UNIVERSITY OF BIRMINGHAM

Artificial Intelligence and Chemical Kinetics Enabled Property-Oriented Fuel Design for Internal Combustion Engine

by

Runzhao Li

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Mechanical Engineering

School of Engineering

College of Engineering and Physical Science

University of Birmingham

June 2021

UNIVERSITYOF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

Fuel Genome Project aims at addressing the forward problem of fuel property prediction and the inverse problems of molecule design, retrosynthesis and reaction condition prediction. This work primarily addresses the forward problem by integrating feature engineering theory, artificial intelligence (AI) technologies, gas-phase chemical kinetics. Group contribution method (GCM) is utilized to establish the GCM-UOB (University of Birmingham) 1.0 system with 22 molecular descriptors and the surrogate formulation is to minimize the difference of functional group fragments between target fuel and surrogate. The improved QSPR (quantitative structure–activity relationship)-UOB 2.0 system with 32 molecular features couples with machine learning (ML) algorithms to establish the regression models for fuel ignition quality prediction. QSPR-UOB 3.0 scheme expands to 42 molecular descriptors to improve the molecular resolution of aromatics and specific fuel types. The obtained structural features combining with ML algorithms enable to predict 15 physicochemical properties with high fidelity and efficiency. In addition to the technical route of ML-QSPR models, another route of deep learning-convolution neural network (DL-CNN) is proposed for property prediction and yield sooting index (YSI) is taken as a case study. The predicted accuracy of DL-CNN is inferior to the ML-QSPR model at its current status, but its benefit of automated feature extraction and rapid advance in classification problem make it a promising solution for regression problem. A high-throughput fuel screening is performed to identify the molecules with desired properties for both spark ignition (SI) and compression ignition (CI) engines which contains the Tier 1 physicochemical properties screening (based on the ML-QSPR models) and Tier 2 chemical kinetic screening (based on the detailed chemical mechanisms). Polyoxymethylene dimethyl ether 3 (PODE3) and diethoxymethane (DEM) are promising carbon-neutral fuels for CI engines and they are recommended by the virtual screening results. Their ignition delay time, laminar flame speed and dominant reactions of PODE3 and DEM are examined by chemical kinetics and a new DEM mechanism including both low and high-temperature reactions is constructed. Concluding remarks and research prospects are summarized in the final section.

Acknowledgments

Q: Why you start a Ph.D. study at the University of Birmingham? Do you achieve the goal?

A: When I start my Ph.D. study at the University of Birmingham, I have a question of "What is the prospect of the fuel combustion science?" After two and a half years of study, my answer is "cheminformatics" and "carbon-neutral fuels" which will become popular topics in the fuel combustion community in the upcoming decade. Cheminformatics refers to use computational methods and information science techniques to solve problems in chemistry. It copes with the challenges of molecular feature extraction and representation, data scarcity. Carbon-neutral fuels denote the fuels synthesized from renewable feedstocks and produced by sustainable energy which have zero net GHG emissions and carbon footprint. So, I have met my academic expectation and enjoy the wonderful journey.

Q: What do you think your biggest reward when studying at UOB?

A: A belief of "the foot is longer than the road". Regardless of the Ph.D. research program and the thesis writing, they are lengthy and challenge processes. But these goals will always achieve by devoting time, patience and commitment. This principle applies to our careers and lives beyond academic research, we can always find more solutions than problems. So, please keep moving forward and be positive in our lives.

Q: How about your academic research in UOB?

A: I conduct my Ph.D. study under the supervision of Prof. Athanasios Tsolakis and Dr. Jose Herreros. They give me the full freedom to explore the interesting area, topics and put them into practice. And I appreciate their guidance on my research direction and scheme construction. I must say a big thank you to both who provide the creative environment in our group and support my research interest in the application of AI in fuel combustion. I also need to acknowledge the school scholarship awarded by UOB.

Q: How do you get along with your colleagues?

A: Dr. Omid Doustdar, Sak Sittichompoo, Felipe Torres make a great influence on my Ph.D. study. Omid is a

research fellow in our group and he does provide great supports to our students. I appreciate Dr. Omid Doustdar for his guidance and assistance on my engine test. Sak Sittichompoo is an expert in engine experiments and we have many insightful discussions during my engine test preparation. I learned a lot of knowledge on engine tests and DOC light-off when helping Felipe's test, I appreciate that valuable opportunity. Feliple also guides our peers on how to conduct the research efficiently and generously share his resource and experience, his academic commitment makes me impressed and respectful.

Q: How the COVID-19 affect your research?

A: To be honest, a lot. Before COVID-19, I have designed a comprehensive engine test plan for fuel design. But the COVID-19 makes us have to work from home and away from the lab for a long time. After discussing with supervisors, I decided to transfer a great deal of research work from experiment to modeling. Occasionally, I find that the increasing computational power and advanced AI technology can implement many applications in the field of fuel combustion science. This is a new research topic and area with some challenges for myself but I am sure I interest in this field. The transition from massive engine tests to AI-aided fuel research turn adversity into opportunity and lead me into the inter-discipline of AI and fuel combustion chemistry. To some extent, COVID-19 change my research direction indirectly and without this unnormal circumstance, I probably won't step into the AI field.

Q: Have you participate in some academic conferences or webinars? Which of them make you impressive and why?

A: I participate in many academic exchangers during my Ph.D. study regardless of a presenter or an audience, I would like to say thank you for those opportunities. I will mention some of them since they benefit me a lot in the long term. The first one is the combustion webinar organized by The Combustion Institute. This weekly webinar presents much cutting-edge research covering fundamental researches and applied research in the combustion community. The launch of the online combustion webinar is caused by COVID-19, and it provides a

valuable opportunity for the students to learn from the high-level presenters. I think this is another case to show how our researchers covert the challenge into an opportunity. The second one is the MathWorks online events, I learn many AI knowledge from them and they showcase how to apply AI in the engineering domain. The third one is the conference "Fuel Science-From Production to Propulsion" organized by RWTH Aachen. I participate in the 8th conference and will join the 9th conference in June 2021. This conference provides many frontier research from fuel production to fuel combustion. It is an excellent window to help the students to know the recent advance in fuel research.

Q: Do you want to say something to your friends?

A: Studying abroad to pursue a Ph.D., is always a challenge, stressful process and daily life sometimes will encounter loneliness and missing the lifestyle, friendships in my home country China. I have regular contact with Wenzhao Yang, Lei Zhang, Jiayao Yan, Kunpeng Ma, Weiwei Li, Wei Chen et. al. in a wide range of topics in academia and beyond. These friendships accompany my 2.5 years of study in UOB, I hope they can receive my thanks.

Q: How about your parents?

A: I am most grateful to my parent because of their financial, spiritual supports and the freedom they provide me to decide my life. Many of my friends said that it is unwise to pursue a second Ph.D. from the perspective of time and finance. I agree and understand their opinion but I am also very clear about the reason why I decided to start my 2nd Ph.D., study. I want to learn more about fuel combustion science in the world-leading institute, work with creative peoples and I am sure this experience will benefit my whole life. I talk to my parents and they decide to support my dream and pursuit. I appreciate the warm support and understanding from my parents so I will accompany them for the rest of my life.

Author Publications

[1]   **Li R**, Herreros JM, Tsolakis A, Yang W. Integrated machine learning-quantitative structure property relationship (ML-QSPR) and chemical kinetics for high throughput fuel screening toward internal combustion engine. Fuel 2022;307.

[2]   **Li R**, Herreros JM, Tsolakis A, Yang W. Machine learning-quantitative structure property relationship (ML-QSPR) method for fuel physicochemical properties prediction of multiple fuel types. Fuel 2021;304.

[3]   **Li R**, Herreros JM, Tsolakis A, Yang W. Chemical kinetic modeling of diethoxymethane oxidation: A carbon–neutral fuel. Fuel 2021;291.

[4]   **Li R**, Herreros JM, Tsolakis A, Yang W. Machine learning regression based group contribution method for cetane and octane numbers prediction of pure fuel compounds and mixtures. Fuel 2020;280.

[5]   **Li R**, Herreros JM, Tsolakis A, Yang W. Chemical kinetic study on ignition and flame characteristic of polyoxymethylene dimethyl ether 3 (PODE3). Fuel 2020;279.

[6]   **Li R**, Herreros JM, Tsolakis A, Yang W. Novel Functional Group Contribution Method for Surrogate Formulation with Accurate Fuel Compositions. Energy & Fuels 2020;34(3):2989-3012.

[7]   Machine learning and deep learning enabled fuel sooting tendency prediction from molecular structure (submitted to journal)

Table of Contents

List of Figures

List of Tables

Nomenclature

Symbols

A	pre-exponential factor
C_p	constant pressure specific heat capacity
D_f	degree of freedom
E_a	apparent activation energy
$fgn_{i,j}$	the i th functional group number of the j th surrogate fuel component
F_i	the normalized difference of the i th functional group
$FGN_{i,targ}$	total number of the i th functional group in the target fuel molecules
$FGN_{i,calc}$	total number of the i th functional group in the surrogate fuel molecules
ΔH_{vap}	enthalpy of vaporization
$\Delta H_{vap,\,norm}$	normalized enthalpy of vaporization
κ	a constant depending on the pressure and temperature evolution in unburned gas or global reaction rate of one step mechanism
m	pressure exponent
n	equivalence ratio exponent or the order of chemical reaction
N_g	number of surrogate functional groups
n_j	molar quantity of the j th surrogate fuel component
n_F	number of filters
N_f	number of the surrogate palettes
P_{init}	initial intake air pressure or mixture initial pressure
p	mixture initial pressure (used in equations)
POSF	fuel designation, not an acronym
R_u	universal gas constant
R	number of response
R^2	R squared, coefficient of determination
S	objective function
S_{ig}	statistical significance
S_L	premixed laminar flame speed
$S_{L\,norm}$	normalized premixed laminar flame speed
T	mixture initial temperature (used in equations)
T_0	unburned gas temperature
T10	10vol.% recovered temperature
T50	50vol.% recovered temperature
T90	90vol.% recovered temperature
T_b	boiling point

$T_{b,norm}$	normalized boiling point
T_{init}	initial intake air temperature or mixture initial temperature
T_i	combustible ignition temperature
T_f	flame temperature
T_m	melting point
$T_{m,norm}$	normalized melting point
$T_{reactor}$	reactor temperature
t_i	target output
W_i	weighting factor of the i th functional group
y_i	network's prediction for response i

Greek letters

α	thermal diffusivity
λ	thermal conductivity
δ_r	length of reaction zone
ε	reaction progress variable
φ	equivalence ratio
τ_{ig}	ignition delay time
$\tau_{resident}$	residence time
γ	surface tension
γ_{norm}	normalized surface tension
ν	kinematic viscosity
μ	dynamic viscosity
ρ	liquid density
η	φ-sensitivity
η_{norm}	normalized φ-sensitivity
η_{brake}	brake thermal efficiency
$\eta_{combustion}$	combustion efficiency
$\eta_{combustion}$	(cycle) thermal efficiency
$\eta_{gas\ exchange}$	gas exchange efficiency
$\eta_{mechanical}$	mechanical efficiency

Abbreviations

ABFIS	adaptive network-based fuzzy inference system
AFT	adiabatic flame temperature
AIChE	American Institute of Chemical Engineers
ANFIS	adaptive neuro-fuzzy inference system
ANN	artificial neural networks
API	American Petroleum Institute
ASM	active subspace method
ASTM	American Society for Testing and Materials
ATDC	after top dead center
atm	standard atmosphere, pressure unit, 1atm equal to 101.325kPa
bar	pressure unit, 1bar equal to 100kPa
BECCS	bioenergy with carbon capture and storage
BTDC	before top dead center
2-BTHF	2-butyltetrahydrofuran
BP-NN	back-propagation neural network
CaL	calcium carbonate looping
CAMD	computer-aided molecular design
CAS	chemical abstract service
CCDB	carbon-carbon double bond
CCS	carbon capture and storage
CCTB	carbon-carbon triple bond
CFD	computational fluid dynamics
CFR	cooperative fuels research
CI	compression ignition
CLC	chemical looping combustion
CN	cetane number
$CN_{blending}$	cetane number of a specific fuel compound when it is blended with a base fuel in particular volume fractions
CNN	convolution neural network
CODESSA	comprehensive descriptors for structural and statistical analysis
Co-Optima	Co-Optimization of Fuels & Engines
CPU	central processing unit
CRC	Chemical Rubber Company
CUDA	compute unified device architecture
CVCC	constant volume combustion chamber
CVV	constant volume vessel
DAC	direct air capture
DCN	derived cetane number
DEM	diethoxymethane
DHA	detailed hydrocarbon analysis
DIPE	diisopropylether
DIPPR	Design Institute for Physical Properties
DL	deep learning

DMM	dimethoxymethane
EOR	CO_2-enhanced oil recovery
ETBE	ethyl tert-butylether
FACE	fuels for advanced combustion engines
FAME	fatty acid methyl ester
FFNN	feed-forward neural network
FIT	fuel ignition tester
FP	flash point
FSC	Fuel Science Center
FTIR	Fourier-transform infrared spectroscopy
GA	genetic algorithm
GC×GC-FID	two-dimensional gas chromatography with flame ionization detection
GC×GC-TOFMS	two-dimensional gas chromatography with time-of-flight mass spectrometry
GCI	gasoline (like fuels) compression ignition
GCM	group contribution method
GC-MS	gas chromatography-mass spectrometry
GCR	group chemistry representative
GCVOL	group contribution method for predicting saturated liquid density
GHG	green house gas
GPR	Gaussian process regression
GPU	graphic processing unit
HCCI	homogeneous charge compression ignition
HRR	heat release rate
HTHR	high-temperature heat release
IC	internal combustion
IDT	ignition delay time
ILSVRC	ImageNet Large-Scale Visual Recognition Challenge
IMEP	indicated mean effective pressure
IQT	ignition quality tester
ISFC	indicated specific fuel consumption
IT	ignition temperature
ITE	indicated thermal efficiency
IUPAC	International Union of Pure and Applied Chemistry
JSR	jet-stirred reactor
KAUST	King Abdullah University of Science and Technology
LANL	Los Alamos National Laboratory
LC	liquid chromatography
LFL	lower flammability limit
LHV	lower heating value
LHV_{norm}	normalized lower heating value
LLNL	Lawrence Livermore National Laboratory
LRM	linear regression model
LTC	low-temperature combustion
LTHR	low-temperature heat release

MAE	mean absolute error
MCCI	mixing controlled compression ignition
MD	molecular dynamics
MDM	molecular dynamics method
MFB	mass fraction burn
ML	machine learning
MLP	multi-layer perceptron
MLR	multiple linear regression
MPI	micropyrolysis index
MSE	mean-squared-error
MON	motor octane number
MTBE	methyl tert-butylether
MON	motor octane number
N/A	not applicable
NIST	National Institute of Standards and Technology
NKRDPC	national key research and development program of china
NMR	nuclear magnetic resonance spectroscopy
NN	neural network
NREL	National Renewable Energy Laboratory
NTC	negative temperature coefficient
OESI	oxygen extended sooting index
OI	octane index
OS	octane sensitivity
OS_{norm}	normalized octane sensitivity
P2X	Power-to-X
PAH	polycyclic aromatic hydrocarbon
PCR	principle component regression
PFR	plug flow reactor
PIONA	paraffins, isoparaffins, olefins, naphthenes, aromatics
PLS	partial least squares
PM	particulate matter
PODE	polyoxymethylene dimethyl ether
PPCI	partially premixed compression ignition
PSO	particle swarm optimization
QSPR	quantitative structure-property relationship
RAS	Russian Academy of Sciences
RBN	radial basis function neural network
RCCI	reactivity controlled compression ignition
RCM	rapid compression machine
ReaxFF	reactive force field
ReLU	rectified linear unit
RME	rapeseed methyl ester
RMSE	root mean square error
RON	research octane number

RON$_{norm}$	normalized research octane number
SCU	Sichuan University
SDSeries38	standard series network with 38 layers
SFN	Solar Fuels Network
SGDM	stochastic gradient descent with momentum
SI	spark ignition
SME	soy methyl ester
SP	smoke point
ST	shock tube
SUPERTRAPP	NIST Thermophysical Properties of Hydrocarbon Mixtures Database
SVM	support vector machines
SwRI	Southwest Research Institute
TAME	tert-amylmethylether
THC	total hydrocarbon
TI	topological indices
TMFB	Tailor-Made Fuels from Biomass
Torr	pressure unit, 1Torr approximately equal to 133.32Pa
TPRF	toluene primary reference fuels (n-heptane-iso-octane-toluene mixture)
TSI	threshold sooting index
UCL	Université catholique de Louvain
UFL	upper flammability limit
UOB	University of Birmingham
USC	University of South Carolina
VP	vapor pressure
YSI	yield sooting index
YSI$_{norm}$	normalized yield sooting index

Chapter 1 Introduction

1.1 Technical background of fuel design and artificial intelligence (AI) application

The engine-fuel interaction is through the combustion process that the combustion connects the engine hardware and fuel compositions. The essence of fuel design is to formulate the fuel compositions to address required properties for specific combustion modes, so it is called "property-oriented fuel design" in this work. The ultimate target of fuel design is to maximize the engine thermal efficiency, increase renewable energy utilization and reduce pollutant emissions. Research efforts are devoted to discovering viable high-performance fuel candidates paired with various combustion modes. The combustion modes for CI engines are mainly classified by the levels of in-cylinder fuel stratification at the combustion initiation [1-3] which order from homogeneous charge to full fuel stratification as HCCI, PPCI, MCCI. Either low and high reactivity fuels can be applied in these combustion modes as shown in Figure 1.1. Another innovative combustion mode of RCCI adopts both low and high reactivity fuels to construct in-cylinder fuel reactivity gradient, φ gradient, temperature gradient while other combustion modes contain only the latter two gradients [4]. Therefore, the property-oriented fuel design should perform toward a specific combustion mode.

The experiment-based property-oriented fuel design method has serious limitations: (1) Facilities should be available to measure the target properties. (2) Enormous compounds without experimental data need to be screened to identify the suitable molecules with desired properties. This process is time-consuming, expensive and even unrealistic for those emerging fuels. One way to address these problems is applying AI technologies to accelerate the property-oriented fuel design by electrically handling the huge amount of data. Therefore, the chemical structures need to be transformed into mathematical representations that can be processed and manipulated by the computer [5]. Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) composes of nucleotides polymers (strand) and nitrogenous bases pairs which can be expressed as single nucleotide polymorphisms (SNP) as shown

in Figure 1.2 (a). The polylactic acid is composed of repeated monomers forming a long chain which can be represented by hierarchical levels of fingerprints (property-based features, fragment-based features, atomic position-based features) as shown in Figure 1.2 (b). Given that the fuel molecule is composed of a set of atoms and bonds, a molecule can be decomposed into a series of atomic fragments and encoded as molecular descriptors as shown in Figure 1.2.

Artificial neural network (ANN) is a branch of AI technologies and its application in chemical engineering is reviewed by Panerati et al. [6] as shown in Figure 1.3. ANN is widely adopted in the oil and energy sector and it is mainly used for modeling, prediction, control and optimization. New application scenarios should be created to promote AI application to the oil and energy sector. The data volume for ANN modeling in chemical engineering mainly has the order of magnitude of 10^2 which indicates that ANN rarely applies in streaming data type and the more training data is needed to improve the model robustness, reliability, extrapolation ability.
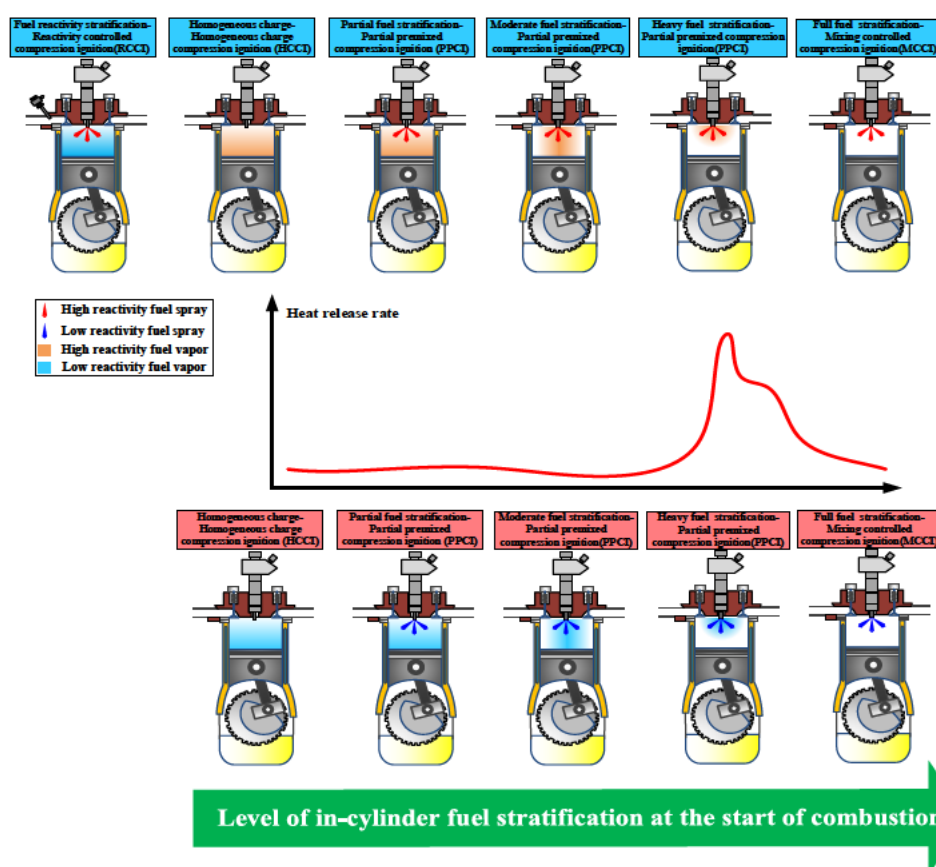
Figure 1.1. Traditional combustion mode and advanced combustion modes for compression ignition engines, strategies are positioned according to the level of fuel stratification at the start of combustion relative to conventional diesel combustion, which is highly stratified.
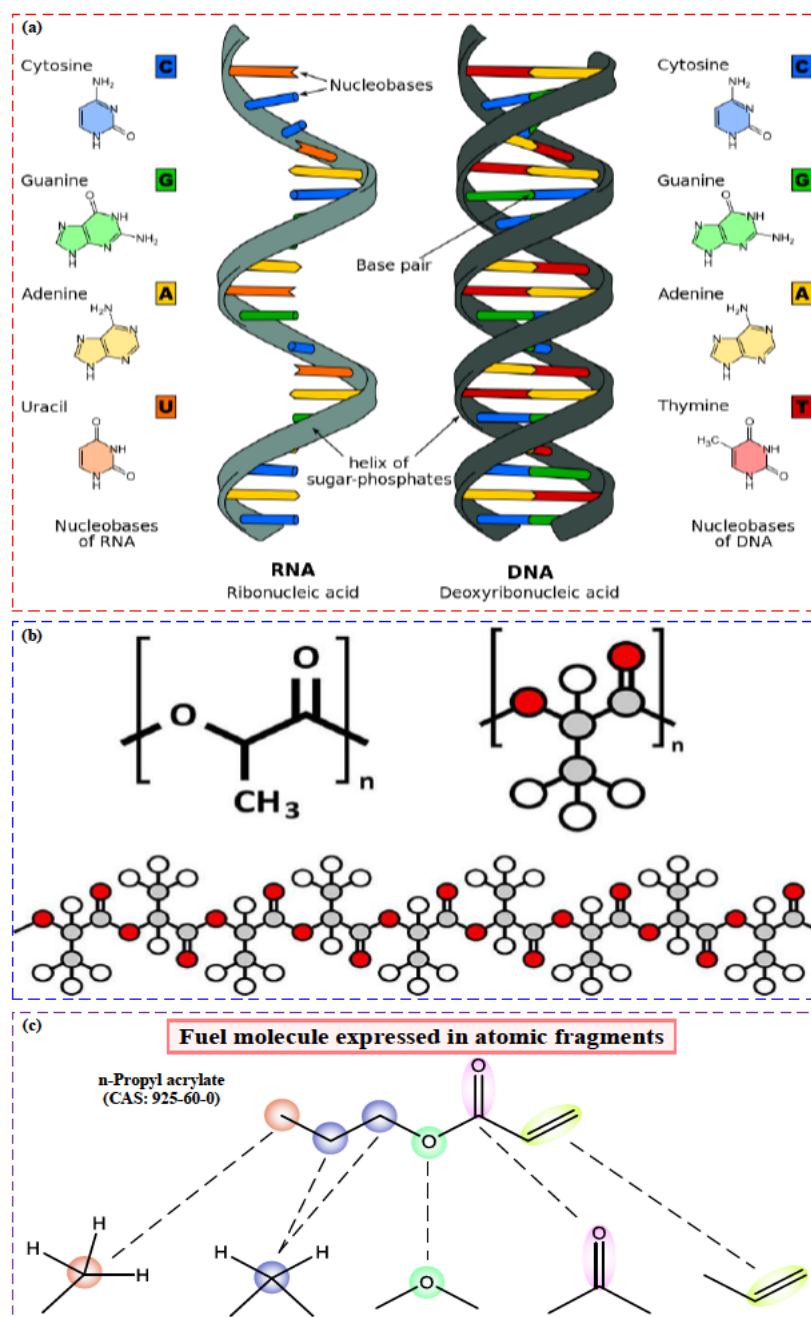
Figure 1.2. (a) Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) compose of nucleotides polymers (strand) and nitrogenous bases pairs. Reproduced from ref. [7]; (b) a polymer(polylactic acid) makes up of repeated monomers. Reproduced from ref. [8];. (c) fuel fingerprint expressed in terms of typical atomic fragments.
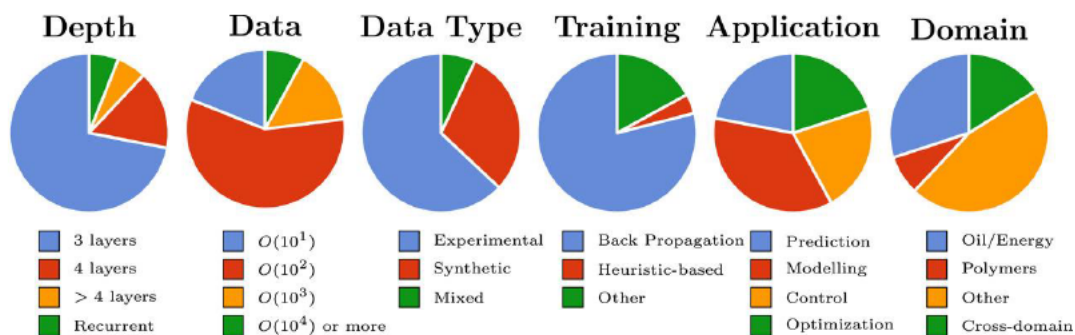
Figure 1.3. Summary of artificial neural networks (ANNs) application in chemical engineering: the network depth, orders of magnitude of data, data type, network organization, task, application domain. Reproduced from ref. [6].

1.2 Historical evolution review of artificial intelligence (AI) application: From Human Genome Project to emerging material discovery

1.2.1 Human Genome Project

The Human Genome Project (HGP) coordinates by the U.S. National Institutes of Health (NIH), the National Academy of Sciences and the Department of Energy (DOE) to decipher and sequence all the genes of human beings at 1988~2003 [9]. The timeline of Human Genome Project is shown in Figure 1.4 and it provides a foundation for enormous scientific projects regarding the relationship between human genes and disease Deoxyribonucleic acid (DNA) molecule carries the organism genetic information which contains four types of bases including adenine (A), cytosine (C), guanine (G), thymine (T). The assembling of the organism building blocks is instructed by the order of the four types of bases. A gene is a sequence of nucleotides in DNA coding the synthesis instruction and a genome denotes a complete set of genetic genes that contains all the synthesis instructions. Human beings have 3 billion bases and 99.9% of these bases are identical for all peoples. The remaining 0.1% of 3 billion bases vary from person to person. A single-nucleotide polymorphism (SNP) denotes a DNA sequence variation in the genome as shown in Figure 1.5 (a). The DNA molecule cannot directly sequence without breakage because the bases identification reactions can only read DNA stand containing less than 1000 bases. The whole genome sequencing is composed of 5 steps as shown in Figure 1.5 (b): (1) DNA sample collection. (2) Massive replication of the DNA sample. (3) Breaking the massive DNA samples into tremendous small, overlapping segments. (4) Sequencing the order of bases by implementing a series of chemical reactions. (5) Assembling the complete set of genes. Genome sequencing facilitates to identify the association between human genes with diseases and these types of studies are called genome-wide association study (GWAS) as shown in Figure 1.5 (c) [10]. To evaluate the impact of SNPs on a given disease, the members are grouped into cases group (patients with specific diseases) and control group (health individuals). The SNPs are sequenced for both case group and the obtained SNPs across the genome are compared and analyzed to identify the caused gene to many rare and severe diseases.
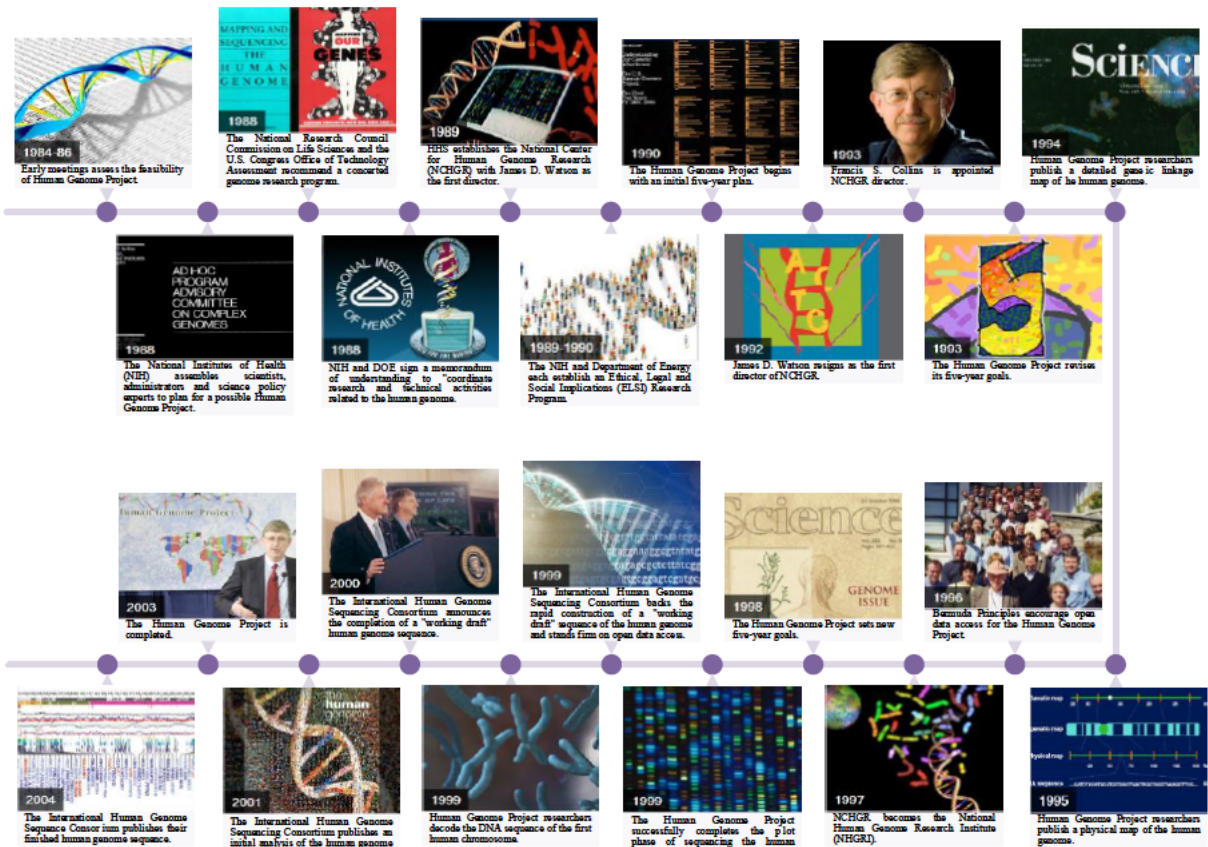
Figure 1.4. Human Genome Project Timeline. Reproduced from ref. [11].
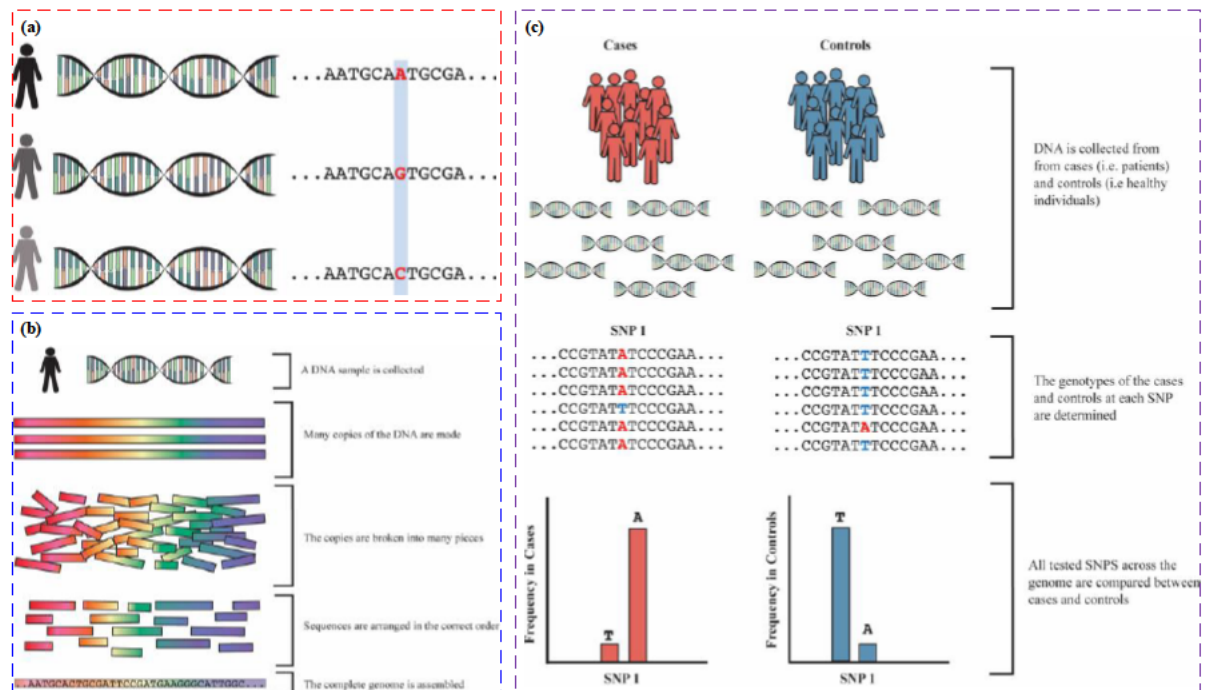
Figure 1.5. (a) Single nucleotide polymorphisms (SNP) of three different people which has a different base at the same spot in the genome; (b) overview of genome shotgun sequencing; (c) overview of the genome-wide association study (GWAS). Reproduced from ref. [12].

1.2.2 Material Genome Initiative

In the biological fields, a genome is a full set of encoded information in DNA language that constitutes the nucleic acid sequences for human beings and acts as the genetic blueprint of an organism's growth. Outside the biological context, the term "genome" denotes a set of build blocks of a particular object [13]. Material Genome Initiative (GMI) aims at creating a material innovation ecosystem to accelerate the new material discovery, manufacture, and deployment in clean energy, human beings, national security by integrating computation techniques, experimental facilities and data informatic tools as shown in Figure 1.6 [14]. Yun Liu et al. [15] and Yue Liu et al [16] propose their understanding of the application of ML technology in material engineering science as shown in Figure 1.7 (a) and Figure 1.7 (b) respectively and their ideas are similar. Advanced computational tools of feature engineering and intelligent algorithms are adopted to accelerate the material property prediction by ML algorithms and discover material with desired properties by high throughput screening. The computational and data-driven material design and screening enable to replace the expensive, time-consuming and high-risk experiments for new material validation and certification [13]. The material genome reveals the intrinsic structure-property relationship of materials by complementary efforts of computation, informatic theory and experiment [17].
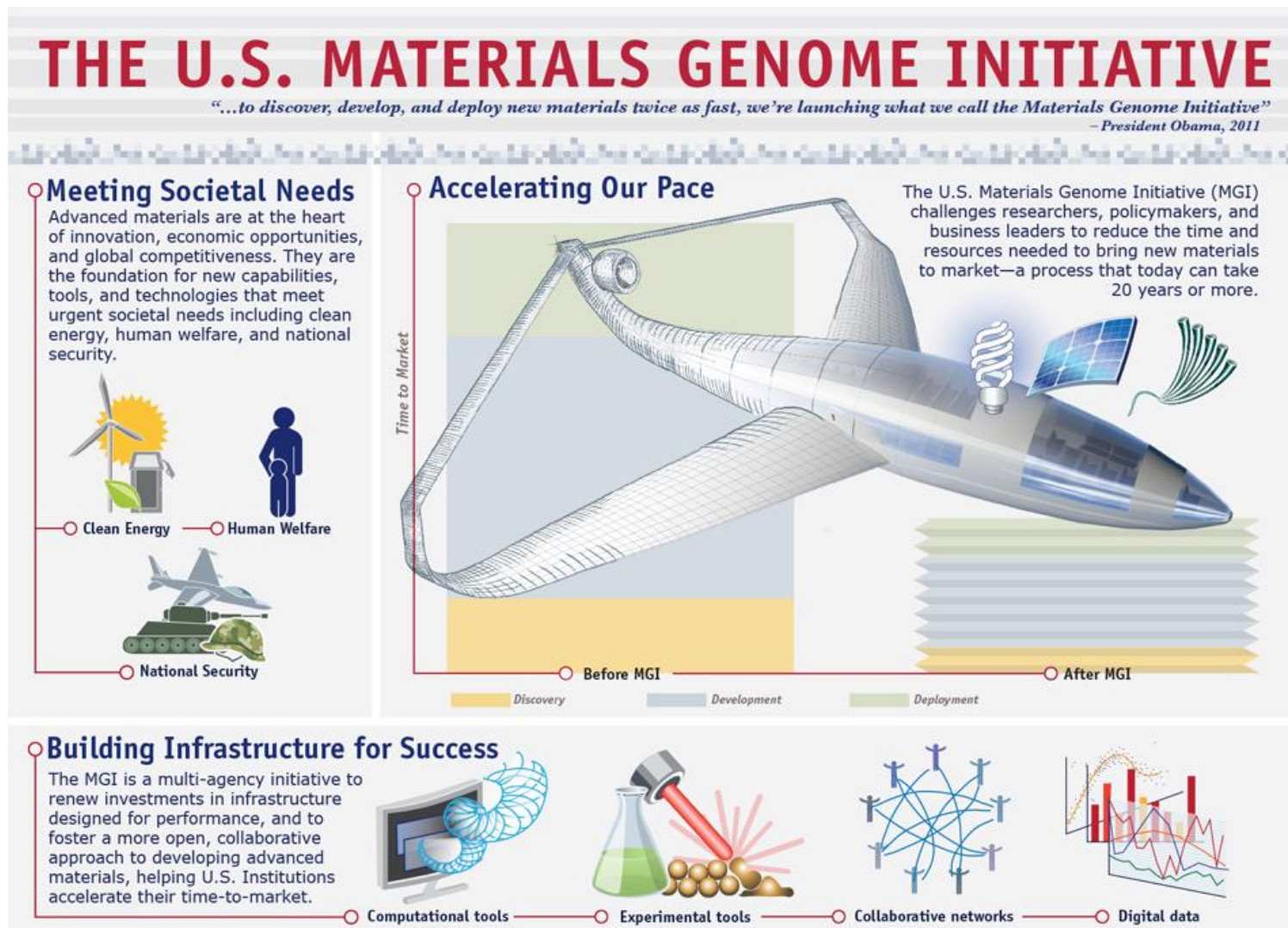
Figure 1.6. The goal of Materials Genome Initiative (MGI): accelerate the discovery, design, development and deployment of new materials for clean energy, national security and human beings by integrating computational techniques, experiment facilities and big data management. Reproduced from ref. [18, 19].
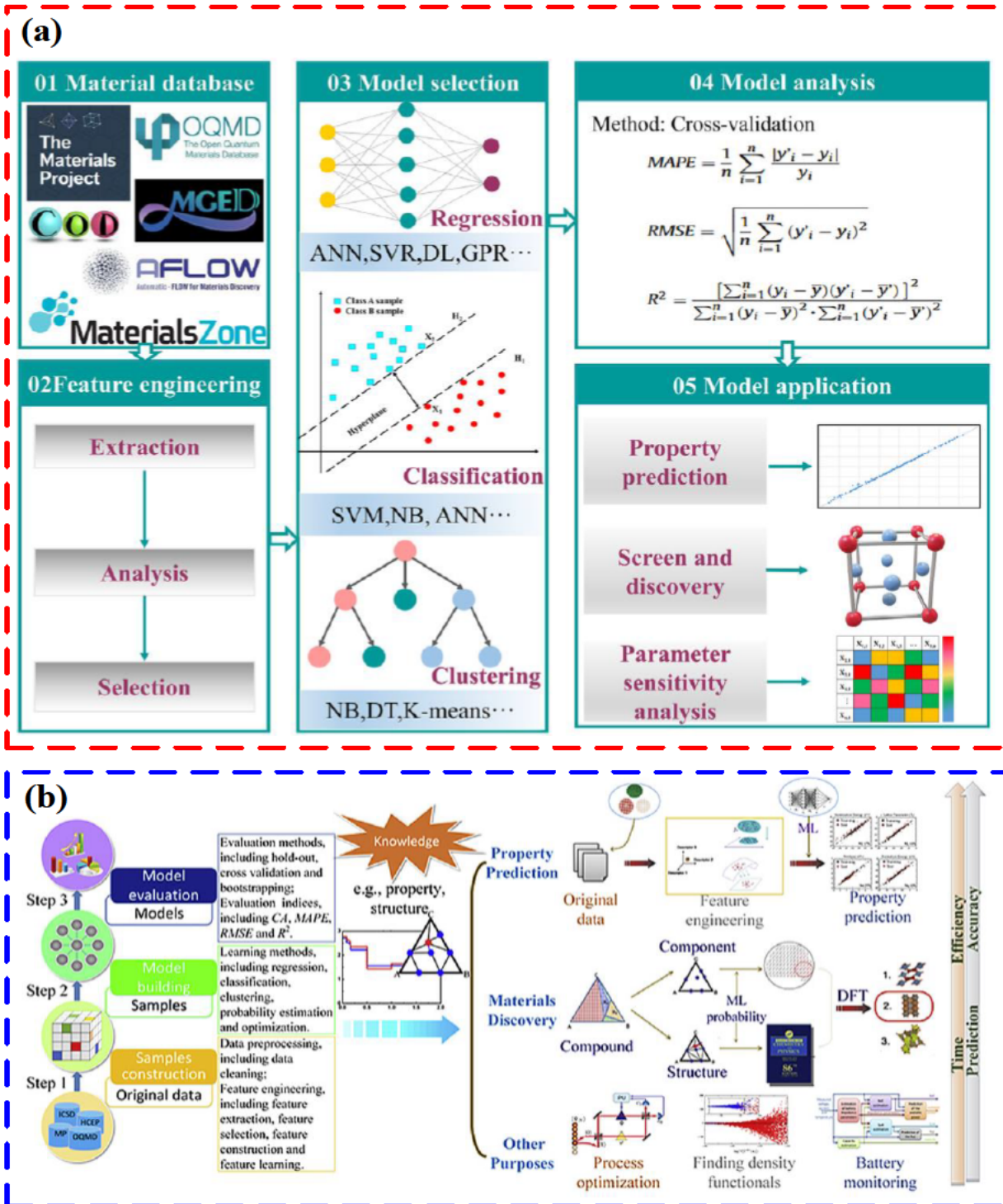
Figure 1.7. Application of AI technology in material engineering science for material property prediction, high-performance material discovery and other purposes. (a) The conceptual graph proposed by Yun Liu et al. [15]. (b) The conceptual graph proposed by Yue Liu et al [16].

1.2.3 Polymer Genome

Polymer Genome provides an informatic platform to predict the polymer properties, design the polymer structure with desired properties, perform retrosynthesis analysis by machine learning and density function theory [20-24]. Polymer Genome aims at addressing the forward problem of the polymers properties prediction and the inverse problem of new polymers generation with desired properties. The technology roadmap of Polymer Genome is shown in Figure 1.8. An online material properties repository is established to store the notable material properties and the database list can be found in ref. [21]. Polymer Genome follows the idea of material genome initiative to develop hierarchical fingerprints (including atomic level, block level and chain level) to represent the polymers in a numerical format making it manageable by the AI tool as shown in Figure 1.9 [25]. Polymer Genome applies ML algorithms and density functional theory (DFT) to predict the polymer electronic properties, dielectric & optical properties, thermal properties, physical & thermodynamic properties, mechanical properties, solubility properties, permeability properties as shown in Figure 1.10, the accessible properties are summarized in Table 1.1. Polymer Genome applies ML algorithms and genetic algorithm to generate polymer structure iteration to meet the performance objective as shown in Figure 1.11 and Figure 1.12 [26]. The recommended polymers are verified by computational and experimental tests. The validated polymers are registered to the established polymer properties repository.
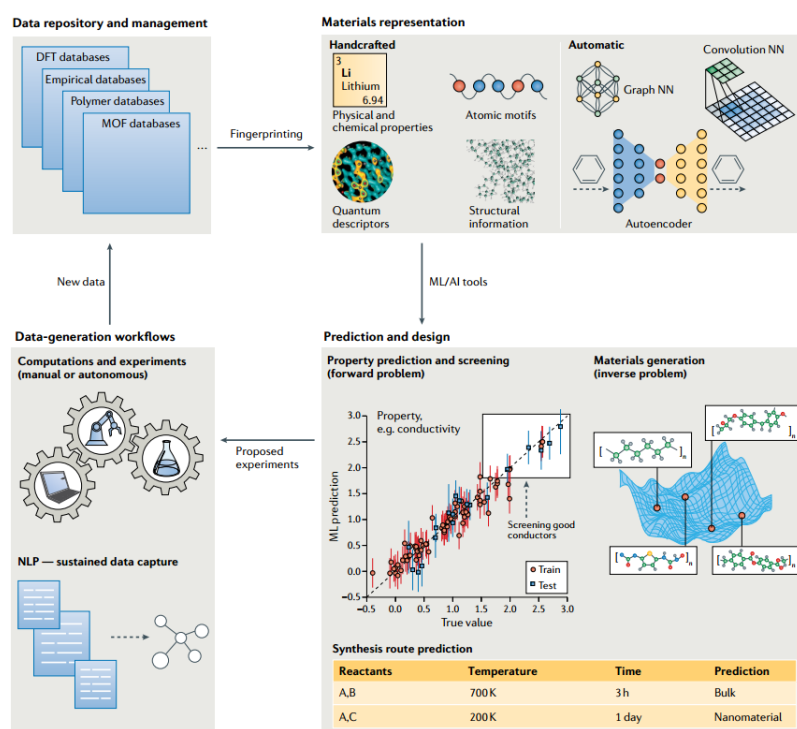
Figure 1.8. Overview of the Polymer Genome project. Reproduced from ref. [20, 21].
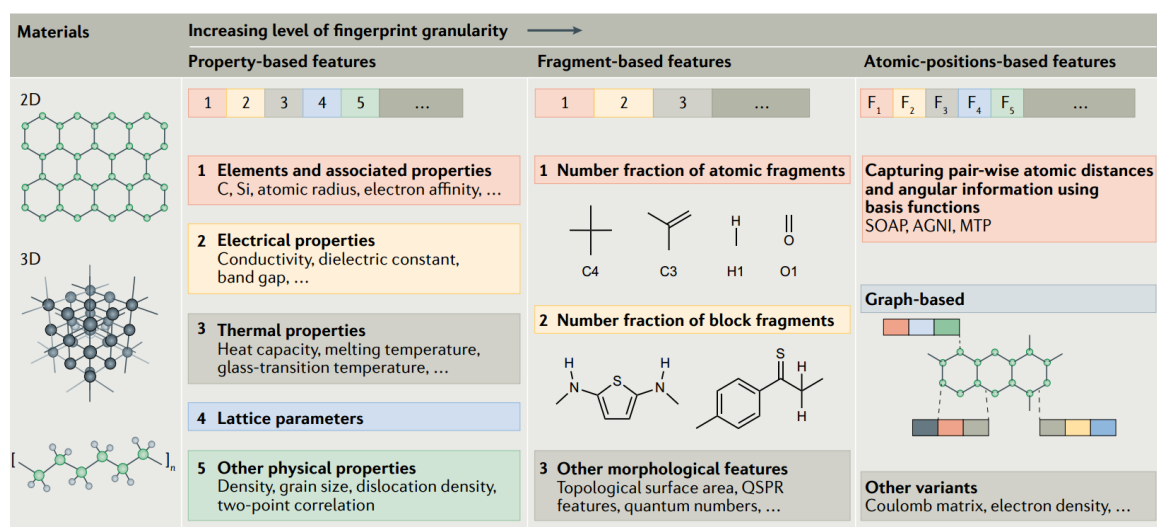
Figure 1.9. Hierarchical material fingerprints in the Polymer Genome project. Reproduced from ref. [21].
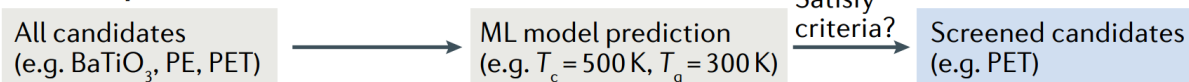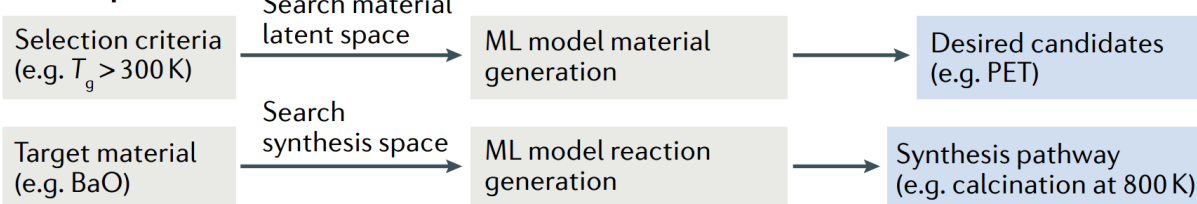
Figure 1.10. The forward and inverse problems of AI-powered material screening. Reproduced from ref. [21].

Figure 1.11. The forward problem of material property prediction by machine learning algorithms. Reproduced from ref. [21].
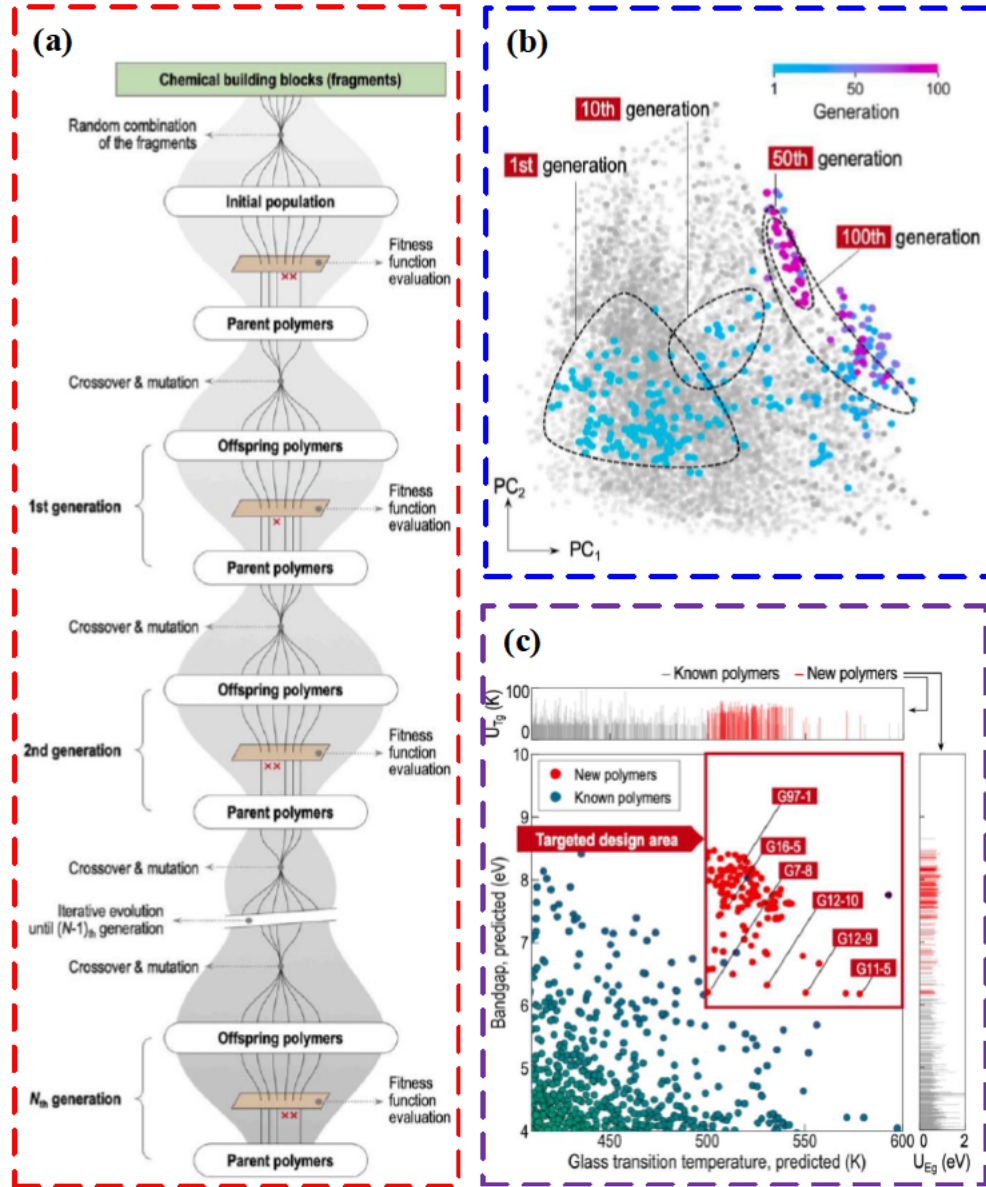
Figure 1.12. The inverse problem of material production with target properties. Reproduced from ref. [26].

Table 1.1 Machine learning-powered polymer property prediction in Polymer Genome platform [27]

Property Type	No.	Properties	Property Type	No.	Properties
Electronic	1	Bandgap (bulk)	Dielectric & Optical	15	Dielectric constant
	2	Bandgap (single chain)		16	Refractive index (bulk resin)
	3	Ionization energy		17	Refractive index (crystal)
	4	Electron affinity	Mechanical	18	Tensile Strength
	5	Electron injection barrier		19	Young's Modulus
	6	Electrical conductivity	Solubility	20	Solvent
Thermal	7	Glass transition temperature		21	Non-solvent
	8	Melting temperature		22	Hildebrand Solubility Parameter
	9	Thermal decomposition temperature	Permeability	23	Permeability (Barrer)
Physical & Thermodynamic	10	Density		24	Selectivity
	11	Atomization energy	Others	25	Tendency to Crystallize
	12	Specific heat		26	Limiting Oxygen Index
	13	Cohesive energy density			
	14	Fractional free volume			

1.3 Fuel Genome Project

1.3.1 Goals and technology roadmap: from property prediction to molecule discovery

Cheminformatics refers to the application of computational methods and information science techniques to solve problems in chemistry. The Fuel Genome Project provides a chemoinformatic platform to predict fuel property (forward problem) and discover/design advanced fuels with desired performance (inverse problem) as shown in Figure 1.13. Especially, the inverse problem comprises of three subproblems: (1) Design new molecules with demand properties. (2) Perform retrosynthesis to design reaction steps going backward from the target molecules to the commercially available feedstocks; (3) Determine appropriate reaction conditions (catalyst, temperature, pressure, reaction/resident time, solvent, reagent, purification method) for the identified reaction steps to improve the product yield. In the Fuel Genome Project, AI technologies are applied to chemistry to accelerate and boost the efficiency of the forward problem and inverse problem as shown in Figure 1.14. For the forward problem of fuel property prediction, QSPR theory is employed to extract the structural features and property features while AI technologies (especially machine learning and deep learning) are used to build a regression model to correlate both features. For the inverse problem of molecule design, the recurrent neural network (RNN)-driven encoder creates a continuous molecular representation (latent space) by converting molecules into vectors and the decoder translates a point in the latent space with desired properties into the promising molecule. For the inverse problem of retrosynthesis and reaction condition design, the AI-based retrosynthesis tool condenses the chemical rules to generate the intermediates, reaction networks and search the suitable reaction conditions.

The technological roadmap of the Fuel Genome Project contains 4 aspects as shown in Figure 1.15. The first aspect is to create a data repository of molecular structure features and a wide variety of fuel properties. The second aspect is feature extraction which includes two ways of manual feature extraction (see chapter 2, 3, 4) and automated feature extraction (see chapter 5). The third aspect is to develop and deploy the property prediction

model, molecular generation tool, retrosynthesis package. The fourth aspect is to employ the computational programs to accelerate the fuel property prediction, molecule design, retrosynthesis and combine with experimentally derived data to boost the data repository.

This thesis adopts the Quantitative structure-property relationship (QSPR) method to address the forward problem of fuel property prediction which correlates the relationship between molecular structure and the properties of interest. In other words, the property is functional of the molecular descriptors as shown in Eq. (1.1) [28, 29]. Once the dependence of fuel properties on the molecular structure is established, the QSPR approach enables the physicochemical properties prediction of new molecules. The molecular descriptors development is to elucidate is section 1.3.2.

$$\text{Macroscopic property} = f(\text{molecular descriptors}) \tag{1.1}$$
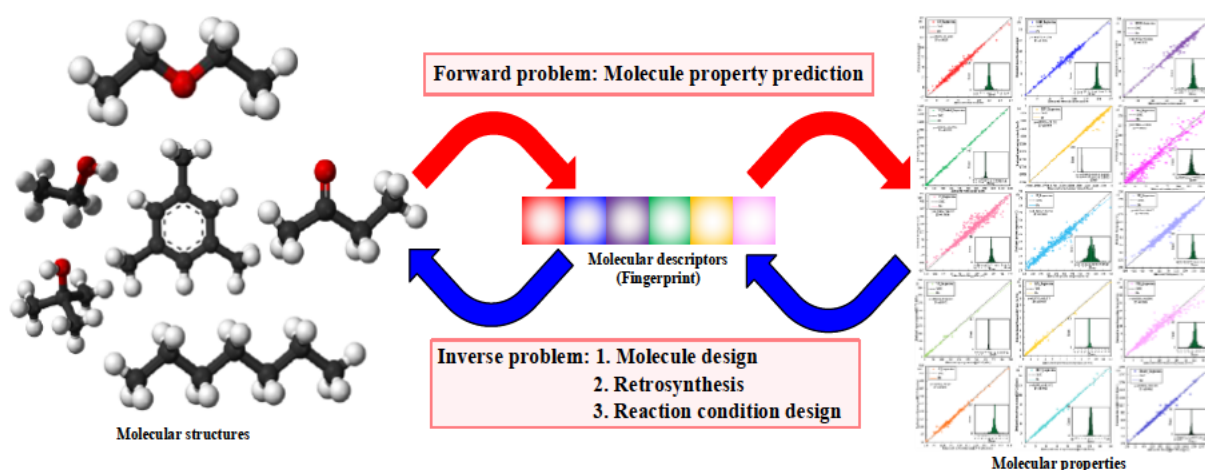
Figure 1.13. Two general problems of fuel design establishing the relationship between molecular structure and its properties: the forward problem of fuel property prediction and the inverse problem of molecule design and retrosynthesis.

