Sep. 18, 2019, named as ISO/PAS 23263:2019. It defines general requirements and serves to confirm the compliant fuels of ISO 8217:2017 with maximum 0.50% sulfur content, and it addresses quality considerations and the range of marine fuels as well. A new category, very low sulfur fuel oil (VLSFO) with sulfur content range between 0.1% and 0.5%, was introduced to serve as the blended IMO fuel oil for non-ECAs in line with ISO 8217:2017. In addition, the proven IMO 2015 ECA grade distillate fuel, marine gas oil with maximum 0.1% sulfur content is still an available option for shipowners to comply with the 2020 Sulfur Cap. The high sulfur fuel oil (HFSO), 3.5% maximum sulfur content, is considered as an

attractive option for shipping industry because of the low fuel price, but the HFSO driven vessels should be equipped with pollutant reduction devices, mostly scrubbers, to continue serving for worldwide trade under the current MARPOL convention. As a clean fossil fuel, liquefied natural gas (LNG) has been applied to marine transportation for one decade. Many new vessels adopt this technique to meet the latest IMO rule. As illustrated in Figure 1.3, these four fuel options are the widely recognized solutions for green shipping strategy of the global commercial fleet.

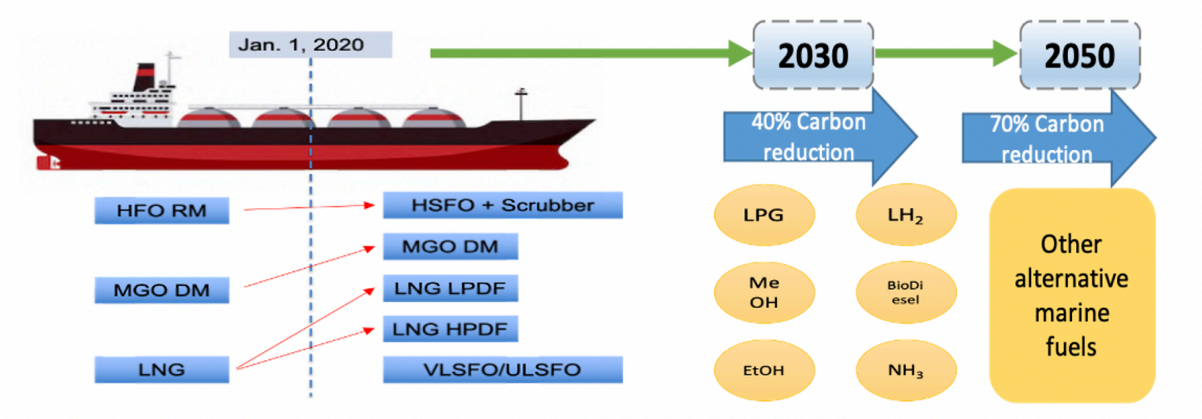


Figure 1.3 IMO 2020 shipping fuel options and promising fuel options

Although some other alternative fuel options, such as liquefied hydrogen, ammonia, methanol, and biodiesel, might be sustainable candidates for marine transportation in the future, these above-mentioned four fuels are ready to use for ocean going vessels, and the merits and demerits of them are listed in Table 1.1.

Table 1.1 Merits and demerits of proven shipping fuel options under IMO 2020 Sulfur Cap

	Pros	Cons	Remarks
VLSFO RM	New fuel, mostly blended; A preferred option by considering both	Stability and compatibility need to be widely tested	Sulfur content 0.1%~0.5%; BP estimates VLSFO will

	sustainability and fuel price		supply over 50% of the market after 2020 (CIMAC, 2018).
MGO DM	No fuel switching required; Good sustainability as a distillated component; Further expenditure may be needed.	Fuel price is the most expensive.	Distillate oil, sulfur content <0.1%; Lowest CAPEX, High OPEX
LNG	Most sustainable option (Low NO_x and SO_x emissions); Low fuel price.	Infrastructures for LNG bunkering are least established; CH_4 slip may bring more GHG negative effects; High CAPEX, unknown OPEX, safety concerns	Long term fuel option; Further technology and infrastructure needed.
HFO with Scrubber	Widely accepted fuel solution to immediately meet the requirements;	Scrubber installation needs time and room on board, and extra cost;	Sulfur content <3.5%, >0.5%;

	Medium CAPEX, low OPEX; Low fuel price.	Not a good sustainable solution; More than 10 countries forbit scrubber use in near shore areas.	A fuel economy option.

Aiming at providing sustainable fuel solutions for the stage of tank to propeller, this chapter evaluated all the compliant shipping fuel options under latest IMO sulphur limit for six most common merchant ship types, focusing on quantitatively comparing the fuel consumptions and emissions of GWP related ship exhaust gases and other major air pollutants.

1.2. TTP carbon capture system design

As many uncertainties lie in the new fuel exploration, the carbon capture techniques are currently attractive for IMO to meet its decarbonization strategy during this energy transition period. The maritime carbon capture, one specific type of combustion carbon capture, whose techniques can be classified as:

- Pre-combustion capture: carbon is separated from H_2, H_2 is used for combustion
- Oxy-fuel combustion: air for combustion is replaced by pure O_2, then CO_2 and H_2O are separated
- Post-combustion capture: CO_2 is removed from flue gases produced during the combustion of fuels using air

When considering technology maturity, the post-combustion carbon capture is the most favorable one to apply for the tank to propeller process. In particular, the tank to propeller post-combustion capture approaches may involve adsorption, cryogenic carbon capture, solvent-based chemical absorption, physical absorption, and membrane-based technologies.

The fuel tank continuously sends the shipping fuel to the ship main engine. Both the shaft work and flue gas are generated by the fuel combustion in the engine chamber. While the effective work constantly provides axial thrust to the propeller to move the ship, the produced flue gas is processed by the post-combustion carbon capture system. The flue gas firstly reacts with the lean solvent solution in the absorber, leaving the treated gas with a lower CO_2 content. Then the rich solvent (high CO_2 stream) is desorbed in the stripper, the top of which is connected to the compressed CO_2 tank for storage. The regenerated solution passes through a heat exchanger for heat recovery and goes back to the absorber.

This chapter aims to provide a holistic way to determine the most efficient and sustainable post-combustion carbon capture solution for tank to propeller processes. The objectives of this work encompass the following: 1) marine engine cylinder modeling and validation, 2) TTPPCC system process model development and pilot plant validation, 3) Optimal absorber and stripper design under variation of solvents, packed type, and liquid gas ratio, 4) Quantitative sustainable evaluation by emission reduction efficiency, energy penalty and carbon cyclic capacity.

1.3. Inherently safer fuel properties for TTP process

Besides exploring clean promising shipping fuels and developing onboard carbon capture systems, safety is a key aspect for the promising marine fuel screening process. Most ships employ 2-stroke diesel compression-ignition (CI) engine as their main power-driven source. Flammability and explosive hazards, which have been well studied, are the major concerns of safety aspects of the TTP process. However, the liquid aerosolization and flame propagation, making bulk liquids more hazardous on combustion and explosion in the cylinder of CI engine, have not widely recognized in industry or academia, thus, it is necessary to identify contributors

of liquid aerosol formulation and to find an effective method of predicting these aerosolization-leading target properties.

The objective of this work is to develop a mechanism to find an optimal algorithm for the three identified liquid aerosol safety parameters of organic compounds by comparing interpretation and prediction accuracy among the proposed machine learning models. One liquid aerosolization controlling factor (LVP) and two SMD dominating parameters (LDV and liquid surface tension) are set as the dependent variables to develop quantitative structure–property relationship (QSPR) models individually.

1.4. Novel combustion risk index for TTP liquids

To fill in gaps for categorizing promising shipping fuels from the perspective of chemical safety and process system engineering, there are two steps to carry on the study of TTP process safety: the first step is to find optimal models for the identified contributors of liquid aerosol formulation and the next step is to adopt the clustering and classification approaches via unsupervised machine learning (ML) algorithms to quantitatively classify the safety level of promising marine fuels.

By integrating the liquid aerosol formulation contributors with liquid flammability and flame propagation, the database of our liquid in-cylinder combustion criterion is built with three evaluation matrix and nine contributors including auto-ignition temperature (AIT), flash point (FP), flammability range (FR), heat capacity (HC), liquid density (LD), and liquid thermal conductivity (LTC). This work employs two unsupervised clustering algorithms to categorize the risk rating of liquid flammability, flame propagation, and aerosol formulation, then a liquid in-cylinder combustion risk index (LICRI) is presented to obtain the overall liquid combustion

safety ratings. The four cluster models are finalized as the optimal ones for the three TTP risk rating categorizes. The liquid organic compound database is clustered into four groups for the three safety matrices, and the low overall rating presents the high-level hazard.

Compared with the flash point driven liquid flammability classification method, this work integrates the major inherent properties of the liquid flammability and flame propagation with aerosol formulation. Unlike the traditional expert judgment dominated risk assessment techniques, the two ML clustering algorithms are firstly executed to evaluate the risk associated with marine fuel combustion in the engine cylinder. Also, the graph theory-based spectral clustering algorithm is the first time to employ in the chemical process safety field based on the authors' knowledge. More than 700 liquid chemicals have been analyzed and rated by the proposed LICRI criterion, proving as a solid reference for marine fuel selection in the TTP process.

To achieve these goals, the following specific tasks will be conducted in the below chapters:

- Determination of emission profiles of IMO 2020 compliant fuel options with focus on equivalent carbon dioxide (CO_{2e}) and other major ship exhaust gases in marine shipping while adopting a holistic approach with clear basis of assumptions, scenarios, and characterization of uncertainties;
- Identification of the most promising marine fuels for designated ship types with consideration different unit of capacity bins;
- Ship dual fuel engine cylinder modeling and validation;
- TTP post-combustion carbon capture process design and pilot plant validation;
- Optimal absorber and stripper design under variation of solvents, packed type and liquid gas ratio;

- Identification of the contributors of the liquified fuel flammability, flame propagation and liquid aerosol formulation;
- Determination of the optimal models of liquid aerosolization leading factors;
- Development of a novel criterion to classify the risk rating of liquified fuels in marine compression ignition engines.

1.5. References

Argonne National Laboratory, n.d. The Greenhouse Gases, Regulated Emissions, and Energy Use in Transportation (GREET) Model [WWW Document]. URL http://greet.es.anl.gov/.

Baldi, F., Bengtsson, S., Andersson, K., 2013. The influence of propulsion system design on the carbon footprint of different marine fuels. Low Carbon Shipp. Conf.

Baresic, D.; Smith, T.; Raucci, C.; Rehmatulla, N.; Narula, K.; Rojon, I., 2019. LNG as a marine fuel in the EU. Market, bunkering infrastructure investments and risks in the context of GHG reductions, University Maritime Advisory Services.

Bengtsson, S., Andersson, K., Fridell, E., 2011. A comparative life cycle assessment of marine fuels: Liquefied natural gas and three other fossil fuels. Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ. 225, 97–110. https://doi.org/10.1177/1475090211402136

Bengtsson, S., Fridell, E., Andersson, K., 2012. Environmental assessment of two pathways towards the use of biofuels in shipping. Energy Policy 44, 451–463. https://doi.org/10.1016/j.enpol.2012.02.030

CIMAC, 2018. Fuel 2020.

El-Halwagi, M.M., 2017. Introduction to Sustainability, Sustainable Design, and Process Integration, in: Sustainable Design Through Process Integration. pp. 1–14. https://doi.org/10.1016/b978-0-12-809823-3.00001-1

El-Houjeiri, H., Monfort, J.C., Bouchard, J., Przesmitzki, S., 2019. Life Cycle Assessment of Greenhouse Gas Emissions from Marine Fuels: A Case Study of Saudi Crude Oil versus Natural Gas in Different Global Regions. J. Ind. Ecol. 23, 374–388. https://doi.org/10.1111/jiec.12751

El-Houjeiri, H.M., Brandt, A.R., Duffy, J.E., 2013. Open-source LCA tool for estimating greenhouse gas emissions from crude oil production using field characteristics. Environ. Sci. Technol. https://doi.org/10.1021/es304570m

Gilbert, P., Walsh, C., Traut, M., Kesieme, U., Pazouki, K., Murphy, A., 2018. Assessment of full life-cycle air emissions of alternative shipping fuels. J. Clean. Prod. 172, 855–866. https://doi.org/10.1016/j.jclepro.2017.10.165

International Maritime Organization, 2019. International Maritime Organization Profile: Overview [WWW Document]. Int. Marit. Organ.

Manganaro, J.L., Lawal, A., 2016. CO2 Life-Cycle Assessment of the Production of Algae-Based Liquid Fuel Compared to Crude Oil to Diesel. Energy and Fuels. https://doi.org/10.1021/acs.energyfuels.6b00207

MEPC 72, I., 2018. INITIAL IMO STRATEGY ON REDUCTION OF GHG EMISSIONS FROM SHIPS 304, 1–11.

Verbeek, Ruud; Kadijk, Gerrit;Mensch, Pim van; Wulffers, Chris; Beemt, Bas van den; Fraga, F., 2011. Environmental and Economic aspects of using LNG as a fuel for shipping in The Netherlands.

2. LITERATURE REVIEW*

2.1. TTP fuel sustainability analysis

The increasingly intense greenhouse effects have brought more concerns than ever before. Motivated by its mission of "safe, secure, clean and sustainable shipping", IMO has implemented many strategies, regulations, and rules to reduce Greenhouse Gases (GHG) emissions from ships. The major Greenhouse Gases, CO_2, CH4, N2O, and O3 have been studied for years to take effective measures to limit their emissions (Brynolf et al., 2014a; Comer et al., 2017; DNV GL, 2019a; IPCC Panel, 2014; Luo and Wang, 2017; Olmer et al., 2017; Pavlenko et al., 2020; Smith et al., 2014). The Global Warming Potential (GWP) was firstly proposed by Intergovernmental Panel on Climate Change (IPCC) and the 100-year GWP (GWP100) was adopted by the Nations Framework Convention on Climate Change and its Kyoto Protocol (IPCC Panel, 2014). However, black carbon aerosol (BCA) has not been widely considered by the shipping industry and related research. Among all particulate phase species, BCA has its uniqueness: stable at high temperature, strong absorption of solar radiation, and insolubility in water, alcohol and other liquids (AMAP, 2015). It is worth noting that an individual soot particle can be aggregated to form external structures, which is subject to forming a mixture of coated particles to reduce the albedo of the surface and to contribute to a continuous warming effect (AMAP, 2015). As the dominant form of light-absorbing particulate matter in the atmosphere, BCA emission has been linked with negative influences on earth's

* Reprinted with permission from "Ji, Chenxi, and Mahmoud M. El-Halwagi. "A data-driven study of IMO compliant fuel emissions with consideration of black carbon aerosols." *Ocean Engineering* 218 (2020): 108241. Ji, Chenxi, Shuai Yuan, Zeren Jiao, Mitchell Huffman, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Predicting flammability-leading properties for liquid aerosol safety via machine learning." *Process Safety and Environmental Protection* 148 (2021): 1357-1366. Ji, Chenxi, Zeren Jiao, Shuai Yuan, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Development of novel combustion risk index for flammable liquids based on unsupervised clustering algorithms." *Journal of Loss Prevention in the Process Industries* 70 (2021): 104422." Ji, Chenxi, Shuai Yuan, Mitchell Huffman, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Post-combustion carbon capture for tank to propeller via process modeling and simulation." *Journal of CO2 Utilization* 51 (2021): 101655.

climate change, especially on the vulnerable Arctic region (IPCC Panel, 2014; Sharafian et al., 2019). One recent study (Olmer et al., 2017) pointed out that BCA is the second largest contributor to ship-induced GWP emissions, larger than CH4 and N2O; Moreover, black carbon is a major contributor to the human heart and lung disease as well (Bond et al., 2013). As a result, there is no reason to ignore BCA emissions, and this paper considers it as a major pollutant for the global warming potential consequence.

In assessing the environmental impacts of shipping fuels, it is necessary to adopt a life cycle assessment (LCA) approach for the whole supply chain of alternative fuels. Several studies (Bengtsson, 2011; Bengtsson et al., 2012; Brynolf et al., 2014b) have considered the extraction of raw material, fuel production, fuel transportation and storage, and bunkering as the well to tank stage and ship operation with the produced fuel as the tank to propeller stage. An important segment of the LCA is the tank to propeller (TTP) stage which is responsible for the bulk of emissions for carbonaceous fuels and is under the direct control of the shipping industry.

Accurate determination of shipping fuel emissions is critically important in benchmarking and accomplishing the long-term carbon reduction goal. To show a more reasonable estimation on the shipping fuel consumption and ship exhaust gas emissions, this work thoroughly searched the relevant works in the topic. The Marine Environment Protection Committee (MEPC), a subcommittee of IMO, has been working to streamline the process of collecting and calculating fuel combustion data. In one of the early studies on estimating emissions from the shipping industry, Eyring et al. (Eyring et al., 2005) analyzed 11 types of ships and 117,500 engines and divided them into 132 subgroups; they estimated total fuel consumptions by the following expression:

$$FC = \sum_{i=1}^{132} FC_i = \sum_{i=1}^{132} P_i \cdot F_{MCR,i} \cdot \tau_i \cdot SFOC_i \quad (2.1)$$

Where P_i is the engine power, $F_{MCR,i}$ is the maximum continuous rated, τ_i is the engine operation time and $SFOC_i$ is the specific fuel oil consumption value.

Following Eyring's study, many research including the two widely-accepted milestone studies, Port of Long Beach emission inventory (Starcrest Consulting Group, 2014) and IMO 3rd GHG study (Smith et al., 2014), have presented similar approaches with minor revise to calculate the shipping fuel consumption. Moreno-Gutiérrez's research team (Moreno-Gutiérrez et al., 2015) determined the engine operation time by calculating the result of average speed divided ship navigation distance, but they only considered the ship's sailing process, and the auxiliary boiler emission was ignored. Rakke (Rakke, 2016) presented the detailed way in his master thesis to calculate the brake power by introducing total resistance, ship speed and sea margin. Moreover, a recent study (Olmer et al., 2017) by the international council on clean transportation (ICCT) proposed a new approach to calculate the fuel consumption by dividing carbon dioxide intensity for specific fuel with total CO_2 emissions for one ship in a year. Since most studies established the mechanism models of ship fuel consumptions by the way from formula 1, the quality of estimated outputs largely relies on the accuracy of ship engine operation time, engine actual delivered power and the specific fuel oil consumption. On the other hand, several studies made their contributions on statistical models of ship exhaust gas emission. A Gaussian process model was proposed by Yuan and Nian (Yuan and Nian, 2018), seven influencing factors, including speed, mean draft, trim, wind speed, wind direction, wave height and wave direction were considered to build their model, but the RMSE value for both training data and validation set was not good. In addition, a multi linear regression method was proposed by Uyanik and his co-workers (Uyanik et al., 2019) to estimate the main engine fuel rate, but the predicted

variables of main engine revolutions per minute and shaft power seemed to be highly collinearity.

Beyond the shipping fuel consumption, many scholars have published great works on estimation of ship exhaust gas emissions. Some previous work evaluated the shipping emission on a basis of geographical characteristics, one way is to calculate shipping emission in a specific area then to expand the estimation around the world, while the other is to estimate the global shipping emission and then assign the specific values to one defined zone. In the meanwhile, more recent studies adopted either the bottom-up approaches or top-down methods to quantitatively analyze emissions occurring the maritime activities. The top-down approach is a method to calculate the total emissions without considering the vessel characteristics and the estimated values are assigned to the different ships in a specific fleet or area; It is an accessible way to evaluate regional emission inventories, but the ship data tend to have a large number of uncertainties as a result of ocean-going vessels travelling among continents, and the fuel oil characteristics and prices cannot be consistent from port to port as well. Consequently, a more close-to-reality approach, the bottom-up approach emerges to obtain the air emission by a ship in a specific position, and the total emissions are estimated by aggregating these estimates over time and over the fleet. As a first milestone study, US EPA (EPA, 2000) proposed the theoretical basis for the ship emission calculation, as shown in equation 2.2.

$$Emissions_{VCC,DWT,Mode} = EF \cdot LF_{mode} \cdot HP \cdot DWT \cdot T_{VCC,DWT,Mode} \qquad (2.2)$$

Where VCC is the vessel class (oil tanker, RORO carrier, etc.), DWT is the deadweight ton, EF is the emissions factor, LF is the mode specific load factor, HP is the horsepower of the engine and T is the running time of engine.

Based on equation 2.2, many studies (Eyring et al., 2005; Kwon et al., 2019; Moreno-Gutiérrez et al., 2019; Rakke, 2016; Smith et al., 2014; Starcrest Consulting Group, 2014; Wan et al., 2019) estimated shipping emissions with different assumptions and updated information to reduce the uncertainties. As a widely accepted way, the ship emission calculation usually begins with inputting ship static information, including ship's identification (ship name, MMSI code, IMO registry), physical properties (deadweight ton, length overall, design speed, draught, etc.) and engine properties (main engine brake horsepower, stroke type, rpm, installed power of auxiliary engine, boiler operation characteristics); Then a great number of valuable ship dynamic information, such as ship's actual speed, position, engine operation status can be provided by the automatic information system (AIS) on board. Combined these two categories of information with ship consumption calculation, a specific ship exhaust gas emission can be determined by multiplying the ship energy consumption by the emission factors of different air pollutants accordingly.

Many uncertainties are associated with the bottom-up calculations of ship emission. Firstly, ships within different types have many options of marine propulsion, among them diesel engines (slow speed engine, medium speed engine, high speed engine), gas turbine engine, electric motors, and LNG fueled engines are the typical ones. Secondly, one ship in a whole trip may experience several operation modes: cruising mode at sea, reduced speed mode in designated areas, maneuvering mode for inbound and outbound operation, and hoteling mode for in-port operation. Uncertainties underlie all the operation modes, particularly during the switch period of two modes because of frequent rudder and engine orders. Thirdly, ship's exhaust gas sources, main engine (ME) and auxiliary engine (AE), are the major ones for cruising mode, while another emission source auxiliary boiler (AB) should never be neglected

when a ship is in maneuvering and hoteling mode (EPA, 2000; Goldsworthy and Galbally, 2011; Olmer et al., 2017; Smith et al., 2014; Starcrest Consulting Group, 2014; Wan et al., 2019). Lastly but not least, ship activities in one trip may involve uncertainties underly ship routes selection, operation time and speed in ECAs and non-ECAs, loading status (in full or ballast), and geographical and meteorological conditions. Many contributions (EPA, 2000; Goldsworthy and Galbally, 2011; Jalkanen et al., 2009; Moreno-Gutiérrez et al., 2019; Olmer et al., 2017; Smith et al., 2014; Starcrest Consulting Group, 2014) have made to reduce the uncertainties of ship emission in the TTP stage, and among them the most effective way is to improve the accuracy of the engine load factor, specific fuel oil consumption and emission factor. This study compares the load factor, SFC, and emission factor for six previous studies, and Table 2.1 shows the summary of the major studies. One can find that most of the studies took the shipping speed as the key parameter for load factor to simplify the calculation, however, a more robust is needed to cover uncertainties during the voyage and to deal with different vessel types, ages, and engines with different leading factors.

Table 2.1 Summary of major studies on load factor, specific fuel consumption and emission factor

Method	Load Factor	SFC	Emission Factor
EPA Approach (EPA, 2000)	$LF = \left(\frac{V_{transient}}{V_{MCR}}\right)^3$	Obtained from data regression between specific emission and fractional load.	$E\ \left(\frac{g}{kWh}\right) = a\ (Fractional\ Load)^{-x} b$ Where E is the emissions rate per unit of work

Goldsworthy and Galbally Approach (Goldsworthy and Galbally, 2011)	$LF = \frac{P}{P_{FL}} = \left(\frac{V_{1,2}}{V_{FL}}\right)^3$	SFC values are obtained from different sources (Cooper and Gustafsson, 2004; ENTEC, 2002; EPA, 2009)	Emission factors are from different sources, covering residual oil, marine distillate oil, ultra-low sulfur diesel.
IMO 3rd study (Smith et al., 2014)	$LF = \frac{\left(\frac{t_{transient}}{t_1}\right)^{\frac{2}{3}} \left(\frac{V_{transient}}{V_1}\right)^3}{\eta_w \cdot \eta_f}$ η_w: the weather adjustment factor η_f: the fouling adjustment factor.	SFC for auxiliary boiler: 305 g/kWh. SFC for ME and AE was referred from IMO 2nd study and IVL 2004; Use STEAM model for MGO SFC.	IMO fuel combustion data for HFO, LSFO, MGO, and LNG.
Port of Long Beach report (Starcrest Consulting Group, 2014)	$LF = \left(\frac{V_{actual}}{V_{maximum}}\right)^3$	Data from ENTEC (ENTEC, 2010)	Normal engine loads: Data from 2004 IVL study (Cooper and Gustafsson, 2004). Low engine loads: $EF = BaseEF \times LLA$ LLA: Low load adjustment multiplier

STEAM Model (Jalkanen et al., 2009)	$LF = \varepsilon_p \left(\frac{V_{transient}}{V_{design} + V_{safety}} \right)^3$ Where ε_p is assumed to be 0.8, $V_{safety}$is assumed as 0.5 knots.	$SFOC(load)$ $= SFOC_{base}$ $\times (0.455 \times Load^2$ $- 0.71 \times Load$ $+ 1.28)$	$EF\left(\frac{g}{kWh}\right) =$ $\begin{cases} 17, & n < 130 \\ 45 \times n^{-2}, & [130,20 \\ 9.8, & n > 2000 \end{cases}$ n is the engine rpm.
Moreno-Gutiérrez et al.'s real-time emission model (Moreno-Gutiérrez et al., 2019)	$LF = \left(\frac{V_{transient}}{V_{MCR}} \right)^{3.5}$	Use STEAM SFC model	Use Goldsworthy and Galbally (Goldsworthy and Galbally, 2011) approach as the emission factor source.

Black carbon aerosol is formed in incomplete fossil fuel combustion process. Recently, the shipping community has become increasingly concerned with the negative impact of black carbon aerosol (BCA) emissions because they contribute to global warming by absorbing solar radiation (Lu et al., 2020; Talukdar et al., 2019). Lack and Corbett (Lack and Corbett, 2012) pointed out that BCA emissions increased at low engine loads, and ship speed, fuel quality, and exhaust gas scrubbing may bring a significant influence on black carbon aerosol emission. Compared to the common shipping residual fuel oil, most low sulfur fuels can be deemed as applicable solutions of BCA abatement, yet the newly compliant VLSFO tells another story. The BCA emission tends to have a positive linear relationship with the fuel's aromatic content, and one latest study proposed by Finland and Germany (Finland and Germany, 2019) to the IMO advised that the blended VLSFO increased BCA emission up to 85 percent due to its high proportion of aromatic compounds. Under the IMO 2020 Sulfur Cap, alternative fuels such as

LNG, low sulfur distillate oil, biodiesel, and methanol, are typical abatement strategies for BCA; besides, the electrostatic precipitators and selective catalytic reduction served as the main exhaust treatment for BCA abatement.

2.2. TTP post-combustion carbon capture system

Common practice of maritime CO_2 emission reduction encompasses the optimal design of propulsion system (Baldi et al., 2018, 2016), cleaner and safer shipping fuels substitution(Ji et al., 2021c, 2021a), maritime carbon capture and storage (Feenstra et al., 2019; Luo and Wang, 2017) and thermal efficiency improvement via onboard waste heat recovery (Shu et al., 2013; Yang and Yeh, 2015). Unlike the widely investigated carbon capture studies for onshore facilities, research on maritime/onboard carbon capture is still in an early stage: it was firstly proposed by Det Norske Veritas (DNV) and Process Systems Enterprise (PSE) in 2013 (DNV, 2013), and only limited publications have focused on this area. As one specific type of combustion carbon capture, the maritime carbon capture techniques can be classified as:

- Pre-combustion capture: carbon is separated from H_2, H_2 is used for combustion
- Oxy-fuel combustion: air for combustion is replaced by pure O_2, then CO_2 and H_2O are separated
- Post-combustion capture: CO_2 is removed from flue gases produced during the combustion of fuels using air

When considering technology maturity, the post-combustion carbon capture is the most favorable one to apply for the tank to propeller process. In particular, the tank to propeller post-combustion capture approaches may involve adsorption, cryogenic carbon capture, solvent-based chemical absorption, physical absorption, and membrane-based

technologies. Figure 2.1 (a) shows the simplified process flow diagram of solvent-based chemical absorption-type tank to propeller post-combustion carbon capture (TTPPCC), while Figure 2.1 (b) illustrates the major process units and the available solvent options of the TTPPCC process. The fuel tank continuously sends the shipping fuel to the ship main engine. Both the shaft work and flue gas are generated by the fuel combustion in the engine chamber. While the effective work constantly provides axial thrust to the propeller to move the ship, the produced flue gas is processed by the post-combustion carbon capture system. The flue gas firstly reacts with the lean solvent solution in the absorber, leaving the treated gas with a lower CO_2 content. Then the rich solvent (high CO_2 stream) is desorbed in the stripper, the top of which is connected to the compressed CO_2 tank for storage. The regenerated solution passes through a heat exchanger for heat recovery and goes back to the absorber.

Since marine vessels have limitations of space and utilities, the TTPPCC system design should consider the dimensions of the absorber and stripper and the supply of heat and electric power. In addition, the reaction kinetics and carbon capture equipment stability must be carefully considered as the tank to propeller CO_2 capture operation is in a constantly moving environment. The first work on maritime post-combustion carbon capture in public domain was presented by Luo et al. (Luo and Wang, 2017), who employed process simulation for one cargo ship under three scenarios, one without carbon capture as a reference case and the other two cases with either monoethanolamine (MEA) based PCC or MEA solvent with an additional diesel gas turbine. In 2019, Feenstra et al. (Feenstra et al., 2019) used two solvent options, 30 wt% aqueous MEA and 30 wt% aqueous piperazine (PZ), for two ships with both diesel and LNG engines. Utilizing Luo and Wang's ship engine model (Luo and Wang, 2017), Awoyomi (Awoyomi et al., 2019) applied 3.5% and 4.1% ammonia to capture CO_2 and SO_2 simultaneously to meet the requirements of both the IMO

2020 sulfur cap and 2030 IMO CO_2 strategy. Another aspect should be noticed in the process design stage is that the CO_2 concentration in the flue gas shares different values from a power plant compared to a TTPPCC system. Before entering into the carbon capture system, the flue gas from an onshore power plant contains 12-15 mol% CO_2 (Van De Haar et al., 2017), while the typical CO_2 mole fraction in the ship-based flue gas is around 4-7 mol% (Awoyomi et al., 2019; Feenstra et al., 2019; Ji et al., 2021b; Luo and Wang, 2017). The lower CO_2 concentration at the TTPPCC system inlet requires a rigorous regime to screen solvent options and solvent concentration to achieve the target carbon capture rate of 90%. This study finalized three solvent options (MEA, DIPA, and MDEA with PZ) with specific concentration by considering solvent solubility, absorption rate, reaction rate, and thermal degradation.

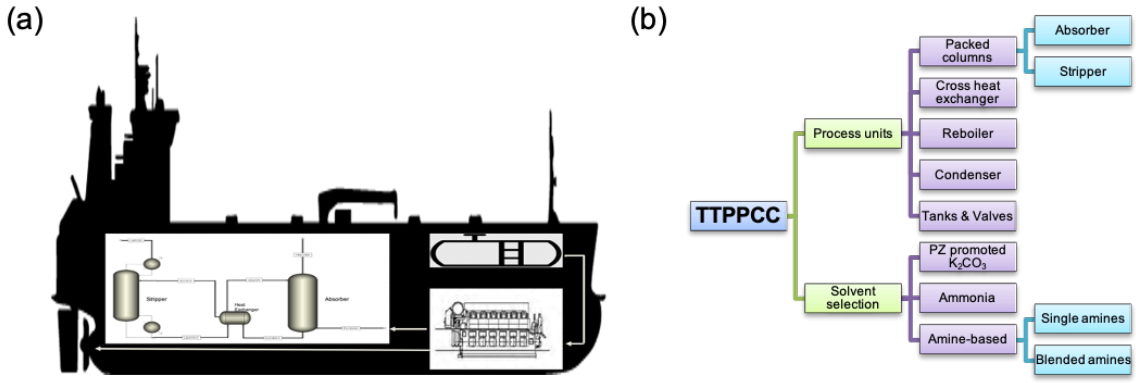


Figure 2.1 (a) Simplified process flow diagram of TTPPCC; (b) Major process units and solvent options of TTPPCC

This chapter aims at providing a holistic way to determine the most efficient and sustainable post-combustion carbon capture solution for tank to propeller processes. The objectives of this work encompass the following: 1) marine engine cylinder modeling and validation, 2) TTPPCC system process model development and pilot plant validation, 3) optimal absorber and stripper design under variation of solvents, packed type, and liquid gas ratio, 4)

Quantitative sustainable evaluation by emission reduction efficiency, energy penalty and carbon cyclic capacity.

As a follow up TTP sustainability study, this chapter will provide the most effective way to contain ship-resulted carbon volume via process intensification techniques. To the best knowledge of the authors, this chapter firstly explored a thorough sustainable performance analysis of a TTPPCC system by employing process modeling methodology. Also, this is the first time a maritime post-combustion carbon capture system has been modeled using a blended amine with PZ as a promoter.

2.3. Predicting flammability-leading properties for liquid aerosol safety

Inherent safety properties such as flash point and upper/lower flammability limit have been well investigated and are believed to be the key to finding safe fuel solutions. Beyond the typical flammability properties, numerous incidents have revealed that a liquid can be ignited below its flash point due to the aerosolization phenomenon (Kohlbrand, 1991; Santon, 2009). Unfortunately, no consensus has been reached on the standardization of aerosol flammability yet, owing to a lack of quantification (Yuan et al., 2020a). It is therefore urgent to quantitatively investigate liquid safety factors, with particular importance placed on quantifying the aerosolization effect.

In 1955, Eichhorn (Eichhorn, 1955) first established that aerosols were able to cause an explosion and pointed out that mist flammability with fuzzy boundaries was completely different from vapor flammability limits. Subsequently, Ballal and Lefebvre (Ballal and Lefebvre, 1979) proposed an experimental model to demonstrate the importance of the aerosolization process to the ignition of liquid fuels (their model consisted of iso-octane, kerosene, gasoline, diesel oil, light oil, and heavy fuel oil). Their model identified that

droplet size was the dominant factor for determining quenching distance and minimum ignition energy. Integrating the experimental results of Ballal and Lefebvre (Ballal and Lefebvre, 1979), Polymeropoulos (Polymeropoulos, 1984) presented a theoretical model for liquid aerosols, and concluded that droplets in the fuel mixtures might have an accelerating effect on flame propagation. As concerns related to the hazards of high flash point liquids continued to rise due to the aerosolization or atomization phenomenon, the study published by Bowen and Shirvill (Bowen and Shirvill, 1994) highlighted that hazards related to liquid aerosolization could be minimized by adopting the minimum practicable pressure for operating systems. However, the aerosolization effect of fuel oils was still underestimated, and was identified as the root cause of many incidents in the shipping industry (Kohlbrand, 1991; Santon, 2009). Thus, it is critical to identify the leading factors of liquid aerosolization effects and to build an accurate mathematical model to predict such properties.

Many literature sources (Ballal and Lefebvre, 1979; Danis et al., 1988; Kiran et al., 2003; Polymeropoulos, 1984; Yuan et al., 2020b, 2019) reached agreement that droplet size is more associated with flame speed than other factors are. As shown in Figure 2.2, a transition range was defined to identify the region of enhanced relationship between burning speed and droplet size (Polymeropoulos, 1984). On the left side of the transition region is homogeneous vapor phase combustion, and the burning speed on the right side of the transition region drops dramatically as a result of insufficient vapor (Krishna et al., 2003). Liquid aerosols are comprised of polydisperse droplets, and the most common mean diameter to apply for heat transfer, combustion, and dispersion modelling of aerosols is the Sauter Mean Diameter (SMD) (Krishna et al., 2003). SMD formulae proposed as an air blast type atomizer (Jasuja, 1979) and as a pressure-swirl type atomizer (Lefebvre and Ballal, 2010) are listed below.

$$SMD = 4.4v^{0.16}\sigma^{0.6}W_f^{0.22}\Delta P^{-0.43} \tag{2.3}$$

$$D_{3,2} = 2.25\sigma^{2.25}\mu_L^{0.25}\dot{m}_L^{0.25}\Delta P_L^{-0.25}\rho_A^{-0.25} \quad (2.4)$$

In the above equations, v is the kinematic viscosity, σ is the surface tension, W_f is the mass flow rate, ΔP is the atomizer pressure drop, μ_L is the liquid dynamic viscosity, $\dot{m}$ is the liquid mass flow rate, ΔP_L is the liquid pressure differential, and ρ_A is the air density.

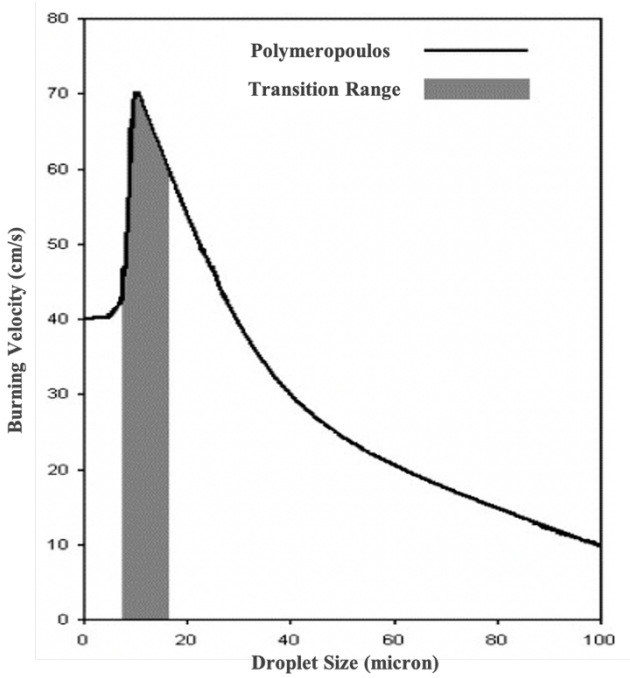


Figure 2.2 Variation of burning velocity with droplet size for medium diesel oil-air sprays (Reprint from (Polymeropoulos, 1984))

Liquid dynamic viscosity (LDV) and surface tension (ST), as the liquid inherent properties, are the determinant parameters for the droplet size of liquid aerosols based on the above two equations. Liquid aerosolization effects often occur at the interface between two different fluids, and the Weber number and Reynolds number both indicate positive correlations with liquid droplet size. Specifically, ST is inversely proportional to the Weber number, and LDV is inversely proportional to the Reynolds number as well. Salehi et al. (Salehi et al., 2017) investigated inertial effects and the shape of droplets in turbulent flows by presenting CFD

simulations, resulting in the identification of fluid viscosity as a key indicator of spherical droplet formulation. Also, Yuan and his colleagues (Yuan et al., 2020b) confirmed viscosity and surface tension as two leading factors for the aerosolization of flammable liquids via an unsupervised machine learning approach. From a safety perspective, liquids with high LDV and ST are preferred since they are less prone to aerosolization. For most practical fuels, any change in dynamic viscosity is always accompanied by a change in volatility, and Ballal and Lefebvre (Ballal and Lefebvre, 1979) indicated that the quenching distance was dependent on fuel volatility. Besides dynamic viscosity, liquid vapor pressure (LVP) is another evidential index that denotes the volatility level of liquids. Large-scale mist explosions revealed the potential hazards associated with the atomization of liquids at vapor pressure. Unlike LDV and ST, LVP has a positive effect on the liquid aerosolization phenomenon, so low LVP liquids are preferred for inherently safer design. Our previous work (Yuan et al., 2019) experimentally measured the maximum explosion pressure and explosibility characteristics of n-octane and n-dodecane liquid aerosols, and established LVP as a controlling step for the liquid aerosol combustion process after comparing the experimental outputs to a three-zone flame model. Another discovery was that the flame thickness, found by integrating the liquid aerosol burning speed with burning time, is a significant indicator to determine the maximum explosion pressure of liquid aerosols. Salehi et al. (Salehi et al., 2015) applied the conditional moment closure model to investigate ignition of lean n-heptane/air and iso-octane/air mixtures with various levels of thermal stratification. This work established that the dominated sources of ignition delay fluctuation are correlations between dissipation and conditional fluctuation, and correlations between reaction and conditional fluctuation. Furthermore, Ji et al. (Ji et al., 2021a) identified three indicators for flame propagation in a compression ignition engine and integrated them with the other six indicators, including surface tension, LVP, and dynamic viscosity, to construct two novel unsupervised machine

learning clustering models. This work exhibited a promising result for classifying flammable liquid safety with inclusion of liquid aerosolization.

Both mechanism models and statistical models have been developed to predict the flammability-leading properties for liquid aerosol safety. As an early study for LVP prediction, Jensen and his colleagues (Jensen et al., 1981) developed the UNIFAC group-contribution approach to predict LVP by calculating the excess Gibbs energies of liquids. Meanwhile, McGarry (Mcgarry, 1983) used constraint regression to determine the coefficients of the Wagner equation, and processed 14 highly accurate existing data to predict 58 other substances. Recently, the quantitative structure–property relationship (QSPR) models have been adopted for predicting multiple chemical properties (Jiao et al., 2020a; Jiao et al., 2019a; Zeng et al., 2019) including lower/upper flammability limit (Jiao et al., 2020b; Jiao et al., 2019b; Pan et al., 2008; Wang et al., 2019), minimum ignition energy (Wang et al., 2017; Chen and Chen, 2020; Chaudhari et al., 2020), liquid vapor pressure (Zeng et al., 2007), and surface tension (Gajewicz et al., 2010; Gardas and Coutinho, 2008). Eight molecular descriptors were presented to develop Gajewicz's (Gajewicz et al., 2010) QSPR model, which was utilized to predict logarithmic values of LVP for halogenated persistent organic pollutants. As for flame-propagation leading factors, Gardas and Coutinho (Gardas and Coutinho, 2008) addressed the lack of ionic liquid surface tension data by introducing parachors, densities, and molecular weight. Nevertheless, rarely has literature focused on the liquid dynamic viscosity. A Gibbs energy additivity approach to QSPR was proposed by Chum-in's group (Chum-in et al., 2017), and although a good fit was shown for biodiesel and fatty acid methyl ester, the only dependent variables were identified as average carbon numbers and number of double bonds, therefore the suitability for more complex systems was unknown.

2.4. Liquid combustion safety in compression ignition engine

New fuel options are in demand to reach the long term decarbonization goal, and safety is a top priority for the promising marine fuel selection. Flammability and explosive hazards, which have been well studied, are the major concerns of safety aspects of the tank to propeller (TTP) process aboard ships. The inherent flammable properties of liquids may involve flash point (FP), auto ignition point (AIP), upper/lower flammability limit (UFL/LFL), and boiling point. Currently, liquid combustion level is commonly determined by flash point. NFPA 704 (National Fire Protection Association, 2017), the widely recognized liquid flammability classification standard, categorizes liquids into five classes. Table 2.2 shows the NFPA 704 standard and flammability ratings of the promising fuels by adopting the NFPA fire diamond.