Forward problem: Molecule property prediction

n-Propyl acrylate (CAS: 925-60-0)

Fingerprint

Property = f (descriptors)
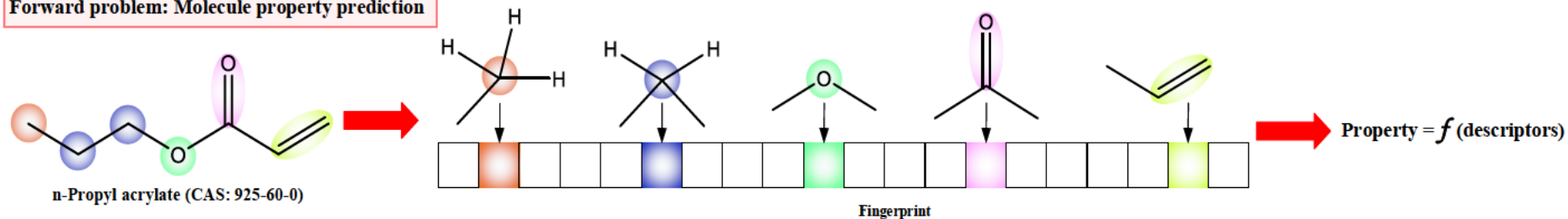
Inverse problem: Molecule design

Benzene
(CAS:71-43-2,C_6H_6)

C1=CC=CC=C1

C1=CC=CC=C1

Benzene
(CAS:71-43-2,C_6H_6)

Molecule input

SMILES input

Encoder (recurrent
neural network, RNN)

Continuous molecular
representation (latent space)

Decoder (recurrent
neural network, RNN)

SMILES output

Molecule output
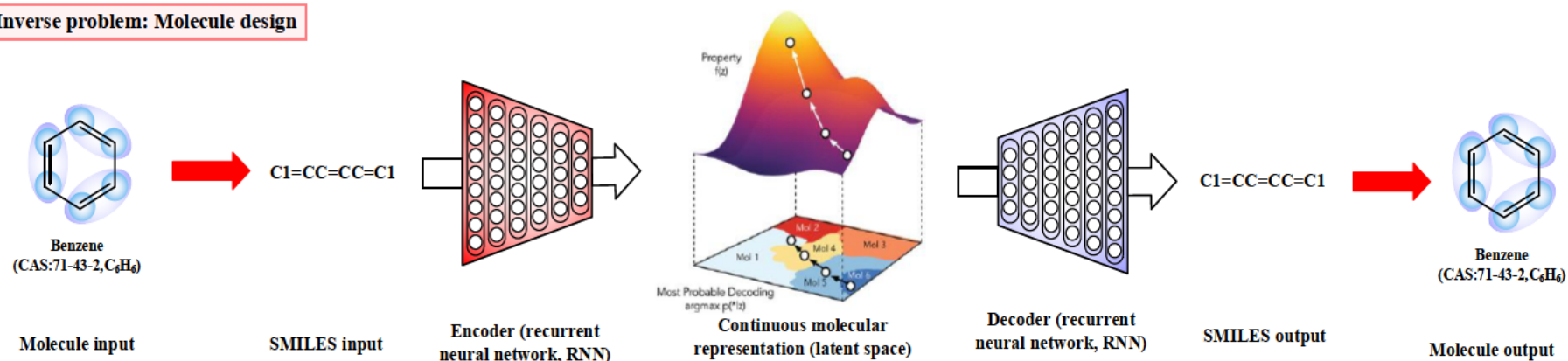
Inverse problem: Retrosynthesis/Reaction condition design

C1=COC=C1

Furan
(CAS:110-00-9,C_4H_4O)

Synthetic target

SMILES input or Draw structure

Chemical-rules based
automatic reaction generation

Produce possible
reaction network

Viable synthetic pathways terminating in
commercially available/known substrates
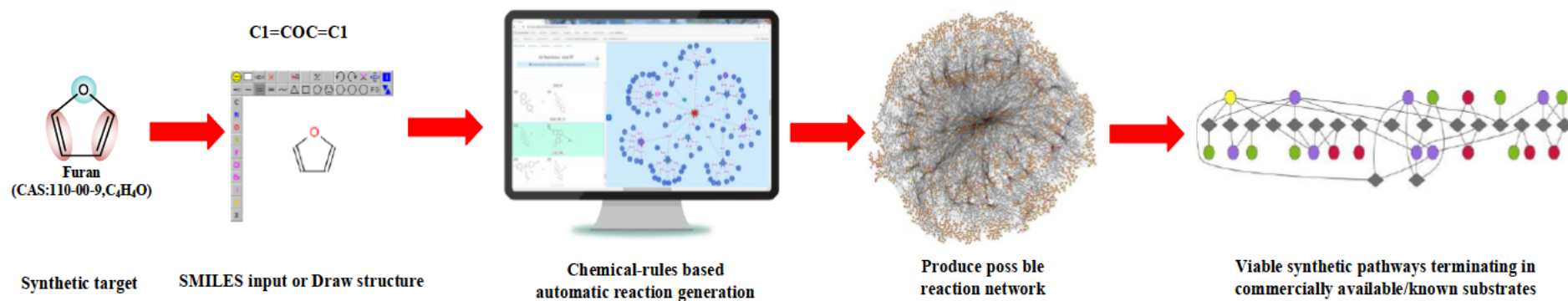
Figure 1.14. AI-powered fuel design: accelerate the forward problem of fuel property prediction and the inverse problem of molecule design, retrosynthesis reaction condition design.
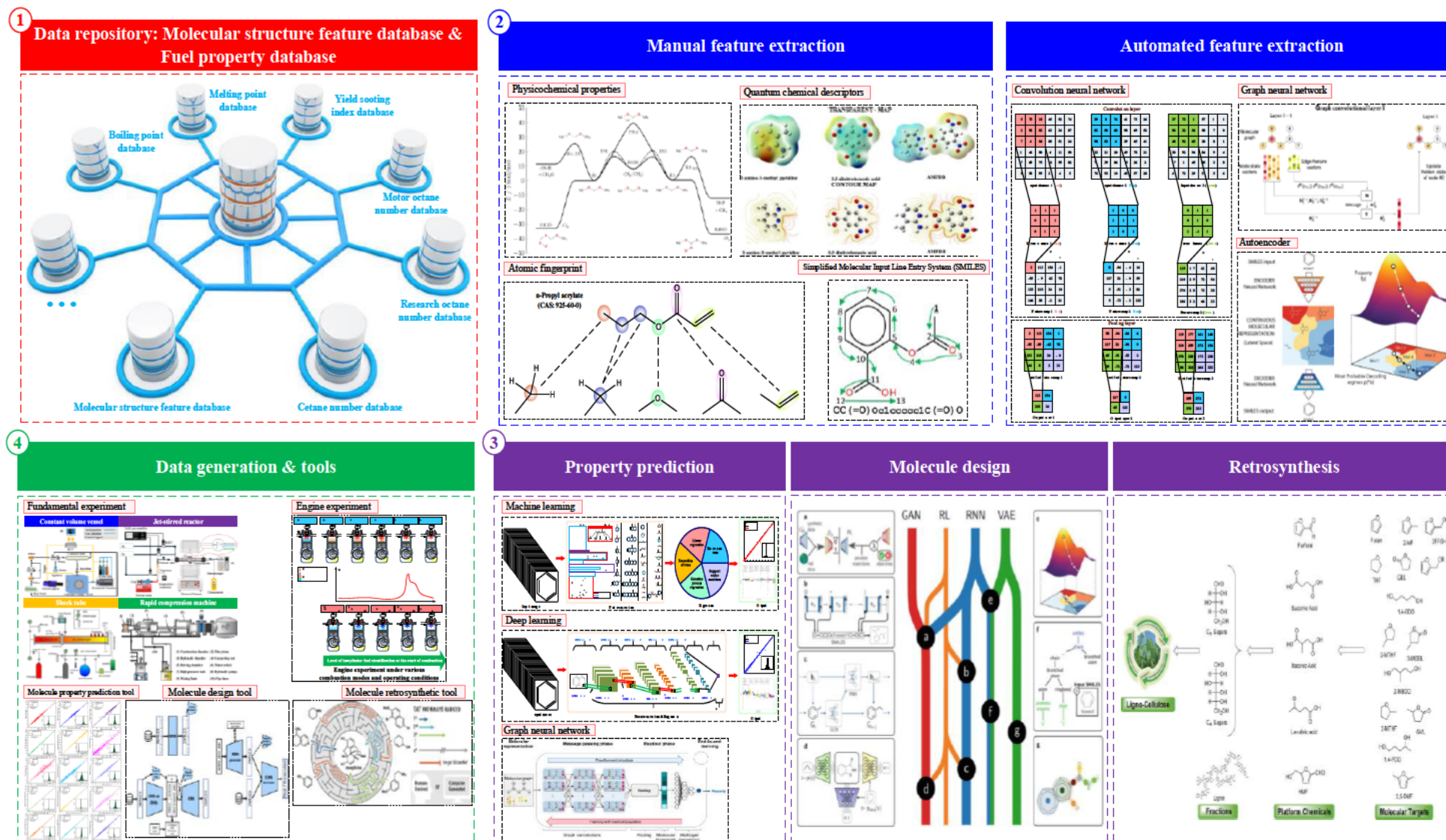
Figure 1.15. Technological roadmap of Fuel Genome Project.

1.3.2 Molecular descriptors design

Before applying AI technologies in chemistry, the molecular structure must be converted into notation that can be processed by the computer. Molecular descriptors (also known as molecular fingerprints or feature vectors) are mathematical representations of molecules that encode the structural features into numerical values [30]. Molecular descriptors development is an open research problem in progress and there is not a representation that is suited to all properties. In short, there are no best molecular descriptors, but the most appropriate representation to the property of interest. The molecular descriptors development should address 7 requirements as below [28, 31]:

(1) They should provide a good correlation performance to the property of interest;

(2) They should contain the appropriate amount of structural information;

(3) They should have a good molecular resolution to discriminate the among isomers;

(4) They should be simple to construct and interpret for the user convenience;

(5) They should not be based on the properties;

(6) They should satisfy the "uniqueness" demand in two directions. The forward direction is that the molecular structure should be expressed in a unique representation or notation. The inverse direction is that each representation corresponds to only one molecule. Therein, the forward direction must be addressed and the inverse direction is not mandatory due to the simplicity requirement.

(7) They should correctly reflect the molecular size variation.

The fuel molecule is composed of atoms and bonds and it can be expressed in five classes of graphics as shown in Figure 1.16. Correspondingly, there are five categories of molecular descriptors including 0D, 1D, 2D, 3D, 4D representation as summarized in Table 1.2. 0D descriptors include the chemical formula, molecular weight and other fuel properties. 1D molecular descriptors (also known as constitutional descriptors) count the chemical fragments. 2D molecular descriptors (also known as topological descriptors) provide information about the connection between atoms and bonds. 3D molecular descriptors (also known as geometrical descriptors) provide additional information about the spatial position of atoms/bonds or the distances and angles between atoms

compared to 2D descriptors. 4D molecular descriptors (also known as quantum descriptors) contain more complex information of electronic features, vibrational frequency levels, and reactivity indices, etc. Molecular descriptors classified by chemical nomenclature compose of 4 categories of "name", "line notations", "graph theory representation", "matrix representation" as shown in Table 1.3. The automated molecular descriptors calculation software is summarized in Table 1.4.
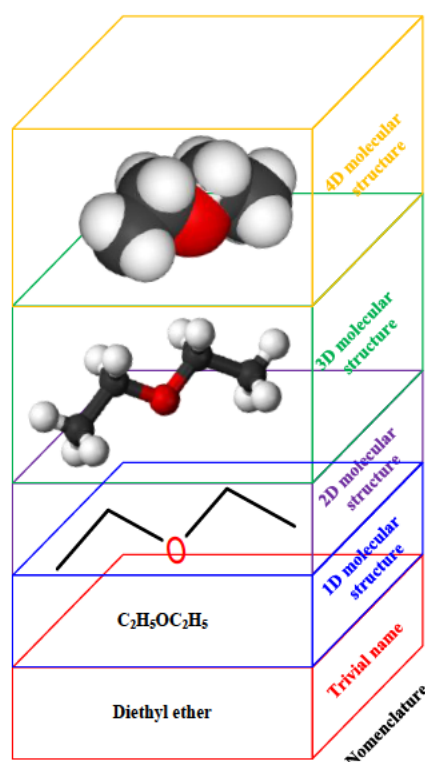
Figure 1.16. Hierarchical scheme of molecular structure with different levels of structural information [5].
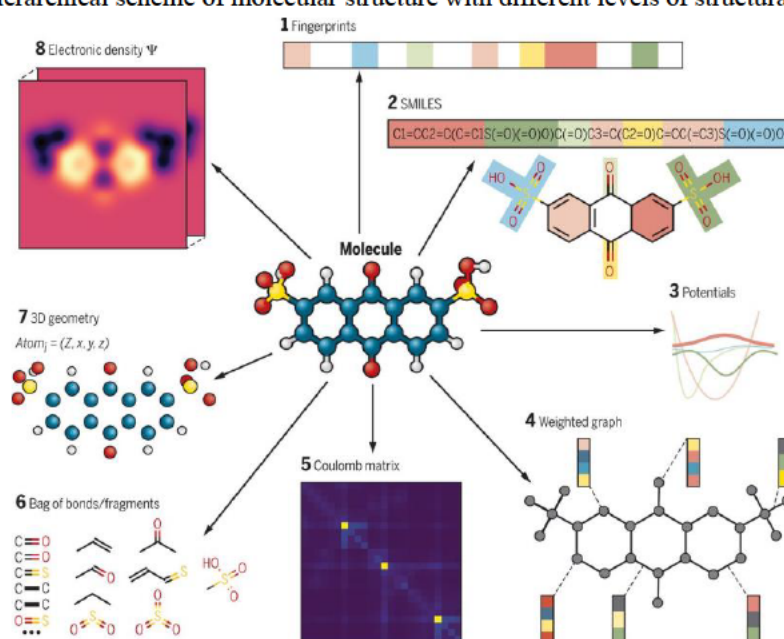
Figure 1.17. Molecular descriptors for chemoinformatics transform the molecule's geometric and chemical information into a numerical representation. Reproduced from ref. [32, 33].

Table 1.2. Molecular descriptors classification by dimensionality of molecular representation [5, 29, 30].

Represented dimension	Descriptor type	Example
0D	Atom/bond counts, molecular weight, atomic properties	Chemical formula, molecular weight, atoms types/numbers atomic or molecular properties
1D	Constitutional descriptors	Fragment counts of primary C, secondary C, tertiary C, quaternary C, secondary C in ring structure, tertiary C in ring structure, quaternary C in ring structure, alcohol/ester/ketone groups etc.
2D	Topological descriptors	Connectivity indices, Zagreb index, Wiener index, Balaban index etc.
3D	Geometrical descriptors	Polar surface area; molecular eccentricity, gyration radius, radial distribution function etc.
4D	Quantum descriptors	Electronic features, vibrational frequency levels, and reactivity indices etc. originating from density functional theory

Table 1.3. Molecular descriptors classification by chemical nomenclature [5]

Chemical nomenclature	Examples
Name	Trivial names, IUPAC (International Union of Pure and Applied Chemistry) nomenclature
Line notations	SMILES (simplified molecular-input line-entry system), WLN (Wiswesser line notation), ROSDAL (Representation of Organic Structures Description Arranged linearly), Sybyl line notation
Graph theory representation	Nodes and edges represent atoms and bonds.
Matrix representation	Adjacency matrix, distance matrix, atom connectivity matrix, incidence matrix, bond matrix, bond-electron matrix, connection table

Table 1.4 Summary of automated molecular descriptors calculation software [28, 34].

No.	Name	Institution	No. of descriptors	Platform requirement	Freely available	Ref.
1	BlueDesc	University of Tübingen	174	JAVA JDK or JRE 1.6.	Yes	[35]
2	ChemDes	Central South University	3679	Web interface	Yes	[36, 37]
3	ChemoPy	Central South University	1135	Windows/Linux	Yes	[38, 39]
4	Cinfony	Cinfony	N/A	Python	Yes	[40]
5	Dragon 7	Kode Chemoinformatics	5270	Windows/Linux	No	[41]
6	Mordred	Osaka University	1825	Python	Yes	[34, 42]
7	PaDEL	National University of Singapore	1875	Java JRE 6 or above	Yes	[43, 44]
8	PyDPI	Central South University	615	Python	Yes	[45, 46]
9	Rcpi	Central South University	307	R 4.1 or above	Yes	[47, 48]

1.3.3 AI technology overview

To better utilize the AI-assisted fuel design (including both forward problem of molecule property prediction and the inverse problem of molecule discovery), the application domain, techniques, tasks of AI-related algorithms are overviewed. AI algorithms are developed to perceive and learn from the environment, make a judgment and take action [49] which can be implemented in a variety of domains including robotics, computer vision, natural language processing, oil/energy sector, material/drug discovery, etc as shown in Figure 1.18 [50, 51]. Machine learning (ML) is a subset of AI techniques that composes of ensemble algorithms, dimensionality reduction algorithms, artificial neural network (ANN) algorithms, decision tree algorithms, Bayesian algorithms, instance-based algorithms. Regardless of the specific application domain, ML application falls into four categories of clustering, classification, regression, control as shown in Figure 1.19 (a). Clustering belongs to unsupervised learning which partitions the data into groups to find the similarity or shared characteristics. Unsupervised learning has only input data without corresponding output responses. Classification and regression belong to supervised learning that learns the inherent pattern from the known input & output data and trains a model to generate a prediction of the new input data. Reinforce learning is a dynamic process to implement control and decision-making of an agent-based on the environment observations and rewards. Figure 1.20 showcases a flow chart to choose the right algorithm for the task based on the data type and size. The data type required for unsupervised learning supervised learning and reinforce learning are compared in Table 1.5. Deep learning (DL) is a subset of ML which refers to a multilayers neural network for classification and regression as shown in Figure 1.19 (b). The ML workflow and DL workflow are compared in Figure 1.21 (a) and the key difference between ML and DL is that DL performs automated feature extraction while the ML requires manual feature extraction as shown in Figure 1.21 (b). Therefore, DL is particularly suited to images or video data input while ML can only process numerical data. ML has six categories of algorism to address regression problem (such as fuel property prediction) as shown in Figure 1.22: (a) linear regression; (b) decision tree; (c) support vector machine (SVM); (d) Gaussian process regression (GPR); (e) ensemble algorithm; (f) artificial neural network (ANN). The principle and application of these algorithms can be found in MATLAB Machine Learning Toolbox User's Guide [52], textbooks [53] and review article [50].

17
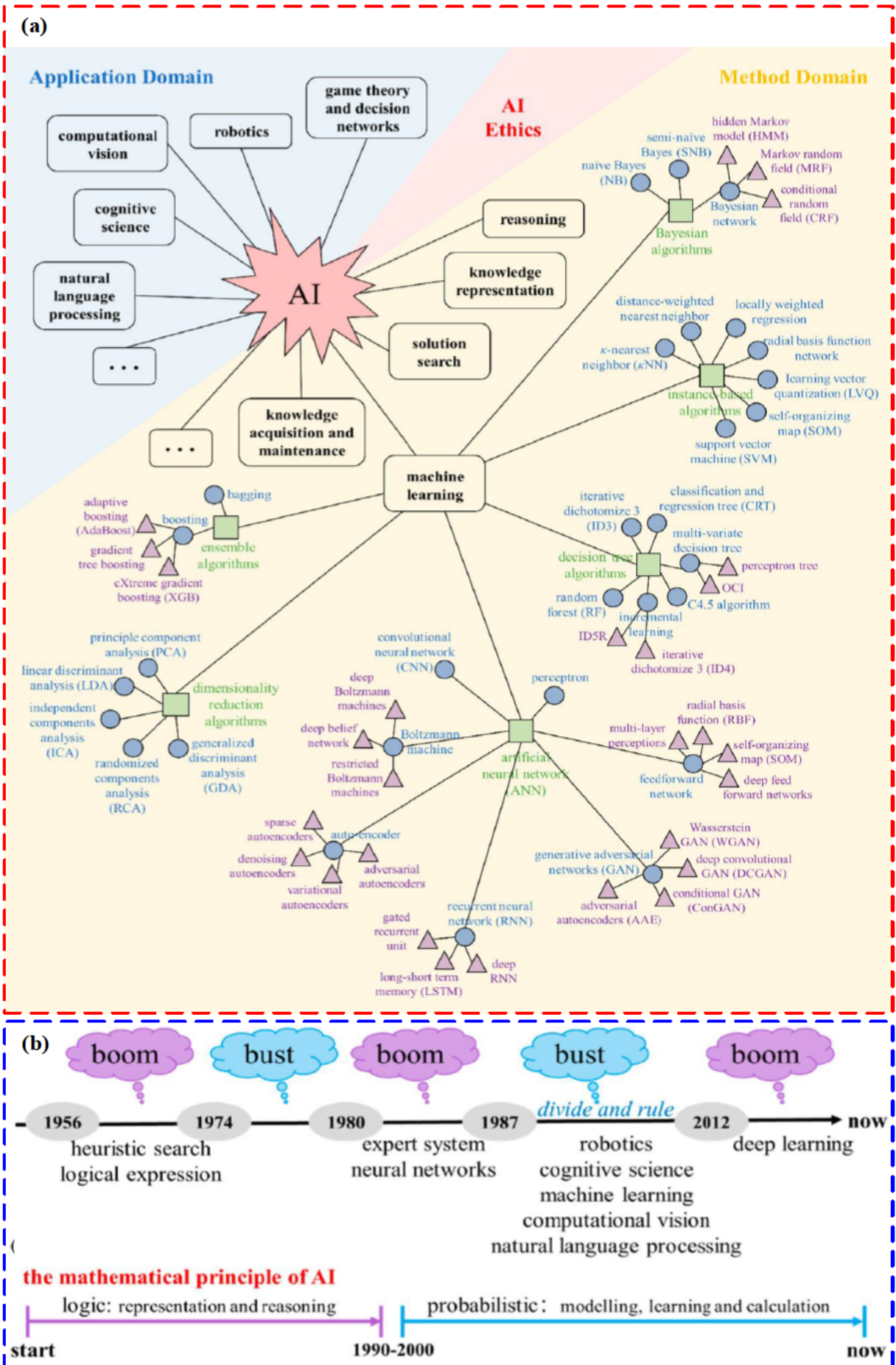
Figure 1.18. (a) Summary of AI-related technologies; (b) Brief AI development history. Reproduced from ref. [50].

18
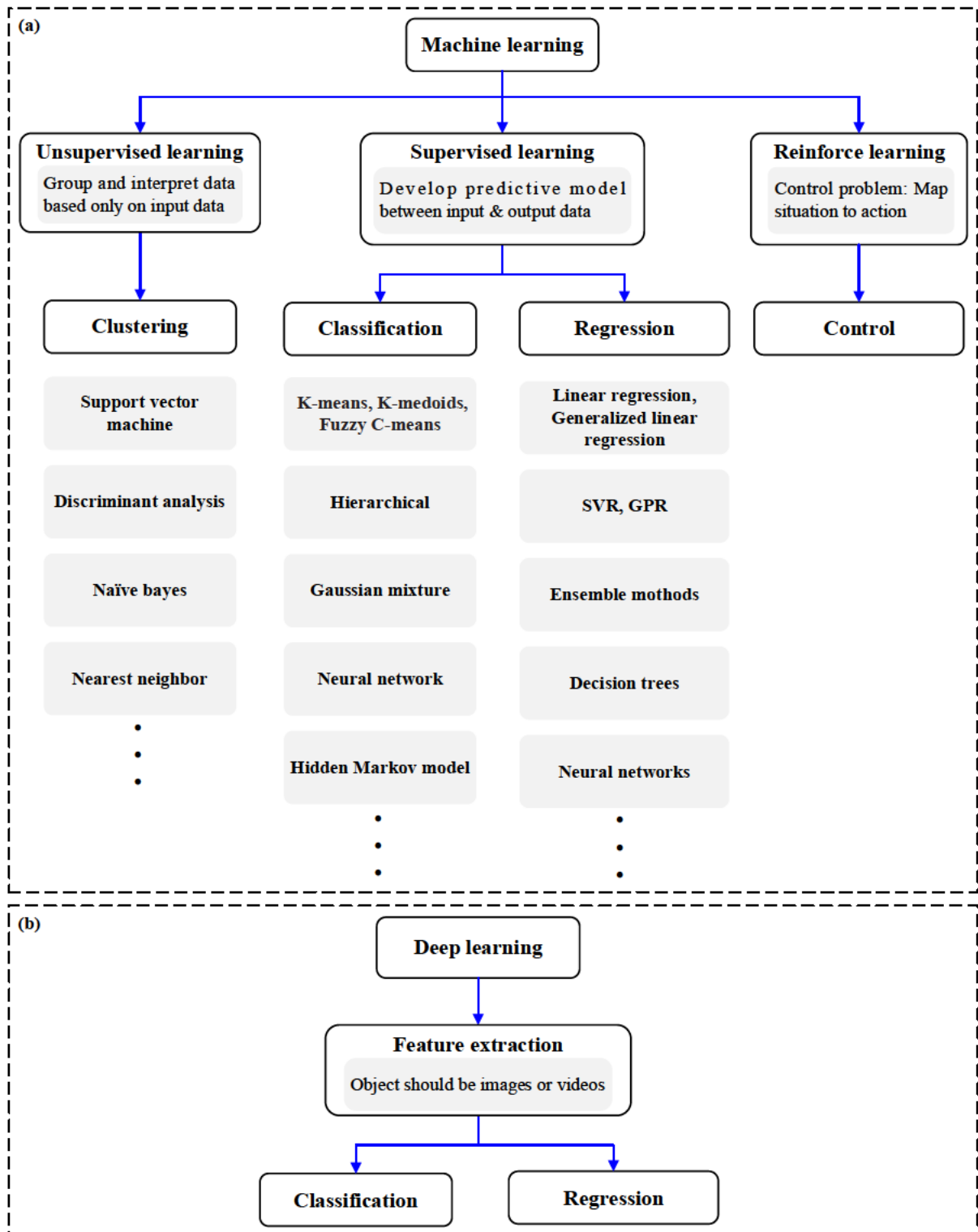
Figure 1.19. (a)Three types of ML technologies of supervised learning, unsupervised learning, reinforce learning and their application domain. (b) A branch of ML Deep learning and its application domain.

19
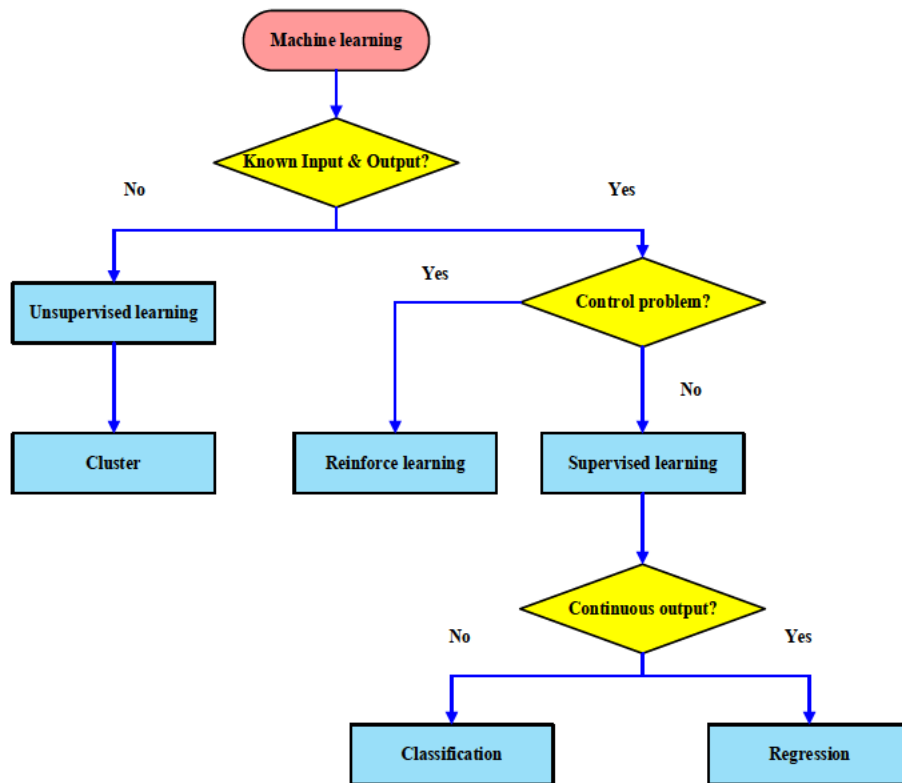
Figure 1.20. The flow chart to decide when to utilize unsupervised learning, supervised learning and reinforce learning.
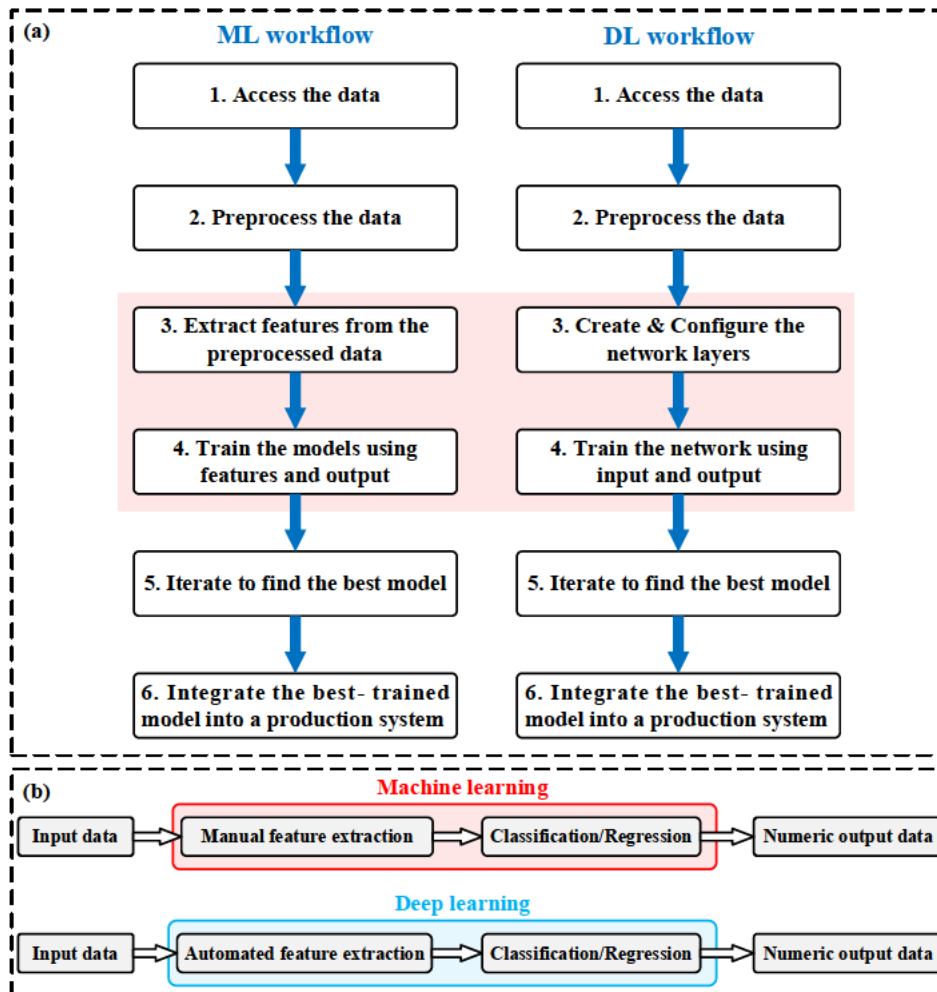
Figure 1.21. (a) The work flow of ML and DL; (b) The difference between ML and DL application.
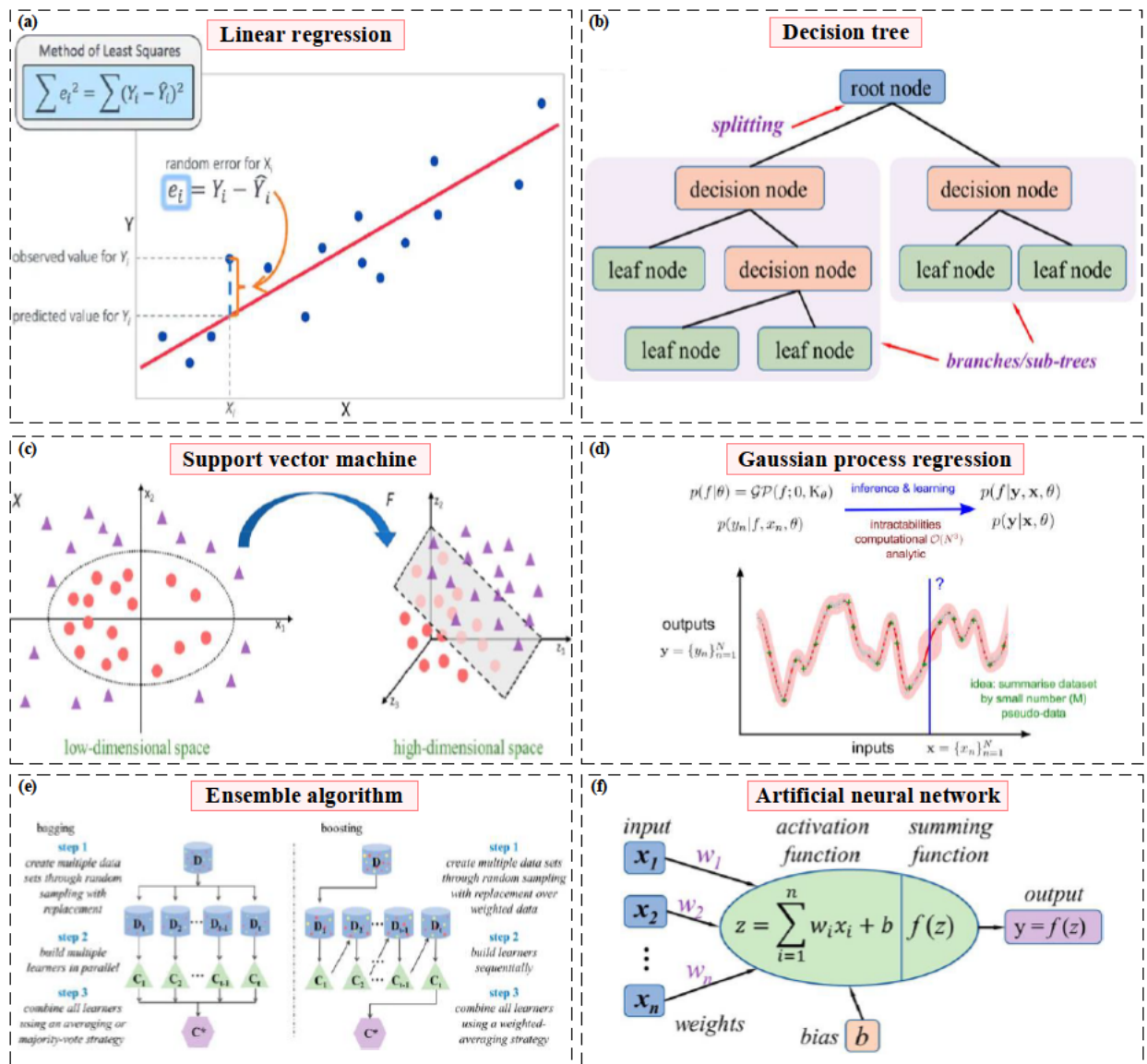
Figure 1.22. Summary of regression algorithm types: (a) linear regression; (b) decision tree; (c) support vector machine (SVM); (d) Gaussian process regression (GPR); (e) ensemble algorithm; (f) artificial neural network (ANN). Reproduced from ref. [50, 54, 55].

Table 1.5. Data type for unsupervised learning, supervised learning, reinforce learning [53].

Branch of ML	Training dataset requirement
Unsupervised learning	1. Input dataset
Supervised learning	1. Input dataset; 2. Output dataset
Reinforce learning	1. Input dataset; 2. Output dataset; 3. Grade of individual outputs

1.3.4 Molecular design and retrosynthesis software

For the inverse problem of molecule design, a variety of computer-aided molecular design software are available to rapidly construct the molecular graphics as summarized in ref. [56]. For the inverse problem of retrosynthesis and reaction condition design, the computer-aided organic synthesis software since 2009 is summarized in Table 1.6.

Table 1.6. Summary of computer-aided organic synthesis software since 2009 [α].

No.	Name	Freely available	Platform requirement	Ref.
1	AiZynthFinder	Yes	Python 3.6~3.8	[57, 58]
2	ASKCOS	Yes	Web interface or Windows	[59, 60]
3	Chemical.AI	Yes	Web interface	[61]
4	IBM RXN	Yes	Web interface	[62]
5	ICSYNTH	No	Windows	[63, 64]
6	Molecule.one	No	Web interface	[65]
7	RetSynth tool	Yes	Python 2.6/2.7	[66]
8	Route Designer	No	Windows	[67]
9	SciFinder[n] (formerly known as ChemPlanner)	No	Windows	[68]
10	Spaya	No	Web interface	[69]
11	SYLVIA	No	Windows	[70, 71]
12	Syntaurus	No	Windows	[72]
13	Synthia (formerly known as Chematica)	No	Windows	[72, 73]

[α] The retrosynthesis programs released before 2009 have been review by Wang et al. [74].

1.4 Thesis outline and research framework

Fuel Genome Project aims at addressing the forward problem of fuel property prediction and the inverse problems of molecule design, retrosynthesis and reaction condition prediction. This work primarily addresses the forward problem by integrating feature engineering theory, artificial intelligence (AI) technologies, gas-phase chemical kinetics. The thesis contains 9 chapters and their contents and inherent relations are described below (see Figure 1.23 as well):

Chapter 1: An overview of the property-oriented fuel design and the AI technology application (Human Genome Project, Material Genome Initiative, Polymer Genome) are provided. The motivation. technology roadmap and necessary technologies (molecular descriptors design, AI algorithms, software for molecular design & retrosynthesis) of the Fuel Genome Project are discussed in detail. The thesis outline and the intrinsic connection are explained.

Chapter 2: Group contribution method (GCM) is proposed for surrogate formulation which used functional group fragments as aligned parameters. The molecular descriptor scheme of GCM-UOB 1.0 contains 22 molecular descriptors which decomposes the molecules into specific atoms or functional groups. POSF 4658 jet fuel, rapeseed methyl ester (RME) biodiesel, diesel, FACE C gasoline are used as target fuels to showcase the application of GCM.

Chapter 3: The principle of quantitative structure−property relationships (QSPR) is that similar molecular structures result in similar physicochemical properties. The improved QSPR-UOB 2.0 system with 32 molecular features is developed to enhance the discrimination of aromatics up to 3 rings. Machine learning (ML) algorithms are utilized to learn a mapping function between molecular structures and properties.

Chapter 4: The ML-QSPR method for property prediction expanded from 3 properties (chapter 3) to 15 properties (chapter 4). The improved QSPR-UOB 3.0 scheme with 42 molecular descriptors is proposed to enhance the discrimination of aromatics (from 3 rings to 6 rings), esters, carboxylic anhydrides, hydroperoxides and

peroxides.

Chapter 5: In addition to the technical route of ML-QSPR regression models, another route of deep learning-convolution neural network (DL-CNN) is proposed for property prediction and yield sooting index (YSI) is taken as a case study. The predicted accuracy of DL-CNN is inferior to the ML-QSPR route at its current status, but its benefit of automated feature extraction and rapid advance in classification problem make it a promising solution for regression problem.

Chapter 6: A high-throughput fuel screening is performed to identify the molecule with desired properties for both spark ignition (SI) and compression ignition (CI) engines which contains the Tier 1 physicochemical properties screening (based on the ML-QSPR models) and Tier 2 chemical kinetic screening (based on detailed chemical mechanisms).

Chapter 7: Polyoxymethylene Dimethyl Ether 3 (PODE3) is a promising carbon-neutral fuel for CI engines which is shortlisted in the fuel screening recommendation in chapter 6. Its ignition delay time, laminar flame speed, dominant reactions are examined by chemical kinetics and are compared with n-heptane.

Chapter 8: Diethoxymethane (DEM) is another promising carbon-neutral fuel for CI engines and it is shortlisted in the fuel screening recommendation in chapter 6. A new DEM mechanism including both low and high-temperature reactions is constructed to compute ignition and flame speed properties.

Chapter 9: Concluding remarks and research prospects are summarized.
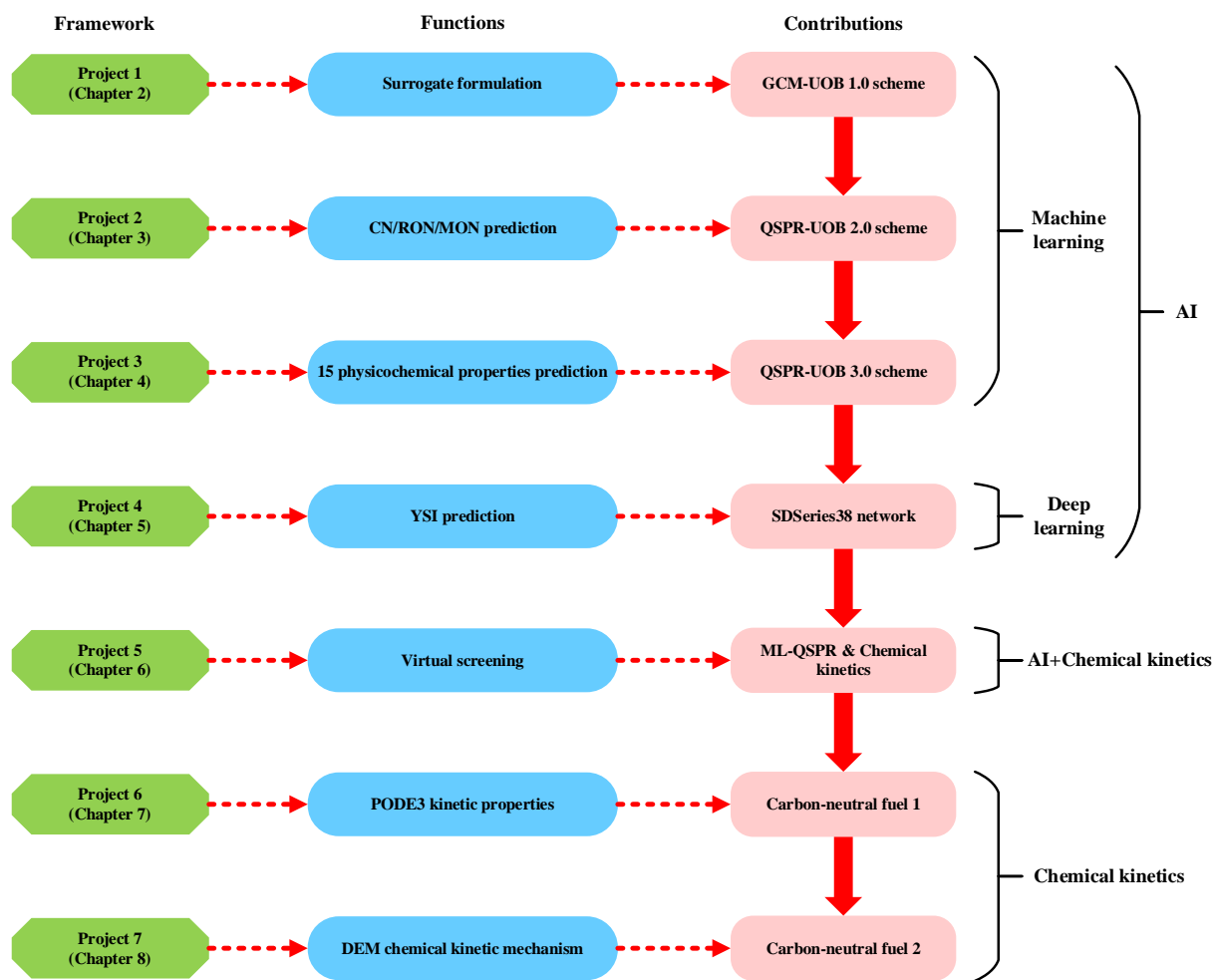
Figure 1.23. Thesis outline and research framework of the property-oriented fuel design for internal combustion engine.

Chapter 2 A Novel Functional Group Contribution Method for Surrogate Formulation with Accurate Fuel Compositions

2.1 Introduction

CFD (computer fluid dynamic) coupling combustion chemistry is an effective tool to develop and understand the application of conventional combustion modes (such as stoichiometric charge spark ignition, mixing controlled compression ignition) and the advanced low temperature combustion (LTC) modes (such as HCCI, PPCI, RCCI) [75] to the internal combustion engine. The CFD code combined with a detailed chemical kinetic mechanism enables the reactive flow simulation to accurately predict the combustion and emission characteristics of practical fuels with the aid of a high-performance computer [76, 77]. However, two problems need to be addressed to provide high fidelity simulation. (1) The complexity of representing the practical fuel with complex components. The detailed hydrocarbon analysis by two-dimensional gas chromatography (GC×GC) [78-80] with a FID (flame ionization detection) or TOFMS (time-of-flight mass spectrometry) quantifies that commercial diesel [81-83], gasoline [84-86], aviation fuel [87] contains individual components/isomers on the order of 5000 (mostly C8 - C25 molecules), 1200 (mostly C4 – C12 molecules), 2500 (mostly C6~C18 molecules). It is unrealistic to construct a detailed chemical kinetic mechanism containing all the detected compounds to mimic the combustion and pollutant formation chemistry, thus the "surrogate fuel" is formulated to address this problem. The surrogate formulation is to choose a single component or a few pure compounds blended together and quantify the proportions following specific criteria to emulate the target properties of the real fuel [88]. The surrogate fuel usually contains less than or equal to 10 components which makes it possible to develop the detailed chemical kinetic mechanism with the compromise of complexity and fidelity. (2) The complexity of detail chemical kinetic mechanism for large molecule fuels. The number of species and reactions grow exponentially with the molecular size and a large molecule contains a large-scale mechanism [89, 90]. For example. the mechanism of 2-methyl alkanes up to C20 developed by LLNL contains approximately 7200 species, 31400 reactions [91, 92] and the mechanism of

SME/RME (soy/rapeseed) developed by LLNL comprises about 4800 species, 20000 reactions. The detailed reaction mechanisms of large molecules cannot be directly applied for CFD simulation and significant mechanism reduction is necessary for computational facilitation. The mechanism reduction techniques are divided into 4 categories [88]: (a) species/reaction elimination method, (b) species lumping method, (c) dimension reduction method, (d) multistage reduction method.

The target properties of the surrogate formulation are used to validate and evaluate the properties between the target fuel and surrogate fuel. The commonly used target properties for surrogate formulation are divided into physical properties and chemical properties as summarized in Table 2.1. The physical properties compose of volatility (e.g. distillation curve), spray properties (e.g. density, viscosity, surface tension), diffusive properties (e.g. molecular weight). The chemical properties comprise chemical class similarity (e.g. proportions of n-/i-paraffins, naphthenes, olefins, aromatics), functional group similarity (e.g. structural fragments), carbon type similarity (e.g. carbon atom types), ignitability (e.g. RON/MON, CN/DCN), energy content (e.g. heating value), H/C ratio & pollutant emission (e.g. TSI), flame phenomenon (e.g. laminar flame speed).

To classify the surrogate formulation method, the concept of "aligned parameters" is introduced which denotes the parameters of the surrogate formulation model to be matched by adjusting the compositional proportions. The "aligned parameters" are used to determine the component compositions in the surrogate formulation model while the "target properties" are sued to validate and examine the predictive capacity of the surrogate fuel. The surrogate formulation methods are divided into 2 categories according to the type of aligned parameters as shown in Table 2.2. (1) Surrogate formulation methods utilize physical & chemical properties as aligned parameters. The typical methods include surrogate blend optimizer [93-95], surrogate formulation toward CFD engine simulation [96], comprehensive surrogate formulation method [81, 87, 97]. For example, the surrogate blend optimizer method combines the physical property of the distillation curve and chemical properties of LHV, H/C ratio, RON/MON/CN as aligned parameters of the surrogate formulation model. (2) Surrogate formulation methods utilize chemical

properties as the aligned parameter. The representative methods compose of group chemistry representative (GCR) method [98], structural compositions matching method [99, 100], minimalist functional group (MFG) [101-103], complex fuel surrogates model (CFSM) [104-107], chemical deconstruction method (CDM) [108] as well as this work. For example. the structural compositions matching method uses structural fragments of CH_3, CH_2, CH, C and phenyl group as aligned parameters to compute the component proportions.

After choosing the surrogate formulation method, another important step is to select the surrogate palette based on the criteria of feasibility, simplicity, similarity, cost [109, 110]. The feasibility requires that the detailed chemical kinetic mechanism should be available for the selected compounds. The simplicity requires adopting small molecules of fewer carbon atoms whenever possible. The similarity requires the surrogate fuel to mimic the target fuel on both physical properties and chemical properties. The cost requirement advocates adopting the fuel molecules with reasonable expense. The surrogate fuel component library for gasoline, jet fuel (kerosene), diesel, biodiesel and their mechanism source are summarized in Table 2.3 and the component library for gasoline, jet fuel and diesel recommended by other researchers is provided in ref. [94, 95, 111-117] , ref. [114, 118-120], ref. [94, 95, 113, 114, 121-123].

Current surrogate formulation methods adopt either physical & chemical properties or only chemical properties as aligned parameters, but the physicochemical properties of emerging fuels are usually not available or difficult to obtain. To address this problem, a surrogate formulation method based only on the molecular structure with reasonable accuracy is needed. This work proposes a novel functional group contribution method (GC<) for a surrogate formulation that minimizes the discrepancy of the functional group fragments between target fuel and surrogate fuel. A GCM-UOB 1.0 functional group classification system is established to decompose the fuel molecules into functional group fragments. The GCM surrogate formulation requires the accurate compositions of the target fuel to achieve high predictive accuracy.

Table 2.1. Target properties for surrogate formulation

Classification	Items	Representative parameters	Ref.
Physical properties	Volatility	Distillation profiles, T10, T50, T90	[81, 87, 93-98]
Physical properties	Spray properties	Density, viscosity, surface tension	[81, 87, 96, 97]
Physical properties	Diffusive properties	Molecular weight	[124, 125]
Chemical properties	Chemical class similarity	Proportions of n-/i-paraffins, naphthenes, olefins, aromatics	[93-96, 98, 126]
Chemical properties	Functional groups similarity	Functional group fragments	[99, 100]
Chemical properties	Carbon types similarity	Carbon atom types	[81, 87, 97]
Chemical properties	Ignition propensity	RON/MON (for gasoline type fuels) or CN/DCN (for diesel type fuels)	[81, 87, 93-98, 124-126]
Chemical properties	Energy content	Heating value, cumulative heat release	[93-96]
Chemical properties	H/C ratio, NO_x/CO/HC/PM emissions	H/C ratio, TSI	[87, 93-96, 124-127]
Chemical properties	Flame phenomena	Premixed laminar flame speed	[128, 129]

Table 2.2. Overview of surrogate formulation methodologies

No.	Method type	Surrogate formulation method	Aligned parameters	Ref.
1	Physical & chemical properties	Surrogate blend optimizer	distillation curve, lower heating value, H/C ratio, and RON/MON/CN	[93-95]
2	Physical & chemical properties	Surrogate formulation toward CFD engine simulation	distillation curve, liquid density, lower heating value (LHV), viscosity, chemical class composition, H/C ratio, CN/RON	[96]
3	Physical & chemical properties	Comprehensive surrogate formulation method	Distillation curve, liquid density; 11 carbon types, CN, TSI	[81, 87, 97]
4	Chemical properties	Group chemistry representative (GCR) method	chemical classes of n-paraffins, naphthenes and aromatics	[98]
5	Chemical properties	Structural compositions matching method	Structural compositions of CH_3, CH_2, CH, C and phenyl group	[99, 100]
6	Chemical properties	Minimalist functional group (MFG)	10 H types	[101-103]
7	Chemical properties	Complex fuel surrogates model (CFSM)	Quasi-components (QCs) and approximate discrete components (ADCs)	[104-107]
8	Chemical properties	Chemical deconstruction method (CDM)	Representative fuel components	[108]
9	Chemical properties	Group contribution method (GCM)	22 functional group fragments	This work

Table 2.3. Surrogate fuel component library and mechanism sources

No.	Fuel class	Components	Formula	CN	Gasoline	Jet fuel	Diesel	Biodiesel	Mechanism source
1	n-Alkanes	n-Butane	C_4H_{10}	20.6	√				[91, 92]
2		n-Pentane	C_5H_{12}	30	√				[91, 92]
3		n-Hexane	C_6H_{14}	47.9	√				[91, 92]
4		n-Heptane	C_7H_{16}	52.8	√	√	√		[91, 92]
5		n-Octane	C_8H_{18}	58.2	√	√	√		[91, 92]
6		n-Nonane	C_9H_{20}	60.9	√	√	√		[91, 92]
7		n-Decane	$C_{10}H_{22}$	65.5	√	√	√		[91, 92]
8		n-Undecane	$C_{11}H_{24}$	69	√	√	√		[91, 92]
9		n-Dodecane	$C_{12}H_{26}$	72.9	√	√	√		[91, 92]
10		n-Tridecane	$C_{13}H_{28}$	79		√	√		[91, 92]
11		n-Tetradecane	$C_{14}H_{30}$	85.1		√	√		[91, 92]
12		n-Pentadecane	$C_{15}H_{32}$	90		√	√		[91, 92]
13		n-Hexadecane	$C_{16}H_{34}$	98.5		√	√		[91, 92]
14		n-Heptadecane	$C_{17}H_{36}$	105			√		N/A
15		n-Octadecane	$C_{18}H_{38}$	106			√		N/A
16		n-Eicosane	$C_{20}H_{42}$	110			√		N/A
17	iso-Alkanes	iso-Butane (2-Methylpropane)	C_4H_{10}	0					[91, 92]
18		iso-Pentane (2-Methylbutane)	C_5H_{12}	25	√				[91, 92]
19		iso-Hexane (2-Methylpentane)	C_6H_{14}	34	√				[91, 92]
20		iso-Heptane (2-Methylhexane)	C_7H_{16}	43.5	√				[91, 92]
21		iso-Octane (2,2,4-Trimethylpentane)	C_8H_{18}	14	√	√	√		[91, 92]
22		iso-Dodecane (2,2,4,6,6-Pentamethylheptane)	$C_{12}H_{26}$	N/A		√	√		[91, 92]
23		2-Methylpentadecane	$C_{16}H_{34}$	N/A		√	√		[91, 92]
24		2,2,4,4,6,8,8-Heptamethylnonane	$C_{16}H_{34}$	15		√	√		[91, 92]
25	Cycloalkanes	Cyclopentane	C_5H_{10}	6.1	√				[130]
26		Cyclohexane	C_6H_{12}	18.5	√	√	√		[131, 132]

27		Methylcyclohexane	C_7H_{14}	22.5	√	√	√		[113, 133, 134]
28		Ethylcyclohexane	C_8H_{16}	35.8	√	√	√		[113]
29		Propylcyclohexane	C_9H_{18}	N/A		√	√		[113]
30		Butylcyclohexane	$C_{10}H_{20}$	47.6		√	√		[113]
31		Dimethylcyclohexane	C_8H_{16}	N/A	√	√			[135]
32		Cyclooctane	C_8H_{16}	22.3		√			N/A
33		Decalin	$C_{10}H_{18}$	44		√	√		[136]
34		n-Pentylcyclohexane	$C_{11}H_{22}$	N/A		√	√		N/A
35		n-Heptylcyclohexane	$C_{13}H_{26}$	N/A		√	√		N/A
36		n-Dodecylcyclohexane	$C_{18}H_{36}$	N/A			√		N/A
37	Alkenes	2-Methyl-2-butene	C_5H_{10}	20					[137]
38		Diisobutylene	C_8H_{16}	N/A	√				[138-140]
39		1-Pentene	C_5H_{10}	24.4	√				[91, 92, 140]
40		2-Pentene	C_5H_{10}	17	√				[91, 92, 140]
41		1-Hexene	C_6H_{12}	27	√		√		[91, 92, 140]
42		2-Hexene	C_6H_{12}	20	√		√		[91, 92, 140]
43		3-Hexene	C_6H_{12}	16	√		√		[91, 92, 140]
44		Diisobutylene	C_8H_{16}	N/A	√				[138]
45		trans-2-Pentene	C_5H_{10}	N/A	√				[140]
46		trans-2-Hexene	C_6H_{12}	N/A	√				[140]
47	Esters	Methyl butanoate	$C_5H_{10}O_2$	30				√	[141]
48		Methyl crotonate	$C_5H_8O_2$	0				√	[141]
49		Methyl decanoate	$C_{11}H_{22}O_2$	52.7				√	[141]
50		Methyl palmitate	$C_{17}H_{34}O_2$	85.9				√	[141]
51		Methyl stearate	$C_{19}H_{38}O_2$	101				√	[141]
52		Methyl oleate	$C_{19}H_{36}O_2$	57				√	[141]
53		Methyl linoleate	$C_{19}H_{34}O_2$	38.2				√	[141]
54		Methyl linolenate	$C_{19}H_{32}O_2$	22.7				√	[141]
55	Alcohols	Methanol	CH_3OH	3	√				[142]

56		Ethanol	C_2H_5OH	12	√			[142]
57		n-Butanol	C_4H_9OH	17	√			[142]
58	Ethers	Dimethyl ether (DME)	C_2H_6O	55			√	[143-145]
59		Methyl tert-butyl ether (MTBE)	$C_5H_{12}O$	24	√			N/A
60		Ethyl tert-butyl ether (ETBE)	$C_6H_{14}O$	24	√			N/A
61		Tert-amyl methyl ether (TAME)	$C_6H_{14}O$	N/A	√			N/A
62		Diisopropyl ether (DIPE)	$C_6H_{14}O$	N/A	√			N/A
63	Cyclic ethers	Tetrahydrofuran	C_4H_8O	18	√			N/A
64		Ethyltetrahydrofurfurylether	$C_7H_{14}O_2$	82	√			N/A
65	Aromatics	Benzene	C_6H_6	10.7	√	√		[146]
66		Toluene	$C_6H_5CH_3$	2.6	√	√	√	[146]
67		Ethylbenzene	$C_6H_5C_2H_5$	7.4	√	√	√	[146]
68		Propylbenzene	$C_6H_5C_3H_7$	7.6	√	√	√	[146]
69		Butylbenzene	$C_6H_5C_4H_9$	12		√	√	[146]
70		Heptylbenzene	$C_6H_5C_7H_{15}$	35		√	√	N/A
71		o-Xylene	$C_6H_4(CH_3)_2$	8.3	√	√	√	[147]
72		m-Xylene	$C_6H_4(CH_3)_2$	2.6	√	√	√	[147]
73		p-Xylene	$C_6H_4(CH_3)_2$	2.6	√	√	√	[147]
74		1,2,3-Trimethyl benzene	C_9H_{12}	10.1	√	√	√	[148]
75		1,2,4-Trimethyl benzene	C_9H_{12}	8.9	√	√	√	[148]
76		1,3,5-trimethylbenzene	C_9H_{12}	8	√	√	√	[148]
77		1,2,4,5-Tetramethylbenzene	$C_{10}H_{14}$	N/A		√		N/A
78		Naphthalene	$C_{10}H_8$	22.6			√	[149, 150]
79		Tetralin	$C_{10}H_{12}$	N/A		√	√	[151]
80		1-Methylnaphthalene	$C_{11}H_{10}$	0		√	√	[152]
81		n-Decylbenzene	$C_{16}H_{26}$	N/A			√	N/A

2.2 Surrogate formulation methodology

2.2.1 Surrogate formulation methodology

Before explaining the surrogate formulation methodology, the terms "target fuel", "surrogate fuel", "target property", "surrogate palette", "functional group fragment" need to be clarified as shown in Figure 2.1. The "target fuel" is a fuel mixture with desired physicochemical properties that are to be matched by a surrogate fuel. It is usually a commercial practical fuel (such as diesel, gasoline, aviation fuel) containing up to thousands of individual components. "Surrogate fuel" is a single fuel component or simple fuel mixture containing several pure compounds blending to reproduce the compositional characteristics and the target physicochemical properties. "Target property" is a certain set of key properties to be emulated by the surrogate fuel. For example, cetane number (CN), research octane number (RON), motor octane number (MON) are the commonly used "target property" to depict the ignition quality of "target fuel". "Surrogate palette" is several pure components blending to formulate a "surrogate fuel" and each component is called a "palette compound". "Functional group fragment", also known as "structural fragment", "chemical fragment", is a set of functional group types or molecular descriptors in the functional group classification system. There are 22 molecular descriptors in the GCM 1.0 functional group classification system of this work which belong to 3 categories of "functional group identifier", "functional group position descriptor", "fuel reactivity descriptor".

The group contribution method (GCM) assumes that the fuel property is the result of the molecular structure, thus the fuel molecules with similar chemical structures have similar properties. The workflow of surrogate fuel formulation comprises 7 steps as shown in Figure 2.2: (1) Specify the target fuel (as well as target properties) and quantify the detailed chemical compositions. A set of ASTM standards are established to characterize the exact chemical composition as summarized in Table 2.4. The individual components of PIONA (paraffins, isoparaffins, olefins, naphthenes, aromatics), olefins, aromatics, oxygenates, ethers/alcohols, FAME (fatty acid methyl esters) can be quantified by ASTM D6730, ASTM D6550, ASTM D5580, ASTM D5599, ASTM D4815, ASTM E2997.

The exact hydrocarbons compositions in the target fuel can be quantified through the followed two methods: (i) two-dimensional gas chromatography with flame ionization detection (GC×GC-FID) [79, 80] or two-dimensional gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) [87]. It is a molecular level composition analysis method that consists of two columns: the nonpolar column separates the constituent compounds by boiling points and the other separates by polarity. Two-column configurations can be used: i. In the "normal" column configuration, the 1st column is a nonpolar column and the 2nd column is a semipolar column. ii. In the "reversed" column configuration, the orders of the nonpolar column and semipolar column reverses. The reversed configuration enables higher separation accuracy and resolution of different hydrocarbon classes and one–ring, two-ring and multi-ring cycloalkanes. Other information provided by the reversed column configuration is similar to the normal one. It yields fuel composition on a per-molecule basis and its principle is described in ref. [78]; (ii) Proton-decoupled 13C (carbon-13) and 1H (proton) Nuclear Magnetic Resonance (NMR) spectroscopy [87, 97, 153, 154]. It measures the mole fractions of different carbon types and yields fuel composition on a per-carbon-atom basis of quantitative carbon spectra. (2) Choose the surrogate palette which constitutes the compositions of surrogate fuels. Each component represents one of the fuel types in the target fuel and contains comparable carbon atoms number, molecular weight. (3) Decompose the fuel molecules into the designated functional group fragments for both target fuel and surrogate fuel. The molecule decomposition rule follows the GCM 1.0 functional group classification system which composes of 22 molecular descriptors as shown in Figure 2.3. No. 1 ~ No. 5, No. 15 ~ No. 22 fragments belong to functional group identifiers which depict the chemical class of the studied molecule. They can discriminate the aromatics, alkanes, alkenes, cycloalkanes, alkynes, alcohols, ethers, ketones, aldehydes, esters, carboxylic acids. No. 2 ~ No. 4, No. 6 ~ No. 7, No. 10, No. 12 ~ No. 14 fragments are functional group descriptors that identify the carbon atom type and corresponding position. They use constitutional information to discriminate the isomers of molecules. The fuel reactivity descriptors include No. 8 ~ No. 12 functional group fragments which represent the fuel reactivity by the number of the methyl group,

methylene group and their ratio as recommended by Dooley et al. [155, 156], Dahmen et al. [157] and Yu et al. [99, 100]. (4) Establish and run the functional group regression model to determine the component proportions. The development of the functional group regression model is explained in the next paragraph. (5) Test the target properties of the formulated surrogate fuel. (6) Compare the target properties of the surrogate fuel with the target fuel and estimate if the property target is achieved. If yes, go to step (7), otherwise, go back to step (2) to adjust the surrogate palette. (7) Complete the surrogate formulation, report the surrogate palette and corresponding proportion.

Running functional group regression model minimizes the object function Eq. (2.1) by adjusting the surrogate palette proportions.

$$S = \sum_{i=1}^{N_g} W_i \cdot F_i \tag{2.1}$$

where S is the objective function to be minimized in the iterative optimization procedure. W_i is the weighting factor of the i th functional group which satisfies the constraint of $\sum_{i=1}^{N_g} W_i = 1$. The weighting factors reflect the significance of each functional group and can be adjusted according to demands. To obtain the global optimum the weighting factors are automatically determined by the regression model iteration in this work. N_g is the total number of the functional group fragments and F_i is the normalized difference of the i th functional group fragment between the surrogate fuel and the target fuel which is calculated by Eq. (2).

$$F_i = \left| \frac{FGN_{i\,calc} - FGN_{i\,ta\,g}}{FGN_{i\,ta\,g}} \right| \tag{2.2}$$

where $FGN_{i,targ}$ and $FGN_{i,calc}$ denote the total number of the i th functional group fragment in the target fuel molecules and the surrogate fuel molecules. Therein, the $FGN_{i,calc}$ is calculated by Eq. (2.3).

$$FGN_{i\,calc} = \sum_{j=1}^{N} n_j \cdot fgn_{i\,j} \tag{2.3}$$

where n_j is the molar fraction of the j th surrogate palette compound which satisfies the constraint of

$\sum_{j=1}^{N} n_j = 1$. N_f is the number of the surrogate palettes and $fgn_{i,j}$ is the i th functional group number of the

j th surrogate palette compound.