Table 2.2 NFPA 704 liquid flammability ratings of the promising marine fuels

Rating	NFPA Criteria	Marine fuel options
0	Materials will not burn in air when exposed to a temperature of 1500°F for a period of 5 minutes	
1	Flash point at or above 200°F	Biodiesel, Ammonia
2	Flash point between 100°F and 200°F	Heavy fuel oil, Marine gas oil, VLSFO
3	Flash point between 73°F and 100°F	Methanol, Ethanol
4	Flash point below 73°F	LNG, LPG, Liquefied Hydrogen

As located in rating 4 from Table 2.2, LNG, LPG, and liquefied hydrogen, which have displayed great potential to be long-term sustainable marine fuel solutions (DNV GL, 2019; Ji and El-Halwagi, 2020), might share the higher flammability as per NFPA 704. Since many concerns were risen on the hazards of high flash point liquid because of the aerosolization or atomization phenomenon, the study published by Bowen and Shirvill (Bowen and Shirvill,

1994) highlighted the liquid aerosolization hazard could be minimized by adopting the minimum practicable pressure for operating systems. However, the aerosolization effect of fuel oils was still underestimated, although it has been identified as the root cause of many incidents in shipping industry (Kohlbrand, 1991; Santon, 2009). Therefore, the flash point driven liquid flammability standard is too simple to classify the safety level of marine fuel options, especially when considering the common combustion scenario of marine fuels.

Most ships employ 2-stroke diesel compression-ignition (CI) engine as their main power-driven source (Klett et al., 2017). Unlike the Otto cycle, the diesel internal combustion engine uses a higher compression ratio, 15 to 20, to ignite the marine fuel (Sivaganesan and Chandrasekaran, 2016). The flame in the cylinder of the CI engine is initially propagated as laminar and later becomes turbulent. Besides, there is a common operation for the combustion of heavy fuel oil, which needs to be heated to bring the viscosity below 20 cst for achieving proper aerosolization (MAN Diesel & Turbo, 2014). Date back to 1955, Eichhorn (Eichhorn, 1955) firstly presented that the aerosols were able to lead an explosion and pointed out that the liquid aerosol flammability with fuzzy boundaries was completely different from vapor flammability limits. Moreover, we experimentally confirmed that the n-dodecane in the aerosol state can be ignited lower than the flash point. There is a pressure rising when the equivalence ratio decreases to 0.1, but neither n-octane nor n-dodecane is supposed to be ignited since the LFL for both n-octane and n-dodecane vapor are 0.57% and 0.54% respectively, illustrating the liquid aerosol has a wider flammability range than the bulk liquid (Yuan et al., 2019).

However, the liquid aerosolization and flame propagation, making bulk liquids more hazardous on combustion in the cylinder of CI engine, have not widely recognized. Thus, it is necessary to take the liquid flammability, flame propagation, and liquid aerosolization effects into

consideration to represent the safety extent of the liquefied fuel options in the CI engine. To fill in gaps for categorizing promising shipping fuels from the perspective of chemical safety, there are two steps to carry on the study of TTP process safety: the first step is to find optimal models for the identified contributors of liquid aerosol formulation and the next step is to adopt the clustering to classify the safety level of promising marine fuels. With the rapid development of machine learning (ML) techniques and algorithms, more work has focused on implementing machine learning into safety-related studies, which demonstrated their successful applications with satisfying accuracy (Jiao et al., 2019a, 2020a, 2020b). For classification and clustering, there are unsupervised learning algorithms like k-means clustering, hierarchical clustering, and spectral clustering which are suitable for this study.

2.5. References

Awoyomi, A., Patchigolla, K., Anthony, E.J., 2019. CO2/SO2 emission reduction in CO2 shipping infrastructure. Int. J. Greenh. Gas Control 88, 57–70. https://doi.org/10.1016/j.ijggc.2019.05.011

Baldi, F., Ahlgren, F., Nguyen, T. Van, Thern, M., Andersson, K., 2018. Energy and exergy analysis of a cruise ship. Energies 11, 1–41. https://doi.org/10.3390/en11102508

Baldi, F., Nguyen, T. Van, Ahlgren, F., 2016. The application of process integration to the optimisation of cruise ship energy systems: A case study. ECOS 2016 - Proc. 29th Int. Conf. Effic. Cost, Optimisation, Simul. Environ. Impact Energy Syst.

Ballal, D.R., Lefebvre, A.H., 1979. Ignition and flame quenching of flowing heterogeneous fuel-air mixtures. Combust. Flame. https://doi.org/10.1016/0010-2180(79)90019-1

Bowen, P.J., Shirvill, L.C., 1994. Combustion hazards posed by the pressurized atomization of high-flashpoint liquids. J. Loss Prev. Process Ind. https://doi.org/10.1016/0950-4230(94)80071-5

Chum-in, T., Sudaprasert, K., Phankosol, S., Lilitchan, S., Aryusuk, K., Krisnangkura, K., 2017. Gibbs energy additivity approaches to QSPR in modeling of high pressure density and kinematic viscosity of FAME and biodiesel. Fuel Process. Technol. 156, 385–393. https://doi.org/10.1016/j.fuproc.2016.09.025

Cooper, D., Gustafsson, T., 2004. Methodology for calculating emissions from ships. 1. Update of emission factors 47. https://doi.org/1652-4179

Danis, A.M., Namer, I., Cernansky, N.P., 1988. Droplet size and equivalence ratio effects on spark ignition of monodisperse N-heptane and methanol sprays. Combust. Flame. https://doi.org/10.1016/0010-2180(88)90074-0

DNV, P., 2013. Ship carbon capture and storage.

Eichhorn, J., 1955. Careful! Mist can explode. Pet. Refin. 34, 194–196.

ENTEC, 2010. UK Ship Emissions Inventory Final Report.

ENTEC, 2002. Quantification of emissions from ships associated with ship movements between ports in the European Community, Environmental Research.

EPA, 2009. Current Methodologies in Preparing Mobile Source Port-Related Emission Inventories, Transportation.

EPA, 2000. Analysis of Commercial Marine Vessel Emissions and Fuel Consumption Data,.

Eyring, V., Köhler, H.W., Van Aardenne, J., Lauer, A., 2005. Emissions from international shipping: 1. The last 50 years. J. Geophys. Res. D Atmos. 110, 171–182. https://doi.org/10.1029/2004JD005619

Feenstra, M., Monteiro, J., van den Akker, J.T., Abu-Zahra, M.R.M., Gilling, E., Goetheer, E., 2019. Ship-based carbon capture onboard of diesel or LNG-fuelled ships. Int. J. Greenh. Gas Control 85, 1–10. https://doi.org/10.1016/j.ijggc.2019.03.008

Gajewicz, A., Haranczyk, M., Puzyn, T., 2010. Predicting logarithmic values of the subcooled liquid vapor pressure of halogenated persistent organic pollutants with QSPR:

How different are chlorinated and brominated congeners? Atmos. Environ. 44, 1428–1436. https://doi.org/10.1016/j.atmosenv.2010.01.041

Gardas, R.L., Coutinho, J.A.P., 2008. Applying a QSPR correlation to the prediction of surface tensions of ionic liquids. Fluid Phase Equilib. 265, 57–65. https://doi.org/10.1016/j.fluid.2008.01.002

Goldsworthy, L., Galbally, I.E., 2011. Ship engine exhaust emissions in waters around Australia – an overview. Air Qual. Clim. 45, 1–20.

Jalkanen, J.P., Brink, A., Kalli, J., Pettersson, H., Kukkonen, J., Stipa, T., 2009. A modelling system for the exhaust emissions of marine traffic and its application in the Baltic Sea area. Atmos. Chem. Phys. 9, 9209–9223. https://doi.org/10.5194/acp-9-9209-2009

Jasuja, A.K., 1979. Atomization of crude and residual fuel oils. J. Eng. Gas Turbines Power. https://doi.org/10.1115/1.3446480

Jensen, T., Fredenslund, A., Rasmussen, P., 1981. Pure-Component Vapor Pressures Using UNIFAC Group Contribution. Ind. Eng. Chem. Fundam. 20, 239–246. https://doi.org/10.1021/i100003a010

Ji, C., El-Halwagi, M.M., 2020. A data-driven study of IMO compliant fuel emissions with consideration of black carbon aerosols. Ocean Eng. 218, 108241. https://doi.org/10.1016/j.oceaneng.2020.108241

Ji, C., Jiao, Z., Yuan, S., El-Halwagi, M.M., Wang, Q., 2021a. Development of novel combustion risk index for flammable liquids based on unsupervised clustering algorithms. J. Loss Prev. Process Ind. https://doi.org/10.1016/j.jlp.2021.104422

Ji, C., Yuan, S., Huffman, M., El-Halwagi, M.M., Wang, Q., 2021b. Post-combustion carbon capture for tank to propeller via process modeling and simulation. J. CO2 Util. https://doi.org/10.1016/j.jcou.2021.101655

Ji, C., Yuan, S., Jiao, Z., Huffman, M., El-Halwagi, M.M., Wang, Q., 2021c. Predicting

flammability-leading properties for liquid aerosol safety via machine learning. Process Saf. Environ. Prot. 148, 1357–1366. https://doi.org/10.1016/j.psep.2021.03.012

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 Saf. Environ. Prot. https://doi.org/10.1016/j.psep.2019.06.027

Jiao, Z., Hu, P., Xu, H., Wang, Q., 2020a. Machine Learning and Deep Learning in Chemical Health and Safety: A Systematic Review of Techniques and Applications. ACS Chem. Heal. Saf. acs.chas.0c00075. https://doi.org/10.1021/acs.chas.0c00075

Jiao, Z., Ji, C., Yuan, S., Zhang, Z., Wang, Q., 2020b. Development of machine learning based prediction models for hazardous properties of chemical mixtures. J. Loss Prev. Process Ind. https://doi.org/10.1016/j.jlp.2020.104226

Jiao, Zeren, Yuan, S., Zhang, Z., Wang, Q., 2019. Machine learning prediction of hydrocarbon mixture lower flammability limits using quantitative structure-property relationship models. Process Saf. Prog. https://doi.org/10.1002/prs.12103

Kohlbrand, H.T., 1991. Case history of a deflagration involving an organic solvent/oxygen system below its flash point. Plant/Operations Prog. 10, 52–54. https://doi.org/10.1002/prsb.720100110

Krishna, K., Kim, T.K., Kihm, K.D., Rogers, W.J., Mannan, M.S., 2003. Predictive correlations for leaking heat transfer fluid aerosols in air. J. Loss Prev. Process Ind. https://doi.org/10.1016/S0950-4230(02)00091-8

Krishna, Kiran, Rogers, W.J., Mannan, M.S., 2003. The use of aerosol formation, flammability, and explosion information for heat-transfer fluid selection, in: Journal of Hazardous Materials. https://doi.org/10.1016/S0304-3894(03)00273-5

Kwon, Y., Lim, H., Lim, Y., Lee, H., 2019. Implication of activity-based vessel emission to

improve regional air inventory in a port area. Atmos. Environ. 203, 262–270. https://doi.org/10.1016/j.atmosenv.2019.01.036

Lefebvre, A.H., Ballal;. D. R., 2010. Gas turbine combustion: alternative fuels and emissions. CPC press.

Luo, X., Wang, M., 2017. Study of solvent-based carbon capture for cargo ships through process modelling and simulation. Appl. Energy 195, 402–413. https://doi.org/10.1016/j.apenergy.2017.03.027

Mcgarry, J., 1983. Correlation and Prediction of the Vapor Pressures of Pure Liquids over Large Pressure Ranges. Ind. Eng. Chem. Process Des. Dev. 22, 313–322. https://doi.org/10.1021/i200021a023

Moreno-Gutiérrez, J., Calderay, F., Saborido, N., Boile, M., Rodríguez Valero, R., Durán-Grados, V., 2015. Methodologies for estimating shipping emissions and energy consumption: A comparative analysis of current methods. Energy 86, 603–616. https://doi.org/10.1016/j.energy.2015.04.083

Moreno-Gutiérrez, J., Pájaro-Velázquez, E., Amado-Sánchez, Y., Rodríguez-Moreno, R., Calderay-Cayetano, F., Durán-Grados, V., 2019. Comparative analysis between different methods for calculating on-board ship's emissions and energy consumption based on operational data. Sci. Total Environ. 650, 575–584. https://doi.org/10.1016/j.scitotenv.2018.09.045

Olmer, N., Comer, B., Roy, B., Mao, X., Rutherford, D., 2017. Greenhouse Gas Emissions From Global Shipping, 2013-2015. Int. Counc. Clean Transp. 1–25.

Pan, Y., Jiang, J., Wang, R., Cao, H., 2008. Advantages of support vector machine in QSPR studies for predicting auto-ignition temperatures of organic compounds. Chemom. Intell. Lab. Syst. https://doi.org/10.1016/j.chemolab.2008.03.002

Polymeropoulos, C.E., 1984. Flame Propagation in Aerosols of Fuel Droplets, Fuel Vapor

and Air. Combust. Sci. Technol. https://doi.org/10.1080/00102208408923807

Rakke, S.G., 2016. Ship emissions calculation from AIS-annotated.

Santon, R.C., 2009. Mist fires and explosions - An incident survey, in: Institution of Chemical Engineers Symposium Series.

Shu, G., Liang, Y., Wei, H., Tian, H., Zhao, J., Liu, L., 2013. A review of waste heat recovery on two-stroke IC engine aboard ships. Renew. Sustain. Energy Rev. 19, 385–401. https://doi.org/10.1016/j.rser.2012.11.034

Smith, T.W.P., Jalkanen, J.P., Anderson, B.A., Corbett, J.J., Faber, J., Hanayama, S., O'Keeffe, E., Parker, S., Johansson, L., Aldous, L., Raucci, C., Traut, M., Ettinger, S., Nelissen, D., Lee, D.S., Ng, S., Agrawal, A., Winebrake, J.J., Hoen, M., A., 2014. Third IMO Greenhouse Gas Study 2014. Int. Marit. Organ. 327. https://doi.org/10.1007/s10584-013-0912-3

Starcrest Consulting Group, L., 2014. Port of Long Beach Air Emissions Inventory in 2013, Journal of Chemical Information and Modeling. https://doi.org/10.1017/CBO9781107415324.004

Uyanik, T., Arslanoglu, Y., Kalenderli, O., 2019. Ship Fuel Consumption Prediction with Machine Learning, in: 4th International Mediterranean Science and Engineering Congress.

Van De Haar, A., Trapp, C., Wellner, K., De Kler, R., Schmitz, G., Colonna, P., 2017. Dynamics of Postcombustion CO2 Capture Plants: Modeling, Validation, and Case Study. Ind. Eng. Chem. Res. 56, 1810–1822. https://doi.org/10.1021/acs.iecr.6b00034

Wan, Z., Zhang, Q., Xu, Z., Chen, J., Wang, Q., 2019. Impact of emission control areas on atmospheric pollutant emissions from major ocean-going ships entering the Shanghai Port, China. Mar. Pollut. Bull. 142, 525–532. https://doi.org/10.1016/j.marpolbul.2019.03.053

Wang, B., Xu, K., Wang, Q., 2019. Prediction of upper flammability limits for fuel mixtures using quantitative structure–property relationship models. Chem. Eng. Commun. 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. Ind. Eng. Chem. Res. 56, 47–51. https://doi.org/10.1021/acs.iecr.6b04347

Yang, M.H., Yeh, R.H., 2015. Thermodynamic and economic performances optimization of an organic Rankine cycle system utilizing exhaust gas of a large marine diesel engine. Appl. Energy 149, 1–12. https://doi.org/10.1016/j.apenergy.2015.03.083

Yuan, J., Nian, V., 2018. Ship energy consumption prediction with Gaussian process metamodel. Energy Procedia 152, 655–660. https://doi.org/10.1016/j.egypro.2018.09.226

Yuan, S., Ji, C., Han, H., Sun, Y., Mashuga, C. V., 2020a. A review of aerosol flammability and explosion related incidents, standards, studies, and risk analysis. Process Saf. Environ. Prot. 146, 499–514. https://doi.org/10.1016/j.psep.2020.11.032

Yuan, S., Ji, C., Monhollen, A., Kwon, J.S.-I., Mashuga, C., 2019. Experimental and thermodynamic study of aerosol explosions in a 36 L apparatus. Fuel 245, 467–477. https://doi.org/10.1016/j.fuel.2019.02.078

Yuan, S., Zhang, Z., Sun, Y., Kwon, J.S.-I., Mashuga, C. V., 2020b. Liquid flammability ratings predicted by machine learning considering aerosolization. J. Hazard. Mater. 386, 121640. https://doi.org/10.1016/j.jhazmat.2019.121640

Zeng, M., Yuan, S., Huang, D., Cheng, Z., 2019. Accelerated Design of Catalytic Water-Cleaning Nanomotors via Machine Learning. ACS Appl. Mater. Interfaces 11, 40099–40106. https://doi.org/10.1021/acsami.9b14792

Zeng, X., Wang, Z., Ge, Z., Liu, H., 2007. Quantitative structure-property relationships for

predicting subcooled liquid vapor pressure (PL) of 209 polychlorinated diphenyl ethers (PCDEs) by DFT and the position of Cl substitution (PCS) methods. Atmos. Environ. 41, 3590–3603. https://doi.org/10.1016/j.atmosenv.2006.12.039

Zhang, Z., Yuan, S., Yu, M., Mannan, M.S., Wang, Q., 2020. A Hazard Index for Chemical Logistic Warehouses with Modified Flammability Rating by Machine Learning Methods. ACS Chem. Heal. Saf. https://doi.org/10.1021/acs.chas.9b00026

3. PROBLEM STATEMENT

In general, this study follows shipping fuel sustainability analysis, ship onboard carbon capture process design, inherent safety properties prediction of promising fuel options, and novel inherent safety criterion for flammable liquids development. Many uncertainties are associated with TTP process, such as dual engine emission factor, SFC and load factor determination. This study overcomes many weaknesses by employing new equations on determination of load factor and fuel consumptions, bringing a novel parameter to compare emissions between different ship classes, adopting the updated emission sources and considering the LNG dual engine flexibility and BCA emission for GWP related ship exhaust gases. Aiming at providing a holistic way to find the most efficient and sustainable PCC solution for an LNG tanker, the TTP carbon capture work proposes a system of the tank to propeller post-combustion carbon capture which integrates ship engine process modeling with chemical absorption/desorption process modeling techniques. Meanwhile, one supervised ML study is conducted for the prediction of liquid aerosolization contributors, while another unsupervised study on fuel safety clustering is completed to classify the marine fuels. Many unique works have been carried out to answer the below questions:

- What are the emission sources for ship cruising mode?
- Which marine fuel has a good emission control for GWP gases/ non-GWP gases/ black carbon aerosol?
- How does the newly blended fuel option perform on emission and safety?
- Which type of merchant ship is preferred for energy saving in the tank to propeller process?
- Is there any uniqueness of TTPPCC process design and simulation?
- Which key factors are dominated to determine the TTPPCC performance?

- What is the next focus on maritime carbon capture system?
- How to classify the shipping fuel combustion safety in CIE?
- Which variables are selected for the prediction of the liquid aerosolization features and how to select the significant variables?
- Are QSPR/QSAR models good to predict the liquid aerosolization features?
- Which thermodynamic properties of the liquid impact the consequence of fuel combustion?
- How is the preference of fuel safety clustering approaches determined?
- Which marine fuel is classified as the safest option? Which fuel option may lead the shipping market to meet the long-term IMO strategies?

Firstly, this study proposes a bottom-up approach to estimate the GWP and non-GWP gases as a result of maritime activities during ship cruising mode. By introducing the ship capacity bins, emissions and fuel consumptions of ship classes with same unit of ship capacity are compared to illustrate the recommended size for each ship type, and a case study based on our proposed mathematical models have been conducted as well. In the second part, the chapter provides a reliable prediction for the identified liquid aerosol safety parameters based on a thorough comparison among machine learning algorithms, with an emphasis on the liquid aerosolization. Secondly, to increase the carbon capture efficiency and decrease the energy penalty, the TTPPCC study provided detailed steps for process model development covering innovative absorber/stripper design under variation of solvents, packed type, and liquid to gas ratio. This work also included a thorough sustainability evaluation based on emission reduction efficiency, energy penalty, and carbon cyclic capacity among two single aqueous amines, MEA and diisopropanolamine (DIPA), and one blended amine with a promoter, methyldiethanolamine (MDEA) with piperazine (PZ).

Moreover, four categories of regression algorithms with consideration of principal component analysis are applied in the liquid aerosol safety prediction study by developing quantitative structure−property relationship (QSPR) models. 1215 liquid chemicals and 14 predictors have been input to train the developed machine learning models via k-fold cross validation with the consideration of principal component analysis. Three rounds of model performance comparisons are conducted to find the optimal models for liquid dynamic viscosity (LDV), surface tension (ST), and liquid vapor pressure (LVP). Lastly, the marine fuel clustering section is proceeding towards nine variables in three categorized liquid combustion safety matrices via two unsupervised algorithms. The overall liquid safety performance is evaluated by a novel combustion risk index via the weight values determined by the information entropy approach. This index can be used to explore inherently safer fuels in the process industries.

4. TTP SUSTAINABILITY: SHIPPING FUEL EMISSIONS WITH CONSIDERATION OF BLACK CARBON AEROSOLS*

This chapter is aiming at estimating the fuel consumption and air pollutant emissions during the cruising mode in non-SECAs by adopting IMO 2020 compliant fuel options. The ships' characteristics and machinery data are collected from the Llyod's Maritime database, with the given data of ship capacity, draught, length between perpendicular, main engine type, power, service speed, *etc.* Multiple sources are employed to extract the important energy-based emission factors and the SFC data, collected from two major sources, recalculated as carbon dioxide intensity to convert energy-based emission factors to fuel-based emission factors.

4.1. Methodology

4.1.1. Description of Methodology

The bottom-up approach is more reliable than the top-down method since it deals with more uncertainties. Figure 4.1 is a flowchart showing the key tasks and collected data in the proposed methodology. We have compared all the milestone works of this field and constructed the fuel consumption and shipping emission models by covering more operational uncertainties (adverse weather impacts for ship cruising, hull fouling effects for shipping fuel consumption and ship draught adjustment), adopting the latest ship dimension and machinery data from Lloyd's List Intelligence and collecting the dynamic data for six major types of merchant ships from shipboard AIS. To provide a close-to-reality shipping emission data, the paper makes a great effort to minimize the deterministic input values and to consider the emission for main engine, auxiliary engine, and auxiliary boiler for a large sample ship size, 337 in total. As the IMO 2020 Sulfur Cap is an ambitious move for global

* Reprinted with permission from "Ji, Chenxi, and Mahmoud M. El-Halwagi. "A data-driven study of IMO compliant fuel emissions with consideration of black carbon aerosols." *Ocean Engineering* 218 (2020): 108241."

shipping, especially for the non-ECAs, the study focuses on the cruising mode at sea, and the ship service speed is defined as the speed over ground, which is 80% of the normal continuous rating (NCR) speed. Six ship types, bulk carriers, general cargo ships, container ships, oil tankers, LNG carriers, and LPG tankers are studied in this research since their shares of transportation capacities top other merchant ships among world fleet (UNCTAD, 2019). In addition, this paper considers the four proven shipping fuel solutions (HFO with scrubber, MGO, VLSFO, and LNG) instead of paying attention on the future-promising fuel options, such as biofuels, ammonia, and liquid hydrogen. There is no doubt that shipbuilding has an evident large-scale trend, and our focus is the oceangoing vessels, so the 2-stroke slow speed engine is our major analysis object for oil-fueled main engine, while the 4-stroke engine is assumed for the auxiliary engine. As for LNG-fueled ships, the common techniques may involve steam turbines, gas turbines, lean burn spark-ignited (LBSI) engines, low-pressure injection dual-fuel (LPDF) engines, and high-pressure injection dual-fuel (HPDF) engines. No gas turbine or steam turbine engines are analyzed in this research because they have limited future international shipping applications and are less efficient compared to other LNG engines. Therefore, the LPDF and HPDF solutions are adopted for the cruising mode emission calculation when LNG is taken as the fuel solution to determine the environmental impacts.

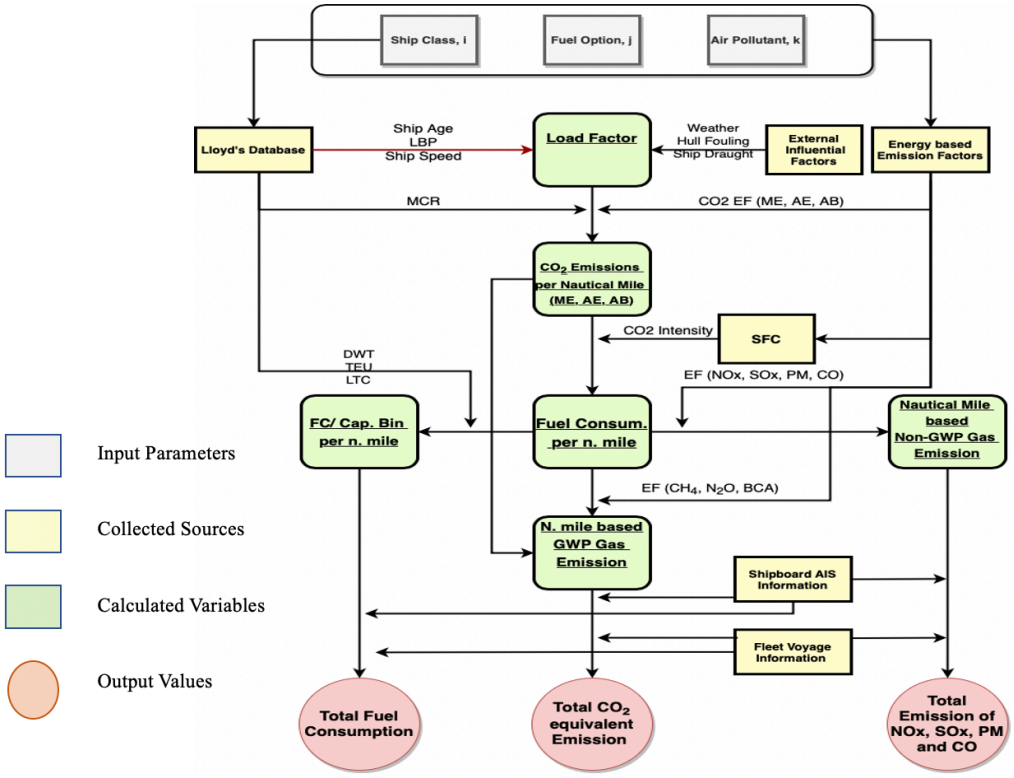


Figure 4.1 A flowchart describing the proposed methodology

As a bottom-up emission approach, this work has completed individual calculations ship by ship using the detailed calculation process shown in Figure 4.1. In general, the given parameters, displaying in the yellow block, cover ship characteristic data, energy-based emission factors, specific fuel consumption (SFC) data, AIS information, and fleet voyage plan. This study firstly preprocesses the collected ship characteristic data from the Lloyd's Maritime Intelligence Unit (LMIU), then the load factors are worked out by integrating ship characteristics and external influential factors, such as the hull fouling factor, the draught adjusted factor, and extreme weather factor; the emission factors are compared and validated among ten sources, and after the AE and AB power demand is calculated case by case, the total CO_2 emissions per nautical mile by one fuel option can be determined. Next, the fuel consumption is obtained by adopting a converted emission factor, which is integrated into the fuel-based emission factor with the SFC data. Then, the study proposed a new variable to show the amounts of fuel consumption and GWP related gas emissions per nautical mile for

one unit of capacity bin so that every unit of transported merchants can be connected with fuel consumption and CO_2 equivalent emissions. Finally, the emission outputs of other pollutants (NOx, SOx, CO, and PM) are finalized by inputting relevant emission factors, fleet voyage information, and real-time data from shipboard AIS. By introducing a unique allocation unit, i.e., shipping air pollutants per nautical mile per capacity bin, the work presents a heuristic approach to reasonably estimate fuel consumption, GWP emissions, and common non-GWP air pollutants. Instead of adopting deterministic values of load factors, this study considered many external factors in the calculation process and the simple linear regression of SFC is avoided by calculating fuel-based emission factors. Also, the proposed model firstly compared four proven IMO 2020 fuel solutions with consideration of LNG dual engines. In addition, this chapter integrated black carbon aerosol emission with other three common GWP air pollutants and the auxiliary boiler is not neglected in this study as well. LMIT provided real-time AIS data for the defined ship types and routes. Python 3.0 was adopted in this study to collect and to preprocess the ship characteristic data. JMP 15 and StatTools 7.6 were applied to process the ship-by-ship calculations for fuel consumption and marine exhaust gas emissions, and they were served to analyze the calculated data and to make regression models for the case study as well.

4.1.2. Merchant ship exhaust gas emissions

Some previous studies used a linear relationship between ship emissions per nautical mile and the total emissions as the shipping speed was presumed as a constant value. This assumption is reasonable for the cruising mode, especially for the Non-ECAs since it is a common practice to adopt auto-pilot mode when the ship is navigating in the stable external environment at the deep sea. The paper works on the cruising mode emissions for marine air pollutants when the selected merchant ships are navigating in the Non-ECAs, comparing the

amount of ship exhaust pollutants emitted by the selected merchant ships when different fuel options are adopted.

$$Em_{pm_{(i,j,k)}} = \sum \Big[(P_{ME_{MCR_{(i,j,k)}}} \cdot LF_{ME_{(i,j,k)}} \cdot EF_{ME_{(i,j,k)}} + P_{AE_{MCR_{(i,j,k)}}} \cdot LF_{AE_{(i,j,k)}} \cdot EF_{AE_{(i,j,k)}} + P_{AB_{ep_{(i,j,k)}}} \cdot EF_{AB_{(i,j,k)}}) \cdot V_{act}^{-1} \Big] \quad (4.1)$$

i: Ship class

j: Fuel option

k: Air pollutant

Em_{pm}: Emissions for ship class i, fuel option j and air pollutant k per nautical mile

$P_{ME_{MCR}}$: main engine maximum continuous rating power

LF_{ME}: main engine load factor

EF_{ME}: main engine energy-based emission factor

$P_{AE_{MCR}}$: auxiliary engine maximum continuous rating power

LF_{AE}: auxiliary engine load factor

EF_{AE}: auxiliary engine energy-based emission factor

$P_{AB_{ep}}$: auxiliary boiler effective power demand

EF_{AB}: auxiliary boiler energy-based emission factor

V_{act}: ship actual speed, equals to ship design speed times speed over ground multiplying speed adjusted factor

As shown in the equation 4.1, the ship exhaust gas emission per nautical mile is integrated with six ship types($i \in \{Bulk\ Carrier, General\ Cargo\ Ship, Container\ Ship, LNG\ Tanker, LPG\ Tanker, Oil\ Tanker\}$), five fuel options ($j \in \{HFO\ with\ Scrubber, MGO, VLSFO, LPDF, HPDF\}$) and eight pollutants ($k \in \{CO_2, CH_4, N_2O, BCA, NO_x, SO_x, PM, CO\}$) for the main engine, auxiliary engine, and auxiliary boiler exhaust gases. Specifically, the main engine maximum

continuous rating (MCR) power is obtained by Lloyd's database dividing 0.9 since NCR conventionally equals to 85~95% MCR; The AE maximum continuous rating power is calculated by using the vessel ratio of Auxiliary Engines/Main Engines, shown in the Table S-3 of Appendix A; The AB effective power demand is determined by the ship capacity bins within the specific ship class, see Tab. S-4 of Appendix 4.

4.1.3. Main engine load factor

This study compares the load factor, SFC, and emission factor for six previous studies, and Table 2.1 shows the summary of the major studies. One can find that most of the studies took the shipping speed as the key parameter for load factor to simplify the calculation, however, a more robust is needed to cover uncertainties during the voyage and to deal with different vessel types, ages, and engines with different leading factors. The main engine load factor tends to be greater than the actual value since seldom of them considered the safety margin and navigation environment impacts. Referred from the ICCT report (Olmer et al., 2017), this study substituted the transient speed with an adjusted ship speed, and the negative effects of hull roughness, weather adjusted factors and draught coefficient are added in the load factor equation, see equation 4.2.

$$LF = \left(\frac{Vact}{V_{MCR}}\right)^3 \cdot HFF_i \cdot WAC \cdot DAF_i \tag{4.2}$$

where V_{act} is the actual ship speed, which equals to the speed over ground multiplying the speed adjusted speed; V_{MCR} is the speed when ME is in maximum continuous rating mode, which is assumed as 1.064 times the normal continuous rating speed because the cruise speeds proposed by Lloyd's data are 94% of the maximum speed (Moreno-Gutiérrez et al., 2015); HFF is the hull fouling factor, which is determined by the ship age, length between perpendiculars and the coefficient of marine biofouling; WAC, weather adjustment coefficient, and DAF, draught adjustment factor.

4.1.4. Ship fuel emission factor

A key element to reflect the actual emissions from theoretical calculations is to adopt the accurate emission factor. Overall, an emission factor is a coefficient which allows to transforming activity data into air polluter emissions. As a critical parameter for the typical fuel converter, emission factor has been analyzed for many decades. IPCC has an emission factor database, IPCC EFDB (IPCC, 2019); The EU has the EMEP EEA database emission factors (EMEP/EEA, 2019) and the USA EPA has AP-42, a compilation of emission factors while in Australia there are some NPI emission factors within the emissions estimation handbooks (NPI, 2017). Besides these open-access data sources, many works have specified emission factor estimation in the maritime sector. The IMO has published three emission studies on shipping fuel options and the emission factors for each fuel have been keeping updating as they have great ambitions to build a low carbon and Sulfur ecofriendly shipping environment. Besides, the International Council on Clean Transportation (ICCT) pays plenty of attention to emission control in both aviation and marine transportation, and their emission factor data for ME, AE, and AB have been adopted by many scholars

As shown in Tab. S-1 of Appendix A, the selected ships are mega ocean-going vessels, thus, the two-stroke SSD engine is the focus for diesel-based main engines. Table 4.1 listed the emission factors of currently compliant fuel options. Emission factor for HFO with the maximum sulfur content of 2.5% for slow speed diesel engine, which is served as a benchmark for the major option for the prior IMO 2020 shipping fuel, is collected from the ICCT study (Olmer et al., 2017). As a widely applied retrofit option to meet the 2020 Sulfur Cap, an open-loop scrubber is adopted in the study, so no addition of alkali is needed. The scrubber is modeled to reduce the emissions of SO_2 from HFO with 2 percent sulfur content to the same levels as the SO_2 emissions from a fuel within 0.5 percent of sulfur content and

the reduction required by the scrubber unit is thus 95 percent. The fuel consumption increases with a scrubber (van der Linde, 2009), and a 2 percent increase is assumed in this study.

The scrubber unit is modeled to remove 50 percent of the particulate matters. As the open-loop scrubber for reducing black carbon aerosol ranging from 25% to 70%, a 40-percentage reduction of BCA is assumed in this study. Moreover, emission factors for HFO with scrubber include a 2 percent energy consumption penalty as well. As for VLSFO, BCA emission factor is assumed to be 35 percent more than that of HFO, since a recent study (Finland and Germany, 2019) criticized VLSFO even increased the BCA emission by 10%~85% comparing with high sulfur fuel oils, making the newly blended fuel limited acceptable for the Arctic transportation, see Tab. S-2 in Appendix A for BCA fuel-based emission factors.

Table 4.1 Emission factors of currently available fuel solutions

	EF CO_2 (ME/AE /AB)	**EF CH_4 (ME/AE/AB)**	**EF N_2O (ME/AE/ AB)**	**EF SO_x (ME/AE /AB)**	**EF PM (ME/AE /AB)**	**EF NO_x (ME/AE /AB)**	**EF CO (ME/AE /AB)**
HFO 2.5% S(SSD)	607/707/950	0.01/0.01/0.002	0.03/0.04/0.05	10.29/11.98/16.1	1.42/1.44/ 0.93	14.4/11.2/ 2.1	0.54/0.54/ 0.2
MGO 0.1% S	519/602/962	0.0053/0.0055 /0.002	0.027/0.029 /0.04	0.37/0.43/0 .57	0.19/0.19/ 0.1	13.54/10.53/2.0	0.54/0.54/ 0.2

VLSFO 0.1~0.5% S	533/557/962	0.0052/0.0055/0.002	0.027/0.028/0.04	0.51/0.6/0.81	0.20/0.20/0.11	13.54/10.53/2.0	0.54/0.54/0.2
HFO with Scrubber	576/602/955	0.0054/0.0056/0.002	0.028/0.029/0.05	0.5/0.6/0.8	0.71/0.72/0.47	14.4/11.2/2.1	0.54/0.54/0.2
Dual Engine I, LPDF	417/445/445	0.13/0.13/0.13	0.011/0.011/0.011	0.14/0.14/0.14	0.1/0.1/0.1	3.4/2.4/2.4	0.3/1.3/1.3
Dual Engine II, HPDF	445/445/445	0.12/0.12/0.12	0.01/0.01/0.01	0.24/0.24/0.24	0.13/0.13/0.13	8.76/2.4/2.4	0.79/1.3/1.3

Moreover, emission factors for CO_2, CH_4 and N_2O of HFO with scrubber, VLSFO, MGO, LPDF, and HPDF are calculated from the GREET model (ANL, 2016) and the ICCT latest report (Pavlenko et al., 2020); while the SO_x, NO_x, PM, and CO are collected from the ICCT study (Olmer et al., 2017).

For LNG-fueled options, the LNG LPDF and HPDF are the main engine types, while LPDF is mostly applied for LNG carriers and large-scale container ship, and HPDF, another slow-speed two-stroke engine, is a proven solution for bulk carriers, general cargo ships and gas tankers. Furthermore, the LPDF engine has a medium-speed version, which will not be discussed in this study because it usually serves as the main power for cruising ships, passenger ferries, and offshore support vessels. Emission factors of CO_2, CH_4, and N_2O for

the dual engines are collected from the latest study (Pavlenko et al., 2020) and the emission factors of SO_x, NOx, PM, and CO for dual engine are collected from two studies (Kristensen, 2015) (Man B&M, 2012).

4.1.5. Fuel consumption per unit of capacity bin

Fuel consumption is determined by many variables, which can be categorized as engine attributes (engine type, load, maintenance state) and fuel types (RM, DM, LNG, *etc.*). To mathematically determine the fuel consumption, it is necessary to convert the energy-based emission factors of marine fuel options to fuel-based emission factors by introducing the SFC. Specifically, the amounts of CO_2 emission combusting from marine fuels are called CO_2 intensity of the fuel, which is listed in the following table.

Table 4.2 CO_2 intensity of selected fuels

Fuel Type	CO_2 intensity of selected fuel (g CO_2/g Fuel)
HFO[1]	3.114
MGO[1]	3.206
VLSFO[1]	3.154
LPDF[2]	2.69
HPDF[2]	2.39

1. Source from the IMO 3rd study (Smith et al., 2014)
2. Calculated from the latest ICCT study (Pavlenko et al., 2020)

This study covers the shipping fuel emissions by all the emission sources on board, and the fuel consumption is calculated by CO_2 emissions of ME, AE, and AB. Then the calculated fuel consumption is adopted to compute the fuel consumption per unit of capacity bin on a distance basis to compare the energy consumption of each ship class. The calculation process is taken on a ship-by-ship basis for our collected ship database, and the equation is listed below.

$$FC_{cb_{i,j,cbt}} = \sum_{i \in Class} \sum_{j \in Fuel} [(\frac{Em_{CO_2 ME\ i,j,cbt} + Em_{CO_2 AE\ i,j,cbt} + Em_{CO_2 AB\ i,j,cbt}}{CO_2\ Intensity}) / Capacity\ Bins] \quad (4.3)$$

$FC_{cb_{i,j,cbt}}$: Fuel consumption per unit of capacity bin per nautical mile for ship class i, fuel option j within the specific capacity bin type.

$Em_{CO_2 ME\ i,j,cbt}$: Amount of CO_2 emission per unit of capacity bin per nautical mile emitted by main engine.

$Em_{CO_2 AE\ i,j,cbt}$: Amount of CO_2 emission per unit of capacity bin per nautical mile emitted by auxiliary engine.

$Em_{CO_2 AB\ i,j,cbt}$: Amount of CO_2 emission per unit of capacity bin per nautical mile emitted by auxiliary boiler.

Since many variances lie on the sample size, ship capacity bins, and ship main power systems, the cumulative fuel consumption and emissions are hard to find the most sustainable fuel options. Thereafter, the capacity bin is introduced among all the six ship classes to calculate the fuel consumption and GWP CO_2 equivalent emissions per nautical mile per unit of capacity bin.

4.1.6. Global warming potential related gas emission calculation

This paper adopted the IPCC fifth report GWP_{100} values for CO_2, CH_4, and N_2O, and the BCA GWP_{100} was referred from Bond's study (Bond et al., 2013). The below equation shows the way to calculate the fleet's total emission of GHGs by considering the emissions of BCA.

$$\sum_{i \in Class}^{6} \sum_{j \in Fuel}^{5} Em_{CO_2 e_{(i,j)}} = \sum Em_{CO_{2(i,j)}} + GWPC_{CH_4} \cdot \sum Em_{CH_{4(i,j)}} + GWPC_{N_2O} \cdot \sum Em_{N_2O_{(i,j)}} + GWPC_{BCA} \cdot \sum Em_{BCA_{(i,j)}} \quad (4.4)$$

$Em_{CO_{2e(i,j)}}$: GWP equivalent CO_2 emission for ship class i, fuel option j

Em_x: Emission of CO_2, CH_4, N_2O, and Black Carbon Aerosol

$GWPC_x$: Global Warming Potential Coefficient of CO_2, CH_4, N_2O, and Black Carbon Aerosol

Following the timeline of IMO's initial GHG strategy, up to 20% reduction in the carbon intensity of international shipping needs to be accomplished this year. The GWP related ship exhaust gases are determined as CO_2, CH_4, N_2O, and BCA in this paper, although other substances like CF_4 and HFC-152a were also deemed as air pollutants for GWP. This research does not consider these non-combustion emissions such as refrigerants, and halogenated hydrocarbons due to the wide variations of the emissions factors of non-combustion sources and little significance to overall shipping GHG emissions, and the non-GWP gas emission estimations follow the typical way to integrate emission factors of SO_x, NO_x, PM, and CO with the calculated fuel consumption.

4.2. Results and discussion

Figure 4.2 shows the fuel consumption and shipping exhaust gas emissions contributing to global warming potential for the six ship types and five designated fuel options under IMO 2020 Sulfur Cap. In general, the total values of fuel consumption per nautical mile for container ships surpass those of the other five ship classes, while the GWP related gas emissions show the same trend as well. Focusing on emission amounts only, CO_2 emissions consistently account for more than 90% among the four GWP related gases, and for the two common abatement option, HFO with Scrubber and VLSFO, the BCA is the second pollutant, then CH_4 and N_2O, while for LNG-based fuel options (LPDF and HPDF), the emission sequence after CO_2 is CH_4, N_2O and BCA.

As for the shipping emission sources, the air pollutants emitted by the main engine are ten times more than those emitted by auxiliary engine and auxiliary boiler. Also, the cruise power demand of boilers for bulk carriers, container ships, and general cargo ships are assumed zero in this study based on the ship maneuvering experiences. Next, one conclusion can be made that the LNG-based fuels have great emission controls on CO_2, N_2O, and BCA, and the amounts of BCA emissions only account for 1%~5% of HFO with scrubber and VLSFO based on our calculation, see Appendix A Tab. S-5 for detail; However, the only issue for LNG-fueled ship is the CH_4 emission, which is around 20 times more than that of oil-based fuel options, and MGO, VLSFO present an obvious strength on methane emission control, while HPDF beats LPDF within the LNG-based fuel options. Specifically, the data for bulk carrier indicates that HFO with scrubber emits the most amounts of CO_2, CH_4 and N_2O among the three oil-based fuels, while the VLSFO emits the most BCA. Compared to HFO with scrubber option, the LPDF and HPDF can reduce CO_2 emissions by 28% and 36%, respectively, and the emissions of N_2O are dropped by 61% and 64% as well. As for general cargo ships, the shipping pollutants emitted by the auxiliary engine have a remarkable rise than other ship classes, that is because most general cargo ships require more power from AE, causing the ratio of AE/ME much higher than other ships. Although the methane emissions from Figure 4.2 indicate a sharp rise for LPDF and HPDF, it does not conclude that those two are not eco-friendly options as the GWP is calculated depending on the four pollutant emissions and CO_2 emissions may make a great difference. On the other hand, the cumulative results between LNG tankers and LPG tankers change dramatically mainly because of the larger ship size of LNG tankers.

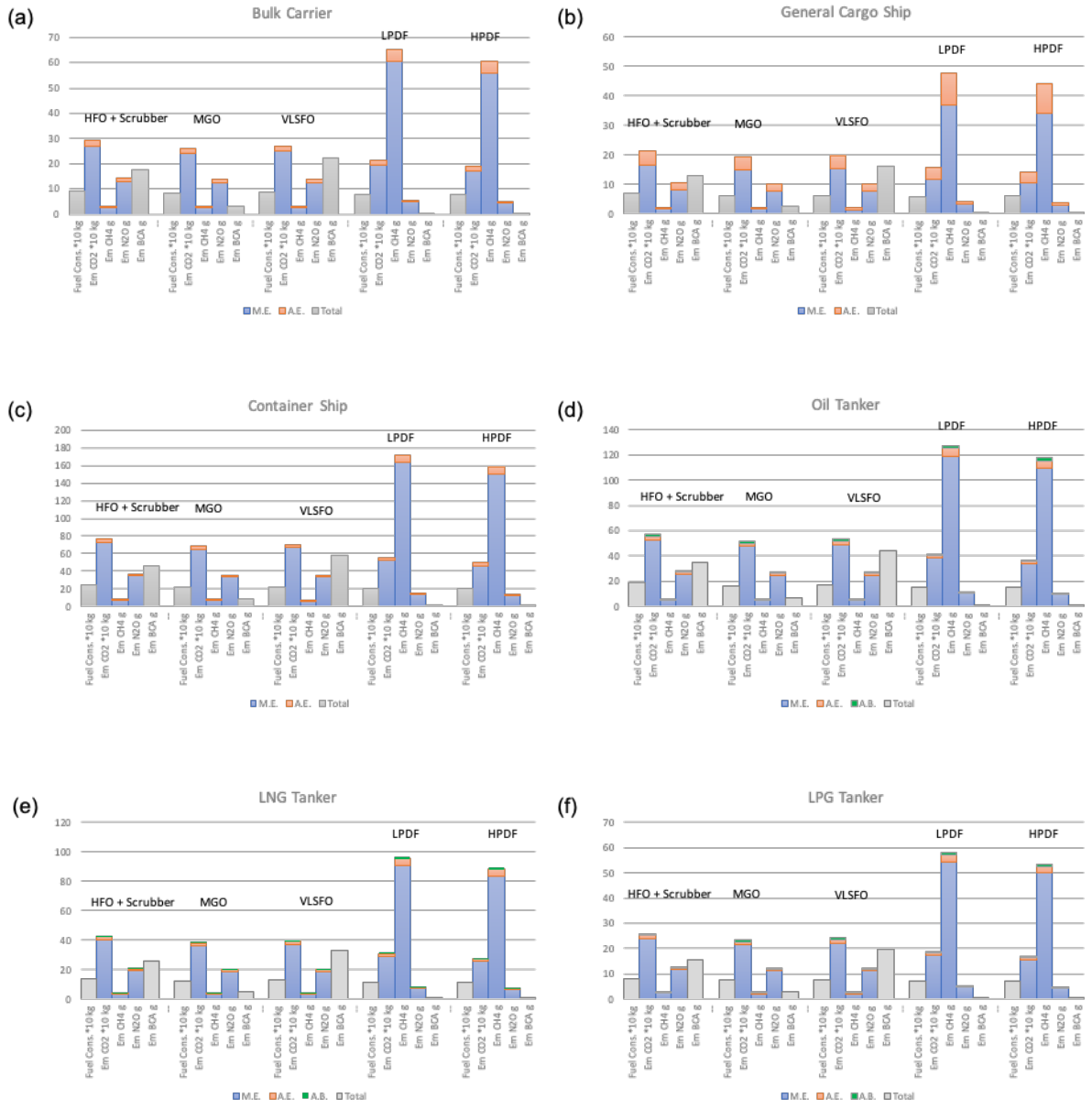


Figure 4.2 Fuel consumption and GWP related gas emissions per nautical mile for six specific ship types

Besides the GWP related gas emissions, other pollutants resulting from the ship activities should be considered as well. This study compared the nautical mile-based emissions of CO, PM, NO_x, and SO_x for five fuel options among six merchant ships, and the calculated mean emission values with 5% confidence interval are shown in Figure 4.3. The amounts of NO_x emission dominate the selected pollutants, and the LNG fueled engines display a less NO_x

emission than that of oil-fueled engines. As for PM, the most amount of emission occurs when HFO with scrubber is chosen, and still, LPDF and HPDF emits less PM than oil-based fuels. Speaking to SO_x emission, VLSFO, and HFO with scrubber top other options, but CO emissions tell another story, the HPDF option shows the worst CO emission. Therefore, one conclusion can be made that the LPDF is the best shipping fuel option based on the cumulative emission results of CO, PM, NO_x, and SO_x.

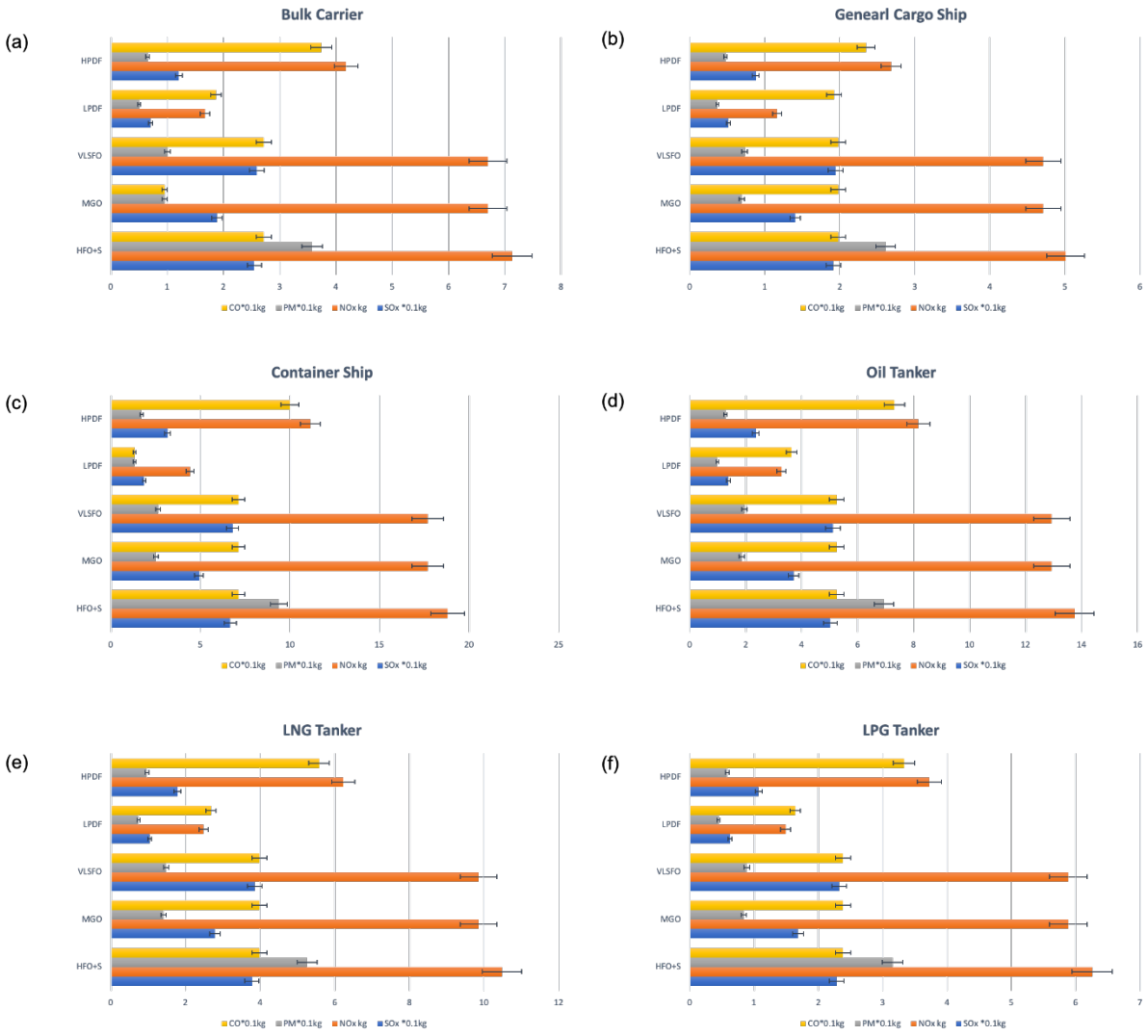


Figure 4.3 Amounts of emissions for CO, PM, NO_x, and SO_x per nautical mile

To compare the fuel consumption and CO_2 equivalent emission between different ship classes, the nautical mile-based fuel consumption/ CO_2 equivalent emission per capacity bin

was brought to categorize six ship types into three groups. Figure 4.4-4.6 presents the relationship between CO_2 equivalent emissions per nautical mile per unit of capacity bin for six ship types and the values of DWT, LTC, and TEU, respectively, and the corresponding fuel consumption are shown in Fig. S-1 and Fig. S-2 in Appendix A. Overall, all unit emission values display a decreasing trend as the capacity bin increases, suggesting larger-size ships exhibit better environmental performance. Specifically, bulk carrier, as shown in Figure 4.4(a), outperforms general cargo ship (Figure 4.4(b)) on exhaust gas emissions. There is no clear pattern for general cargo ships with less than 15,000 DWT capacity, but the unit emission drops along with the increment of capacity bins when the DWT value grows over 15,000. Moreover, LNG-based options are more eco-friendly than oil-based options on either bulk carrier or general cargo ship.

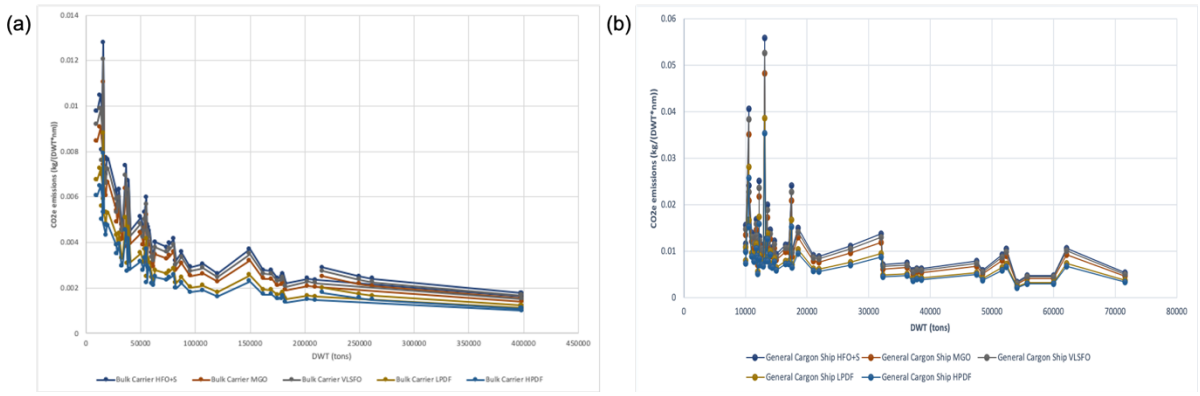


Figure 4.4 Nautical mile-based CO_2 equivalent emissions per DWT for bulk carrier (4.4a) and general cargo ship (4.4b)

Liquid tank capacity is set as the capacity bin unit for the three tankers in this study. Figure 4.5 illustrates the most sustainable option is the oil tanker as a result of its overall large ship scale. In general, the oil-based fuels do not perform as well as the two LNG-fueled engines, and HPDF shows a better performance than LPDF on CO_2 equivalent emissions among these three tankers. Furthermore, the difference between oil-based engines and LNG-based engines is tremendous when the ship size is small. Based on our calculated outputs, LNG tankers with

LTC of 30,000 cubic meters presents the most unit CO_2 equivalent emission, indicating the investment of LNG tanker in this scale may not be a wise choice; When LTC grows to 150,000 LTC, the value is dropped below 0.004 kg/n.m./LTC, and there is no big difference when the LTC locates in the range of 150,000 and 180,000 except for two abnormal ship types with their LTC of 173,400. The unit emission value for LPG tanker, shown in Figure 4.5b, plunges to 0.005kg/n.m./LTC when the x-axis value exceeds 23000 cubic meters, but there are two exceptional data points, *i.e*., LTC=36,000 and LTC=38,000. For oil tankers, the CO_2 equivalent unit emission is below 0.004 kg/n.m./LTC at the time when LTC surpasses 180,000 cubic meters or stays in the range between 100,000 and 130,000. Therefore, the recommended tanker sizes from the perspective of CO_2 equivalent emission savings are as follows: LNG tanker with liquid tank capacity larger than 150,000 cubic meters; LPG tanker with LTC greater than 23,000; Oil tanker whose LTC in the range between 100 thousand and 130 thousand and greater than 180 thousand cubic meters.

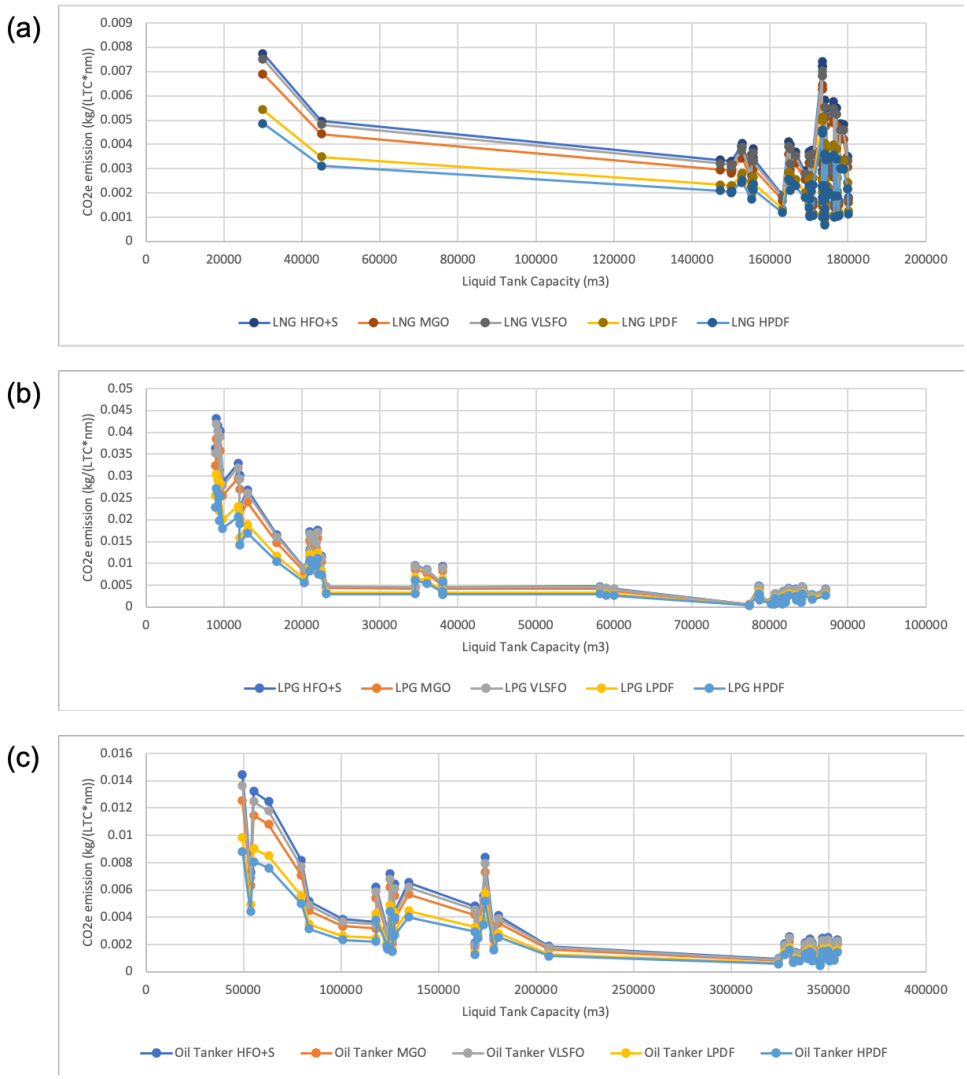


Figure 4.5 CO_2 equivalent emissions per nautical mile per unit of capacity bin for LNG tanker (4.5a), LPG tanker (4.5b) and oil tanker (4.5c)

The capacity bin of container ships is unique among all the six merchant ships, and the unit of the fuel consumption, which is the y-axis in Figure 4.6a, is kilogram fuel per TEU per nautical mile. 0.04 kg fuel per TEU every nautical mile is achieved when the shipping capacity surpasses 5,000 TEU; the lowest value of fuel consumption occurs when the TEU is 18,982, of which the values of MGO, VLSFO, LPDF, and HPDF are 87.5%, 91.4%, 83.9%, and 84.4%, respectively, as much as that of HFO with scrubber. As for Figure 4.6b, the capacity unit CO_2 equivalent emission goes below 0.1 kg/n.m./TEU once the capacity bin reaches 5000 TEU; At that point, when taking HFO with scrubber as a benchmark, the MGO

emission can be saved by 13.6%, while the unit emission saving for VLSFO, LPDF and HPDF are 5.8%, 31.1%, and 38.4%, respectively. Thus, the small size container ship, whose TEU is less than 5,000, is not a recommended choice when considering the fuel cost and GWP gas emission.

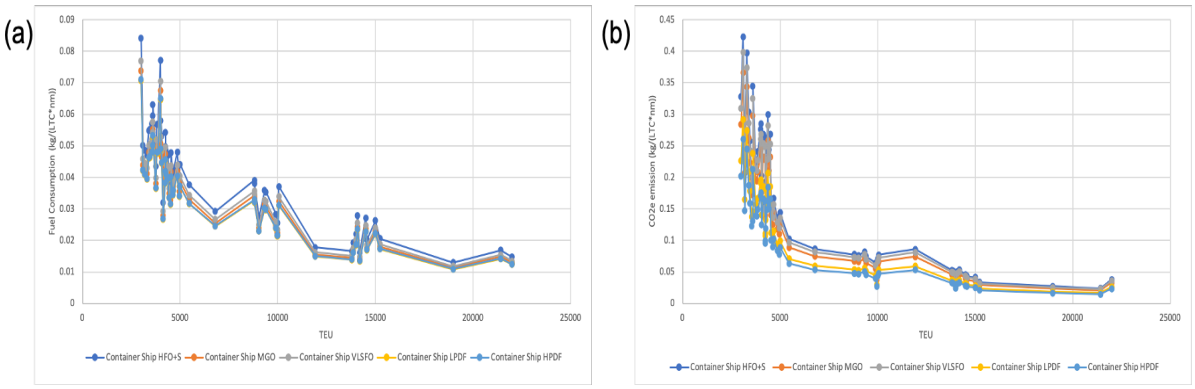


Figure 4.6 Nautical mile-based fuel consumption (4.6a) and CO_2 equivalent emissions (4.6b) per TEU for container ship

Next, the BCA emissions for our database were calculated based on the unit of capacity bin for every nautical mile. Figure 4.7 shows the BCA emissions for two capacity bin groups, one can conclude that the two LNG fuel options can reduce BCA emissions into a neglectable level. Specifically for DWT basis fleets shown in Figure 4.7a, general cargo ship may emit around 2.5 times more black carbons than bulk carriers, while the BCA emission can be saved up to 94 times by replacing oil-based fuels with LNG fuel options; as for the LTC basis shown in Figure 4.7b, the best BCA emission control tanker is LNG tankers, it emits 65% less black carbons than LPG tankers, meanwhile oil tankers only emit 7% more than LNG tankers when employing same fuel options. For container ships, suppose the benchmark choice is HFO with scrubber, by substituting it with other options, the BCA emissions decreased 82% for MGO, 98% for LPDF and HPDF, however, the VLSFO emits 25% more black carbons than the benchmark fuel option.

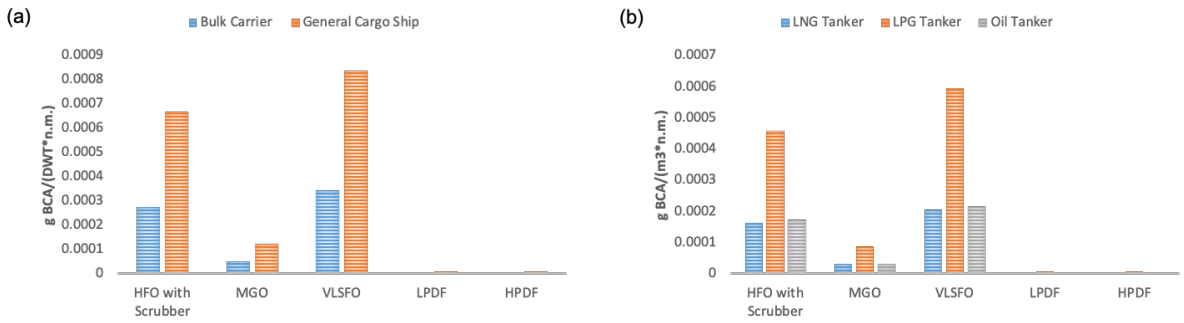


Figure 4.7 BCA emissions per DWT for bulk carrier and general cargo ship (4.7a) and BCA emissions per LTC for LNG, LPG, and oil tankers (4.7b)

To sum up, GHG emissions follow the rank of HFO with scrubber > VLSFO > MGO > LPDF > HPDF, while the order for non GHG emissions is HFO with scrubber > VLSFO > MGO > HPDF > LPDF. Specifically, the two LNG options outperform the three oil options on emissions of non-GWP emissions except for the CH_4 emission, which is reasonable because of the concern of methane slip, and CO emission for HPDF is as well interior than others. On the other hand, the between-group calculated results confirmed the large-size ships of our database tend to have a good performance on fuel saving and GWP gas emission control. In detail, bulk carriers generally have lower values of the GWP gas emissions per DWT than general cargo ships; The three tankers, which share the liquid tank capacity as the capacity bin, have little difference on emission control for large-size tankers, but LPG tankers have better CO_2 equivalent emission records for medium-size tankers, ranging from 20,000 to 100,000 cubic meters; Set TEU as a unique capacity bin unit, container ships emit more air pollutants per unit than other ship types, which is attributed to its fast serving speed and large transport capability per TEU. For black carbon emissions, VLSFO, serving as a newly blended IMO 2020 compliant option, is surprisingly to emit the largest amount of BCA among all the fuel options; while HPDF and LPDF have overwhelming performance in BCA emission control over the two residual oil options, 60 to 100 times BCA emissions will be saved by taking LNG fuel options rather than HFO and VLSFO. Based on our calculations,

the recommended ship types when considering BCA emission are bulk carriers and LNG tankers respectively for the DWT group and the LTC group.

4.3. Case study

The proposed mathematical models for fuel consumption and GWP gas emissions are employed on a specific case for trans-ocean cruising from Houston (located in the North America ECA) to Rotterdam (located in North Sea SECA) from Oct. 1 to Oct. 31, 2019, aiming to find an optimized fuel option by minimizing the fuel consumption and environmental impacts.

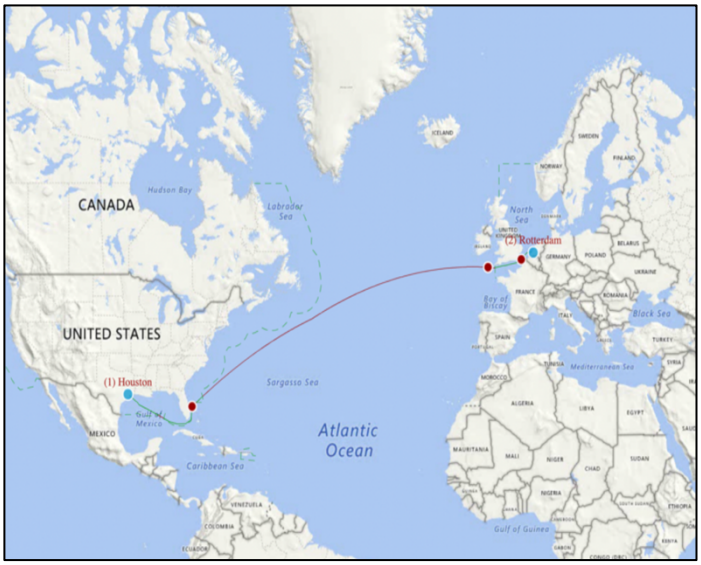


Figure 4.8 Route plan for the Houston to Rotterdam voyage

Based on the AIS collected data, there were 234 ships navigating from Houston to Rotterdam. The total distance of this voyage is 5,089 nautical miles of which 3,615 nm for non-ECA cruising. Collected data from Lloyd's Database, the DWT of six major merchant ships accounts for 90% of all the listed vessels as shown in the outer circle of Figure 4.9 and the inner circle represents the vessel numbers of each ship type, see Tab. S-6 in Appendix A for details. Next, this study will focus on calculating the total fuel consumption and ship exhaust gas emissions for the cruising mode of the six-merchant ship class in the non-ECA areas, shown as the red line in the Figure 4.8, based on the model previously proposed.

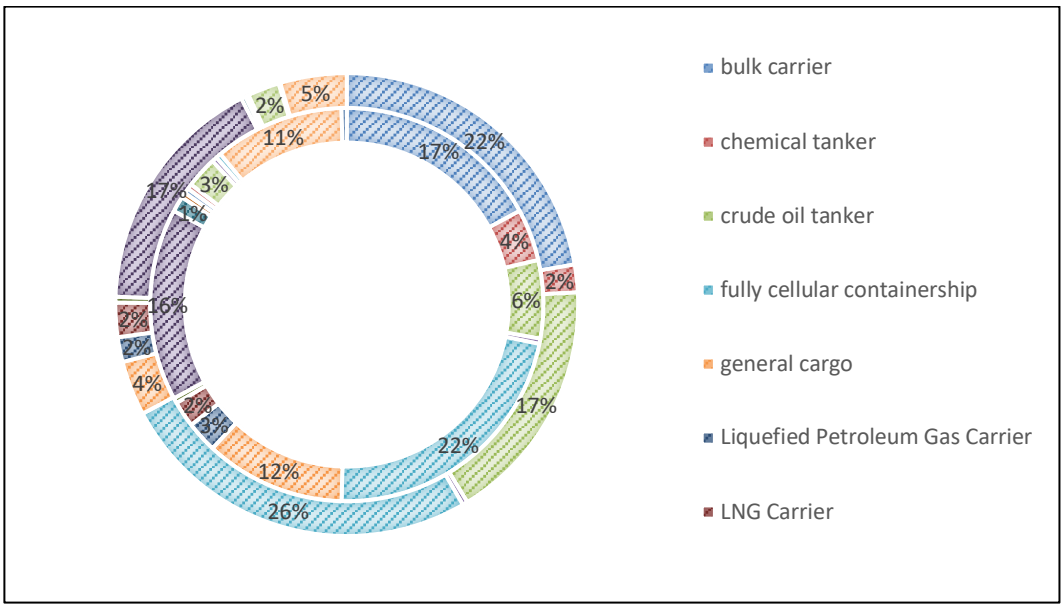


Figure 4.9 Ship number and DWT composition based on ship class from AIS data

As DWT is the unit of capacity bin for bulk carrier and general cargo ship, the relationship between DWT and the capacity bin units of the container ships and the three tankers should be found to proceed with the fuel consumption. This study adopted the regression model proposed by to calculate TEU from DWT (Abramowski et al., 2018); In the meanwhile, a fit regression model was built by the liquid tank capacity and DWT for LNG, LPG and oil tankers, see Fig. S-3 in Appendix A.

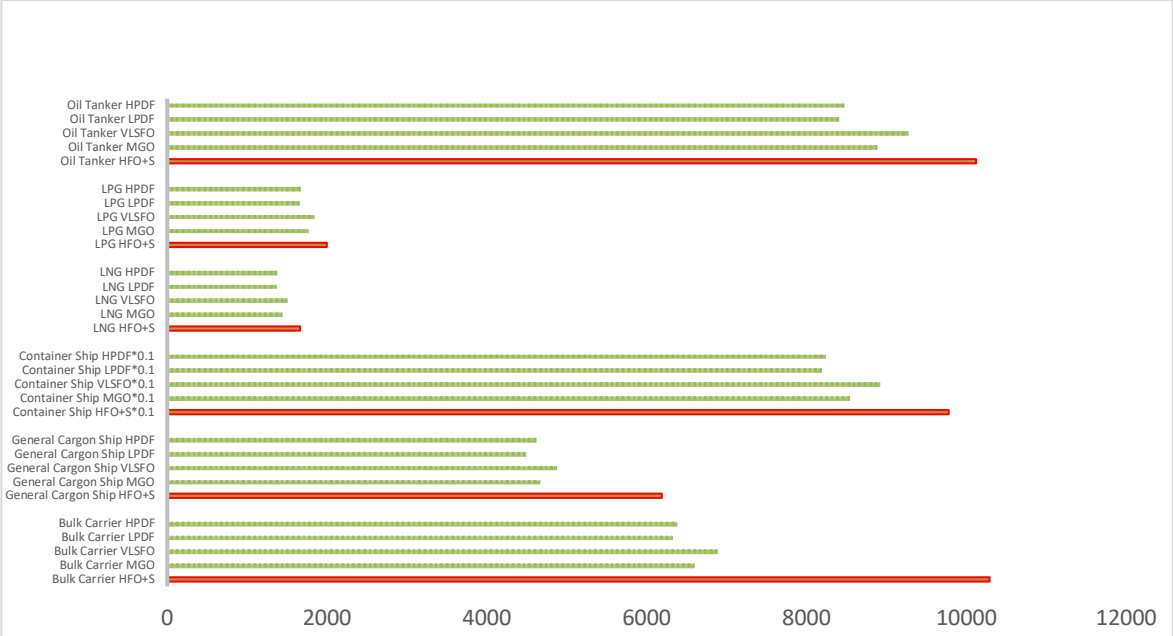


Figure 4.10 Cumulative fuel consumption for non-ECA cruising

From the perspective of fuel consumption, the least energy saving-solution for each ship class is always the HFO with scrubber, shown in Figure 4.10, particularly for bulk carriers, other options can save up to 39 % fuel by discarding this scrubber solution. The container ships take the most value of fuel consumption, ranging from 81,936 to 97,768 tons since they contribute the most transported DWT and their design speed is much higher than that of other options. On the other side, LNG carriers and LPG tankers share the least fuel consumption as there were 6 LNG tankers and 5 LPG tankers navigating from US ECA to North Sea SECA at that time, leading no evident difference between the LNG fueled engines and oil fueled engines. By considering the operating expense only, the LNG based fuel options, LPDF and HPDF, are the best strategy because of the least amount of fuel consumption and the lowest energy price of LNG.

When using the emission model to calculate the GWP-related exhaust gas emissions, it can be deduced that the HFO with scrubber is the option with the highest amounts of CO_2

equivalent emissions, so it is set as the benchmark to show the emission reduction of other four fuel options, see Figure 4.11.

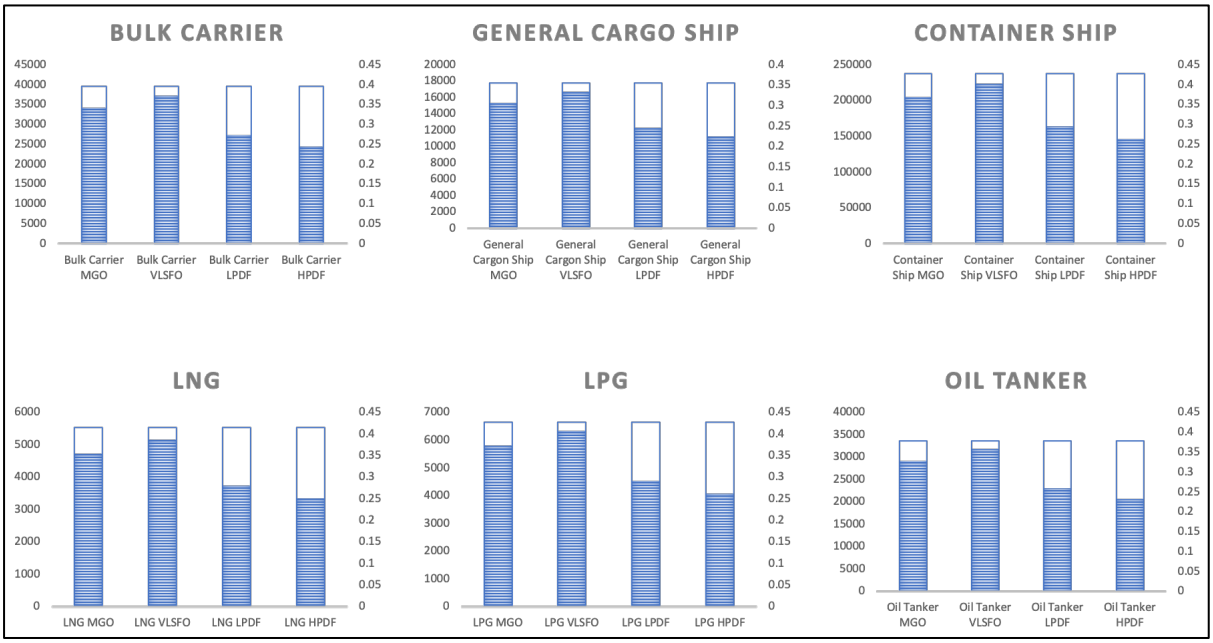


Figure 4.11 Total CO_2 equivalent emissions during the whole non-ECA cruising voyage

HPDF, as the most effective fuel option for emission abatement, can reduce the emission of CO_2 equivalent by around 36~39% for the six ship types. Taken bulk carrier as an instance, the GWP related gases emissions for HFO with scrubber is 39,502 tons while that of HPDF is 24,433 tons. Furthermore, the VLSFO option does not show much difference as the benchmark solution, the percentage of reduced CO_2 equivalent ranges from 4% to 7%. As for the most CO_2 equivalent ship class, the container ship emits around 237,278 tons of GWP related gases when HFO with scrubber is taken for cruising in the non ECA of this study, and when the MGO is adopted, 205,235 tons of CO_2 equivalent will be exhausted, while for VLSFO, LPDF and HPDF are 223,640, 163,696, 146,312 tons, respectively.

Suppose the first scenario employs HFO with scrubber for the six merchant ships, the total CO_2 equivalent emission may surge to 340,344 tons for the whole non-ECA cruising voyage;

alternatively, the number may reduce to 209,990 tons when all the merchant ships employ HPDF engines as the most eco-friendly scenario, 38% GWP related gases may be dropped compared to the first scenario, see Figure 4.12.

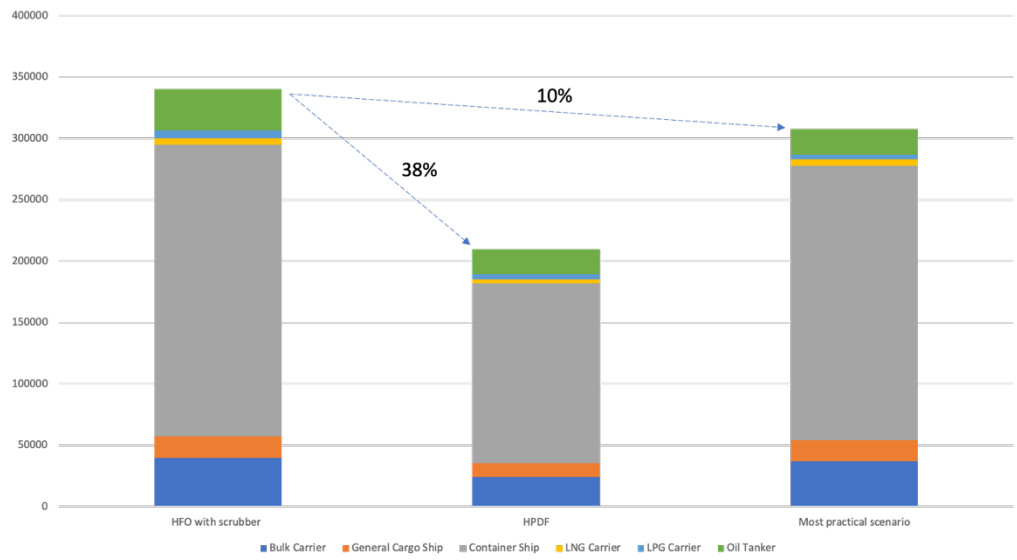


Figure 4.12 CO_2 equivalent emission comparisons among the three scenarios

It is widely recognized that the newly blended marine fuel option, VLSFO, is a compliant solution with high cost-performance ratio under IMO 2020 Sulfur Cap; more and more liquefied gas tankers, either newly building orders or existed ones, tend to install or retrofit LNG fueled engines to meet the long term marine environmental requirements of IMO. As a result, this case study defined a most practical fuel scenario which assumed VLSFO as the major option for bulker carriers, general cargo ships, container ships, and oil tankers, in the meanwhile, the LNG HPDF engines were assumed to be installed on the LNG and LPG carriers. Then the total amount of CO_2 equivalent emission for the case is 3,077,411 tons, with the 37,232 tons contributed by bulk carriers, 16718 tons by general cargo ships, 223,640 tons by container ships, 5,158 tons by LNG carriers, 4,067 tons by LPG carriers and 20,596 tons by oil tankers.

4.4. Concluding remarks

Towards enhancing the sustainable development of the shipping industry, fuel economy and safety concerns are among the most important factors to create future roadmaps for the operation and the growth of the global ship fleets. Focusing on optimizing calculations of shipping fuel consumption and ship exhaust gas emissions, this study provides a holistic bottom-up methodology by integrating and evolving several approaches to reduce the uncertainties. To our knowledge, this chapter is the first work to cover both the newly blended fuel and the BCA emission impacts, and the fuel consumption and exhaust gas emissions from 337 ships within six major merchant ship types have been calculated and analyzed to find the most sustainable solutions by in-group comparison and between-group comparison. In general, the calculated values from the mathematical model confirmed that the large size ships are preferred from the perspective of maritime sustainability, and the two LNG fuel options displayed great potential for fuel savings and reductions in the emissions of CO_2, N_2O, and BCA. Although CH_4 emission may be an issue for LNG fueled ships, their overall GWP gas emissions were better than other options. As for the sustainable preferred ship fleet, bulk carriers outperformed general cargo ships on both nautical mile-based fuel consumption and CO_2 equivalent emissions per unit of DWT, while for the unit of liquid tank capacity, there is little difference in GHG emissions among LNG tanker, LPG tanker and oil tanker when the liquid tank size surpasses 100 thousand cubic meters. The recommended tanker size is LNG tanker with LTC larger than 150,000 cubic meters, LPG tanker greater than 23,000, and oil tanker whose LTC in the range between 100,000 and 130,000, and greater than 180,000 m^3. Among all the proven fuel options under IMO 2020 Sulfur Cap, the HFO with scrubber is the least favorable for both GWP and non-GWP gases, while the LNG fueled engines, LPDF and HPDF, are recommended to meet the green shipping requirement. HPDF took a slight lead for GWP related gas emission control whereas LPDF offers superior

emission control for CO, PM, NO_x, and SO_x. The case study confirmed 36~39% CO_2 equivalent emissions were reduced by replacing HFO with scrubber option with HPDF. As a second dominated GWP gas following CO_2, BCA emission follows the rank: VLSFO > HFO with scrubber > MGO > HPDF > LPDF, from high to low. Because of BCA's high emission rate, VLSFO fails to show great improvement on GWP gas emission control compared to HFO with scrubber. Meanwhile, the two LNG based fuel options show the least global warming potential benefiting from lower BCA and CO_2 emissions.

4.5. References

Abramowski, T., Cepowski, T., Zvolenský, P., 2018. Determination of Regression Formulas for Key Design Characteristics of Container Ships at Preliminary Design Stage. New Trends Prod. Eng. 1, 247–257. https://doi.org/10.2478/ntpe-2018-0031

ANL, 2016. GREET Life-Cycle Model User Guide 121. https://doi.org/10.4337/9781782545583.00006

Bond, T.C., Doherty, S.J., Fahey, D.W., Forster, P.M., Berntsen, T., Deangelo, B.J., Flanner, M.G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P.K., Sarofim, M.C., Schultz, M.G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S.K., Hopke, P.K., Jacobson, M.Z., Kaiser, J.W., Klimont, Z., Lohmann, U., Schwarz, J.P., Shindell, D., Storelvmo, T., Warren, S.G., Zender, C.S., 2013. Bounding the role of black carbon in the climate system: A scientific assessment. J. Geophys. Res. Atmos. 118, 5380–5552. https://doi.org/10.1002/jgrd.50171

EMEP/EEA, 2019. EMEP/EEA air pollutant emission inventory guidebook 2019 [WWW Document]. URL http://efdb.apps.eea.europa.eu/?source=%7B%22query%22%3A%7B%22bool%22%3A%7B%22must%22%3A%5B%7B%22term%22%3A%7B%22Fuel%22%3A%22Marine

diesel oil%2Fmarine gas oil (MDO%2FMGO)%22%7D%7D%5D%7D%7D%2C%22display_type%22%3A%22tabular%22%7D

Finland, Germany, 2019. REDUCTION OF THE IMPACT ON THE ARCTIC OF BLACK CARBON EMISSIONS.

IPCC, 2019. EFDB emission factor database [WWW Document]. URL https://www.ipcc-nggip.iges.or.jp/EFDB/main.php

Kristensen, H.O., 2015. Energy demand and exhaust Gas emissions of marine engines. Clean Shipp. Curr. 1, 18–26.

Man B&M, 2012. ME-GI Dual Fuel MAN B&W Engines. A Technical, Operational and Cost-effective Solution for Ships Fuelled by Gas Content, MAN Diesel & Turbo.

Moreno-Gutiérrez, J., Calderay, F., Saborido, N., Boile, M., Rodríguez Valero, R., Durán-Grados, V., 2015. Methodologies for estimating shipping emissions and energy consumption: A comparative analysis of current methods. Energy 86, 603–616. https://doi.org/10.1016/j.energy.2015.04.083

NPI, 2017. National Pollutant Inventory Emission Data [WWW Document]. URL https://data.gov.au/dataset/ds-dga-043f58e0-a188-4458-b61c-04e5b540aea4/details

Olmer, N., Comer, B., Roy, B., Mao, X., Rutherford, D., 2017. Greenhouse Gas Emissions From Global Shipping, 2013-2015. Int. Counc. Clean Transp. 1–25.