The surrogate formulation by GCM is a what-if analysis and the Microsoft Excel add-in program-Solver can be employed to find an optimal value [158, 159]. The objective function S in Eq. (2.2) is objective which subject to the constrains of of $\sum_{i=1}^{N_g} W_i = 1$ and $\sum_{j=1}^{N} n_j = 1$. The molar fraction n_j and the weighting factor W_i of the surrogate palette are set as the decision variables to adjust the vales to produce the optimal results of the objective and satisfy the limits.

Target fuel: A fuel with selected properties that are to be matched by a surrogate fuel. Target fuel is usually produced in a refinery or other large-scale production process and typically contains tens to thousands of individual compounds.

Surrogate fuel: A fuel mixture containing a limited number of pure compounds blending together or a single compound to match specific properties of the target fuel.
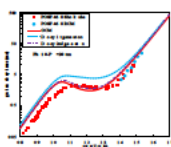
Target property: A selected property of the target fuel that is to be matched by the surrogate fuel. Cetane number (CN) is an example of **target property** to describe the ignition propensity.
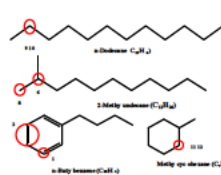
Surrogate palette: A set of pure compounds that are blended together to create a surrogate fuel. Each individual pure compound in the surrogate palette is called a **palette compound**.

Functional group fragment: There are 22 functional group types in the functional group classification system which contains **functional group identifier**, **functional group position descriptor** and **fuel reactivity descriptor**.

Figure 2.1. Definitions of terms used in this study.
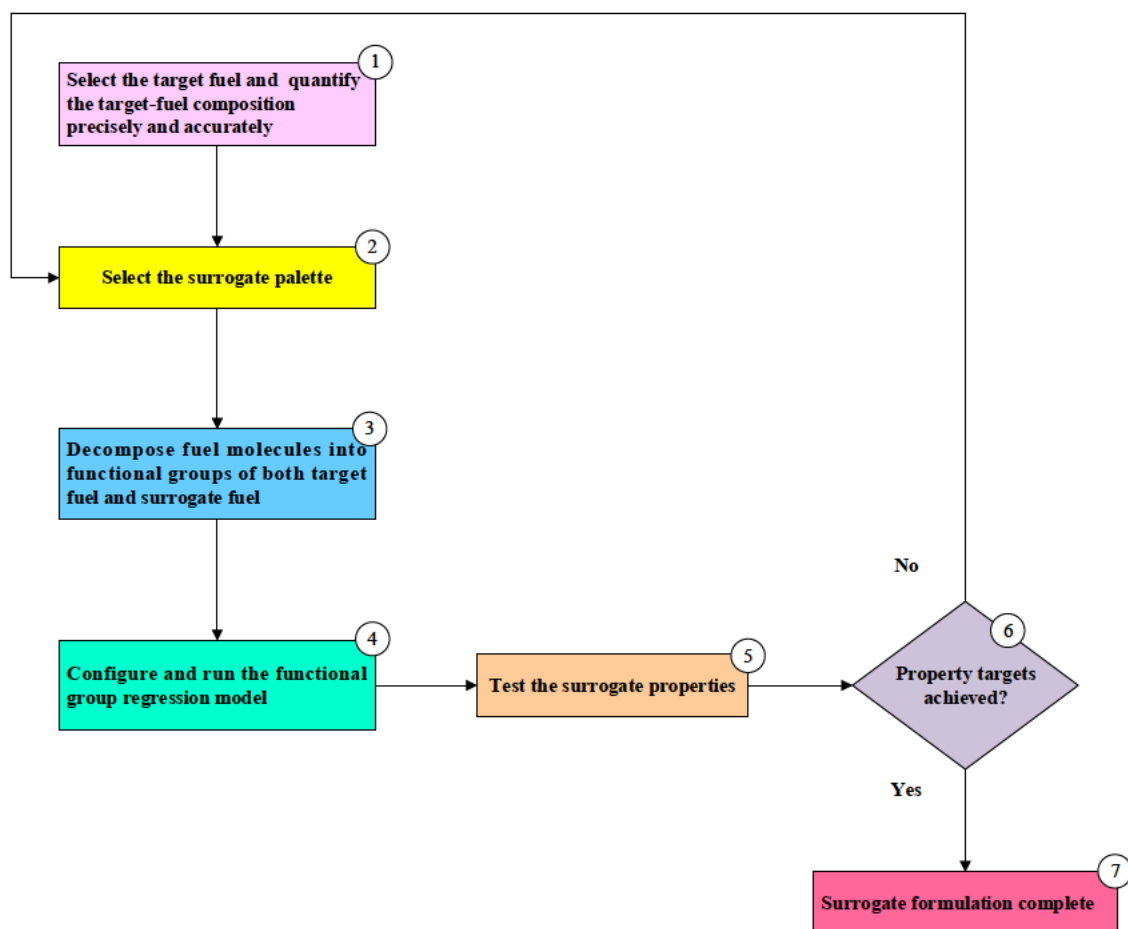
Figure 2.2. Workflow of surrogate fuel formulation. The order of each step in the sequence is provided in the upper-right region of each box.

FUNCTIONAL GROUP CLASSIFICATION SYSTEM

1. **Aromatic bond**
2. **Carbon-carbon double bond (CCDB) of aromatic, CH$_2$=CH$_2$**
3. **Carbon-carbon double bond (CCDB) of ring, CH$_2$=CH$_2$**
4. **Carbon-carbon double bond (CCDB) of non-aromatic, non-ring, CH$_2$=CH$_2$**
5. **Carbon-carbon triple bond (CCTB)**
6. **Tertiary carbon, >CH-**
7. **Quaternary carbon, >C<**
8. **Primary carbon (methyl radical), -CH$_3$**
9. **Maximal quantity of secondary carbon in series (non-ring) (methylene), >(CH$_2$)$_m$ (non-ring)**
10. **Secondary carbon (non-ring) (methylene), >CH$_2$ (non-ring)**
11. **Maximal quantity of secondary carbon in series (ring) (methylene), >(CH$_2$)$_m$ (ring)**
12. **Secondary carbon (ring) (methylene), >CH$_2$ (ring)**
13. **>CH-, non-Tertiary carbon**
14. **>C<, non-Quaternary carbon**
15. **Hydroxyl radical, -OH**
16. **Ether group (non-ring), –O–(non-ring)**
17. **Ether group (ring), –O–(ring)**
18. **Ketone group(non-ring), >C=O(non-ring)**
19. **Ketone group(ring), >C=O(ring)**
20. **Aldehyde group, -CH=O**
21. **Ester group, -C(=O)O-**
22. **Carboxylic acid, -C(=O)OH**

— Functional group identifier

••••••• Functional group position descriptor

—•—•— Fuel reactivity descriptor

Benzene (C$_6$H$_6$)

1,3-Cyclohexadiene (C$_6$H$_8$)

1-Hexene (C$_6$H$_{12}$)

1-Pentyne

2-Methylbutane

2,2-Dimethylbutane

Cyclopentane (C$_5$H$_{10}$)

2-Propanol (C$_3$H$_8$O)

tert-Butanol (C$_4$H$_{10}$O)

Methanol (CH$_3$OH)

Dimethyl ether (C$_2$H$_6$O)

Furan (C$_4$H$_4$O)

2-Propanone (C$_3$H$_6$O)

Cyclopentanone (C$_5$H$_8$O)

Acetaldehyde (C$_2$H$_4$O)

Methyl acetate (C$_3$H$_6$O$_2$)

Acetic acid (C$_2$H$_4$O$_2$)

Figure 2.3. The functional group classification system is used to match compositional characteristics between target and surrogate fuels. The functional groups are listed on the left, and an example of each functional group is circled in the molecular structure on the right.

Target fuel   **Surrogate fuel**

$$\min S = \sum_{i=1}^{N_g} W_i \cdot F_i$$

$$F_i = \left| \frac{FGN_{i\,calc} - FGN_{i\,m\,g}}{FGN_{i\,m\,g}} \right|$$

$$FGN_{i\,calc} = \sum_{j=1}^{N} n_j \cdot fgn_{i\,j}$$

$$\sum_{i=1}^{N_g} W_i = 1$$

$$\sum_{j=1}^{N} n_j = 1$$
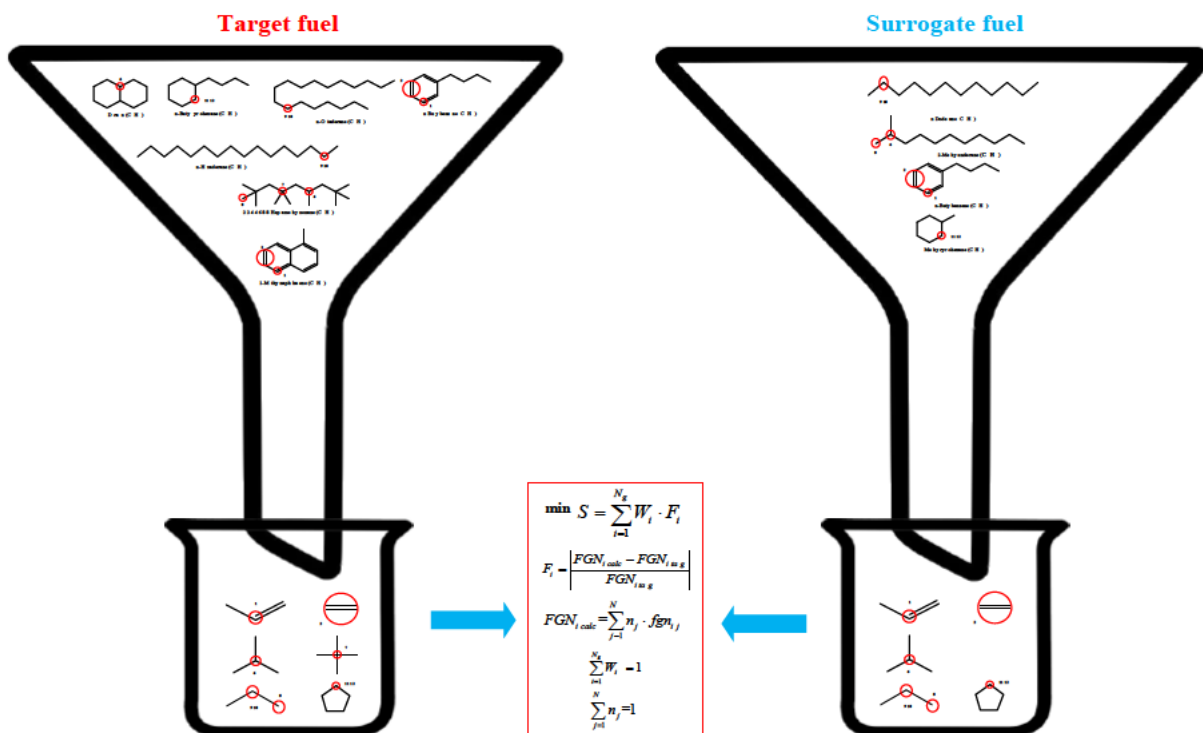
Figure 2.4. Surrogate formulation by decomposing molecules into chemical fragments and functional group regression model development, take diesel as an example.

Table 2.4. ASTM standards for exact chemical composition characterization

Item	Compositions	Scope	Ref.
ASTM D6730	PIONA	Individual hydrocarbon components (PIONA) of spark-ignition engine fuels and their mixtures containing oxygenate blends (MTBE, ETBE, ethanol, and so forth) with boiling ranges up to 225 °C	[85]
ASTM D6550	Olefins	The total amount of olefins in blended motor gasoline and gasoline blending stocks, the application range is from 1 mass % to 25 mass % total olefins	[160]
ASTM D5580	Aromatics	Benzene, toluene, ethylbenzene, the xylenes (p-xylene and m-xylene), C9 and heavier aromatics and total aromatics in finished motor gasoline	[161]
ASTM D5599	Oxygenates	Organic oxygenated compounds in gasoline having a final boiling point not greater than 220 °C and oxygenates having a boiling point limit of 130°C	[162]
ASTM D4815	Ethers and alcohols	Ethers (0.20mass%~20.0mass%): MTBE, ETBE, TAME, DIPE; Alcohols (0.20 mass%~12.0 mass%): methanol, ethanol, isopropanol, n-propanol, isobutanol, tert-butanol, sec-butanol, n-butanol and tert-pentanol	[163]
ASTM E2997	FAME	Fatty acid methyl esters (FAMEs) and petroleum distillate components of biodiesel products	[164]

Table 2.5. Count of functional groups of the real fuel and selected surrogate palette

Palette compound	Target fuel	Occurrence of functional group types										
		1	2	4	6	7	8	9	10	11	12	21
n-Dodecane	POSF 4658	0	0	0	0	0	2	10	10	0	0	0
2,2,4-Trimethylpentane	POSF 4658/FACE C gasoline	0	0	0	1	1	5	1	1	0	0	0
n-Propylbenzene	POSF 4658	6	3	0	0	0	1	2	2	0	0	0
1,3,5-Trimethylbenzene	POSF 4658	6	3	0	0	0	3	0	0	0	0	0
Toluene	POSF 4658/FACE C gasoline	6	3	0	0	0	1	0	0	0	0	0
Decane	POSF 4658	0	0	0	0	0	2	8	8	0	0	0
Methyl decanoate	POSF 4658	0	0	0	0	0	0	8	8	0	0	1
Methyl-9-decenoate	POSF 4658	0	0	1	0	0	0	7	7	0	0	1
Methyl-5-decenoate	POSF 4658	0	0	1	0	0	0	3	6	0	0	1
n-Hexadecane	POSF 4658	0	0	0	0	0	2	14	14	0	0	0
Methyl palmitate	RME	0	0	0	0	0	2	14	14	0	0	1
Methyl stearate	RME	0	0	0	0	0	2	16	16	0	0	1
Methyl oleate	RME	0	0	1	0	0	2	7	14	0	0	1
Methyl linoleate	RME	0	0	2	0	0	2	7	12	0	0	1
Methyl linolenate	RME	0	0	3	0	0	2	7	10	0	0	1
2,2,4,4,6,8,8-Heptamethylnonane	Diesel	0	0	1	3	9	1	3	0	0	0	0
1-Methylnaphthalene	Diesel	10	5	0	0	1	0	0	0	0	0	0
cis-Decalin	Diesel	0	0	2	0	0	0	0	4	8	0	0
n-Octadecane	Diesel	0	0	0	0	2	16	16	0	0	0	0
Butylcyclohexane	Diesel	0	0	1	0	1	3	3	5	5	0	0
2-Methylundecane	Diesel	0	0	0	1	0	3	8	8	0	0	0
n-Butylbenzene	Diesel	6	3	0	0	0	1	3	3	0	0	0
Methylcyclohexane	Diesel	0	0	0	1	0	1	0	0	5	5	0
Methylhexadecane	Diesel	0	0	0	1	0	3	13	13	0	0	0
m-Xylene	Diesel	6	3	0	0	0	2	0	0	0	0	0
n-Butane	FACE C gasoline	0	0	0	0	0	2	2	2	0	0	0
2-Methylbutane	FACE C gasoline	0	0	0	1	0	3	1	1	0	0	0
2-Methylhexane	FACE C gasoline	0	0	0	1	0	3	3	3	0	0	0
n-Heptane	FACE C gasoline	0	0	0	0	0	2	5	5	0	0	0

2.2.2 Chemical kinetic modeling of POSF 4658/RME/Diesel/FACE C gasoline surrogates

The chemical kinetic mechanisms of POSF 4658/RME/Diesel/FACE C gasoline surrogates are built as hierarchical structure as shown in the red block of Figure 2.5. Thus, a modular approach can be employed to assemble the sub-mechanisms of different components into the target mechanisms of interest. The C0-C4 core mechanism is the base mechanism for POSF 4658/RME/Diesel/FACE C gasoline surrogates which contains H_2-O_2 sub-mechanism, C1-C4 (methane, ethane, ethylene, acetylene, formaldehyde, acetaldehyde, propane, propene, butane, etc.) sub-mechanism, CO sub-mechanism. C5-C6 sub-mechanism, NOx sub-mechanism, PRF/TPRF (n-heptane/iso-octane/toluene) sub-mechanism, alcohols sub-mechanism, aldehydes sub-mechanism, esters sub-mechanism, aromatics & PAH sub-mechanism, target fuel sub-mechanism and other necessary sub-mechanisms are combined with the C0-C4 core mechanism to produce the target surrogate mechanisms. The sub-mechanisms developed by Lawrence Livermore National Laboratory (LLNL) are chosen with priority to ensure the compatibility of the sub-mechanisms.

For the jet fuel POSF 4658, the Dooley 1st generation surrogate [124] and Dooley 2nd generation surrogate [125, 148] use the mechanisms provided in ref. [124] and ref. [30]. Especially, the GCM surrogate shares the same mechanism with the Dooley 2nd generation surrogate [125, 148] since their chemical compounds are identical.

For the biodiesel rapeseed methyl ester (RME), GCM1 surrogate components are recommended by Herbinet et al. [165], therefore, the corresponding mechanism is adopted in this study. The compositions of RME real fuel are revealed by Westbrook et al. [141] and the mechanism developed by Creck Modelling Group [166] is adopted for kinetic simulation. And the GCM2 surrogate and hexadecane utilize the Creck Mechanism [166] as well.

For the automotive fuel diesel, Pei surrogate [147] composes of 64.97%n-Dodecane-35.03%m-Xylene and the published mechanism is adopted. But there are no existing detailed mechanisms available for the GCM1 surrogate (22.06865%n-Dodecane [147]-33.11807%2-Methylundecane [147]-30.93333%n-Butylbenzene [146]-13.87995%Methylcyclohexane [134]) and GCM2 (19.666667%n-Hexadecane [91]-29.04%2-Methylhexadecane

[91]-30.933333%n-Butylbenzene   [146]-20.36%Methylcyclohexane   [134]) surrogate, therefore, kinetic

mechanisms need to be built for these two surrogates. For diesel GCM1 surrogate, the published n-dodecane

mechanism developed by Sarathy et al. [91] is used as a backbone mechanism and it also contains 2-

methylundecane. The oxidation of n-butylbenzene and methylcyclohexane are incorporated based on the

mechanisms developed by Nakamura et al. [146] and Pitz et. al. [134] and they all place on a higher hierarchical

level of $C_0 \sim C_4$ core module. For diesel GCM2 surrogate, the n-hexadecane and 2-methylhexadecane developed by

Sarathy et al. [91] act as the backbone mechanism and the n-butylbenzene [146], methylcyclohexane [134] modular

mechanisms are assembled into the skeleton mechanism.

The diesel GCM1 surrogate mechanism contains 3701 species, 15598 reactions and covers n-alkanes from n-

pentane to n-nonane, 2-methylalkanes from 2-Methylhexane to 2-Methylundecane, naphthenes including

cyclohexane, cyclopentane and methylcyclohexane, aromatics from benzene to n-butylbenzene. The diesel GCM2

surrogate mechanism contains 8050 species, 35703 reactions and covers n-alkanes from n-pentane to n-

heptadecane, 2-methylalkanes from 2-Methylhexane to 2-Methylhexadecane, naphthenes including cyclohexane,

cyclopentane and methylcyclohexane, aromatics from benzene to n-butylbenzene. These two mechanisms can be

downloaded from the supporting information. For the FACE C gasoline, the mechanism developed by Sarathy

[167] contains n-butane, 2-methyl butane, 2-methyl hexane, cyclopentane, 1,2,4-trimethylbenzene, 1-hexene, n-

heptane, 2,2,4-trimethylpentane, toluene.
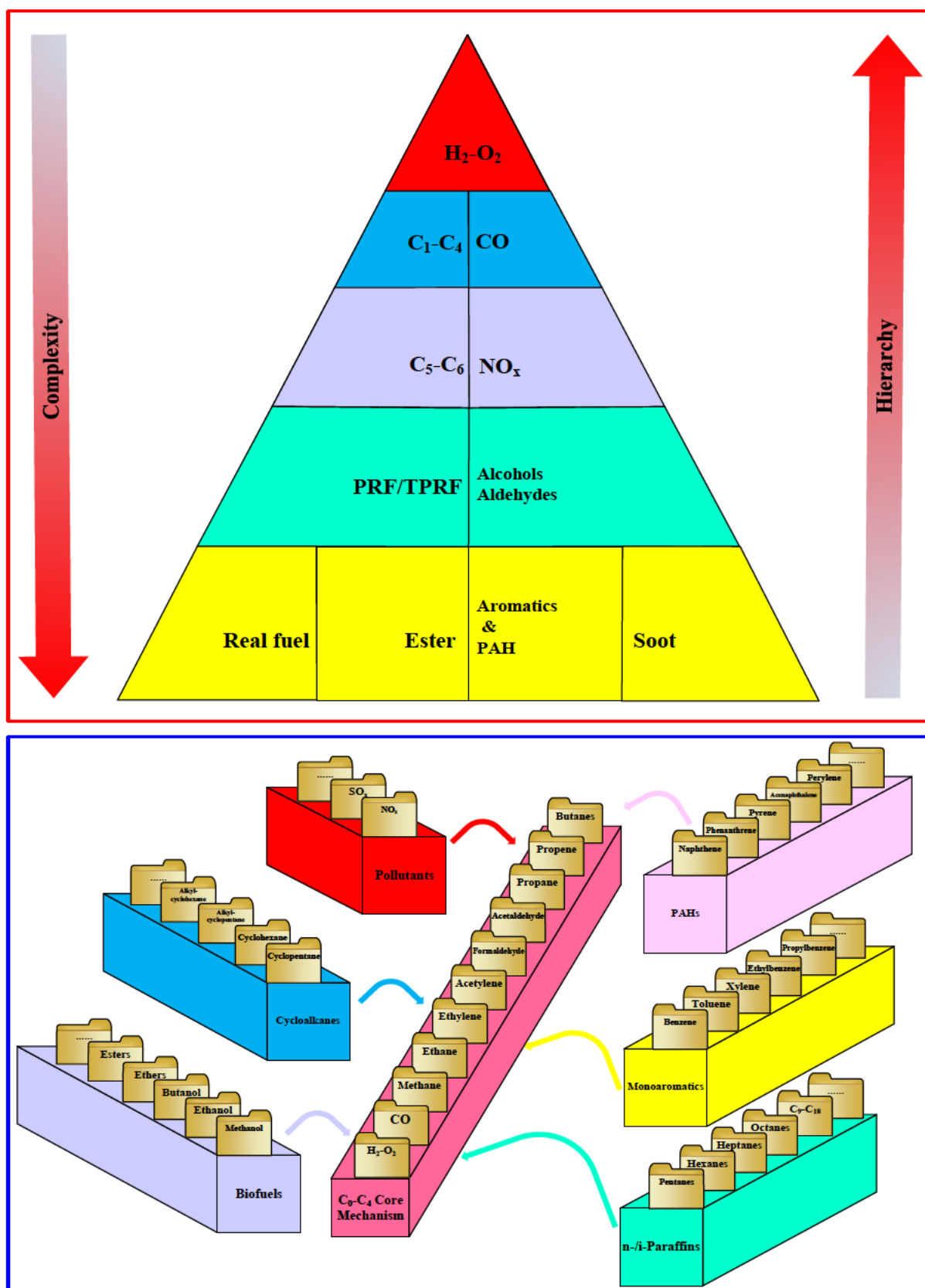
Figure 2.5. Hierarchical formulation of the combustion chemistry (red block) and modular component library framework for detail mechanism construction (blue block).

2.3 Results and discussion

The proposed GCM is applied to formulate surrogates of POSF 4658, RME, Diesel, FACE C gasoline in section 2.3.1, section 2.3.2, section 2.3.3, section 2.3.4 as case studies. The surrogate properties of ignition delay times, laminar flame speed, oxidation species profiles, density, sound speed, kinematic viscosity are compared with the target fuels as well as the published surrogates.

2.3.1 POSF 4658

2.3.1.1 Surrogate formulation of POSF 4658

POSF 4658 is a kerosene-based jet fuel and the main constituents (46 compounds) accounting for more than 40% have been measured by GC-MS [168]. Dooley et al. [124] choose n-decane, iso-octane and toluene as the surrogate palette and use DCN (derived cetane number), H/C ratio, MW (molecular weight), TSI (threshold sooting index) as target properties for surrogate formulation. The Dooley 1st generation surrogate composes of 42.7%n-decane-33.0%2,2,4-trimethylpentane-24.3%toluene, but it fails to simultaneously accommodate the H/C, DCN, TSI properties. Afterward, Dooley et al. [125] propose the 2nd generation POSF 4658 surrogate which replaces the n-decane with n-dodecane to increases the MW and replaces toluene with n-propylbenzene and 1,3,5-trimethylbenzene to raise the TSI. The 2nd generation POSF 4658 surrogate constitutes 40.41%n-dodecane-29.48%2,2,4-trimethylpentane-22.83%n-propylbenzene-7.28%1,3,5-trimethylbenzene.

The GCM decomposes the 46 components into typical functional group fragments following the GCM1.0 system (see Figure 2.3). The decomposed aromatic bond (No. 1 functional group fragment), carbon-carbon double bond (CCDB) in aromatic (No 2 functional group fragment), tertiary carbon (No. 6 functional group fragment), quaternary carbon (No. 7 functional group fragment), methyl radical (No. 8 functional group fragment), maximum quantity of methylene group in series for non-ring structure (No. 9 functional group fragment), methylene group for non-ring structure (No. 10 functional group fragment), maximum quantity of methylene group in series for ring structure (No. 11 functional group fragment), methylene group for ring structure (No. 12 functional group fragment)

account for 10.32 mol.%, 5.16 mol.%, 2.48 mol.%, 10.36 mol.%, 31.08 mol.%, 32.45 mol.%, 3.09 mol.%, 4.91

mol.% as shown in Figure 2.6. GCM surrogate chooses n-dodecane, 2,2,4-trimethylpentane, n-propylbenzene,

1,3,5-trimethylbenzene as surrogate palette which is identical to Dooley 2nd generation surrogate [125, 148]. GCM

surrogate formulation model is running to compute the compositions of POSF 4658 and the output result is

54.92%n-dodecane-12.61%2,2,4-trimethylpentane-25.53%n-propylbenzene-6.94%1,3,5-trimethylbenzene   as

shown in Table 2.6 and the proportions of the decomposed functional group fragments are presented in Figure 2.6.

The GCM surrogate cannot account for the contribution of maximum quantity of methylene group in series for

ring structure (No. 11 functional group fragment), methylene group for ring structure (No. 12 functional group

fragment) since it does not include cycloalkanes compared to the real fuel.
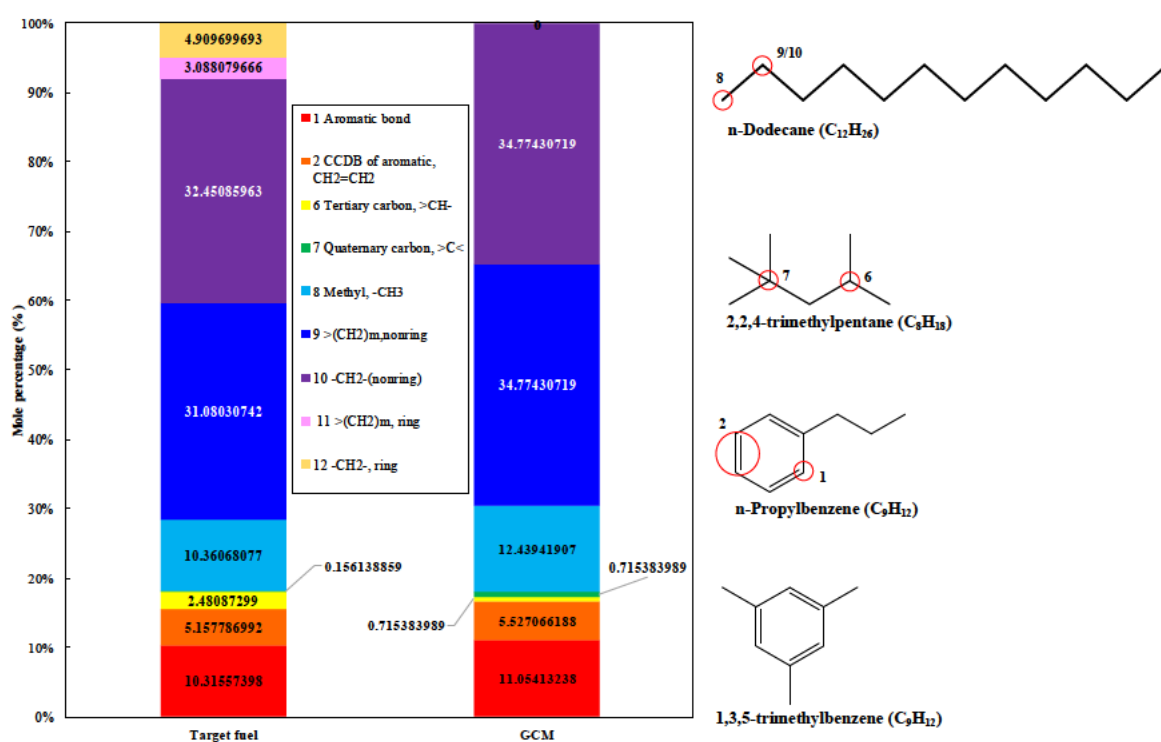
Figure 2.6. Functional group mole fractions for POSF 4658 target fuel [168] and GCM surrogate.

Table 2.6. POSF 4658 surrogate fuel compositions of Dooley 1st generation surrogate [124], Dooley 2nd generation surrogate [148] and GCM surrogate

POSF 4658 surrogate	Fuel compositions, mol.%, and mechanism source	Species/Reactions	Formula	H/C ratio
GCM surrogate	54.92%n-dodecane-12.61%2,2,4-trimethylpentane-25.53%n-propylbenzene-6.94%1,3,5-trimethylbenzene [148]	2080/8310	$C_{10.52}H_{20.45}$	1.9439
Dooely 1st surrogate [124]	42.7%n-decane-33.0%2,2,4-trimethylpentane-24.3%toluene [124]	1599/6633	$C_{8.61}H_{17.28}$	2.0070
Dooley 2nd surrogate [125, 148]	40.41%n-dodecane-29.48%2,2,4-trimethylpentane-22.83%n-propylbenzene-7.28%1,3,5-trimethylbenzene [148]	2080/8310	$C_{9.92}H_{19.24}$	1.9395

2.3.1.2 Ignition delay times emulation

The simulated ignition delay times of POSF 4658 jet fuel by GCM surrogate, Dooley 1st generation surrogate, Dooley 2nd generation surrogate are compared with the observed values (by shock tube and rapid compression machine) at φ=1.0, P_{init}=20 atm as shown in Figure 2.7. The initial gas fractions of Dooley 1st generation surrogate [124], Dooley 2nd generation surrogate[125, 148] and GCM surrogate are presented in Table 2.7 and the diluent (N2) concentration keeps constant as 77.92 mol.% for comparison. The GCM surrogate and Dooley 2nd generation surrogate[125, 148] outperform Dooley 1st generation surrogate [124] because they are closer to the measured ignition delay time. The GCM surrogate and Dooley 2nd generation surrogate [125, 148] almost coincide at the temperature regimes of 570~710K and 970~1250K and both accurately capture the NTC (negative temperature coefficient) behavior at the temperature range of 770~950K as shown in Figure 2.7. The GCM surrogate estimates a shorter ignition delay time at 710~970K than Dooley 2nd generation surrogate [125, 148] which is closer to the experimental profiles. But both GCM surrogate and Dooley 2nd generation surrogate [125, 148] significantly overestimate the ignition delay time at 930~1250K and this issue cannot be addressed by compositional proportion optimization alone. To further improve the ignition delay time predictive accuracy at 930~1250K, the measures are either to replace the existing palette compounds with new constituents or expand the number of surrogate palettes to enhance high temperature reactivity. Increasing the number of surrogate compounds will benefit the physicochemical properties reproduction at the expense of additional complexity.
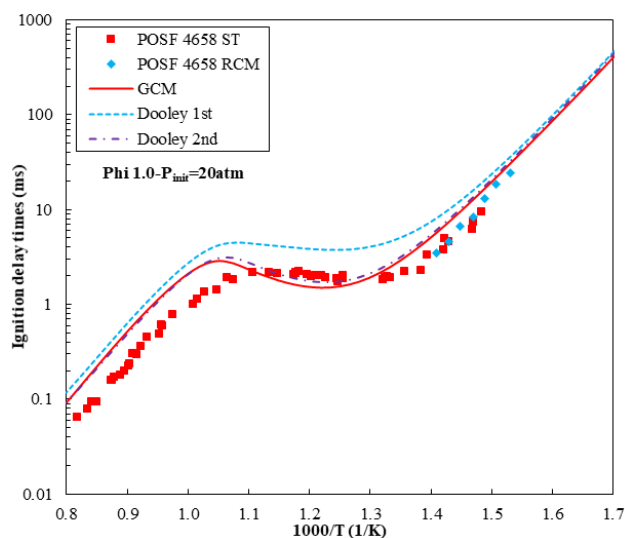
Figure 2.7. Ignition delay times of POSF 4658 (symbols) and surrogates simulation (lines) at φ=1.0, P_{init}=20 atm. Square and diamond symbols correspond to shock tube and rapid compression machine measurements [124]; solid line, dashed line and dash-dotted line correspond to GCM surrogate, Dooley 1st generation surrogate [124], Dooley 2nd generation surrogate [125, 148].

Table 2.7. Initial gas fractions of POSF 4658 surrogates (Dooley 1st generation surrogate [124], Dooley 2nd generation surrogate [125, 148] and GCM surrogate) for ignition delay time simulation at $\varphi=1.0$, $P_{init}=20$ atm

Compositions (mol.%)	Fuel				Oxidizer	Diluent
	n-Decane	2,2,4-Trimethylpentane	Toluene		O_2	N_2
Dooley 1st surrogate [124]	0.67680	0.52305	0.38516		20.49499	77.92
	n-Dodecane	2,2,4-Trimethylpentane	n-Propylbenzene	1,3,5-Trimethylbenzene	O_2	N_2
Dooley 2nd surrogare [125, 148]	0.56565	0.41265	0.31957	0.10190	20.68023	77.92
GCM surrogate	0.72901	0.16738	0.33888	0.09212	20.75261	77.92

2.3.1.3 Laminar flame speed emulation

The laminar flame speeds of POSF 4658-air mixture are measured at φ=0.6~1.6, T_{init}=400K, 470K, P_{init}=1 atm by Dooley et al. [124] and the predictive performances of GCM surrogate (solid line), Dooley 2nd generation surrogate (dashed line) [125, 148] are presented in Figure 2.8 (a). The Dooley 1st generation surrogate [124] does not provide the transport property of the species in the published model, so it is excluded in the flame speed validation. At T_{init}=400K, the Dooley 2nd generation surrogate [125, 148] can better reproduce the laminar flame speed at φ=0.7~1.4 than GCM surrogate and reaches higher R^2 as 0.9312 as shown in Figure Figure 2.8 (b). The GCM surrogate obtains better predictive accuracy than Dooley 2nd generation surrogate [125, 148] at T_{init}=470K and the R^2 reaches 0.9848 as shown in Figure Figure 2.8 (c). But both surrogates significantly overestimate the laminar flame speed at φ=0.9~1.3 and the maximal deviation reaches 14.67 cm/sec at φ=1.1 by GCM surrogate. The future work of surrogate optimization should improve the laminar flame speed at φ=0.9~1.3 and high initial mixture temperature (for example around 470K).
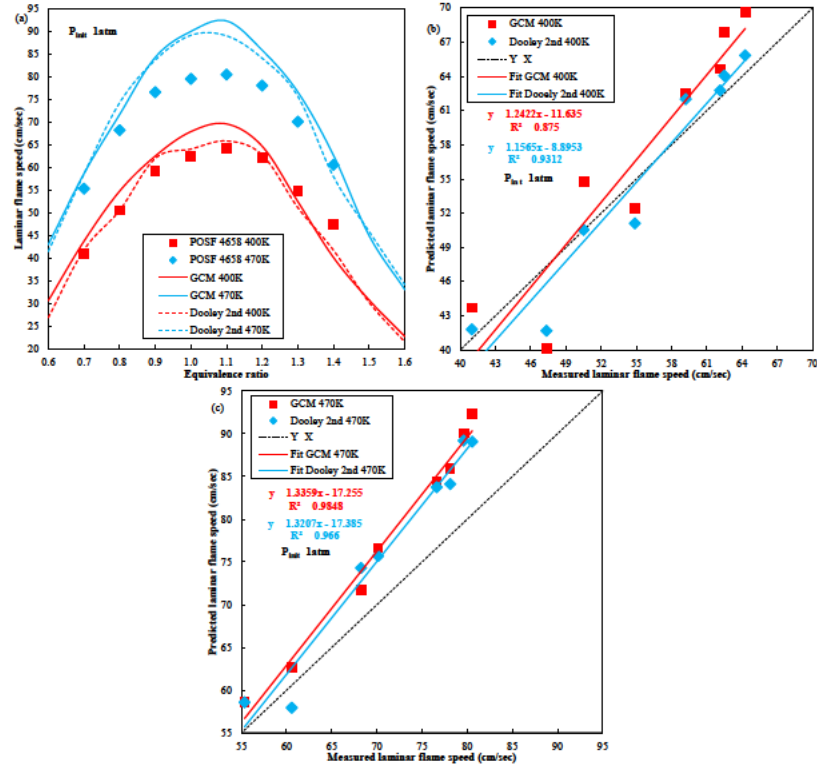
Figure 2.8. (a) Laminar flame speeds of POSF 4658-air mixture (symbols) [148] and surrogates (lines) at φ=0.6~1.6, T_{init}=400K, 470K, P_{init}=1 atm; solid line and dashed line correspond to GCM surrogate, Dooley 2nd generation surrogate [125, 148]; parity plot between measured and simulated values at (b) T_{init}=400K; (c) T_{init}=470K.

2.3.1.4 Flow reactor oxidation emulation

The CO, CO_2, O_2, H_2O profiles of POSF 4658 oxidation in variable pressure flow reactor are measured at 0.3% carbon, $\varphi = 1.0$, $T_{reactor}=500\sim1100K$, $P_{int}=12.5$ atm, $\tau_{resident}=1.8s$, $V=250cm^3$ by Dooley et al. [124]. The predictive species profiles by GCM surrogate, Dooely 1st generation surrogate [124], Dooley 2nd generation surrogate [125, 148] are compared with observed values as plotted in Figure 2.9. The initial gas fractions of these surrogates are provided in Table 2.8 and the carbon atom concentration keeps constant as 0.3 mol.% following the experimental program published by Dooley et al. [124]. The GCM surrogate accurately captures the POSF 4658 oxidation onset of 580K and the subsequent low temperature oxidation behavior at 580~640K. Both GCM surrogate and Dooley 2nd generation surrogate [125, 148] overestimate the oxidation reactivity at NTC regime of 640~780K compared to Dooely 1st generation surrogate [124], thus the O_2 concentration and $CO/CO_2/H_2O$ concentrations are lower and greater than the observed concentrations respectively. This is because the GCM surrogate replaces n-decane with n-dodecane and increases the proportion from 42.7mol.% to 54.92mol.% Dooely 1st generation surrogate [124] as shown in Table 2.6. GCM surrogate and Dooley 2nd generation surrogate [125, 148] almost overlap at high temperature oxidation regime of 780~1100K, but they underestimate the fuel oxidation reactivity at 880~1040K. Therefore, the predictive species of O_2 and $CO/CO_2/H_2O$ are higher and lower than the measured values respectively as shown in Figure 2.9. Even though the Dooely 1st generation surrogate [124] fails to capture the onset of high temperature oxidation onset at 800K, it better reproduces the high temperature oxidation reactivity than other surrogates at 960~1100K. The GCM surrogate replaces toluene with n-propylbenzene and 1,3,5-trimethylbenzene and increases the aromatic proportion compared to Dooely 1st generation surrogate [124] which reduces the high temperature oxidation reactivity.
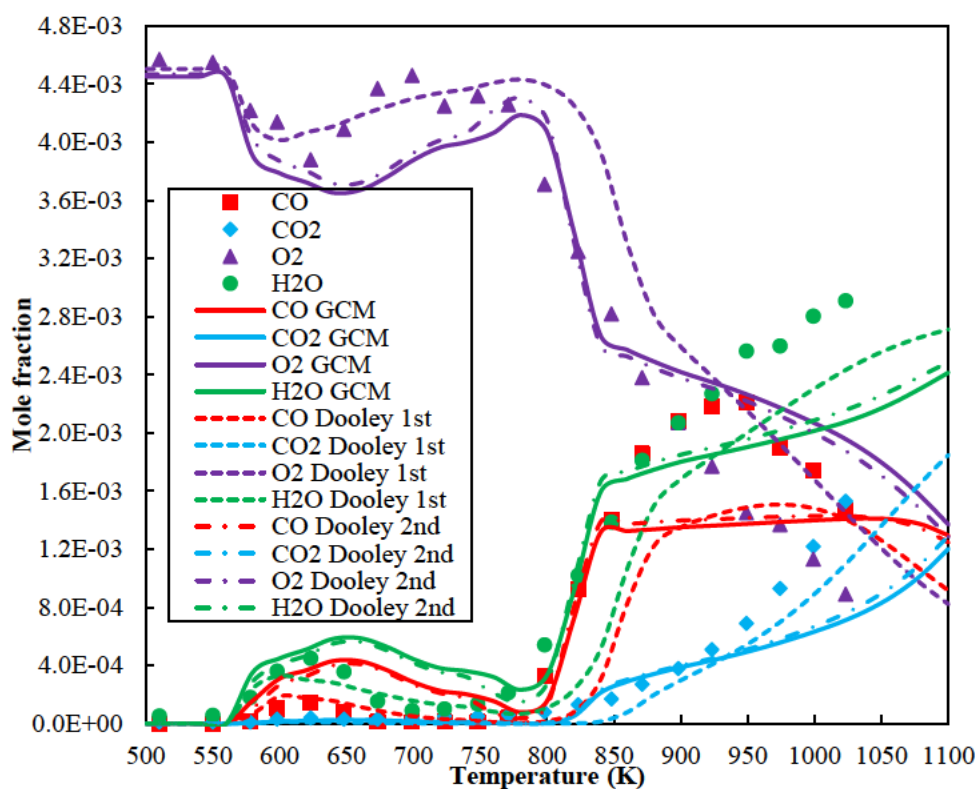
Figure 2.9. Speciation in flow reactor of POSF 4658 (symbols) [124] and surrogates simulation (lines) at 0.3% carbon, $\varphi = 1.0$, $T_{reactor}=500\sim1100K$, $P_{int}=12.5$ atm, $\tau_{resident}=1.8s$, $V=250cm^3$, solid line, dashed line and dash-dotted line correspond to GCM surrogate, Dooely 1st generation surrogate [124] , Dooley 2nd generation surrogate [125, 148].

Table 2.8. Initial gas fractions of POSF 4658 surrogates (Dooley 1st generation surrogate [124], Dooley 2nd generation surrogate[125, 148] and GCM surrogate) for flow reactor simulation at $\varphi = 1.0$, $T_{reactor}$=500~1100K, P_{int}=12.5 atm

Composition (mol.%)	Fuel (C=0.3mol.%)				Oxidizer	Diluent
	n-Decane	2,2,4-Trimethylpentane	Toluene		O_2	N_2
Dooley 1st surrogate [124]	0.014876321	0.011496923	0.008465916		0.450487748	99.51467309
	n-Dodecane	2,2,4-Trimethylpentane	n-Propylbenzene	1,3,5-Trimethylbenzene	O_2	N_2
Dooley 2nd surrogate [125, 148]	0.012223847	0.00891757	0.006905974	0.002202168	0.446908495	99.52284195
GCM surrogate	0.015644049	0.003591979	0.007272261	0.00197687	0.445342676	99.52617216

2.3.1.5 Liquid phase density, sound speed and kinematic viscosity emulation

The measured liquid phase density, sound speed, kinematic viscosity of POSF 4658 jet fuel at P_{int}=0.083MPa

are reported by Bruno et al. [169] and the predictive values of GCM surrogate, Dooley 1st generation surrogate

[124], Dooley 2nd generation surrogate[125, 148] are computed by NIST REFPROP (Reference Fluid

Thermodynamic and Transport Properties Database) program version 9.1 [170] as shown in Figure 2.10.

REFPROP calculates the mixture thermodynamic properties by applying the mixing rule to the equation of state

in Helmholtz energy of mixture constituents. The discrepancy between the real mixture and ideal mixing is

considered by the departure function. The viscosity is estimated by the extended corresponding states (ECS) model.

For the parameters of liquid phase density, sound speed and kinematic viscosity, the predictive accuracy of

surrogates arranges from high to low as: GCM surrogate > Dooley 2nd generation surrogate[125, 148] > Dooley

1st generation surrogate [124].
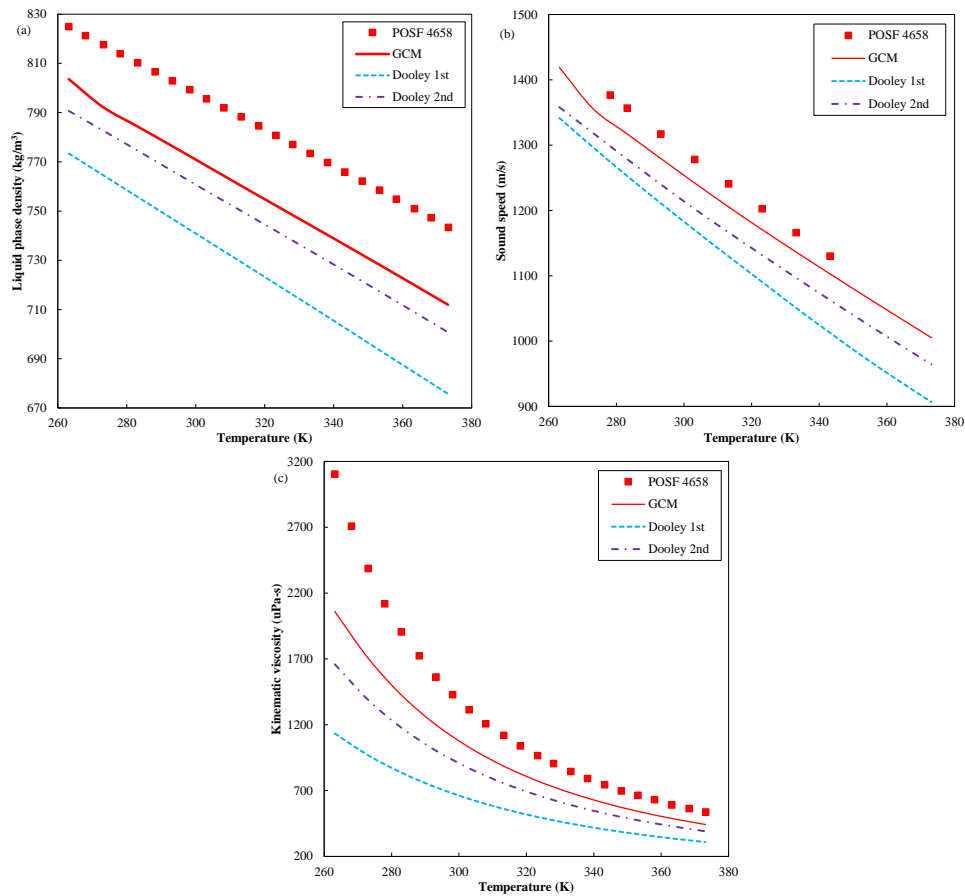
Figure 2.10. (a) Liquid density, (b) sound speed and (c) kinematic viscosity of POSF 4658 (symbols) [169] and
surrogates simulation (lines) at P_{int}=0.083MPa, solid line, dashed line and dash-dotted line correspond to GCM
surrogate, Dooley 1st generation surrogate [124], Dooley 2nd generation surrogate [125, 148].

2.3.2 Rapeseed methyl ester (RME)

2.3.2.1 Surrogate formulation of RME

Rapeseed methyl ester (RME) biodiesel is a mixture of FAME (fatty acid methyl ester) containing 4.3%methyl palmitate-1.3%methyl stearate-59.9%methyl oleate-21.2%methyl linoleate-13.3%methyl linolenate [141] and the CRECK mechanism [166] is used to simulate the real fuel oxidation. The RME has high average carbon atom numbers of 18.91 as shown in Table 2.9 and the surrogate formulation needs to balance the accuracy of having comparable MW and complexity of the chemical kinetic model. Dagaut et al. [171] simply employ n-hexadecane as RME surrogate which contains comparable carbon atom numbers, but the contribution of the ester group to the oxidation process is omitted. To develop the GCM surrogates, the pure compounds in the RME real fuel are disassembled into CCDB of non-aromatic for non-ring structure (No. 4 functional group fragments), methyl group (No. 8 functional group fragments), maximum quantity of methylene group in series for non-ring structure (No. 9 functional group fragment), methylene group for non-ring structure (No. 10 functional group fragment), ester group (No. 10 functional group fragment) which account for 5.71 mol.%, 29.78 mol.%, 52.47 mol.%, 4.01 mol.%. Two GCM surrogates (GCM1, GCM2) are designed for the RME which contain different levels of MW. The GCM1 surrogate introduces methyl decanoate, methyl-9-decenoate, methyl-5-decenoate as the surrogate palettes to address the required functional group fragments. The carbon atom number of each compound is limited to 11 to avoid the complex chemical kinetic model. On the contrary, the GCM2 surrogate replaces methyl-9-decenoate with methyl decanoate to increases the average MW at the expense of increasing the complexity of the corresponding kinetic model. The compositions of GCM1 and GCM2 surrogates are formulated as 59.7727%methyl decanoate-29.7394%methyl-9-decenoate-10.4879%methyl-5-decenoate and 37.42749%n-hexadecane-38.46876%methyl decanoate-24.10376%methyl-5-decenoate as shown in Table 2.9 and the proportions of decomposed functional group fragments are demonstrated in Figure 2.11 and Figure 2.12. After replacing the palette compounds, the average formula of GCM1 and GCM2 surrogates are C11H21.20O2 and C12.87H26.01O1.25 respectively. Their predictive performance on ignition delay time and oxidation in the jet-stirred reactor are examined in section 2.3.2.2 and section 2.3.2.3. The experimental data of liquid density, sound speed and kinematic viscosity for RME are not available for validation. Furthermore, the RME components of

methyl palmitate, methyl stearate, methyl oleate, methyl linoleate, methyl linolenate are not in the substance database of REFPROP software and thus the computational values are not available as well. Therefore, theses 3 properties cannot be used to validate RME surrogate at their current status.
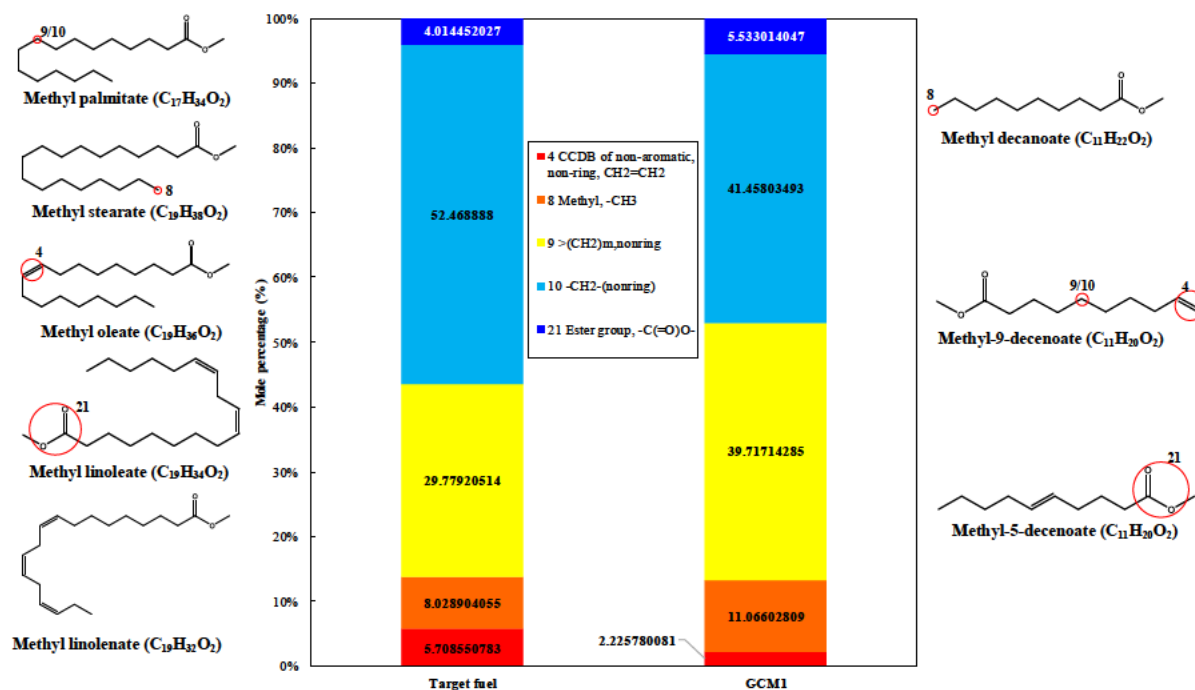
Figure 2.11. Functional group mole fractions for RME target fuel [141] and GCM1 surrogate.
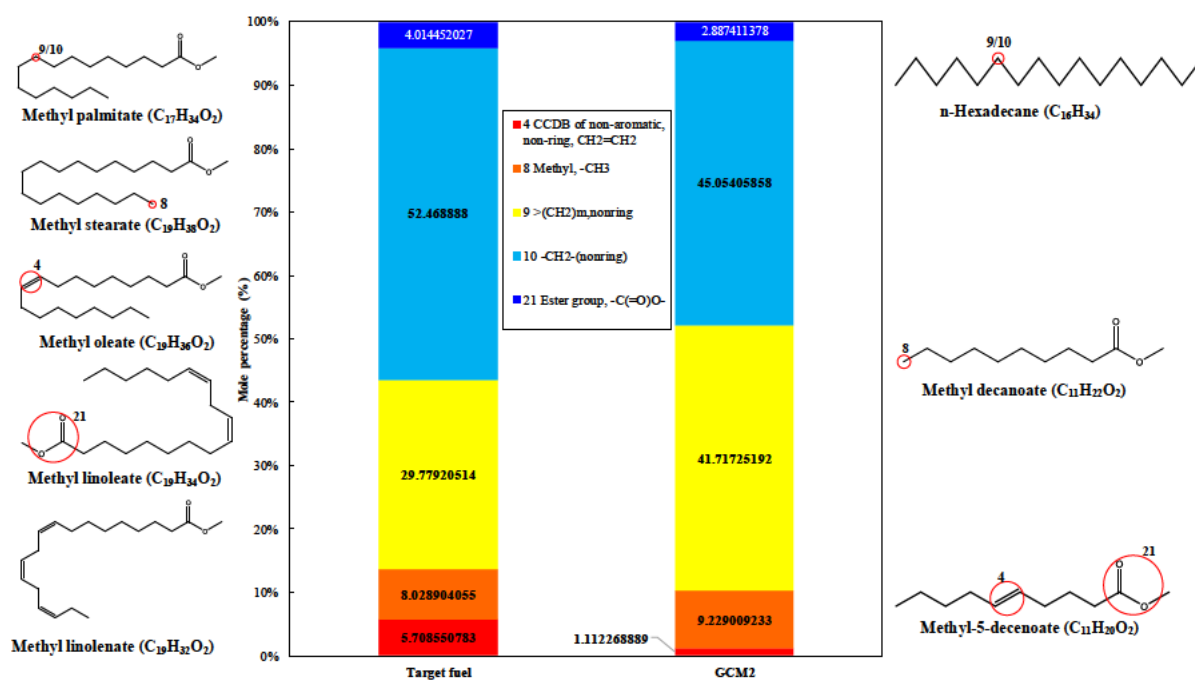
Figure 2.12. Functional group mole fractions for RME target fuel and GCM2 surrogate.

Table 2.9. RME surrogate fuel compositions of GCM1 surrogate, GCM2 surrogate, RME Real fuel [141] and Hexadecane [171]

RME surrogate	Fuel compositions, mol.% and mechanism source	Species/Reactions	Formula	H/C, O/C ratios
GCM1 surrogate	59.7727%methyl decanoate-29.7394%methyl-9-decenoate-10.4879%methyl-5-decenoate [165]	3299/10806	$C_{11}H_{21\,20}O_2$	1.9273, 0.1818
GCM2 surrogate	37.42749%n-hexadecane-38.46876%methyl decanoate-24.10376%methyl-5-decenoate [166]	484/19341	$C_{12\,87}H_{26\,01}O_{1\,25}$	2.0210, 0.0971
RME real fuel [141]	4.3%methyl palmitate-1.3%methyl stearate-59.9%methyl oleate-21.2%methyl linoleate-13.3%methyl linolenate [166]	484/19341	$C_{18\,91}H_{34\,98}O_2$	1.8498, 0.1058
Dagaut surrogate [171]	100%hexadecane [166]	484/19341	$C_{16}H_{34}$	2.125, 0

2.3.2.2 Ignition delay times emulation

The measured ignition delay time of RME is not reported in the literature because its high distillation temperatures (320~360°C) and kinematic viscosity (6.9~8.2mm²/s@20°C) [172] make it difficult to vaporize into gas phase to be tested in shock tube and rapid compression ignition machine. Hence, Westbrook et al. [141] develop a detailed chemical kinetic mechanism for RME and simulate the ignition delay times at $\varphi=1.0$, $T_{init}=625~1250K$, $P_{init}=13.5bar$ which are used as a benchmark in this section (the symbols in Figure 2.13). The predicted ignition delay times of GCM1 surrogate, GCM2 surrogate, real fuel (CRECK model [166]), Dagaut surrogate (hexadecane) [171] are validated against Westbrook mechanism [141] as shown in Figure 2.13. The ignition delay times of real fuel (CRECK model [166]) almost overlap with RME data predicted by Westbrook mechanism [141] because it contains all the RME components of   methyl palmitate, methyl stearate, methyl oleate, methyl linoleate, methyl linolenate. The surrogate predictive accuracy orders from high to low as: GCM2 ($R^2=0.98$) >GCM1 ($R^2=0.9744$) >Dagaut surrogate [171] ($R^2=0.8134$). The GCM2 surrogate accurately reproduces the high temperature ignition at 900~1238K. It overestimates the fuel reactivity at low temperature ignition regime at 615~770K and NTC behavior at 770~888K because the proportion of maximum quantity of methylene group in series for non-ring structure (No. 9 functional group fragment) is 11.94% higher than the target fuel which is the principal influence factor of fuel reactivity. Similar to GCM2 surrogate, the high temperature ignition at 900~1238K is successfully captured by GCM1 surrogate. The low temperature ignition delay time at 575~800K of GCM1 surrogate is closer to the taegett fuel than GCM2 surrogate and Dagaut surrogate [171] because it obtains the minimal propsotional discrepancy of No. 9 functional group fragment. Even thoguh the GCM1 surrogate overestmates the intermediates temperature ignition at 800~900K, the slope of the NTC regime is acucrately captured by GCM1 surrogate due to the minor deviation of No. 9 functional group fragment. As expected, the Daugaut surrogate [171] significantly overestimates the fuel reactivity at low temperature ignition regime of 625~800K and intermediate temperature ignition regime of 800~950K because the proportion of No. 9 functional

group fragment accounts for 46.67% which is even higher than GCM2 surrogate (41.72%).
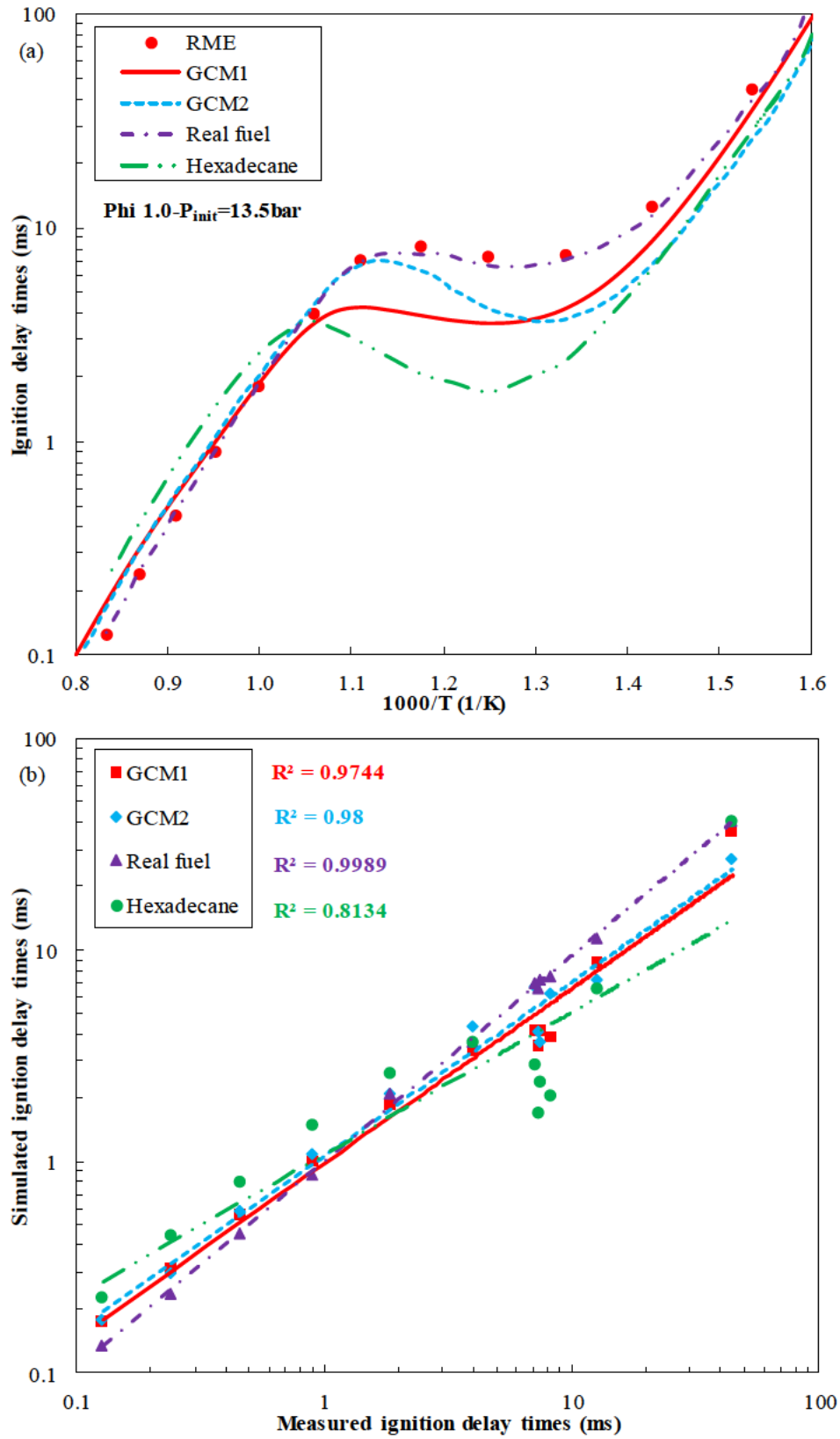
Figure 2.13. (a) Ignition delay times of RME-air mixture LLNL (symbols) [141] and surrogates simulation (lines) at φ=1.0, T_{init}=625~1250K, P_{init}=13.5bar, solid line, dashed line, dashed-dotted line and double dotted dashed line correspond to GCM1 surrogate, GCM2 surrogate, real fuel [141], hexadecane surrogate [171]; (b) Comparison between measured and simulated ignition delay times.