Pavlenko, N., Comer, B., Zhou, Y., Clark, N., Rutherford, D., 2020. The climate implications of using LNG as a marine fuel.

Smith, T.W.P., Jalkanen, J.P., Anderson, B.A., Corbett, J.J., Faber, J., Hanayama, S., O'Keeffe, E., Parker, S., Johansson, L., Aldous, L., Raucci, C., Traut, M., Ettinger, S., Nelissen, D., Lee, D.S., Ng, S., Agrawal, A., Winebrake, J.J., Hoen, M., A., 2014. Third IMO Greenhouse Gas Study 2014. Int. Marit. Organ. 327.

https://doi.org/10.1007/s10584-013-0912-3

van der Linde, G., 2009. Efficient exhaust gas scrubbing.

5. TTP SUSTAINABILITY: POST-COMBUSTION CARBON CAPTURE FOR TANK TO PROPELLER VIA PROCESS MODELING AND SIMULATION[*]

Following Chapter 4, this chapter aims at providing a holistic way to determine the most efficient and sustainable post-combustion carbon capture solution for tank to propeller processes. The objectives of this work encompass the following: 1) marine engine cylinder modeling and validation, 2) TTPPCC system process model development and pilot plant validation, 3) Optimal absorber and stripper design under variation of solvents, packed type and liquid gas ratio, 4) Quantitative sustainable evaluation by emission reduction efficiency, energy penalty and carbon cyclic capacity.

5.1. Maritime system design

For decades, the larger-scale ship has displayed a great potential for improved fuel efficiency and emission control per unit of cargo, as well as a good return on investment. There is a clear trend of embracing the era of larger-scale ships. However, no study has focused on simulation of the carbon capture system for larger sized ships. The size of the reference ships for the maritime PCC studies are in a surprisingly small range, with a maximum displacement of 20,550 cubic meters for an LPG tanker (Awoyomi et al., 2019). Moreover, the ship engine is another perspective to consider in the sustainable maritime solution. The two-stroke slow speed diesel engine still dominates the ship engine market, while many dual-fuel engine options (mainly marine diesel fuel and LNG) are available to meet the current IMO emission standard. Based on our collected data, no study in the maritime carbon capture field has focused on process modelling for the 2-stroke dual-fuel engine type. As a milestone study in

[*] Reprinted with permission from "Ji, Chenxi, Shuai Yuan, Mitchell Huffman, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Post-combustion carbon capture for tank to propeller via process modeling and simulation." *Journal of CO2 Utilization* 51 (2021): 101655."

the maritime post-combustion carbon capture, Luo and Wang (Luo and Wang, 2017) established the process model for a two-diesel-engine power system, while most other studies pay more attention to modeling the single-diesel main engine.

5.1.1. Reference LNG tanker

Three strategies of sulfur emission control including ship exhaust gas scrubber, oil change to the low sulfur shipping fuels, and engine change to LNG-driven type are currently available for ship owners under IMO 2020 sulfur cap. When considering long-term carbon emission reduction, the LNG-fueled engine is the most attractive option. Therefore, the study targeted the LNG dual-fuel engine to establish the modeling of the tanker to propeller process. Tab. S-1 in Appendix B lists the parameters of the reference LNG tanker (Wärtsilä, 2019), which used to be the largest LNG carrier before the Q-flex and Q-max era. The ship can be driven by either marine diesel fuel or LNG with three V-type dual engines and one L-type dual engine (Wärtsilä, 2019). The fuel consumed in the LNG tanker is set as the marine diesel oil and its inherent properties are listed in Table 5.1.

Table 5.1 Inherent properties for marine diesel oil (Reprint from (Xie, 2017))

Types	Name	LHV (MJ/kg)	Density (g/cm3)	RON	Weight %
Branched alkanes	Iso-cetane/ HMN	44.38	0.321	98.9	0.1786
Aromatic HCs	1-Methylnaphthalene	40.27	0.548	120	0.1776
Naphthenic HCs	Decalin	43.02	0.569	46	0.36
N-alkanes	N-hexadecane	45.23	0.268	-30	0.2838

5.1.2. Tank to propeller process model development

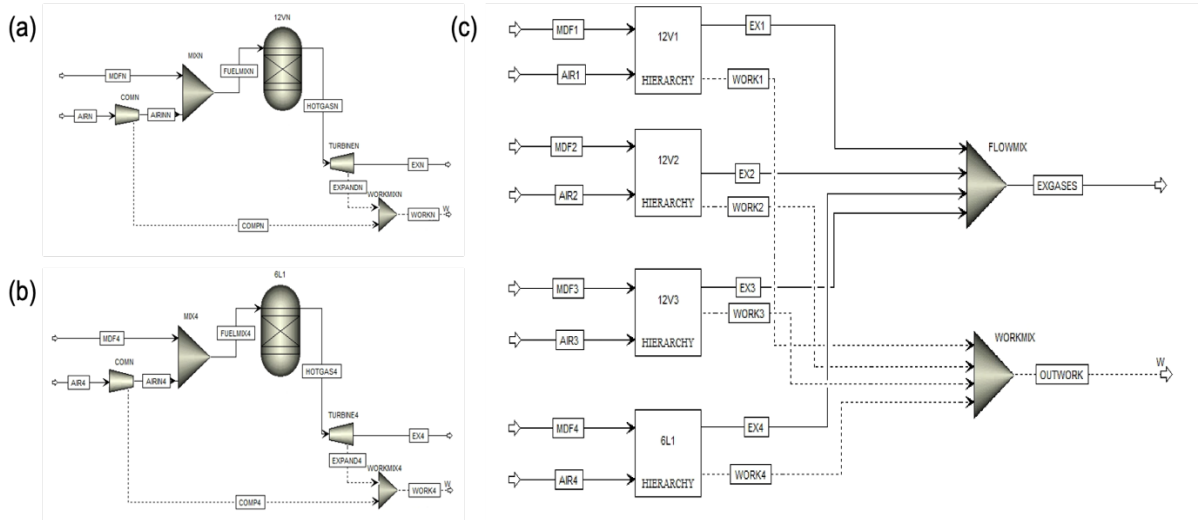


Figure 5.1 Process flow charts for two types of marine engines: (a) 12V50DF; (b) 6L50DF; (c) Simplified process modeling for the target ship engine system

The tank to propeller process of the targeted ship was modeled by Aspen Plus V11. At the property stage, the airflow (mainly nitrogen and oxygen) and marine diesel oil flow were input using the Peng-Robinson equation of state as the property method with Boston-Mathias modifications. The TTP process in the simulation stage has four functional modules: compression module, fuel combustion module, mixing module, and integration module. The block models of *Compr, RGibbs, Mixer,* and *Hierarchy* were employed for the four modules, respectively.

Referenced from the manual book of Wärtsilä 12V50DF and 6L50DF (Wärtsilä, 2019), the input value of fuel flow and airflow can be calculated accordingly, as well as the shipping

fuel temperature and mean effective pressure in the engine cylinder. Figure 5.1 (a) shows the tank to propeller flow chart of 12V50DF. The fresh air is preprocessed by filter, compressor, and cooler before it is mixed and reacted with injected marine diesel fuel in the engine cylinder, then one part of the high-temperature exhaust gas is cooled down and another part is entering the turbocharger. Using a similar theory, the model of 6L50DF is built as shown in Figure 5.1 (b). The detailed process models for the two dual engines can be found in the Fig. S-1 of Appendix B. The 4-engine ship power system is integrated by four *Hierarchy* blocks, displayed in Figure 5.1 (c). The effective work is then sent to the propeller and the total exhaust gas enters the post-combustion carbon capture system.

To validate the proposed TTP process model, the simulated results for the two engines (see Tab. S-2 from Appendix B for details) are compared with the data abstracted from the Wärtsilä 50DF product guidebook. The validation results displayed in Table 5.2 present good agreement for the values of the simulated engine output power and exhaust gas flow rate across all load levels. The maximum prediction deviation for the V-type engine is -9.92% at 50% load level, while that for the L-type is -9.85% at 50% as well. 85% workload simulation outputs will be applied to the following TTPPCC system process modeling since it is a common practice to assume 85% engine load when the ship is in the navigation mode.

Table 5.2 Marine engine cylinder process simulation output and validation

Engine Type	Load	SFOC (kg/s)	Comp. Air Flow Rate (kg/s)	Validation	Engine Output (kw)	Ex. Gas Flow Rate (kg/s)
12V50DF	**100**	0.62	21.9	Manual	11700	23.8
				Model	11623	22.52
				Prediction Dev. (%)	-0.66	-5.38
	85	0.604	18.61	Manual	9945	19.7
				Model	9877	19.22
				Prediction Dev. (%)	-0.68	-2.44
	75	0.599	16.42	Manual	8775	18.2
				Model	8714	17.02
				Prediction Dev. (%)	-0.69	-6.48
	50	0.618	11.9	Manual	5850	13.9
				Model	6315	12.52
				Prediction Dev. (%)	7.94	-9.92
6L50DF	**100**	0.308	11.0	Manual	5850	12.4
				Model	5838	11.31
				Prediction Dev. (%)	-0.21	-8.79
	85	0.302	9.35	Manual	4972	10.1

			Model	4962	9.66
			Prediction Dev. (%)	-0.20	-4.36
75	0.3	8.25	Manual	4387	9.2
			Model	4378	8.55
			Prediction Dev. (%)	-0.21	-7.06
50	0.309	6.0	Manual	2925	7.1
			Model	3184	6.4
			Prediction Dev. (%)	8.96	-9.85

5.2. TTPPCC process model development

Extensive studies have focused on the only commercialized CO_2 capture technology, post-combustion carbon capture system. For ships powered by fossil fuels, the solvent based TTPPCC methodology is the most promising process for implementing this technology. Many previous milestone works in the field of PCC have pointed out the MEA as an ideal solution for carbon capture because of its high CO_2 solubility, acceptable reaction kinetics, and fair price. In this chapter, the MEA with molality of 7 m (30% in weight percent) was employed to establish the base case process model for the TTPPCC system. Furthermore, the unique features and operation limitations of TTPPCC have been identified and shown in Table 5.3, and the corresponding process modeling considerations are listed as well.

Table 5.3 Features, limitations, and modelling considerations of TTPPCC

Features of TTP	Limitations of TTPPCC	TTPPCC modelling considerations

Limited space	Size/ height of equipment	Constraints in absorber/ stripper dimension determination
Limited utilities	Supply of heat, electric power, *etc.*	Energy penalty/ Required regeneration energy
Long lasting constant movement	Fast reaction rate, Equipment effectiveness & stability	Solvent selection preference: stable, fast reaction kinetics, not subject to degradation
Multiple operation modes	Integrated carbon capture stategies	Scenario based process models, variant of L/G ratio
Vulnerable marine environment	Toxic/high corrosive substance release	Low toxic & corrosive solvent preferred, packed materials with high packing factor preferred

Based on Table 5.3, the major considerations to improve the performance of the TTPPCC process with a maritime process intensification study are: the optimized dimension of process units, variant lean solution flows, and intensified absorption/desorption reactions with the least negative effects.

5.2.1. Physical solubility and Henry's constant

Using the same approach of molecular interaction calculations as the electrolyte NRTL, the eNRTL-Redlich-Kwong (RK) property method is applicable for handling aqueous and mixed solvent systems with a wide range of concentrations (Aspen Technology, 2006). This work adopted eNRTL-RK equation of state (EOS) to compute the liquid phase thermodynamic properties, while the PC-SAFT was utilized for vapor property calculations of the MEA-H_2O-

CO_2 system. The Henry's constant of the mixture is calculated from the binary Henry's constants of pure solvents (see equation 5.1):

$$\ln\left(\frac{H_i}{\gamma_i^{\infty}}\right) = \sum_A w_A \ln\left(\frac{H_{iA}}{\gamma_{iA}^{\infty}}\right) \tag{5.1}$$

where H_i is Henry's constant for binary components, γ_i^{∞} is the infinite dilution activity coefficient of molecular solute i in the mixed solvent, H_{iA} Henry's constant of molecular solute i in pure solvent A, γ_{iA}^{∞} is the infinite dilution activity coefficient of molecular solute i in pure solvent A. w_A, the weighting factor, is calculated by equation 5.2:

$$w_A = \frac{x_A (V_{iA}^{\infty})^{2/3}}{\sum_B x_B (V_{iB}^{\infty})^{2/3}} \tag{5.2}$$

where x_A is the mole fraction of solvent A on solute-free basis, V_{iA}^{∞} is the partial molar volume of molecular solute i at infinite dilution in pure solvent A and its calculation process is referred to as the Brelvi-O'Connell model (Brelvi and O'Connell, 1972). Henry's law constants H_{i_1,i_2} of the binary components (i_1, i_2) in this study follow the below equation:

$$H_{i_1,i_2} = \exp\left(C_1 + \frac{C_2}{T} + C_3 lnT + C_4 T\right) \tag{5.3}$$

where C_1, C_2, C_3, C_4 are the correlations for binary Henry's law constants, and T is the system temperature. Table 5.4 displays the correlation coefficients for the determination of Henry's law constants on a molality basis. This study adopted the model of Yan and Chen (Yan and Chen, 2010) to calculate the Henry's constant of CO_2 and H_2O, and CO_2-MEA was extracted from the work of Liu et al. (Liu et al., 1999). The correlation coefficients of N_2- H_2O and O_2-H_2O were both from the APV Binary database, while the APV ENRTL-RK database was applied for the H_2S-H_2O binary pair. The default values from the Aspen Databank were used for other Henry's constants of the binary components (Aspen Technology Inc, 2019).

Table 5.4 Correlation coefficients of Henry's law constants

Binary components	CO_2-H_2O	CO_2-MEA	N_2- H_2O	O_2-H_2O	H_2S-H_2O
Source	Yan and Chen (Yan and Chen, 2010)	Liu et al. (Liu et al., 1999)	APV Binary (Aspen Technology Inc, 2019)	APV Binary (Aspen Technology Inc, 2019)	APV ENRTL-RK (Aspen Technology Inc, 2019)
Unit	N/sqm	N/sqm	N/sqm	N/sqm	N/sqm
T	273-473 K	280-600 K	273-346 K	274-348 K	273-423 K
C1	100.65	89.452	176.507	155.921	358.138
C2	-6147.7	-2934.6	-8432.77	-7775.06	-13236.8
C3	-10.191	-11.59	-21.558	-18.3974	-55.0551
C4	0	0.001644	-0.008436	-0.009444	0.05957

5.2.2. Carbon capture chemical reaction mechanism

The equilibrium reactions and kinetics-controlled reactions of the MEA-CO_2-H_2O mixture have been well investigated, and Table 5.5 summarizes the two categories of the aqueous phase chemical reactions in this work.

Table 5.5 Reactions in the kinetic model for MEA-CO_2-H_2O system

Reaction No.	Reaction Type	Stoichiometry
1	Equilibrium	$MEAH^+ + H_2O \leftrightarrow H_3O^+ + MEA$
2	Equilibrium	$2H_2O \leftrightarrow H_3O^+ + OH^-$
3	Equilibrium	$HCO_3^- + H_2O \leftrightarrow H_3O^+ + CO_3^{2-}$
4	Equilibrium	$H_2S + H_2O \leftrightarrow HS^- + H_3O^+$
5	Equilibrium	$H_2O + HS^- \leftrightarrow S^{2-} + H_3O^+$

6	Kinetic	$CO_2 + OH^- \rightarrow HCO_3^-$
7	Kinetic	$HCO_3^- \rightarrow CO_2 + OH^-$
8	Kinetic	$MEA + H_2O + CO_2 \rightarrow MEACOO^- + H_3O^+$
9	Kinetic	$MEACOO^- + H_3O^+ \rightarrow MEA + H_2O + CO_2$

The equilibrium constants (K_i) for reactions 1-5 in MEA were calculated from the reference state Gibbs free energies of the participating components by equation 5.4 (Zhang et al., 2011):

$$-RTlnK_p = \Delta G_p^\circ \tag{5.4}$$

where R is the universal gas constant, T is the system temperature, K_p is the chemical equilibrium constant of reaction p, and ΔG_p° is the reference state Gibbs energy change for reaction p. The kinetics-controlled reactions (6-9 in Table 5.5) are governed by the power law expressions expressed in the below equation:

$$r_q = k_q^\circ T^n \exp\left(-\frac{E_q}{RT}\right) \prod_{i=1}^{N} (x_i \gamma_i)^{\varepsilon_{iq}} \tag{5.5}$$

where r_q is the reaction rate for reaction q, k_q° is the pre-exponential factor, T^n is the system temperature with temperature factor n, E_q is the activation energy, R is the gas constant, x_i is the molar fraction of component i, γ_i is the activity coefficient of component i, and ε_{iq} is the stoichiometric of component i in reaction q.

The rate-based mass transfer correlations can be found in the Tab. S-3 of Appendix B, and the parameters of the other components were collected from the databank of Aspen Plus (Aspen Technology Inc, 2019).

5.2.3. Base case design and validation

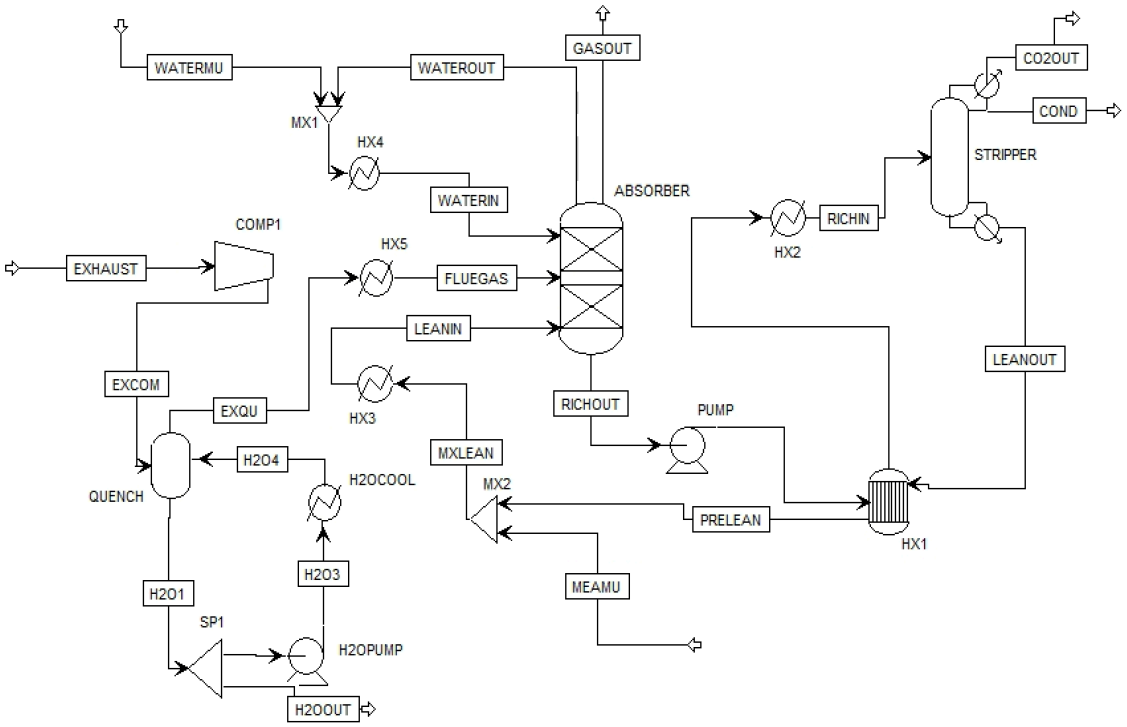


Figure 5.2 Base case design of the TTPPCC process

Figure 5.2 shows the base case process model of TTPPCC with the major columns of absorber and stripper. Two water makeup streams (one for the pre-cooling module and another for the absorption module) and one MEA replenishment before the *leanin* flow were incorporated in the PCC process, and the operation conditions of the pilot plant were extracted from experiment 1 of the Notz study to validate the proposed TTPPCC model. The key parameters of TTPPCC validation were identified as Leanin/Richout flow CO_2 loading, CO_2 emission reduction efficiency (CRE), captured CO_2 flow rate, and reboiler or specific reboiler duty. In particular, the Leanin/Richout flows are expressed per amount of CO_2 absorbed, the CRL is calculated from the CO_2 mass difference between the flue gas and the exhausted gas, and the captured CO_2 is the amount of CO_2 regenerated after the desorption operation. The energy required by the reboiler was determined by the sum of the latent heat

water condensation (Q_{cond}), the sensible heat for solvent to reach reboiler temperature (Q_{sh}) and the heat of CO_2 desorption (Q_{des}), then the penalty energy is calculated by integrating the reboiler duty and the captured CO_2 flow rate. The equations of the validation parameters are represented below:

$$\alpha_{leanin\ (richout)} = \frac{[CO_2]+[HCO_3^-]+[CO_3^{2-}]+[MEACOO^-]}{[MEA]+[MEAH^+]+[MEACOO^-]} \tag{5.6}$$

$$CRE = \frac{y_{co_2,FG}\cdot F_{FG} - y_{co_2,EG}\cdot F_{EG}}{y_{co_2,FG}\cdot F_{FG}} \tag{5.7}$$

$$Q_{reb} = Q_{cond} + Q_{sh} + Q_{des} = m_w \Delta H_w + m_s c_p \left(T_{bottom} - T_{top}\right) - m_{co_2}\Delta H_{co_2} \tag{5.8}$$

$$Q_{spe}\left(\frac{MJ}{kg}\ CO_2\right) = \frac{Q_{reb}}{F_{cap\ CO_2}} \tag{5.9}$$

Table 5.6 Pilot plant carbon capture process simulation outputs and validation

Variable	**Experiment output** (Notz et al., 2012)	**Rate-based model output**	**Absolute relative deviation, %**
Lean in CO_2 loading (mol CO_2/mol MEA)	0.265	0.273	3.0
Rich out CO_2 loading (mol CO_2/mol MEA)	0.386	0.386	0
CO_2 emission reduction efficiency (%)	76.0%	73.9%	2.7
Captured CO_2 rate (kg/h)	4.65	4.51	3.0

Reboiler duty (kw)	6.48	6.80	4.9
Specific reboiler duty (MJ/kg CO_2)	5.02	5.42	7.9

The experimental data of the pilot plant (Mellapak 250Y packing type) were extracted from case 1 of Notz et al.'s study (Notz et al., 2012). The method from Bravo's research group (Bravo et al., 1985) was applied to determine corrections for the mass transfer coefficients and effective interfacial area on the liquid and gas side. The results of the flows of *leanin, richout, gasout,* and *co_2out*, and the reboiler duty value was obtained by the proposed model. Employing equation 5.6-5.9, the validation results are calculated and summarized in Table 5.6. The largest absolute relative deviation of 7.9% is in the specific reboiler duty, and others are consistent with the experimental data as well, leading to the conclusion that the proposed model appears to be in line with experimental results. Moreover, the validation process for Sulzer BX, Mellapak plus 252Y, and Flexipac 2Y were completed with acceptable results as well, see Tab. S-4 in Appendix B for details.

5.3. TTPPCC process simulation and performance analysis

After the process synthesis, design, and analysis, the well-known MEA-based PCC operation has been set as the base case of the TTPPCC model with good validation results. As per the framework of the sustainable process synthesis-intensification (Babi et al., 2014), the innovation design should be the next focus to find a more eco-friendly solution for our defined maritime system. Referenced from the above-mentioned considerations in Table 5.3 of TTPPCC modelling, the alternative design for intensified unit operations would involve sustainable solvent selection, optimal column dimension determination, variance of L/G ratio, and minimal energy penalty for the TTPPCC system.

5.3.1. Intensified TTPPCC design

Following the 7 m MEA case, the alteration process design is processed with varying inputs of *leanin* flow rates and different structured packed types to find the optimal L/G ratio and design diameter of the absorber under acceptable flooding percentage, which is assumed to be 80% in this work.

In the contribution of Agbonghae et al. (Agbonghae et al., 2014), the range of L/G (kg/kg) ratio is from 0.70 to 2.75 for gas-fired power plants and is from 2.00 to 5.50 for coal-fired power plants. This study selected the point values 1 to 4 as the variant L/G ratio to conduct the calculation of column diameter of different structured packed types, involving BX, Mellapak plus 252Y, Mellapak 250Y and Flexipac 2Y.

The packed column parameters such as liquid/gas (L/G) ratio, diameter, packed height and total height were determined using physical modeling with two–film gas–liquid absorption theory. The input values of MEA-based on-board PCC are displayed in Table 5.7.

Table 5.7 Major input values of MEA -based maritime PCC

Entering gas	Flow rate (kg/s)	67.32
	Average molecular weight (g/mol)	28.91
	Temperature (℃)	48
	Pressure (bar)	1
	Density (kg/m^3)	1.089
Entering solvent	Flow rate (kg/s)	67.32/134.64/201.96/269.28
	Temperature (℃)	40

	Density (kg/m^3)	1053.39
	Viscosity at 40 ℃ (cP)	3

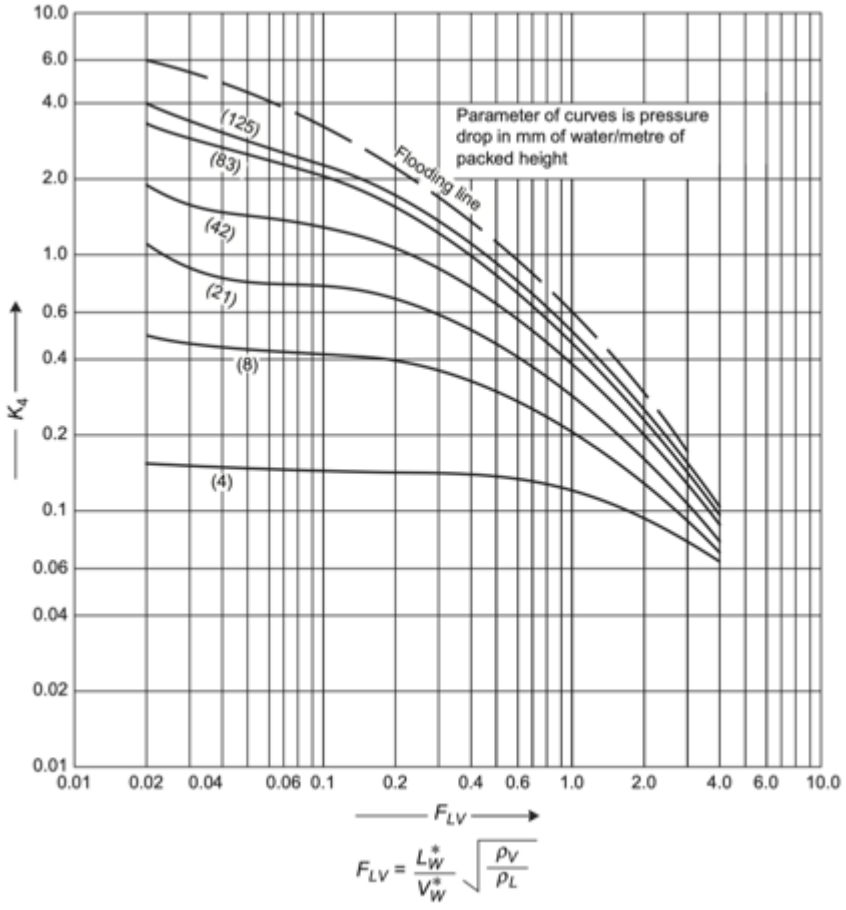


Figure 5.3 Generalized pressure drop correlation (Reprint from (Stichlmair and Fair, 1998))

For the absorber design, the recommended value of liquid pressure drop is between 15 to 50 mm H_2O per meter packed area (Towler and Sinnott, 2013). This study assumed the drop line of 20.83 mmH_2O per meter packing for the present design of absorber and stripper. Figure 5.3 shows the generalized pressure drop correlation (Stichlmair and Fair, 1998), the relationship between flooding correlation factor (K_4) and the flow parameter (F_{LV}), and their equations are listed below:

$$F_{LV} = \frac{L}{G}\sqrt{\frac{\rho_V}{\rho_L}} \tag{5.10}$$

$$K_4 = \frac{13\cdot(V_w^*)^2 F_p(\frac{\mu_L}{\rho_L})^{0.1}}{\rho_v(\rho_L-\rho_v)} \tag{5.11}$$

where V_w^* is the gas mass flow rate per unit column cross-sectional area, F_p is the packing factor, μ_L is the liquid viscosity, ρ_L and ρ_v are the liquid density and vapor density, respectively.

Besides the flooding correlation factor (K_4), its corresponding flooding factor (K_{fl}) can also be found from Figure 5.3, and the column flooding percentage (CFP) is calculated by:

$$CFP = \sqrt{\frac{K_4}{K_{fl}}} \tag{5.12}$$

After the V_w^* is calculated, the trial cross-sectional area and trial diameter can be determined. The designed diameter and cross-sectional area for varying L/G ratio and multiple packing types are shown in Table 5.8.

Table 5.8 The designed diameter and cross-sectional area for variant L/G ratio and multiple packing types

Liquid gas ratio	Flow para., F_{LV}	Flooding corr. factor, k_4	Packed type	Packing factor (m-1), F_p	Gas mass flowrate per unit column C/S area (kg/m²s), V_w^*	Trial C/S area (m^2), a	Trial diameter (m), d	Flooding factor, K_{fl}	Trial flooding percentage (%)	Design diameter (m)	Column C/S area (m^2)	Column flooding percentage (%)
1.0	0.032	0.90	Mellapak plus 252Y	39.0	2.701	24.927	5.635	5.4	0.408	6.0	19.63	0.518
			Flexipac 2Y	49.0	2.409	27.941	5.966			6.0	23.76	0.480
			Mellapak 250 Y	66.0	2.076	32.427	6.427			6.5	28.27	0.468
			Sulzer BX	90.0	1.778	37.867	6.945			7.0	33.18	0.466
2.0	0.064	0.80	Mellapak plus 252Y	39.0	2.546	26.439	5.803	4.0	0.447	6.0	19.63	0.602
			Flexipac 2Y	49.0	2.272	29.635	6.144			6.5	23.76	0.558
			Mellapak 250 Y	66.0	1.957	34.394	6.619			7.0	28.27	0.544
			Sulzer BX	90.0	1.676	40.164	7.153			7.5	33.18	0.541

			Mellapak plus 252Y	39.0	2.514	26.776	5.840			6.0	23.76	0.540
3.0	0.096	0.78	Flexipac 2Y	49.0	2.243	30.013	6.183	3.4	0.479	6.5	23.76	0.605
			Mellapak 250 Y	66.0	1.933	34.832	6.661			7.0	28.27	0.590
			Sulzer BX	90.0	1.655	40.675	7.198			7.5	33.18	0.587
			Mellapak plus 252Y	39.0	2.382	28.265	6.000			6.5	23.76	0.565
4.0	0.129	0.70	Flexipac 2Y	49.0	2.125	31.682	6.353	3.1	0.475	6.5	28.27	0.533
			Mellapak 250 Y	66.0	1.831	36.769	6.844			7.0	28.27	0.618
			Sulzer BX	90.0	1.568	42.937	7.396			7.5	33.18	0.615

5.3.2. Analysis of solvent selection

Due to the unique features of maritime PCC (see Table 5.3), the criteria of solvent selection are as follows: fast kinetics, minimal energy penalty, less prone to degradation, less corrosive, and neglectable toxicity. In addition, to reduce the sensible heat loss and the dimension of maritime PCC system, the solvent should have high cyclic capacity, which refers to the difference between the CO_2 concentration in the rich and lean solution. The molar cyclic CO_2 absorption capacity (the cyclic capacity in this work) of aqueous single and blended amines is defined as:

$$\Delta\alpha = \alpha_{richout} - \alpha_{leanin} \tag{5.13}$$

where α_{leanin} represents the CO_2 loading of the initial amine solution, and $\alpha_{richout}$ is the CO_2 loading of the amine solution after absorption.

One single aqueous amine and one blended amine were selected to find an optimal solution of the proposed TTPPCC system. Diisopropanolamine (DIPA) was chosen as the single amine for the TTPPCC system since the aqueous DIPA solution requires less energy to regenerate, removes CO_2 without degradation of the solution, and also is less corrosive and has greater selectivity of H_2S toward CO_2 (Haghtalab et al., 2014). Moreover, the methyldiethanolamine (MDEA) is another alkanol-amine which has less regeneration cost, lower volatility, greater thermal stability, and higher CO_2 cyclic capacity. However, both DIPA and MDEA have one common weakness for maritime carbon capture, a lower second order CO_2 absorption rate constant (K_2) than MEA (Liang et al., 2015), leading to their inferior absorption reaction rate. Many studies have

pointed out the blended amine with the promotor PZ is a superior solution to resist to degradation with higher kinetic rates. Therefore, this study proposed an intensified maritime carbon capture process using a MDEA-PZ-CO_2-H_2O system.

This study completed 40 process simulations in total: 16 groups for MEA, 16 groups for DIPA, and 8 groups for MDEA-PZ (since the flooding percentage of absorber and stripper exceeded 80% when L/G ratio was set as 3 and 4). The input values of selected variables for the TTPPCC system, including the composition of the selected solvent in *leanin* flow, the calculated *richout* loading range, and the diameter, stages and height of absorber and stripper, are listed in Table 5.9. Unlike the PCC system of onshore facilities, the limited space (particularly the height of columns) on the ship should be a solid constraint for TTPPCC system design rather than setting the 90% carbon capture amount as the abatement goal.

Table 5.9 TTPPCC process simulation input variables

Solvent	**MEA**	**DIPA**	**MDEA-PZ**
Composition (on molality scale)	7 m	6 m	5 m + 5 m
Leanin loading (mole CO_2/mole amine)	0.199	0.029	0.127
Richout loading range (mole CO_2/mole amine)	[0.309, 0.489]	[0.195, 0.318]	[0.406, 0.489]
Diameter of absorber (stripper)	BX: 7.5(4.5)	BX: 7.5(4.5)	BX: 7.5(6.5)
	M 250: 7(4.5)	M 250: 7(4.5)	M 250: 7(6.5)
	Mp 252: 6.5(4.5)	Mp 252: 6.5(4.5)	Mp 252: 6.5(6.5)
	2Y: 6.5(4.5)	2Y: 6.5(4.5)	2Y: 6.5(6.5)

Stages of absorber (stripper)	20 (8)	20 (8)	20 (8)
Height of absorber (stripper)	10 (6)	10 (6)	10 (6)

5.3.3. TTPPCC system performance analysis

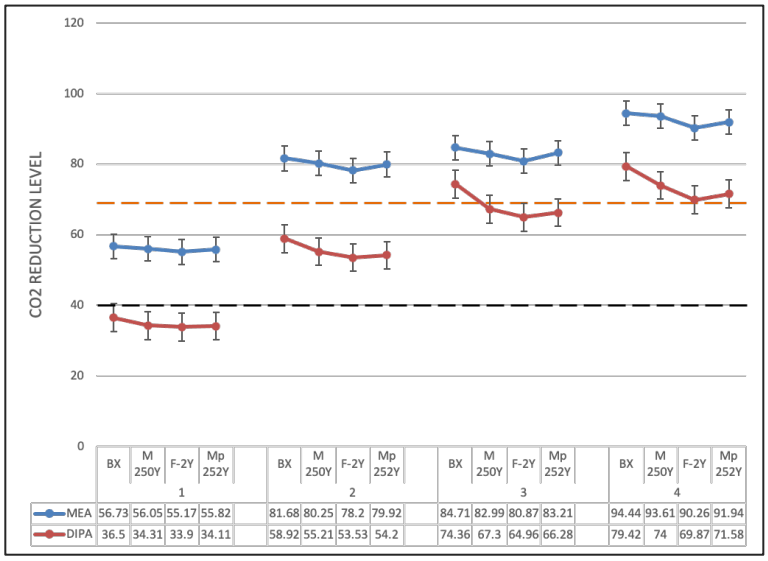


Figure 5.4 CO_2 reduction efficiency of single amines: MEA and DIPA

The black dashed line and orange dashed line in Figure 5.4 represent IMO 2030 and 2050 carbon reduction goals, respectively. In general, the MEA outperforms DIPA in every defined scenario. There is a clear trend that the carbon reduction level is increasing when the L/G ratio goes up for the two single-amine solvents. Furthermore, the sustainable-preferred packing type is as follows: Sulzer BX > Mellapak 250 Y> Mellapak plus 252Y > Flexipac 2Y

A conclusion can be made that only the maximum L/G ratio of DIPA cases can satisfy the IMO 2050 carbon reduction strategy, and when the L/G ratio equals 1, DIPA solvent is not even eligible for the 2030 IMO carbon reduction plan. When L/G ratio exceeds 2, the MEA solvent can meet the IMO long-term strategy well, but not when the L/G ratio falls below 2.

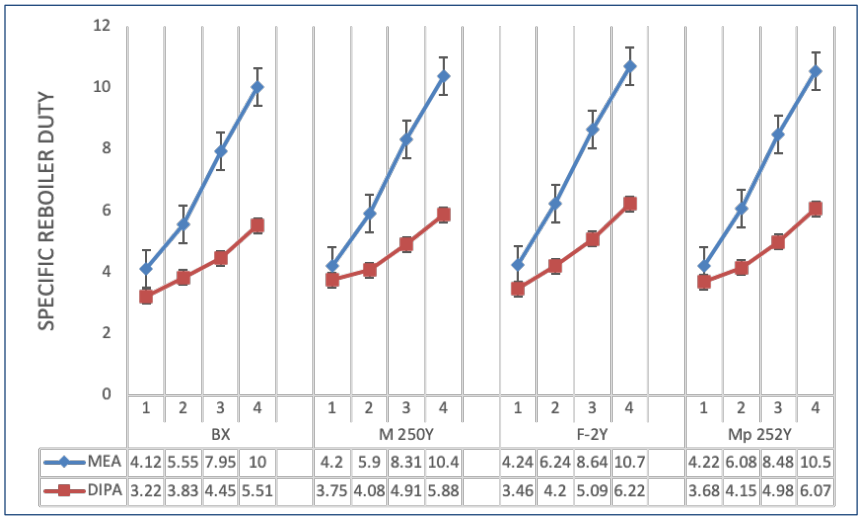


Figure 5.5 Specific reboiler duty of MEA and DIPA

In Figure 5.5, the DIPA solution saves more energy than MEA, and when the L/G ratio goes up, more energy will be saved by adopting DIPA over MEA. From the perspective of energy saving, the preferred packing type would correspond to the highest carbon capture efficiency in most cases. The best energy-saving performance is 3.22 MJ/Kg CO_2 which is obtained by DIPA solvent with BX under L/G=1, while the worst case is the MEA solvent with MEA with Flexipac 2Y under L/G=4, which yields a value of 10.7 MJ/Kg CO_2.

To summarize, there will be a tradeoff between carbon capture performance and specific reboiler duty for the two single amines. Although DIPA is applicable for short-term IMO strategy with less energy consumption when the L/G ratio exceeds 2, MEA is recommended for most cases since its carbon capture performance is much better than that of DIPA.

Next, the TTPPCC process was conducted by introducing the blended MDEA and PZ solvent. The carbon capture performance and the specific reboiler duty of the three solvents are shown in Figure 5.6.

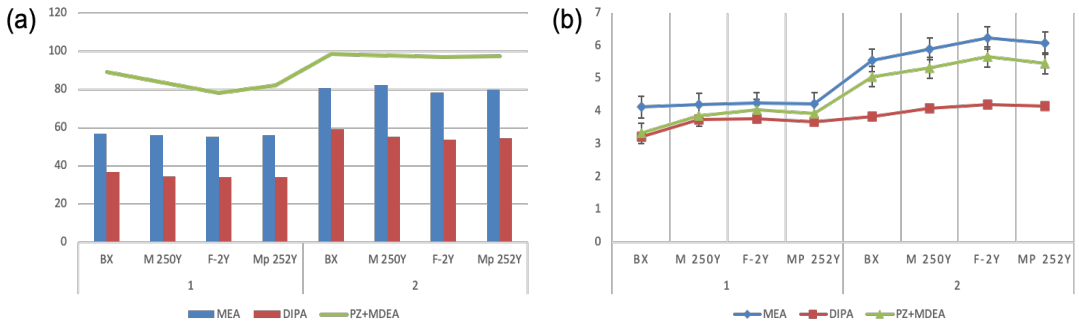


Figure 5.6 (a) Carbon capture efficiency of MEA, DIPA and MDEA-PZ; (b) Specific reboiler duty of MEA, DIPA and MDEA-PZ

Overall, the CO_2 capture efficiency of MDEA-PZ tops the other options, and its regeneration energy is between MEA and DIPA (see Figure 5.6 (a)). When the liquid flow equals to gas flow (L/G =1 cases), the MDEA-PZ is still capable of meeting the 2050 IMO carbon strategy with the only exception being the Flexipac 2Y case. As for L/G=2, all the blend amine cases generated great carbon capture results, as more than 95% CO_2 from the ship engine flue gas was reduced by installing the TTPPCC system.

DIPA is the best option for specific reboiler duty, and the blend-amine option presented performance similar to DIPA for the L/G =1 cases; when the *leanin* flow rate is twice of that of the flue gas, the MDEA-PZ option can save more than 10% energy for every mole captured CO_2 compared to the benchmark solvent MEA.

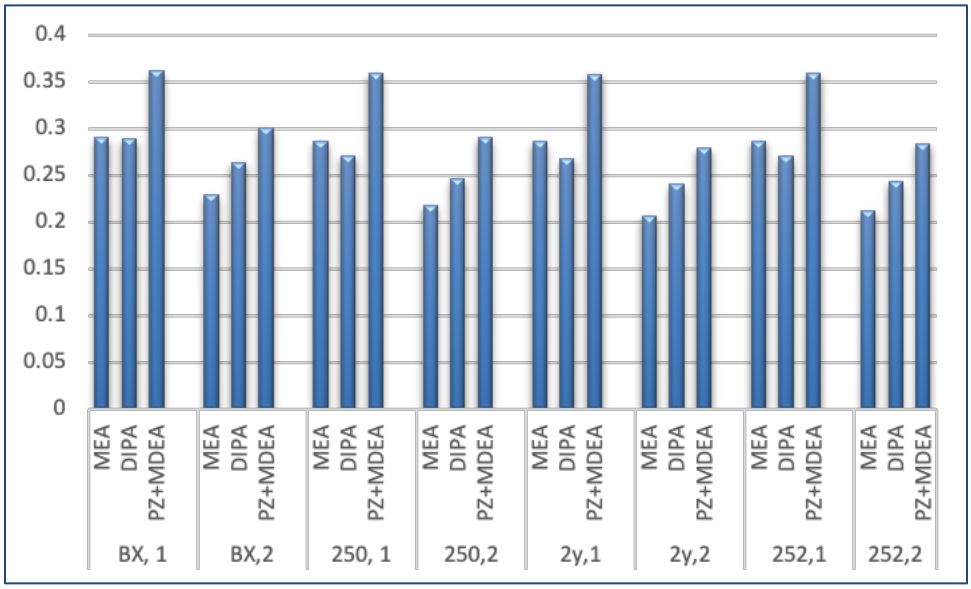


Figure 5.7 Cyclic capacity of MEA, DIPA and MDEA-PZ

Lastly, Figure 5.7 illustrates the cyclic capacity of the three solvents, and the detailed calculated results are listed in Tab.S-5 of Appendix B. The cyclic capacity has an inverse relationship with the lean solution flow rate, which can be reasonably predicted by the smaller holding time in the absorber led by the larger liquid flow rate, resulting in insufficient CO_2 absorbed by the selected solvents. The MDEA-PZ option exhibits superior cyclic capacity performance based on our process model, and the Sulzer BX with a lower liquid flow rate is recommended to reach a high-level cyclic capacity.