2.3.2.3 Jet stirred reactor oxidation emulation

Daugaut et al. [171] measured the major and intermediate species of RME oxidation in JSR (jet-stirred reactor) at a wide range of conditions to reveal the fuel oxidation reactivity and species evolution at low/intermediate/high temperature regimes for chemical kinetic model development. The predicted species profiles of GCM1 surrogate, GCM2 surrogate, Daugaut surrogate [171], the real fuel (CRECK model [166]) are applied to mimic the RME oxidation in JSR at the condition of T_{init}=500~1600K, P_{init}=1atm, V=30.5cm^3 and Figure 2.14, Figure 2.15, Figure 2.16, Figure 2.17 correspond to φ=0.25, $\tau_{resident}$=0.07s; φ=0.5, $\tau_{resident}$=0.07s; φ=1.0, $\tau_{resident}$=0.07s; φ=1.5, $\tau_{resident}$=0.1s respectively. The initial gas fractions of RME surrogates are provided in Table 2.10 and the fuel concentration keeps constant as 0.05 mol.% with nitrogen dilution. The real fuel (CRECK model [166]) can accurately predict the major species of H2, CO, CO2, CH2O (formaldehyde) at the studied conditions, Daugaut surrogate [171] comes second, GCM2 surrogate is next and GCM1 surrogate obtains the worst accuracy. This is because the real fuel (CRECK model [166], $C_{18\,91}H_{34\,98}O_2$) contains identical components and the Daugaut surrogate [171] has a comparable chemical formula as $C_{16}H_{34}$, while there are great discrepancies in formula between GCM2 surrogate ($C_{12\,87}H_{26\,01}O_{1\,25}$), GCM1 surrogate ($C_{11}H_{21\,20}O_2$) and target fuel. For similar reasons, the real fuel (CRECK model [166]) and Daugaut surrogate [171] achieve better predictive accuracy than GCM1 surrogate, GCM2 surrogate for intermediate concentration of CH4 (methane), C2H6 (ethane), C2H2 (acetylene), C2H4 (ethylene), C3H6 (propene), C4H6 (but-1-yne) as shown in Figure 2.14. There is no measured data available in the literature for the oxygenated species of methanol (CH3OH), acetaldehyde (CH3CHO), propionaldehyde (C2H5CHO) and acetone (CH3COCH3), thus the real fuel (CRECK model [166]) is regarded as the baseline for comparison. The peak concentration and corresponding temperature are broadly captured by the GCM2 surrogate and Daugaut surrogate [171] while the GCM1 surrogate fails to reproduce the correct orders of magnitude for propionaldehyde (C2H5CHO) and acetone (CH3COCH3) concentration due to low average carbon atom numbers. In summary, the speciation predictive performance orders from high to low as: real fuel (CRECK model [166]) >

Daugaut surrogate [171] > GCM2 surrogate > GCM1 surrogate. The GCM surrogate formulation enables to utilize simple compounds (such as methyl decanoate, methyl-9-decenoate, methyl-5-decenoate) to reproduce the gas phase combustion process of complex fuel mixtures (such as methyl palmitate, methyl stearate, methyl oleate, methyl linoleate, methyl linolenate) by providing similar proportions of functional group fragments compared. The ester group in the molecule forms asymmetric structure and thus a larger numbers of species, reactions are required to describe the gas phase combustion process than corresponding n-alkanes with identical carbon atom numbers.
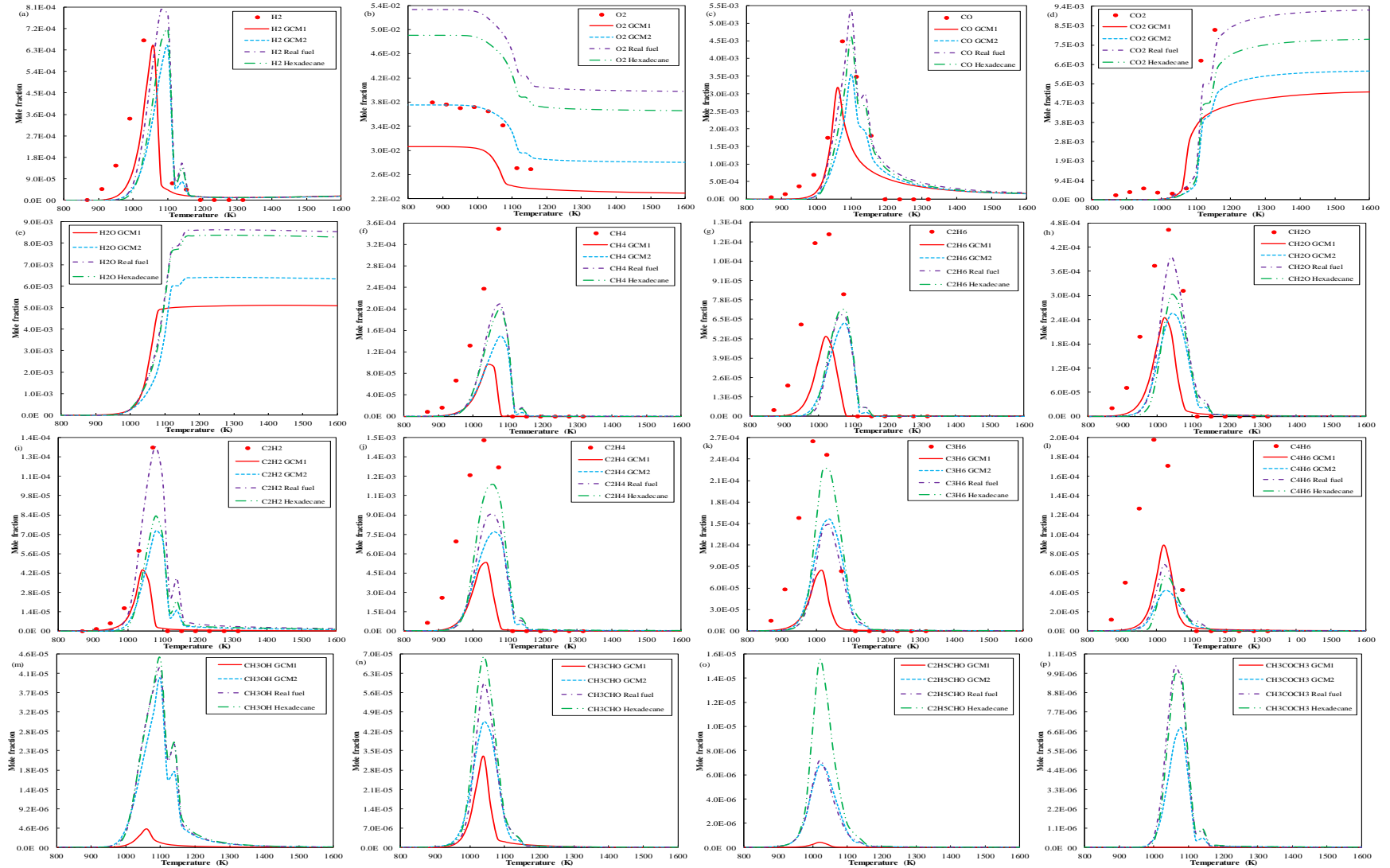
Figure 2.14. Speciation in JSR of RME (symbols) [171] and surrogates simulation (lines) at $\varphi=0.25$, $T_{init}=500\sim1600K$, $P_{init}=1atm$, $\tau_{resident}=0.07s$, $V=30.5cm^3$, 0.05mol.% fuel diluted by nitrogen, solid line, dashed line, dash-dotted line and double dot-dash line correspond to GCM1 surrogate, GCM2 surrogate, real fuel [141] and hexadecane [171].
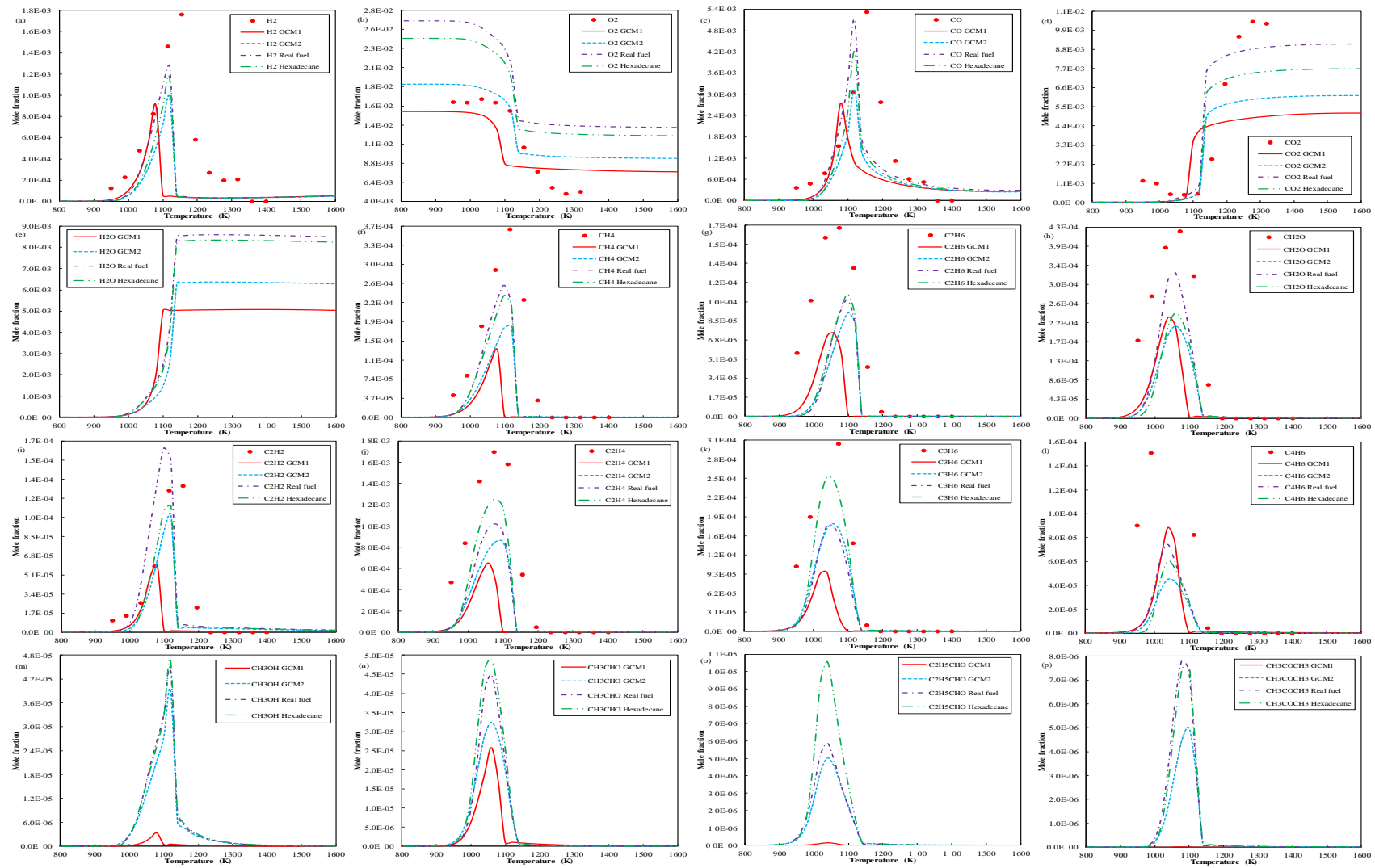
Figure 2.15. Speciation in JSR of RME (symbols) [171] and surrogates simulation (lines) at φ=0.5, T_{init}=500~1600K, P_{init}=1atm, $\tau_{resident}$=0.07s, V=30.5cm^3, 0.05mol.% fuel diluted by nitrogen, solid line, dashed line, dash-dotted line and double dot-dash line correspond to GCM1 surrogate, GCM2 surrogate, real fuel [141] and hexadecane surrogate [171].
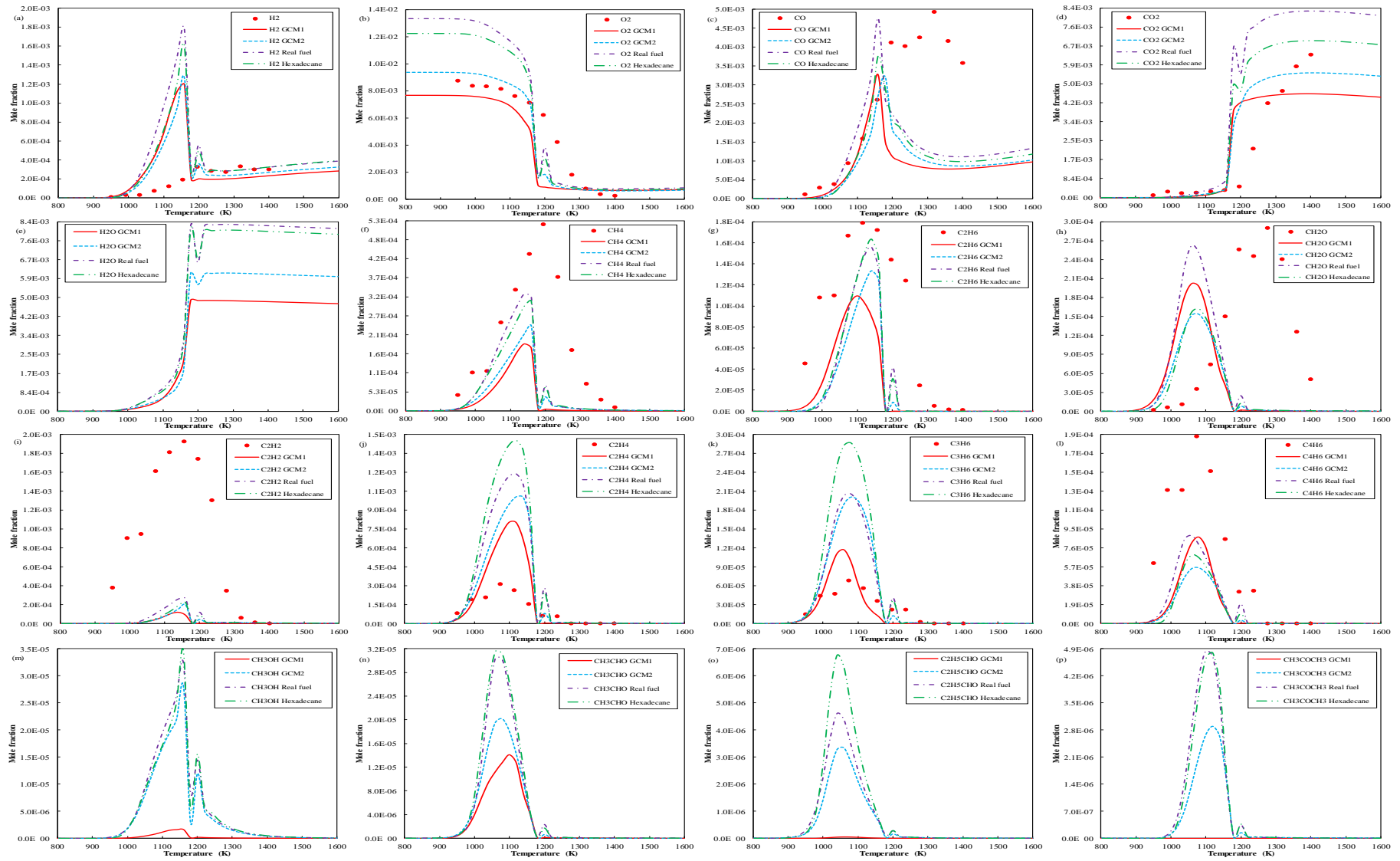
Figure 2.16. Speciation in JSR of RME (symbols) [171] and surrogates simulation (lines) at φ=1.0, T_{init}=500~1600K, P_{init}=1atm, 0.05mol.% fuel diluted by nitrogen, solid line, dashed line, dash-dotted line and double dot-dash line correspond to GCM1 surrogate, GCM2 surrogate, real fuel [141] and hexadecane surrogate [171].

63
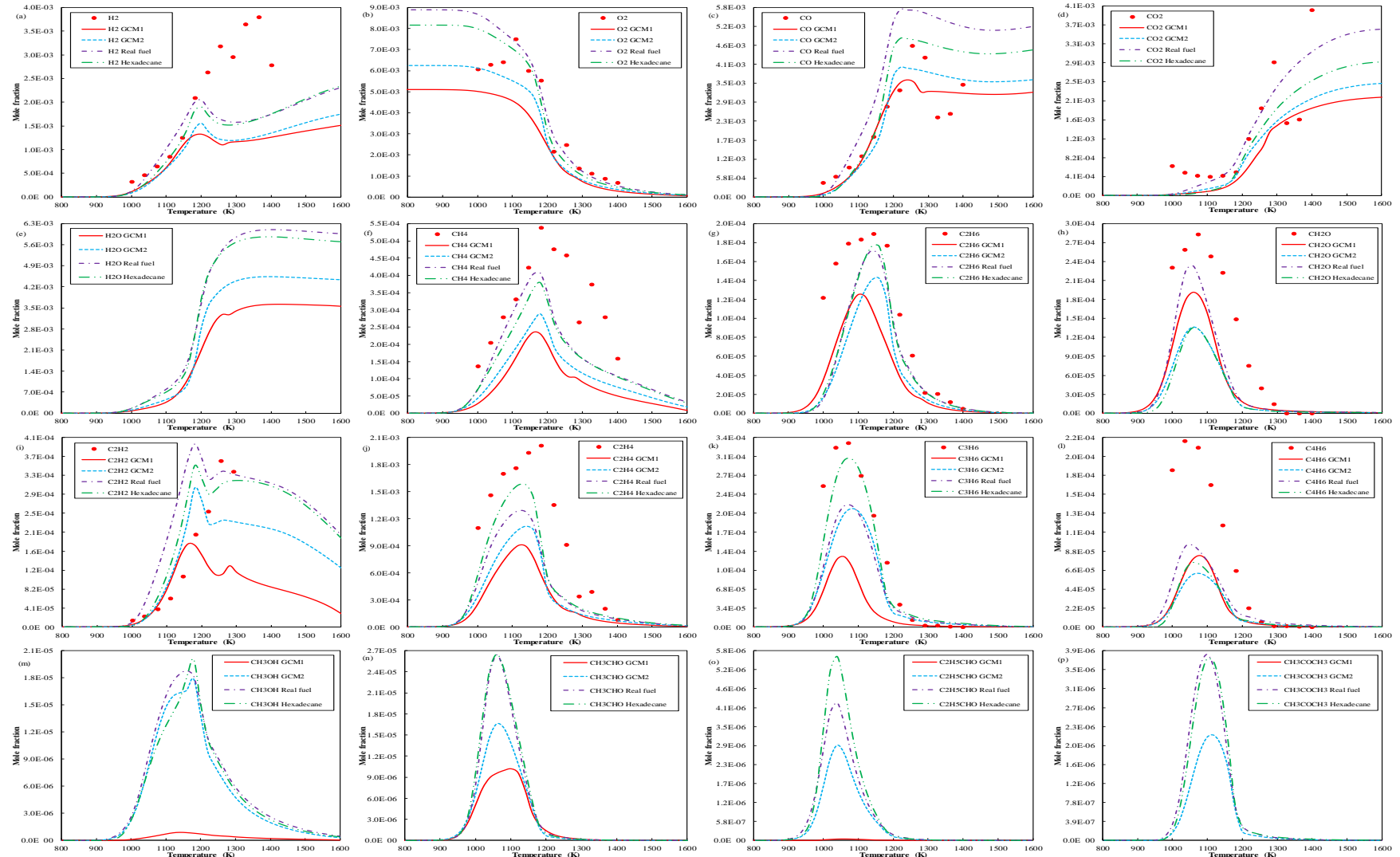
Figure 2.17. Speciation in JSR of RME (symbols) [171] and surrogates simulation (lines) at $\varphi=1.5$, $T_{init}=500\sim1600K$, $P_{init}=1atm$, $\tau_{resident}=0.1s$, $V=30.5cm^3$, 0.05mol.% fuel diluted by nitrogen, solid line, dashed line, dash-dotted line and double dot-dash line correspond to GCM1 surrogate, GCM2 surrogate, real fuel [141] and hexadecane surrogate [171].

Table 2.10. Initial gas fractions of RME surrogates (GCM1 surrogate, GCM2 surrogate, real fuel [141] and hexadecane surrogate [171]) for jet stirred reactor simulation at T_{init}=500~1600K, P_{init}=1atm

φ	$\tau_{resident}$	X_{Fuel} (mol.%)	X_{O2} (mol.%, GCM1/GCM2/real fuel [141]/hexadecane surrogate [171])	X_{N2} (mol.%, GCM1/GCM2/real fuel [141]/hexadecane surrogate [171])
0.25	0.07	0.05	3.059775006/3.749591059/5.332/4.9	96.89022499/96.20040894/94.618/95.05
0.5	0.07	0.05	1.529887503/3.749591059/2.666/2.45	98.4201125/98.07520447/97.284/97.5
1	0.07	0.05	0.764943751/0.937397765/1.333/1.225	99.18505625/99.01260224/98.617/98.725
1.5	0.1	0.05	0.509962501/0.624931843/0.888666667/0.816666667	99.4400375/99.32506816/99.06133333/99.13333333

2.3.3 Diesel

2.3.3.1 Surrogate formulation of diesel

High-resolution chemcials identification of commercial petroleum can be performed by GC-MS with the aid of NIST (National Institute of Standards and Technology)/EPA (United States Environmental Protection Agency)/NIH (National Institutes of Health) mass spectral library, NIST Tandem mass spectral library as well as NIST GC method/retention index library [173]. But the detailed compounds and corresponding proportion of diesel are not reported in the literature which fails to meet the requirement of step 1 of GCM surrogate formulation (see Figure 2.2). To work out a compromise, Qian surrogate [174] containing 21.5%n-hexadecane-25.8%2,2,4,4,6,8,8-heptamethylnonane(HMN)-13.7%1-methylnaphthalene%-8.1%decalin-15.4%n-octadecane-8.1%n-butylbenzene-7.4%n-butylcyclohexane is regarded as target fuel in this section because it has been approved in the single-cylinder compression ignition engine and is composed of up to 7 components to represent the real fuel. There are 4 chemical families of n-alkanes, iso-alkanes, aromatics, cycloalkanes in Qian surrogate [174] to represent the China stage V 0# diesel, thus the GCM1 surrogate and GCM2 surrogate contain 4 components and each of them represents the corresponding chemical class. The Qian surrogate [174] probably underestimates the molecular weight (198.2g/mol vs 206 ± 20g/mol [175]) and overestimates the H/C ratio (1.9268 vs 1.82 ± 0.1 [175]) of the commercial diesel. GCM1 surrogate select n-dodecane, 2-methylundecane, n-butylbenzene, methylcyclohexane as surrogate palette as the simpler case while the GCM2 surrogate replaces n-dodecane with n-hexadecane to increases the average MW for better representing the real fuel. The target fuel of Qian surrogate [174] is decomposed into 8 functional group fragments of aromatic bond (No. 1 functional group fragment), CCDB of aromatic (No. 2 functional group fragment), tertiary carbon (No. 6 functional group fragment), methyl group (No. 8 functional group fragment), maximum quantity of methylene group in series for non-ring structure (No. 9 functional group fragment), methylene group for non-ring structure (No. 10 functional group fragment), maximum quantity of methylene group in series for ring structure (No. 11 functional group fragment), methylene group for

ring structure (No. 12 functional group fragment) account for 8.73 mol.%, 4.37 mol.%, 2.32 mol.%, 15.77 mol.%, 29.16 mol.%, 31.59 mol.%, 3.27 mol.%, 4.79 mol.% as shown in Figure 2.18. By adjusting the proportions of palette compounds to minimize the target function Eq. (2.1), the formulated GCM1 surrogate and GCM2 surrogate are 22.06865%n-dodecane-33.11807%2-methylundecane-30.93333%n-butylbenzene-13.87995%methylcyclohexane and 19.666667%n-hexadecane-29.04%2-methylhexadecane-30.933333%n-butylbenzene-20.36%methylcyclohexane as shown in Table 2.11. The access of the diesel GCM1 surrogate mechanism and diesel GCM2 surrogate mechanisms are illustrated in the section of Data and Software Availability. The decomposed functional group fragments of GCM1 surrogate and GCM2 surrogate are provided in Figure 2.18 and Figure 2.19. Pei surrogate [147] which contains 64.97%n-dodecane-35.03%m-xylene is included for comparison. It should be noted that there is no published detailed chemical kinetic mechanism for Qian surrogate [174].
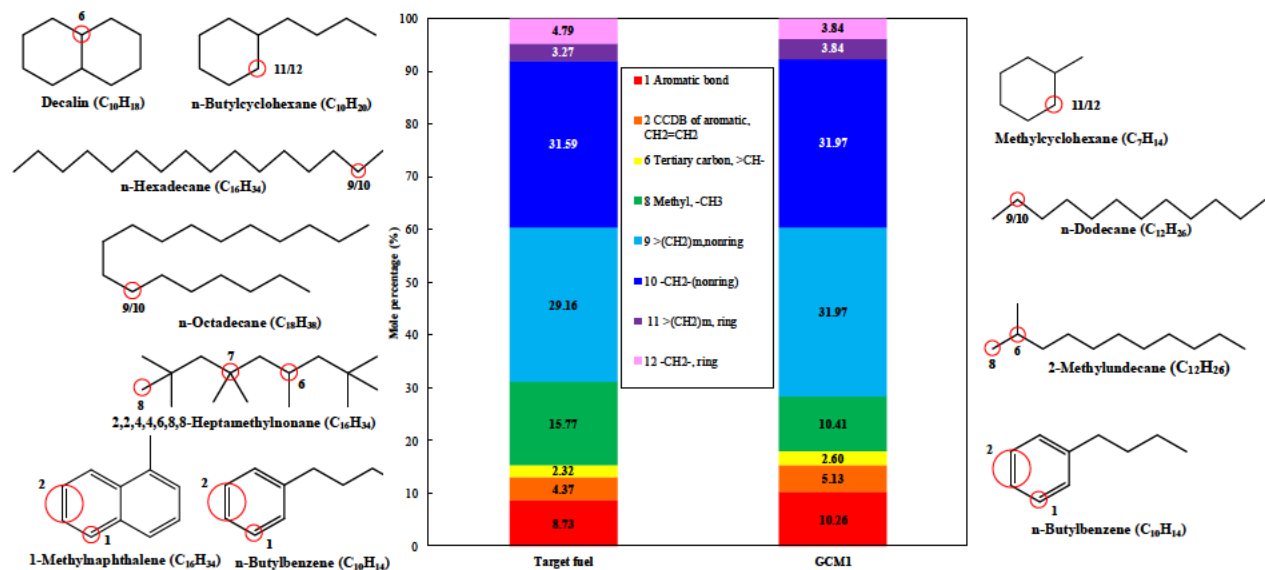
Figure 2.18. Functional group mole fractions for diesel target fuel [174] and GCM1 surrogate.
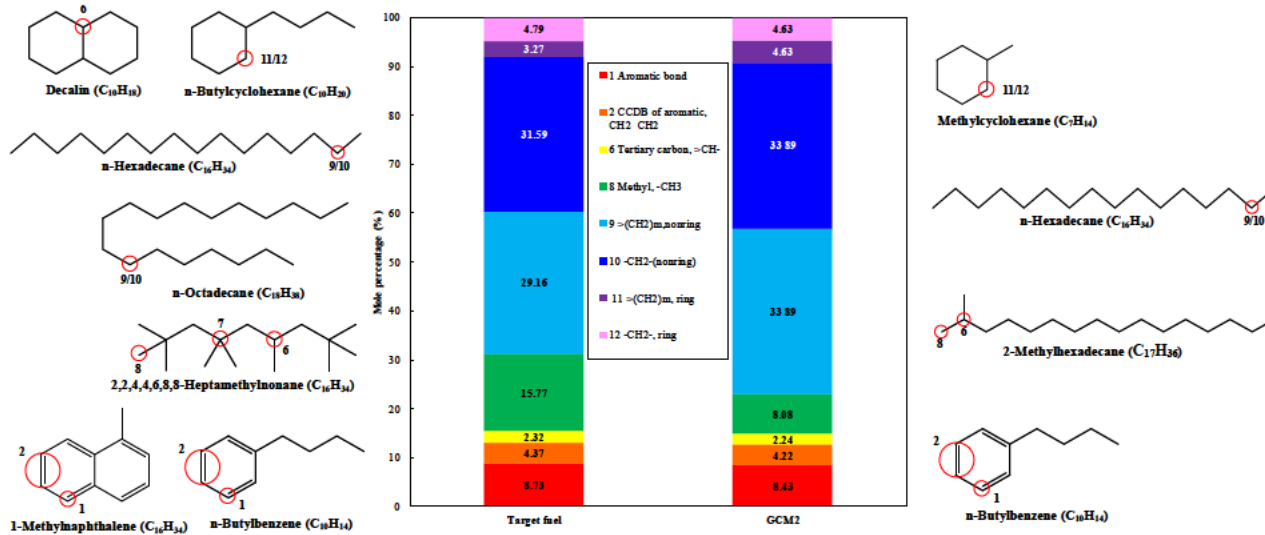
Figure 2.19. Functional group mole fractions for diesel target fuel [174] and GCM2 surrogate.

68

Table 2.11. Diesel surrogate compositions of Qian surrogate [174], GCM1 surrogate, GCM2 surrogate and Pei surrogate [147]

Diesel surrogate	Fuel compositions, mol.% and mechanism source	Species/Reactions	Formula	H/C ratio
Qian surrogate [174]	21.5%n-hexadecane-25.8%2,2,4,4,6,8,8-heptamethylnonane(HMN)-13.7%1-methylnaphthalene%-8.1%decalin-15.4%n-octadecane-8.1%n-butylbenzene-7.4%n-butylcyclohexane	N/A	$C_{14\,21}H_{27\,38}$	1.9268
GCM1 surrogate	22.06865%n-dodecane [147]-33.11807%2-methylundecane [147]-30.93333%n-butylbenzene [146]-13.87995%methylcyclohexane [134]	3701/15598	$C_{10\,69}H_{20\,62}$	1.9289
GCM2 surrogate	19.666667%n-hexadecane [91]-29.04%2-methylhexadecane [91]-30.933333%n-butylbenzene [146]-20.36%methylcyclohexane [134]	8050/35703	$C_{12\,60}H_{24\,32}$	1.9302
Pei surrogate [147]	64.97%n-dodecane-35.03%m-xylene [147]	3701/15598	$C_{10\,60}H_{20\,40}$	1.9245

2.3.3.2 Ignition delay times emulation

The measured ignition delay times of commercial diesel at at T_{init}=600~1250K, (a) φ=0.37, P_{init}=10bar, (b) φ=0.37, P_{init}=15bar, (c) φ=0.5, P_{init}=10bar, (d) φ=0.5, P_{init}=15bar, (e) φ=0.67, P_{init}=10bar, (f) φ=0.67, P_{init}=15bar, (g) φ=0.67, P_{init}=20bar, (h) φ=1.0, P_{init}=15bar, (i) φ=1.0, P_{init}=20bar are reported by Yu et al. [175, 176]. The GCM1 surrogate, GCM2 surrogate and Pei surrogate [147] are applied to predict the ignition delay times at the mentioned conditions and expand the temperature range of 600~1250K as shown in Figure 2.20 (a) ~ Figure 2.20 (i). The initial gas fractions are presented in Table 2.12. The GCM1 surrogate, GCM2 surrogate and Pei surrogate [147] underestimate the fuel reactivity and thus result in higher predictive ignition delay time than the observed values. The ignition delay times profiles of GCM1 surrogate and GCM2 surrogate almost overlap at the full temperature range of 600~1200K which indicates that the GCM enables a consistent surrogate formulation with different palette compounds. The ignition delay times of GCM1 surrogate and GCM2 surrogate indicate that the target fuel of Qian surrogate [174] possibly underestimates the fuel reactivity compared to commercial diesel. Both GCM surrogates successfully reproduce distinct NTC behavior at intermediate temperature regime of 725~850K while the Pei surrogate [147] exhibits monotonous ignition delay times with increasing temperature. The ignition delay times of diesel fuel decrease with increasing pressure and equivalence ratio which are well captured by the GCM1 surrogate and GCM2 surrogate. The ignition delay times predictive accuracy of four studied surrogate rank from high to low as: GCM2 surrogate ≈ GCM1 surrogate ≈ Qian surrogate [174]> Pei surrogate [147]. These four surrogates cannot represent the high reactivity of the real diesel fuel resulting in higher ignition delay times, new surrogates with higher reactivity are needed. This problem can be addressed by the current GCM formulation given that the detailed components of commercial diesel are revealed by GC-MS. The GCM surrogate formulation enables a rapid quantization of compositional proportions and provides a reasonable property reproduction.
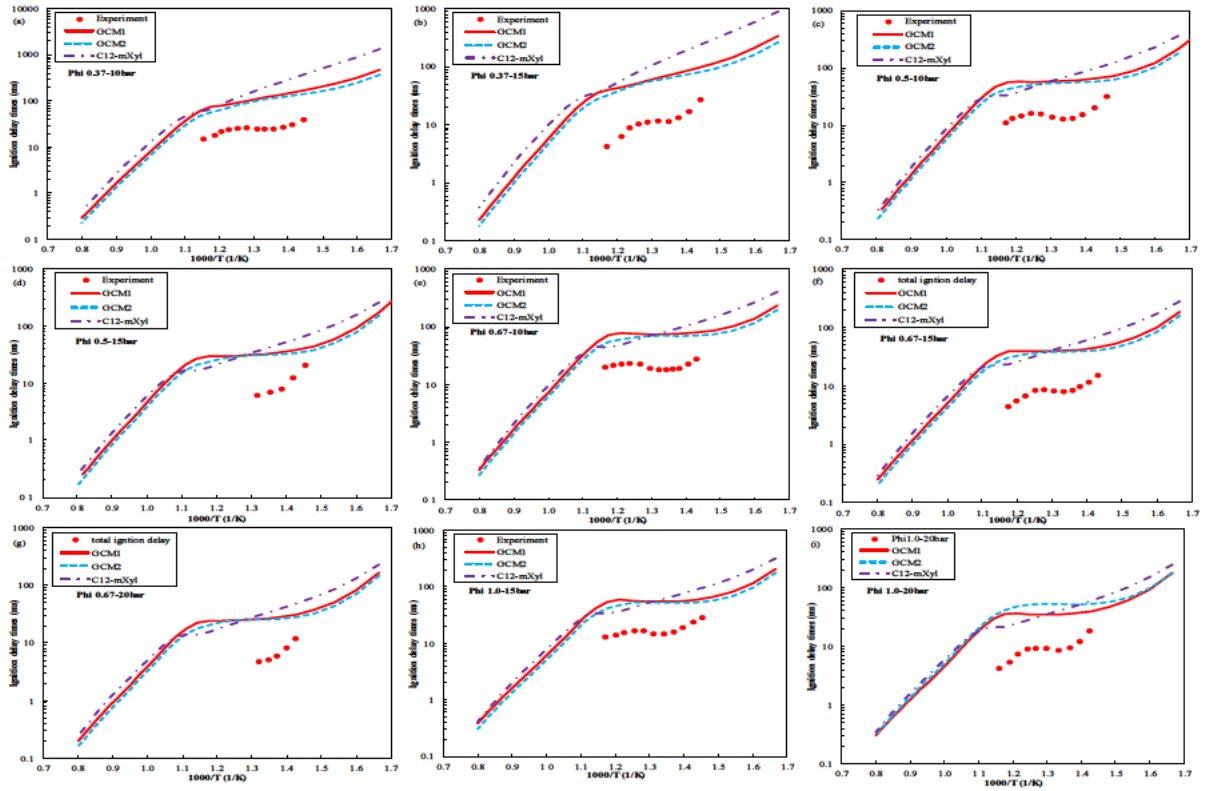
Figure 2.20. Ignition delay times of diesel (symbols) [175, 176] and surrogates simulation (lines) at T_{init}=600~1250K, (a) φ=0.37, P_{init}=10bar, (b) φ=0.37, P_{init}=15bar, (c) φ=0.5, P_{init}=10bar, (d) φ=0.5, P_{init}=15bar, (e) φ=0.67, P_{init}=10bar, (f) φ=0.67, P_{init}=15bar, (g) φ=0.67, P_{init}=20bar, (h) φ=1.0, P_{init}=15bar, (i) φ=1.0, P_{init}=20bar, solid line, dashed line and dashed-dotted line correspond to GCM1 surrogate, GCM2 surrogate and Pei surrogate [147].

Table 2.12. Initial gas fractions of diesel surrogates (GCM1 surrogate, GCM2 surrogate and Pei surrogate [147]) for ignition delay times simulation at T_{init}=600~1250K.

φ	P_{init} (atm)	X_{Fuel} (mol.%)	X_{O2} (mol.%, GCM1/GCM2/Pei [147] surrogate)	X_{N2} (mol.%, GCM1/GCM2/Pei [147] surrogate)
0.37	10	0.35	14.98656272/17.67266667/14.84908108	84.66343728/81.97733333/84.80091892
0.37	15	0.35	14.98656272/17.67266667/14.84908108	84.66343728/81.97733333/84.80091892
0.5	10	0.473	14.98741909/17.67367653/14.8499296	84.53958091/81.85332347/84.6770704
0.5	15	0.473	14.98741909/17.67367653/14.8499296	84.53958091/81.85332347/84.6770704
0.67	10	0.473	11.18464111/13.18931085/11.08203701	88.34235889/86.33768915/88.44496299
0.67	15	0.473	11.18464111/13.18931085/11.08203701	88.34235889/86.33768915/88.44496299
0.67	20	0.473	11.18464111/13.18931085/11.08203701	88.34235889/86.33768915/88.44496299
1.0	15	0.473	7.493709546/8.836838267/7.4249648	92.03329045/90.69016173/92.1020352
1.0	20	0.473	7.493709546/8.836838267/7.4249648	92.03329045/90.69016173/92.1020352

2.3.3.3 Jet stirred reactor oxidation emulation

The diesel oxidation kinetic in a JSR is studied at $\varphi=0.5\sim2.0$, $T_{init}=800\sim1400K$, $P_{init}=1\sim10atm$ by Mati et al. [177] and the GCM1 surrogate, GCM2 surrogate and Pei surrogate [147] are validated against the experimental values at the studied conditions as shown in Figure 2.21 ~ Figure 2.26. The initial gas fractions of these surrogates are provided in Table 2.13 and the average chemical formula of the "real fuel" (diesel) is $C_{15\,5}H_{30}$ [177]. The chemical formulas of GCM1 surrogate, GCM2 surrogate and Pei surrogate [147] are $C_{10\,69}H_{20\,62}$, $C_{12\,60}H_{24\,32}$, $C_{10\,60}H_{20\,40}$ which are smaller than the real fuel, thus they require less O_2 concentration than the real fuel as shown in Figure 2.21 ~ Figure 2.26. The molecules of H_2, CO, CO_2, CH_2O (formaldehyde) are accurately described by GCM2 surrogate, GCM1 surrogate comes second, and Pei surrogate [147] is the worst. For example, the GCM1 surrogate and GCM2 surrogate successfully predict the peak CO concentration at around $1120\sim1140K$ and the sequential rapid oxidation at $1140\sim1400K$ under the fuel-lean condition of $\varphi=0.5$, $\tau_{resident}=0.1s$ as shown in Figure 2.21 (c). Under the fuel-rich condition of $\varphi=2.0$, $\tau_{resident}=0.5s$, the GCM1 surrogate and GCM2 surrogate forecast the CO emission concentration rapidly increases at $840\sim960K$ while the rapid oxidation occurs at $870\sim1127K$ for the real fuel. Both GCM surrogates successfully predict the CO_2 formation plateau at high temperature oxidation regime of $1100\sim1400K$ because the oxygen concentration is insufficient to further oxidize CO into CO2 as shown in Figure 2.26 (c). Correspondingly, the CO2 production curve exhibits a plateau at $1260\sim1400$ which is accurately reproduced by both GCM surogates. In addition, the alkanes species of CH_4 (methane), C_2H_6 (ethane), alkenes species of C2H4 (ethene), C5H10-1 (1-pentene), C6H12-1 (1-hexene), aromatic species of C6H5CH3 (toleene), C6H6 (benzene) are well represented under the studied conditions by both GCM surrogates. The concentrations of C3H4-P, C3H4-A, IC4H8 species have distinct deviations at 900-1200K for all the studied surrogates, but the GCM2 surrogate obtains better accuracy than the rest surrogates. In summary, the species evolution predictive accuracy of the studied surrogates ranks from high to low as: GCM2 surrogate > GCM1 surrogate > Pei surrogate [147].
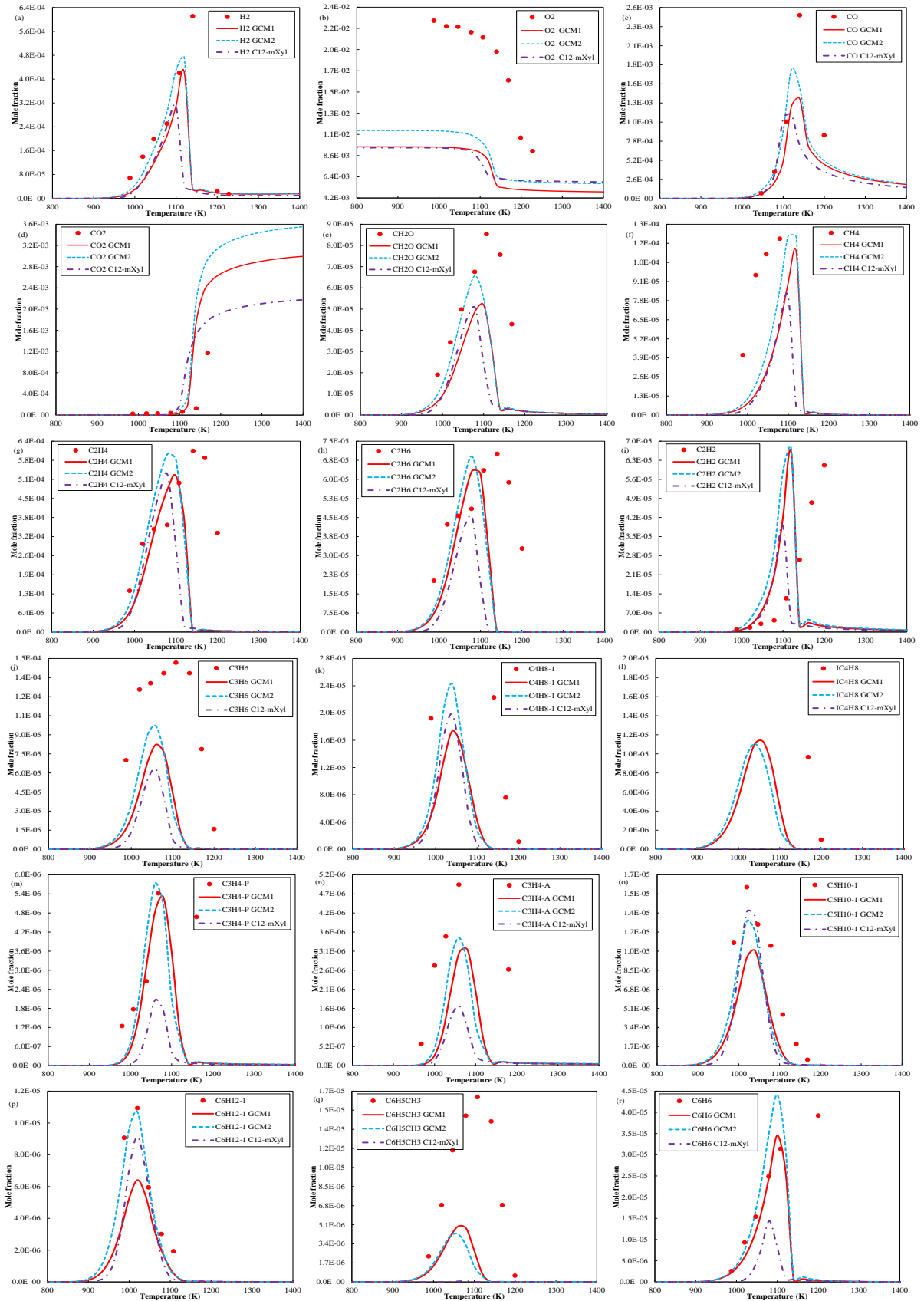
Figure 2.21. Speciation in JSR of diesel (symbols) [177] and surrogates simulation (lines) at φ=0.5, T_{init}=500~1600K, P_{init}=1atm, $\tau_{resident}$=0.1s, V=39cm^3, 0.03mol.% fuel diluted by nitrogen; solid line, dashed line and dash-dotted line correspond to GCM1 surrogate, GCM2 surrogate and Pei surrogate [147].
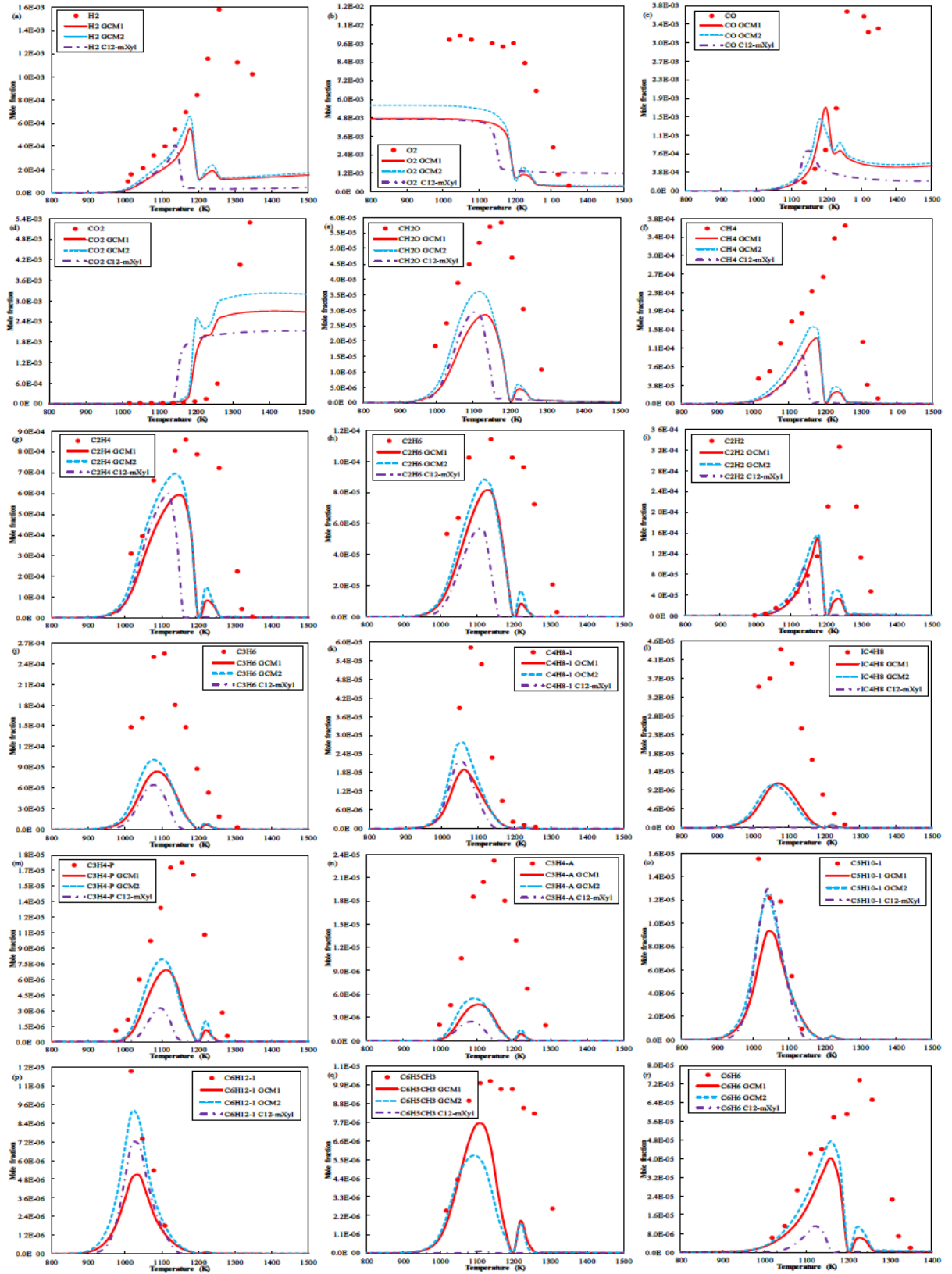
Figure 2.22. Speciation in JSR of diesel (symbols) [177] and surrogates simulation (lines) at φ=1.0, T_{init}=500~1600K, P_{init}=1atm, $\tau_{resident}$=0.1s, V=39cm^3, 0.03mol.% fuel diluted by nitrogen; solid line, dashed line and dash-dotted line correspond to GCM1 surrogate, GCM2 surrogate and Pei surrogate [147].
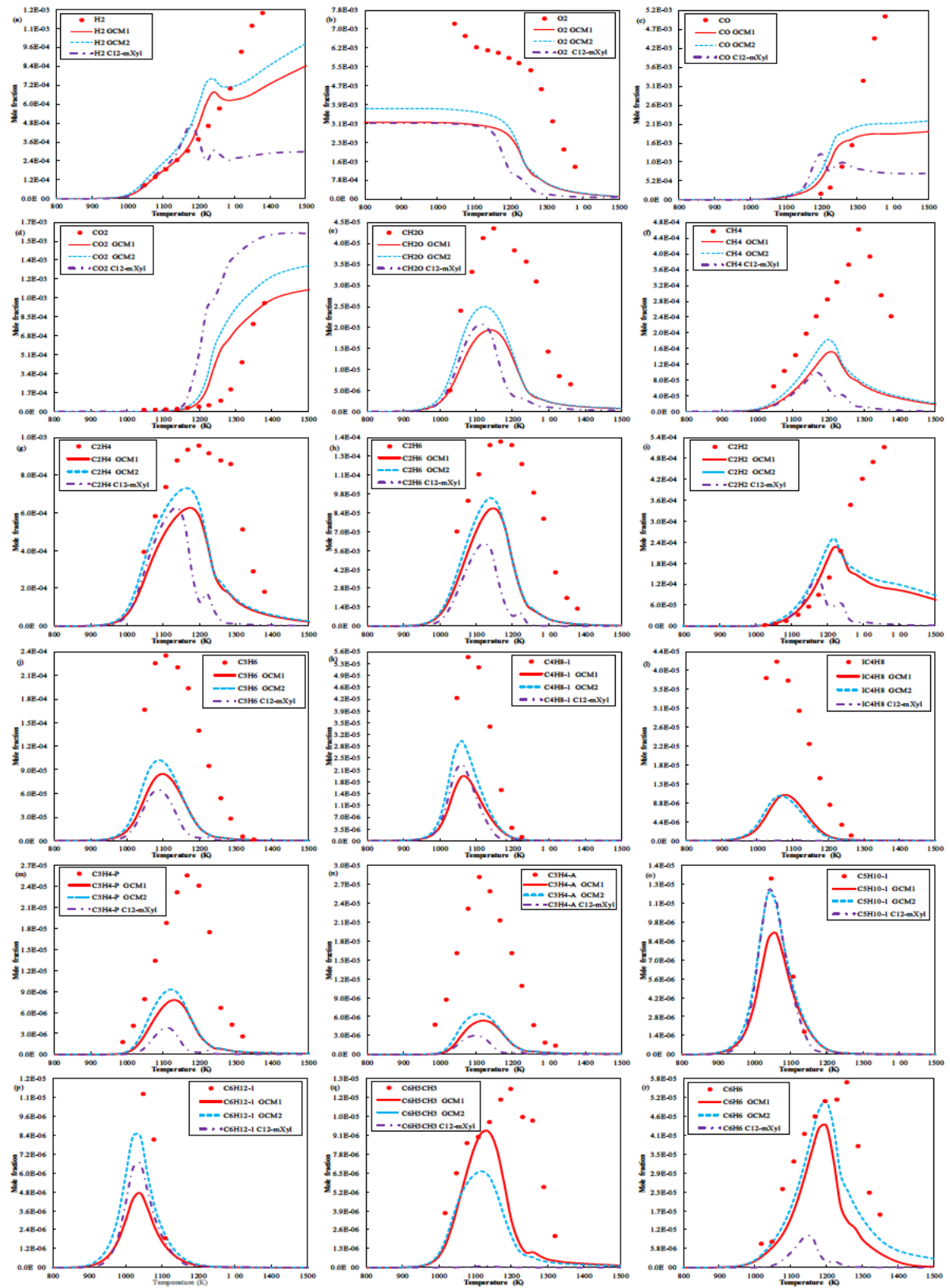
Figure 2.23. Speciation in JSR of diesel (symbols) [177] and surrogates simulation (lines) at φ=1.5, T_{init}=500~1600K, P_{init}=1atm, $τ_{resident}$=0.1s, V=39cm^3, 0.03mol.% fuel diluted by nitrogen; solid line, dashed line and dash-dotted line correspond to GCM1 surrogate, GCM2 surrogate and Pei surrogate [147].
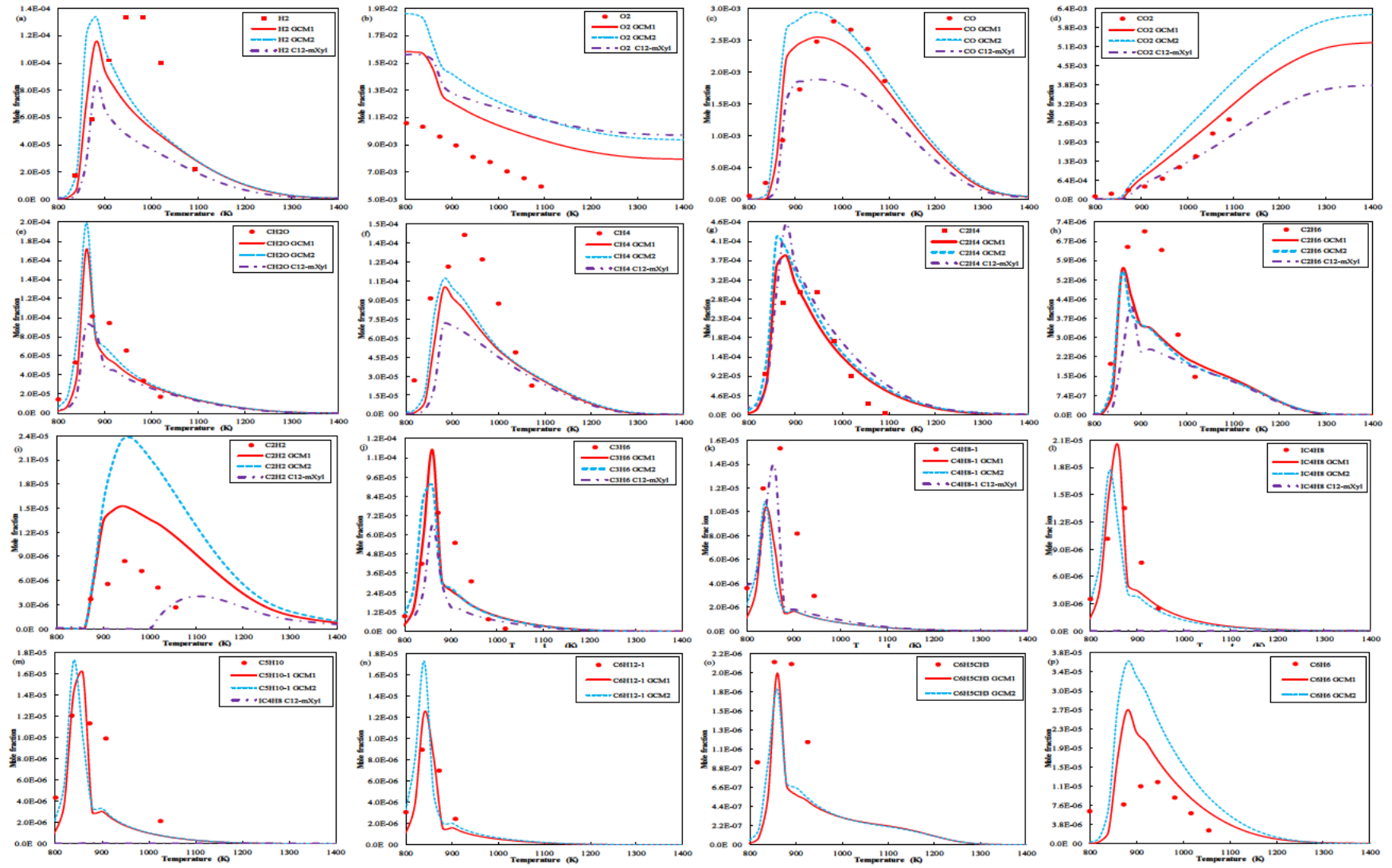
Figure 2.24. Speciation in JSR of diesel (symbols) [177] and surrogates simulation (lines) at φ=0.5, T_{init}=500~1600K, P_{init}=10atm, $τ_{resident}$=0.5s, V=39cm^3, 0.05mol.% fuel diluted by nitrogen; solid line, dashed line and dash-dotted line correspond to GCM1 surrogate, GCM2 surrogate and Pei surrogate [147].

Figure 2.25. Speciation in JSR of diesel [177] (symbols) and surrogates simulation (lines) at φ=1.0, T_{init}=500~1600K, P_{init}=10atm, $τ_{resident}$=0.5s, V=39cm^3, 0.05mol.% fuel diluted by nitrogen; solid line, dashed line and dash-dotted line correspond to GCM1 surrogate, GCM2 surrogate and Pei surrogate [147].

Figure 2.26. Speciation in JSR of diesel (symbols) [177] and surrogates simulation (lines) at φ=2.0, T_{init}=500~1600K, P_{init}=10atm, , $\tau_{resident}$=0.5s, V=39cm^3, 0.05mol.% fuel diluted by nitrogen; solid line, dashed line and dash-dotted line correspond to GCM1 surrogate, GCM2 surrogate and Pei surrogate [147].

Table 2.13. Initial gas fraction of diesel surrogate (GCM1 surrogate, GCM2 surrogate and Pei surrogate [147]) for jet stirred reactor simulation at T_{init}=500~1600K

φ	P_{init} (atm)	$\tau_{resident\ (s)}$	X_{Fuel} (mol.%)	X_{O2} (mol.%, real diesell/GCM1/GCM2/Pei [147] surrogates)	X_{N2} (mol.%, real diesel/GCM1/GCM2/Pei [147] surrogate)
0.5	1.0	0.1	0.03	1.38/0.950576/1.120952/0.941856	98.59/99.019424/98.849048/99.028144
1.0	1.0	0.1	0.03	0.69/0.475288/0.560476/0.470928	99.28/99.494712/99.409524/99.499072
1.5	1.0	0.1	0.03	0.46/0.316859/0.373651/0.313952	99.51/99.653141/99.596349/99.656048
0.5	10	0.5	0.05	2.3/1.584294/1.868253/1.56976	97.65/98.365706/98.081747/98.38024
1.0	10	0.5	0.05	1.15/0.792147/0.934127/0.78488	98.8/99.157853/99.015873/99.16512
2.0	10	0.5	0.05	0.575/0.396073/0.467063/0.39244	99.375/99.553927/99.482937/99.55756

2.3.4 FACE C gasoline

2.3.4.1 Surrogate formulation of FACE C gasoline

Fuels for Advanced Combustion Engines (FACE) are designed and certified by a subgroup of the Coordinating Research Council's (CRC's) Advanced Vehicle, Fuel, and Lubricants Committee's FACE Working Group. The FACE objectives are to design, advocate a set of advanced research fuels and quantify the influence of fuel physicochemical properties on the combustion and emission characteristics of advanced combustion modes [178, 179]. The designed FCAE diesel [82, 83] for compression ignition engines are toward advanced low temperature combustion (LTC) mode including HCCI (homogeneous charge compression ignition), PPCI (partially premixed compression ignition), CAI (controlled auto-ignition), etc. The designed FACE gasoline [84] for spark-ignition engines are toward stoichiometric high-efficiency combustion. The FACE C gasoline contains 24.43 vol.% n-paraffins, 69.73 vol.% iso-paraffins, 0.36 vol.% cyclo-paraffins, 3.92 vol.% aromatics, 1.27 vol.% olefins with RON of 84.3, MON of 83, OS of 1.3. The species concentrations of FACE C gasoline are measured by gas chromatography with flame ionization detectors (GC‐FID) which are reported in CRC Report No. AVFL-24 [84]. n-Butane, n-heptane, 2-methylbutane, 2-methylhexane, 2,2,4-trimethylpentane , toluene are selected as surrogate palette of GCM surrogate which represent n-paraffins, iso-paraffins, aromatics. The effects of olefins, cyclo-paraffins are omitted by the GCM surrogate to avoid complex chemical kinetic model because the sum proportion of these compounds is only 1.63 vol.%. Based on the result of detialed hydrocarbon analysis, the major components of the FACE C gasoline are decomposed into designed functional group fragments. The mole fractions of surrogate components are determined by the GCM model as 15.54%n-butane-9.29%2-methylbutane-0.41%2-methylhexane-4.8%toluene-15.36%n-heptane-54.60%2,2,4-trimethylpentane and the average chemcial formula is $C_{6\ 894}H_{15\ 404}$ as shwon in Table 2.14.