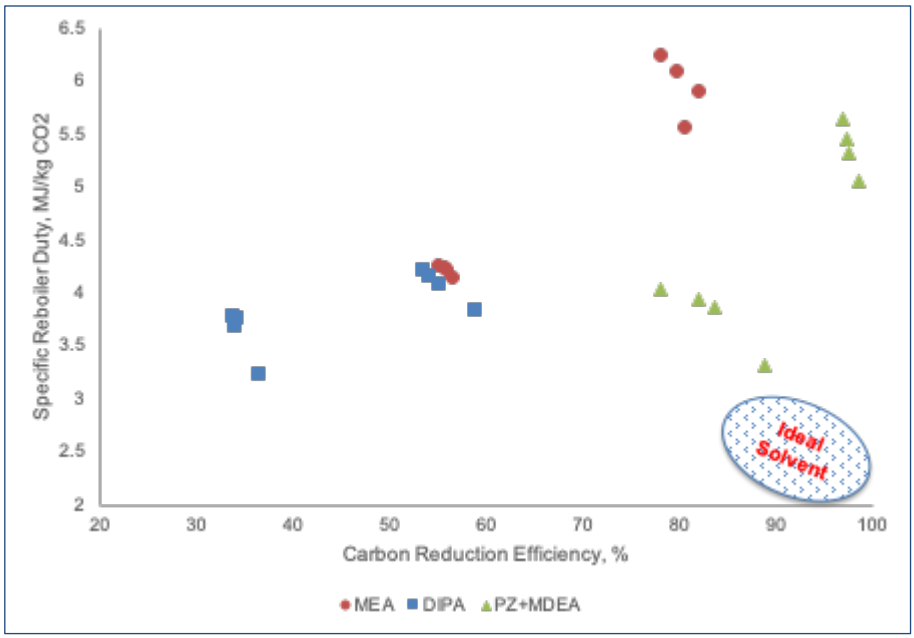


Figure 5.8 CO_2 emission control and energy penalty tradeoff analysis of MEA, DIPA and MDEA-PZ

In summary, although a large carbon capture rate can be achieved by increasing the L/G ratio, the associated energy penalty increase rate is even higher, probably due to the inverse relationship between the liquid-gas ratio and the chemical absorption cyclic capacity. The ideal solvent for the TTPPCC system can be identified as low specific reboiler duty and high carbon reduction efficiency, as shown in the right corner of Figure 5.8. Additionally, one can easily conclude that the cases of MDEA-PZ solvent with L/G equals to one have the best overall emission control-energy saving performance, and the MEA (L/G=1) group and DIPA (L/G=2) group share similar TTPPCC performance.

5.4. Concluding remarks

This study successfully established the process model of a 2-stroke ship engine system consisting of three Wärtsilä 12V50DF and one 6L50DF and presented a good validation result. Following the stages of sustainable process intensification, a rate-based aqueous MEA process model was developed and validated, then scaled up and modified to capture CO_2 from the flue gas. Also, a blended amine process model was established to simulate the onboard PCC system for the first time. The innovative design covered the absorber /stripper design with variation of solvents, packed type, and liquid gas ratio. A thorough sustainability evaluation of emission reduction efficiency, energy penalty, and carbon cyclic capacity was conducted among three aqueous amine options: MEA, DIPA, and MDEA-PZ. Compared to the benchmark aqueous amine MEA, the MDEA-PZ option could capture more than 57% CO_2 while around 25% specific reboiler duty can be saved. The generated results confirmed that the Sulzer BX was preferred among the selected four packing types from the perspective of carbon capture efficiency and energy requirement for solvent regeneration, and the cyclic capacity has an inverse relationship with the lean solution flow rate.

5.5. References

Agbonghae, E.O., Hughes, K.J., Ingham, D.B., Ma, L., Pourkashanian, M., 2014. Optimal process design of commercial-scale amine-based CO2 capture plants. Ind. Eng. Chem. Res. 53, 14815–14829. https://doi.org/10.1021/ie5023767

Aspen Technology Inc, 2019. Aspen Physical Property System, version 11.

Awoyomi, A., Patchigolla, K., Anthony, E.J., 2019. CO2/SO2 emission reduction in CO2 shipping infrastructure. Int. J. Greenh. Gas Control 88, 57–70.

https://doi.org/10.1016/j.ijggc.2019.05.011

Babi, D.K., Lutze, P., Woodley, J.M., Gani, R., 2014. A process synthesis-intensification framework for the development of sustainable membrane-based operations. Chem. Eng. Process. Process Intensif. 86, 173–195. https://doi.org/10.1016/j.cep.2014.07.001

Bravo, J.L., Rocha, J.A., Fair, J.R., 1985. MASS TRANSFER IN GAUZE PACKINGS. Hydrocarb. Process.

Brelvi, S.W., O'Connell, J.P., 1972. Correspondling states correlations for liquid compressibility and partial molal volumes of gases at infinite dilution in liquids. AIChE J. https://doi.org/10.1002/aic.690180622

Haghtalab, A., Eghbali, H., Shojaeian, A., 2014. Experiment and modeling solubility of CO2 in aqueous solutions of Diisopropanolamine + 2-amino-2-methyl-1-propanol + Piperazine at high pressures. J. Chem. Thermodyn. 71, 71–83. https://doi.org/10.1016/j.jct.2013.11.025

Liang, Z. (Henry), Rongwong, W., Liu, H., Fu, K., Gao, H., Cao, F., Zhang, R., Sema, T., Henni, A., Sumon, K., Nath, D., Gelowitz, D., Srisang, W., Saiwan, C., Benamor, A., Al-Marri, M., Shi, H., Supap, T., Chan, C., Zhou, Q., Abu-Zahra, M., Wilson, M., Olson, W., Idem, R., Tontiwachwuthikul, P. (PT), 2015. Recent progress and new developments in post-combustion carbon-capture technology with amine based solvents. Int. J. Greenh. Gas Control 40, 26–54. https://doi.org/10.1016/j.ijggc.2015.06.017

Liu, Y., Zhang, L., Watanasiri, S., 1999. Representing vapor-liquid equilibrium for an

aqueous MEA-CO2 system using the electrolyte nonrandom-two-liquid model. Ind. Eng. Chem. Res. https://doi.org/10.1021/ie980600v

Luo, X., Wang, M., 2017. Study of solvent-based carbon capture for cargo ships through process modelling and simulation. Appl. Energy 195, 402–413. https://doi.org/10.1016/j.apenergy.2017.03.027

Notz, R., Mangalapally, H.P., Hasse, H., 2012. Post combustion CO2 capture by reactive absorption: Pilot plant description and results of systematic studies with MEA. Int. J. Greenh. Gas Control 6, 84–112. https://doi.org/10.1016/j.ijggc.2011.11.004

Stichlmair, J.G., Fair, J.R., 1998. Distillation: Principles and Practices, Chemistry Biology.

Technology, A., 2006. Aspen Physical Property System Physical Property Methods and Models.

Towler, G., Sinnott, R.K., 2013. Chemical Engineering Design - Principles, Practice and Economics of Plant and Process Design (2nd Edition), Elsevier.

Wärtsilä, 2019. Wärtsilä 50DF Product Guide.

Xie, Q., 2017. Simulation of combustion characteristics in a LNG/diesel dual fuel engine. Wuhan University of Technology.

Yan, Y., Chen, C.C., 2010. Thermodynamic modeling of CO2 solubility in aqueous solutions of NaCl and Na2SO4. J. Supercrit. Fluids. https://doi.org/10.1016/j.supflu.2010.09.039

Zhang, Y., Que, H., Chen, C.C., 2011. Thermodynamic modeling for CO2 absorption in aqueous MEA solution with electrolyte NRTL model. Fluid Phase Equilib.

https://doi.org/10.1016/j.fluid.2011.08.025

6. TTP SAFETY: PREDICTING FLAMMABILITY-LEADING PROPERTIES FOR LIQUID AEROSOL SAFETY*

A limited number of studies have attached importance to the liquid aerosol formulation and flame speed, leading to insufficient data for liquid aerosol combustion and explosion parameters (*i.e.*, LVP, LDV, and liquid surface tension). Therefore, it is urgent to find an effective method of predicting these aerosolization-leading target properties. The objective of this work is to develop a mechanism to find an optimal algorithm for the three identified liquid aerosol safety parameters of organic compounds by comparing interpretation and prediction accuracy among the proposed machine learning models. One liquid aerosolization controlling factor (LVP) and two SMD dominating parameters (LDV and liquid surface tension) are set as the dependent variables to develop machine learning models individually, as shown in Figure 6.1. 14 predictors, including 3 intersectional items, are selected through forward stepwise feature selection to ensure statistical significance of proposed models and to avoid collinearity issues between any pair of predictors. Since these parameters are all temperature related, 298 K is the temperature used to pre-process the collected data and to compare the performance of the prediction models.

* Reprinted with permission from "Ji, Chenxi, Shuai Yuan, Zeren Jiao, Mitchell Huffman, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Predicting flammability-leading properties for liquid aerosol safety via machine learning." *Process Safety and Environmental Protection* 148 (2021): 1357-1366."

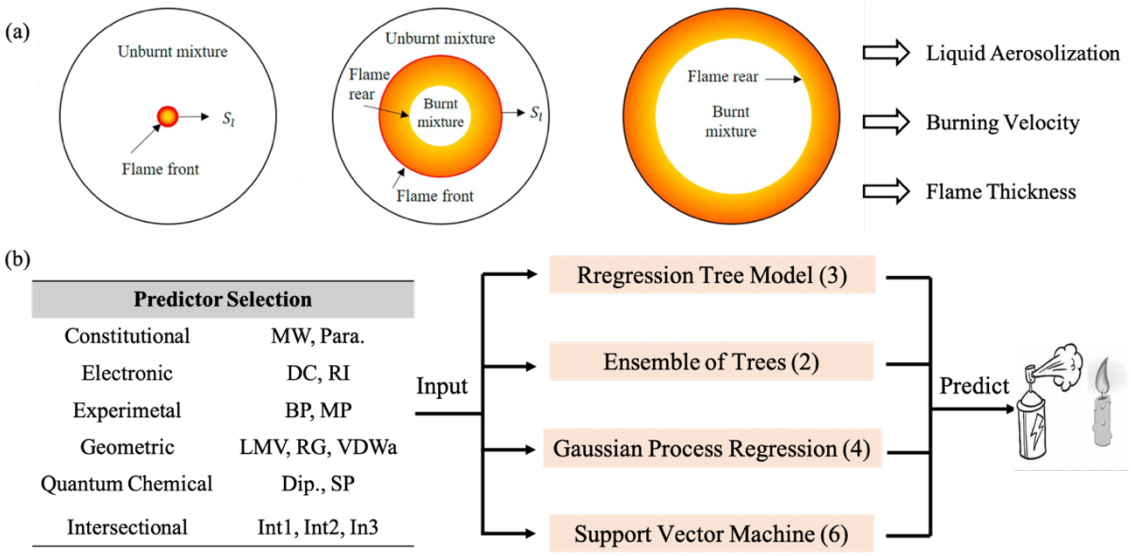


Figure 6.1 (a) Three-zone flame model of a flame thickness less than the explosion vessel radius (b) Feature selection and predictive models for liquid aerosol safety contributors

6.1. Methodology

6.1.1. Data set and response variables

The dataset of LVP, LDV and surface tension were collected from the Design Institute for Physical Properties (DIPPR) 801 database (Knovel, 2019). The liquid aerosolization dataset, consisting of 1215 data points for organic compounds, was randomly split into training data and test data by k-fold cross-validation to find a more stable prediction accuracy than the dataset-split validation approach. k-fold cross validation is applied in this research because of its more desirable trade-off between prediction and interpretation. There is no denying that the leave-one-out cross-validation (a special case of k-fold cross-validation) would show an even greater fit to the training data, however, its prediction variance will be suboptimal in comparison to k-fold validation. The liquid

aerosolization data set is split into k number of sections, and the proposed machine learning (ML) models are developed k times using one-fold cross validation as the test set in each iteration. The k-fold statistical assessment value is calculated by averaging through all developed models. As a common practice, 5 to 10 are the ideal k values for k-fold cross validation, while $k = 10$ has become the standard value in practical machine learning and data mining circles (Ceri et al., 2003).

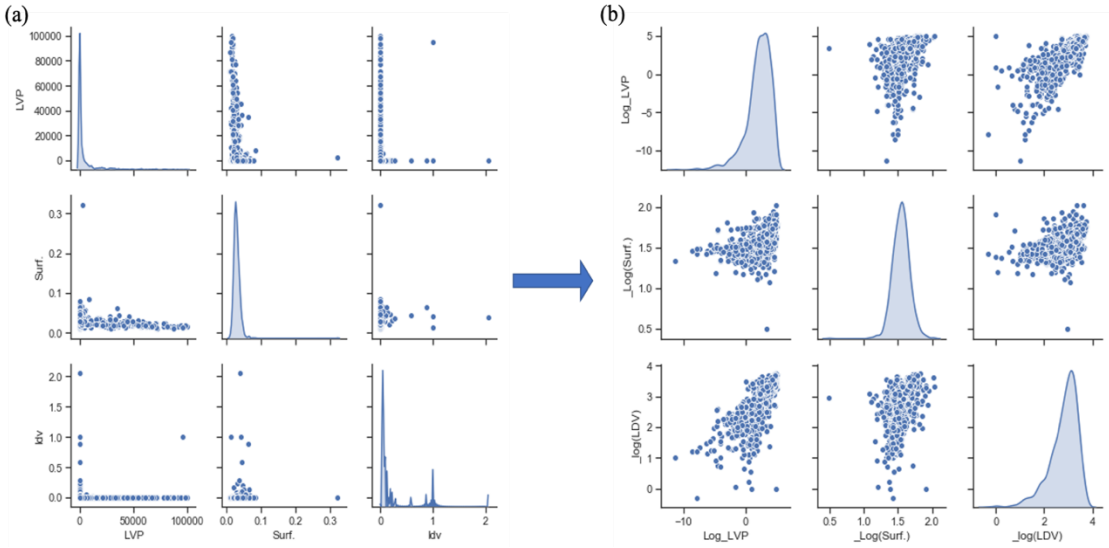


Figure 6.2 (a). Original experimental data distribution of response variables, y_1, y_2, y_3. (b). Logarithmic transformation of response variables, $\pm\log(y_n)$

As indicated by Figure 6.2 (a), the distributions of LVP (y_1,), ST (y_2) and LDV (y_3) show clearly right-skewed trends. This is especially true for LDV, which has a highly centralized distribution close to zero with extensive noise on the right side. A logarithmic transformation was performed to obtain an acceptable distribution of the dependent parameters (see Figure 6.2 (b)). Table 6.1 displays the statistics summary for the three response variables before and after transformation. The outputs of transformed

response variables in Table 6.1 confirms the improvement of the original data (see Figure S-1 and S-2 in Appendix C for detailed statistical analysis). No clear linear relationship between these three response variables exists.

Table 6.1 Summary of statistics for original/ transformed data for response variables

Original	Mean	Median	Std Dev.	Std Err Mean	Upper/ Lower 95% Mean
LVP	6669.4	284.1	16087.1	461.5	7574.9/5764.0
ST	0.030	0.028	0.012	3.4E-4	0.030/0.029
LDV	9.7E-3	1.1E-3	0.079	2.3E-3	0.014/0.005
Transformed	Mean	Median	Std Dev.	Std Err Mean	Upper/ Lower 95% Mean
Log (LVP)	2.16	2.45	2.04	0.059	2.27/2.04
Log (ST)	1.55	1.55	0.13	3.6E-3	1.55/1.54
Log (LDV)	2.81	2.94	0.59	0.017	2.85/2.78

6.1.2. Predictors and correlations

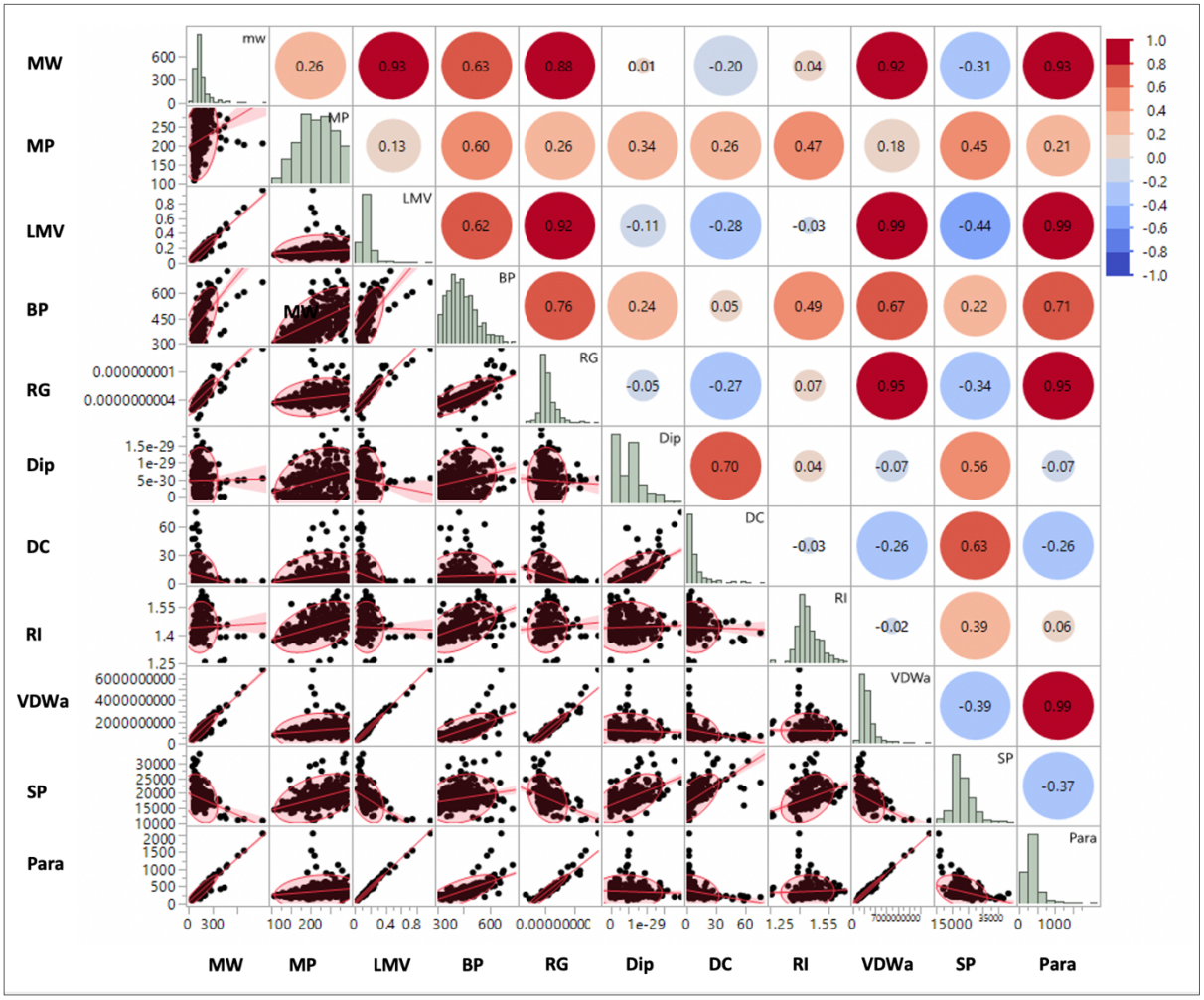


Figure 6.3 Scatterplot matrix with pairwise correlation heat map

The simplified molecular-input line-entry system (SMILES) for the organic compounds is obtained from Pubchem (PubChem, 2020). Then, the predictor screening process is conducted by the open-source Python modules, RDKit, Psi4, and Mordred. First, the SMILES is translated to a molecular structure using RDKit (RDKit, 2020). Then, two rounds of optimization are carried out: the MMFF94 force filed optimization conducted by RDKit followed by the density functional optimization by Psi4 (Nikolaienko et al.,

2016). After the calculation process using Mordred package is completed (Moriwaki et al., 2018), the molecular descriptors are selected by forward stepwise selection.

Classified by the dimensionality of structure representation, three categories of molecular descriptors, 1-dimension to 3-dimension, are employed in this study. Besides the commonly adopted quantum chemical descriptor dipole moment, geometric indices such as radius of gyration, Van Der Waals surface area, and electric information such as refractive index and dielectric constant were introduced in this study as well (see Table S-1 from Appendix C for details). Before fitting regression models, the potential flaws in assumptions regarding non-constant error variance, outliers, abnormal leverage points and collinearity between independent variables should be avoided in advance. The residual plots could be utilized to check the heteroscedasticity, outliers, and uncommon leverage points. The covariance and correlation between any two predictors $(x_1, \dots, x_n)$ are calculated using the below equations.

$$Cov(x_j, x_k) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{j_i} - \overline{x_j})(x_{k_i} - \overline{x_k}) \tag{6.1}$$

$$r_{x_j x_k} = r_{x_k x_j} = \frac{Cov(x_j, x_k)}{s_{x_j} s_{x_k}} \tag{6.2}$$

In the above equations, $Cov(x_j, x_k)$ and $r_{x_j x_k}$ are the covariance and correlation, respectively, between x_j and x_k, and s_{x_j} and s_{x_k} denote the standard deviations for x_j and x_k, respectively. Figure 6.3 shows the scatterplot matrix between eleven predictors. Collinearity issues might exist between parachor and Van Der Waals surface area, parachor and liquid molar volume, or parachor and radius of gyration. In general, there

are three ways to deal with highly collinear items: remove the highly correlated predictor directly, perform a principal component analysis (PCA) to reduce predictor numbers, add extra linearly combined pairwise independent variables. This work added three products from the identified pairs of collinear predictors to the set of independent variables. This is necessary because PCA will reduce the interpretation of the proposed models and removing predictors would cause less statistical significance for the response variables.

6.1.3. Model evaluation

To evaluate the performance of machine learning regression models, two aspects of the index are calculated to consider the model's bias and variance, those aspects being the coefficient of determination and the model fitting error (Pedregosa et al., 2011). R squared value is a common criterion to determine what percentage of the training data can be represented by the proposed model, but the weakness of this value lies in its continuously increasing value as the number of predictors grows. This chapter therefore utilizes adjusted R squared value to avoid redundant insignificant predictors in the models (see equation 6.3 below).

$$R^2_{adj.} = 1 - \left(\frac{n-1}{n-p-1}\right)\left[1 - \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\right] \tag{6.3}$$

In the above equation, n is the number of data points in the sample, p is the number of predictors, $\hat{y}_i$ is the predicted value, y_i is the observed value, and $\bar{y}$ is the average of the observed values. In addition to the adjusted R squared value, the root mean squared error

(RMSE), mean squared error (MSE) and mean absolute error (MAE) are chosen to be the performance values for statistical errors. As previously mentioned, k-fold cross validation is adopted in this work, so the modified formulae are listed below:

$$RMSE_{k-Fold} = \frac{1}{k}\left(\sum_{n=1}^{k}\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_{ki}-y_{ki})^2}{n}}\right) \tag{6.4}$$

$$MSE_{k-Fold} = \frac{1}{k}\left(\sum_{n=1}^{k}\frac{\sum_{i=1}^{n}(y_{ki}-\hat{y}_{ki})^2}{n}\right) \tag{6.5}$$

$$MAE_{k-Fold} = \frac{1}{k}\left(\sum_{n=1}^{k}\frac{\sum_{i=1}^{n}|\hat{y}_{ki}-y_{ki}|}{n}\right) \tag{6.6}$$

where y_{ki} is the observed value in the kth iteration and $\hat{y}_{ki}$ is the corresponding predicted value of k-fold cross validation (here k=10).

In general, there are three steps to obtain the final optimal prediction models for the three identified contributors. Based on the reported values for interpretation and prediction, the first comparison was taken to determine the best model for each regression algorithm, and then the comparison among groups was employed to find the optimized models without dimensionality reduction techniques. The final model is determined by comparing the optimized models with and without the optimal number of principal components.

Four categories of machine learning algorithms have been investigated in this work: regression tree, ensemble of trees, Gaussian process regression, and SVM. Regression tree is an algorithm that follows the tree from a root node down to a leaf node, which contains the value of the response. The values of the predictors are checked in each node

to decide the correct branch to follow, and the response variable value is set until the branches reach a leaf node. In this work, three types of regression tree models were presented, with a minimum leaf size of 4 (fine tree), 12 (medium tree), and 36 (coarse tree), and a maximum number of surrogates per node of 10. Two ensembles of trees, boosted trees and bagged trees, are employed to develop preliminary models. Boosted tree is an advanced method of integrating least-squares boosting with traditional regression trees, while bootstrap aggregating techniques are combined with regressions trees to construct Bagged trees. The minimum leaf size and number of learners of Bagged trees are set as 8 and 30, respectively, and the Boosted trees employ the same value of leaf size and learners' number, with a learning rate of 0.1 at the model development stage. As for Gaussian process regression, a probability distribution is utilized over a space of functions to determine the dependent variables. Four kernel functions are adopted to determine the correlation in their response to the distance between the values of independent variables: rational quadratic, Matern 5/2, exponential, and squared exponential functions. The sigma values of each kernel function are all set as 1.44. The support vectors, defined as the data points whose errors are larger than ε, are the way of the nonlinear transformation applied to the dataset before SVM is trained; The SVM kernel functions covers linear, quadratic, cubic, and Gaussian functions. In this work, the box constraint is set as 1.67 with an ε of 0.167; the kernel scale of fine, medium, and coarse Gaussian SVM models is input as 0.94, 3.7, and 15, respectively.

At the last stage, PCA is employed to seek the possibility of model improvement. A scree test (Cattell, 1966) is conducted in this study to find the optimal number of principal components. By comparing the reported performance values of the machine learning models with and without PCAs, the optimized model for liquid aerosol safety properties can be established.

6.2. Results and Discussion

6.2.1. Liquid dynamic viscosity

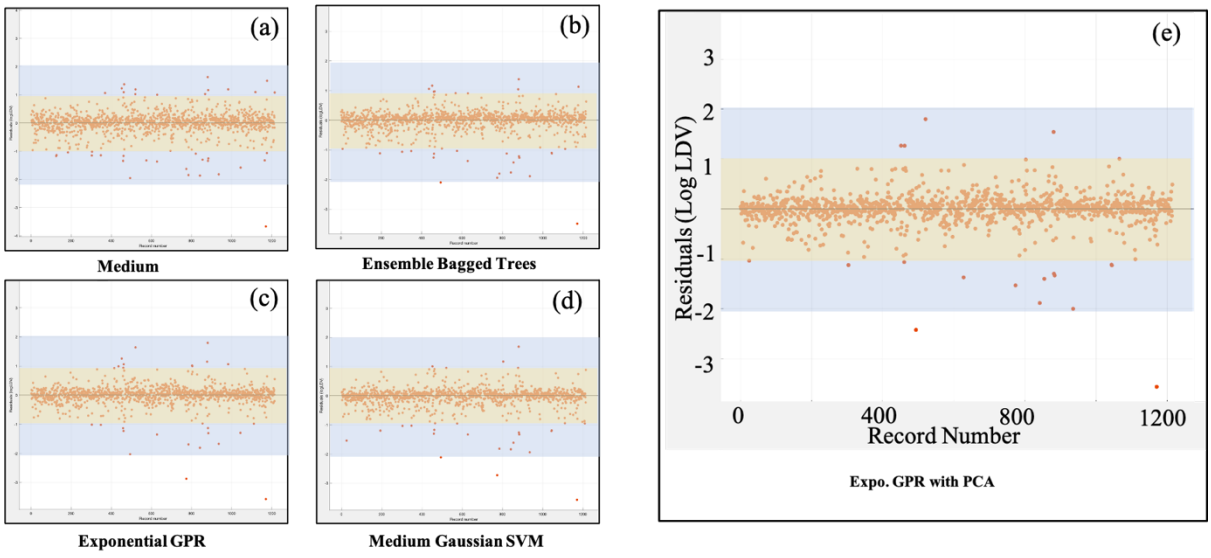


Figure 6.4 Response plots and residual plots: (a) Medium regression tree; (b) Bagged Ensemble of Trees; (c) Exponential Gaussian process regression; (d) Medium Gaussian support vector machine model; (e) Exponential GPR with 7 PCAs

To begin, a total of 15 machine learning models have been developed for the prediction of LDV. Then, four models among them (see Figure 6.4 (a)~(d)) were identified as the

most persuasive models for the regression tree, ensemble of trees, Gaussian process regression (GPR), and support vector machine (SVM) models, respectively. The best adjusted R squared performance is found in the Exponential GPR model (0.69), while the SVM and ensemble bagged trees have similar values (both 0.68). As for the values of RMSE and MAE, the medium regression tree shows results of 0.38 and 0.24, unacceptable compared to other models. Again, the Exponential GPR outperforms others with the least RMSE (0.32) and MAE (0.19). Then, the PCA technique is employed to build another model which optimizes the exponential GPR model. The seven-principal-component increases both the interpretation and the prediction accuracy, as 71% of the training data can be represented and the MSE value is reduced to 0.10. In addition, the overall shapes of figure 6.4 show that there is no evidence of heteroscedasticity of the proposed models, verifying that the models meet the requirement of constant error variance. To verify another assumption of statistical significance, the outliers are checked based on an acceptable range of (-2, 2). Only three outliers from the medium Gaussian SVM model are located outside the acceptable range, therefore it is negligible as the sample size is greater than 1000. The performance of the Gaussian SVM model and ensemble bagged tree model provided MSE values similar to the GPR model, so these two types of ML models can also be utilized to predict liquid dynamic viscosity effectively. Figure S-3 in Appendix C displays the response plots for the proposed ML models, and Figure S-4 from Appendix C shows the predicted response plots compared to the experimental response plots.

6.2.2. Liquid surface tension

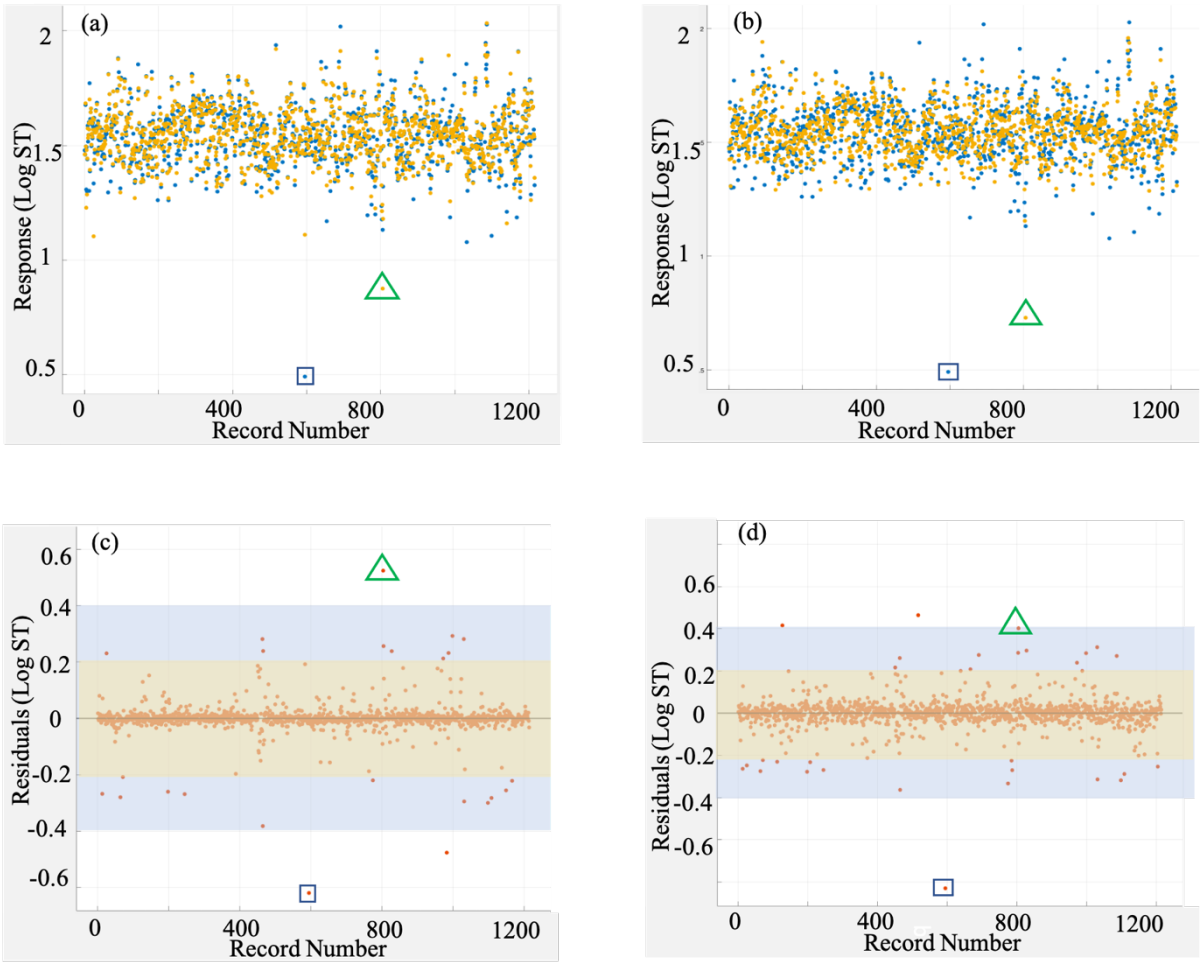


Figure 6.5 Response plots: (a) Matern 5/2 GPR model; (b) Matern 5/2 GPR with PCA; Residual plots: (c) Matern 5/2 GPR model; (d) Matern 5/2 GPR with PCA

Similar to LDV, the preliminary results first found the optimal models of each regression group using adjusted R squared value and fitting errors. For the comparison among groups, the coefficient of determination value follows the order of Matern 5/2 GPR > Coarse Gaussian SVM > Ensemble Bagged Tree > Coarse Regression Tree. One notable result is that the proposed models share satisfactory values of mean absolute error,

ranging from 0.022 to 0.049, likely due to a highly normalized distribution of the data for surface tension after logarithm transformation. Matern 5/2 GPR is selected as the optimal ML model for surface tension, then it is compared to its PCA model (see Figure 6.5). Two distinct outliers reside within the upper diagram of Figure 6.5 (a). The outlier within the rectangle is the abnormal observed data point, and the outlier in the green triangle is the uncommon data point as a predicted value. These two data points are marked on the lower chart of Figure 6.5 (a) as well. The blue area is deemed as the range of the residual plot from -0.4 to 0.4, and the yellow area is a narrow band to limit the range from -0.2 to 0.2 By introducing PCA, the outlier in the green triangle indeed moves into the acceptable region, but two other points are moving out, making the performance of the model worse based on the total number of outliers. Furthermore, there are more data points outside the narrow band of the model with PCA than the model without PCA. The reported results support our observation that the MAE of the model without PCA outperforms the model with PCA, with MAE values of 0.0022 and 0.039 for the two models with and without principal components, respectively.

6.2.3. Liquid vapor pressure

Liquid vapor pressure is an essential indicator for a liquid's evaporation rate, as well as a leading factor for liquid aerosolization. Figure 6.6 shows the relationship charts between predicted values and experimental values for the data set, and their response plots and residual plots can be found in the Appendix C (see Figure S-5). The first two models have many data points that are located outside the acceptable error range (-5%, 5%),

especially for the low-value zone. An interesting result is that when the PCA technique is coupled with GPR model, the prediction accuracy seems to drop. This can be attributed to the two emerging prediction flaws compared to the presented models in Figure 6.6 (c) and Figure 6.6 (d).

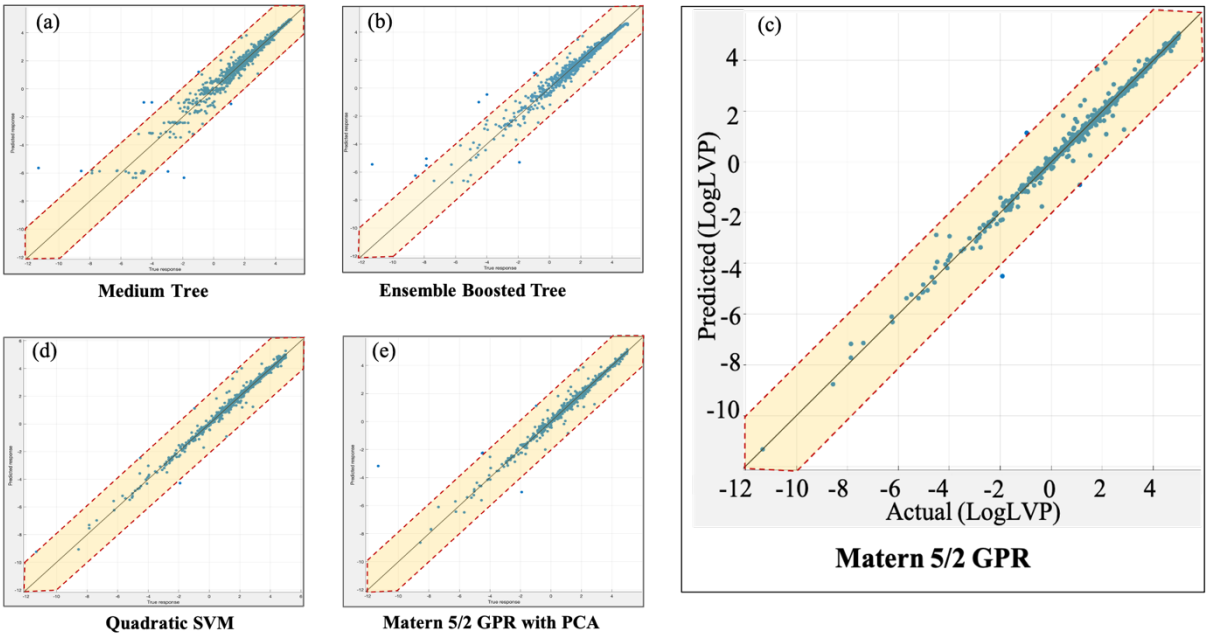


Figure 6.6 Predicted vs experimental plots: (a) Medium regression tree; (b) Ensemble boosted tree; (c) Matern 5/2 GPR; (d) Quadratic SVM; (e) Matern 5/2 GPR with PCA

Overall, the Matern GPR model and the quadratic SVM model outperformed the regression tree and ensemble boosted tree models; they showed excellent interpretability and great prediction results with corresponding MAE values of 0.11 and 0.10. A closer look at the results along the axis indicate that the prediction will be more validated when the dependent variable ($-\log(LVP)$) falls within the range of (-2, 4). A possible explanation of the prediction error, which is higher than the other two response

variables, is that the logarithmic transformation switches the original data distribution from highly right-skewed to slightly left-skewed. The optimized ML model for LVP is again obtained by the Matern 5/2 GPR approach without PCA, with a determination coefficient of 0.99 and an MSE value of 0.054.

6.2.4. Model performance and discussion

Table 6.2 Calculated Performance Parameters of optimal models for the liquid aerosol safety contributors on five ML algorithms

		RMSE	Adjusted R^2	MSE	MAE
	Liquid Dynamic Viscosity				
1	Medium Regression Tree	0.38	0.58	0.14	0.24
2	Ensemble Bagged Tree	0.33	0.68	0.11	0.20
3	Exponential GPR	0.32	0.69	0.11	0.19
4	Medium Gaussian SVM	0.33	0.68	0.11	0.19
5	Expo. GPR with PCA	0.32	0.71	0.10	0.18
	Surface Tension				
1	Coarse Regression Tree	0.078	0.61	0.0061	0.049
2	Ensemble Bagged Tree	0.073	0.66	0.0053	0.040
3	Matern 5/2 GPR	0.054	0.82	0.0029	0.022
4	Coarse Gaussian SVM	0.072	0.67	0.0052	0.040
5	Matern 5/2 GPR with PCA	0.070	0.69	0.0049	0.039
	Liquid Vapor Pressure				
1	Medium Regression Tree	0.45	0.95	0.20	0.22
2	Ensemble Boosted Tree	0.42	0.96	0.18	0.25
3	Matern 5/2 GPR	0.21	0.99	0.054	0.11

4	Quadratic SVM	0.23	0.99	0.11	0.10
5	Matern 5/2 with PCA	0.34	0.97	0.11	0.10

As seen in Table 6.2, all the proposed ML models exhibit acceptable prediction and interpretation performance for the three liquid aerosol safety contributors since their RMSE values are below 0.5 and the adjusted R squared values are above 0.6 (Witten et al., 2011). In particular, the optimal model for LDV is determined to be the exponential GPR with 7 PCAs, while the optimal models for ST and LVP are both determined to be the Matern 5/2 GPR model. Compared to other ML models, the Gaussian process regression algorithm presents superior performance for the liquid aerosolization database. GPR is a nonparametric kernel based multivariate distribution method used for a finite collection of random variables (Zhang et al., 2018). The basic idea of GPR is to employ a linear model with normalized noise to obtain the predictive distribution by coupling with the likelihood function and posterior probability function (Rasmussen and Williams, 2018). Therefore, this study utilized the predictive distribution functions shown in equations 6.7-6.10 to construct the GPR machine learning models.

$$y_i = X_i^T w + \varepsilon, \varepsilon = N(0, \sigma_n^2 I) \;\; i = 1, \dots, n. \tag{6.7}$$

$$p\left(\tilde{y} \middle| X, y, \tilde{X}\right) = N(\hat{\mu}, \hat{\Sigma}) \tag{6.8}$$

$$\hat{\mu} = K\left(\tilde{X}, X\right)^T (K(X, X) + \sigma_n^2 I)^{-1} (y - \mu(\tilde{X})) \tag{6.9}$$

$$\hat{\Sigma} = K\left(\tilde{X}, \tilde{X}\right) - K\left(\tilde{X}, X\right)^T (K(X, X) + \sigma_n^2 I)^{-1} K\left(\tilde{X}, \tilde{X}\right) \tag{6.10}$$

From the above functions, $y_i = (y_1, y_2, \dots, y_n)^T$ is the response variables, X_i^T is the independent variables, ε is the Gaussian noise, K is the $n \times n$ covariance matrix, μ is the

mean vector matrix, $\tilde{X}$ is the test locations, $\mu(\tilde{X}) = [\mu(x_{n+1}), \dots, \mu(x_{n+m})]^T$, $\mathrm{K}(X, X) = K$, and $K(\tilde{X}, \tilde{X})$ is an $m \times m$ matrix. The kernel functions for the optimized models, exponential GPR and Matern 5/2 GPR, are shown below in equations 6.11 and 6.12.

For exponential GPR,

$$K(x_j, x_k|\theta) = \sigma_f^2 \exp\left(-\frac{r}{\sigma_l}\right) \tag{6.11}$$

Matern 5/2 GPR,

$$K(x_j, x_k|\theta) = \sigma_f^2\left(1 + \frac{\sqrt{3r}}{\sigma_l}\right)\exp\left(-\frac{\sqrt{3r}}{\sigma_l}\right) \tag{6.12}$$

where $r = \sqrt{(x_j - x_k)(x_j - x_k)^T}$, σ_f is the signal standard deviation, and θ is the maximum posterior estimated value.