Table 2.14. FACE C surrogate compositions of GCM surrogate

FACE C surrogate	Fuel compositions, mol.% and mechanism source	Species/Reactions	Formula	H/C ratio
GCM surrogate	15.54%n-butane-9.29%2-methylbutane-0.41%2-methylhexane-4.8%toluene-15.36%n-heptane-54.60%2,2,4-trimethylpentane [167]	2406/9633	$C_{6\,894}H_{15\,404}$	2.2344

2.3.4.2 Ignition delay times emulation

The measured ignition delay times of FACE C gasoline at φ=0.5, 1.0, T_{init}=550~1250K, P_{init}=20bar, 40bar are reported by Sarathy et al. [180]. Therein the low temperature and high temperature ignition delay times are measured by RCM (400~1200K) and ST (800~2500K) based on their operating conditions [181]. The simulated ignition delay times of the GCM surrogate are validated against the measured values [180] as shown in Figure 2.27. At φ=0.5, P_{init}=20bar, the GCM surrogate accurately reproduces the measured ignition delay times at low temperature (570~770K)/intermediate temperature (770~850K)/high temperature (850~1250K) regimes. The NTC behavior is also captured by the GCM surrogate at the intermediate temperature of 770~850K that the ignition delay times increase with enhanced temperature. At φ=1.0, P_{init}=20bar, the GCM surrogate overestimates the ignition delay times at 770~1050 and the measured curve demonstrates a plateau at 774~929K. The slope of NTC behavior declines as increasing equivalence ratio and pressure [182, 183] which also supports by the GCM surrogate. At φ=0.5, P_{init}=40bar, the GCM surrogate obtains a reasonable predictive accuracy at the full temperature regime. It reveals that the NTC intensity decreases as increasing initial gas pressure and the ignition delay times curve of φ=0.5, P_{init}=40bar becomes flattered compared to that of φ=0.5, P_{init}=20bar. At φ=1.0, P_{init}=40bar, the GCM surrogate successfully emulates the ignition delay times at 570~730K and 910~1250K but overestimates the fuel reactivity at an intermediate temperature regime of 730~910K. The NTC intensity weakens as increasing temperature and thus the ignition delay times curve becomes flat as shown in Figure 2.27. The dependence of equivalence ratio, pressure on the NTC behavior is accurately captured by GCM surrogate.
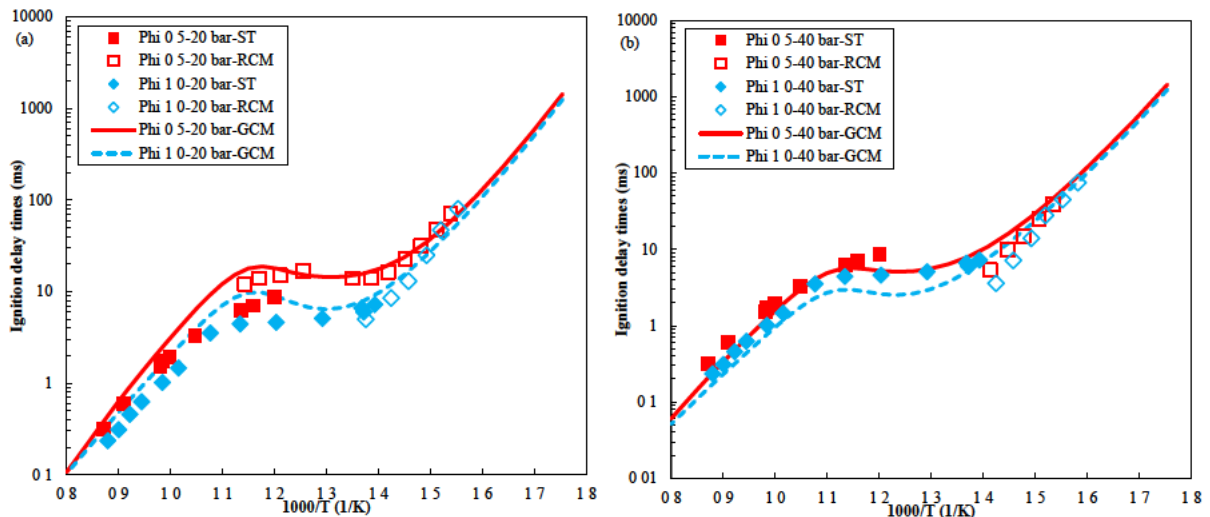
Figure 2.27. Ignition delay times of FACE C gasoline-air mixture (symbols) [180] and surrogates simulation (lines) at φ=0.5/1.0, T_{init}=550~1250K, (a) P_{init}=20bar, (b) P_{init}=40bar.

2.3.4.3 Jet stirred reactor oxidation emulation

The FACE C gasoline oxidation in a JSR is studied at φ=0.5/1.0/2.0, T_{init}=500~1100K, P_{init}=1atm, $\tau_{resident}$=0.7s by Chen et al. [184]. The measured mole fraction profiles of GACE C gasoline and the predicted profiles of GCM surrogate are compared in Figure 2.28. The measured and predicted concentrations of major species (H_2O, CO, CO_2, H_2, CH_4, O_2), C1-C2 oxygenated species (CH_2O, CH_3CHO, CH_3OH), C2-C4 olefins (C_2H_4, C_3H_6, IC_4H_8) at φ=0.5, 1.0, 2.0 are plotted in Figure 2.29, Figure 2.30, Figure 2.31 respectively. The initial gas fractions of the FACE C gasoline surrogate (GCM surrogate) at φ=0.5, 1.0, 2.0 are presented in Table 2.15. The low temperature oxidation reactivity enhances as decreasing equivalence ratio from 2.0 to 0.5 which is successfully reproduced by GCM surrogate as shown in Figure 2.28. The GCM surrogate also well forecasts the fuel consumption at high temperature regime of 750~1100K compared to the observed data. The GCM surrogate satisfactorily predicts the major species of H2O, CO, CO2, H2, CH4, O2 at the entire temperature regime (500~1100K). The H2O, CO profiles exhibit the low temperature oxidation at temperature of 575~750K and the high temperature oxidation starts at around 800K, the bimodal profiles at fuel-lean, stoichiometric, fuel-rich conditions are well captured by GCM surrogate. The oxygenate species of CH_2O (formaldehyde), CH_3CHO (acetaldehyde), CH_3OH (methanol) have two peaks corresponding to low and high temperature oxidation respectively. The first peak is caused by the reaction of oxygen addition to alkyl radical at low temperature of 575~750K and the high temperature oxidation dominates the second peak. The C-C bond β-scission of alkyl radical reaction is prevalent at high temperature above 800K and produce significant amount of olefins of C_2H_4 (ethylene), C_3H_6 (propene), IC_4H_8 (iso-butene).
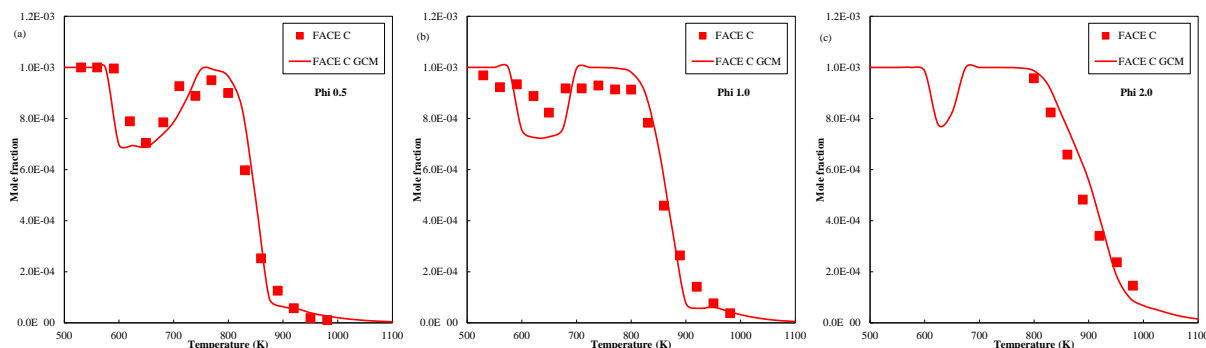
Figure 2.28. Reactant profiles in JSR of FACE C gasolines (symbols) [184] and surrogates simulation (lines) at T_{init}=500~1100K, P_{init}=1atm, $\tau_{resident}$=0.7s, V=38cm^3, 0.1mol.% fuel diluted by nitrogen; (a) φ=0.5, (b) φ=1.0, (c) φ=2.0.
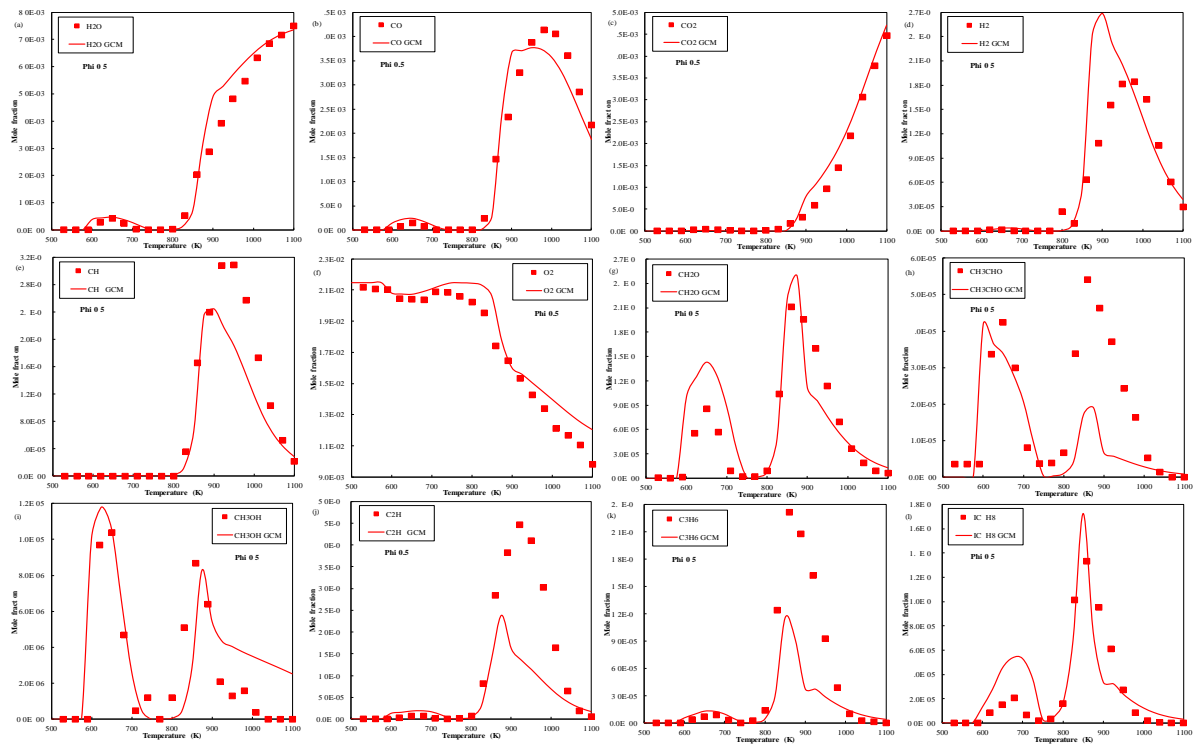
Figure 2.29. Speciation in JSR of FACE C gasoline (symbols) [184] and surrogates simulation (lines) at φ=0.5, T_{init}=500~1100K, P_{init}=1atm, $\tau_{resident}$=0.7s, V=38cm^3, 0.1mol.% fuel diluted by nitrogen.
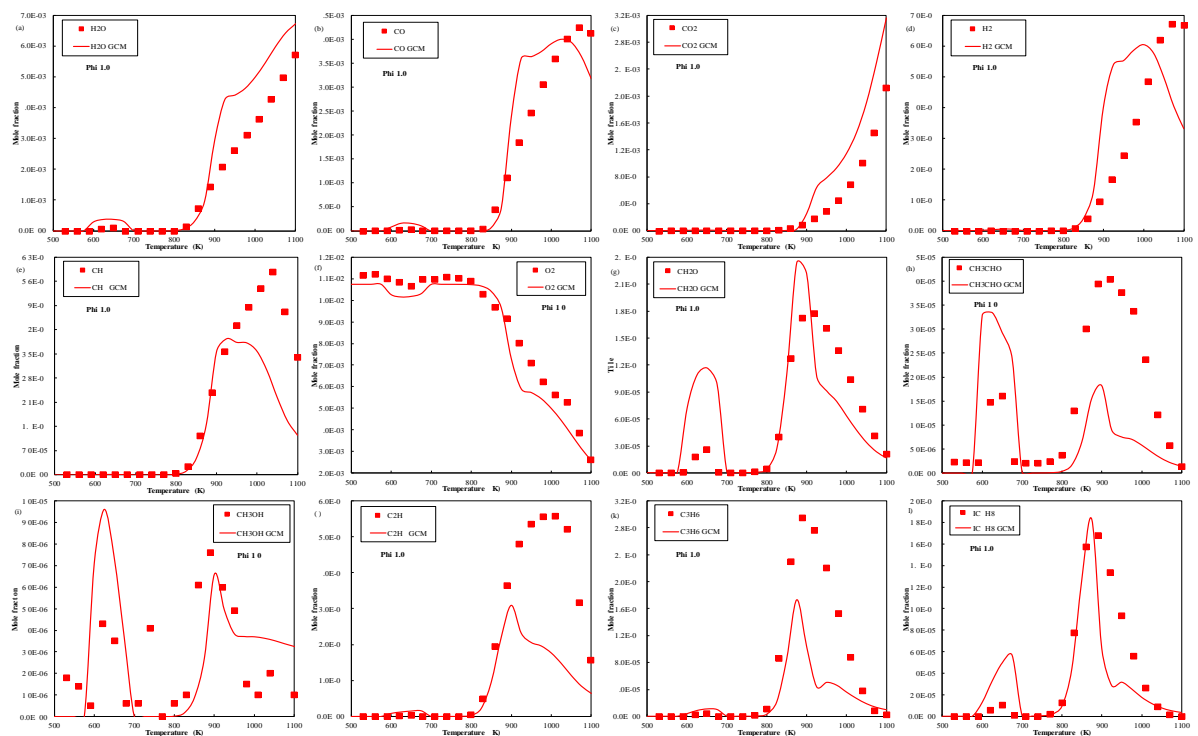
Figure 2.30. Speciation in JSR of FACE C gasoline (symbols) [184] and surrogates simulation (lines) at φ=1.0, T_{init}=500~1100K, P_{init}=1atm, $\tau_{resident}$=0.7s, V=38cm^3, 0.1mol.% fuel diluted by nitrogen.
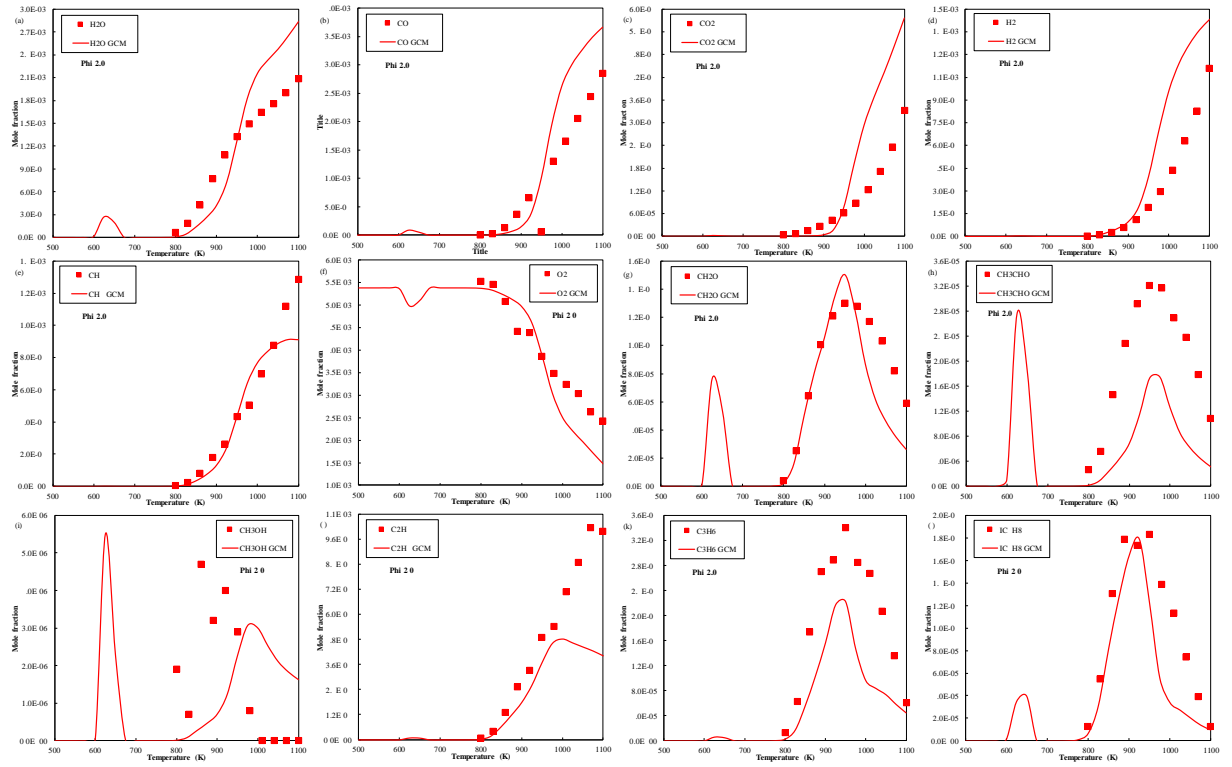
Figure 2.31. Speciation in JSR of FACE C gasoline (symbols) [184] and surrogates simulation (lines) at φ=2.0, T_{init}=500~1100K, P_{init}=1atm, $\tau_{resident}$=0.7s, V=38cm^3, 0.1mol.% fuel diluted by nitrogen.

Table 2.15. Initial gas fractions of FACE C gasoline surrogate (GCM surrogate) for JSR simulation at T_{init}=500~1100K, P_{init}=1atm

φ	Fuel (0.1 mol %)						O$_2$	N$_2$
	n-Butane	2-Methylbutane	2-Methylhexane	Toluene	n-Heptane	2,2,4-Trimethylpentane		
0 5	0.015539974	0.00929003	0.000409959	0.004800001	0.015360028	0.054600008	2.149	97.751
1.0	0.015539974	0.00929003	0.000409959	0.004800001	0.015360028	0.054600008	1.0745	98.8255
2.0	0.015539974	0.00929003	0.000409959	0.004800001	0.015360028	0.054600008	0.53725	99.36275

2.4 Conclusion

This work proposes a novel group contribution method (GCM) for surrogate formulation and the core idea is to minimize the difference of functional group fragments between the target fuel and surrogate fuel. It is based on the assumption that the fuel molecular structure determines the fuel physicochemical properties, thus the fuels with similar structures have similar properties. A GCM-UOB 1.0 functional group classification system containing 22 structural fragments is established to decompose the fuel molecules into functional group fragments. The GCM based surrogate formulation contains 7 steps: (1) Identify the target fuel and its detailed composition. (2) Select the surrogate palette to represent the major chemical classes of the target fuel. (3) Decompose the fuel molecules into typical functional group fragments for both target fuel and surrogate fuel. (4) Run the GCM regression model to minimize the difference of functional group fragments between target fuel and surrogate fuel. (5) Examine the surrogate fuel properties and validate against the target fuel. (6) Evaluate if the surrogate properties meet the requirement. (7) Finalize and output the surrogate formulation result.

GCM surrogate formulation applies to POSF 4658 aviation fuel, RME biodiesel, Diesel, FACE C gasoline as case studies. A satisfactory agreement on ignition delay times, species evolution, laminar flame speed, liquid density, sound speed, kinematic viscosity between the target fuel and surrogate fuel. The success of GCM surrogate formulation is attributed to the following reasons: (1) The reasonable GCM hypothesis that the fuel physicochemical properties are the sum result of the fuel molecular structure and functional groups. (2) The GCM surrogate formulation simplifies the aligned parameters from complex physicochemical properties to functional group fragments of fuel molecules. (3) The GCM-UOB 1.0 system considers the functional group interaction (by functional group identifier, functional group position descriptor) and the contribution to fuel reactivity (fuel reactivity descriptor). The GCM provides an efficient and reliable tool to determine the compositional proportions for surrogate formulation.

Chapter 3 Machine Learning Regression Quantitative Structure-Property Relationship (ML-QSPR) for Cetane and Octane Numbers Prediction of Pure Fuel Compounds and Mixtures

3.1 Introduction

Ignition quality is an important property of fuel molecules for property-oriented fuel screening and cetane number (CN) [185], derived cetane number (DCN) [186, 187], indicated cetane number (ICN) [188], research octane number (RON) [189], motor octane number (MON) [190] are the commonly used measurements. The scope of these indexes are different. CN, DCN and ICN are used to describe the ignition quality of diesel fuel oils or diesel-like fuels with high autoignition propensity while RON, MON depict the ignition quality of spark-ignition engine fuels or gasoline-like fuel with high anti-knock property. Especially, the CN, RON, MON are measured by CFR (Cooperative Fuel Research) engine test while the DCN and ICN are determined by constant volume chamber. The apparatus, calibrated range, sample consumption, test time of these ASTM fuel ignition quality test standards are compared in Table 3.1. There are some challenges in determining the fuel ignition quality by combustion experiments (CFR engine test or combustion vessel test) which are summarized below: (1) It is difficult to simultaneously determine the CN, ROM, MON by CFR engine test because the CN loosely negatively correlates with RON/MON [157, 191, 192] and the calibrated ranges of ASTM standards are limited (see Table 3.1). In other words, those fuel molecules are suited to be characterized by CN/DCN are usually not adapted to RON/MON scale. For example, the dibutyl ether (CAS No.: 142-96-1) is prone to autoignition with DCN of 115.4 [193] but its RON/MON cannot determine by the CFR engine test because of the extremely high reactivity. Similarly, the toluene has good knock resistance with RON of 120 but its CN cannot be measured by CFR engine test or constant volume chamber test. For comparison, it is good to place molecules on the unified ignition quality scale either CN or RON/MON. (2) 500mL fuel sample is required to determine the CN/RON/MON by CFR engine test and this quantity is hard to prepare for the new fuels due to technical difficulty and cost. The DCN (100 mL or 370 mL)

and ICN (40 mL) determination by combustion chamber consume fewer sample quantity but it is not cost-effective to conduct large-scale studies during the preliminary stage of fuel screening. For example, polyoxymethylene dimethyl ether 3 (PODE3, CAS No.: 13353-03-2) is a promising carbon-neutral fuel for CI engine, but there are incredibly high price ($32/1g) and limited stocks in AmBeed [194]. It is unreasonable and unrealistic to perform a large-scale experimental screening for the emerging fuels at the early stage of fuel screening due to the cost and sample availability. (3) The CFR engine test (40 min per sample) and combustion vessel test (25~30 min per sample) are time-consuming and large-scale experimental fuel screening on emerging fuels from biomass or waste plastics is not realistic. Some solutions are proposed to address the above challenges which are summarized as follow: (1) Unify the CN and RON/MON scales using conversion formulas as shown in Table 3.2. But their scopes are limited to typical hydrocarbons and the predictive accuracy is modest, so they cannot satisfy the requirement of large-scale fuel screening. (2) Model-based ignition quality predictive method. According to the feature extraction method, the CN/RON/MON predictive models can be divided as: (i) GCM (group contribution method), (ii) QSPR (quantitative structure-property relationship), (iii) Simulated IDT (ignition delay time), (iv) Measured chemical compositions information such as LC (liquid chromatography) & GC-MS (gas chromatography-mass spectrometry), H NMR (nuclear magnetic resonance) spectroscopy, FTIR spectra as summarized in Table 3.3.

GCM is proposed by Benson et al. [195-197] to estimate the thermodynamic properties and rate parameters of pure compounds which belongs to an empirical method. GCM is based on the group additivity principle that a molecule is decomposed into typical functional groups (building blocks) and the fuel property is a sim of the contributions from these structural fragments [29]. It applies to any possible combination of the present functional groups but the relevant contributions (weights) are empirical [198] and the model performance heavily relies on the studied compounds [199]. QSPR theory assumes that similar chemical structure results in similar properties and the macroscopic properties change reflects the molecular structure variation [29, 200]. The object of QSPR is to relate the macroscopic property and molecular structure features (encoded by molecular descriptors) by

mathematical regression model as shown in Eq. (1.1) [29, 201]. The molecular descriptors to characterize the chemical structure of components are divided into 4 categories [29, 200]: (1) Constitutional descriptors which represent the occurrence of typical atom types, bonds, functional groups of molecules based on the 1D chemical formula or 2D chemical structure. (2) Topological descriptors which use vertices and edges to describe typical atoms and bonds based on the 2D molecular structure. (3) Geometric descriptors which depict the interatomic distances, angles, dihedral angles, molecular volume and surface area based on 3D molecular structure. (4) Quantum chemical descriptors which sketch electronic features, vibrational frequency levels, reactivity indices, orbital energies based on 3D chemical structure. A wide variety of mathematical tools can be used to build the regression models such as artificial neural network (ANN), genetic algorithm (GA), multilinear regression (MLR), non-linear regression (NLR), partial least squares (PLS) and machine learning (ML) algorithms [52] (linear regression, regression trees, support learning machines, Gaussian process regression, ensembles of trees) [29, 199]. The simulated IDT method is used to correlate the simulated IDT with RON/MON while the IDT is computed by a detailed chemical kinetic mechanism. This method has been validated against the fuel types of alkanes, alkenes, cycloalkane, aromatics, alcohols, ketones, esters, acids, furans, TPRF mixtures but limited number of compounds are involved. In addition, the detail chemical kinetic mechanism must be available for the studied compounds of interest. Measured chemical compositions information method uses LC, GC-MS, NMR, FTIR to reflect the information of molecules and atoms types which used as the regression model input.

This work proposes a machine learning quantitative structure-property relationship (ML-QSPR) method to predict the ignition quality (CN/RON/MON) of pure compounds and mixtures. A novel constitutional molecular descriptor of QSPR-UOB 3.0 is developed to extract the structural features and transforms them into the fuel molecular structural matrix. A fuel ignition quality data containing 869 pure compounds and 432 mixtures is established to prepare the fuel property matrix and train the regression model. ML algorithm is employed to build the regression model to connect the fuel molecular structure matrix and fuel property matrix. The model predictive

performance is examined by 10-fold cross validation and prevent over-fitting.

Table 3.1. Comparison of ASTM fuel ignition quality test standards

Parameters	Standard	Apparatus	Calibrated range	Vol. (mL/sample)	Test time (min/sample)	Ref.
CN	ASTM D613	CFR engine	30~65	500	40	[185]
DCN	ASTM D6890	CVV	31.5~75.1	100	20	[186]
DCN	ASTM D7668	CVV	30~70	370	30	[187]
ICN	ASTM D8183	CVV	35~85	40	25	[188]
RON	ASTM D2699	CFR engine	40~120.3	500	40	[189]
MON	ASTM D2700	CFR engine	40~120	500	40	[190]

Table 3.2. Conversion formula between cetane and octane numbers

No.	Formula, R-square	Institute	Validation scope	Ref.
1	CN=56–0.39×RON, 0.87	NREL	Hydrocarbons and oxygenates	[193]
2	CN=68.54-0.59×RON, 0.93 (general expression)	SwRI	66 gasoline samples, MON: 75~ 94, CN: 24~71	[202]
	CN=67.49-0.59×RON, 0.94 (high OS fuel)			
	CN=68.51-0.59×RON, 0.92 (low OS fuel)			
	CN=60.96–0.56×MON, 0.82 (general expression)			
	CN=64.21-0.61×MON, 0.893 (high OS fuel)			
	CN=67.86-0.63×MON, 0.903 (low OS fuel)			
3	CN=54.633-0.4208×RON, 0.9799	Shell Global Solutions	n-Heptane-iso-octane mxitures, n-heptane-toluene mxitures	[203]
4	ON=125-1.96×CN, 0.99	University of Wisconsin-Madison	secondary reference fuels T and U mixtures, CN: 19.4~75.2	[204]
5	ON=146-3.7×CN+0.08946×CN^2-0.001263×CN^3	SwRI	PRF mixtures	[205]
6	CN=(120-RON)/2	Toyota Central R&D Labs., Inc.	Diesel fuel samples	[206]
7	CN=60-0.5×MON	Umweltbundesamt	Hydrocarbons and gasoline stocks	[207]

Table 3.3. Overview of CN and ON forecasting approaches

Objective	Feature extraction method	Regression model	Model inputs	Optimal R^2	RMSE	Dataset	Scope	Ref.
CN/RON/MON	GCM	ANN	38 functional groups	0.90	N/A	449	Alkanes, alkenes, alkynes, cycloalkanes, cycloalkenes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans	[208, 209]
CN	GCM	IQT ignition delay model	13 functional groups, IQT ignition delay, vapor pressure	0.98	8.75	162	As above	[157]
CN	GCM	LRM	^{13}C NMR spectroscopy and 7 group descriptors	0.64	N/A	127	34 pure alkanes, 93 hydrocarbon mixtures	[210]
CN	GCM	BP-NN	4 functional groups and boiling point for isoparaffins	0.97	N/A	141	iso-Paraffins and diesel fuels	[211]
CN	QSPR	SVM	28 functional groups	0.934	6.3	229	Hydrocarbons, alcohols, esters	[212]
CN	QSPR	GA	150 molecular descriptors	0.978	N/A	147	Alkanes, alkenes, cycloalkanes, aromatics	[213]
CN	QSPR	ASM	9 topological indices and 5 carbon-chain related descriptors	0.93	N/A	110	Alkanes, alkenes, cycloalkanes, aromatics	[214]
CN	QSPR	ABFIS	4 evaporation relevant descriptors and 6 combustion relevant descriptors	0.986	3.38	496	204 hydrocarbons and 292 oxygenates, no further detail available	[215]
CN	QSPR	ANN & ReLU	15 QSPR descriptors	0.963.	7.94	N/A	Alkanes, alkenes, alkynes, cycloalkanes, aromatics, alcohols, aldehydes/ketones, ethers, esters	[216]
CN	QSPR	ANN	10 molecular descriptors	0.934	N/A	349	Alkanes, alkenes, aromatics, alcohols, esters, others (3 ketones, 1 aldehyde, 8 ethers and 4 acids)	[217]
CN	QSPR	ANN	15 molecular descriptors	N/A	9.1	284	Alkanes, alkenes, alkynes, cycloalkanes, cycloalkenes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans	[218]
CN	QSPR	Regression equations	2D TI	0.99998	N/A	71	Alkanes, cycloalkanes	[219]
CN	LC & GC-MS	General regression NN	12 hydrocarbon groups	0.97	N/A	69	Alkanes, cycloalkanes, aromatics	[220]
CN	H NMR spectroscopy	MLR	H NMR spectroscopy	0.95	N/A	125	Alkanes, alkenes, alkynes, cycloalkanes, aromatics, hydrocarbon mixtures	[156]
RON/MON	QSPR	SVM	Molecular descriptors: 12 for RON, 23 for MON	0.92	N/A	552	279 for RON, 273 for MON, alkanes, alkenes, alkynes, cycloalkanes, cycloalkenes, aromatics, alcohols, esters, furans	[221]

RON/MON	QSPR	MLR	Molecular mass, hydration energy, boiling point, molar refractivity, octanol/water distribution coefficient, critical pressure, critical volume, critical temperature	0.9419	N/A	65	Alkanes, cycloalkanes	[222]
RON/MON	QSPR	Regression equations	2D TI	0.9643	N/A	27	Heptane isomers, octane isomers	[223]
RON/MON	QSPR	Regression equations	2D TI	0.9966	N/A	78	46 samples of alkanes, 32 samples of cycloalkanes	[224]
RON/MON	Simulated IDT	IDT & ON fitting model	Computed ignition delay curve	N/A	N/A	N/A	Alkanes, alkenes, cycloalkane, aromatics, alcohols, ketones, esters, acids, furans	[225]
RON/MON	Simulated IDT	IDT & ON fitting model	RON: constant volume IDT at 750K, 25bar MON: constant volume IDT at 825K, 25bar	0.9932	N/A	N/A	Alkanes, alkenes, aromatics and their mixtures	[226]
RON/MON	Simulated IDT	IDT & ON fitting model	Compression ratio dependent variable volume IQT	0.9726	N/A	25	TPRF mixtures	[227]
RON/MON	H NMR spectroscopy	ANN	15 H types in H NMR spectroscopy	0.99	2.2	251	128 pure hydrocarbons: alkanes, alkenes, cycloalkanes, cycloalkenes, aromatics; 123 hydrocarbon blends: n-heptane, iso-octane toluene, trimethylbenzene, cyclopentane, 1-hexene, ethanol	[228]
RON	FTIR spectra	PCR	Fourier-transform infrared absorption spectra	N/A	N/A	34	Alkanes, alkenes, cycloalkanes, aromatics	[229]

3.2 Modeling approach

3.2.1. Methodological overview

ML-QSPR models are developed to predict the CN, RON, MON simultaneously from the molecular structure which provides an insightful understanding of the impact of molecular structure on the ignition quality. A QSPR-UOB 2.0 functional group classification system is proposed to decompose the fuel molecule into component fragments and transform them into a fuel molecular structure matrix. Fuel ignition quality database containing 869 pure compounds and 432 mixtures is established to provide the fuel property matrix and train the regression model. Fuel molecular structure matrix and fuel property matrix are mapped by 19 ML algorithms. 10-fold cross validation is implemented to examine the predictive accuracy and choose the best model with minimum RMSE.

3.2.2. Fuel ignition quality database development

Fuel ignition quality database is established to store the CN/RON/MON data and fuel molecular structure matrix of 869 pure compounds and 432 mixtures. The CN/RON/MON properties are mainly obtained from Co-Optimization of Fuels & Engines: Fuel Properties Database [230] and Los Alamos National Laboratory Report No. LA-UR-16-25529 [208, 209]. Another major data source of CN is the Compendium of Experimental Cetane Numbers [193] and the RON/MON is provided by API Data Book [231] and American Petroleum Institute Research Project 45 [232]. The detailed data sources are summarized in Table 3.4 and the fuel ignition quality database contains 603, 374, 371 data of CN, RON, MON respectively as shown in Table 3.5.

Table 3.4. The data source of measured CN/RON/MON for pure compounds and fuel mixtures

Items	Fuel type	Institute	Ref.
CN/RON/MON	Alkanes, alkenes, alkynes, cycloalkanes, cycloalkenes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans	NREL	[230]
CN/RON/MON	Alkanes, alkenes, alkynes, cycloalkanes, cycloalkenes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans	LANL	[208, 209]
CN	Alkanes, alkenes, alkynes, cycloalkanes, cycloalkenes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans	NREL	[193]
CN	Alkanes, alkenes, alkynes, cycloalkanes, cycloalkenes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans	RWTH Aachen University	[157]
CN	Alkanes, alkenes, alkynes, cycloalkanes, aromatics, hydrocarbon mixtures	KAUST	[156]
CN	Alkanes, cycloalkanes	RAS	[233]
CN	Alkanes, cycloalkanes, aromatics	Hokkaido University	[234]
CN	Alkanes, aromatics, hydrocarbon mixtures	USC	[235]
CN	Cycloalkanes, n-heptane-cycloalkane mixtures	USC	[236]
CN	Hydrocarbon mixtures	Princeton University,	[155]
CN	Hydrocarbon mixtures	Stanford University	[237]
RON/MON	Alkanes, alkenes, alkynes, cycloalkanes, cycloalkenes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans	AIChE	[231]
RON/MON	Alkanes, alkenes, alkynes, cycloalkanes, aromatics	ASTM	[232]
RON/MON	TPRF mixtures	Saudi Aramco	[238]
RON/MON	TPRF mixtures	University of Cambridge	[239]
RON/MON	TPRF mixtures	Saudi Aramco	[227]
RON/MON	TPRF mixtures	KAUST	[240]
RON/MON	TPRF-ethanol mixtures	University of Melbourne	[241]
RON/MON	Hydrocarbon mixtures	KAUST	[226]
RON/MON	Hydrocarbon-ethanol mixtures	LLNL	[225]
RON/MON	Alkanes, alkenes, cycloalkanes, cycloalkenes, aromatics, ethanol and their mixtures	KAUST	[228]

Table 3.5. Number of compounds of different chemical classes in the ignition quality database for model training

Compound class	Number of compounds (measured data)		
	CN	RON	MON
Alkanes	74	46	46
Alkenes	35	70	72
Alkynes	0	4	2
Naphthenes	52	40	35
Aromatics	56	35	37
Total oxygenates	266	24	23
Alcohols	52	13	12
Aldehydes/Ketones	19	2	2
Saturated esters	66	3	3
Unsaturated esters	19	N/A	N/A
Ethers	66	6	6
Carboxylic acids	5	N/A	N/A
Polyfunctionals	39	N/A	N/A
Fuel mixtures	120	156	156
Total	603	375	371

3.2.3. Structural features extraction by QSPR-UOB 2.0

QSPR-UOB 2.0 functional group classification system is used to decompose fuel molecules into component

fragments and transform structural features into a fuel molecular structure matrix. The fuel molecular structure

matrix contains *32* columns, *n* rows and each fuel molecule corresponds to 1 row. Each column in a row represents

the occurrence of the corresponding functional group in QSPR-UOB 2.0. The QSPR-UOB 2.0 system upgrades

from the GCM 1.0 system of Chapter 1 which complements aromatic bond at position 1~9 and unbranched

aromatic bond (functional group type 1.1~1.10 in Figure 3.1) to describe the aromatics with one benzene ring, two

fused benzene rings (naphthyl group), three fused benzene rings. Similar to the GCM 1.0 system, the QSPR-UOB

2.0 system composes of 13 functional group identifiers, 18 functional group position descriptors and 6 fuel

reactivity descriptors as shown in Figure 3.1. The roles of the functional group identifiers, functional group position

descriptors and fuel reactivity descriptors have been discussed in section 2.2.1.

Figure 3.1. QSPR-UOB 2.0 for structural features extraction. The functional groups are listed on the left, and an example of each functional group is circled in the molecular structure on the right.

3.2.4 Training and validation of ML regression model

The fuel molecular structure matrix and fuel property matrix are prepared by UOB-QSPR 2.0 system and fuel

ignition quality database and then correlated by 19 ML algorithms as shown in Figure 3.2. Five regression model

types of linear regression, regression trees, support vector machines, Gaussian process regression, ensemble trees

are included which possess different model interpretability and flexibility as shown in Table 3.6. 10-fold cross

validation is adopted to examine the predictive accuracy of 19 ML regression models and use RMSE as criteria to

choose the best model. The cross validation scheme can not only examine the model performance to predict new

data but also protect against over-fitting especially for the flexible models. Advanced model training option can

be adjusted in the MATLAB regression learner APP to further improve the predictive accuracy [52].
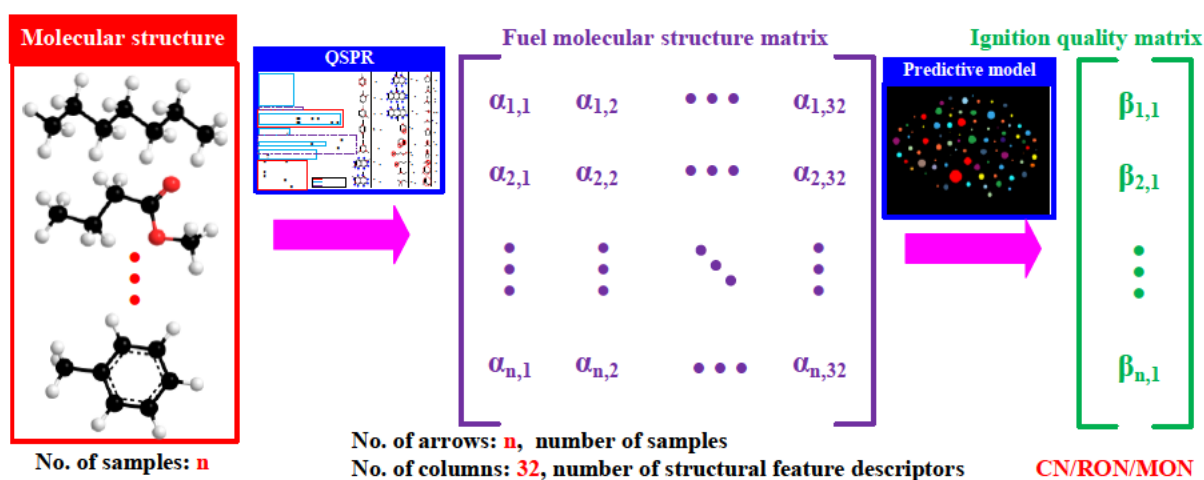
Figure 3.2. Flow chart of the CN/RON/MON prediction by coupling QSPR and ML regression model.

Table 3.6. 19 ML algorithms used to train the regression model [242]

Regression algorithm	Type	Interpretability	Model Flexibility
Linear regression	①Linear	Easy	Very low
	②Interactions linear	Easy	Medium
	③Robust linear	Easy	Very low. Less sensitive to outliers, but can be slow to train.
	④Stepwise linear	Easy	Medium
Regression trees	⑤Fine tree	Easy	High. Many small leaves for a highly flexible response function (Minimum leaf size is 4.)
	⑥Medium tree	Easy	Medium. Medium-sized leaves for a less flexible response function (Minimum leaf size is 12.)
	⑦Coarse tree	Easy	Low. Few large leaves for a coarse response function (Minimum leaf size is 36.)
Support vector machines	⑧Linear SVM	Easy	Low
	⑨Quadratic SVM	Hard	Medium
	⑩Cubic SVM	Hard	Medium
	⑪Fine Gaussian SVM	Hard	High. Allows rapid variations in the response function. Kernel scale is set to sqrt(P)/4, where P is the number of predictors.
	⑫Medium Gaussian SVM	Hard	Medium. Gives a less flexible response function. Kernel scale is set to sqrt(P).
	⑬Coarse Gaussian SVM	Hard	Low. Gives a rigid response function. Kernel scale is set to sqrt(P)*4.
Gaussian process regression	⑭Rational quadratic	Hard	Automatic
	⑮Squared exponential	Hard	Automatic
	⑯Matern 5/2	Hard	Automatic
	⑰Exponential	Hard	Automatic
Ensembles of trees	⑱Boosted trees	Hard	Medium to high. Least-squares boosting with regression tree learners.
	⑲Bagged trees	Hard	High. Bootstrap aggregating or bagging, with regression tree learners.

3.3 Results and discussion

3.3.1 Predictive accuracy of CN/RON/MON

The R^2 of CN, RON, MON between measured and predicted values are shown in Figure 3.3 and the left column and right column are the model performances trained by the pure compound dataset and pure compound & mixture dataset. The functions and parameters of these 6 models are presented in Table 3.7. The models trained by pre compound & mixture dataset outperform those trained by the pure compound dataset and their R^2 & RMSE of CN. RON, MON reach 0.9911 & 2.526, 0.9874 & 2.454, 0.9731 & 2.765 as shown in Table 3.8. Based on the ML theory of the bigger the data the better the model, the full dataset contains additional 120, 156, 156 samples of CN, RON, MON than the pure compound dataset. These mixture samples facilitate the model training by refining the contribution of the functional groups in QSPR-UOB 2.0 on the ignition quality. The models trained by pure compounds & mixtures dataset are adopted in the following sections unless otherwise specifies in the following context. The subgroup R^2 of ML-QSPR models for CN/RON/MON prediction is compared with various published methods as shown in Table 3.9. The ML-QSPR models obtain the best overall as well as subgroup predictive accuracy among these models. The GCM model proposed by Kubic et al. [208, 209] achieves R^2 of 0.90, 0.93, 0.91 for CN, RON, MON and their RMSE data are not available, but it is reasonable to infer that their RMSE are greater than the current ML-QSPR models.

The predictive residuals of typical compound groups (left column) and within specified ranges (right column) for CN, RON, MON are presented in Figure 3.4. The left column intends to examine the model predictive accuracy for different chemical families while the right column aims at exploring the model forecasting accuracy for various ranges. The box and whisker chart have the greatest maximum & minimum residuals for CN, RON, MON of 6.91 & -6.65 (contributed by alkanes), 5.42 & -18.85 (contributed by alkynes), 4.24 & -6.32 (contributed by alkanes) as shown in Figure 3.4 (a), Figure 3.4 (b), Figure 3.4 (c). The large RON predictive residuals of alkynes are caused by the insufficient samples (4 compounds) as shown in Table 3.5. The minimum and maximum RON predicted

residual of other compound groups are within the range of -6.65~6.91 as shown in Figure 3.4 (b) and it indicates that the ML-QSPR models enable accurate forecasting for different chemical classes. If more experimental data are available for typical compound groups, it would enhance the model interpolation and extrapolation capability and reduce the predictive residual. The predictive residuals at the range of -20~-10 (two samples, propane and 3,3-dimethylpentane) for CN are abnormally high as shown in Figure 3.4 (d). The outliers may contain measurement uncertainty in addition to predictive error, thus repeatability and reproducibility tests are needed for these compounds. For example, the CN of propane (-20) reported in the Compendium of Experimental Cetane Numbers [193] are collected from ref. [243] which is computed by the Ab initio molecular orbital calculations. It combines both Hartree–Fock density functional theory (HF–DFT) method and the 6-311G(p,d) basis set but it should be further verified by the experimental test [243].
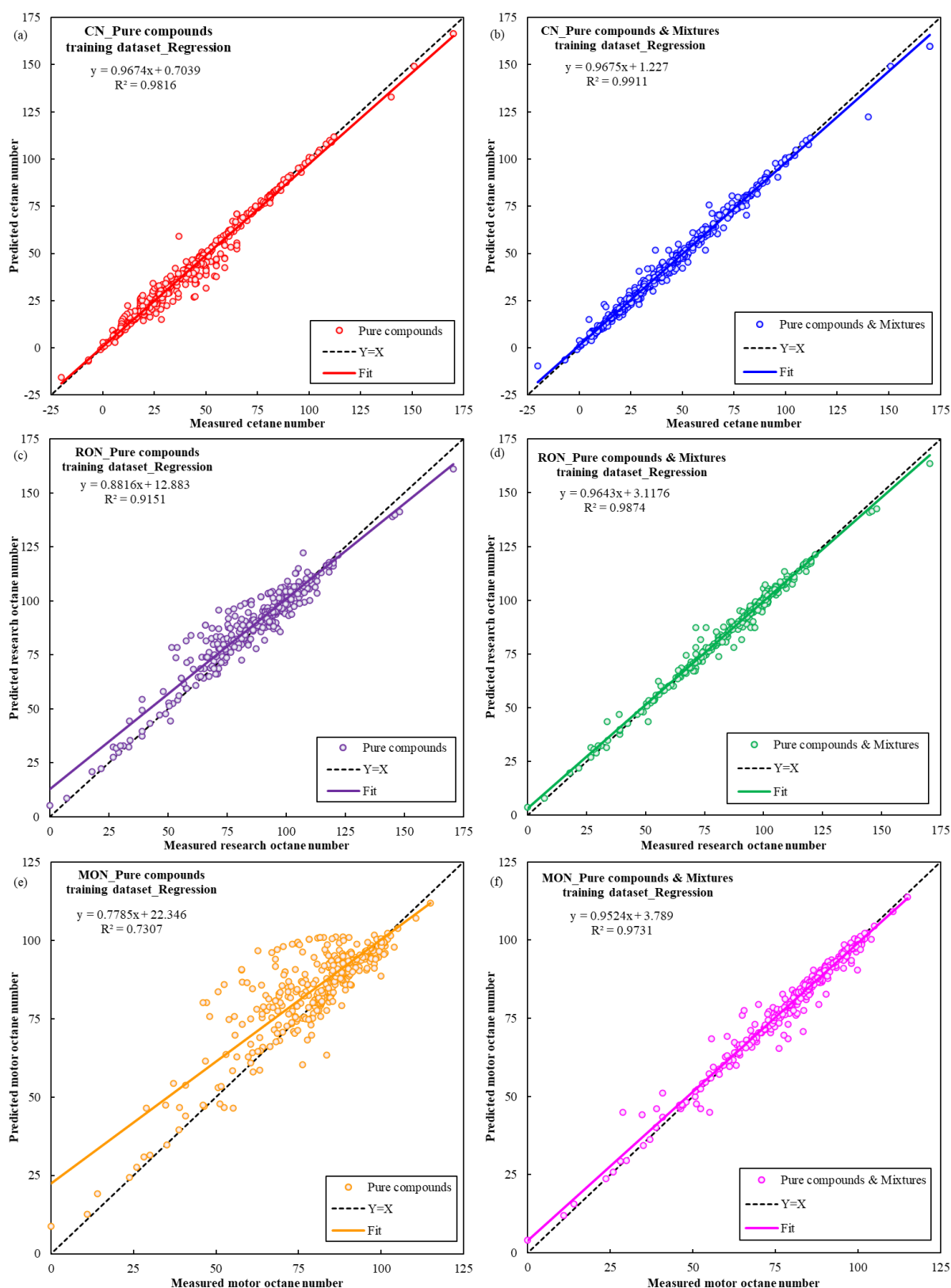
Figure 3.3. Parity plots for (a)-(b) CN, (c)-(d) RON, (e)-(f) MON between measured and predictive values by ML regression model. The regression models in the left and right columns are trained by the pure compounds dataset and full (pure compounds & mixtures) dataset.
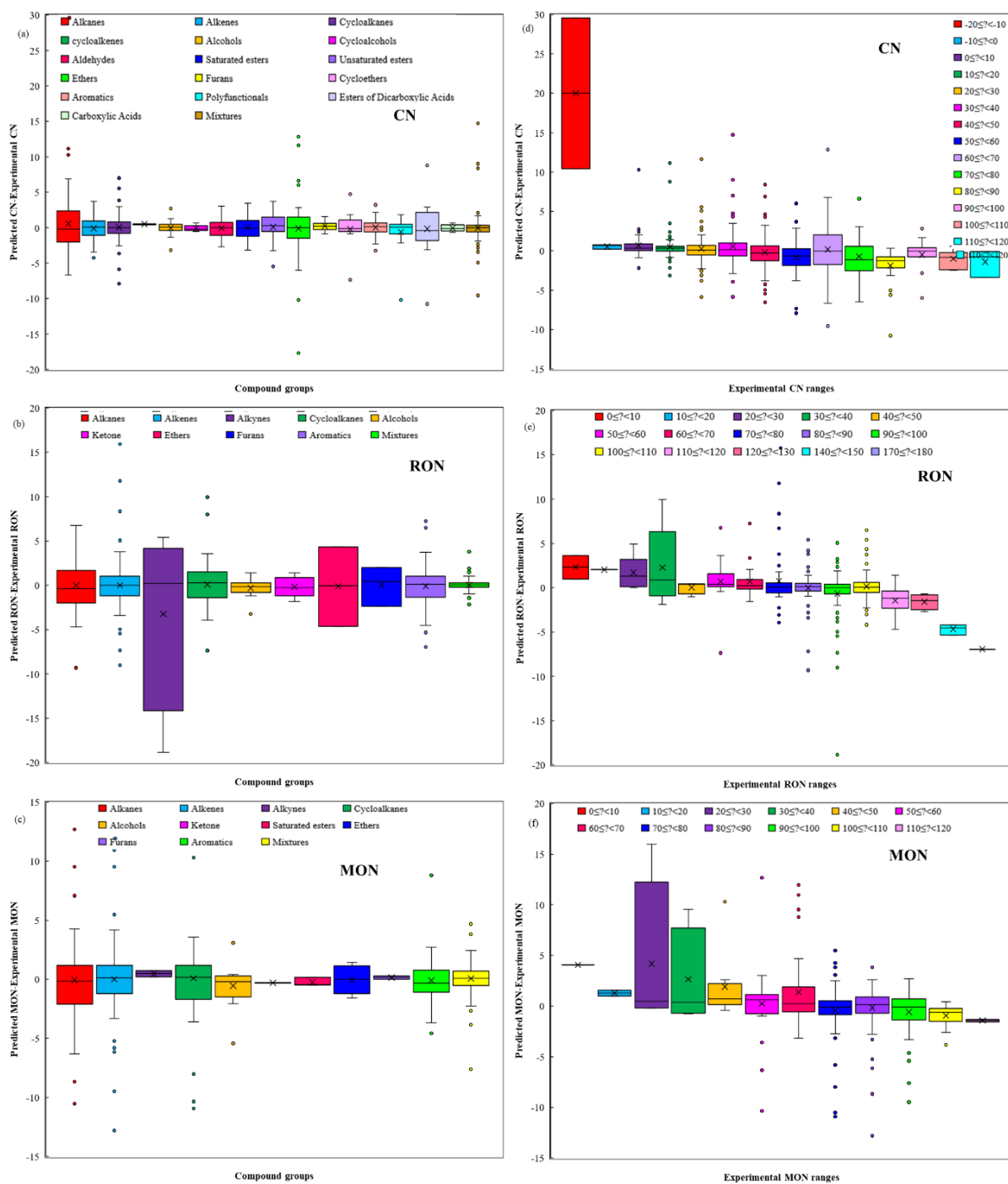
Figure 3.4. Predictive residuals of typical compound groups and within specified ranges for (a), (d) CN; (b), (e) RON; (c), (f) MON.

Table 3.7. Functions and parameters of the ML regression models

Property	Training compounds (No.)	MATLAB module	Algorithm	Kernel function	Basic function	Optimizer	Nonlinear programming solver
CN	Pure compounds (483)	Regression learner	GPR	Exponential	Linear	Quasinewton	Active-Set method
	Pure compounds & Mixtures (603)	Regression learner	GPR	Exponential	Constant	Quasinewton	Active-Set method
RON	Pure compounds (217)	Regression learner	GPR	Exponential	Constant	Quasinewton	Active-Set method
	Pure compounds & Mixtures (373)	Regression learner	GPR	Exponential	Constant	Quasinewton	Active-Set method
MON	Pure compounds (215)	Regression learner	GPR	Matern 5/2	Constant	Quasinewton	Active-Set method
	Pure compounds & Mixtures (371)	Regression learner	GPR	Matern 5/2	Constant	Quasinewton	Active-Set method

Table 3.8. Statistical analysis of predictive performance for the machine learning regression models

Property	Training dataset	R^2	MAE	RMSE
CN	Pure compounds (483)	0.9816	1.891	3.580
	Pure compounds & Mixtures (603)	0.9911	1.460	2.526
RON	Pure compounds (217)	0.9151	4.543	6.795
	Pure compounds & Mixtures (373)	0.9874	1.386	2.454
MON	Pure compounds (215)	0.7307	6.500	9.922
	Pure compounds & Mixtures (371)	0.9731	1.567	2.765

Table 3.9. Comparison of correlation coefficients of different compound groups between the current study and published methods

R^2	CN					RON			MON		
Compound class	Current	Saldana et al. [212]	Kubic et al. [208, 209]	DeFries et al. [210]	Dahmen et al. [157]	Current	Kubic et al. [208, 209]	Albahri [244]	Current	Kubic et al. [208, 209]	Albahri [244]
Paraffins	0.9866	N/A	0.91	0.73	0.53	0.9902	0.94	0.86	0.9765	0.95	0.87
Olefins and alkynes	0.992	N/A	0.90	N/A	-0.48	0.9279	0.90	0.53	0.8839	0.65	-1.55
Naphthenes	0.9599	N/A	0.81	N/A	0.25	0.9839	0.85	0.75	0.9504	0.89	-0.40
Aromatics	0.9946	N/A	0.87	0.44	0.58	0.9896	0.76	-3.28	0.9722	N/A	N/A
Oxygenates	0.993	N/A	0.85	N/A	0.41	0.9821	0.62	N/A	0.9767	0.56	N/A
Alcohol	0.9977	N/A	N/A	N/A	N/A	0.9945	N/A	N/A	0.9433	N/A	N/A
Aldehydes/Ketones	0.9973	N/A	N/A	N/A	N/A	1	N/A	N/A	1	N/A	N/A
Saturated esters	0.9956	N/A	N/A	N/A	N/A	0.9977	N/A	N/A	0.9991	N/A	N/A
Unsaturated esters	0.9881	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ethers	0.9905	N/A	N/A	N/A	N/A	0.9414	N/A	N/A	0.9943	N/A	N/A
Carboxylic acids	0.9996	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Polyfunctionals	0.9937	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Fuel mixtures	0.98	N/A	N/A	N/A	N/A	0.9982	N/A	N/A	0.9908	N/A	N/A
Overall	0.9911	0.934	0.90	0.64	0.53	0.9874	0.93	0.55	0.9731	0.91	-1.16

3.3.2 Method application: characterizing the impact of fuel molecular structure on the ignition quality

This section displays the functionality of characterizing the impact of molecular structure on the ignition quality for alkanes (section 3.3.2.2), alkenes (section 3.3.2.3), naphthenes (section 3.3.2.4), aromatics (section 3.3.2.5), alcohols (section 3.3.2.6), esters (section 3.3.2.7). The ignition quality comparison of aldehyde, n-alkane, alcohol, ether, alkene, ketone, ester is showcased in section 3.3.2.1. Unless otherwise specified, RON is chosen as ignition quality evaluation index for the C1~C11 molecules because they have good ani-autoignition propensity and the measured RON rather than measured CN. If the studied compounds contain 12 carbon atoms or greater, CN would be used to evaluate the ignition quality because these molecules are prone to autoignition and their measured CN are usually available compared to RON.

3.3.2.1 Comparison of ignition quality for different fuel types

Current ML-QSPR models provide an efficient way to compare the ignition quality of multiple fuel types when rare experimental data is available. A case study is conducted to investigate the ignition quality of aldehyde (pentanal), n-alkane (n-pentane), alcohol (1-pentanol), ether (methyl butyl ether), alkene (cis 2-pentene), ketone (2-pentanone), ester (methyl butanoate) with 5 carbon atoms as shown in Figure 3.5. Since the carbon chain lengths of these fuel molecules are relatively short and they are less prone to occur autoignition, thus RON is used as an ignition quality indicator. The RON of these fuels orders from lowest to highest as: pentanal<n-pentane<1-pentanol<methyl butyl ether<cis 2-pentene<2-pentanone<methyl butanoate. Therein, the measured RON of pentanal, methyl butyl ether, cis 2-pentene, 2-pentanone, methyl butanoate have not been reported in the literature and this work successfully overcome this dilemma and it is extremely important for the fuel screening. The results in Figure 3.5 also indicate that hydrothy group, ether group, carbon-carbon double bond, ketone group, and ester group can boost the RON compared to the equivalent n-alkane with 5 carbon atoms. On the contrary, the aldehyde group slightly decreases the RON compared to the C5 n-alkane. But the major influence factors of ignition quality

are the carbon atom number (carbon chain length for hydrocarbons) and branching degree of the fuel molecules.

By this ML-QSPR model, the users can easily obtain the ignition quality information of the interested fuel molecules since the functional group interaction has been considered by the Gaussian process regression model.
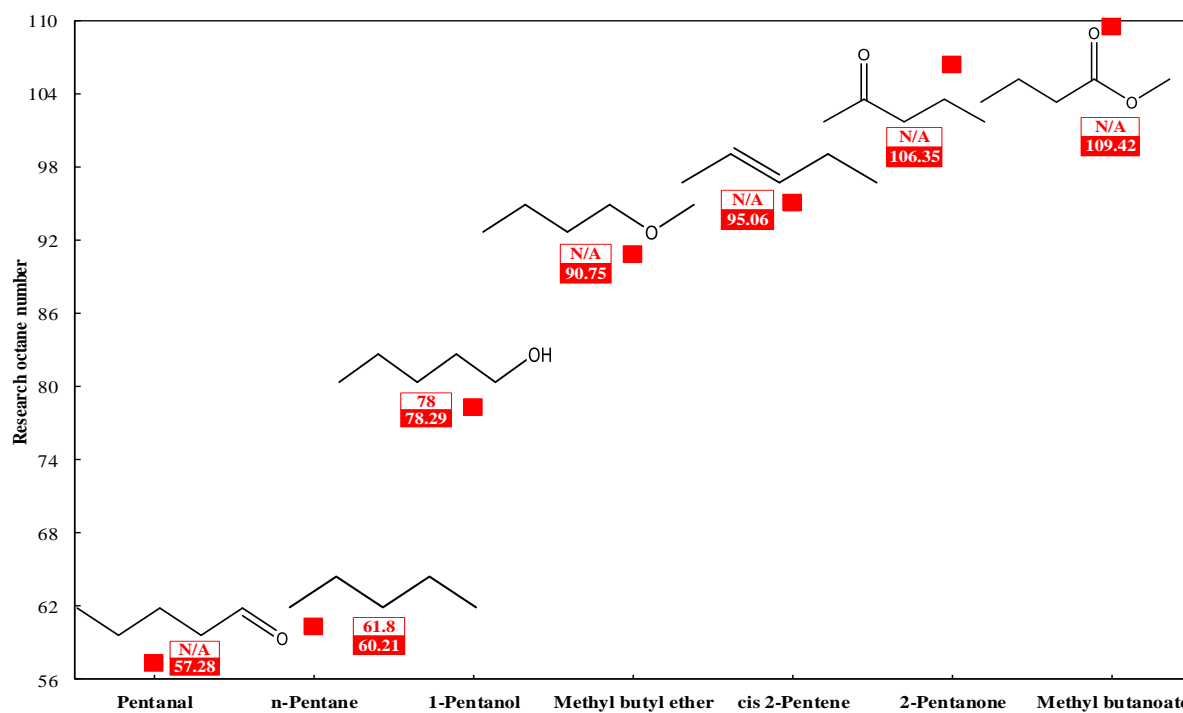
Figure 3.5. RON of different fuel types with 5 carbon atoms, numbers with red frames and red backgrounds are measured values and predictive values.

3.3.2.2 Impact of alkanes structural features on ignition quality

The RON of C1~C8 n-alkanes and corresponding isomers are predicted by the ML-QSPR model and the C4, C5, C6, C7 and C8 alkanes have 2, 3, 5, 9, 18 isomers respectively as shown in Figure 3.6. It indicates that the proposed ML-QSPR model enables to study the impact of molecular structure variation (increasing carbon chain length, molecular centralization, methyl group addition, etc.) on the ignitability property. There are four ways to increases the RON of fuel molecules by modifying the molecular structure: (1) Shorten the carbon chain length or reduce the number of carbon atoms to increases RON. For example, the measured RON of n-heptane, n-hexane, n-pentane, n-butane, propane, ethane increase as 0, 29, 61.8, 94, 111, 114.9 and the predicted values can accurately reproduce this tendency. The predicted RON of n-octane increases slightly to 8.53 but the true value should be negative. iso-Octane and n-heptane are the primary reference fuels of the RON rating which corresponds to 100 and 0 respectively [189]. The n-octane (CN=64.4) is more reactive than n-heptane (CN=56) based on their CN rating [193], thus the RON of n-octane should be lower than that of n-heptane (RON=0). More measured RON data for those molecules with 8 carbon atoms or greater are needed to improve predictive accuracy. (2) Adding methyl group into fuel molecule would either increases or decreases RON, thus concrete analysis should be made according to concrete circumstance. For example, adding methyl group into 2-methylbutane (measured RON: 93) at position 2 and position 3 produce 2,2-dimethylbutane (measured RON: 91.8) and 2,3-dimethylbutane (measured RON: 104.3). (3) Increasing the branching degree or the number of side-chain can improve the RON at a given carbon atom number. For example, the predicted RON of n-octane, 2-methylheptan, 2,3-dimethylhexane, 2,3,4-trimethylpentane, 2,2,3,3-tetramethylbutane increases as 8.53, 21.83, 68.56, 103.27, 118.54 as shown in Figure 3.54. (4) Moving the side chain toward the center of the fuel molecule can increase the RON. For example, the predicted RON of n-octane, 2-methylheptane, 3-methylheptane, 4-methylheptane, 3-ethylhexane, 3,3-dimethylhexane increases as 8.53, 21.83, 27.06, 31.6, 31.6, 62.26 as shown in Figure 3.7.

Figure 3.6. RON of C1~C8 alkanes, numbers with red frames and red backgrounds are measured values and predictive values.

Figure 3.7. Impact of branching/centralization on RON of C8 alkanes, numbers with red frames and red backgrounds are measured values and predictive values.
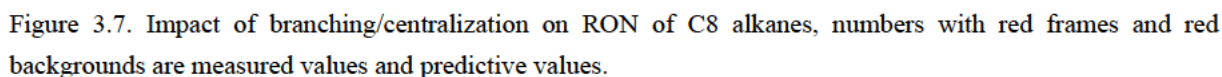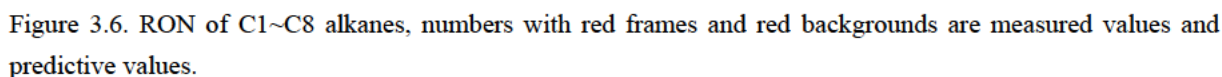
3.3.2.3 Impact of alkenes structural features on ignition quality

Alkenes are the main components of gasoline up to 10 vol.% which is a good component to increases octane level [245, 246]. It results in deposit formation in injector and combustion chamber and increases reactive hydrocarbons and toxic compounds emissions, thus the gasoline regulation constrains the maximum allowed olefin content. Appropriate design of the alkenes can increase the gasoline anti-knock performance [115, 247]. The measured and predicted RON of C2~C11 straight-chain alkenes are shown in Figure 3.8 and the results are summarized below: (1) The RON decreasing slope of 1-alkenes is smaller than that of alkanes. The RON of C2~C3 1-alkenes are smaller than counterpart n-alkanes. The RON of 1-alkenes are greater than corresponding n-alkanes as the number of carbon atoms is equal to or greater than 4. Thus, replacing n-alkanes with counterpart 1-alkenes in the gasoline can boost the anti-knock property, For example, the predicted RON of 1-butene, 1-pentene, 1-hexene, 1-heptene are 6.09, 28.17, 47.22, 52.26 greater than the counterpart n-alkanes which accurately reproduces the observed values. (2) The RON of straight-chain alkenes increases as the carbon-carbon double bond moving to the molecular center. For example, the predicted RON of 1-undecene, 2-undecene, 3-undecene, 4-undecene, 5-undecene increase as 26.88, 27.97, 28.28, 30.6, 34.84 as shown in Figure 3.8.

The measured and predicted RON of C4~C8 branched chain alkenes are shown in Figure 3.9. Carbon-carbon double bond centralization would increase or decrease the RON compared to the counterpart alkanes. For example, the predicted RON of 2,4,4-trimethyl-1-pentene and 2,4,4-trimethyl-2-pentene are 105.73 and 99.46 respectively and that of 2,2,4-trimethylpentane is 105.42 as shown in Figure 3.9. In general, the impact of carbon-carbon bond centralization on RON varies from molecule to molecule and this ML-QSPR model can draw a full picture of the ignition quality for all the isomers at a given carbon atom number.

Figure 3.8. RON of C2~C11 straight chain alkenes, numbers with red frames and red backgrounds are measured values and predictive values.

Figure 3.9. RON of C4~C8 branched chain alkenes, numbers with red frames and red backgrounds are measured values and predictive values.
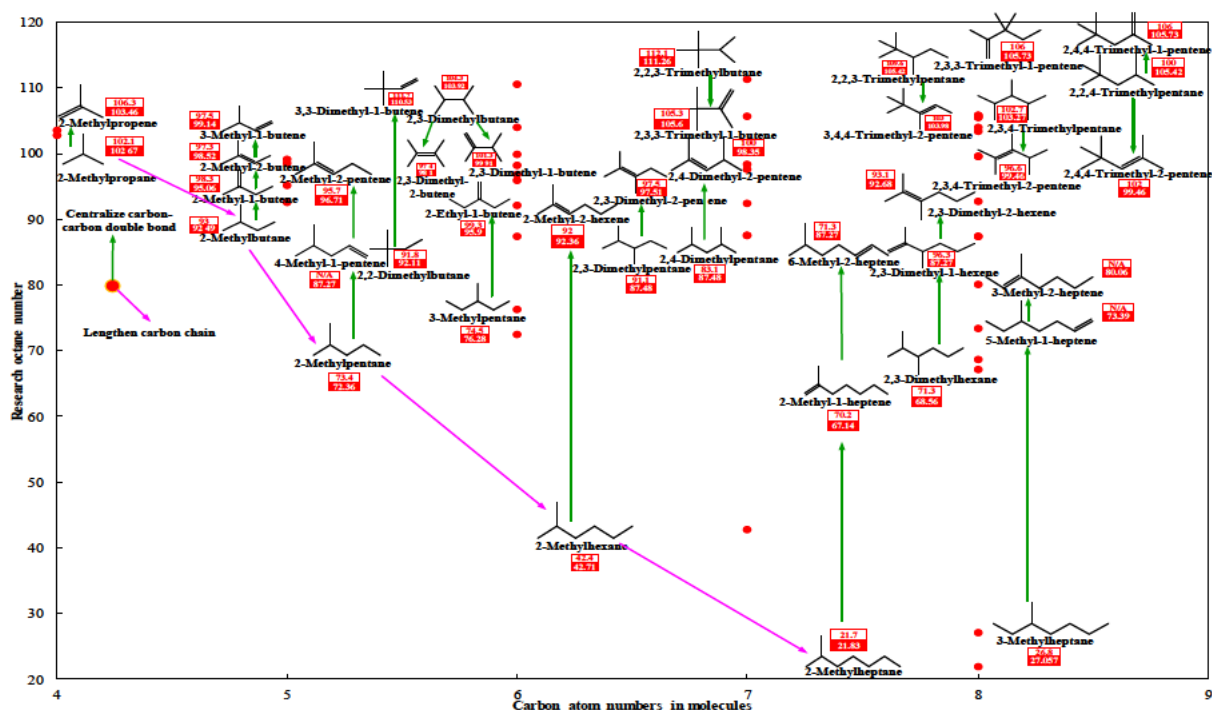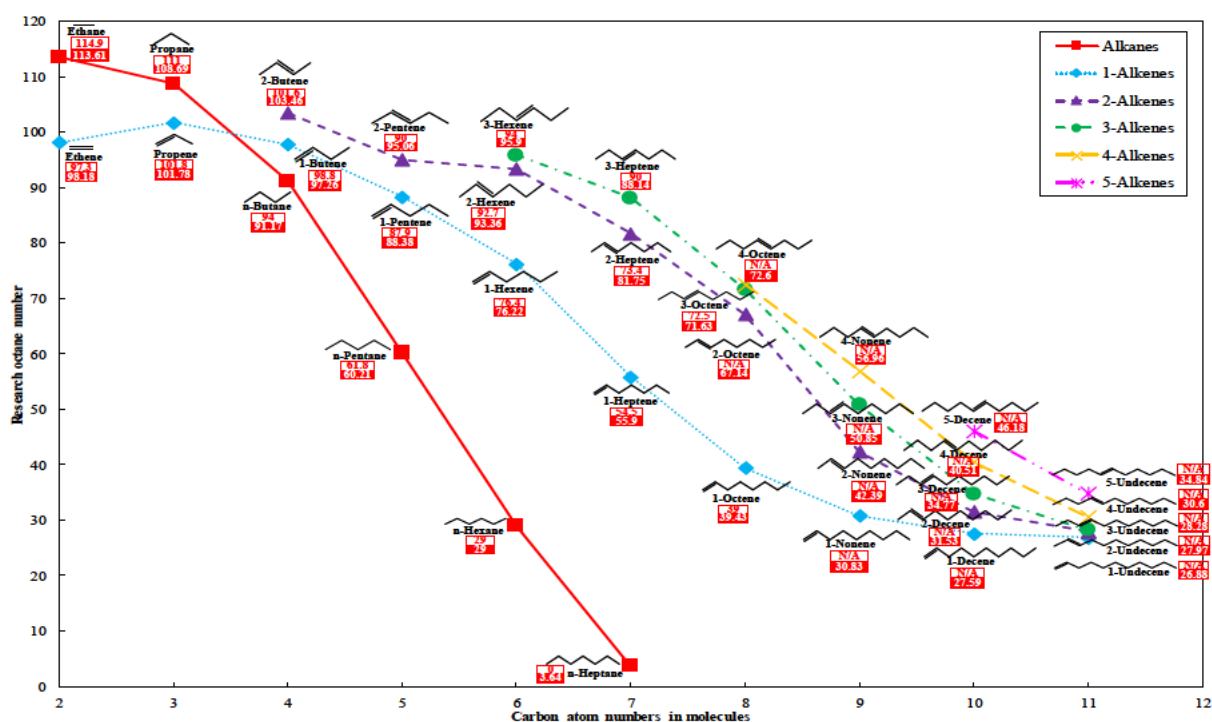
3.3.2.4 Impact of naphthenes structural features on ignition quality

The most prevalent napththenes (also called cycloalkanas or cycloparaffins) in gasoline, diesel and aviation fuels have ring with five and six carbon atoms [248-250]. The naphthenes have a wide range of octane levels similar to alkanes and the RON of C3~C7 alkyl cyclopropanes, C5~C9 alkyl cyclopentanes, C6~C10 alkyl cyclohexanes are compared in Figure 3.10, Figure 3.11, Figure 3.12 respectively. The impacts of molecular structure on RON for naphthenes are summarized below: (1) The RON of the simplest nathelenes are greater than the corresponding n-alkanes except for cyclooctane. For example, the predicted RON of Cyclopentane (98.6), Cyclohexane (80.59) are 38.39 and 51.59 greater than n-pentane (60.21), n-hexane (29) while the RON of Cyclopropane (101.82) is 6.87 lower than n-propane (108.69). (2) The RON decreases with increasing ring size. Take the methyl-cycloalkanes as an example, the predicted RON of methylcyclopropane, methylcyclopentane, methylcyclohexane decreases as 95.09, 88.45, 74.05. Ethyl-cycloalkanes show a similar tendency, the predicted RON of ethylcyclopropane, ethylcyclopentane, ethylcyclohexane decreases as 78.57, 65.91, 46.93. The impact of side-chain size. number and position on the RON of naphthenes are summarized below: (1) Shortening the side chain length and increasing the branching degree benefit the RON enhancement. For example, the predicted RON of butylcyclohexane, propylbenzene, ethylcyclohexane, methylcyclohexane increases as 13.13, 19.84, 46.93, 74.05 as increasing chain length. The predicted RON of Butylcyclohexane, isoButylcyclohexane, Tert-butylcyclohexane increases as 13.13, 43.63, 95.84 with increasing branching degree. (4) Splitting the carbon atoms in a single side chain into two or more side chains increases the octane rating. For example, the RON of butylcyclohexane, 1-methyl-2-propylcyclohexane, 1-ethyl-3,5-dimethylcyclohexane, 1,2,3,5-tetramethylcyclohexane as 13.13, 32.24, 61.69, 85.63 as side-chain distribution. (5) Combining two side chains on the same carbon atom on the ring has high octane rating, For example, the predicted RON of 1,2,3,5-tetramethylcyclohexane, 1,1,3,5-tetramethylcyclohexane, 1,1,4,4-tetramethylcyclohexane increases as 85.63, 89.64, 100.14. (6) The naphthenes with the less symmetric structure have greater RON. For those naphthenes with

two side groups, the RON increases in the order of 1,4-naphthene, 1,3-naphthene, 1,2-naphthene. For example, the

1,4-dimethylcyclohexane, 1,3-dimethylcyclohexane, 1,2-dimethylcyclohexane have RON of 68.43, 71.36, 80.16.
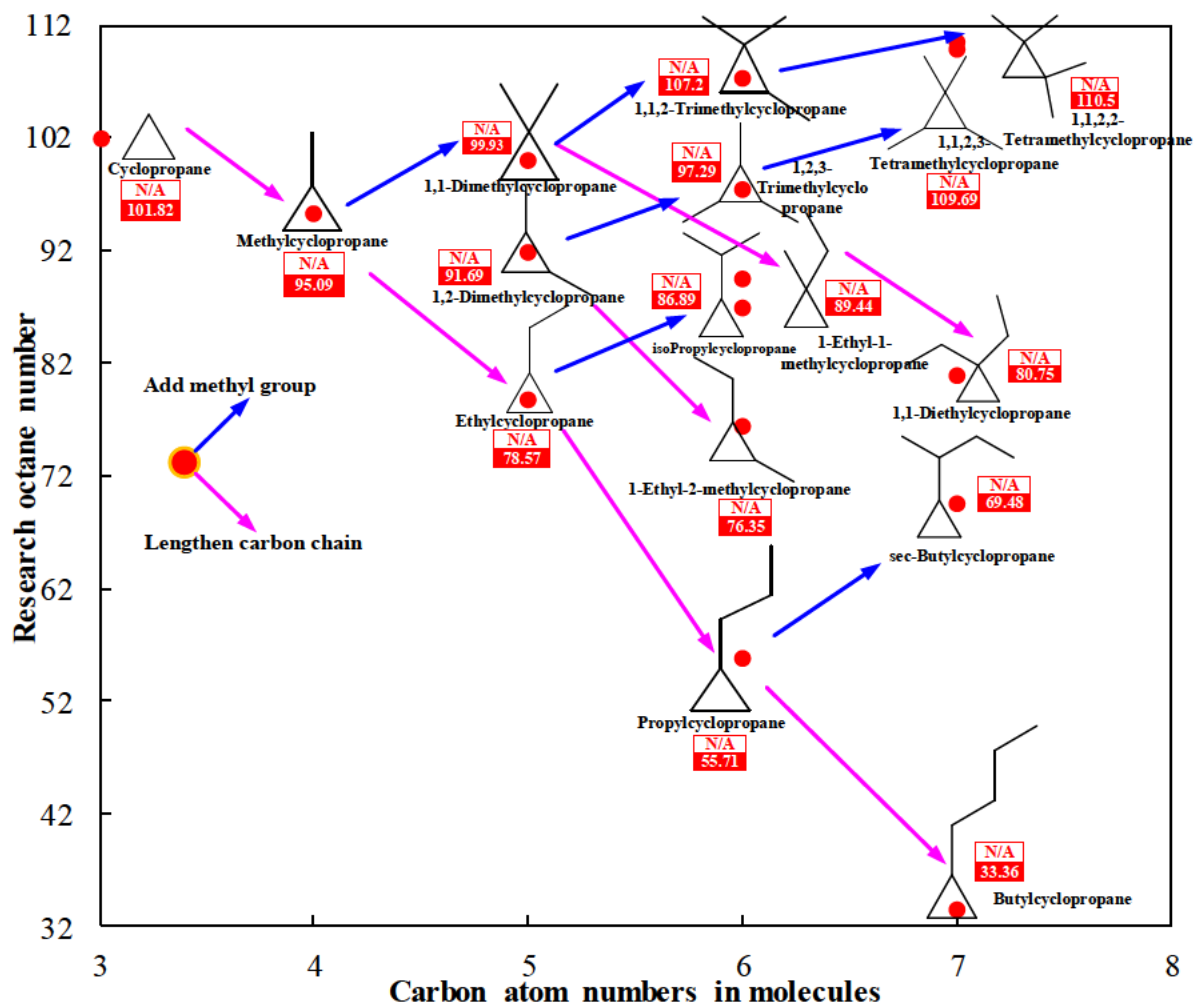
Figure 3.10. RON of C3~C7 alkyl cyclopropanes, numbers with red frames and red backgrounds are measured values and predictive values.

Figure 3.11. RON of C5~C9 alkyl cyclopentanes, numbers with red frames and red backgrounds are measured values and predictive values.

Figure 3.12. RON of C6~C10 alkyl cyclohexanes, numbers with red frames and red backgrounds are measured values and predictive values.
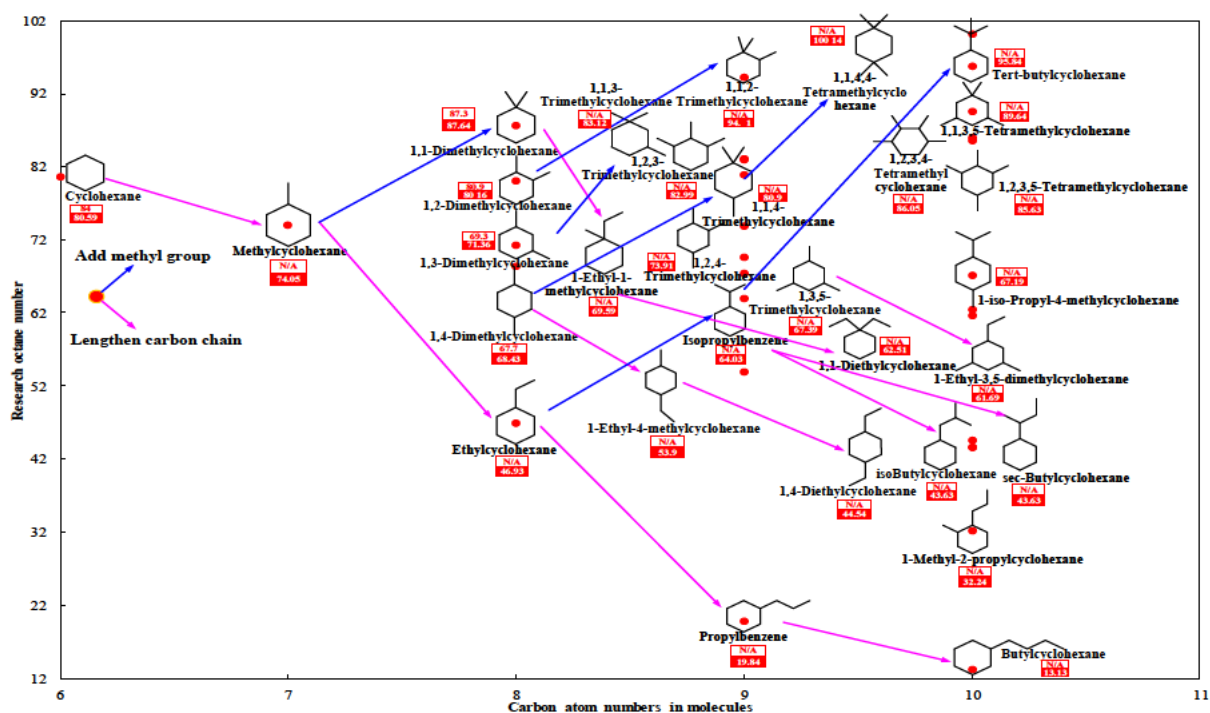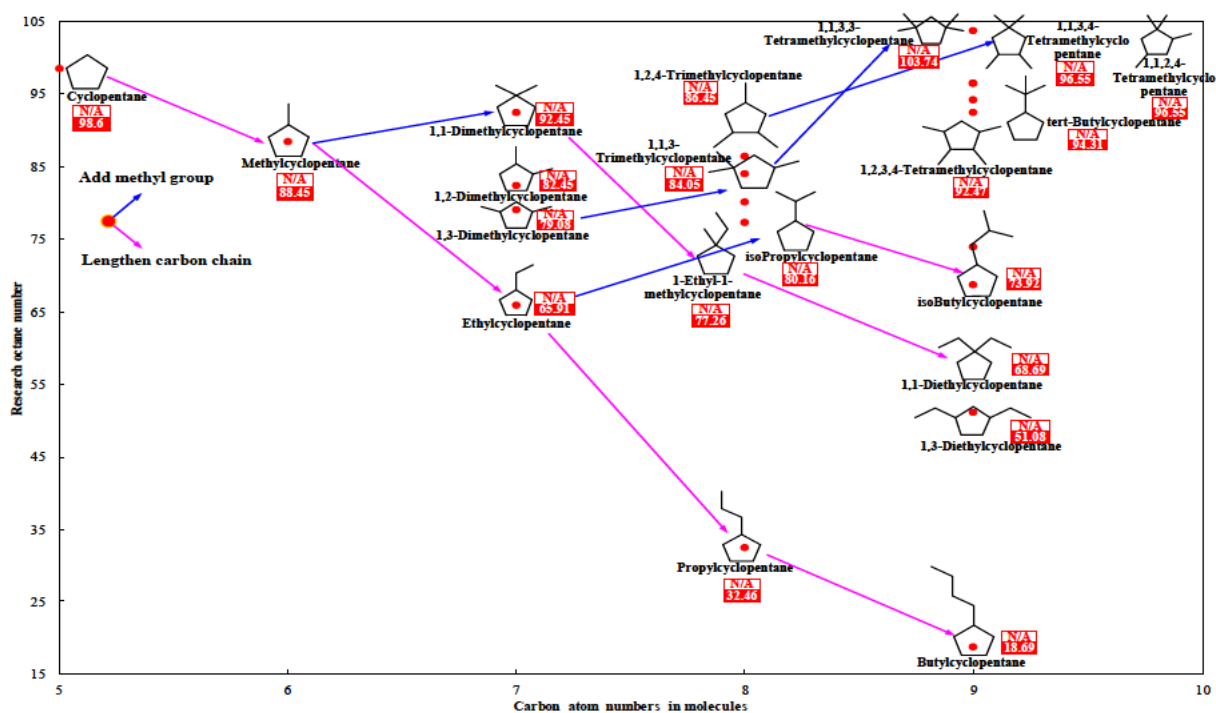
3.3.2.5 Impact of aromatics structural features on ignition quality

Aromatics represent a class of hydrocarbons with six-membered ring which contains three conjugated double bonds [251]. Especially, the components with two or more aromatic rings fused on shared carbon atoms are called polycyclic aromatics [249]. The resonance structure of aromatics results in a low reactivity and thus most of them have a high RON above 100 as shown in Figure 3.13. The impact of the size, number, position of the aromatic substituted groups on the RON are summarized below: (1) For aromatics with one side chain, the RON increases as decreasing chain length and increasing branching degree. For example, the predicted RON of butylbenzene, propylbenzene, ethylbenzene, toluene increases as 97.91, 101.17, 107.62, 117.33 with shortening substituted group chain length. The butylbenzene, isobutylbenzene, tert-butylbenzene have RON of 97.91, 99.34, 104.43 as increasing branching degree of the side chain. (2) For aromatics with two side chains, the RON decreases as para-isomer>meta-isomer>ortho-isomer. For example, the predicted RON of p-xylene, m-xylene, o-xylene decreases as 141.49>140.79>118.2. (3) For aromatics with three side chains, the most symmetric compounds have the highest RON. For example, the predicted RON of 1,2,3-trimethylbenzene, 1,2,4-trimethylbenzene, 1,3,5-trimethylbenzene increases as 117.15, 142.68, 163.65 with a more symmetric structure.
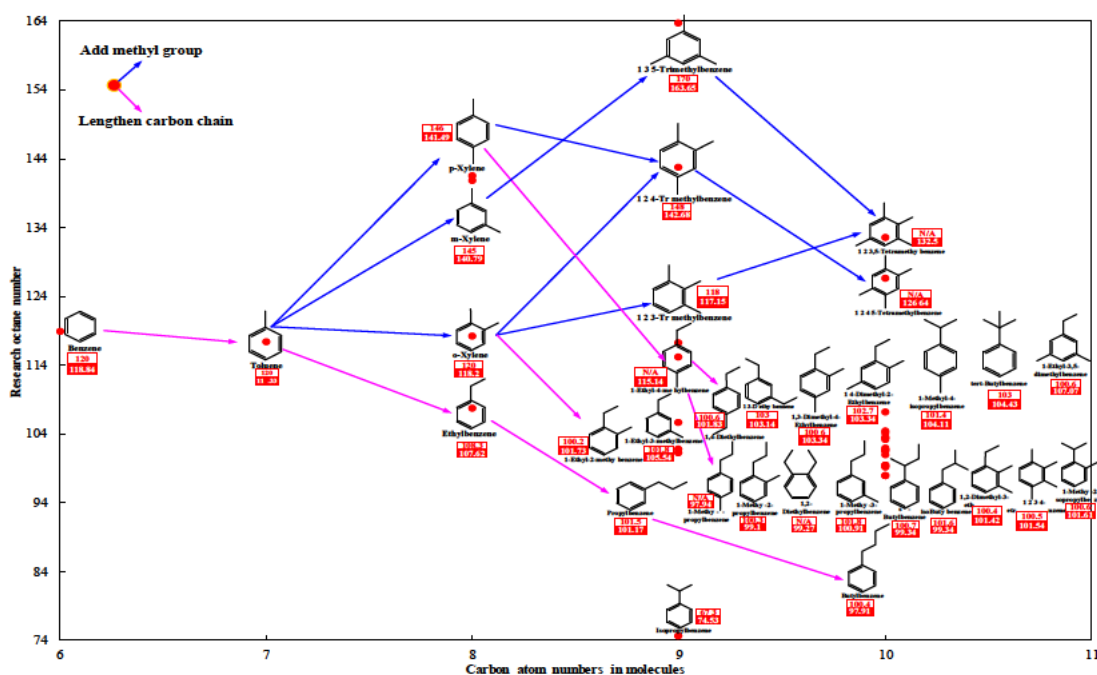
Figure 3.13. RON of C6~C11 aromatic hydrocarbons, numbers with red frames and red backgrounds are measured values and predictive values.

3.3.2.6 Impact of alcohols structural features on ignition quality

The measured and predicted RON of C1~C5 alcohols are compared in Figure 3.14 and the C3, C4, C5 alcohols have 2, 4, 8 isomers respectively. The impacts of molecular structure on the RON of alcohols are summarized below: (1) For the straight chain alcohols, the octane rating increases with shortening chain length. For example, the predicted RON of 1-pentanol, 1-butanol, 1-propanol, ethanol, methanol increases as 78.3, 94.81, 102.82, 110.83, 121.3 with decreasing chain length. (2) The effect of adding methyl group on RON varies with various types of alcohols. For example, the predicted RON decreases from 116.62 of 2-propanol to 106.84 of tert-butanol after adding methyl group. Another situation is that the RON increases from 102.82 of 1-propanol to 106.53 of 2-methyl-1-propanol after adding methyl group. (3) The influence of molecular centralization on RON of alcohols varies from case to case. For example, the predicted RON reduces from 109.76 of 3-Methyl-1-butanol to 104.74 of 2-Methyl-1-butanol as the methyl group moving from the 3rd carbon atom to .2nd carbon position. On the contrary, the predicted RON increases from 109.76 of 3-methyl-1-butanol to 114.18 of 2,2-dimethylpropanol with molecular centralization because the backbone chain length reduces from 4 to 3.
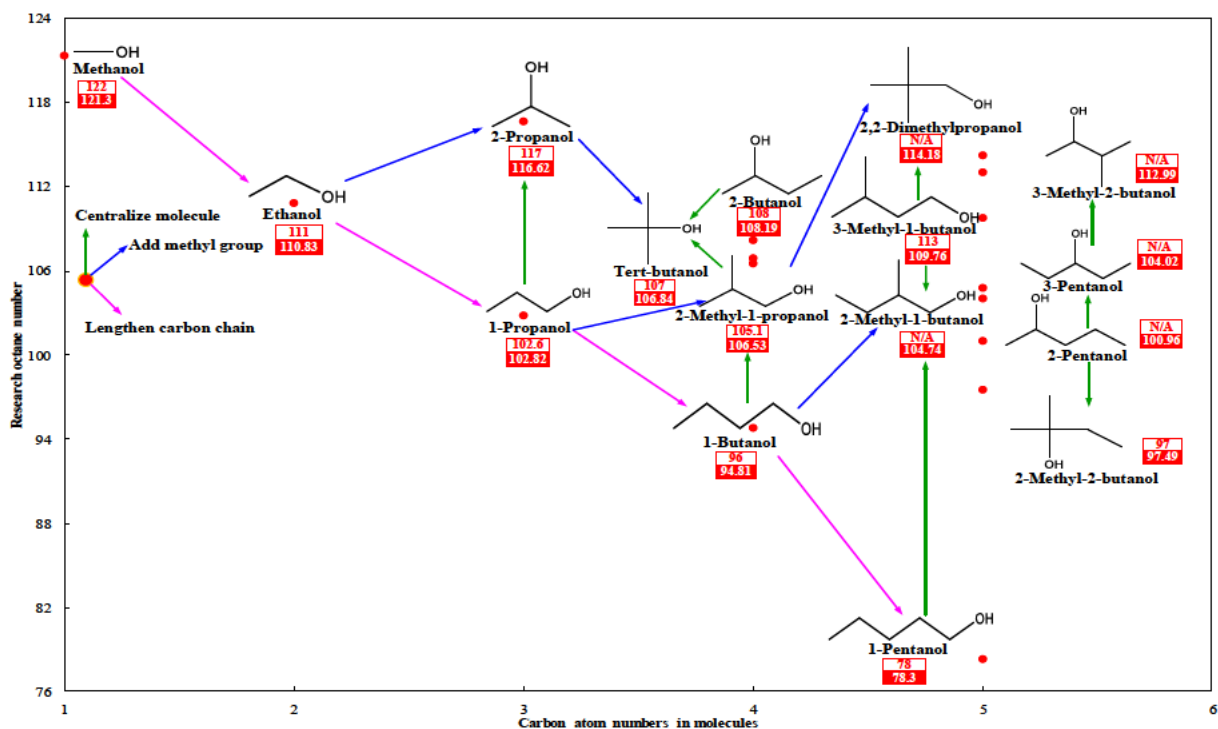
Figure 3.14. RON of C1~C5 alcohols, numbers with red frames and red backgrounds are measured values and predictive values.

3.3.2.7 Impact of esters structural features on ignition quality

Long-chain methyl esters of methyl palmitate, methyl stearate, methyl oleate, methyl linoleate, methyl linolenate are the main components of the soy and rapeseed biodiesels [141]. The methyl decanoate [252], ethyl-5-decenoate and methyl-9-decenoate [165] with mediate chain length are often used as biodiesel surrogate. The measured ignition quality of FAME is usually reported as CN rather than RON/MON as shown in Table 3.10, so the CN is adopted for ignition quality comparison in this section. The impact of carbon chain length, unsaturated bond position, unsaturated degree on the CN of methyl esters are investigated in Figure 3.15 which are summarized as below: (1) The CN increases with carbon chain length. For example, the predicted CN of methyl butanoate, methyl pentanoate, methyl hexanoate, methyl heptanoate, methyl octanoate, methyl nonanoate, methyl decanoate, methyl undecanoate, methyl laurate, methyl palmitate, methyl stearate increases with carbon chain length as 8.80, 13.72, 21.47, 32.77, 36.93, 43.00, 48.84, 59.64, 64.13, 83.91, 85.51. (2) Increasing the number of unsaturated carbon-carbon double bond decreases the CN. The predicted CN of methyl stearate, methyl oleate, methyl linoleate, methyl linolenate decreases with increasing number of carbon-carbon double bond as 85.51, 57.2, 42.5, 40.44. (3) Moving the carbon-carbon double bond toward the molecular edge increases the CN at an identical unsaturated degree. For example, the predicted CN increases from 35.51 of methyl-5-dodecenoate to 46.59 of methyl-2-dodecenoate as the carbon-carbon double bond moving away from the center of the molecule. The measured CN of unsaturated methyl esters are rarely reported and this work can enrich the fuel ignition quality dataset for fuel screening.
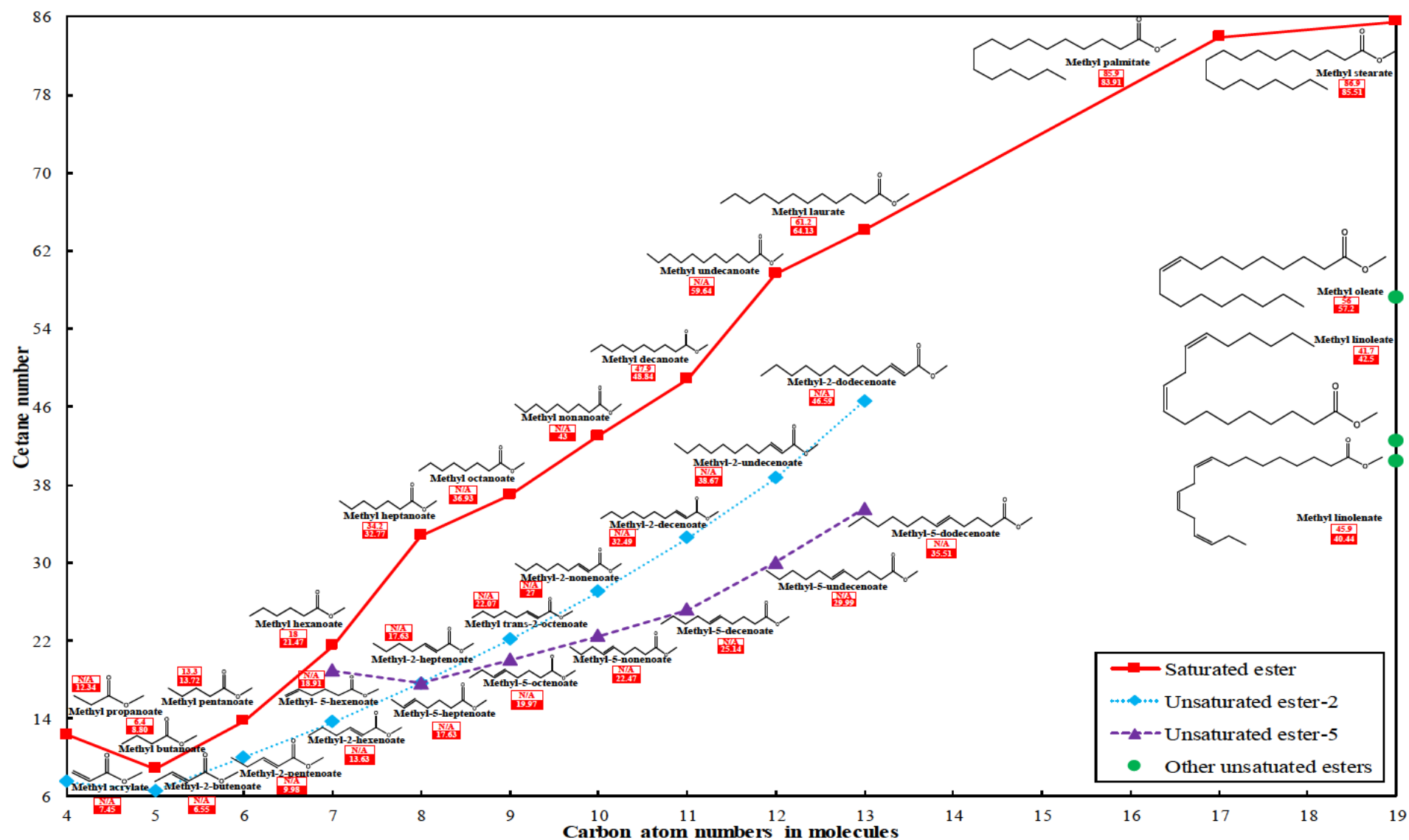
Figure 3.15. CN of C4~C19 esters, numbers with red frames and red backgrounds are measured values and predictive values.

Table 3.10. CN/RON/MON data for the C4~C19 FAME

Type	Compound name	CAS	Formula	CN		RON		MON	
				Measured	Predicted	Measured	Predicted	Measured	Predicted
Saturated ester	Methyl propanoate	554-12-1	C4H8O2	N/A	12.3374537	N/A	114.5810038	N/A	99.70427057
	Methyl butanoate	623-42-7	C5H10O2	6.4	8.803907176	N/A	109.4152828	N/A	103.1973804
	Methyl pentanoate	624-24-8	C6H12O2	13.3	13.72257375	105	104.4873508	105	104.5531942
	Methyl hexanoate	106-70-7	C7H14O2	18	21.47406229	N/A	95.97552268	N/A	102.4482313
	Methyl heptanoate	106-73-0	C8H16O2	34.2	32.76689424	N/A	88.12779085	N/A	98.35823618
	Methyl octanoate	111-11-5	C9H18O2	N/A	36.93403167	N/A	81.09446911	N/A	94.23497525
	Methyl nonanoate	1731-84-6	C10H20O2	N/A	42.99989239	N/A	74.95562214	N/A	91.01511649
	Methyl decanoate	110-42-9	C11H22O2	47.9	48.84417181	N/A	69.72134462	N/A	88.85634686
	Methyl undecanoate	1731-86-8	C12H24O2	N/A	59.63656042	N/A	65.34623513	N/A	87.56586154
	Methyl laurate	111-82-0	C13H26O2	61.2	64.12724379	N/A	61.74985414	N/A	86.87033775
	Methyl palmitate	112-39-0	C17H34O2	85.9	83.90988485	N/A	53.24419747	N/A	86.35316755
	Methyl stearate	112-61-8	C19H38O2	86.9	85.50863183	N/A	51.36357188	N/A	86.37935827
Unsaturated ester-2	Methyl acrylate	96-33-3	C4H6O2	N/A	7.451941853	N/A	119.541674	N/A	93.60812733
	Methyl but-2-enoate	N/A	C5H8O2	N/A	6.547661437	N/A	120.0847422	N/A	93.87104261
	2-Pentenoic acid methyl ester	N/A	C6H10O2	N/A	9.977274701	N/A	114.8447465	N/A	96.56159128
	Methyl 2-hexenoate	13894-63-8	C7H12O2	N/A	13.62774544	N/A	109.2792792	N/A	98.85318241
	Methyl 2-heptenoate	38693-91-3	C8H14O2	N/A	17.63048788	N/A	103.1724211	N/A	99.6628692
	Methyl trans-2-octenoate	7367-81-9	C9H16O2	N/A	22.07129927	N/A	96.427612	N/A	98.42163841
	Methyl 2-nonenoate	111-79-5	C10H18O2	N/A	26.99665496	N/A	89.51556936	N/A	95.78147928
	Methyl (E)-2-decenoate	2482-39-5	C11H20O2	N/A	32.48884001	N/A	82.97180469	N/A	92.87887352
	Methyl-2-undecenoate	22104-71-8	C12H22O2	N/A	38.67122643	N/A	77.07701433	N/A	90.45134162
	Methyl (E)-2-dodecenoate	N/A	C13H24O2	N/A	46.59097091	N/A	71.93732974	N/A	88.72413925
Unsaturated ester-5	Methyl 5-hexenoate	2396-80-7	C7H12O2	N/A	18.91313174	N/A	102.4863373	N/A	99.01555011
	Methyl(E)-5-heptenoate	54004-28-3	C8H14O2	N/A	17.63048788	N/A	103.1724211	N/A	99.6628692
	Methyl (Z)-5-octenoate	41654-15-3	C9H16O2	N/A	19.96845447	N/A	99.20752041	N/A	99.04192494
	Methyl non-5-enoate	N/A	C10H18O2	N/A	22.4689945	N/A	95.06759219	N/A	97.61074494
	Methyl 5-decenoate	79837-87-9	C11H20O2	N/A	25.14293194	N/A	90.9463543	N/A	95.72250187
	Methyl-5-undecenoate	N/A	C12H22O2	N/A	29.98755308	N/A	84.45378128	N/A	93.42007995
	Methyl (Z)-5-dodecenoate	N/A	C13H24O2	N/A	35.50509499	N/A	78.41689467	N/A	91.11998166
Unsaturated ester-9	Methyl oleate	112-62-9	C19H36O2	56	57.20057352	N/A	60.6025195	N/A	86.65434945
Unsaturated ester-9, 12	Methyl linoleate	112-63-0	C19H34O2	41.7	42.4973983	N/A	66.88578885	N/A	86.99091671
Unsaturated ester-9, 12, 15	Methyl linolenate	301-00-8	C19H32O2	45.9	40.44218212	N/A	74.5217454	N/A	87.09859343

3.3.3 Method application to fuel mixtures

The QSPR method decomposes the fuel molecule into component fragments following the QSPR-UOB 2.0 system and it operates at the atomic level rather than the molecular level. As a result, the ML-QSPR model can apply to pure compounds as well as fuel mixtures. The fuel ignition quality database contains 120, 156, 156 mixture samples of CN, RON, MON as shown in Table 3.5 which are used to train the ML regression model to handle the interaction of functional groups. The octane sensitivity (OS) is the difference between RON and MON. Practical gasoline with higher OS has a better anti-knock property for modern spark-ignition engine [253]. The measured, predicted and residuals of CN, RON, MON for TPRF (n-heptane-iso-octane-toluene) mixtures are compared in Figure 3.16 and the ML-QSPR models enable precise forecast. The residuals of CN, RON, MON, OS are within -1.22~3.46, -5.42~2.68, -4.7~7.7 (see Figure 3.16) and the R^2 are 0.9933, 0.9984, 0.991, 0.8849 (see Figure 3.17) respectively. The predicted CN, RON, MON, OS of TPRF mixtures are validated against 30, 87, 87, 87 samples and their RMSE is 0.732, 0.655, 1.368, 1.139 as shown in Table 3.11. The ML-QSPR model does not limit to the TPRF mixture but can also apply to other fuel mixtures. This work takes n-heptane-dibutyl ether-ethanol mixtures as an example, the predicted CN, RON, MON are plotted in Figure 3.18, but there is no ignition quality data available for these new mixtures. This evidence supports that the ML-QSPR model can enrich the fuel ignition quality dataset and provide a priori and reasonable prediction for fuel screening.
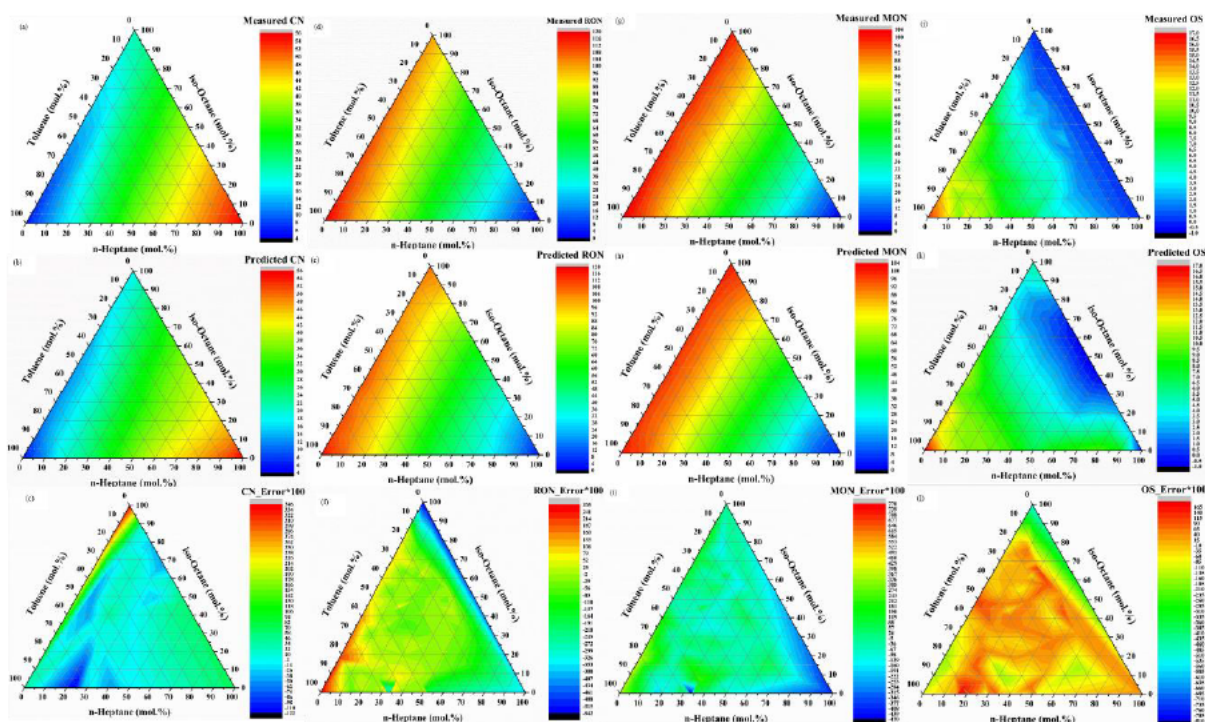
Figure 3.16. Comparison between measured, predicted values and errors of (a)~(c) CN, (d)~(f) RON, (g)~(i) MON, (j)~(l) OS of TPRF mixtures.
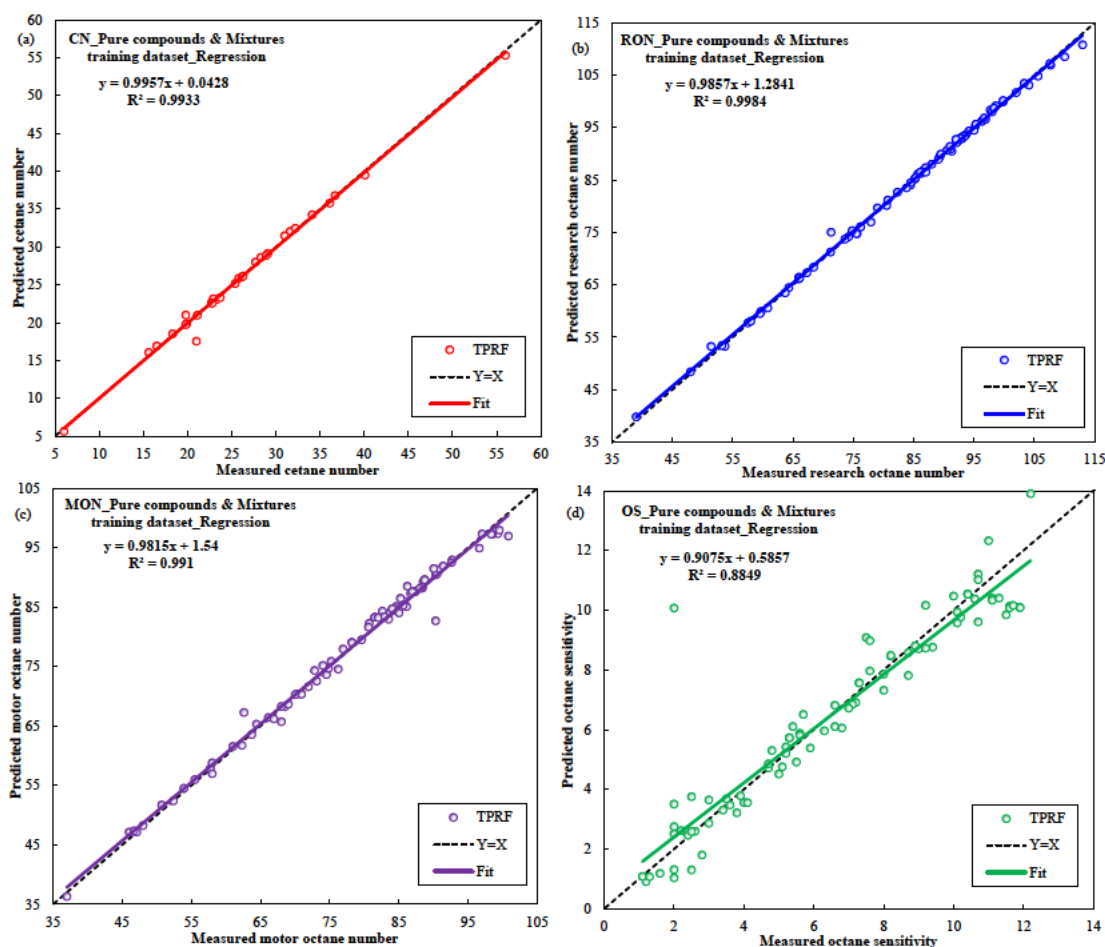
Figure 3.17. Parity plots for (a) CN, (b) RON, (c) MON, (d) OS of TPRF mixtures between measured and predictive values by machine learning regression model.
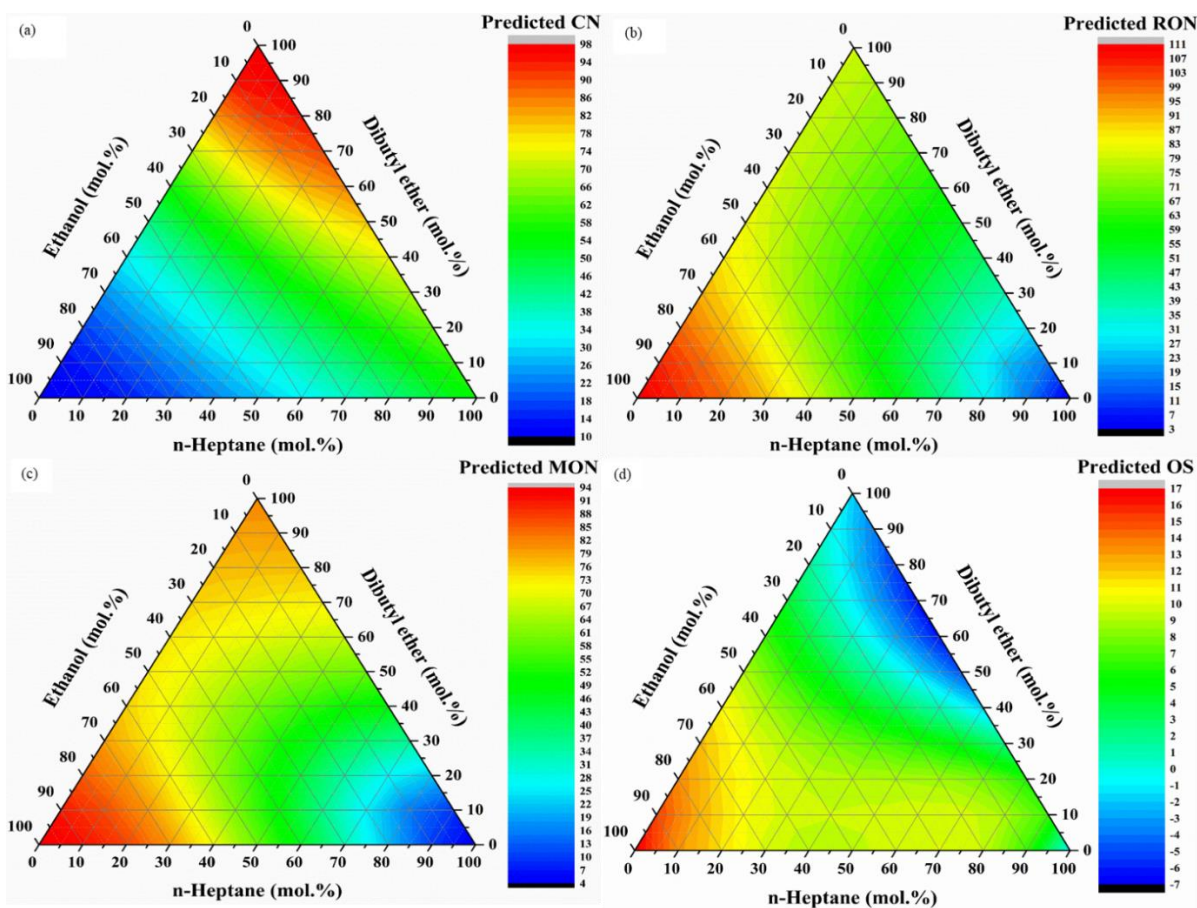
Figure 3.18. Predicted (a) CN, (b) RON, (c) MON, (d) OS of n-heptane-dibutyl ether-ethanol mixtures by machine learning regression model.

Table 3.11. Predictive performance of TPRF mixtures for ML-QSPR models

Property	Fuel mixtures (No. of measured compounds)	R^2	MAE	RMSE
CN	TPRF (30)	0.9933	0.395	0.732
RON	TPRF (87)	0.9984	0.4	0.655
MON	TPRF (87)	0.991	0.869	1.368
OS	TPRF (87)	0.8849	0.661	1.139

3.4 Conclusions

It is difficult to determine the CN, RON, MON of pure compounds simultaneously because the reactive fuels are well suited to CN and those with a high anti-knock property are described by RON/MON in CFR engine test. In addition, few methods can be applied to predict the ignition quality of fuel mixtures and their scopes are limited to typical fuel types. This work proposes a machine learning quantitative structure-property relationship (ML-QSPR) method for fuel ignition quality (CN/RON/MON) prediction of pure compounds and mixtures. It applies to 13 fuel types of alkanes, alkenes, alkynes, naphthenes, aromatics, alcohols, aldehydes, ketones, saturated esters, unsaturated esters, ethers, carboxylic acids, polyfunctionals, fuel mixtures. QSPR-UOB 2.0 functional group classification system is proposed to extract the structural features and transform them into a fuel molecular structure matrix. A fuel ignition quality database is developed to store the 603, 375, 371 samples of CN, RON, MON data and provides the ignition quality matrix. 19 ML algorithms are utilized to map the fuel molecular structure matrix and ignition quality matrix and develop the regression model. The ML-QSPR regression models provide high predictive fidelity and their model performance is better than the published models. The R^2 & RMSE of CN, RON, MON for ML-QSPR models reach 0.9911 & 2.526, 0.9874 & 2.454, 0.9731 & 2.765 respectively while the R^2 of published GCM models are 0.90, 0.93, 0.91. This method takes advantage of QSPR theory which decomposes the fuel molecules into typical component fragments and transforms them into chemical representation. These atomic-level features rather than molecular level are adopted for regression model development, so the proposed ML-QSPR model enables to predict the ignition quality of fuel mixtures. High validation accuracy of CN, RON, MON for TPRF mixtures are obtained and their R^2 reach 0.9933, 0.9984, 0.991. Most importantly, the ML-QSPR models can enrich the sample set in the fuel ignition quality database for fuel screening and enable understanding the impact of the molecular structure on the fuel ignition quality. This tool benefits fuel formulation to adjust the fuel reactivity adapting to the combustion mode requirement.

Chapter 4 Machine Learning-Quantitative Structure Property Relationship (ML-QSPR) Method for Fuel Physicochemical Properties Prediction of Multiple Fuel Types

4.1 Introduction

Combustion modes optimization and fuel components design are the two major approaches to realize efficient and clean combustion in IC engines. In the category of combustion modes optimization, the major solution is low temperature combustion (LTC) [254]. It exploits high dilution ratio and charge density to reduce peak combustion temperature and increase the absolute oxygen content which benefits NOx, PM and specific fuel consumption reduction. The represented LTC combustion modes include HCCI [255, 256], PPC [257], RCCI [258] etc. In the category of fuel component design, the solution is utilizing fuel physicochemical properties [259] and combustion chemistry [76, 89, 260] of typical fuel types to control the combustion rate and timing. The impacts of molecular structure on the ignition and oxidation behavior of gasoline [115], diesel [113, 114, 261], kerosene [118] and biofuels [262, 263] are explored by the chemical kinetics. Relative independence between the combustion modes optimization and fuel component design is observed and they lack an intrinsic interaction until the concept of "property-oriented fuel design" comes up.

The property-oriented fuel design concept is applied to the Cluster of Excellence "Tailor-Made Fuels from Biomass" (TMFB) [261, 264, 265] and the U.S. Department of "Co-Optimization of Fuels & Engines" [266, 267] for the first time. At the first step, use engine combustion (droplet spray and atomization, charge preparation, fuel-air mixing, combustion timing/duration, HRR) and emission requirement (NOx/HC/CO/PM) as design constraints to determine the combustion mode and required fuel physicochemical properties. In the second step, fuel screening is performed to identify the fuel molecules with desired physicochemical properties [268, 269]. The fuel screening usually involves 5 categories of properties: (1) volatility; (2) atomization; (3) energy density; (4) sooting propensity; (5) ignitability. The volatility and atomization are described by the T_m, T_b, ΔH_{vap}, FP, VP and the γ respectively. The LHV, ρ and YSI are the indicators of energy density and sooting tendency. The CN, RON, MON, LFL,

UFL evaluate the fuel ignition quality. But the experimental data of these properties are limited and a part of them are not in interest, so fuel physicochemical property prediction models are needed to accelerate the fuel screening process and the models are summarized in Table 4.1. The predictive models for CN/RON/MON have been summarized in Table 3.3 and do not repeat here. Group contribution method (GCM) and quantitative structure-property relationship (QSPR) are the most commonly used methods for molecular structure features extraction and the regression methods are diverse including correlation equations, convolution neural network (CNN), machine learning (ML) algorithms, commercial software, etc. There are 4 knowledge gaps of these fuel properties predictive models: (1) None of the published GCM/QSPR system can apply to all 15 properties and diverse fuel types. A generic molecular structure features extraction method that can apply to multiple fuel properties and compound groups is needed. (2) Some GCM/QSPR methods contain a large set of molecular descriptors which is complex for feature extraction and reducing the model interpretability. A simple and efficient feature extraction method is required for high fidelity prediction. (3) The scopes of published property predictive models are limited to typical fuel types and they should be expanded to a wider range of chemical classes. (4) Some models obtain modest predictive performance and their R^2 are below 0.9. The model forecasting capability can be further improved by optimizing feature extraction systems and regression models. (5) The model training dataset should be expanded to improve the model interpolation and extrapolation capacity.

This work develops the machine learning quantitative structure-property structure (ML-QSPR) models to predict 15 physicochemical properties for 24 fuel types. A novel QSPR-UOB 3.0 functional classification system is proposed for molecular structure feature extraction which contains 42 component fragments. ML algorithm is exploited to correlate the molecular structure features and fuel properties. A UOB Fuel Property database containing 1797 pure compounds, 465 mixtures is established for model training and validation. 10-fold cross validation is implemented to protect against over-fitting. ML-QSPR models enable property-oriented fuel screening despite limited experimental data and provide insights into the impact of molecular structure on the fuel properties.