Besides the above-mentioned GPR approaches, this chapter applied rational quadratic and squared exponential kernel functions as well, but they did not produce the high quality of results that Matern 5/2 and Exponential options produced. Moreover, the overall performance of regression tree models is inferior to other models based on the outputs from Table 6.2, and this may be attributed to the low-robust inherent feature of regression trees, as a tiny change in the data may result in an obvious change in the final predicted tree. Undoubtedly, the prediction ML models of the three response variables have acceptable overall predictability and interpretability. Specifically, the surface tension prediction models indicate a great prediction accuracy with the lowest RMSE, MSE, and MAE values, while the LVP models have the best interpretability due to their

high adjusted R square values, and the LDV models share good values of RMSE and coefficient of determination. These results make the exponential GPR mode with optimal principal components a robust tool to predict dynamic viscosity values of organic compounds. Regarding the application of PCA, the LDV model with PCA does exhibit improved performance due to the introduction of seven principal components. However, the PCA is unnecessary for the LVP and ST models, as the dimension reduction technique brought negative effects on the prediction performance. Thus, it can be deduced that the dimensional reduction approach is not always useful to improve the GPR models. Moreover, as many sources established that the SVM has exceptional prediction performance on classification problems (Chih-Wei Hsu, Chih-Chung Chang, 2008; Huang et al., 2012; Ivanciuc, 2007), this work employed six types of SVM algorithms (linear, quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian kernel equations), and they provided close-to-optimal results, demonstrating that SVM methods are applicable for multivariate regression problems as well. Based on the outputs from the best model for each dependent variable (see Table 6.3), the most accurate prediction may come from the surface tension Matern 5/2 model since it performs best on RMSE, and therefore minimizes the variance limit to a narrow interval. Meanwhile, the LVP optimized model has the best trade-off between bias and variance, illustrating that the LVP Matern 5/2 model can be robust enough for vapor pressure prediction of chemical substances. One conclusion to be drawn from this research is that the GPR models are often the optimal models for quantitative structure−property relationship studies based on the ML outputs.

Finally, 315 organic compounds were randomly selected as the validation set to verify the prediction accuracy of our presented models for LDV, Surface Tension, and LVP. As illustrated in Table 6.3, the predicted values are close to the observed ones, resulting in good quality of prediction errors. The box plots of the three response parameters are presented in Figure S-6 in Appendix C, and clearly demonstrate that the surface tension model has very accurate prediction, with a median value of 0.05% located within a 95% confidence interval of [-2.35%, 2.19%], while the median value and confidential interval for the LVP model and LDV model are -0.44% within [-8.29%, 7.56%] and 0.45% within [-10.84%, 11.17%], respectively. Again, these values highly confirmed the prediction robustness of the presented models for the liquid aerosol safety contributors.

Table 6.3 Excerpt from the supporting information data table showing the prediction of the LDV, ST, and LVP using the optimized models

Name	Log(LDV)	Pred (Ex GPR with PCA)	Pred. Err.	Log(Surf)	Pred(Matern 5/2 GPR)	Pred. Err.	Log(LVP)	Pred(Matern 5/2 GPR)	Pred. Err.
3-methylcyclopentanone	2.92	3.01	-8.84%	1.53	1.53	-0.75%	2.87	2.82	-5.02%
3-methylundecane	2.90	2.91	-1.03%	1.60	1.60	0.29%	1.52	1.54	1.98%
3-methyltridecane	2.71	2.71	-0.49%	1.58	1.58	0.36%	0.49	0.56	7.02%
diethyl sulfide	3.37	3.36	1.15%	1.61	1.60	-0.52%	3.90	3.90	-0.53%
n-butyl formate	3.19	3.20	-1.22%	1.61	1.59	-2.31%	3.58	3.54	-4.62%
n-butyl mercaptan	3.32	3.32	-0.32%	1.60	1.59	-0.47%	3.79	3.77	-1.41%
Hhexachlorosilane	2.81	2.82	-0.66%	1.63	1.64	0.55%	2.78	2.75	-2.88%
Eethyl benzoate	2.70	2.73	-3.30%	1.46	1.45	-1.33%	1.58	1.58	0.05%

6.3. Concluding remarks

In summary, this work provides a reliable prediction for the identified liquid aerosol safety parameters based on a thorough comparison among multiple machine learning algorithms, with an emphasis on liquid aerosolization, aerosol flame speed, and flame thickness. Based on the liquid aerosolization database, four categories of regression algorithms (*i.e.*, regression tree, SVM, ensemble of trees, and GPR) with consideration of principal component analysis are applied in this study. The performance values have revealed the advantage of the Gaussian process regression approach in predicting liquid dynamic viscosity, surface tension, and liquid vapor pressure. After integrating 11 molecular predictors with 3 interaction items, the final models showed excellent prediction robustness regarding the major contributors of liquid aerosol combustion and explosion hazards. The outputs serve as an effective tool to address the lack of

quantification for liquid aerosol safety. The output properties can assist an inherently safer process design when dealing with liquid aerosols.

6.4. References

Cattell, R.B., 1966. The scree test for the number of factors. Multivariate Behav. Res. https://doi.org/10.1207/s15327906mbr0102_10

Chih-Wei Hsu, Chih-Chung Chang, and C.-J.L., 2008. A Practical Guide to Support Vector Classification. BJU Int.

Huang, G. Bin, Zhou, H., Ding, X., Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man, Cybern. Part B Cybern. https://doi.org/10.1109/TSMCB.2011.2168604

Ivanciuc, O., 2007. Applications of Support Vector Machines in Chemistry. https://doi.org/10.1002/9780470116449.ch6

Knovel, 2019. DIPPR Project 801 - Full Version - Physical Constants - Knovel [WWW Document]. URL https://app.knovel.com/web/view/itable/show.v/rcid:kpDIPPRPF7/cid:kt00CZDUQI/viewerType:itble//root_slug:physical-constants/url_slug:physical-constants?filter=table&b-toc-cid=kpDIPPRPF7&b-toc-root-slug=&b-toc-url-slug=physical-constants&b-toc-title=DIPPR%25 (accessed 3.28.20).

Moriwaki, H., Tian, Y.S., Kawashita, N., Takagi, T., 2018. Mordred: A molecular descriptor calculator. J. Cheminform. https://doi.org/10.1186/s13321-018-0258-y

Nikolaienko, T.Y., Bulavin, L.A., Hovorun, D.M., 2016. Can we treat ab initio atomic

charges and bond orders as conformation-independent electronic structure descriptors? RSC Adv. https://doi.org/10.1039/c6ra17055b

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res.

Rasmussen, C.E., Williams, C.K.I., 2018. Gaussian Processes for Machine Learning, Gaussian Processes for Machine Learning. https://doi.org/10.7551/mitpress/3206.001.0001

RDKit: Open-source cheminformatics [WWW Document], n.d.

The PubChem Project [WWW Document], n.d. URL https://pubchem.ncbi.nlm.nih.gov/#opennewwindow

Witten, I.H., Frank, E., Hall, M. a., 2011. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition, Annals of Physics. https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C

Zhang, N., Xiong, J., Zhong, J., Leatham, K., 2018. Gaussian process regression method for classification for high-dimensional data with limited samples. 8th Int. Conf. Inf. Sci. Technol. ICIST 2018 358–363. https://doi.org/10.1109/ICIST.2018.8426077

7. TTP SAFETY: DEVELOPMENT OF NOVEL COMBUSTION RISK INDEX FOR FLAMMABLE LIQUIDS*

Compared with the flash point driven liquid flammability classification method, this work integrates the major inherent properties of the liquid flammability and flame propagation with aerosol formulation. Unlike the traditional expert judgment dominated risk assessment techniques, the two ML clustering algorithms are firstly executed to evaluate the risk associated with fuel combustion in the engine cylinder. Also, the graph theory-based spectral clustering algorithm is the first time to employ in the chemical process safety field based on the authors' knowledge. A new flammability rating called liquid in-cylinder combustion risk index (LICRI) will be calculated to show the overall liquid safety performance, which can be applied as a reasonable reference when considering the marine fuel selection issue in the TTP process.

7.1. Methodology

In this chapter, the methodology will be introduced following: Identification of liquid in-cylinder combustion contributors, liquid in-cylinder combustion safety clustering approaches, safety indicator weight value determination and clustering performance evaluation, shown in the process flow chart Figure 7.1.

* Reprinted with permission from "Ji, Chenxi, Zeren Jiao, Shuai Yuan, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Development of novel combustion risk index for flammable liquids based on unsupervised clustering algorithms." *Journal of Loss Prevention in the Process Industries* 70 (2021): 104422."

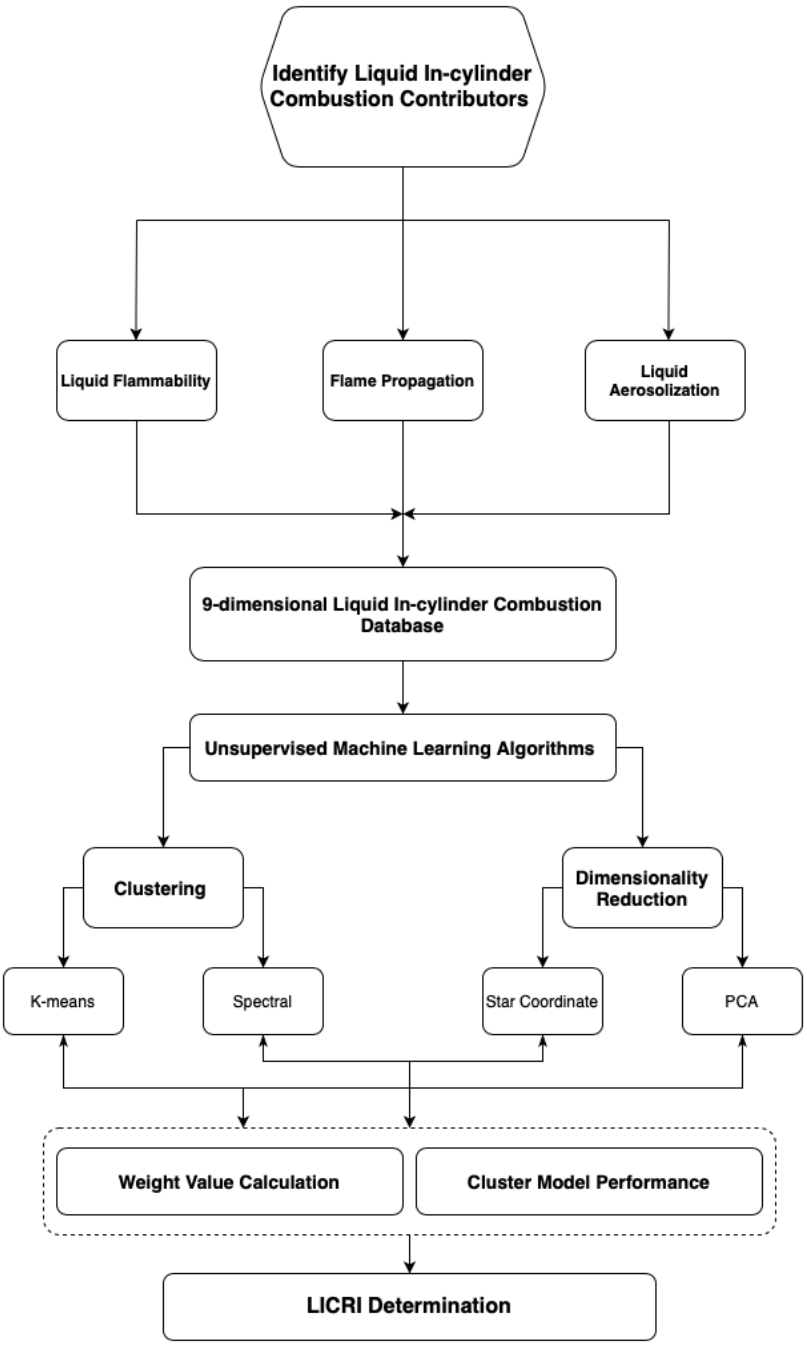


Figure 7.1 Overall methodology process flowchart

7.1.1. Identification of liquid in-cylinder combustion contributors

The field of fluid flammability characteristics has been well studied while few work has focused on the inherent property identification for liquid in-cylinder flame propagation and liquid aerosolization. Thus, it is critical to identify the leading factors for liquid fuel aerosolization and flame propagation effects so that a reasonable liquified fuel safety criterion in CI engines may be established accordingly.

This work adopts AIT, FP, and flammability range (FR), the range between LFL and UFL, as the contributors for liquid flammability matrix. Since the liquid in-cylinder flame has a combination feature of both premixed laminar and turbulent, the theoretical models of these two flames are analyzed to identify the significant parameters.

The well known "two zones" model proposed by Mallard and Le Chatelier (Mallard and Le Chatelier, 1883) is:

$$S_l = \sqrt{\alpha \dot{\omega} (\frac{T_b - T_i}{T_i - T_u})} \tag{7.1}$$

$$\alpha = \frac{k}{\rho C_p} \tag{7.2}$$

α is the thermal diffusivity, $\dot{\omega}$ is the reaction rate, k is the thermal conductivity, C_p is the specific heat capacity and ρ is the density. The relationship, shown in equation 7.3, between laminar flame and turbulent flame speed presented by Peters (Peters, 2000) has been widely accepted and shown a good performance.

$$S_T = S_l + \mu' \left\{ -\frac{a_4 b_3^2}{2b_1} Da + [\left(\frac{a_4 b_3^2}{2b_1} Da\right)^2 + a_4 b_3^2 Da]^{1/2} \right\}, \text{ where } Da = \frac{S_L l}{\mu' \delta_L} \tag{7.3}$$

As the above equation shows, μ' is the turbulence intensity; b_1, b_4 and a_4 are the turbulence modeling constants with value of 2.0, 1.0, and 0.78; Da is the Damkohler number and l is the turbulence integral length scale, δ_L that denotes the flame thickness is a function of heat capacity, heat conductivity, density, and laminar flame speed. Hence, both laminar and turbulent flame equations pointed out the dependent variables of flame propagation for CI engines are heat capacity (HC), liquid density (LD), and liquid thermal conductivity (LTC), and these three variables construct our liquid in-cylinder flame propagation matrix.

Many literature (Ballal and Lefebvre, 1979; Danis et al., 1988; Kiran Krishna et al., 2003; Polymeropoulos, 1984; Yuan et al., 2020b, 2020a, 2019; Zhang et al., 2020) have pointed out the key parameter to determine liquid aerosolization is the droplet size. Among all the theoretical mean diameters of aerosols, the Sauter Mean Diameter (SMD) is the most common one to apply for heat transfer, combustion, and dispersion modeling (K. Krishna et al., 2003). In addition, the diesel engine fuel injector can be deemed as an electrospray type of aerosol generator (Yuan et al., 2020a). Most studies conducted on pressure atomizers have focused on the type of injector used in compression ignition engines (Lefebvre and McDonell, 2017). Two SMD formulae proposed by Harmon (Harmon, 1955) and Elkotb (Elkotb, 1982) for plain-orifice type pressure atomizers are listed as below:

$$\mathrm{SMD} = 3300 d_o^{0.3} \sigma^{-0.15} \rho_L^{-0.648} \mu_L^{0.07} U_L^{-0.55} \rho_G^{-0.052} \mu_G^{0.78} \tag{7.4}$$

where d_o is the discharge orifice diameter, μ_L is the liquid kinematic viscosity, σ is the surface tension, U_L is the liquid flow rate and ρ_L is the liquid density.

$$D_{3,2} = 3.08(\sigma\rho_L)^{0.737} v_L^{0.385} \Delta P_L^{-0.54} \rho_A^{0.06} \tag{7.5}$$

From the above equation, σ is the surface tension, v_L is the liquid dynamic viscosity, ρ_L is the liquid density, ΔP_L is the liquid pressure differential and ρ_A is the air density.

Liquid dynamic viscosity (LDV) and surface tension (ST), as the inherent properties of fuels, are the determinant parameters for the droplet size of liquid aerosols based on the above two equations. For most practical fuels, any change in dynamic viscosity is always accompanied by a change in volatility, and Ballal and Lefebvre (Ballal and Lefebvre, 1979) indicated as well the quenching distance was dependent on fuel volatility. Besides dynamic viscosity, liquid vapor pressure (LVP) is an evidential index to tell the volatility level of liquids. Therefore, the identified contributors for liquid aerosolization are surface tension, liquid dynamic viscosity, and liquid vapor pressure.

7.1.2. Liquid in-cylinder combustion safety clustering approaches

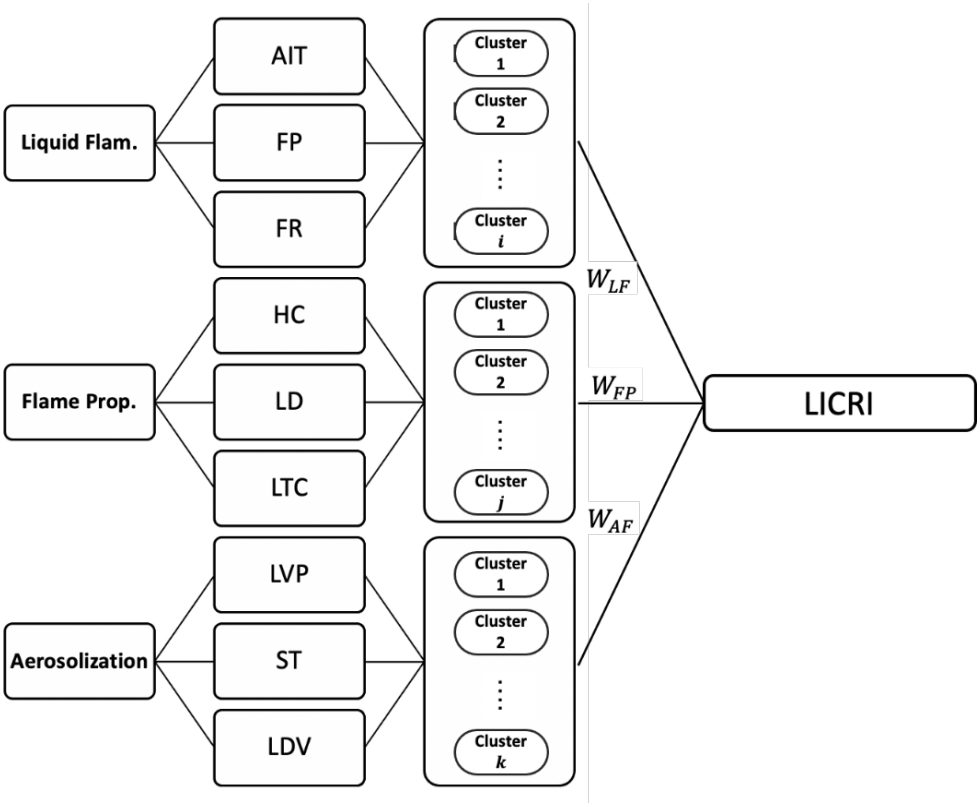


Figure 7.2 Network of liquid in-cylinder combustion risk index

By integrating the liquid aerosol formulation contributors with liquid flammability and flame propagation, the database of our liquid in-cylinder combustion criterion is built with three evaluation matrix and nine contributors, shown in Figure 7.2. This work employs two unsupervised clustering algorithms, the network of which is shown in the above figure to categorize the risk rating of liquid flammability, flame propagation, and aerosol formulation, then a liquid in-cylinder combustion risk index (LICRI) is presented to obtain the overall liquid combustion safety ratings.

$$\text{LICRI} = \sum_{i=1,j\in\{1,2,\ldots,n\}}^{3} W_i^{C_{i,j}} \tag{7.6}$$

As shown in the above equation, the weight value W_i for the three safety matrices should be determined and normalized before implementing the ML clustering algorithms. The range of LICRI is between 0 to 1, as the values of cluster numbers $C_{i,j}$ are increasing, the less value of LICRI value for one substance would be, illustrating its high risk for liquid in-cylinder combustion. In this study, the DIPPR 801 database (Knovel, 2019) are preprocessed to collect 703 effective organic compounds in the liquid state under specific temperatures with values on the nine-dimensional data.

7.1.2.1. Information entropy approach

Statistically, information entropy can be referred as the expectation of the amount of information contained in an event. Thus, the entropy value is a useful tool to give the degree of dispersion. The smaller the entropy value, the greater the degree of dispersion of the indicator, and the greater the influence, *i.e.*, weight vector, of the indicator on the comprehensive evaluation. Compared with other weight value determination approaches, the information entropy method, as an objective approach, has an outstanding capability to distinguish indicators, and it always brings high credibility and precision to avoid subject weight determination (Li et al., 2011). The typical information entropy procedures are summarized as:

- Step 1 Determination of the evaluation matrix:

Herein, three parameters within a total of nine indicators construct the evaluation matrix.

- Step 2 Normalization of the evaluation matrix:

This study employs the critical value approach to normalize the evaluation matrix, and the normalized indicators are calculated by:

$$x'_{ij} = \frac{x_{ij} - \min x_j}{\max x_j - \min x_j} \ , \ x'_{ij} = \frac{\max x_j - x_{ij}}{\max x_j - \min x_j} \tag{7.7}$$

- Step 3 Calculation of information entropy:

$$e_j = \frac{-\sum_{i=1}^{m} \frac{x'_{ij}}{\sum_{i=1}^{m} x'_{ij}} \ln\left(\frac{x'_{ij}}{\sum_{i=1}^{m} x'_{ij}}\right)}{\ln m} \tag{7.8}$$

- Step 4 Calculation of weight vectors:

$$W_j = \frac{1 - e_j}{\sum_j (1 - e_j)} \tag{7.9}$$

Therefore, the LICRI can be updated after the weight vectors are determined, please check supporting information (Table S-1 in Appendix D) for calculation details.

$$\text{LICRI} = 0.480^{Cluster_i} + 0.323^{Cluster_j} + 0.197^{Cluster_k} \tag{7.10}$$

7.1.2.2. *k*-means clustering

For clustering, the application of unsupervised clustering methods is still limited because of their unlabeled nature as well as the difficulty in boundary determination. Among those clustering methods, *k*-means clustering is the go-to algorithm that can benchmark the clustering algorithms performance (Jiao et al., 2020c; Zeren Jiao et al., 2019b). *k*-means clustering is based on the distance between objects and centroids with the actual observations as the input. The identified cluster shape is assumed as spheroidal with equal diagonal covariance. As a traditional clustering algorithm, the core idea of the *k*-means clustering approach is to minimize the total with-in cluster variation, as shown in equation

7.11. Firstly, the initial cluster assignment is proceeded by randomly assigning a number from 1 to K; then the cluster centroid for each datum of the K clusters will be calculated and each data point will be distributed to the closest Euclidean distance (Likas et al., 2003). The centroids calculation for each cluster will be repeated until the clustering assignments complete, and the objective function of *k*-means algorithm is shown below:

$$\min C_1, \dots, C_k \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x'_{ij})^2 \right\} \quad (7.11)$$

Where $C_1, \dots, C_k$ denote cluster 1 to k, $|C_k|$ is the number of samples in the kth cluster, p is the number of predictors, and $\sum_{j=1}^{p}(x_{ij} - x'_{ij})^2$ represents the Euclidean distance between two observations in the kth cluster. The study applies Python package Scikit-learn (Pedregosa et al., 2011) to process the *k*-means cluster algorithm, and the silhouette analysis (scikit-learn, 2017) is adopted to determine the number of clusters, the corresponding codes are listed in the supporting document.

7.1.2.3. Spectral clustering

In contrast to the traditional *k*-means algorithm, spectral clustering is more adaptable to data distribution with an excellent clustering effect and less computational cost. Spectral clustering is an algorithm that evolved from graph theory. The main idea is to treat all data as points in space, and these points can be connected by edges. The edge weight value between two points farther away is lower, and the weight value between two points closer is higher. With the eigenvectors of matrices as the input algorithm, spectral clustering adopts graph distance geometry and arbitrarily identified cluster shape. By cutting the graph composed of all data points, the difference after cutting the sum of the edge weights

between the subgraphs is low, while the sum of the edge weights in the subgraphs is quite high, so the purpose of clustering can be achieved (Bürk, 2012; Von Luxburg, 2007). Referred from Ng's work (Ng et al., 2002), the normalized spectral clustering algorithm for LICRI is formulated as Table 7.1.

Table 7.1 LICRI spectral clustering model development procedures

Step	Explanation
1	Split the LICRI database into three data sets and normalize each data set;
2	Construct similarity graph with adjacency matrix by normal *k*-nearest neighbor approach by setting the number of neighbors to 15 and sigma value of 1;
3	Compute the normalized graph Laplacian L and its first eigenvectors $v_1, \dots, v_k$;
4	Set $V \in \mathbb{R}^{n \times k}$ as the matrix containing the vectors $v_1, \dots, v_k$ and formulate the matrix $U \in \mathbb{R}^{n \times k}$ by normalizing the matrix V;
5	Let $y_i \in \mathbb{R}^k, i \in \{1,2, \dots, n\}$ be the vector corresponding the *i*-th row of matrix U;
6	Cluster the points $(y_i)_{i=1,\dots,n}$ with the k-means algorithm into clusters $C_1, \dots, C_k$

The relationship between the vectors in matrix V and matrix U is:

$$u_{ij} = \frac{v_{ij}}{\sqrt{\sum_k v_{ik}^2}} \tag{7.12}$$

The key to spectral clustering algorithm is to construct similarity graph for a given set of data points $x_1, x_2, \dots, x_n$. Many methods, including the fully connected graph, ε-neighborbood graph and normal/mutual k-nearest neighbor graphs (Von Luxburg, 2007), can be applied to transform the given data set with pairwise similarities s_{ij} into a similarity graph. As mentioned in step 2, this study employs normal *k*-nearest neighbor approach to construct the similarity graphs. The spectral clustering algorithm is implemented with the help of *Matlab* statistics and machine learning toolbox and the fast and efficient spectral clustering package (Ingo, 2020). In contrast to the convex data set shape of *k*-means algorithm, spectral clustering tends to be useful for hard non-convex problems (Hocking et al., 2011).

7.1.2.4. Clustered data visualization

Data visualization is a challenge for model performance comparison since the LICRI database has a three-dimensional feature. One available technique as previously discussed is to utilize principal component analysis (PCA) to reduce the dimension of the data sets, and this work employs PCA to automatically consider weight values of principal components and to visualize *k*-means cluster models; while another method is to build star coordinates (SC), converting high-dimensional database to 2-dimensional coordinates.

Basically, the SC system is a curvilinear coordinate system. By defining the origin as a 2d point $O_n(x, y) = (o_x, o_y)$ and a series of n 2d vectors $A_n = \langle \overrightarrow{a_1}, \overrightarrow{a_2}, \dots, \overrightarrow{a_i}, \dots, \overrightarrow{a_n} \rangle$, the axes can be established and mapped to the Cartesian Coordinates (Kandogan, 2000). The

data points D_j from a high dimensional dataset D are converted to data points D_j' of the established 2d Cartesian Coordinates by the sum of all unit vectors $\overrightarrow{u_i} = (u_{x_i}, u_{y_i})$ on each coordinate, and the relationship between the original and converted data points are shown below:

$$D_j'(x,y) = [o_x + \sum_{i=1}^{n} u_{x_i} \cdot (d_{ji} - min_i),\ o_y + \sum_{i=1}^{n} u_{y_i} \cdot (d_{ji} - min_i)\ ] \quad (7.13)$$

where $$D_j = (d_{j0},\ d_{j1}, \dots,\ d_{ji}, \dots,\ d_{jn}),\ |\overrightarrow{u_i}| = \frac{|\overrightarrow{a_i}|}{max_i - min_i}$$

$$max_i = \max\{d_{ji}, 0\ \leq j\ <\ |D|\}, min_i = \min\{d_{ji}, 0\ \leq j\ <\ |D|\}$$

Moreover, the cluster projection diagram of any two response variables can be applied to find the optimal clustering model as well. The star coordinates and cluster projection diagrams are integrated with the spectral clustering algorithm to visualize the clustered data of the LICRI database. The silhouette plot, which has been widely applied to show the optimal number of clusters for unsupervised algorithms, is employed to find the better clustering models among three safety features between *k*-means and spectral clustering algorithms.

7.1.2.5. Cluster validation criterion

To evaluate which algorithm has a better clustering performance, cluster validation criteria are introduced in the work. Focusing on measuring the fit of a clustering structure itself, the study employs the internal validation indices to consider both cluster cohesion and cluster separation. Three common internal measures of cluster validation have been surveyed, including the Dunn index (Dunn, 1974), Davies-Bouldin index (Davies and

Bouldin, 1979), and Silhouette index (Rousseeuw, 1987). All the three indices have presented as robust strategies to predict the optimal clustering partitions. This work utilizes the Silhouette index as the cluster validation criterion because of its interpretation and validation of consistency within clusters of the liquid in-cylinder combustion database. The silhouette validation criterion applies a concise graphical representation, the silhouette plot, to display how well each data points have been clustered. Like the Dunn index, the higher the silhouette index is, the better the clustering performance would be. For the data point j in the cluster C_j, the mean distance between I and other data point within the same cluster is defined as:

$$a(j) = \frac{1}{|C_j|-1} \sum_{k \in C_j, j \neq k} d(j,k) \tag{7.14}$$

where $d(j,k)$ is the distance between two data points j and k within the cluster C_j. The reason for adding the item $\frac{1}{|C_j|-1}$ is because the distance $d(j,j)$ is excluded in the sum.

Then the distance of j to the points in some cluster C_l ($C_l \neq C_j$) other than C_j is defined as:

$$b(j) = \min_{l \neq j} \frac{1}{|C_l|} \sum_{k \in C_l} d(j,k) \tag{7.15}$$

Next, the silhouette value of one data point j can be expressed as:

$$s(j) = \begin{bmatrix} 1 - a(j)/b(j), & if\ a(j) < b(j) \\ 0, & if\ a(j) = b(j) \\ b(j)/a(j) - 1 & if\ a(j) > b(j) \end{bmatrix} \tag{7.16}$$

From the above expression, one can find the range of $s(j)$ is between -1 and 1: a vale close to 1 indicates the data point is clustered to the correct cluster whereas the value -1 tells the

data point is affected to the wrong cluster. This study applies the mean silhouette value to show the performance for a given cluster C_l, which is denoted as $\overline{s_l}$:

$$\overline{s_l} = \frac{1}{|C_l|} \sum_{j \in C_l} s(j) \tag{7.17}$$

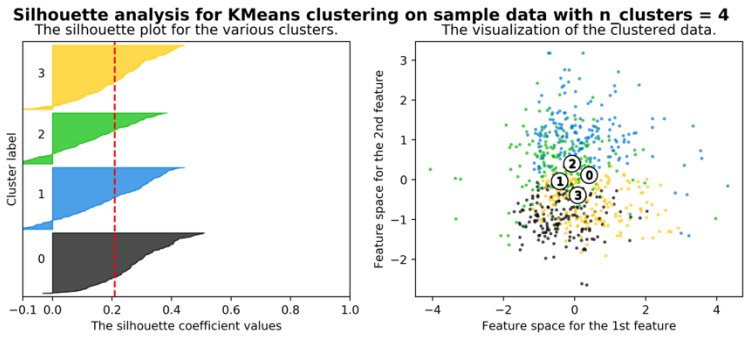
Finally, the overall performance of one specific model is able to be evaluate by the global silhouette index, the mean of the average silhouette values through all the clusters with the cluster number L:

$$\bar{S} = \frac{1}{L} \sum_{L=1}^{L} \overline{s_l} \tag{7.18}$$

7.2. Results and discussion

This study employs silhouette analysis to study the separation distance between the final clusters of k-means and spectral clustering algorithms. Also, the silhouette value is adopted to determine the optimal numbers of clusters for the LICRI database. The performance of the unsupervised clustering models is evaluated with the visualization of the clustered data and the silhouette plot.

7.2.1. *k*-means clustering



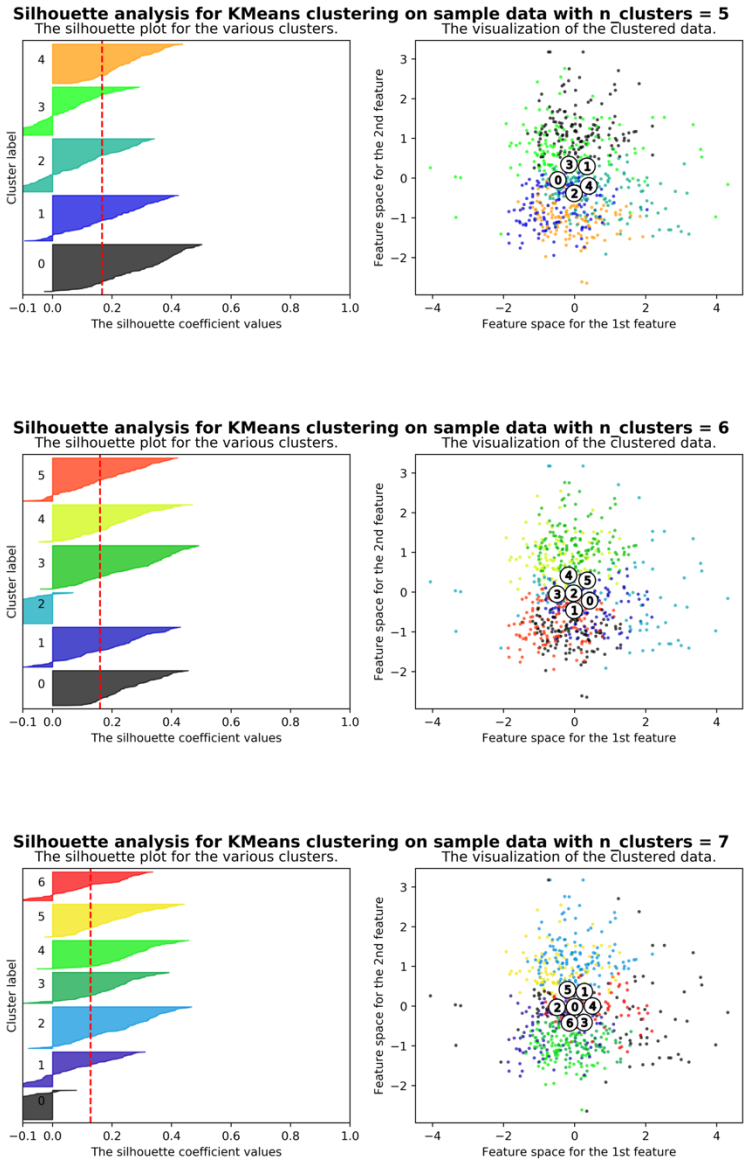


Figure 7.3 Silhouette plots and the n cluster labeled scatter plots for liquid flammability indicators by integrating PCA ($n \in \{4, 5, 6, 7\}$)

As shown in Figure 7.3, the silhouette plots of the k-means clustering show the silhouette coefficient values of n clusters for the three indicators of liquid flammability. The n clustering value of 6 and 7 are bad picks because each of them has a negative dominated cluster, indicating those samples might have been assigned to the wrong cluster. The right

figure is the n cluster labeled scatter plots, and its horizontal axis is PCA 1 and the vertical axis is PCA 2. The 4-cluster model seems to have fewer overlap points than others, and this statement is supported by the highest average silhouette value, 0.209, locating in the 4-cluster model and its narrow fluctuations in the size of the silhouette plots. These bring to the conclusion that the 4-cluster model is the optimal clustering model. However, many data points are locating in the unclear clusters from the 4-cluster labeled scatter plot.

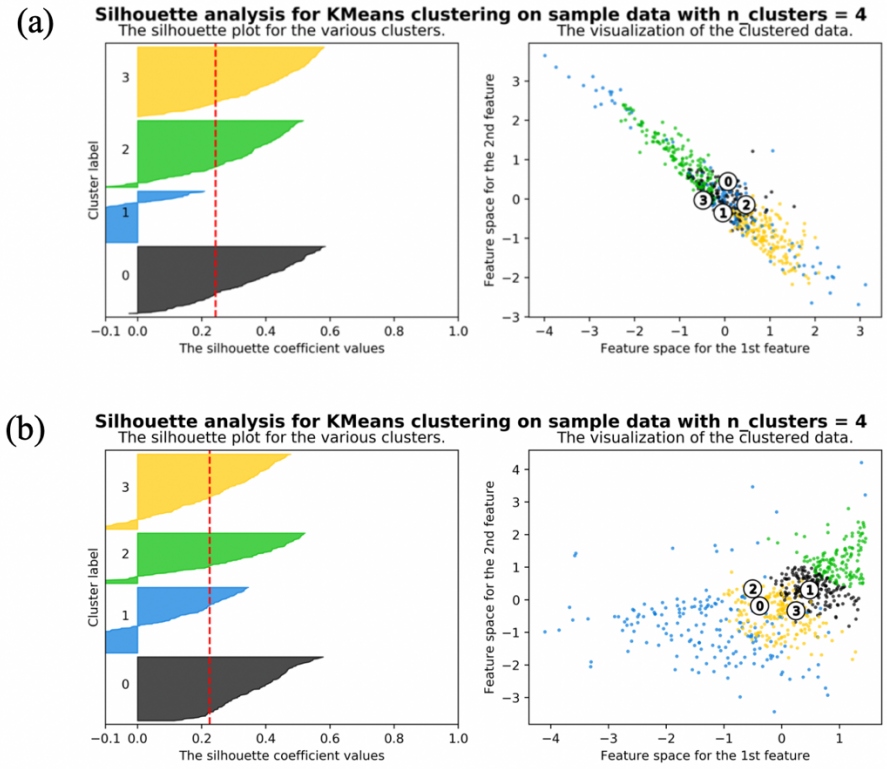


Figure 7.4 Silhouette plots and the n cluster labeled scatter plots for optimal flame propagation model (7.4a) and optimal liquid aerosolization model (7.4b)

Similarly, the study analyzed the silhouette plots and the n cluster labeled scatter plots for flame propagation and liquid aerosolization indicators by integrating PCA, and the optimal models are shown in Figure 7.4. Surprisingly, even the optimal clustering models have less acceptable cluster labeled scatter plots and wide fluctuation in the values of the silhouette coefficient. Therefore, a more accurate clustering algorithm is needed to establish a reasonable liquid combustion safety criterion.

7.2.2. Spectral clustering

By adopting the normal k-nearest neighbor approach, the two dimensional and three-dimensional similarity graphs for liquid flammability are shown in Figure 7.5. Also, the 3D similarity graphs for flame propagation and aerosol formulation can be found in Figure S-4 of Appendix D.

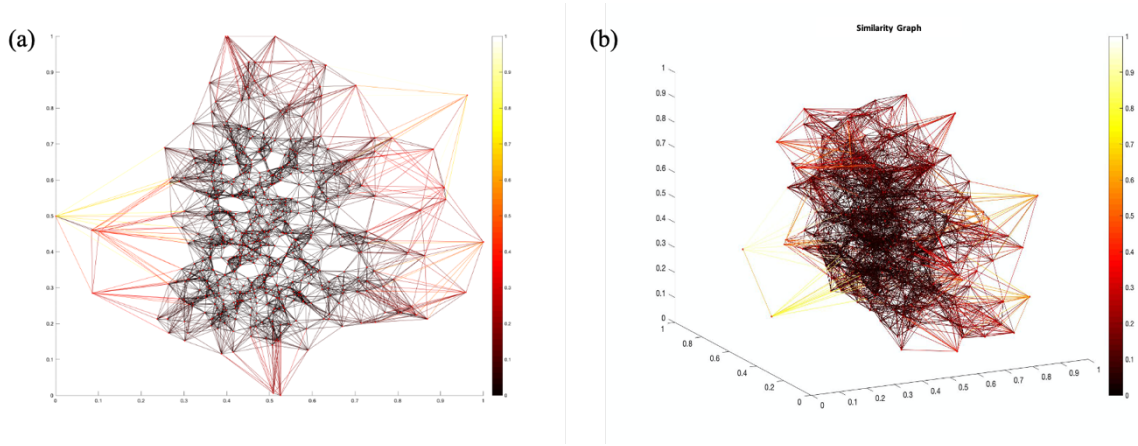


Figure 7.5 The two-dimensional (7.5a) and three-dimensional (7.5b) similarity graphs

Then the Ncut approach is employed to obtain the optimal number of connected components, and the n ($n \in \{4,5,6,7\}$) clustering performance is illustrated in Figure 7.6, the silhouette plots for liquid flammability indicators.

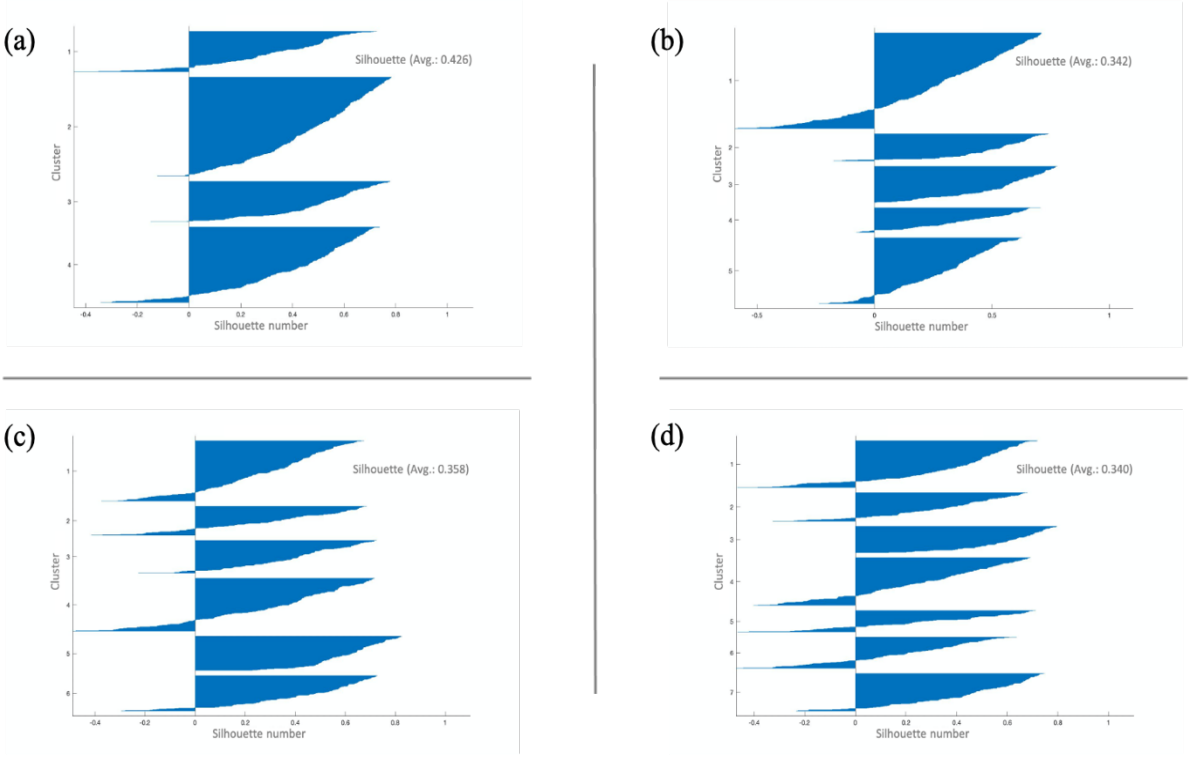


Figure 7.6 n cluster silhouette plots for liquid flammability indicators ($n \in \{4, 5, 6, 7\}$)

By considering the average silhouette value, the negative silhouette values, and the fluctuation of the size of each cluster, the four-cluster model is determined as the optimal spectral clustering model. Based on the optimal model, the clustered data for liquid flammability, flame propagation, and aerosol formulation with all combinations of dimensions are drawn, see Appendix D (Figure S-1, S-2, S-3) for details.