Table 4.1. Prediction of 13 fuel physicochemical properties by computer-aided molecular design (CAMD) methodology

Object	Feature extraction method	Regression model	Model inputs	R^2*	Dataset	Scope	Ref.
T_m	QSPR	Correlation equations	639 molecular descriptors	0.8373	443	Substituted benzenes	[270]
	QSPR	Graph-based convolutional NN	molecular tensor	N/A	3041	Aliphatic compounds	[271]
	QSPR	Extreme learning machine	145 2D descriptors	0.63	4173	Organic compounds	[272]
	QSPR	Nonlinear NN	40 molecular descriptors	0.99	1250	Organic compounds	[273]
	QSPR	Machine learning	23 molecular descriptors, 27 functional group count descriptors	0.851	1097	hydrocarbons, alcohols and esters	[274]
	GCM	position distribution function	95 molecular descriptors	N/A	730	Covalent compounds	[275]
	GCM	Enthalpy/entropy transition	43 molecular descriptors	N/A	596	Aliphatic, non-hydrogen-bonding compounds	[276]
	GCM	Enthalpy/ entropy transition	46 molecular descriptors	N/A	1040	Aliphatic compounds	[277]
	MDM	Potential energy function	Polymer consistent force-field	N/A	8	n-Alkanes	[278]
T_b	QSPR	Boosting regression tree	330 2D and 432 3D descriptors	0.9565	2475	Hydrocarbons, oxygenates	[279]
	QSPR	MLR/MLP-ANN	1666 molecular descriptors	0.9999	223	pure hydrocarbons	[280]
	QSPR	Back-propagation NN	8 Molecular descriptors	0.9999	327	Alkanes, alkenes, alkynes	[281]
	QSPR	MLR	2 electro-negativity descriptors	0.9993	215	Alkanes, unsaturated hydrocarbons, alcohols	[282]
	QSPR	7-parameter equation	~800 molecular descriptors	0.9517	612	C-H-O-N-S-F-Cl-Br-I compounds	[283]
	QSPR	4-parameter equation	~600 molecular descriptors	0.9598	298	Organic compounds	[284]
	QSPR	RBN	1666 molecular descriptors	0.989	240	Acyclic oxygen organic compounds	[285]
	QSPR	RBN	1666 molecular descriptors	0.99	432	C-H-O compounds	[286]
	QSPR	MLR	119 Topological indices	0.9939	119	Aliphatic esters	[287]
	GCM	Correlation equation	293 molecular descriptors	0.9836	1141	Organic compounds	[288]
ΔH_{vap}	QSPR	Correlation equation	22 norm descriptors	0.967	573	Hydrocarbons, oxygenates	[289]
	QSPR	Correlation equation	3 norm indexes	0.9503	480	Hydrocarbons, oxygenates	[290]
	QSPR	GA-MLR	~3000 molecular descriptors	0.9814	4879	Pure chemical compounds	[291]

	QSPR	ANN	11 input properties	0.9998	281	Saturated/unsaturated hydrocarbons	[292]
	GCM	ANN	147 functional groups	0.993	4907	pure chemical compounds	[293]
	GCM,	4-parameter equation	125 molecular descriptors	N/A	831	organic compounds	[294]
	GCM	3-constant equations	39 molecular descriptors	N/A	509	C-H-O-N-F-Cl-S compounds	[295]
γ	QSPR	ANN	COSMO-RS sigma moments	0.963	1275	organic compounds	[296]
	QSPR	6-parameter correlation equation	6 molecular descriptors	0.96	320	C-H-O-N-F-Cl-Br-S compounds	[297]
	QSPR	Macleod-Sugden-Quayle method	5 types of molecular descriptors	N/A	649	C-H-O-N-F-Cl-Br-S compounds	[298]
	QSPR	MLR	2 atom-type topological indices	0.9829	92	Saturated/unsaturated compounds	[299]
	QSPR	SVM	Constitutional/topological/geometrical/electrostatic/quantum chemical descriptors	0.9348	196	hydrocarbons, halogenated aliphatics, oxygenates	[300]
	QSPR	ANN	5 physicochemical parameters	0.9997	210	C1~C20 Pure Hydrocarbons	[301]
	GCM	Esmaeilzadeh-Roshanfekr equation of state	pressure, temperature, molar volume	N/A	N/A	Alkane, alkene, cycloalkane, aromatic	[302]
	GCM	ANN	151 molecular descriptors	0.997	4672	C-H-O-N-Br-I-F-Cl-S compounds	[303]
ν	QSPR	ANN & SVM	26 molecular descriptors	0.993	407	DIPPR database	[304]
	QSPR	CODESSA software	over 600 molecular descriptors	0.854	361	C-H-O-N-S compounds	[305]
	QSPR	CODESSA software	579 molecular descriptors	0.8464	337	C-H-O-N-S compounds	[306]
	QSPR	GA-MLR	over 3000 molecular descriptors	0.9697	2748	C-H-O-N-Br-I-F-Cl-S compounds	[307]
	QSPR	ANN	36 chemical features	0.9938/6	261	n-paraffins, isoparaffins, olefins, alkynes, cycloalkanes, and aromatics	[308]
	QSPR	Random Forest	116 molecular descriptors	0.9	403	C-H-O-N-S compounds	[309]
	QSPR	ANN	11 input properties	0.99986	281	16 family classes DIPPR database	[292]
LHV	QSPR	MLR	1481 molecular descriptors	0.996	1650	organic compounds	[310]
	QSPR	GA–MLR	1664 molecular descriptors	0.9954	1714	Pure compounds in DIPPR 801 database	[311]
	QSPR	ML	23 molecular descriptors, 27 functional group count descriptors	0.999	2767	hydrocarbons, alcohols and esters	[274]
	GCM	ANN	47 structural groups	0.999	586	Alkanes, alkenes, aromatics, cyclic compounds	[312]

	Method	Approach	Descriptors	R^2	N	System	Ref
	GCM	Robust regression	1st, 2nd, 3rd order functional groups	0.99	794	Pure substances in DIPPR 801 Database	[313]
	GCM	Least square method	32 atom-type structural groups	0.9982	452	pure hydrocarbons	[314]
	GCM	ANN	142 chemical groups	0.99999	4590	Compounds in Yaws' Handbook	[315]
ρ	QSPR	MLR or MLP-ANN	20 molecular descriptors	0.9993	222	pure hydrocarbons	[280]
	QSPR	ANN or SVM	22 molecular descriptors, 26 functional group descriptors	0.997	730	Hydrocarbons, oxygenates	[304]
	QSPR	ANN	11 input properties	0.99981	281	Saturated/unsaturated hydrocarbons	[292]
	QSPR	2-parameter correlation	constitutional, topologic, geometric, electrostatic descriptors	0.9749	303	C-H-O-N-S-F-Cl-Br-I compounds	[316]
	GCM	GCVOL density equation	60 molecular descriptors	N/A	1040	Hydrocarbons, oxygenates	[317]
	GCM	equation-of-state	1st, 2nd order functional groups	0.983	334	Hydrocarbons, oxygenates	[318]
YSI	QSPR	MLP	5270 molecular descriptors	N/A	297	Oxygenated bioblendstocks	[319]
	QSPR	ANN	25 molecular descriptors	0.978	421	Hydrocarbons, oxygenates	[320]
	GCM	Bayesian linear regression	66 fragment types.	N/A	441	oxygenates, alkanes, alkenes, cycloalkanes, aromatics	[321]
	GCM	kernel ridge regression	33 molecular descriptors	0.9858	204	Esters, ketones, aldehydes, ethers, alcohols	[322]
	MD	2 parameters equation	ReaxFF Molecular Dynamics simulations	N/A	4	aromatic	[323]
IT	QSPR	GA-MLR or FFNN	3224 molecular descriptors	0.8317	813	69 different chemical families.	[324]
	QSPR	GA-PLS or SVM	605 descriptors	0.901	446	C-H-O-N-S-F-Cl-Br-I compounds	[325]
	QSPR	MLR or BPNN or SVM	6 atom types descriptors	0.9274	142	hydrocarbons, halogenated aliphatics, aromatics, alcohols, ethers, esters, ketones, amine	[326]
	QSPR	ANN or MLR	16 atom-type electrotopological-state indices	0.9063	118	alkanes, olefins, alkynes, aromatics	[327]
	GCM	3 layers ANN	46 functional groups descriptors	0.984	1025	78 different chemical families	[328]
	GCM	Robust regression	1st, 2nd, 3rd order functional groups	0.76	513	Alkanes, aldehydes, alcohols, acids	[329]
	GCM	ANN + PSO	42 molecular descriptors	0.9899	343	C-H-O-N-Cl-Br compounds	[330]
	Empirical model	Correlation equation	size and branches of different classes of hydrocarbons	0.955	274	alkanes, alkenes, cycloalkanes, cycloalkenes, alkynes, and aromatics	[331]
FP	QSPR	ANN	geometrical, topological, quantum mechanical, electronic descriptors	0.978	758	organic compounds	[332]
	QSPR	3 parameter equation	Boiling point, relative negative charge, H-donors charged surface	0.902	271	C-H-O-N-S-F-Cl-Br-I	[333]

| | | | | | | area | |
						compounds	
	QSPR	Radial basis function	26 molecular descriptors	0.9879	400	C-H-O-N-S-F-Cl-Br-I compounds	[334]
	QSPR	FFNN	1926 molecular descriptors	N/A	87	pure compounds in DIPPR 801 database	[335]
	QSPR	PSO-SVM	22 categories of molecular descriptors	0.885	1651	pure compounds in DIPPR 801 database	[336]
	GCM	ANN	42 molecular descriptors	0.9933	740	pure compounds in DIPPR 801 database	[337]
	GCM	GA-MLR	20 types of molecular descriptors	0.9487	1294	pure compounds in DIPPR 801 database	[338]
	GCM	Robust regression	1st, 2nd, 3rd order functional groups	0.99	927	Alkanes, aldehydes, alcohols, acids	[329]
	GCM	Correlation equation	66 molecular descriptors	N/A	1062	C-H-O-N-S-X-Si compounds	[339]
	Empirical model	Correlation equation	distillation temperatures, density	0.87	N/A	pure hydrocarbons and undefined petroleum fractions	[340]
VP	QSPR	3-layer FFNN	29 molecular descriptors and temperature	0.99	~1500	81 chemical families in DIPPR 801 database	[341]
	QSPR	BPNN	4 valance molecular connectivity indices, molecular weight, temperature	0.9967	274	Alkanes, alkenes, alkynes, aromatics, cyclic compounds	[342]
	QSPR	quantum mechanics/neural net	5 descriptors	0.74	1085	Organic compounds	[343]
	QSPR	MLR	17 molecular structures	0.937	645	Organic compounds	[344]
	GCM	nonlinear regression	24 molecular descriptors	N/A	62	Alkenes, aromatics	[345]
LFL	QSPR	SVM	578 molecular descriptors	0.979	1038	C-H-O-N-S-F-Cl-Br-I compounds	[346]
	QSPR	MLR-ANFIS	22 categories descriptors	0.93	1615	2 diverse chemical material classes	[347]
	QSPR	MLR	4 descriptors	0.9137	458	pure compounds in DIPPR 801 database	[348]
	QSPR	GA-MLR	29 types and 4885 kinds of molecular descriptors	0.964	181	binary hydrocarbon gases	[349]
	GCM	3-layer FFNN	105 functional groups	0.986	1057	pure compounds in DIPPR 801 database	[350]
	GCM	Robust regression	1st, 2nd, 3rd order functional groups	0.99	443	Alkanes, aldehydes, alcohols, acids	[329]
	GCM	ANN + PSO	42 molecular descriptors	0.9865	418	C-H-O-N-S compounds	[351]
	GCM	3-layer ANN	30 molecular descriptors	0.9996	543	pure compounds in DIPPR 801 database	[352]
UFL	QSPR	GA-MLR	1664 molecular descriptors	0.92	865	pure compounds in DIPPR 801	[353]

					database	
QSPR	GA-MLR	1664 molecular descriptors	0.898	278	pure compounds in DIPPR 801 database	[354]
QSPR	GA-MLR	6 types of molecular descriptors	0.758	588	C-H-O-N-S-F-Cl-Br-I compounds	[355]
GCM	Robust regression	1st, 2nd, 3rd order functional groups	0.91	351	Alkanes, aldehydes, alcohols, acids	[329]
GCM	ANN-PSO	42 molecular descriptors	0.9818	418	C-H-O-N-S compounds	[351]
GCM	3-layer FFNN	113 functional groups	0.9469	867	pure compounds in DIPPR 801 database	[356]

* Maximal R^2 is selected.

4.2 Modeling approach

4.2.1 Methodological overview

The ML-QSPR models of 15 physicochemical properties enable the Tier 1 fuel physicochemical property screening as shown in Figure 4.1. The property-oriented fuel design contains Tier 1 fuel physicochemical property screening and Tier 2 chemical kinetics screening which will be discussed in Chapter 6. The studied properties include melting point T_m, boiling point T_b, vapor pressure (VP), enthalpy of vaporization ΔH_{vap}, cetane number (CN), research octane number (RON), motor octane number (MON), ignition temperature (IT), flash point (FP), yield sooting index (YSI), liquid density ρ, lower heating value (LHV), surface tension γ, lower flammability limit (LFL), upper flammability limit (UFL). A UOB Fuel Property Database containing 1797 pure compounds and 465 mixtures is established for model training and validation. A novel functional group classification system QSPR-UOB 3.0 (see section 4.2.3) is proposed to extract the structural features and transform them into a fuel molecular structure matrix according to the occurrence of typical component fragments. Machine learning (ML) algorithm is adopted to map the fuel molecular structure matrix and fuel property matrix. 10-fold cross validation is applied to examine the model performance on new data prediction and avoid over-fitting.
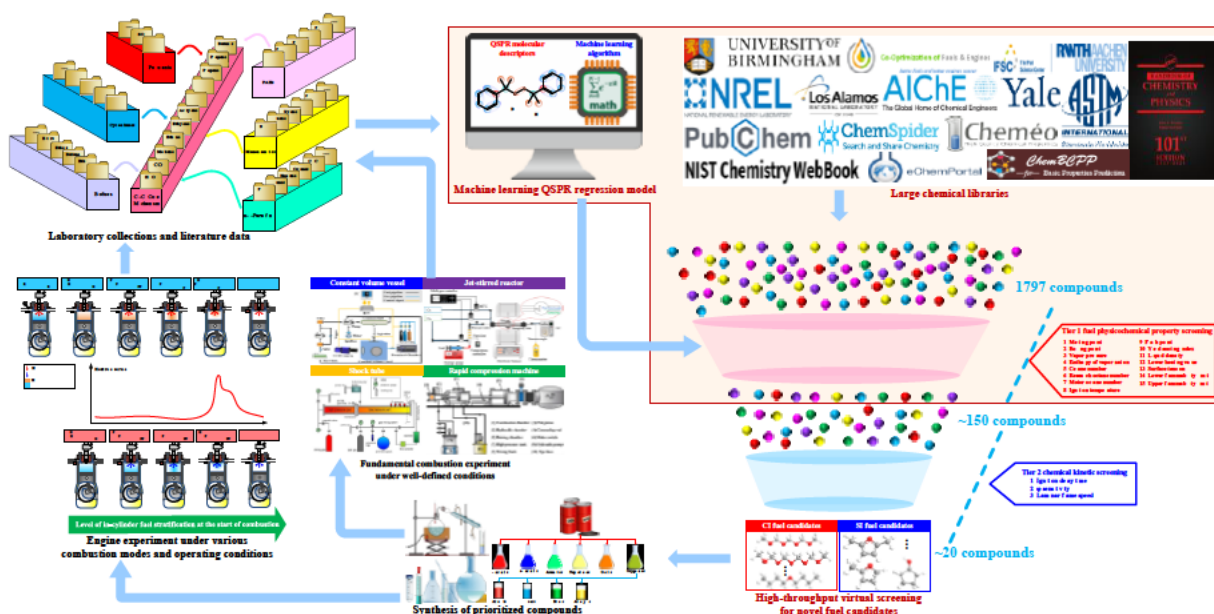
Figure 4.1. The workflow of virtual fuel screening by ML-QSPR and chemical kinetics.

4.2.2 Fuel physicochemical properties database development

The UOB Fuel Property Database covering 15 properties is established for model training and validation. The total number of compounds for CN, RON; MON, T_m, T_b. ΔH_{vap}, γ, LHV, ρ, YSI, IT, FP, VP, LFL, UFL and the numbers of different compound groups are shown in Table 4.2. It covers 24 fuel types of n-alkanes, iso-alkanes, cycloalkanes, alkenes, cyclic alkenes, alkadienes, alkynes, alcohols, cycloalcohols, aldehydes, ketones, cyclic ketone, saturated esters, unsaturated esters, acyclic ethers, furans, other cyclic ethers, aromatics, carbonate ester, carboxylic anhydride, peroxide, hydroperoxide, polyfunctionals, carboxylic acids as shown in Table 4.2. Particularly, the term "polyfunctionals" refers to component with more than one designated functional groups (aromatic bond, carbon-carbon double bond, carbon-carbon triple bond, hydroxyl group, carbonyl group, aldehyde group, ether group, ester group, carboxylic acid group, carbonate ester group, carboxylic anhydride group, hydroperoxide group, peroxide group). The model training dataset is measured values and the data source is summarized in Table 4.3.

Table 4.2. Number of compounds of different chemical classes in the model training dataset

Compound class	Number of compounds (measured data)														
	CN	RON	MON	T_m	T_b	ΔH_{vap}	γ	LHV	ρ	YSI	IT	FP	VP	LFL	UFL
Alkanes	75	46	46	93	111	61	88	102	106	30	48	98	31	94	56
Cycloalkanes	50	38	34	75	79	26	89	101	79	20	36	47	13	30	24
Alkenes	25	70	72	85	106	49	97	107	105	36	32	73	29	61	62
Cycloalkenes	5	1	N/A	16	20	1	9	19	19	12	5	17	5	14	12
Alkadienes	4	N/A	N/A	22	28	2	20	32	28	1	6	28	3	23	13
Alkynes	1	4	2	19	24	5	10	18	23	3	2	15	5	11	N/A
Aromatics	49	35	37	156	167	31	120	170	165	122	47	142	27	135	106
Alcohols	38	14	13	107	137	71	16	124	130	49	47	133	38	109	86
Cycloalcohols	2	N/A	N/A	8	9	2	N/A	8	7	1	7	9	2	2	1
Aldehydes	8	N/A	N/A	26	40	2	6	37	38	18	14	41	14	35	25
Ketones	10	2	1	38	47	22	6	31	50	28	13	42	15	30	17
Cycloketones	4	1	2	8	9	2	N/A	7	10	2	2	7	3	7	2
Saturated esters	87	1	1	124	151	41	10	93	147	41	36	116	38	88	30
Unsaturated esters	34	N/A	N/A	28	44	N/A	1	28	40	20	9	33	9	24	16
Acyclic ethers	35	3	4	50	69	31	4	51	70	34	20	60	23	48	31
Furans	14	3	2	7	13	3	1	6	13	3	1	12	5	7	3
Other cyclic ethers	7	N/A	N/A	13	17	6	N/A	12	14	4	6	13	6	9	2
Carbonate ester	N/A	N/A	N/A	4	5	1	N/A	5	5	3	6	5	3	3	1
Carboxylic anhydride	N/A	N/A	N/A	9	9	N/A	N/A	8	7	N/A	N/A	9	3	6	3
Peroxide	N/A	N/A	N/A	1	1	N/A	N/A	1	N/A	N/A	N/A	1	N/A	1	N/A
Hydroperoxide	N/A	N/A	N/A	1	3	N/A	N/A	3	3	N/A	N/A	3	N/A	2	N/A
Polyfunctionals	52	2	2	57	97	8	3	70	87	14	23	102	24	69	56
Carboxylic acids	5	N/A	N/A	56	52	7	8	62	41	4	18	59	11	49	33
Mixture	130	167	167	N/A	N/A	N/A	N/A	N/A	N/A	12	N/A	N/A	N/A	N/A	N/A
Total oxygenates	296	26	25	537	703	196	55	546	662	221	202	645	194	486	306
Total	635	387	383	1003	1238	371	488	1095	1187	457	378	1065	307	857	579

Table 4.3. Data sources of 15 fuel physicochemical properties in the model training dataset

Property	Fuel type	Institute	Ref.
CN	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, multi-oxygen compounds, mixtures	University of Birmingham	[357]
RON	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, multi-oxygen compounds, mixtures	University of Birmingham	[357]
MON	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, multi-oxygen compounds, mixtures	University of Birmingham	[357]
T_m	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, peroxides, hydroperoxides, multi-oxygen compounds	NIST	[358]
T_b	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, peroxides, hydroperoxides, multi-oxygen compounds	NIST	[358]
ΔH_{vap}	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, multi-oxygen compounds	NIST	[358]
γ	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, alcohols, aldehydes/ketones, ethers, esters,	Kuwait University	[359]
LHV	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, peroxides, hydroperoxide, multi-oxygen compounds	Kuwait University	[312]
LHV	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes, ethers, esters, furans	Nanjing University of Technology	[310]
LHV	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, peroxides, hydroperoxide, multi-oxygen compounds	DIPPR 801 database	[360]
ρ	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, hydroperoxides, multi-oxygen compounds	NIST	[358]
YSI	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, furans, multi-oxygen compounds, mixtures	Yale University	[321, 361]
IT	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, multi-oxygen compounds	NIST	[358]
FP	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, peroxides, hydroperoxides, multi-oxygen compounds	NIST	[358]

VP	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, multi-oxygen compounds	NIST	[358]
LFL	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, peroxides, hydroperoxides, multi-oxygen compounds	University of Tehran	[350]
LFL	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, peroxides, hydroperoxides, multi-oxygen compounds	National Kaohsiung First University of Science and Technology	[348]
LFL	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, peroxides, hydroperoxide, multi-oxygen compounds	DIPPR 801 database	[360]
UFL	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, multi-oxygen compounds	University of Tehran	[353]
UFL	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, multi-oxygen compounds	University of Tehran	[356]
UFL	Alkanes, cycloalkanes, alkenes, cycloalkenes, alkynes/alkadienes, aromatics, alcohols, aldehydes/ketones, ethers, esters, acids, furans, carboxylic anhydride, multi-oxygen compounds	DIPPR 801 database	[360]

4.2.3 Features extraction of molecular structure by QSPR-UOB 3.0

QSPR method is applied to extract chemical structure features and transforms them into a fuel molecular structure matrix. A novel QSPR-UOB 3.0 functional group classification system is established to further improve the fuel molecule resolution compared to QSPR-UOB 2.0 and the major modifications are below: (1) Add component fragments of 1.10, 1.11, 1.12, 1.13, 1.14 which enables to describe aromatics with 3 fused benzene rings, 4 fused benzene rings, 5 fused benzene rings accompanying substituents at position 10, 11, 12, 13, 14. For example, it can recognize the 9,10-dimethylanthracene (CAS:781-43-1, $C_{16}H_{14}$), 11-methyltetracene (CAS:6111-78-0, $C_{19}H_{14}$), 12-methyltetracene (CAS:2422-79-9, $C_{19}H_{14}$), 13-methylpentacene ($C_{23}H_{16}$), 14-methylpentacene ($C_{23}H_{16}$) and convert structural features into fuel molecular structure matrix as shown in Figure 4.2. (2) Complement ester group in ring structure (fragment component 22 in Figure 4.2) to distinguish ester group in ring structure or non-ring structure such as 2(3H)-furanone, dihydro-5-methyl- (CAS:108-29-2, $C_5H_8O_2$). (3) Add fragment components of carbonate ester (functional group 24), carboxylic anhydride (functional group 25), hydroperoxide (functional group 26) and peroxide (functional group 27) to identify corresponding fuel types such as dimethyl carbonate (CAS:616-38-6, $C_3H_6O_3$), acetic anhydride (CAS:108-24-7, $C_4H_6O_3$), tert-butyl hydroperoxide (CAS:75-91-2, $C_4H_{10}O_2$), dicumyl peroxide (CAS:80-43-3, $C_{18}H_{22}O_2$). The functional groups are divided into 3 categories of functional group identifier, functional group position descriptor and fuel reactivity descriptor as shown in Figure 4.2.

The fuel molecular structure matrix has 42 columns and the element in each column corresponding to the occurrence of the 42 functional groups in QSPR-UOB 3.0 system. For example, 2-(2-hexoxyethoxy)ethanol (CAS: 112-59-4, $C_{10}H_{22}O_3$) has 1 methyl radical (functional group 8), 2 non-ring ether groups (functional group 16), 1 hydroxyl group (functional group 15), 5 maximum quantity non-ring methylene groups in series (functional group 9), 9 non-ring methylene groups in total (functional group 10). Thus, the the fuel molecular structure matrix of 2-(2-hexoxyethoxy)ethanol is express as [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 5 9 0 0 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0]. Each fuel molecule occupies a row in the matrix.

FUNCTIONAL GROUP CLASSIFICATION SYSTEM: QSPR-UOB 3.0

1.1. Aromatic bond 1-branched
1.2. Aromatic bond 2-branched
1.3. Aromatic bond 3-branched
1.4. Aromatic bond 4-branched
1.5. Aromatic bond 5-branched
1.6. Aromatic bond 6-branched
1.7. Aromatic bond 7-branched
1.8. Aromatic bond 8-branched
1.9. Aromatic bond 9-branched
1.10. Aromatic bond 10-branched
1.11. Aromatic bond 11-branched
1.12. Aromatic bond 12-branched
1.13. Aromatic bond 13-branched
1.14. Aromatic bond 14-branched
1.15. Sum of unbranched aromatic bond

1. Aromatic bond
2. Carbon–carbon double bond (CCDB) of aromatic, $CH_2=CH_2$
3. Carbon–carbon double bond (CCDB) of ring, $CH_2=CH_2$
4. Carbon–carbon double bond (CCDB) of non-aromatic, non-ring, $CH_2=CH_2$
5. Carbon–carbon triple bond (CCTB)
6. Tertiary carbon, >CH-
7. Quaternary carbon, >C<
8. Primary carbon (methyl radical), -CH_3
9. Maximal quantity of secondary carbon in series (non-ring) (methylene), >$(CH_2)_m$ (non-ring)
10. Secondary carbon (non-ring) (methylene), >CH_2 (non-ring)
11. Maximal quantity of secondary carbon in series (ring) (methylene), >$(CH_2)_m$ (ring)
12. Secondary carbon (ring) (methylene), >CH_2 (ring)
13. >CH-, non-Tertiary carbon
14. >C<, non-Quaternary carbon
15. Hydroxyl radical, -OH
16. Ether group (non-ring), –O–(non-ring)
17. Ether group (ring), –O–(ring)
18. Ketone group(non-ring), >C=O(non-ring)
19. Ketone group(ring), >C=O(ring)
20. Aldehyde group, -CH=O
21. Ester group, -C(=O)O-(non-ring)
22. Ester group, -C(=O)O-(ring)
23. Carboxylic acid, -C(=O)OH
24. Carbonate ester, R_1O(C=O)OR_2
25. Carboxylic anhydride, R_1-(O=)COC(=O)-R_2
26. Hydroperoxide, R-O-O-H
27. Peroxide, R_1-O-O-R_2

Figure 4.2. Conversion of molecules into fuel molecular structure matrix by QSPR-UOB 3.0 functional group classification system. The molecular descriptor computation of n-propyl acrylate in this figure is simply for display and illustrative purposes. The true fingerprint is [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 2 2 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0].
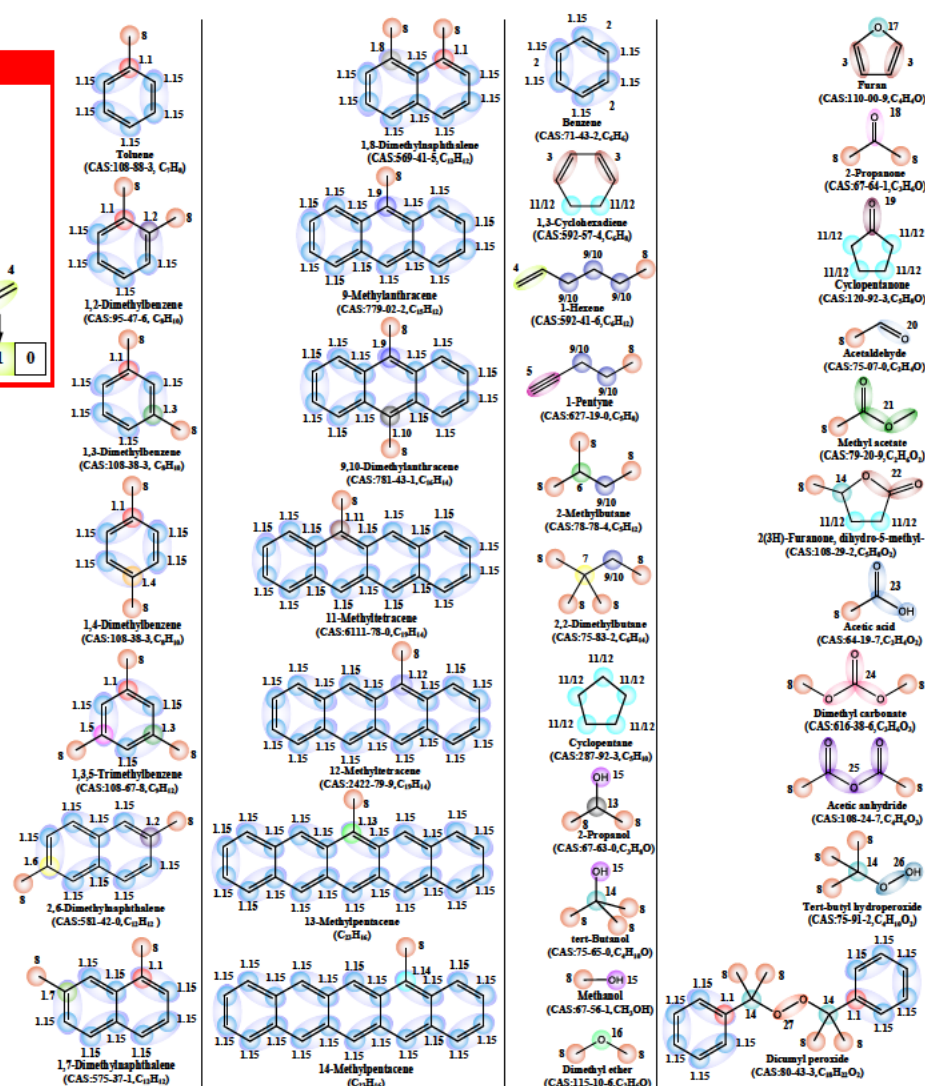
137

4.2.4 Training and validation of ML-QSPR models

The workflow of ML-QSPR method is shown in Figure 4.3. (1) Uses QSPR-UOB 3.0 system to manually extract the structural features from molecule images and convert them into a fuel molecular structure matrix. (2) Prepare the fuel property matrix based on the UOB Fuel Property Database. (3) Apply ML algorithms to map the fuel molecular structure matrix and fuel property matrix and perform 10-fold cross validation. By using the Regression Learner APP in MATLAB, 19 ML algorithms (as shown in Table 3.6) are used to train the ML-QSPR models in parallel and the one with minimal RMSE is chosen. The principles of these ML algorithms can be found in MATLAB Statistics and Machine Learning Toolbox [52]. 10-fold cross validation is used to examine the model capacity to predict new data and prevent over-fitting.
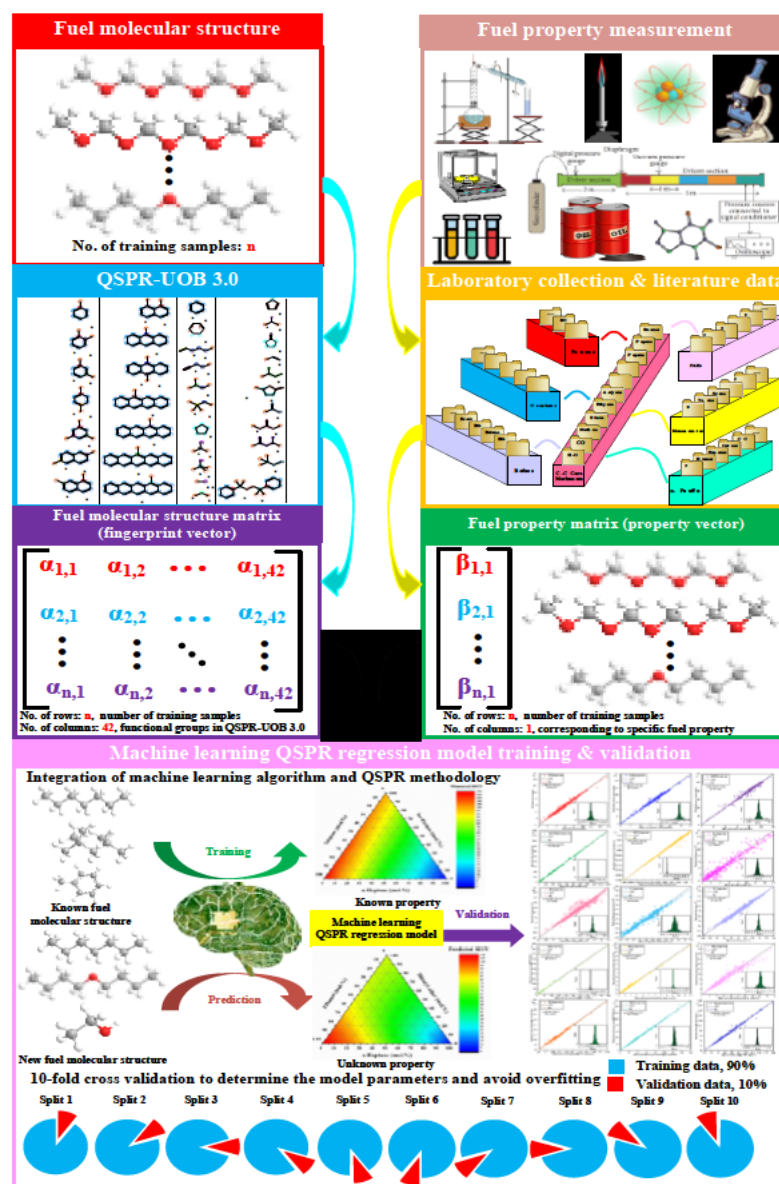
Figure 4.3. Integration ML algorithm and QSPR methodology to predict fuel physicochemical properties.

4.3 Results and discussion

4.3.1 General training settings and predictive accuracy of 15 fuel physicochemical properties

The optimal ML-QSPR models for 15 fuel physicochemical properties are determined by the RMSE and the model functions, parameters are demonstrated in Table 4.4. LHV is proportional to carbon atom number and it does not exhibit a strong non-linear relationship with fuel types, so the linear regression algorithm is well suited to LHV regression. Gaussian process regression algorithm is a nonparametric kernel-based probabilistic model which is suited for the other 14 properties with a strong non-linear relationship. The R^2 of CN, RON, MON, T_m, T_b, ΔH_{vap}, γ, LHV, ρ, YSI, IT, FP, VP, LFL, UFL are 0.9898, 0.9884, 0.9758, 0.9653, 0.9484, 0.9968, 0.9898, 0.9959, 0.9946, 0.9993, 0.9603, 0.9798, 0.9972, 0.9935, 0.9486 and they reach reasonable overall predictive accuracy as shown in Figure 4.4. The T_b and UFL need further improvement ($R^2 < 0.95$), especially the former is an important fuel volatility property for fuel screening task. The predictive residuals of different chemical classes for 15 properties are shown in Figure 4.5 which describe the deviation between predictive value and observed value. The box-and-whisker plots also provide information of the outliers and the uncertainty may come from measurement error and predictive error.

Combining Table 4.2, Table 4.5 and Figure 4.5 enable the users to evaluate the confidence of the predictive value. Even though the ML-QSPR models applied to 24 fuel types, the greater number of samples, the higher R^2 and the smaller predicted residuals of particular compound groups result in better predictive reliability of specific property and vice versa. For example, the R^2 of ketones, polyfunctionals for RON are 1 in Table 4.5 but they don't represent a good accuracy because of insufficient samples of typical compound groups. The number of compounds of ketones and polyfunctinals for RON is 2 as shown in Table 4.2 and their box-and-whisker plots become lines as shown in Figure 4.5. In summary, current ML-QSPR models enable interpolation and extrapolation prediction of 15 properties, but more measured property data are demand to improve predictive accuracy and confidence for

typical compound groups.
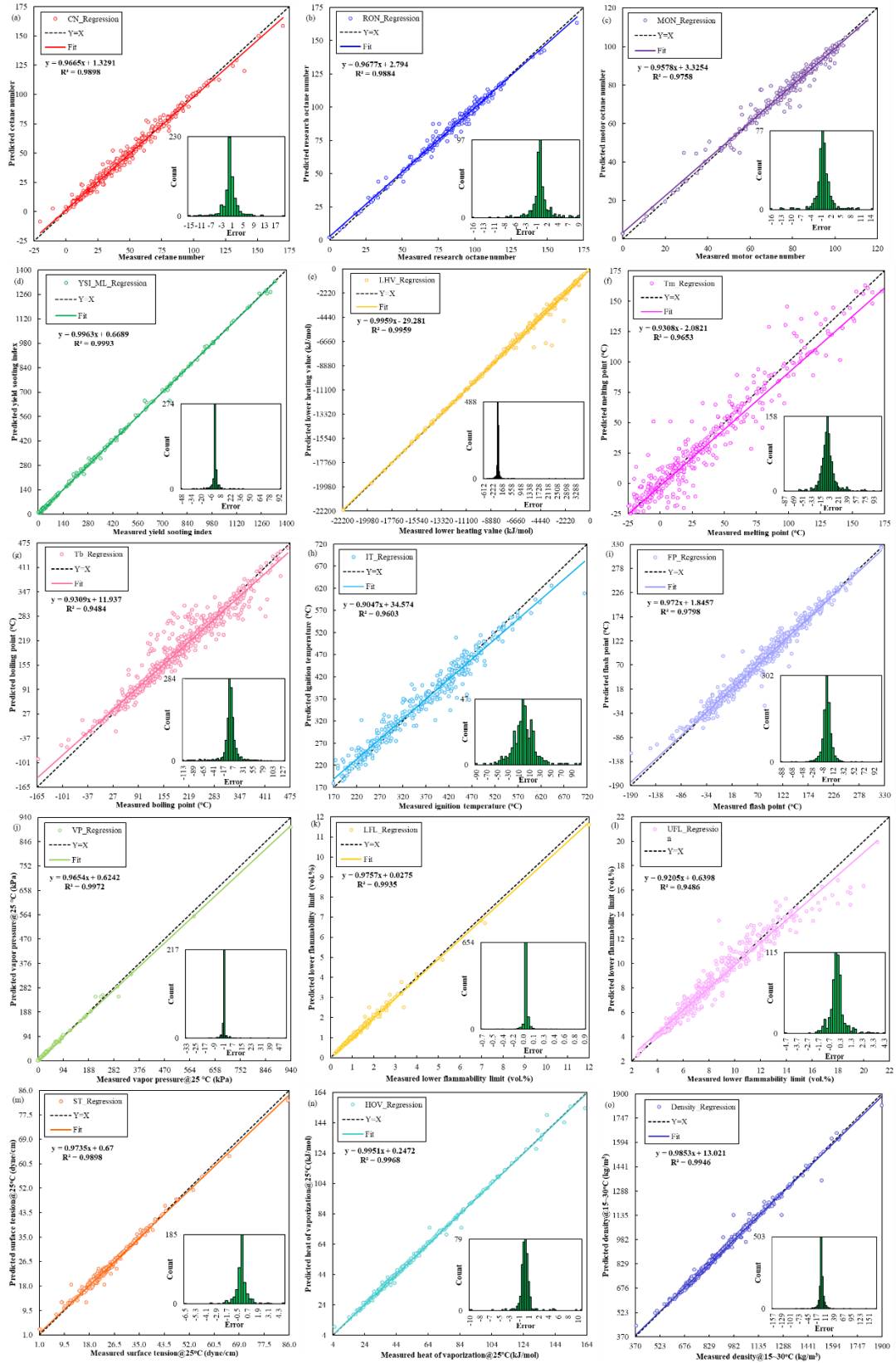
Figure 4.4. Parity plots between measured and predictive values and error distribution histograms for (a) CN, (b) RON, (c) MON, (d) YSI, (e) LHV, (f) T_m, (g) T_b, (h) IT, (i) FP, (j) VP, (k) LFL, (l) UFL, (m) γ, (n) ΔH_{vap}, (o) ρ.

Figure 4.5. Predictive residuals of typical compound groups for (a) CN, (b) RON, (c) MON, (d) YSI, (e) LHV, (f) T_m, (g) T_b, (h) IT, (i) FP, (j) VP, (k) LFL, (l) UFL, (m) γ, (n) ΔH_{vap}, (o) ρ.

Table 4.4. Functions and parameters of the ML-QSPR models for 15 fuel physicochemical properties

Property	Compounds No.	MATLAB module	Type of algorithm		Kernel function	Basic function	Beta	Sigma	Optimizer	Nonlinear solver	programming
CN	635	Regression learner	Gaussian regression	process	Exponential	None	N/A	5.0584	Quasinewton	Active-Set Random	method-
RON	387	Regression learner	Gaussian regression	process	Exponential	None	N/A	4.0483	Quasinewton	Active-Set Random	method-
MON	383	Regression learner	Gaussian regression	process	Rational quadratic	None	N/A	3.6833	Quasinewton	Active-Set Random	method-
T_m	1003	Regression learner	Gaussian regression	process	Exponential	Constant	109.8691	21.1167	Quasinewton	Active-Set Random	method-
T_b	1238	Regression learner	Gaussian regression	process	Rational quadratic	None	N/A	23.1803	Quasinewton	Active-Set Random	method-
ΔH_{vap}	371	Regression learner	Gaussian regression	process	Rational quadratic	None	N/A	1.7424	Quasinewton	Active-Set Random	method-
γ	488	Regression learner	Gaussian regression	process	Exponential	None	N/A	1.2709	Quasinewton	Active-Set Random	method-
LHV	1095	Regression learner	Linear regression		N/A	N/A	N/A	N/A	N/A	N/A	
ρ	1187	Regression learner	Gaussian regression	process	Exponential	None	N/A	0.0196	Quasinewton	Active-Set Random	method-
YSI	457	Regression learner	Gaussian regression	process	Exponential	None	N/A	16.8289	Quasinewton	Active-Set Random	method-
AIT	378	Regression learner	Gaussian regression	process	Exponential	None	N/A	32.7374	Quasinewton	Active-Set Random	method-
FP	1065	Regression learner	Gaussian regression	process	Rational quadratic	None	N/A	11.9396	Quasinewton	Active-Set Random	method-
VP	307	Regression learner	Gaussian regression	process	Exponential	None	N/A	14.4922	Quasinewton	Active-Set Random	method-
LFL	857	Regression learner	Gaussian regression	process	Rational quadratic	None	N/A	0.1169	Quasinewton	Active-Set Random	method-
UFL	579	Regression learner	Gaussian regression	process	Rational quadratic	None	N/A	0.8949	Quasinewton	Active-Set Random	method-

Table 4.5. Correlation coefficients of different compound groups

R^2	CN	RON	MON	T_m	T_b	ΔH_{vap}	γ	LHV	ρ	YSI	IT	FP	VP	LFL	UFL
n-Alkanes	0.9945	0.9995	0.9979	0.9959	0.9915	0.9988	0.9991	1	0.9966	0.9999	0.988	0.9881	0.9995	0.9997	0.979
iso-Alkanes	0.9653	0.9867	0.9671	0.8426	0.9683	0.9783	0.9119	0.9954	0.9368	0.9979	0.8712	0.9505	0.9993	0.9919	0.7855
Cycloalkanes	0.9549	0.983	0.9515	0.936	0.9962	0.9897	0.9814	0.9994	0.9879	0.9968	0.9574	0.9691	0.9994	0.9922	0.9548
Alkenes	0.9903	0.9136	0.8387	0.9547	0.9564	0.9976	0.8958	0.9989	0.9815	0.99947	0.7726	0.9811	0.9798	0.9954	0.9383
Cycloalkenes	0.9994	N/A	N/A	0.9933	0.985	N/A	0.9921	0.5035	0.9894	0.9999	0.9771	0.9535	1	0.9989	0.96
Alkadienes	0.9925	N/A	N/A	0.8383	0.932	1	0.9723	0.9702	0.9908	N/A	0.9239	0.925	1	0.9961	0.686
Alkynes	N/A	0.9916	1	0.8129	0.9811	0.9954	0.9963	0.9998	0.9886	0.9942	1	0.9497	0.9998	0.9994	N/A
Aromatics	0.9932	0.9886	0.9704	0.9301	0.9091	0.9994	0.995	0.9917	0.9662	0.9979	0.9417	0.9592	0.9998	0.9939	0.906
Alcohols	0.9977	0.9924	0.9527	0.9653	0.9342	0.9922	0.9944	0.9997	0.9954	0.9973	0.9463	0.9604	0.9926	0.9926	0.9316
Cycloalcohols	1	N/A	N/A	0.9065	0.8915	1	N/A	0.9892	0.986	N/A	0.5379	0.8254	1	1	N/A
Aldehydes	0.9957	N/A	N/A	0.976	0.9016	1	0.8661	0.9989	0.995	0.9946	0.88	0.7611	0.9995	0.9991	0.9482
Ketones	0.996	1	N/A	0.9695	0.9419	0.9901	0.9149	0.8738	0.983	0.9999	0.8244	0.9643	0.9994	0.9915	0.8906
Cycloketones	0.9924	N/A	1	0.9993	0.9951	1	N/A	0.9984	0.9999	1	1	0.9985	0.9999	1	1
Saturated esters	0.9904	N/A	N/A	0.9636	0.8292	0.9988	0.9354	0.9997	0.9742	0.8535	0.9098	0.9821	0.9989	0.9935	0.9431
Unsaturated esters	0.9914	N/A	N/A	0.801	0.8315	N/A	N/A	0.9998	0.9934	0.7967	0.9016	0.9808	0.9523	0.9802	0.91
Acyclic ethers	0.9842	0.25	0.9885	0.8987	0.9951	0.9886	0.9748	0.9999	0.9979	0.9986	0.9781	0.9459	0.9992	0.9919	0.9381
Furans	0.9961	0.9843	1	0.9854	0.9938	0.9984	N/A	0.9964	0.9979	0.9999	N/A	0.9835	0.9999	1	0.9996
Other cycloethers	0.9998	N/A	N/A	0.9603	0.9913	0.9956	N/A	0.668	0.9995	0.9879	0.9919	0.9868	1	1	1
Carbonate esters	N/A	N/A	N/A	0.9454	0.956	N/A	N/A	0.974	0.9985	1	N/A	0.9938	1	1	N/A
Carboxylic anhydrides	N/A	N/A	N/A	0.9764	0.6742	N/A	N/A	0.9884	0.9962	N/A	0.999	0.8913	1	0.997	0.7156
Peroxides	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Hydroperoxides	N/A	N/A	N/A	N/A	0.9986	N/A	N/A	0.9996	0.9998	N/A	N/A	1	N/A	1	N/A
Polyfunctionals	0.9936	1	1	0.9602	0.9359	0.9865	0.9987	0.9962	0.9982	0.998	0.9765	0.9745	0.9059	0.922	0.9296
Carboxylic acids	0.9993	N/A	N/A	0.9739	0.8527	0.9892	0.997	0.9999	0.997	0.9972	0.9433	0.9869	0.9995	0.9969	0.9086
Mixture	0.9841	0.9985	0.9946	N/A	N/A	N/A	N/A	N/A	N/A	0.9973	N/A	N/A	N/A	N/A	N/A
Overall R^2	0.9898	0.9884	0.9758	0.9653	0.9484	0.9968	0.9898	0.9959	0.9946	0.9993	0.9603	0.9798	0.9972	0.9935	0.9486
Overall MAE	1.565	1.391	1.581	9.539	10.609	0.6566	0.4420	44.014	5.2808	2.711	15.456	6.234	1.235	0.021	0.4556
Overall RMSE	2.776	2.468	2.805	15.214	20.097	1.3991	0.7993	189.563	11.9453	7.567	21.951	10.142	4.798	0.062	0.7251

4.3.2 Method application

Current ML-QSPR models enable to predict the 15 fuel physicochemical properties from the molecular structure. The models possess the capability of interpolation and extrapolation through the 10-fold cross validation. This section comprises 4 case studies of CN/RON/MON, YSI, LHV, T_b to provide insight into the impact of the chemical structure on the properties.

4.3.2.1 Case study of CN/RON/MON

As discussed in Chapter 3, the fuel ignition quality is quantified by CN (DCN), RON and MON. To determine the ignition quality by CFR engine test, CN is applicable to diesel fuel oil or the fuels prone to autoignition [185] while RON/MON is suited to the spark-ignition engine fuel or the fuels with high anti-knock propensity [189, 190] as shown in Table 3.1. Therefore, it is difficult to determine the CN, RON, MON simultaneously by CFR engine test which poses a great challenge to fuel screening of ignition quality. There are two measures to address this issue: (1) Use the conversion formula between CN and RON/MON as shown in Table 3.2. But this method limits to particular fuel types and the predictive accuracy is modest. (2) Develop CN/RON/MON predictive models from the molecular structure or other fuel properties as summarized in Table 3.3. The other fuel properties include ignition delay time (IDT), Fourier-transform infrared absorption spectra, NMR spectroscopy, Molecular mass, hydration energy, boiling point, molar refractivity, octanol/water distribution coefficient, critical pressure, critical volume, critical temperature, etc. The ML-QSPR models belong to the second technology roadmap and for more details on ignition quality characterization please refer to section 3.1.

The measured and predicted results of CN/RON for n-alkanes and 1-alkanols are compared in Figure 4.6 and the ML-QSPR models successfully capture the impact of fuel types and carbon atom numbers on CN and MON. Figure 4.6 reveals two strong non-linear relationships for particular fuel types: (1) the non-linear relationship between CN/RON and carbon atom numbers; (2) the non-linear relationship between CN and RON. These non-linear relationships substantially reduce the predictive accuracy of conversion formulas between CN and

RON/MON. The ML-QSPR models enable the user to forecast the interested properties of specific fuel compounds without measured values or the data is no public access. The measured and predicted CN of 18 fuel types with 7 carbon atom numbers are shown in Figure 4.7 and they are ordered from lowest to highest as: toluene (aromatic) <2,2,3-trimethylbutane(iso-alkane) <methyl-2-hexenoate (unsaturated ester) <1-methylcyclohexanol (cycloalcohol) <cycloheptanone (cycloketone) <methyl hexanoate (saturated ester) <1-heptyne (alkyne) <2-methyl-1,5-hexadiene (alkadiene) <cycloheptane (cycloalkane) <cycloheptene (cycloalkene) <1-heptanol (alcohol) <2-heptanone (ketone) <1-heptene (alkene) <dipropyl carbonate (carbonate ester) <1-butoxy-2-propanol (polyfunctional) <n-heptane (n-alkane) <heptanal (aldehyde) <1-methoxyhexane (acyclic ether). The ML-QSPR models also enable the CN/RON/MON/OS prediction of fuel mixtures as shown in Figure 4.8 and the high predictive accuracy (R^2: 0.9929 for CN, 0.9979 for RON, 0.9932 for MON, 0.8896 for OS) is achieved as shown in Figure 4.9. The QSPR method decomposes the fuel molecules into component fragments and the regression model development is based on the atom level rather than the molecular level. This characteristic of QSPR method enables handling not only the pure compounds but also the fuel mixtures.
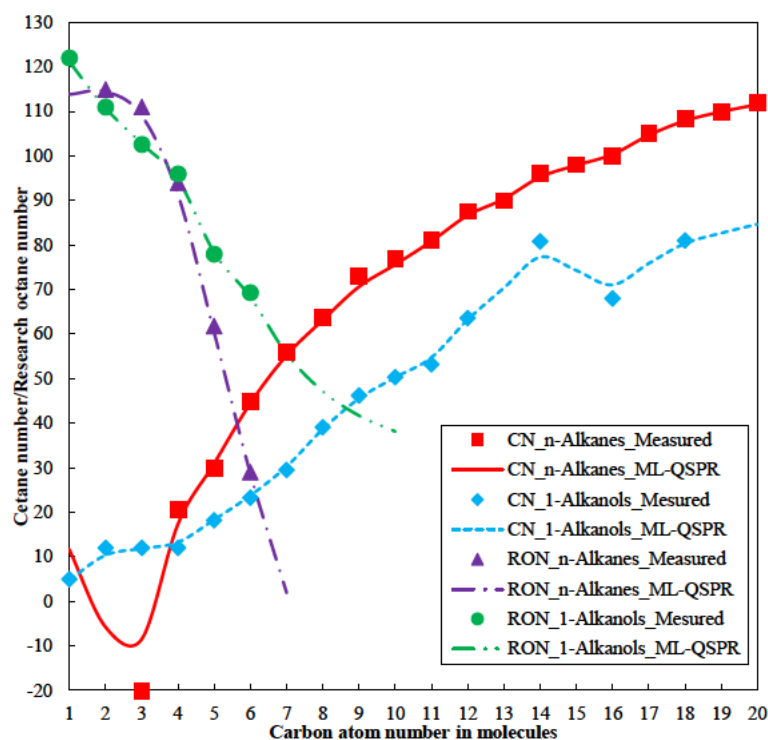
Figure 4.6. Measured (symbols) and predicted (by ML-QSPR method, lines) results of CN/RON for n-alkanes and 1-alkanols.
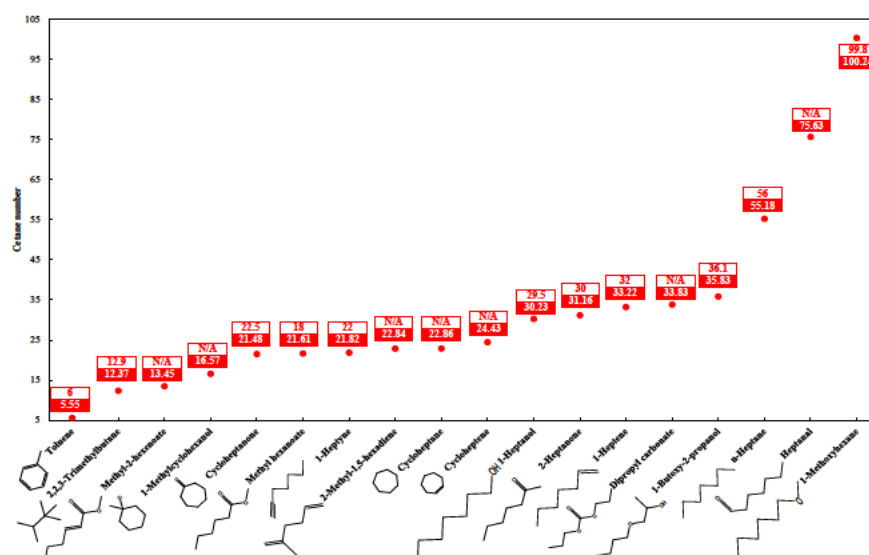
Figure 4.7. CN of different fuel types, numbers with red frames and red backgrounds are measured values and predictive values.
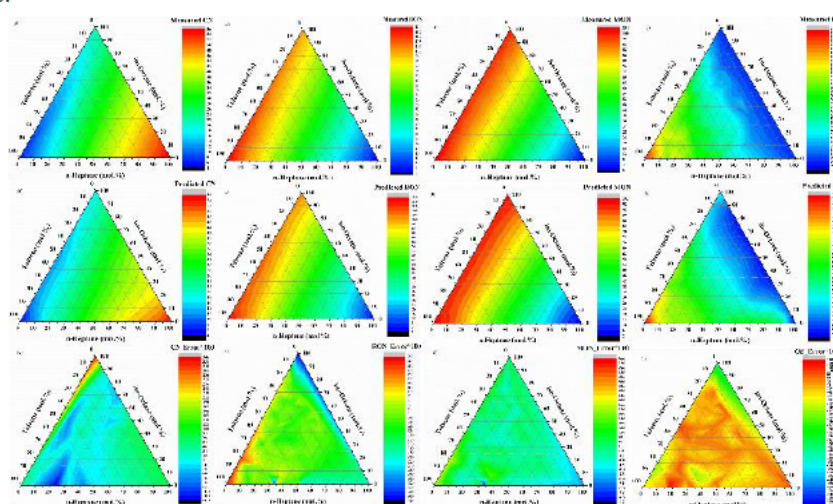
Figure 4.8. Comparison between measured, predicted values and errors of (a)–(c) CN, (d)–(f) RON, (g)–(i) MON, (j)–(l) OS of TPRF mixtures.
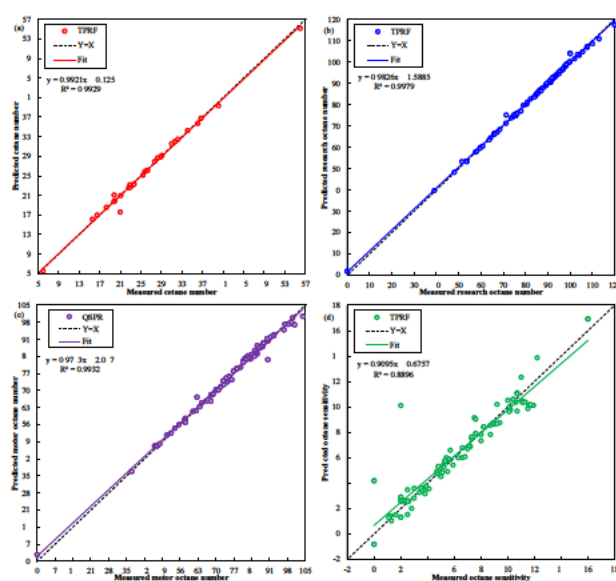
Figure 4.9. Parity plots for (a) CN, (b) RON, (c) MON, (d) OS of TPRF mixtures between measured and predictive values by ML-QSPR method.

146

4.3.2.2 Case study of YSI

Similar to section 4.3.2.1, the ML-QSPR model predict the YSI of 18 fuel types with 7 carbon atom numbers as shown in Figure 4.10 and they are ranked from lowest to highest as: 1-methoxyhexane (acyclic ether) <1-butoxy-2-propanol (polyfunctional) <2-heptanone (ketone) <heptanal (aldehyde) <methyl hexanoate (saturated ester) <dipropyl carbonate (carbonate ester) <1-heptanol (alcohol) <n-heptane (n-alkane) <methyl-2-hexenoate (unsaturated ester) <1-heptene (alkene) <cycloheptane (cycloalkane) <cycloheptanone (cycloketone) <2,2,3-trimethylbutane (iso-alkane) <1-heptyne (alkyne) <2-methyl-1,5-hexadiene (alkadiene) <cycloheptene (cycloalkene) <1-methylcyclohexanol (cycloalcohol) <toluene (aromatic). The ring structure (cyclic hydrocarbon, aromatic) and unsaturated bond (carbon-carbon double bond, carbon-carbon triple bond) dramatically increase the YSI. The sooting tendency becomes more complex as increasing carbon atom numbers and molecular isomerization degree. The ML-QSPR model can predict the impact of one side carbon chain, both sides carbon chain and branching degree on YSI as shown in Figure 4.11. There are 3 distinct characteristics: (1) The YSI increases with carbon atom number and the YSI of one side chain ethers are similar to both side-chain ethers with identical carbon atom numbers. For example, the side chain ethers of methyl propyl ether (17.30), methyl pentyl ether (25.32), methylheptyl ether (29.91), 1-methoxynonane (40.84) are corresponding to diethyl ether (15.38), dipropyl ether (28.06), dibutyl ether (38.27), dipentyl ether (43.36). Many of the C6 or above ethers don't have measured YSI and fore measured data is needed to upgrade the extrapolation capability of the ML-QSPR model. (2) Isomerization of fuel molecules can either increase or decreases the YSI compared to the straight-chain ether which depends on the typical molecular structure. (3) For a given carbon atom number, YSI increases as the methyl group moving toward the oxygen atom until forming the ring structure. For example, the predicted YSI of 2-methoxy-3-methylbutane, 3-methoxypentane, 2-methoxypentane, tert-amyl methyl ether, cyclopentyl methyl ether are 29.55, 31.36, 31.59, 37.75, 61.28 respectively as shown in Figure 4.11.
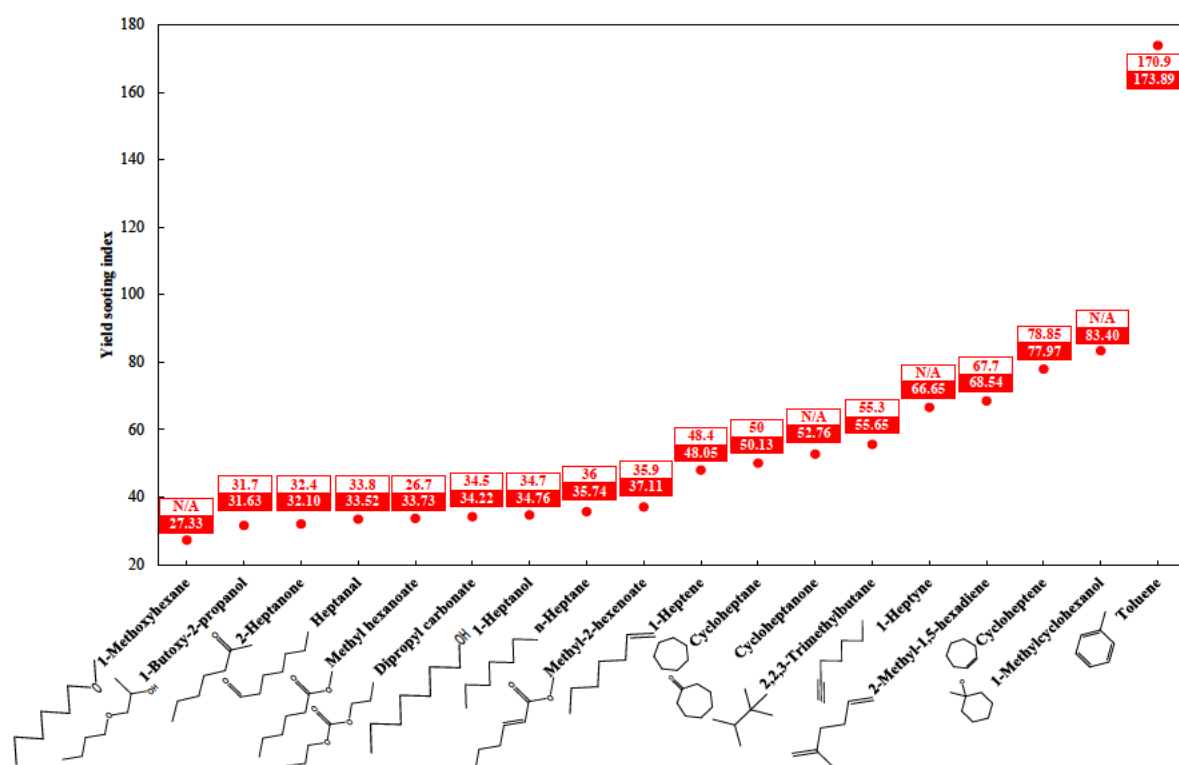
Figure 4.10. YSI of different fuel types with 8 carbon numbers with red frames and red backgrounds are measured values and predictive values.
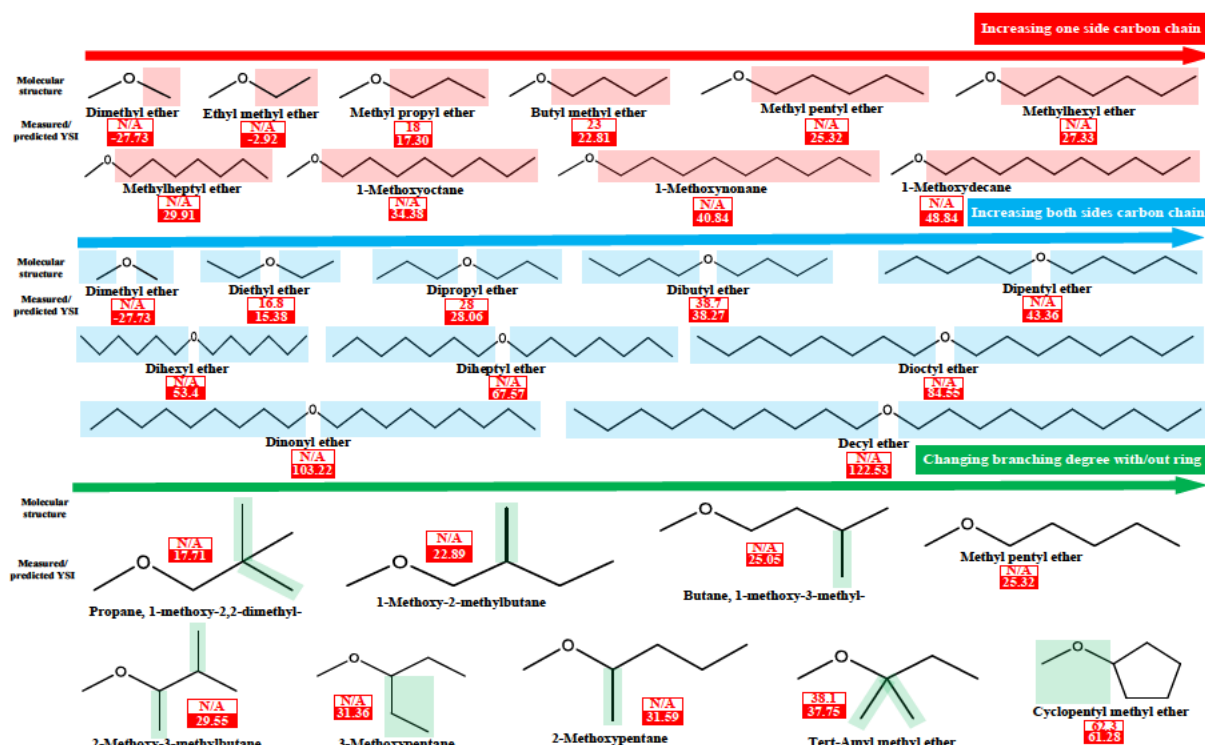
Figure 4.11. YSI of typical ethers, numbers with red frames and red backgrounds are measured values and predictive values.