Figure 7.7(a) shows the 3-dimensional labeled scatter plot with little clustering effects, so the theory of star coordinate (Kandogan, 2000) was presented to visualize the performance of the proposed spectral clustering model, see Figure 7.7(b).

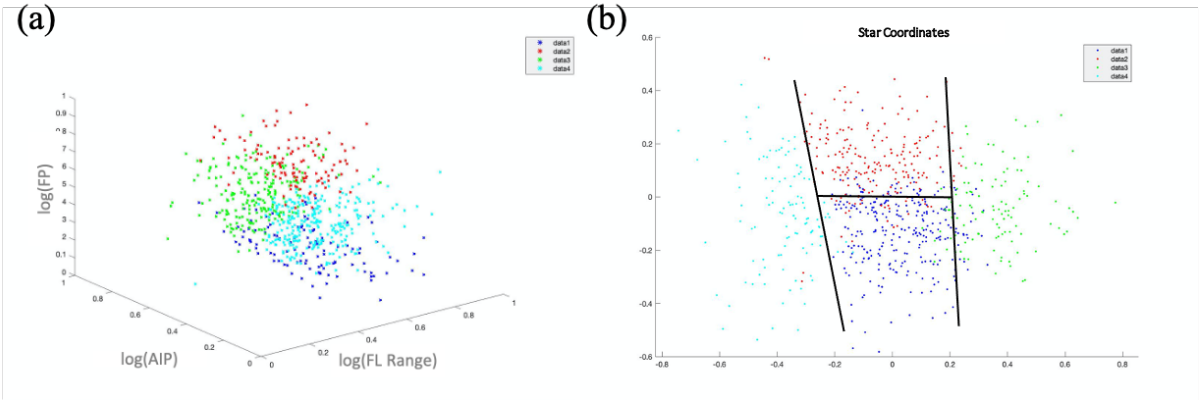


Figure 7.7 Three-dimensional labeled scatter plot (7.7a) and the corresponding star coordinate plot (7.7b)

Clearly, the spectral model performs great on clustering liquid flammability indicators, *i.e.*, FT, AIT, and FR, only a limited number of misclassified points locate in cluster 1 and cluster 2, while cluster 3 and 4 have excellent clustering feature. Compared with the *k*-means clustering model with 2 principal components, the spectral clustering model increases the average silhouette value from 0.209 to 0.426, and the clustering effect on the labeled scatter plot improves a lot as well.

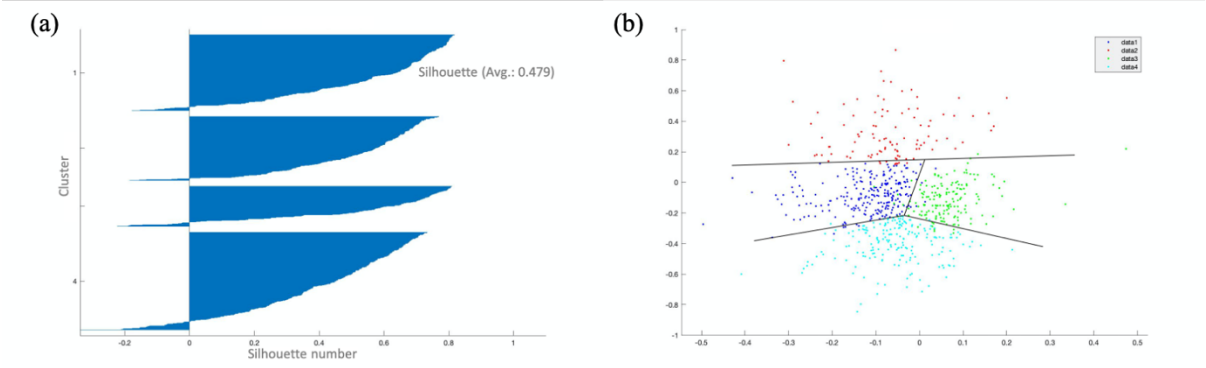


Figure 7.8 Optimal silhouette plot (7.8a) and optimal cluster star coordinate plot (7.8b) for flame propagation

The spectral clustering model performs even better in the dataset of flame propagation. Illustrated in Figure 7.8, the silhouette plot and the optimal cluster star coordinates plot show fairly good results on clustering HC, LD, and LTC, while the silhouette coefficient value is 0.479 and few outliers in the star coordinate plot have crossed the boundary of each cluster. Meanwhile, the silhouette and clustered SC plot for aerosol formulation which can be found in Appendix D (Figure S-5, S-6) point out the 4-cluster model as the optimal one.

To summarize, the spectral clustering models outperform k-means clustering models with two principal components for each liquid combustion safety matrix, see Table 7.2. Also, the silhouette plots of the spectral clustering models present better performance than those of the k-means clustering models, consistent with the values of global average silhouette numbers. The optimal cluster models can be determined by the highest value of silhouette coefficient value, and the 4-cluster models are selected as the optimal clustering models for the three liquid in-cylinder combustion safety features: liquid flammability, flame propagation, and aerosol formulation.

Table 7.2 Average silhouette coefficient value of three liquid combustion safety matrices for two clustering models

	Liquid Flammability				Flame Propagation				Aerosol Formulation			
	4 cluster model	5 cluster model	6 cluster model	7 cluster model	4 cluster model	5 cluster model	6 cluster model	7 cluster model	4 cluster model	5 cluster model	6 cluster model	7 cluster model
k-means clustering	0.209	0.167	0.160	0.128	0.243	0.181	0.153	0.142	0.225	0.219	0.156	0.145

models with 2 PCAs												
3-dimensional spectral clustering models	0.426	0.342	0.358	0.340	0.479	0.432	0.455	0.453	0.436	0.379	0.291	0.301

By employing the optimal clustering models, the risk ratings of each collected liquid flammability, flame propagation, and aerosol formulation are generated by the *Matlab* codes with the help of the calculated information entropy values. The whole clustered data and the weight value calculation can be found in supporting documents. The following table shows the example liquids with NFPA 704 flammability level 3 and 4 but different ratings based on our proposed clustering models.

Table 7.3 NFPA flammable and highly flammable liquids with different liquid in-cylinder combustion ratings

	NFPA Flammability	Liquid Flammability	Flame Propagation	Liquid Aerosolization	LICRI Value
Methanol	3	2	4	2	0.280
Ethanol	3	4	2	3	0.165
Methoxy acetone	3	4	4	2	0.103
o-Ethyl aniline	4	4	1	4	0.378
Di-(2-Chloroethoxy) Methane	4	3	2	3	0.223

As shown in Table 7.3, methanol and ethanol share the same flammability level in the NFPA 704 standard, but the proposed model points out that methanol is a less risky marine fuel on overall liquid in-cylinder risk combustion value than that of ethanol, although methanol has a high-risk rating on flame propagation. Based on the LICRI values of the extracted substances in Table 7.3, the safety preferences can be ranked as o-Ethyl aniline, methanol, di-(2-Chloroethoxy) methane, ethanol, and methoxy acetone. Following the same way, more promising fuels can be evaluated from the LICRI value of the spectral clustering model.

7.3. Concluding remarks

In this study, a novel combustion safety criterion for promising liquid fuel options is carried out successfully with acceptable clustering outputs. This work confirms that the graph theory-based spectral clustering performs better than *k*-means clustering algorithm in the non-convex liquid in-cylinder combustion database. The four cluster models are finalized as the optimal ones for liquid flammability, flame propagation, and liquid aerosolization. The liquid organic compound database is clustered into four groups for the three safety matrices, the low overall rating presents the high-level hazard. More than 700 liquid chemicals have been analyzed and rated by the proposed LICRI criterion. The *k*-means algorithm integrates PCA to automatically optimize weight values of each principal component; while the spectral clustering algorithm employs the star coordinate plots to only reduce the high dimensional data into two-dimensional data in the visualization stage.

The star coordinate plots provide a great way to visualize high dimensional data sets, and it can solve the most obvious disadvantage of PCA, lack of interpretability.

This work can be a solid reference for exploring more promising fuel options from the perspective of inherent safety. Compared with the NFPA flammability standard, this criterion embodies more safety information on fuel combustion in the CI engines. Also, the global mean silhouette value is reliable to find the optimal number of clusters in this work and it can be served as a robust approach to quantitatively evaluate the goodness of the clustering algorithms. Although the clustered results do show good performance on cluster 1 and cluster 4, the results still need to be improved on the boundary determination of cluster 2 and cluster 3. Nevertheless, the unsupervised clustering models with information entropy determined weight values give a more objective way to evaluate the risk associated with the liquid in-cylinder combustion since it avoids subjective human judgment for establishing the safety matrices.

7.4. References

Ballal, D.R., Lefebvre, A.H., 1979. Ignition and flame quenching of flowing heterogeneous fuel-air mixtures. Combust. Flame. https://doi.org/10.1016/0010-2180(79)90019-1

Bürk, I., 2012. Spectral Clustering. University of Stuttgart. https://doi.org/10.1201/9781315373515-8

Danis, A.M., Namer, I., Cernansky, N.P., 1988. Droplet size and equivalence ratio

effects on spark ignition of monodisperse N-heptane and methanol sprays. Combust. Flame. https://doi.org/10.1016/0010-2180(88)90074-0

Davies, D.L., Bouldin, D.W., 1979. A Cluster Separation Measure. IEEE Trans. Pattern Anal. Mach. Intell. https://doi.org/10.1109/TPAMI.1979.4766909

Dunn, J.C., 1974. Well-separated clusters and optimal fuzzy partitions. J. Cybern. https://doi.org/10.1080/01969727408546059

Elkotb, M.M., 1982. Fuel atomization for spray modelling. Prog. Energy Combust. Sci. 8, 61–91. https://doi.org/10.1016/0360-1285(82)90009-0

Harmon, D.B., 1955. Drop sizes from low speed jets. J. Franklin Inst. https://doi.org/10.1016/0016-0032(55)90098-3

Hocking, T.D., Armand;, J., Bach, F., Jean-Philppe, V., 2011. Clusterpath: An Algorithm for Clustering using Convex Fusion Penalties 1–8.

Ingo, 2020. Fast and efficient spectral clustering,. Retrieved May 9, 2020. [WWW Document]. MATLAB Cent. File Exch. URL https://www.mathworks.com/matlabcentral/fileexchange/34412-fast-and-efficient-spectral-clustering

Jiao, Z., Sun, Y., Hong, Y., Parker, T., Hu, P., Mannan, M.S., Wang, Q., 2020. Development of flammable dispersion quantitative property-consequence relationship models using extreme gradient boosting. Ind. Eng. Chem. Res. https://doi.org/10.1021/acs.iecr.0c02822

Jiao, Z., Yuan, S., Zhang, Z., Wang, Q., 2019. Machine learning prediction of hydrocarbon mixture lower flammability limits using quantitative structure-

property relationship models. Process Saf. Prog. https://doi.org/10.1002/prs.12103

Kandogan, E., 2000. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. Proc. IEEE Inf. Vis. Symp. Late Break. Hot Top. https://doi.org/10.1.1.4.8909

Knovel, 2019. DIPPR Project 801 - Full Version - Physical Constants - Knovel [WWW Document]. URL https://app.knovel.com/web/view/itable/show.v/rcid:kpDIPPRPF7/cid:kt00CZDUQI/viewerType:itble//root_slug:physical-constants/url_slug:physical-constants?filter=table&b-toc-cid=kpDIPPRPF7&b-toc-root-slug=&b-toc-url-slug=physical-constants&b-toc-title=DIPPR%25 (accessed 3.28.20).

Krishna, K., Kim, T.K., Kihm, K.D., Rogers, W.J., Mannan, M.S., 2003. Predictive correlations for leaking heat transfer fluid aerosols in air. J. Loss Prev. Process Ind. https://doi.org/10.1016/S0950-4230(02)00091-8

Krishna, Kiran, Rogers, W.J., Mannan, M.S., 2003. The use of aerosol formation, flammability, and explosion information for heat-transfer fluid selection, in: Journal of Hazardous Materials. https://doi.org/10.1016/S0304-3894(03)00273-5

Lefebvre, A.H., McDonell, V.G., 2017. Atomization and sprays, Atomization and Sprays. https://doi.org/10.1201/9781315120911

Li, X., Wang, K., Liuz, L., Xin, J., Yang, H., Gao, C., 2011. Application of the entropy weight and TOPSIS method in safety evaluation of coal mines, in: Procedia Engineering. https://doi.org/10.1016/j.proeng.2011.11.2410

Likas, A., Vlassis, N., J. Verbeek, J., 2003. The global k-means clustering algorithm.

Pattern Recognit. https://doi.org/10.1016/S0031-3203(02)00060-2

Mallard, E., Le Chatelier, H.L., 1883. Thermal Model for Flame Propagation, in: Annals of Mines. https://doi.org/10.1029/90JD02231

Ng, A.Y., Jordan, M.I., Weiss, Y., 2002. On spectral clustering: Analysis and an algorithm, in: Advances in Neural Information Processing Systems.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res.

Peters, N., 2000. Turbulent Combustion, Turbulent Combustion. https://doi.org/10.1017/cbo9780511612701

Polymeropoulos, C.E., 1984. Flame Propagation in Aerosols of Fuel Droplets, Fuel Vapor and Air. Combust. Sci. Technol. https://doi.org/10.1080/00102208408923807

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. https://doi.org/10.1016/0377-0427(87)90125-7

scikit-learn, 2017. Selecting the number of clusters with silhouette analysis on KMeans clustering [WWW Document]. Scikit-learn.

Von Luxburg, U., 2007. A tutorial on spectral clustering. Stat. Comput. https://doi.org/10.1007/s11222-007-9033-z

Yuan, S., Ji, C., Han, H., Sun, Y., Mashuga, C. V., 2020a. A review of aerosol

flammability and explosion related incidents, standards, studies, and risk analysis. Process Saf. Environ. Prot. 146, 499–514. https://doi.org/10.1016/j.psep.2020.11.032

Yuan, S., Ji, C., Monhollen, A., Kwon, J.S.-I., Mashuga, C., 2019. Experimental and thermodynamic study of aerosol explosions in a 36 L apparatus. Fuel 245, 467–477. https://doi.org/10.1016/j.fuel.2019.02.078

Yuan, S., Zhang, Z., Sun, Y., Kwon, J.S.-I., Mashuga, C. V., 2020b. Liquid flammability ratings predicted by machine learning considering aerosolization. J. Hazard. Mater. 386, 121640. https://doi.org/10.1016/j.jhazmat.2019.121640

Zhang, Z., Yuan, S., Yu, M., Mannan, M.S., Wang, Q., 2020. A Hazard Index for Chemical Logistic Warehouses with Modified Flammability Rating by Machine Learning Methods. ACS Chem. Heal. Saf. https://doi.org/10.1021/acs.chas.9b00026

8. CONCLUSIONS*

To meet the IMO 2050 decarbonization goals during this maritime energy transition period, this work applied a holistic approach by both mechnism and data-driven models to provide solutions covering IMO-compliant fuel emission analysis, onboard carbon capture system design, fuel aerosolization contributor prediction, and inherent safety property based in-cylinder combustion risk index development.

To summarize the sustainability study for TTP process, the two LNG fuel options displayed great potential for fuel savings and reductions in the emissions of CO_2, N_2O, and BCA. Although CH_4 emission may be an issue for LNG fueled ships, their overall GWP gas emissions were better than other options. As for the sustainable preferred ship fleet, bulk carriers outperformed general cargo ships on both nautical mile-based fuel consumption and CO_2 equivalent emissions per unit of DWT, while for the unit of liquid tank capacity, there is little difference in GHG emissions among LNG tanker, LPG tanker and oil tanker when the liquid tank size surpasses 100 thousand cubic meters.

* Reprinted with permission from “Ji, Chenxi, and Mahmoud M. El-Halwagi. "A data-driven study of IMO compliant fuel emissions with consideration of black carbon aerosols." *Ocean Engineering* 218 (2020): 108241. Ji, Chenxi, Shuai Yuan, Zeren Jiao, Mitchell Huffman, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Predicting flammability-leading properties for liquid aerosol safety via machine learning." *Process Safety and Environmental Protection* 148 (2021): 1357-1366. Ji, Chenxi, Zeren Jiao, Shuai Yuan, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Development of novel combustion risk index for flammable liquids based on unsupervised clustering algorithms." *Journal of Loss Prevention in the Process Industries* 70 (2021): 104422.” Ji, Chenxi, Shuai Yuan, Mitchell Huffman, Mahmoud M. El-Halwagi, and Qingsheng Wang. "Post-combustion carbon capture for tank to propeller via process modeling and simulation." *Journal of CO2 Utilization* 51 (2021): 101655.

As for TTPPCC, the MDEA-PZ system outperforms MEA and DIPA systems since it could capture more CO_2 while more specific reboiler duty can be saved. Newly blended aqueous amines, particularly ternary mixtures, might be the next focus on the solvent selection of the chemical absorption-based maritime capture system, despite the binary blended amines displaying satisfactory PCC performance in this study. The TTPPCC system needs more studies on the kinetics modeling, the second-order absorption rate constant, and activation energy to meet its requirement of fast reaction rate and low energy consumption. Moreover, process design for intensified unit operations involving rotating packed bed and printed circuit heat exchangers is another direction to expand on this study. Other post-combustion carbon capture techniques such as membrane-based approach, adsorption, and physical absorption can be integrated to the TTPPCC system as well.

The two safety studies for TTP process can be served as solid references for exploring more promising fuel options from the perspective of inherent safety. Liquid dynamic viscosity (LDV), surface tension (ST), and liquid vapor pressure (LVP) were identified as the most contributed factors for liquid aerosolization. Due to their good interpretation and prediction performance, the optimized machine learning models can be used to predict flammability-leading properties for aerosols and can therefore help design inherently safer processes involving potential liquid aerosolization. With reliable prediction models of the defined set of response variables, future studies can be expanded to formulate an aerosol flammability standard test or to establish a new safety

categorical criterion for the liquids. After successfully implementing machine-learning models, one can directly use the developed optimal models to expand the inherent safety-related properties for organic compounds in the liquid state. Future liquid aerosol safety studies may integrate the clustering and classification techniques via unsupervised and semi-supervised ML algorithms to discover the knowledge necessary to categorize emerging sustainable fuels.

The proposed criterion in the second TTP safety study embodies more safety information on fuel combustion in the CI engines compared with the NFPA flammability standard. Also, the safety work provides a reliable prediction for the identified liquid aerosol safety parameters based on a thorough comparison among multiple machine learning algorithms, with an emphasis on liquid aerosolization, aerosol flame speed, and flame thickness. The future safety work may focus on improving the clustering results by adopting other clustering techniques such as hierarchical clustering and density-based spatial clustering of algorithms with noise and expand this work to integrate sustainability parameters of the promising marine fuels. In this way, the greener and safer fuel solutions can be determined to meet the long-term sustainable and safe strategy.

APPENDIX A

SUPPLEMENTAL DOCUMENT FOR SHIP EMISSION AND FUEL CONSUMPTION

1. Data sample description

This research collected 337 ship data from the Lloyd's List Intelligence, the detailed information is as follows:

Tab. S-1 Features of collected ship database

Ship Class	Ship Capacity Bin	Age
Bulk carrier	above 10000 DWT	after 2005
General cargo ship	above 10000 DWT	after 2005
Container ship	above 3000 TEU	after 2015
LNG carrier	above 10000 LTC	after 2005
LPG carrier	above 10000 LTC	after 2005
Oil tanker	above 10000 LTC	after 2005

Tab. S-2 BCA fuel based emission factors (kg BCA/ ton fuel) from (Comer et al., 2017; Pavlenko et al., 2020)

HFO with Scrubber	0.19
MGO	0.04
VLSFO	0.26
LPDF	0.003

HPDF	0.003

2. AE and AB emission determination

The total ME power and capacity bins of the collected ship samples, associating with the ratio of AE/ME and AE load factors.

Tab. S-3 Total ME power, capacity bins, AE/ME ratio, AE LF

Ship class	Total ME(kw)	Total Capacity Bin	AE/ME[1]	AE Load Factor for Cruising[2]
BC	572078	5355659 dwt	0.222	0.17
CS	2185165	422149 teu	0.220	0.13
GCS	295087	295087 dwt	0.191	0.80
LNG	1285532	12000195 m3	0.211	0.13
LPG	532948	2640167 m3	0.211	0.13
OT	1058849	11921208 m3	0.211	0.13

1. Source from ENTEC final report (ENTEC, 2010);
2. Source from (Moreno-Gutiérrez et al., 2015)

The power demand of AB varies with the ship type and ship capacity bin. Referred from the (Olmer et al., 2017), the crusing boiler power demand relationship for our selected ship type is shown in the below table.

Tab. S-4 Boiler power demand for cruising mode by ship type and capacity bin

Ship class	Ship Capacity Bin	Effective Power Demand(kw)
Bulk Carrier	Any	0
Container ship	Any	0
General cargo ship	Any	0
LNG/LPG	<50000	100
	[50000,200000]	150
	>200000	300
Oil tanker	<20000	0
	[20000,80000]	150
	(80000,120000)	200
	[120000,200000]	250
	>200000	300

3. Ship fuel consumption

Fuel consumption per hour for cruising mode is calculated as follows:

$$FC_{pm_{i,j}} = \sum_{j} \left(\frac{CO_{2_{ME,i,j}} + CO_{2_{AE,i,j}} + CO_{2_{AB,i,j}}}{CO_2\ Intensity}\right)$$

$FC_{pm_{i,j}}$: Fuel consumption per nautical mile for ship class i, fuel option j.

$CO_{2_{ME,i,j}}$: Amout of CO2 emission per nautical mile emitted by main engine.

$CO_{2_{AE,i,j}}$: Amout of CO2 emission per nautical mile emitted by auxiliary engine.

$CO_{2_{AB,i,j}}$: Amout of CO2 emission per nautical mile emitted by auxiliary boiler.

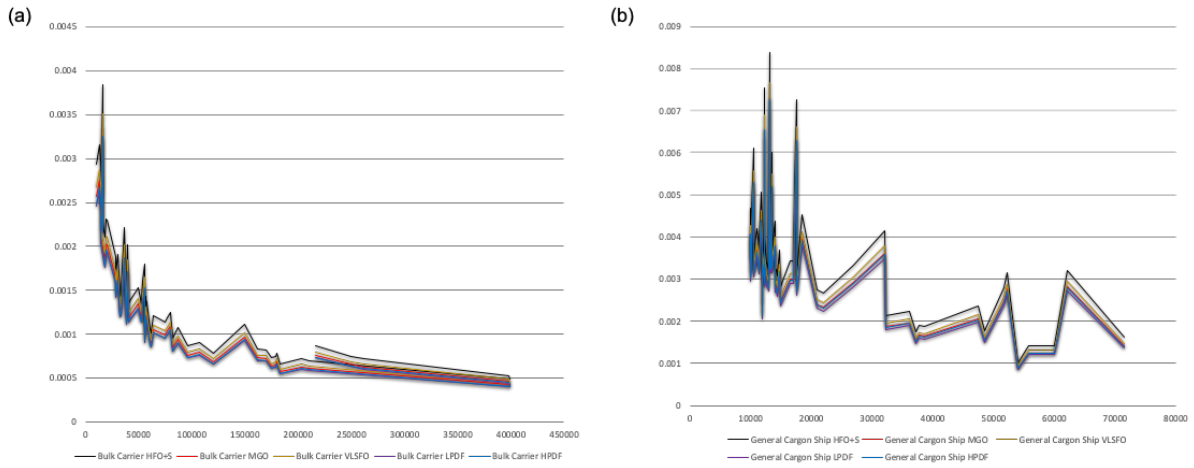


Fig. S-1 Fuel consumption per nautical mile per DWT for bulk carrier and general cargo ship

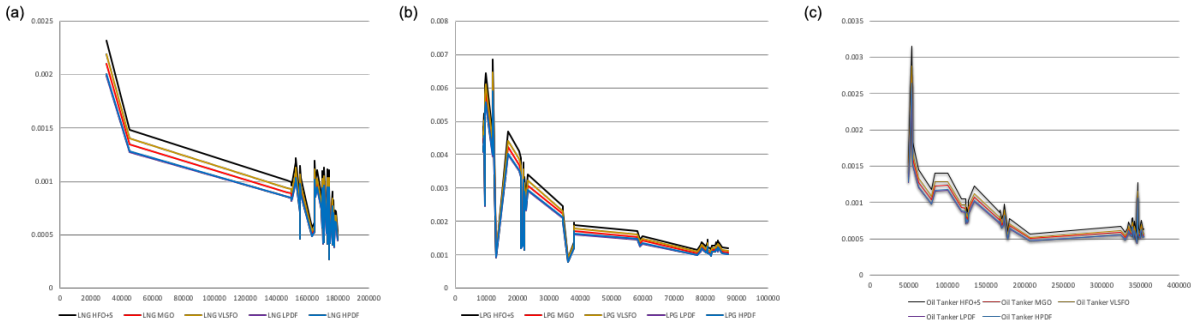


Fig. S-2 Fuel consumption per nautical mile per LTC for LNG tanker, LPG tanker and oil tanker

Tab. S-5 Calculated data for Figure 2 in the paper

	Bulk Carrier			Container Ship			General Cargo Ship			LNG Tanker				LPG Tanker				Oil Tanker			
HFO+S	M.E.	A.E.	Total	M.E.	A.E.	Total	M.E.	A.E.	Total	M.E.	A.E.	A.B.	Total	M.E.	A.E.	A.B.	Total	M.E.	A.E.	A.B.	Total
uel Cons.*10 kg			9.34			24.48			6.84				13.61				8.12				18.46
m CO2 *10 k	26.93	2.14		72.24	4.00		16.35	4.97		40.24	2.15	0.01		24.01	1.26	0.01		52.65	2.93	1.90	
Em CH4 g	2.52	0.20	--	6.77	0.37	--	1.53	0.46	--	3.77	0.20	0.02	--	2.25	0.12	0.02	--	4.94	0.27	0.04	--
Em N2O g	13.09	1.03		35.12	1.93		7.95	2.39		19.56	1.03	0.48		11.67	0.61	0.44		25.59	1.41	1.00	
Em BCA g	--	--	17.74	--	--	46.52	--	--	13.01	--	--		25.87	--	--		15.43	--	--		35.08
MGO	--	--		--	--		--	--		--	--			--	--			--	--		
uel Cons.*10 kg			8.17			21.43			5.99				12.20				7.37				16.22
m CO2 *10 k	24.27	1.93		65.09	3.60		14.73	4.47		36.26	1.93	0.93		21.64	1.14	0.84		47.44	2.64	1.92	
Em CH4 g	2.48	0.20	--	6.65	0.37	--	1.50	0.45	--	3.70	0.20	0.02	--	2.21	0.12	0.02	--	4.84	0.27	0.04	--
Em N2O g	12.62	1.03		33.86	1.93		7.66	2.39		18.86	1.03	0.39		11.26	0.61	0.35		24.68	1.41	0.80	
Em BCA g			3.27			8.57			2.40				4.88				2.95				6.49
VLSFO	--	--		--	--		--	--		--	--			--	--			--	--		
uel Cons.*10 kg			8.53			22.37			6.25				12.73				7.68				16.92
m CO2 *10 k	24.92	1.98		66.85	3.70		15.13	4.59		37.23	1.99	0.93		22.22	1.17	0.84		48.72	2.71	1.92	
Em CH4 g	2.43	0.20	--	6.52	0.37	--	1.48	0.45	--	3.63	0.20	0.19	--	2.17	0.12	0.17	--	4.75	0.27	0.40	--
Em N2O g	12.62	1.00		33.86	1.86		7.66	2.31		18.86	1.00	0.39		11.26	0.59	0.35		24.68	1.36	0.80	
Em BCA g			22.18			58.16			16.26				33.10				19.98				43.98
LPDF	--	--		--	--		--	--		--	--			--	--			--	--		
uel Cons.*10 kg			7.84			20.54			5.76				11.58				6.96				15.31
m CO2 *10 k	19.50	1.58		52.30	2.96		11.84	3.67		29.13	1.59	0.43		17.39	0.93	0.39		38.12	2.17	0.89	
Em CH4 g	60.78	4.63	--	163.04	8.64	--	36.90	10.72	--	90.81	4.63	1.26	--	54.20	2.73	1.14	--	118.83	6.33	2.59	--
Em N2O g	5.14	0.39		13.80	0.73		3.12	0.91		7.68	0.39	0.11		4.59	0.23	0.10		10.06	0.54	0.22	
Em BCA g			0.24			0.62			0.17				0.35				0.21				0.46
HPDF	--	--		--	--		--	--		--	--			--	--			--	--		
uel Cons.*10 kg			7.90			20.65			5.93				11.66				7.01				15.43
m CO2 *10 k	17.30	1.58		46.40	2.96		10.50	3.67		25.85	1.59	0.43		15.43	0.93	0.39		33.82	2.17	0.89	
Em CH4 g	56.11	4.27	--	150.50	7.97	--	34.06	9.90	--	83.83	4.28	1.16	--	50.03	2.52	1.05	--	109.69	5.85	2.39	--
Em N2O g	4.68	0.36		12.54	0.66		2.84	0.82		6.99	0.36	0.10		4.17	0.21	0.09		9.14	0.49	0.20	
Em BCA g			0.24			0.62			0.18				0.35				0.21				0.46

4. Case study data description

Ship class and DWT distribution based on the AIS collected data for the ships transited from Houston to Rotterdam.

Tab. S-6 Summary of AIS data for ship travelling from Houston to Rotterdam during Oct. 2019

Row Labels	Count of Vessel Type	Sum of DWT
bulk carrier	40	2680417
chemical tanker	10	233661
crude oil tanker	15	2064658
fruit juice tanker	1	43067
fully cellular containership	52	3081899
general cargo	27	461168

Liquefied Petroleum Gas Carrier	6	208343
LNG Carrier	5	289186
Naval Vessel	1	49000
oil tanker	37	2042693
passenger (cruise)	3	36878
reefer	1	13202
research	1	1000
Roll On Roll Off	1	14800
roll on roll off with container capacity	6	267746
semi-sub HL vessel	1	13300
support	1	
vehicle carrier	25	560109
Vessel Type (unspecified)	1	
Grand Total	**234**	**12061127**

Two regression model to convert DWT to TEU for container ships and DWT to LTC for tankers

Referred from the (Abramowski et al., 2018), the regression formula between TEU and DWT is as follows:

$$TEU = 372.53 + 7.94 \times 10^{-3} \cdot DWT^{1.2}$$

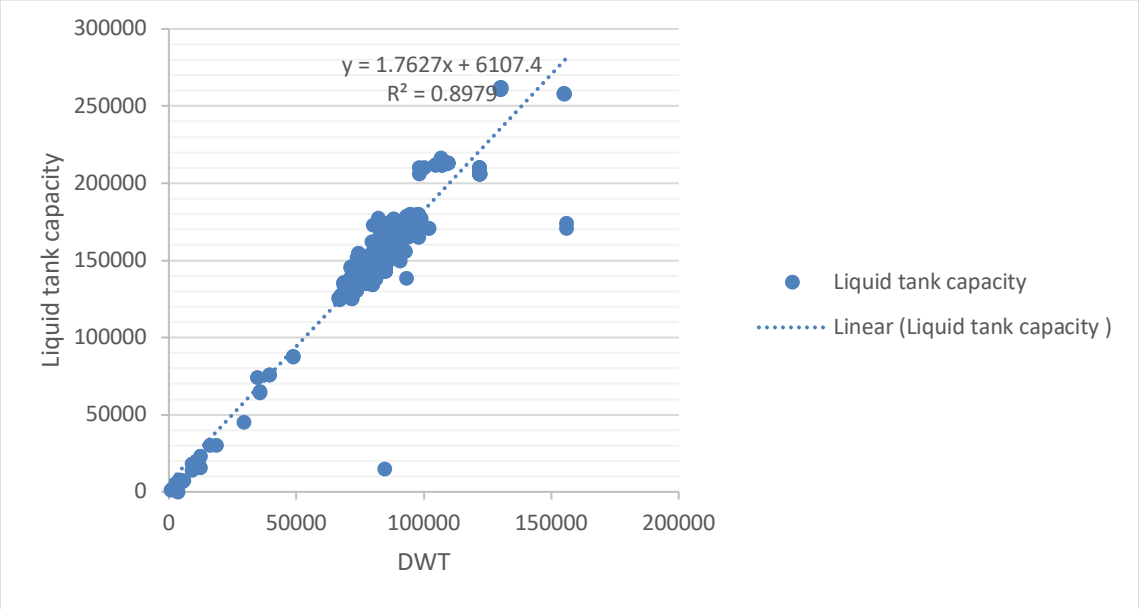


Fig. S-3 Fit regression model between LTC and DWT for tankers

References

Abramowski, T., Cepowski, T., Zvolenský, P., 2018. Determination of Regression Formulas for Key Design Characteristics of Container Ships at Preliminary Design Stage. New Trends Prod. Eng. 1, 247–257. https://doi.org/10.2478/ntpe-2018-0031

Comer, B., Olmer, N., Mao, X., Roy, B., Rutherford, D., 2017. Black Carbon Emissions and Fuel Use in 2015. Int. Counc. Clean Transp.

ENTEC, 2010. UK Ship Emissions Inventory Final Report.

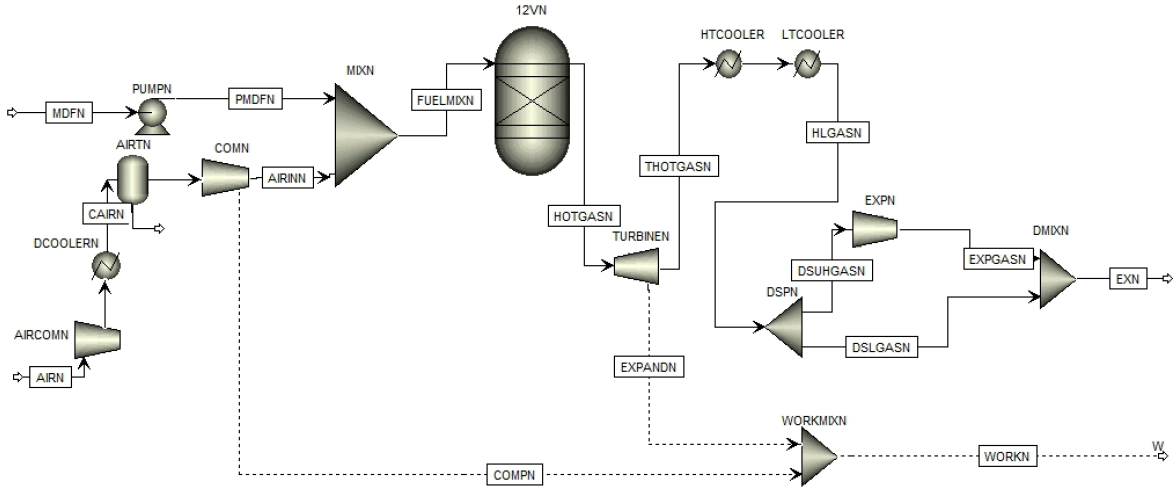
Moreno-Gutiérrez, J., Calderay, F., Saborido, N., Boile, M., Rodríguez Valero, R., Durán-Grados, V., 2015. Methodologies for estimating shipping emissions and energy consumption: A comparative analysis of current methods. Energy 86, 603–616. https://doi.org/10.1016/j.energy.2015.04.083

Olmer, N., Comer, B., Roy, B., Mao, X., Rutherford, D., 2017. Greenhouse Gas Emissions From Global Shipping, 2013-2015. Int. Counc. Clean Transp. 1–25.

APPENDIX B

SUPPLEMENTAL DOCUMENT FOR MARITIME CAPTURE PROCESS DESIGN

(a)



(b)

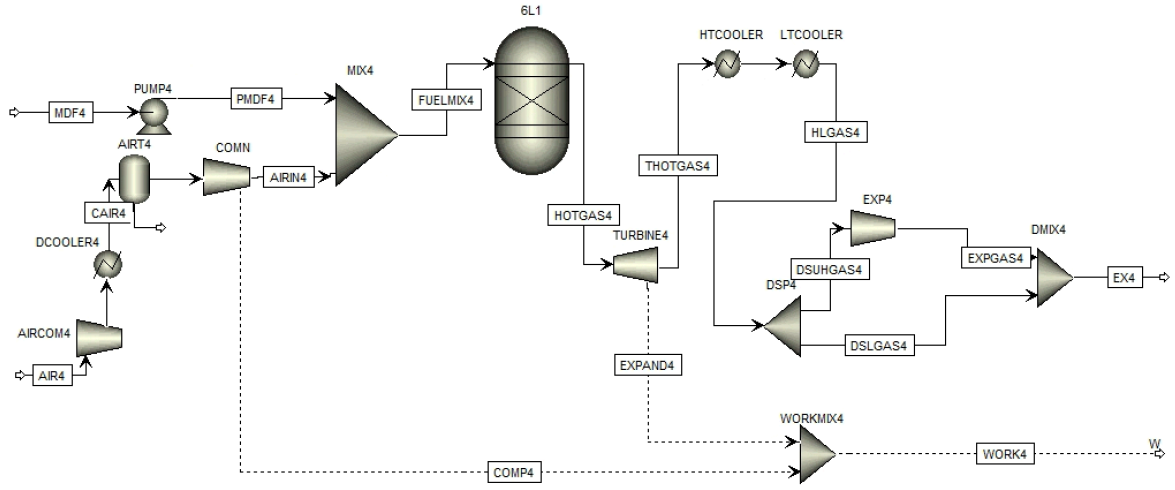


Fig. S-1 Detailed process design of the two dual engines: a) Wärtsilä 12V50DF and b) 6L50DF

Table S-1. Parameters of reference ship

Item	Value
Size (m^3)	154,472
Length (m)	289.6
Beam (m)	45.35
Draft (m)	12.5
Propulsion engine	3* Wärtsilä 12V50DF & 1*Wärtsilä 6L50DF
Deadweight (mt)	74,300
Propulsion power (MW)	39.9

Tab. S-2 Marine engine (12V50DF) simulation output

Flue gas flow rate, kg/s	19.22
Flue gas inlet temperature, c	355.9
Steam pressure, kpa	900
CO2 production, kg/s	1.93
Available heat in exhaust gas, kw	1446
CO2 concentration in exhaust gas	10.02%
Heat available per kg co2 produced (kwh/kg)	0.21

Tab.S-3 Parameters and correlations selection for mass transfer in RateSep model

Parameters	Correlations
Flow model	Vplug
Film discretization points	20
Mass transfer coefficients	Brf-85
Heat transfer coefficient	Chilton and Colburn

Tab. S-4 Pilot plant carbon capture process simulation outputs and validation

For Mellapak plus 252Y (experimental data from a recent work by Morgan et al. [1], Case 1A)

Variable	**Experiment output**	**Sim. model output**	**Absolute relative deviation, %**
Lean in CO_2 loading (mol CO_2/mol MEA)	0.24	0.253	5.4
Rich out CO_2 loading (mol CO_2/mol MEA)	0.46	0.486	5.7
CO_2 emission reduction efficiency (%)	97.5%	95.9%	1.6
Specific reboiler duty (MJ/kg CO_2)	4.451	4.781	7.4

For Sulzer BX (experimental data from a study by Mangalapally and Hasse [2], Case 3)

Variable	**Experiment output**	**Sim. model output**	**Absolute relative deviation, %**
Lean in CO_2 loading (mol CO_2/mol MEA)	0.205	0.195	4.9
Rich out CO_2 loading (mol CO_2/mol MEA)	0..465	0.449	3.4
CO_2 emission reduction efficiency (%)	90.9%	89.7%	1.3
Specific reboiler duty (MJ/kg CO_2)	4.20	4.28	1.9

For Flexipac 2Y (experimental data from a study by Hamborg et al. [3], Case C1-1a)

Variable	Experiment output	Sim. model output	Absolute relative deviation, %
Lean in CO_2 loading (mol CO_2/mol MEA)	0.230	0.225	2.2
Rich out CO_2 loading (mol CO_2/mol MEA)	0.480	0.476	0.8
CO_2 emission reduction efficiency (%)	91.5%	90.9%	0.7
Specific reboiler duty (MJ/kg CO_2)	4.16	4.32	3.8

Tab. S-5 Molar cyclic capacity under variance of packing materials and L/G

Scenario (packing type, L/G ratio)	Solvent	cyclic capacity mole co2/mole amine
BX, 1	MEA	0.29
	DIPA	0.289
	PZ+MDEA	0.362
BX,2	MEA	0.229
	DIPA	0.264
	PZ+MDEA	0.301
250, 1	MEA	0.287
	DIPA	0.271
	PZ+MDEA	0.359

250,2	MEA	0.218
	DIPA	0.247
	PZ+MDEA	0.29
2y,1	MEA	0.286
	DIPA	0.268
	PZ+MDEA	0.358
2y,2	MEA	0.207
	DIPA	0.24
	PZ+MDEA	0.279
252,1	MEA	0.287
	DIPA	0.27
	PZ+MDEA	0.359
252,2	MEA	0.212
	DIPA	0.243
	PZ+MDEA	0.284

References

[1] J.C. Morgan, A.S. Chinen, C. Anderson-Cook, C. Tong, J. Carroll, C. Saha, B. Omell, D. Bhattacharyya, M. Matuszewski, K.S. Bhat, D.C. Miller, Development of a framework for sequential Bayesian design of experiments: Application to a pilot-scale solvent-based CO2 capture process, Appl. Energy. 262 (2020) 114533. https://doi.org/10.1016/j.apenergy.2020.114533.

[2] H.P. Mangalapally, H. Hasse, Pilot plant study of post-combustion carbon dioxide capture by reactive absorption: Methodology, comparison of different structured packings, and comprehensive results for monoethanolamine, Chem. Eng. Res. Des. 89 (2011) 1216–1228. https://doi.org/10.1016/j.cherd.2011.01.013.

[3] E.S. Hamborg, V. Smith, T. Cents, N. Brigman, O.F.- Pedersen, T. De Cazenove, M. Chhaganlal, J.K. Feste, Ø. Ullestad, H. Ulvatn, O. Gorset, I. Askestad, L.K. Gram, B.F. Fostås, M.I. Shah, A. Maxson, D. Thimsen, Results from MEA testing at the CO2 Technology Centre Mongstad. Part II: Verification of baseline results, Energy Procedia. 63 (2014) 5994–6011. https://doi.org/10.1016/j.egypro.2014.11.634.

APPENDIX C

SUPPLEMENTAL DOCUMENT FOR LIQUID AEROSOL FLAMMABILITY

Table S-1. Predictors selected for constructing QSPR model

Predictor	Type	Definition
MW	Constitutional descriptor	Molecular weight
Para.	Constitutional descriptor	Parachor
DC	Electronic information	Dielectric constant
RI	Electronic information	Refractive index
BP	Experimental data	Boiling point (K)
MP	Experimental data	Melting point (K)
LMV	Geometrical descriptor	Liquid molar volume (m3/kmol)
RG	Geometrical descriptor	radius of gyration (m)
VDWa	Geometrical descriptor	Van Der Waals surface area (m2/kmol)
SP	Quantum descriptor	Solubility parameter ((J/m3)1/2)
Dip.	Quantum descriptor	Dipole moment (C m)
Int1	Interaction item	Between VDWa and Para.
Int2	Interaction item	Between Para. and LMV
Int3	Interaction item	Between RG and Para.

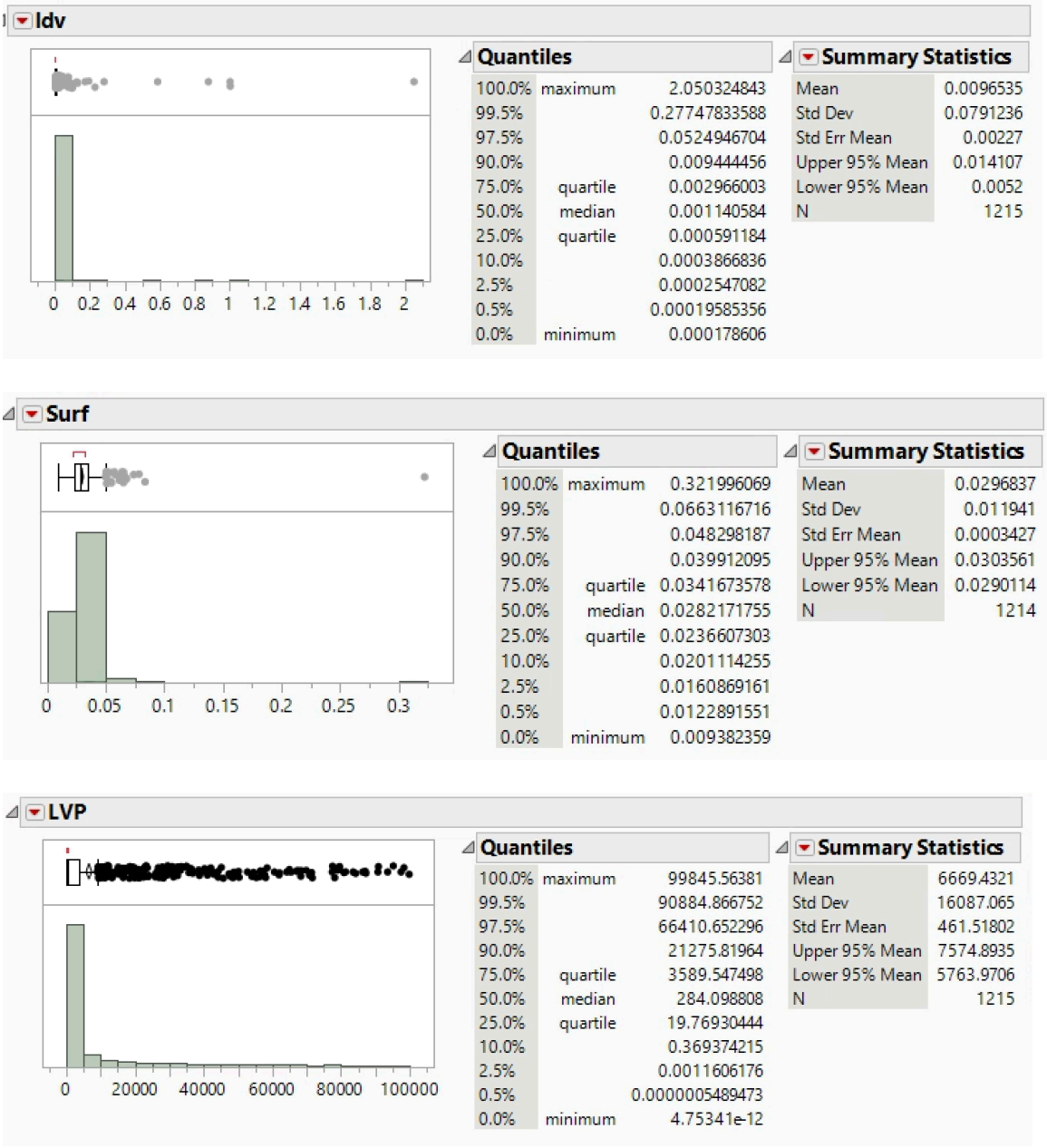


Figure S-1. Summary of statistics for original data for response variables

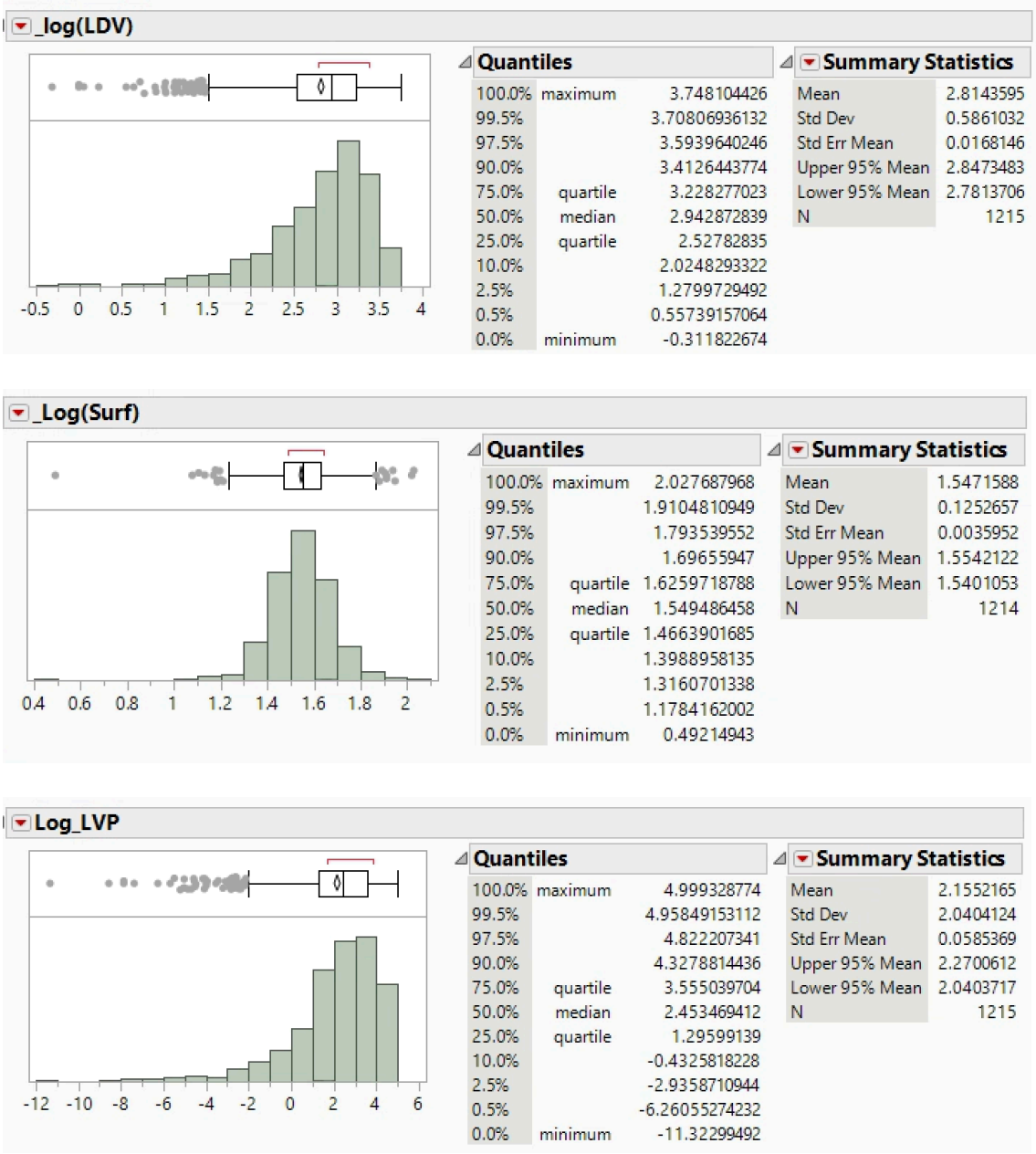


Figure S-2. Summary of statistics for transformed data for response variables

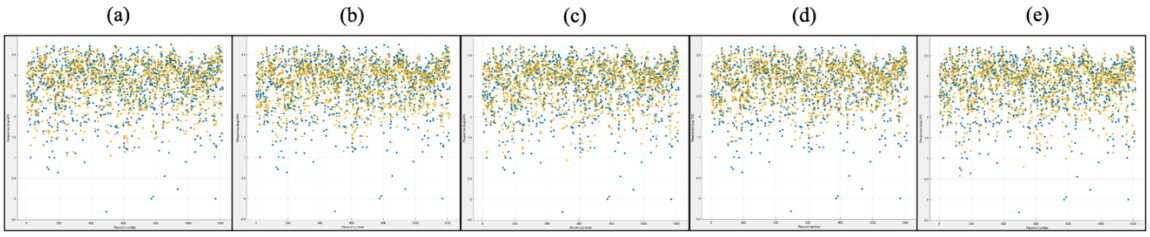


Figure S-3. Response plots for the proposed ML LDV models, with the blue dots serving as the experimental data and yellow dots for predicted values.

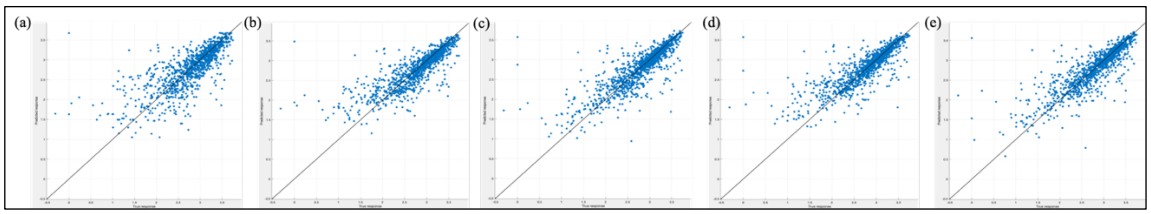


Figure S-4. Predicted vs experimental plots: (a) Medium regression tree; (b) Bagged Ensemble of Trees; (c) Exponential Gaussian process regression; (d) Medium Gaussian support vector machine model; (e) Exponential GPR with 7 PCAs

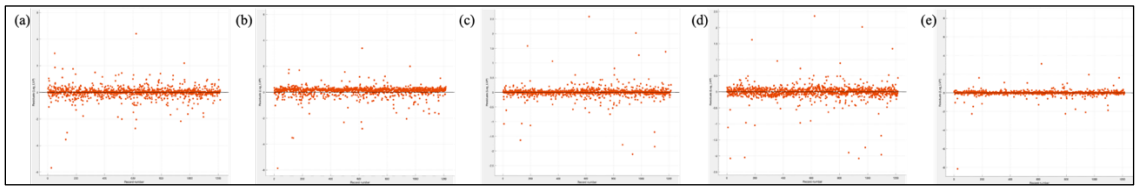


Figure S-5. Residual plots: (a) Medium regression tree; (b) Ensemble boosted tree; (c) Matern 5/2 GPR; (d) Quadratic SVM; (e) Matern 5/2 GPR with PCA

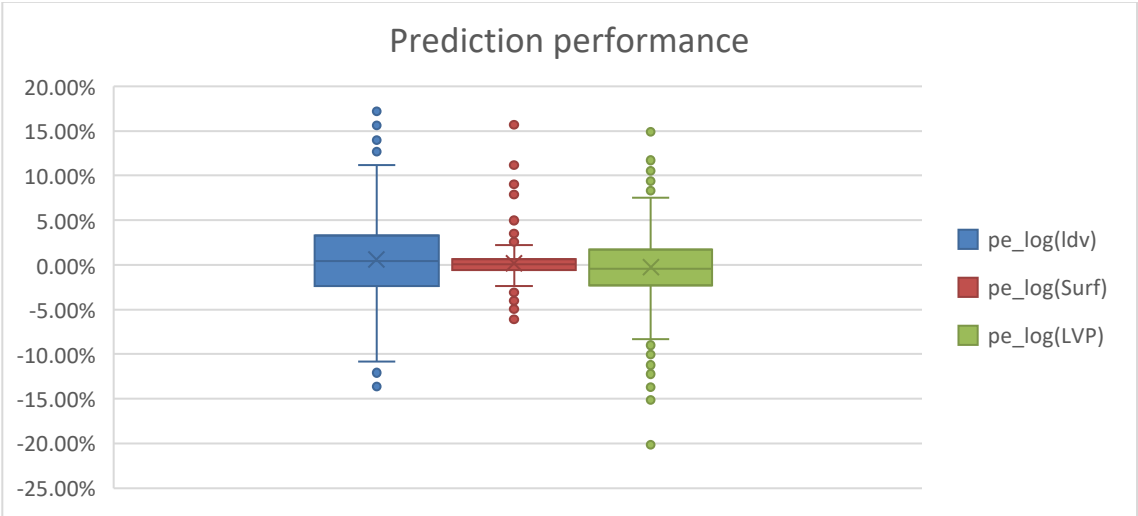


Figure S-6. Box plots of the three response parameters

Supplemental codes

Matlab code

```
%% Import data from spreadsheet
% Script for importing data from the following spreadsheet:
%
%    Workbook: /Users/jichenxi/Desktop/research
everyday/data/TableData-DIPPR python with st.xlsx
%    Worksheet: Table
%
% To extend the code for use with different selected data or a
different
% spreadsheet, generate a function instead of a script.

% Auto-generated by MATLAB on 2020/04/01 21:08:01

%% Import the data
[~, ~, raw] = xlsread('/Users/jichenxi/Desktop/research
everyday/data/TableData-DIPPR python with st.xlsx','Table');
raw = raw(2:end,:);
raw(cellfun(@(x) ~isempty(x) && isnumeric(x) && isnan(x),raw)) = {''};
stringVectors = string(raw(:,1));
stringVectors(ismissing(stringVectors)) = '';
raw =
raw(:,[2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24]);

%% Replace non-numeric cells with NaN
R = cellfun(@(x) ~isnumeric(x) && ~islogical(x),raw); % Find non-
numeric cells
raw(R) = {NaN}; % Replace non-numeric cells

%% Create output variable
data = reshape([raw{:}],size(raw));

%% Create table
TableDataDIPPRpythonwithst = table;

%% Allocate imported array to column variable names
TableDataDIPPRpythonwithst.Name = stringVectors(:,1);
TableDataDIPPRpythonwithst.mw = data(:,1);
TableDataDIPPRpythonwithst.MP = data(:,2);
TableDataDIPPRpythonwithst.LMV = data(:,3);
TableDataDIPPRpythonwithst.BP = data(:,4);
TableDataDIPPRpythonwithst.RG = data(:,5);
TableDataDIPPRpythonwithst.Dip = data(:,6);
TableDataDIPPRpythonwithst.DC = data(:,7);
TableDataDIPPRpythonwithst.RI = data(:,8);
TableDataDIPPRpythonwithst.VDWa = data(:,9);
TableDataDIPPRpythonwithst.SP = data(:,10);
TableDataDIPPRpythonwithst.Para = data(:,11);
TableDataDIPPRpythonwithst.LTC = data(:,12);
TableDataDIPPRpythonwithst.LVP = data(:,13);
```

```
TableDataDIPPRpythonwithst.Log_LVP = data(:,14);
TableDataDIPPRpythonwithst.log_LTC = data(:,15);
TableDataDIPPRpythonwithst.ln_LTC = data(:,16);
TableDataDIPPRpythonwithst.int1 = data(:,17);
TableDataDIPPRpythonwithst.int2 = data(:,18);
TableDataDIPPRpythonwithst.int3 = data(:,19);
TableDataDIPPRpythonwithst.Surf = data(:,20);
TableDataDIPPRpythonwithst.LogSurf = data(:,21);
TableDataDIPPRpythonwithst.ldv = data(:,22);
TableDataDIPPRpythonwithst.logLDV = data(:,23);

%% Clear temporary variables
clearvars data raw stringVectors R;

%% Machine Learning regression model code
function [trainedModel, validationRMSE] =
trainRegressionModel(trainingData)
% [trainedModel, validationRMSE] = trainRegressionModel(trainingData)
% returns a trained regression model and its RMSE. This code recreates
the
% model trained in Regression Learner app. Use the generated code to
% automate training the same model with new data, or to learn how to
% programmatically train models.
%
%  Input:
%      trainingData: a table containing the same predictor and response
%       columns as imported into the app.
%
%  Output:
%      trainedModel: a struct containing the trained regression model.
The
%       struct contains various fields with information about the
trained
%       model.
%
%      trainedModel.predictFcn: a function to make predictions on new
data.
%
%      validationRMSE: a double containing the RMSE. In the app, the
%       History list displays the RMSE for each model.
%
% Use the code to train the model with new data. To retrain your model,
% call the function from the command line with your original data or
new
% data as the input argument trainingData.
%
% For example, to retrain a regression model trained with the original
data
% set T, enter:
%   [trainedModel, validationRMSE] = trainRegressionModel(T)
%
% To make predictions with the returned 'trainedModel' on new data T2,
use
%   yfit = trainedModel.predictFcn(T2)
```

```
%
% T2 must be a table containing at least the same predictor columns as
used
% during training. For details, enter:
%   trainedModel.HowToPredict

% Auto-generated by MATLAB on 01-Apr-2020 23:14:40


% Extract predictors and response
% This code processes the data into the right shape for training the
% model.
inputTable = trainingData;
predictorNames = {'mw', 'MP', 'LMV', 'BP', 'RG', 'Dip', 'DC', 'RI',
'VDWa', 'SP', 'Para', 'int1', 'int2', 'int3'};
predictors = inputTable(:, predictorNames);
response = inputTable.Log_LVP;
isCategoricalPredictor = [false, false, false, false, false, false,
false, false, false, false, false, false, false, false];

% Apply a PCA to the predictor matrix.
% Run PCA on numeric predictors only. Categorical predictors are passed
through PCA untouched.
isCategoricalPredictorBeforePCA = isCategoricalPredictor;
numericPredictors = predictors(:, ~isCategoricalPredictor);
numericPredictors = table2array(varfun(@double, numericPredictors));
% 'inf' values have to be treated as missing data for PCA.
numericPredictors(isinf(numericPredictors)) = NaN;
numComponentsToKeep = min(size(numericPredictors,2), 7);
[pcaCoefficients, pcaScores, ~, ~, explained, pcaCenters] = pca(...
    numericPredictors, ...
    'NumComponents', numComponentsToKeep);
predictors = [array2table(pcaScores(:,:)), predictors(:,
isCategoricalPredictor)];
isCategoricalPredictor = [false(1,numComponentsToKeep),
true(1,sum(isCategoricalPredictor))];

% Train a regression model
% This code specifies all the model options and trains the model.
regressionGP = fitrgp(...
    predictors, ...
    response, ...
    'BasisFunction', 'constant', ...
    'KernelFunction', 'matern52', ...
    'Standardize', true);

% Create the result struct with predict function
predictorExtractionFcn = @(t) t(:, predictorNames);
pcaTransformationFcn = @(x) [ array2table((table2array(varfun(@double,
x(:, ~isCategoricalPredictorBeforePCA))) - pcaCenters) *
pcaCoefficients), x(:,isCategoricalPredictorBeforePCA) ];
gpPredictFcn = @(x) predict(regressionGP, x);
trainedModel.predictFcn = @(x)
gpPredictFcn(pcaTransformationFcn(predictorExtractionFcn(x)));
```

```
% Add additional fields to the result struct
trainedModel.RequiredVariables = {'mw', 'MP', 'LMV', 'BP', 'RG', 'Dip',
'DC', 'RI', 'VDWa', 'SP', 'Para', 'int1', 'int2', 'int3'};
trainedModel.PCACenters = pcaCenters;
trainedModel.PCACoefficients = pcaCoefficients;
trainedModel.RegressionGP = regressionGP;
trainedModel.About = 'This struct is a trained model exported from
Regression Learner R2017b.';
trainedModel.HowToPredict = sprintf('To make predictions on a new
table, T, use: \n  yfit = c.predictFcn(T) \nreplacing ''c'' with the
name of the variable that is this struct, e.g. ''trainedModel''. \n
\nThe table, T, must contain the variables returned by: \n
c.RequiredVariables \nVariable formats (e.g. matrix/vector, datatype)
must match the original training data. \nAdditional variables are
ignored. \n \nFor more information, see <a
href="matlab:helpview(fullfile(docroot, ''stats'', ''stats.map''),
''appregression_exportmodeltoworkspace'')">How to predict using an
exported model</a>.');

% Extract predictors and response
% This code processes the data into the right shape for training the
% model.
inputTable = trainingData;
predictorNames = {'mw', 'MP', 'LMV', 'BP', 'RG', 'Dip', 'DC', 'RI',
'VDWa', 'SP', 'Para', 'int1', 'int2', 'int3'};
predictors = inputTable(:, predictorNames);
response = inputTable.Log_LVP;
isCategoricalPredictor = [false, false, false, false, false, false,
false, false, false, false, false, false, false, false];

% Perform cross-validation
KFolds = 10;
cvp = cvpartition(size(response, 1), 'KFold', KFolds);
% Initialize the predictions to the proper sizes
validationPredictions = response;
for fold = 1:KFolds
    trainingPredictors = predictors(cvp.training(fold), :);
    trainingResponse = response(cvp.training(fold), :);
    foldIsCategoricalPredictor = isCategoricalPredictor;

    % Apply a PCA to the predictor matrix.
    % Run PCA on numeric predictors only. Categorical predictors are
passed through PCA untouched.
    isCategoricalPredictorBeforePCA = foldIsCategoricalPredictor;
    numericPredictors = trainingPredictors(:,
~foldIsCategoricalPredictor);
    numericPredictors = table2array(varfun(@double,
numericPredictors));
    % 'inf' values have to be treated as missing data for PCA.
    numericPredictors(isinf(numericPredictors)) = NaN;
    numComponentsToKeep = min(size(numericPredictors,2), 7);
    [pcaCoefficients, pcaScores, ~, ~, explained, pcaCenters] = pca(...
        numericPredictors, ...
```

```
        'NumComponents', numComponentsToKeep);
    trainingPredictors = [array2table(pcaScores(:,:)),
trainingPredictors(:, foldIsCategoricalPredictor)];
    foldIsCategoricalPredictor = [false(1,numComponentsToKeep),
true(1,sum(foldIsCategoricalPredictor))];

    % Train a regression model
    % This code specifies all the model options and trains the model.
    regressionGP = fitrgp(...
        trainingPredictors, ...
        trainingResponse, ...
        'BasisFunction', 'constant', ...
        'KernelFunction', 'matern52', ...
        'Standardize', true);

    % Create the result struct with predict function
    pcaTransformationFcn = @(x) [
array2table((table2array(varfun(@double, x(:,
~isCategoricalPredictorBeforePCA))) - pcaCenters) * pcaCoefficients),
x(:,isCategoricalPredictorBeforePCA) ];
    gpPredictFcn = @(x) predict(regressionGP, x);
    validationPredictFcn = @(x) gpPredictFcn(pcaTransformationFcn(x));

    % Add additional fields to the result struct

    % Compute validation predictions
    validationPredictors = predictors(cvp.test(fold), :);
    foldPredictions = validationPredictFcn(validationPredictors);

    % Store predictions in the original order
    validationPredictions(cvp.test(fold), :) = foldPredictions;
end

% Compute validation RMSE
isNotMissing = ~isnan(validationPredictions) & ~isnan(response);
validationRMSE = sqrt(nansum(( validationPredictions - response ).^2) /
numel(response(isNotMissing) ));

% Structure  'trainedModelaeo' exported from Regression Learner.
To make predictions on a new table, T:
T=readtable("/Users/jichenxi/Desktop/research everyday/data/validate
new.xlsx")
yfit = trainedModelaeo.predictFcn(T)
```

Part of Python code

```
%matplotlib inline
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import numpy as np

df = pd.read_excel('/Users/ log transform.xlsx')
df.head()
des=df[['mw', 'MP', 'LMV', 'BP', 'RG', 'Dip.', 'DC', 'RI', 'VDWa', 'SP',
      'Para.','Log_LVP','log_LTC']]
des.head()
df.corr()

df.corr().style.background_gradient(cmap='coolwarm')
plt.figure(figsize=(15,15))

sns.heatmap(des.corr(),annot=True,cmap='PuOr')

sns.set(style='ticks',color_codes=True)

g=sns.pairplot(des)

g=sns.pairplot(df, diag_kind="kde")
```

APPENDIX D

SUPPLEMENTAL DOCUMENT FOR LIQUID COMBUSTION RISK INDEX

Weight value determination

Table S-1. Information entropy, information effectiveness, weight values for nine indicators and final weight value for three features

	lvp	**ltc**	**Surf**	**ldv**	**heat capa**	**fl range**	**AIT**	**flash p**	**densit y**
informatio n entropy	0.997 9	0.995 3	0.993 1	0.996 4	0.989 2	0.995 5	0.988 9	0.984 8	0.994 7
information effectiveness	0.002 1	0.004 7	0.006 9	0.003 6	0.010 8	0.004 5	0.011 1	0.015 2	0.005 3
weight value	0.033 2	0.072 7	0.107 3	0.056 0	0.168 5	0.070 4	0.172 1	0.237 3	0.082 4

liquid aerosolization	0.19657416
flame propagation	0.32362415
flammability	0.47980169

Spectral clustering

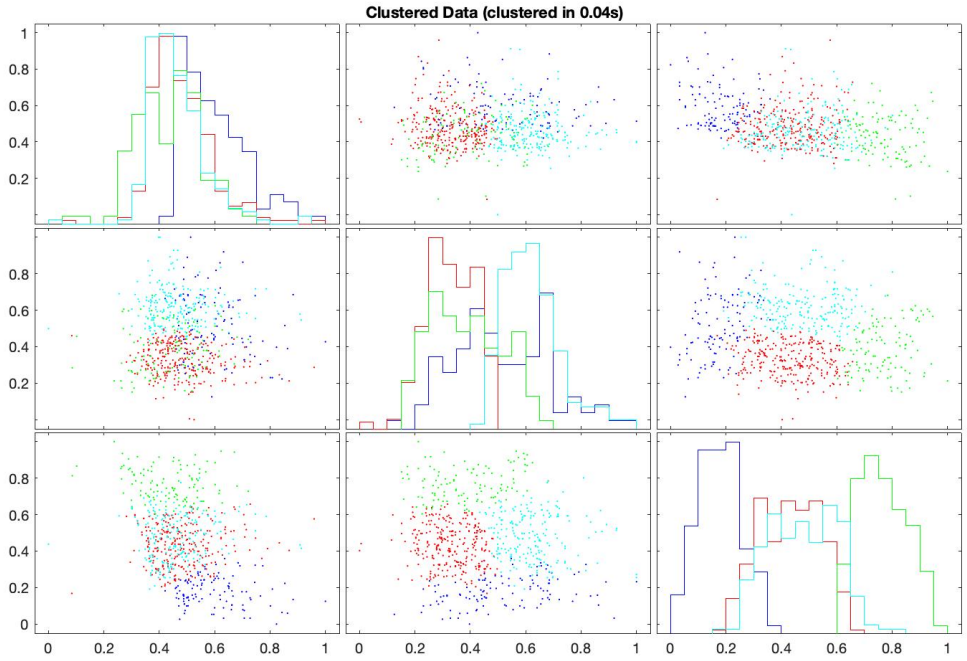


Figure S-1. Clustered data for liquid flammability using optimal model

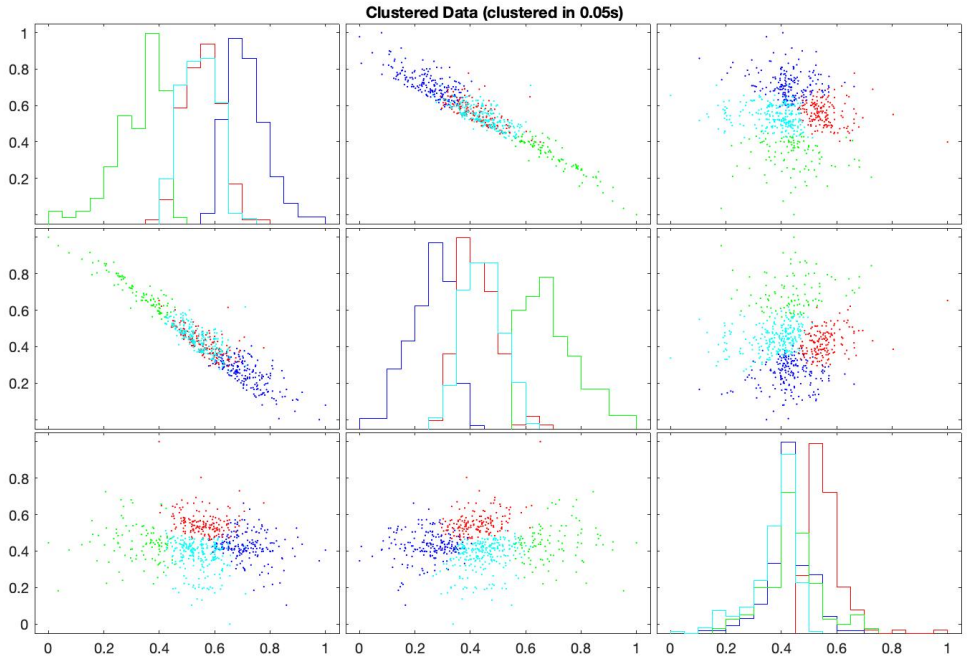


Figure S-2. Clustered data for flame propagation using optimal model

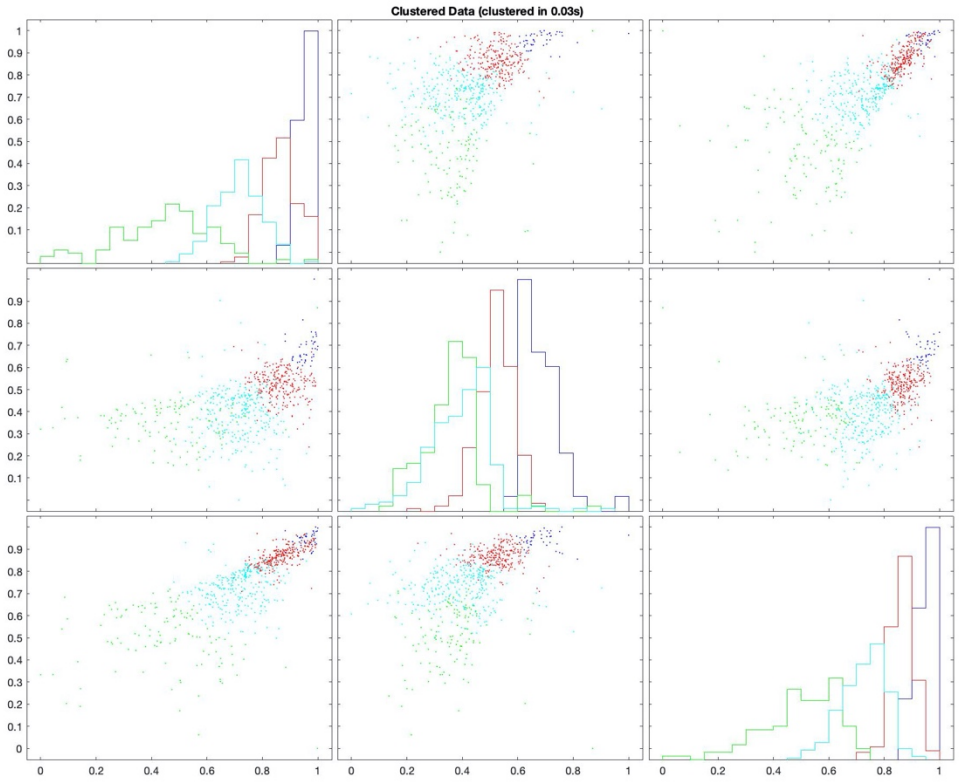


Figure S-3. Clustered data for aerosol formulation using optimal model

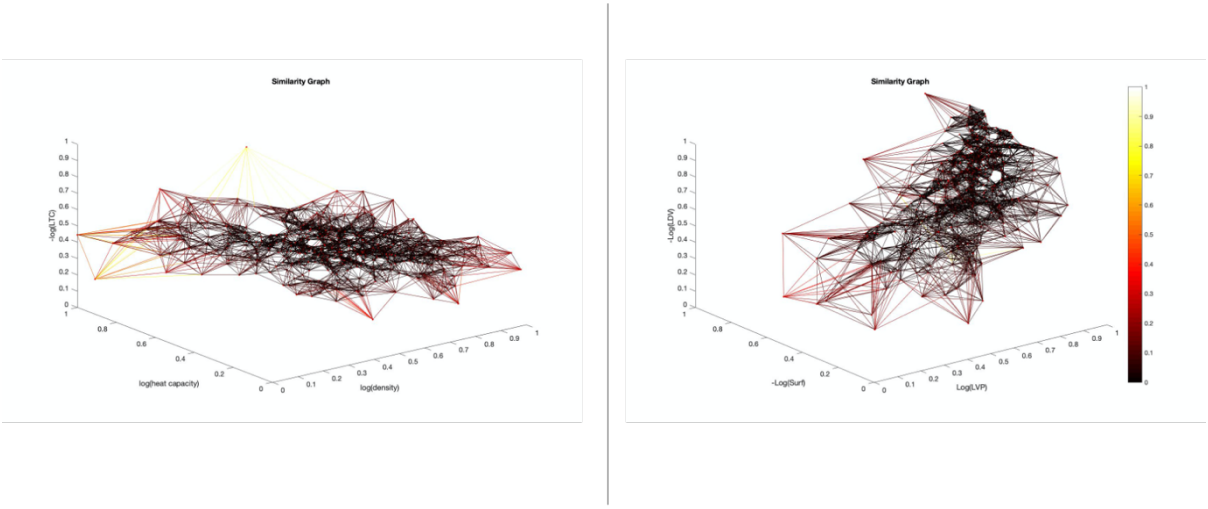


Figure S-4. The three-dimensional similarity graphs for flame propagation and aerosol formulation

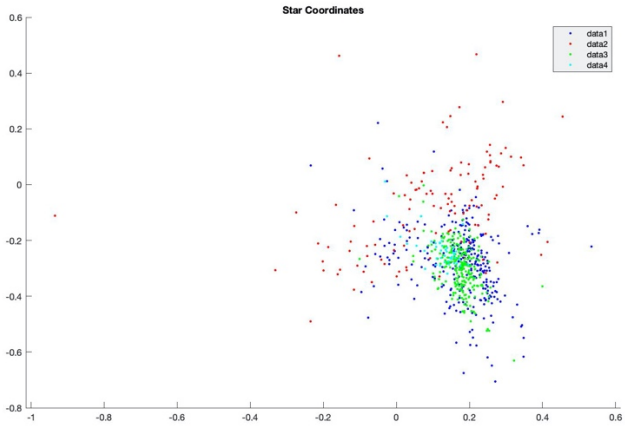


Figure S-5. Star coordinate plot for aerosol formulation clustered data

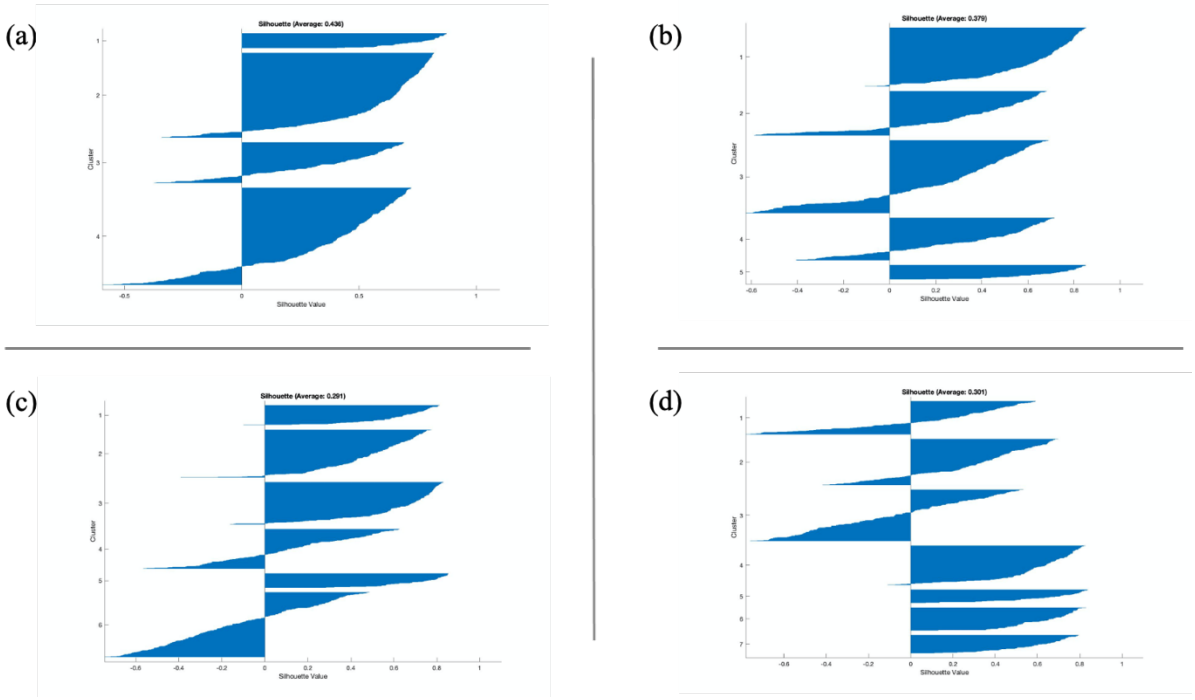


Figure S-6. n cluster silhouette plots for aerosol formulation indicators ($n \in \{4, 5, 6, 7\}$)

Clustering Algorithm Code

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
# Though the following import is not directly being used, it is required
# for 3D projection to work
from mpl_toolkits.mplot3d import Axes3D

from sklearn.cluster import KMeans

dataset = pd.read_excel('/Users/jichenxi/Desktop/research
everyday/data/unsupervise/flam.xlsx')
dataset

X = dataset.iloc[:, [1,2,3]].values

from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X = sc.fit_transform(X)

from sklearn.decomposition import KernelPCA
kpca = KernelPCA(n_components = 2, kernel = 'rbf')
X_new = kpca.fit_transform(X)
a = kpca.alphas_

from sklearn.metrics import silhouette_samples, silhouette_score
import matplotlib.cm as cm

for n_clusters in range(2, 8):
    kmeans = KMeans(n_clusters=n_clusters, random_state=10)
    y_kmeans = kmeans.fit_predict(X_new)

    fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(11,4), dpi=500)
    # lie within [-0.1, 1]
    ax1.set_xlim([-0.1, 1])
    # The (n_clusters+1)*10 is for inserting blank space between silhouette
    # plots of individual clusters, to demarcate them clearly.
    ax1.set_ylim([0, len(X) + (n_clusters + 1) * 10])

    silhouette_avg = silhouette_score(X, y_kmeans)
    print("For n_clusters =", n_clusters,
```

```
      "The average silhouette_score is :", silhouette_avg)
# Compute the silhouette scores for each sample
sample_silhouette_values = silhouette_samples(X, y_kmeans)

y_lower = 10
for i in range(n_clusters):
    # Aggregate the silhouette scores for samples belonging to
    # cluster i, and sort them
    ith_cluster_silhouette_values = \
        sample_silhouette_values[y_kmeans == i]

    ith_cluster_silhouette_values.sort()

    size_cluster_i = ith_cluster_silhouette_values.shape[0]
    y_upper = y_lower + size_cluster_i

    color = cm.nipy_spectral(float(i) / n_clusters)
    ax1.fill_betweenx(np.arange(y_lower, y_upper),
                      0, ith_cluster_silhouette_values,
                      facecolor=color, edgecolor=color, alpha=0.7)

    # Label the silhouette plots with their cluster numbers at the middle
    ax1.text(-0.05, y_lower + 0.5 * size_cluster_i, str(i))

    # Compute the new y_lower for next plot
    y_lower = y_upper + 10  # 10 for the 0 samples

ax1.set_title("The silhouette plot for the various clusters.")
ax1.set_xlabel("The silhouette coefficient values")
ax1.set_ylabel("Cluster label")

# The vertical line for average silhouette score of all the values
ax1.axvline(x=silhouette_avg, color="red", linestyle="--")

ax1.set_yticks([])  # Clear the yaxis labels / ticks
ax1.set_xticks([-0.1, 0, 0.2, 0.4, 0.6, 0.8, 1])

# 2nd Plot showing the actual clusters formed
colors = cm.nipy_spectral(y_kmeans.astype(float) / n_clusters)
ax2.scatter(X[:, 0], X[:, 1], marker='.', s=30, lw=0, alpha=0.7,
            c=colors, edgecolor='k')

# Labeling the clusters
centers = kmeans.cluster_centers_
```

```
    # Draw white circles at cluster centers
    ax2.scatter(centers[:, 0], centers[:, 1], marker='o',
                c="white", alpha=1, s=200, edgecolor='k')

    for i, c in enumerate(centers):
        ax2.scatter(c[0], c[1], marker='$%d$' % i, alpha=1,
                    s=50, edgecolor='k')

    ax2.set_title("The visualization of the clustered data.")
    ax2.set_xlabel("Feature space for the 1st feature")
    ax2.set_ylabel("Feature space for the 2nd feature")

    plt.suptitle(("Silhouette analysis for KMeans clustering on sample data "
                  "with n_clusters = %d" % n_clusters),fontsize=14, fontweight='bold')

plt.show()

kmeans = KMeans(n_clusters = 5, init = 'k-means++')
y_kmeans = kmeans.fit_predict(X)

fig = plt.figure(figsize=(12,8))
ax = Axes3D(fig, rect=[0, 0, 1, 1], elev=48, azim=134)
plt.scatter(X[y_kmeans == 3, 0],X[y_kmeans == 3, 1], X[y_kmeans == 3, 2], c = 'black',
label = 'Cluster 1')
plt.scatter(X[y_kmeans == 4, 0],X[y_kmeans == 4, 1], X[y_kmeans == 4, 2], c = 'red',
label = 'Cluster 2')
plt.scatter(X[y_kmeans == 0, 0],X[y_kmeans == 0, 1], X[y_kmeans == 0, 2], c = 'cyan',
label = 'Cluster 3')
plt.scatter(X[y_kmeans == 1, 0],X[y_kmeans == 1, 1], X[y_kmeans == 1, 2], c = 'green',
label = 'Cluster 4')
#plt.scatter(X[y_kmeans == 5, 0], X[y_kmeans == 5, 1], s = 20, c = 'magenta', label =
'Cluster 5')
plt.scatter(X[y_kmeans == 2, 0],X[y_kmeans == 2, 1], X[y_kmeans == 2, 2], c = 'blue',
label = 'Cluster 5')
#plt.scatter(X[y_kmeans == 6, 0], X[y_kmeans == 6, 1], s = 20, c = 'gray', label =
'Cluster 7')
#plt.scatter(X[y_kmeans == 7, 0], X[y_kmeans == 7, 1], s = 20, c = 'orange', label =
'Cluster 8')
# plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 300, c =
'yellow', label = 'Centroids')

plt.legend(loc='upper right', fontsize = 16)
plt.show()
```

```
dataset.loc[dataset.y_kmeans==0,'cluster']=3
dataset.loc[dataset.y_kmeans==3,'cluster']=1
dataset.loc[dataset.y_kmeans==4,'cluster']=2
dataset.loc[dataset.y_kmeans==2,'cluster']=5
dataset.loc[dataset.y_kmeans==1,'cluster']=4
dataset
```