4.3.2.3 Case study of LHV

The ML-QSPR model is applied to predict the LHV of 15 fuel types up to 12 carbon atom numbers as shown in Figure 4.13 and the comparison between experimental and predicted values are provided in Appendix A. It displays two trends: (1) The LHV increases linearly with carbon atom number for specific fuel types. This phenomenon supports that the linear regression algorithm is the best-suited model to correlate the fuel molecular structure features and LHV as shown in Table 4.4. (2) For the fuel types with the same carbon atom number the LHV increases in the following order: alkylbenzenes <methyl esters <cycloalkanes <2-ketones <aldehydes <2-alkanols <1-alkynes <1-alkanols <2-methylalkenes <1-alkenes <acyclic ethers <2,2-dimetylalkanes <2-methylalkanes <3-methylalkane < n-alkanes. Take the fuel molecules with 12 carbon atom numbers as an example, the following order is obtained: n-hexylbenzene (6781.51 kJ/mol)<methyl undecanoate (6943.83 kJ/mol)<cyclododecane (7015.92 kJ/mol)<2-dodecanone (7140.01 kJ/mol)<dodecanal (7168.07 kJ/mol)<2-dodecanol (7310.07 kJ/mol)<1-dodecyne (7312.56 kJ/mol)<1-dodecanol (7322.92 kJ/mol)<2-methyl-1-undecene (7386.59 kJ/mol)<1-dodecene (7389.97 kJ/mol)<di-n-hexyl ether (7395.25 kJ/mol)<2,2-dimethyldecane (7505.22 kJ/mol)<2-methylundecane (7506.29 kJ/mol)<3-methylundecane (7509.12 kJ/mol)<n-dodecane (7509.78 kJ/mol). The orders may change slightly as varying carbon atoms numbers due to the small difference for typical fuel types (such as 3-methylalkane and n-alkanes).
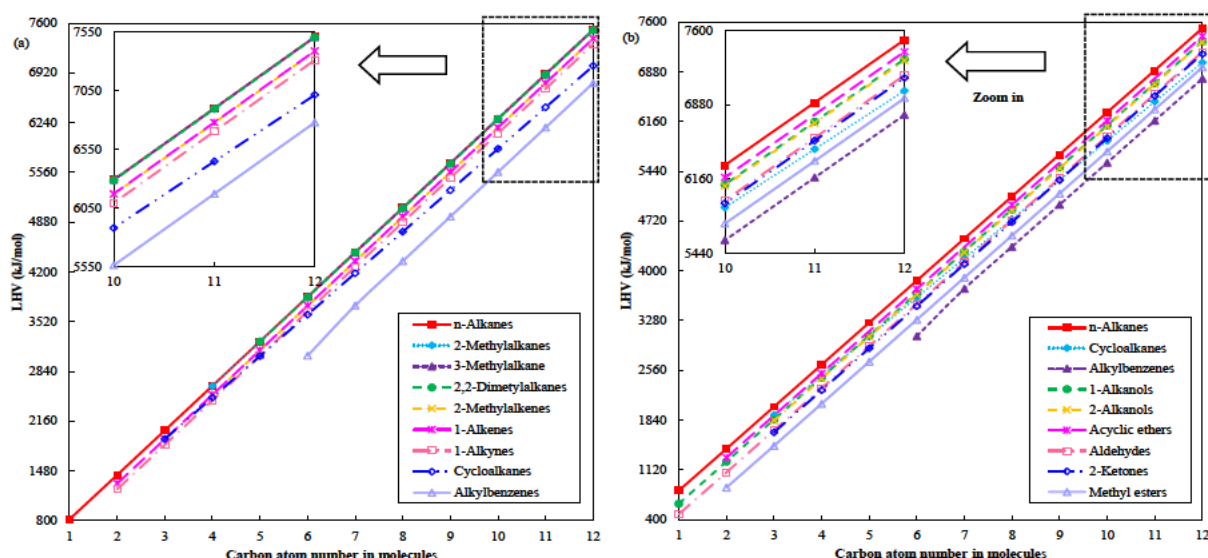
Figure 4.12. Predicted LHV of C_1~C_{12} (a) hydrocarbons, (b) n-alkanes, cycloalkanes, alkylbenzenes, 1-/2-alkanols, acyclic ethers, aldehydes, 2-ketones, methyl esters.

4.3.2.4 Case study of T_b

The predicted T_b of 15 fuel types from C1 to C12 is plotted in Figure 4.13 and the comparison between experimental and predicted values are provided in Appendix A. The fuel molecules with 12 carbon atom number rank from lowest to highest as: dodecanal (aldehydes) < 2,2-dimethyldecane (2,2-dimetylalkanes) < 3-methylundecane (3-methylalkane) < 1-dodecene (1-alkenes) < 2-methylundecane (2-methylalkanes) < 2-methyl-1-undecene (2-methylalkenes) < n-dodecane (n-alkanes) < 1-dodecyne (1-alkynes) < di-n-hexyl ether (acyclic ethers) < n-hexylbenzene (alkylbenzenes) < methyl undecanoate (methyl esters) < 2-dodecanol (2-alkanols) < cyclododecane (cycloalkanes) < 2-dodecanone (2-ketones) < 1-dodecanol (1-alkanols). The T_b and carbon atom number display strong non-linear relationship which is different from LHV, so the T_b orders of various fuel types change significantly at varying carbon atom numbers. ML-QSPR model can provide the users a quantitative prediction on T_b properties for fuel screening.
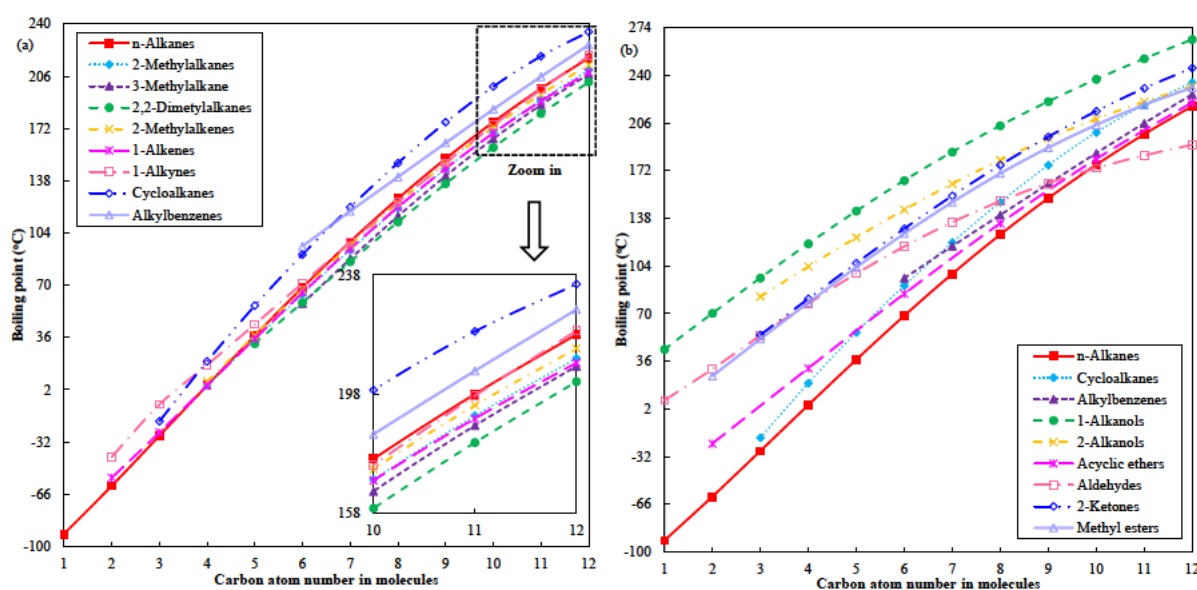
Figure 4.13. Predicted T_b of C1~C12 (a) hydrocarbons, (b) n-alkanes, cycloalkanes, alkylbenzenes, 1-/2-alkanols, acyclic ethers, aldehydes, 2-ketones, methyl esters.

4.4 Conclusions

This work develops ML-QSPR models to predict 15 physicochemical properties of 24 fuel types from molecular structure. They can accelerate property-oriented fuel screening and provide insight into the impact of chemical structure on properties. UOB Ful Property Database containing 1797 pure compounds and 465 mixtures is established to train the regression models and prepare the fuel property matrix. QSPR-UOB 3.0 is proposed to manually extract the structural features and transforms them into a fuel molecular structure matrix. 19 ML algorithms are applied to map the fuel molecular structure matrix and fuel property matrix. 10-fold cross validation is used to examine the model performance to predict new data and prevent over-fitting. The model with the minimum RMSE is chosen as the optimal model for a particular property. Linear regression algorithm is the most suitable for LHV and the Gaussian process regression algorithm is suited to the other 14 properties. Current ML-QSPR models obtain reasonable predictive accuracy and the R^2 of CN, RON, MON, T_m, T_b, ΔH_{vap}, γ, LHV, ρ, YSI, IT, FP, VP, LFL, UFL are 0.9898, 0.9884, 0.9758, 0.9653, 0.9484, 0.9968, 0.9898, 0.9959, 0.9946, 0.9993, 0.9603, 0.9798, 0.9972, 0.9935, 0.9486 respectively.

The success of ML-QSPR models can be attributed to the following reasons: (1) The QSPR-UOB 3.0 functional group classification system substantially increases molecular resolution by complementing aromatic bond at position 10~14, ester group in ring structure, carbonate ester group, carboxylic anhydride group, hydroperoxide group, peroxide group compared to QSPR-UOB 2.0 system. The QSPR-UOB 3.0 system contains only 42 component fragments covering 24 fuel types and it is simple to quantify the molecular structure features compared to the commercial QSPR system with thousands of descriptors. (2) The non-linear systems of various properties are described by nonparametric models created by ML algorithms. The ML algorithms facilitate high predictive accuracy, fast training speed, reasonable memory usage. (3) UOB Fuel Property Database provides a large set of samples for model training and validation. But more experimental data for typical chemical families are needed to further improve the interpolation and extrapolation capacity of the ML-QSPR models.

Chapter 5 Machine Learning and Deep Learning Enabled Fuel Sooting Tendency Prediction from Molecular Structure

5.1 Introduction

Fuel sooting tendency is an important property for fuel screening to identify the suitable molecule and it can be quantified by smoke point (SP) [362], threshold sooting index (TSI) [363], oxygen extended sooting index (OESI) [364], micropyrolysis index (MPI) [365] and yield sooting index (YSI) [321, 361, 366]. SP measures the sooting tendency of kerosene and aviation turbine fuel which represents the maximum height of a smokeless flame of fuel burned in a wick-fed lamp [363]. TSI is an index free of apparatus dependence modified from the SP and takes into account the effect of molecular weight on sooting tendency [363]. OESI accounts for the impact of oxygen atom in fuel molecules on stoichiometric air [364]. These SP-based indices encounter problems to determine the proper flame shape for compounds with high SP. MPI and YSI are independent of SP. MPI accounts for the dependence of sooting tendency on oxygen concentration [365]. YSI is a robust measure of the sooting tendency for a wide variety of fuel types including alkanes, alkenes, cycloalkanes, aromatics and oxygenates [321], thus it is chosen as the sooting tendency indicator in this work. Soot formation model can provide insightful understanding into the impact of chemical structure on the sooting tendency and screen the renewable fuels with soot-reducing benefits. There are 3 technology roadmaps to build the YSI predicted model as summarized in Table 5.1: (1) Machine learning (ML) roadmap. Group contribution method (GCM) or quantitative structure-property relationship (QSPR) method is applied to manually extract chemical structure features. By using the GCM or QSPR method, it is possible to extend the model application domain from pure compounds to fuel mixtures. ML algorithms are used to correlate the structural features and target property. There are two ways to improve the model predictive capacity of fuel properties: one is to propose an innovative functional group classification system to increases the component fragment resolution; the other is to adopt new ML algorithms to fit the model input and output. (2) Deep learning (DL) roadmap. A deep neural network is used to automatically learn the molecular

structure features and perform regression. The commonly used deep network is convolution neural network (CNN), but it is usually utilized in computer vision domains such as image classification, object detection, semantic segmentation, super-resolution, etc. The popular CNN architectures and their designed application domains are summarized in Table 5.2 and it indicates that few CNN is specially developed for a regression problem. Without tailor-made CNN architectures for a regression problem, the only way to conduct regression operation is by transfer learning but the model performances contain considerable uncertainty. Schweidtmann et al. [367] for the first time to utilize CNN into regression task to predict the fuel properties of cetane number, research/motor octane number. Application of DL into fuel properties prediction is still in its infancy, there is no existing DL-CNN model for YSI prediction. (3) Molecular Dynamics (MD) simulation. It uses ReaxFF reactive MD simulation to mimic the fuel oxidation, pyrolysis and soot formation process. The advantage of this method is that the molecular dynamic simulation does not require the chemical kinetic mechanism as an input, it is particular suitable to those fuel molecules with poor-known or even unknown chemistry [323].

This work applies both ML and DL roadmap to build the YSI prediction model from the molecular structure and conduct a systematic performance evaluation. In the ML route, QSPR-UOB 3.0 functional group classification system is developed to manually extract chemical structure features and transform them into a fuel molecular structure matrix. ML algorithm is used to map the fuel molecular structure matrix and YSI matrix. In DL, a tailor-made CNN network of SDSeries38 is developed for automated features learning and regression operation based on the molecule images. Transfer learning is performed on 10 classical CNN (which are designed for image classification or object detection) for YSI prediction and they are compared with the SDSeries38 network in predictive accuracy and training speed.

Table 5.1. Comparison of YSI predicted method in this study with published methods

Feature extraction method	Regression method	Year	Compound No.	Training dataset	R^2	RMSE	Ref.
GCM (35 chemical groups)	Linear regression	2015	265	YSI database Volume 1 [368]	0.95	N/A	[369]
QSAR (5270 molecular descriptors)	Multilayer perceptron	2017	297	YSI database Volume 1 [368]	N/A	N/A	[319]
GCM (66 fragment types)	Bayesian linear regression	2018	441	YSI database Volume 2 [370]	N/A	N/A	[321]
GCM (37 structural groups)	Kernel ridge regression	2019	204	YSI database Volume 2 [370]	0.9018	N/A	[322]
QSPR (15 descriptors)	Artificial neural network	2019	421	YSI database Volume 1 [368]	N/A	13.478	[320]
ReaxFF software	Molecular dynamics simulation	2020	2	YSI database Volume 2 [370]	N/A	N/A	[323]
GCM (37 chemical groups)	Artificial neural network	2021	449	YSI database Volume 2 [370]	0.99	N/A	[371]
QSPR-UOB 3.0 (42 chemical groups)	Gaussian process regression	2021	456	YSI database Volume 2 [370]	0.9993	7.567	This work
CNN-SDSeies38	CNN-SDSeies38	2021	456	YSI database Volume 2 [370]	0.9953	19.58	This work

Table 5.2. Popular CNN architectures developed by the DL research community for feature extraction, classification, regression and transfer learning

Feature[α]	CNN	Purpose	Training dataset	Year	Top-1 accuracy (%)	Top-5 accuracy (%)	Depth	Size (MB)	Parameters (M)	Input size	Ref.
Versatile	SDSeries38	Regression	YSI database Volume 2 [370]	2021	N/A[β]	N/A[β]	10	22.7	6.03	150×300×3	This work
Accuracy	Xception	Image classification	ImageNet[372, 373]	2017	79	94.5	71	85	22.9	299×299×3	[374, 375]
Accuracy	Densenet201	Image classification	ImageNet[372, 373]	2017	78.54	94.46	201	77	20	224×224×3	[376, 377]
Accuracy	Inceptionv3	Image classification	ImageNet[372, 373]	2015	78.2	94.1	48	89	23.9	299×299×3	[378, 379]
Accuracy	ResNet18	Image classification/ Object detection	ImageNet[372, 373]	2015	72.12	91.8	18	44	11.7	224×224×3	[380, 381]
Accuracy	ResNet50	Image classification/ Object detection	ImageNet[372, 373]	2015	77.15	93.29	50	96	25.6	224×224×3	[381, 382]
Accuracy	ResNet101	Image classification	ImageNet[372, 373]	2016	78.25	93.29	101	167	44.6	224×224×3	[380, 381]
Accuracy	DarkNet19	Image classification	ImageNet[372, 373]	2016	27.1	8.8	19	78	20.8	256×256×3	[383, 384]
Accuracy	DarkNet53	Image classification	ImageNet[372, 373]	2016	22.8	6.2	53	155	41.6	256×256×3	[383, 384]
Accuracy	InceptionResNetV2	Image classification	ImageNet[372, 373]	2017	80.4	95.3	164	209	55.9	299×299×3	[385]
Speed	AlexNet	Image classification	ImageNet[372, 373]	2012	59.3	81.8	8	227	61	227×227×3	[386, 387]
Speed	GoogleNet	Image classification/ Object detection	ImageNet[372, 373]	2014	68.7	93.33	22	27	7	224×224×3	[388, 389]
Speed	VGG-16	Image classification	ImageNet[372, 373]	2014	74.4	91.9	16	515	138	224×224×3	[390, 391]
Speed	VGG-19	Image classification	ImageNet[372, 373]	2014	74.5	92	19	535	144	224×224×3	[390, 391]
Size	Shufflenet	Image classification	ImageNet[372, 373]	2018	73.7	91.23	50	5.4	1.4	224×224×3	[392-395]
Size	Mobilenetv2	Image classification	ImageNet[372, 373]	2018	72	91.76	53	13	3.5	224×224×3	[396, 397]
Size	Squeezenet	Image classification	ImageNet[372, 373]	2016	57.5	80.3	18	5.2	1.24	227×227×3	[398, 399]
Size	EfficientnetB0	Image classification	ImageNet[372, 373]	2019	77.1	93.3	82	20	5.3	224×224×3	[400, 401]

[α] The network features are divided according to network size (network requires low memory footprint and can be deployed to low-compute, low-power devices), accuracy (network enables learn informative features and achieves high accuracy score), speed (simple network enables fast iterations and seeks a balance between speed and accuracy).

[β] Top-1 accuracy and Top-5 accuracy are used to evaluate the precision of image classification, SDSeries38 network is specially designed for regression tasks and it is not yet tested in classification problem.

5.2. Modeling approach

5.2.1 Methodological overview

Two technology roadmaps of ML and DL are developed for YSI prediction as shown in Figure 5.1. (1) In the ML technology roadmap, quantitative structure–property relationship (QSPR) is used to extract and digitalize the fuel molecular structure features. 19 ML algorithms are implemented to train the regression models in parallel. The model with minimum RMSE is chosen as the optimal ML-QSPR model for YSI regression. Especially, the QSPR-UOB 3.0 functional group classification system is adopted for structural feature extraction. The ML-QSPR model development is described in section 5.2.2. (2) In the DL technology roadmap, a convolution neural network (CNN)-SDSeries38 is designed to perform automated feature extraction and regression task. The SDSeries38 network has a standard series network architecture containing 9 feature learning modules and 1 regression module in series. The feature learning modules start with simples features and increase in complexity to the features that define the sooting tendency as the modules progress [402]. The regression module flattens the network's 2D spatial features into a 1 D vector of each image and maps the vectors with responses for regression purpose. The DL-CNN model development and relevant model information are explained detaily in section 5.2.3. The YSI dataset is established for regression model training and validation which contains 444 pure compounds and 12 mixtures covering 24 fuel types as shown in Table 5.3. The YSI data of pure compounds are derived from Yield Sooting Index Database Volume 2 [370] while those of mixtures are collected from [321, 361].

Machine learning based YSI prediction

Input images

Feature extraction

Regression

Output

Deep learning based YSI prediction

Input images

Feature extraction & Regression

Output
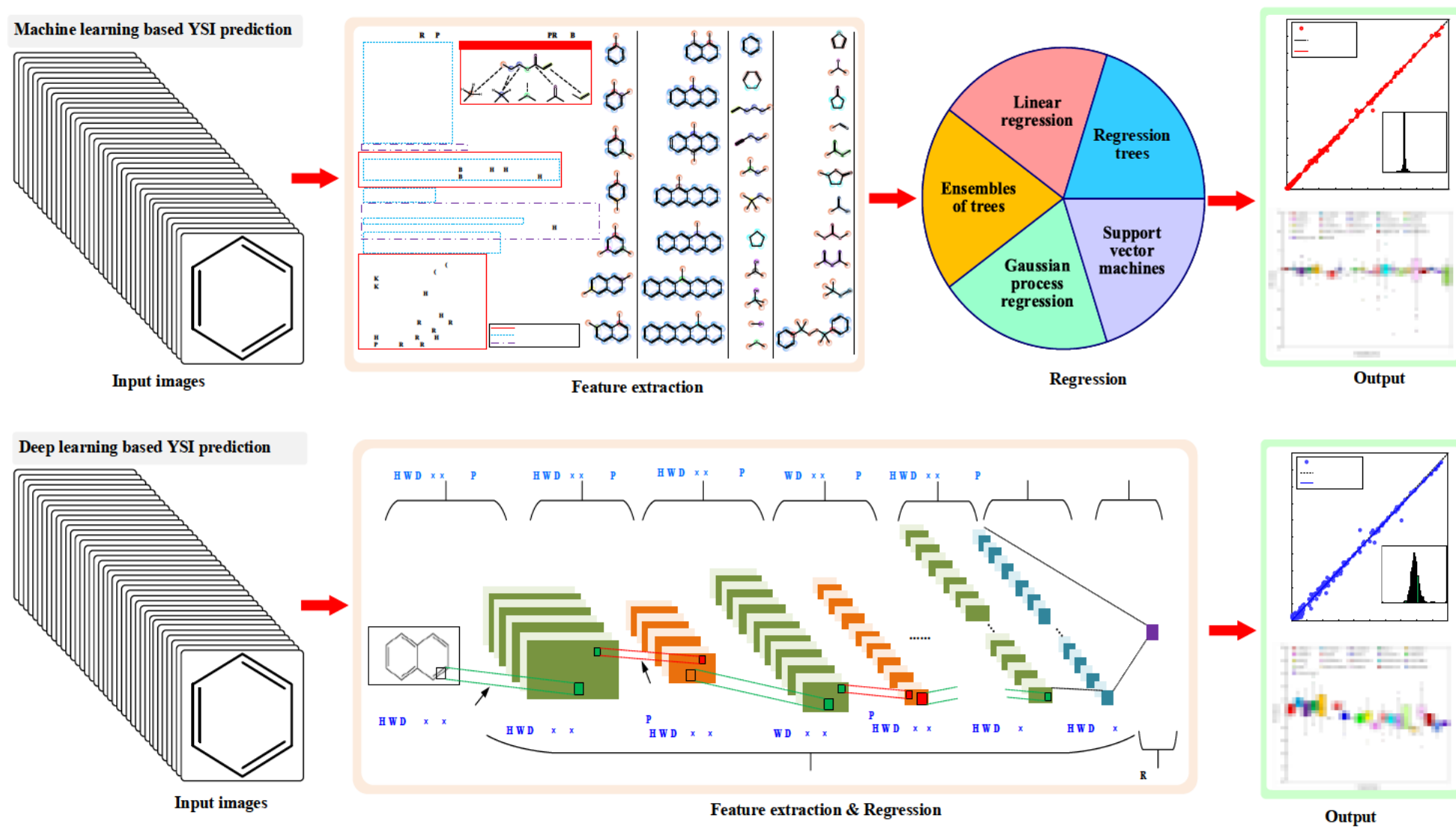
Figure 5.1. Frameworks showing two YSI prediction roadmaps of ML regression and DL regression.

Table 5.3. Number of compounds of different chemical classes in the YSI dataset

Compound class	No. of compounds (measured data)
n-Alkanes	8
iso-Alkanes	22
Cycloalkanes	20
Alkenes	36
Cycloalkenes	12
Alkadienes	1
Alkynes	3
Aromatics	122
Alcohols	49
Cycloalcohols	1
Aldehydes	18
Ketones	28
Cycloketones	2
Saturated esters	40
Unsaturated esters	20
Acyclic ethers	34
Furans	3
Other cyclic ethers	4
Carbonate ester	3
Carboxylic anhydride	N/A
Peroxide	N/A
Hydroperoxide	N/A
Polyfunctionals	14
Carboxylic acids	4
Total oxygenates	220
Mixture	12
Total	456 (444 pure compounds+12 mixtures)

5.2.2 ML-QSPR regression model development

The development of ML-QSPR models for 15 fuel properties has been explained detailly in Chapter 4 and the

ML workflow is illustrated in Figure 4.3, so this section only gives a necessary instruction of the model principle

for YSI prediction. QSPR-UOB 3.0 functional group classification system is implemented to manually extract

molecular structure features and transforms these features into a fuel molecular structure matrix. The elements of

the fuel molecular structure matrix of a particular fuel molecule represent the occurrence of the component

fragments listed in QSPR-UOB 3.0 system. The YSI training dataset provides the fuel property matrix for

regression purpose. 19 ML algorithms are adopted to map the fuel molecular structure matrix and fuel property

matrix and build the regression models as shown in Table 3.6. 10-fold cross validation is applied to validate the

model performance to predict new data and prevent over-fitting. The model with minimum RMSE is selected as

the optimal ML-QSPR model for YSI prediction. The model functions and parameters are demonstrated in Table

4.4.

5.2.3 DL-CNN regression model development

The DL workflow to develop a deep neural network is illustrated in Figure 5.2 which comprises 3 steps. (1) Data access and preparation. YSI dataset containing 444 pure compounds is established based on Yield Sooting Index Database Volume 2 [370]. Accordingly, the images of 444 fuel molecules are prepared by Chemdraw software [403] and the image size would be resized to 150×300×3 (height×width×depth) before entering the SDSeries38 network. The images and YSI values are the DL-CNN model input and output respectively. (2) Develop and train the CNN predictive model. MATLAB deep network designer toolbox is used to build the CNN architecture and the model training is perform by Experiment Manager toolbox or MATLAB command lines. In this work, the model training is performed on a device with Intel® Core™ i5-4200M Processor (2 cores, 4 threads, 3.1 GHz), 32GB memory, Intel® HD Graphics 4600 integrated graphics, GeForce GT 755M independent graphics. (3) Deploy the trained network into production systems on desktops, mobiles, embedded devices as an application. There are 4 most common ways to deploy the DL-CNN model depends on the intended purposes: (a) Deploy as an APP on desktops or mobile devices by MATLAB Compiler. (b) Deploy to the cloud or a server by MATLAB Production Server. (c) Deploy to desktop-based GPUs by MATLAB GPU Coder. (d) Deploy to embedded devices such as GPUs or processors.

A standard series CNN with 38 layers, SDSeries38, is built for YSI prediction which contains 1 image input layer, 9 feature learning modules and 1 regression module as shown in the red box of Figure 5.3. The image input layer imports the graphs into the network and performs data normalization. Each feature learning module is a stack of convolution layer, batch normalization layer, rectified linear unit (ReLU) layer, max pooling layer. The convolution layer appliers sliding convolutional filter to the input and the feature map forms as a result of the filter moving along the layer input as shown in the red box of Figure 5.4. The filter size is $3×3× n_F$ (height×width×number of filters) and n_F doubles when advances to the next convolution layer until reaching 512. n_F is equal to 8, 16, 32, 64, 128, 256, 512 for 9 convolution layers which generates the corresponding number of

feature map to store the features (local information) as feature map size scaling down by the max pooling layers. The filters scan through the input with a stride of 1 and padding of 0, thus the feature map size remains consistent before and after the convolution layer. The batch normalization layer normalizes the layer input across a mini-batch which is used to accelerate the network training and weaken the network initialization sensitivity. The ReLU layer implements a threshold operation to all the input elements, where all the negative values are set to zero as shown in Eq. (5.1). To reduce the connection numbers to the sequential layer, max pooling layer is adopted to down-sample the layer input. Max pooling layer divides the input into a few rectangular pooling areas and extracts the maximal value of each area as shown in the blue box of Figure 5.4. In other words, the max pooling layer down-sample the input feature maps by dividing the input into rectangular pooling regions of height 2, width 2 and returns the maximal values of each region as shown in the blue box. The dropout layer randomly sets the input elements to zero with a given probability. By applying 9 feature learning modules, the feature map size shrinks progressively from 150×300×3 to 2×3×512 and the feature map size of each layer is listed in Table 5.4. The drop of the feature map size alongside the deeper feature learning modules reduces the capacity requirement of computer memory. The required memory decreases from 1598.4 MB of the 1st convolution layer to 13.64 MB of the 9th convolution layer as shown in Table 5.4. The number of filters at each convolution layers increases as 8, 16, 32, 64, 128, 256, 512 and the numbers of feature map increases correspondingly to store the learned features as shown in the blue box of Figure 5.3. Thus, the model parameters increase from 224 of the 1st convolution layer to 2359808 of the 9th convolution layer as shown in Table 5.4. The regression module consists of 1 fully connected layer and 1 regression layer. The fully connected layer combines all the learned features from the previous layers to recognize the larger patterns. The regression layer calculates the half-mean-squared-error loss of the predictive response by Eq. (5.2).

$$f(x) = \begin{cases} x, & x \geq 0 \\ x, & x < 0 \end{cases} \tag{5.1}$$

$$loss = \frac{1}{2} \sum_{i=1}^{R} \frac{(t_i - y_i)^2}{R} \tag{5.2}$$

The model parameters set up and SDSeries38 network training compose 6 steps. (1) Define the solver and a maximum number of epochs. Stochastic gradient descent with momentum (SGDM) is adopted as a solver to move toward the negative gradient of the loss function. SGDM can overcome local minima and saddle points to avoid gradient descent to zero [404]. The maximum number of epochs is set as 40 and the training data is shuffled and looped over mini-batches for each epoch for SDSeries38 network training as shown in Table 5.5. (2) Specify and modify the learning rate. The learning rate initializes as 1E-05 until epoch 20 and then it drops to 1E-06 until the maximum number of epochs. (3) Select hardware resources. (4) Train the CNN. The tuned hyperparameters of the SDSeries38 network are presented in Table 5.5.
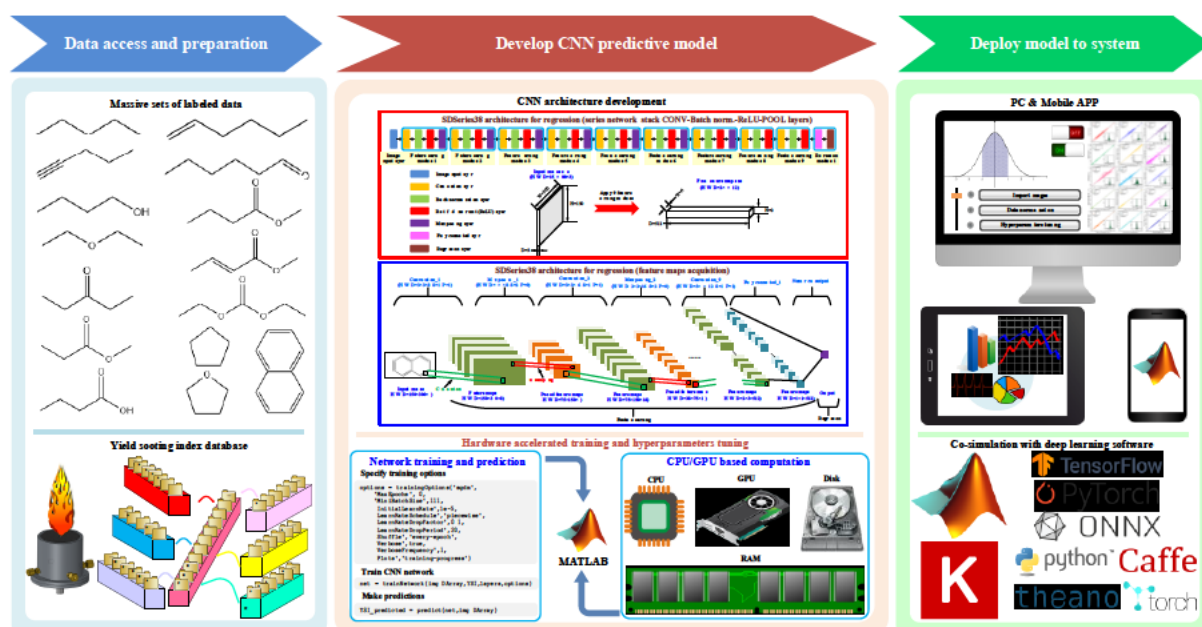
Figure 5.2. DL workflow to develop a deep neural network for regression.
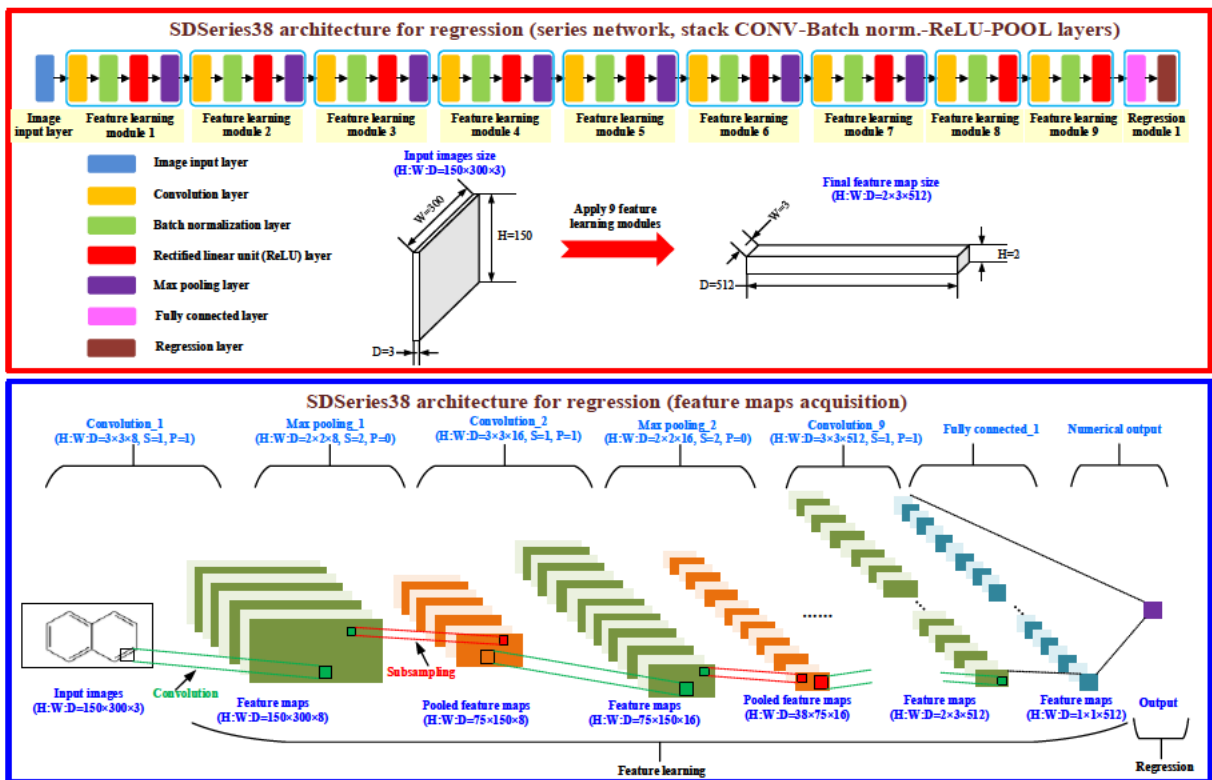
Figure 5.3. The network architecture of SDSeries38 for regression. The sequence of deep learning layers (red box) and obtained feature maps (blue box).
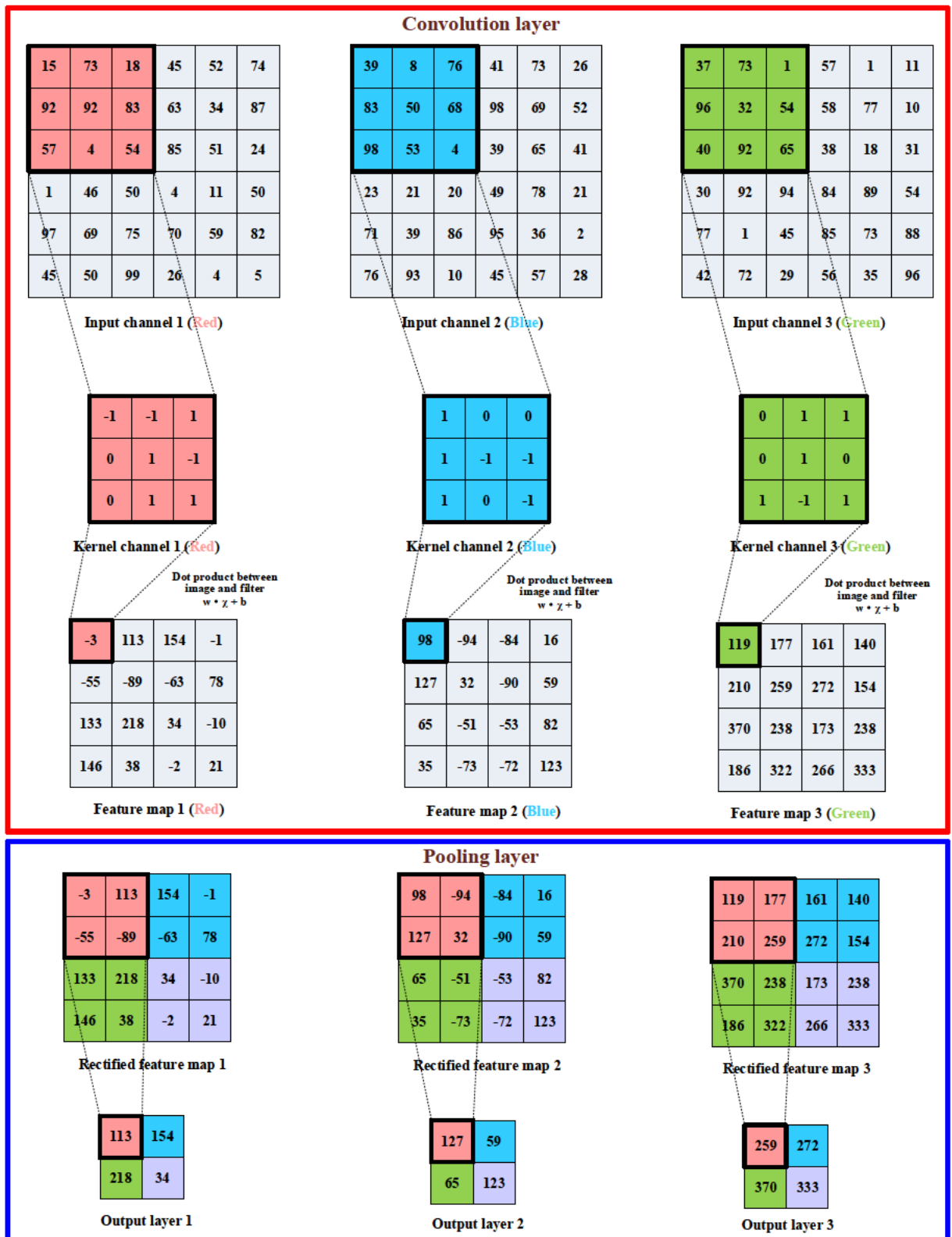
Figure 5.4. Convolution operation on a M×N×3 image matrix with a 3x3x3 filter (red box) and 2×2 max pooling operation over convolved feature map (blue box).

Table 5.4. SDSeries38 architectural dimensions

No.	Layer type	Filter size			Stride		Padding		Activations size			Parameters				Memory	Convolution
		Height	Width	No.	Vertical	Horizontal	Top/Bottom	Left/Right	Height	Width	Depth	Weights	Bias	Offset	Scale	(MB) [β]	operation
1	Image input	N/A	N/A	N/A	N/A	N/A	N/A	N/A	150	300	3	N/A	N/A	N/A	N/A	599.4	0
2	Convolution	3	3	8	1	1	Same [α]	Same	150	300	8	3×3×3×8	1×1×8	N/A	N/A	1598.4	9720000
3	Batch normalization	N/A	N/A	N/A	N/A	N/A	N/A	N/A	150	300	8	N/A	N/A	1×1×8	1×1×8	N/A	N/A
4	ReLU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	150	300	8	N/A	N/A	N/A	N/A	N/A	N/A
5	Max pooling	2	2	8	2	2	0	0	75	150	8	N/A	N/A	N/A	N/A	399.6	0
6	Convolution	3	3	16	1	1	Same	Same	75	150	16	3×3×8×16	1×1×16	N/A	N/A	799.2	12960000
7	Batch normalization	N/A	N/A	N/A	N/A	N/A	N/A	N/A	75	150	16	N/A	N/A	1×1×16	1×1×16	N/A	0
8	ReLU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	75	150	16	N/A	N/A	N/A	N/A	N/A	0
9	Max pooling	2	2	16	2	2	0	0	38	75	16	N/A	N/A	N/A	N/A	202.464	0
10	Convolution	3	3	32	1	1	Same	Same	38	75	32	3×3×16×32	1×1×32	N/A	N/A	404.928	13132800
11	Batch normalization	N/A	N/A	N/A	N/A	N/A	N/A	N/A	38	75	32	N/A	N/A	1×1×32	1×1×32	N/A	0
12	ReLU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	38	75	32	N/A	N/A	N/A	N/A	N/A	0
13	Max pooling	2	2	32	2	2	0	0	19	38	32	N/A	N/A	N/A	N/A	102.58176	0
14	Convolution	3	3	64	1	1	Same	Same	19	38	64	3×3×32×64	1×1×64	N/A	N/A	205.16352	13307904
15	Batch normalization	N/A	N/A	N/A	N/A	N/A	N/A	N/A	19	38	64	N/A	N/A	1×1×64	1×1×64	N/A	0
16	ReLU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	19	38	64	N/A	N/A	N/A	N/A	N/A	0
17	Max pooling	2	2	64	2	2	0	0	10	19	64	N/A	N/A	N/A	N/A	53.9904	0
18	Convolution	3	3	128	1	1	Same	Same	10	19	128	3×3×64×128	1×1×128	N/A	N/A	107.9808	14008320
19	Batch normalization	N/A	N/A	N/A	N/A	N/A	N/A	N/A	10	19	128	N/A	N/A	1×1×128	1×1×128	N/A	0
20	ReLU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	10	19	128	N/A	N/A	N/A	N/A	N/A	0

21	Max pooling	2	2	128	2	2	0	0	5	10	128	N/A	N/A	N/A	N/A	28.416	0
22	Convolution	3	3	256	1	1	Same	Same	5	10	256	3×3×128×256	1×1×256	N/A	N/A	56.832	14745600
23	Batch normalization	N/A	N/A	N/A	N/A	N/A	N/A	N/A	5	10	256	N/A	N/A	1×1×256	1×1×256	N/A	0
24	ReLU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	5	10	256	N/A	N/A	N/A	N/A	N/A	0
25	Max pooling	2	2	256	2	2	0	0	3	5	256	N/A	N/A	N/A	N/A	17.0496	0
26	Convolution	3	3	512	1	1	Same	Same	3	5	512	3×3×256×512	1×1×512	N/A	N/A	34.0992	17694720
27	Batch normalization	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3	5	512	N/A	N/A	1×1×512	1×1×512	N/A	0
28	ReLU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3	5	512	N/A	N/A	N/A	N/A	N/A	0
29	Max pooling	2	2	512	2	2	0	0	2	3	512	N/A	N/A	N/A	N/A	13.63968	0
30	Convolution	3	3	512	1	1	Same	Same	2	3	512	3×3×512×512	1×1×512	N/A	N/A	13.63968	14155776
31	Batch normalization	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2	3	512	N/A	N/A	1×1×512	1×1×512	N/A	0
32	ReLU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2	3	512	N/A	N/A	N/A	N/A	N/A	0
33	Convolution	3	3	512	1	1	Same	Same	2	3	512	3×3×512×512	1×1×512	N/A	N/A	13.63968	14155776
34	Batch normalization	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2	3	512	N/A	N/A	1×1×512	1×1×512	N/A	0
35	ReLU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2	3	512	N/A	N/A	N/A	N/A	N/A	0
36	Dropout	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2	3	512	N/A	N/A	N/A	N/A	N/A	0
37	Fully connected	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1	1	3072	1×3072	1×1	N/A	N/A	0.00444	0
38	Regression output	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0

[α] Add padding of size at training/prediction time so that the output has the same size as the input when the stride equals 1. If the stride is larger than 1, then the output size is ceil(inputSize/stride), where input size is the height or width of the input and stride is the stride in the corresponding dimension.

[β] The required memory is calculated based on the YSI training dataset of 444 fuel molecule graphs.

Table 5.5. Tuned hyperparameters of SDSeries38 network for YSI prediction

Specification	Item
Solver	SGDM optimizer
Momentum	0.9
Initial learn rate	1.0E-05
Learn rate schedule	Piecewise
Learn rate drop factor	0.1
Learn rate drop period	20
L2 Regularization	1.0E-05
Gradient threshold method	L2norm
Gradient threshold	Inf
Max epochs	40
Mini batch size	111
Verbose	1
Verbose frequency	1
Shuffle	Every-epoch
Execution environment	auto
Plots	Training-progress
Sequence length	Longest
Sequence padding value	0
Sequence padding direction	Right
Dispatch in background	0
Reset input normalization	1

5.3 Results and discussion

5.3.1 Predictive accuracy of ML-QSPR model and DL-CNN model

The correlation coefficients of the ML-QSPR model and DL-CNN model-SDSeries38 are 0.9993 and 0.9953 respectively as shown in Figure 5.5. The RMSE of the ML-QSPR model and DL-CNN model-SDSeries38 are 7.567 and 19.58 as shown in Table 5.6. The predictive accuracy of the ML-QSPR model outperforms the SDSeries38 network. Especially, the ML-QSPR model obtains much lower YSI predicted residuals for most of the compound groups than the SDSeries38 network as shown in Figure 5.6. The R^2 of different compound groups also supports that ML-QSPR obtains good overall and subgroup predictive performance as shown in Table 5.6. The R^2 of cycloalkanes, alkenes, alcohols, saturated esters, unsaturated esters, furans, other cyclic ethers, polyfunctionals for SDSeries38 network are 0.4992, 0.2389, 0.1726, 0.00484, 0.2086, 0.0008, 0.3858, 0.0715 as shown in Table 5.6. The DL roadmap implements CNN to learn the structural features but it still needs significant improvement to capture the dependence of YSI on molecular structures for typical chemical classes.
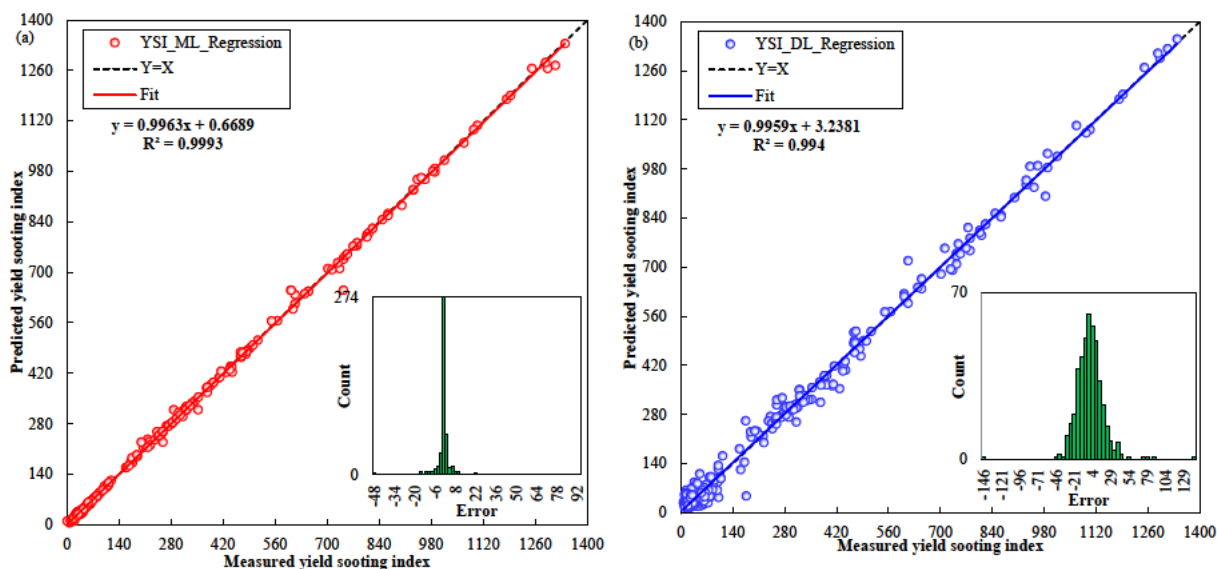
Figure 5.5. Parity plots of YSI between measured and predictive values and error distribution histogram by (a) ML-QSPR model and (b) DL-CNN model of SDSeries38 network.
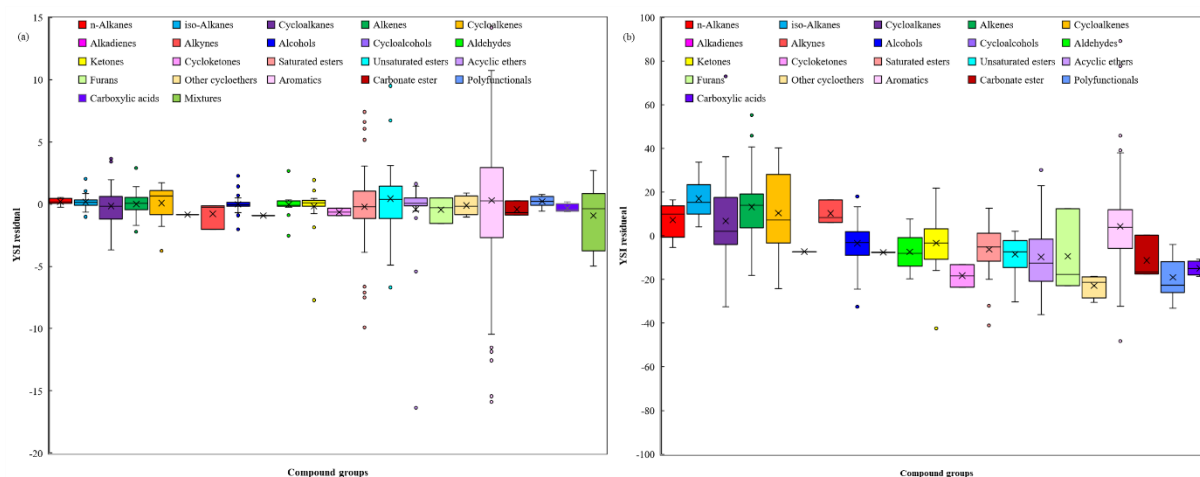
Figure 5.6. YSI predicted residuals of typical compound groups by (a) ML-QSPR model and (b) DL-CNN model of SDSeries38 network.

Table 5.6. Correlation coefficients and RMSE of YSI prediction by ML-QSPR model and DL-CNN model of SDSeries38 network

R^2	ML-QSPR	DL-CNN
n-Alkanes	0.9999	0.8656
iso-Alkanes	0.9979	0.5503
Cycloalkanes	0.9968	0.4992
Alkenes	0.99947	0.2389
Cycloalkenes	0.9999	0.9764
Alkadienes	N/A	N/A
Alkynes	0.9942	0.9176
Aromatics	0.9979	0.9927
Alcohols	0.9973	0.1726
Cycloalcohols	N/A	N/A
Aldehydes	0.9946	0.6713
Ketones	0.9999	0.9717
Cycloketones	1	1
Saturated esters	0.8535	0.00484
Unsaturated esters	0.7967	0.2086
Acyclic ethers	0.9986	0.8746
Furans	0.9999	0.0008
Other cycloethers	0.9879	0.3858
Carbonate esters	1	0.6123
Carboxylic anhydrides	N/A	N/A
Peroxides	N/A	N/A
Hydroperoxides	N/A	N/A
Polyfunctionals	0.998	0.0715
Carboxylic acids	0.9972	0.7624
Mixture	0.9973	N/A
Overall R^2	0.9993	0.9953
Overall MAE	2.711	13.352
Overall RMSE	7.567	19.58

5.3.2 Transfer learning of 10 classical CNN architectures

10 classical CNN of AlexNet [386, 387], Densenet201 [376, 377], GoogleNet [388, 389], Inceptionv3 [378, 379], Mobilenetv2 [396, 397], ResNet18 [380, 381], Resnet50 [381, 382], Shufflenet [392-395], Squeezenet [398, 399], Xception [374, 375] are applied to predict YSI and compared with SDSeries38 network of this work. The max epochs of these 11 networks are set as 40 and the mini batch size varies depending on the available/required memory. The traditional transfer learning is to replace the final layers with new layers to learn the new dataset, then set a faster learning rate in the new layers than those in the transferred layers [404]. However, it is not adapted to the YSI prediction problem because these 10 classical CNN are developed for image classification. To make these CNNs fit the current problem, the CNN architectures are maintained unchanged but the learnable parameters (weights and bias) in the convolution layer and fully connected layer are trained from the scratch. The network training progress of the SDSeries38 network and 10 classical CNN are illustrated in Figure 5.7. The SDSeries38, Xception, Shufflenet, Resnet50, Resnet18, Mobilenetv2, Inception v3 obtain lower mini-batch size RMSE after 40 epochs training while other networks have little effect to reduce the RMSE. The predictive accuracy versus model training speed of these 11 networks are shown in Figure 5.8 and it indicates that the SDSeries38 network achiebes the the lowest RMSE and the shortest time elapsed per epoch among the studied networks. The model performance (based on predictive accuracy and model training speed) for YSI prediction orders from high to low as:

SDSeries38>ResNet18>ResNet50>Inceptionv3>Xception>Densenet201>Shufflenet≈GoogleNet≈Squeezenet≈AlexNet.
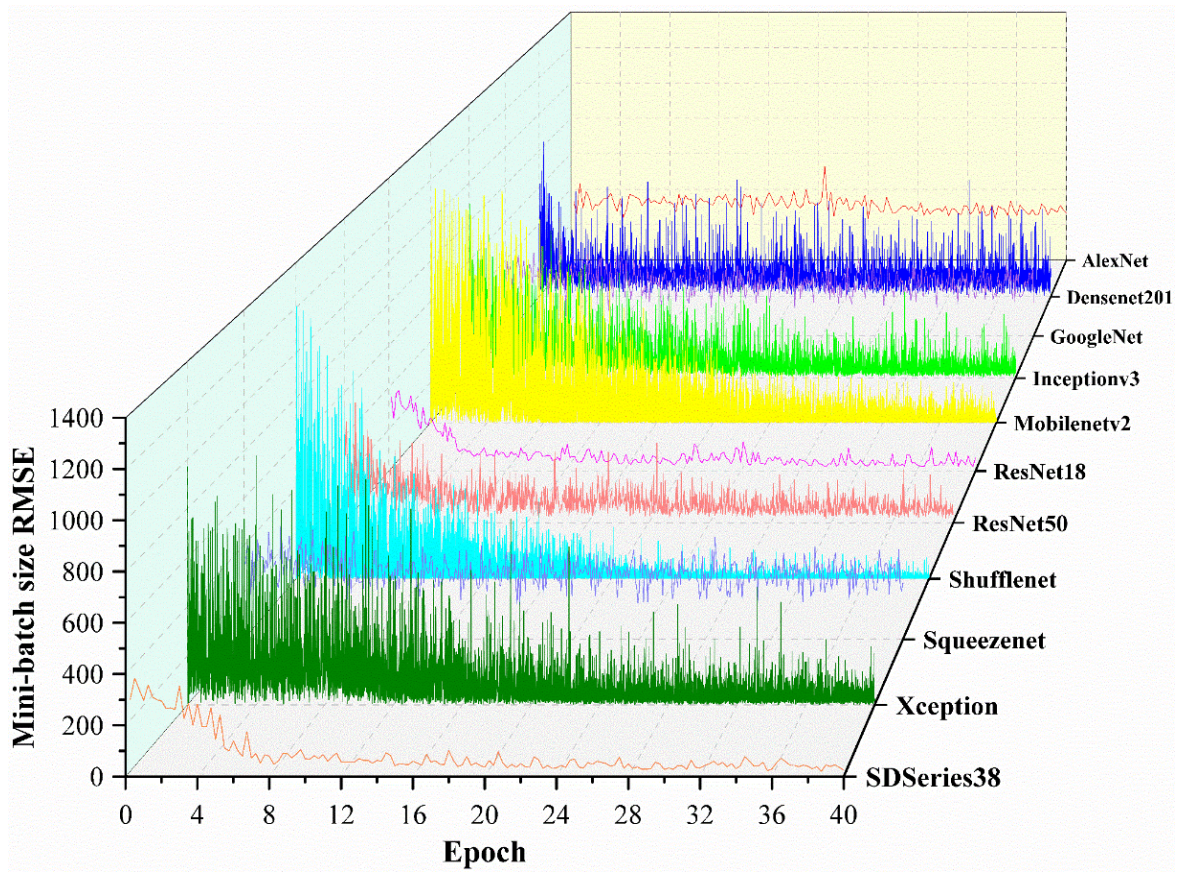
Figure 5.7. Network training progress of SDSeries38 network and 10 classical CNN architectures.
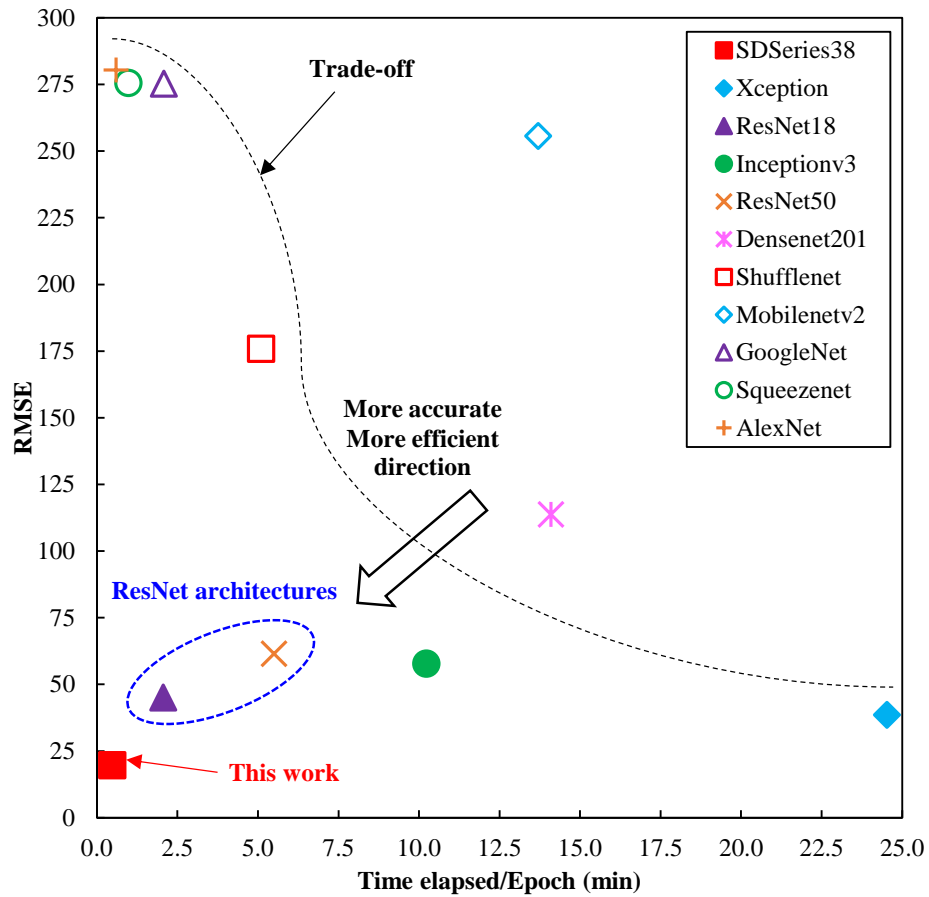
Figure 5.8. Predictive accuracy and model training speed of SDSeries38 network and 10 classical CNN architectures.

5.3.3 Challenge of application DL into regression problem

As discussed in section 5.3.1 and section 5.3.2, the predictive accuracy of DL-CNN is lower than the ML-QSPR method and the SDSeries38 network obtains better accuracy than 10 classical CNN. It is necessary to examine how to further improve the predictive accuracy of the DL roadmap by optimizing the CNN architectures. There are a few reasons that the 10 classical CNN do not obtain high YSI predictive accuracy: (1) These networks are originally designed for image classification rather than regression. Rare CNN architectures are specially designed for regression problem and it is still in its infancy. (2) Even for the image classification problem, their predictive accuracy are still far from perfect. The top-1 accuracy of theses networks ranks from highest to lowest as: Xception(79%) > Densenet201 (78.54%) > Inceptionv3 (78.2%) > ResNet50 (77.15%) > Shufflenet (73.7%) > ResNet18 (72.12%) > Mobilenetv2 (72%) > GoogleNet (68.7%) > AlexNet (59.3%) > Squeezenet (57.5%) as shown in Figure 5.9. Breakthrough technology is needed for CNN architecture in DL to substantially promote predictive accuracy. Top-1 accuracy and top-5 accuracy represent the fraction of test images for which the correct label is amongst the model's top-1 and top 5 predictions respectively [372, 373]. They are the commonly used evaluation index of image classification in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition [372, 373].

Even though image classification problem and regression problem are different types of problems, their evaluation indexes can represent the models' performances well and exist inherent relationships. The model with a lower top-1 error rate for image classification also obtains lower RMSE in YSI regression and the R^2 is 0.6333 as shown in Figure 5.10. According to this result, the CNN with higher top-1 accuracy is preferred when conducting transfer learning from image classification problem to regression problem. Given that the specially designed SDSeries38 network in this work outperforms 10 classical CNN networks in predictive accuracy and training speed for YSI prediction (see Figure 5.8). It indicates that tailor-made CNN is needed for typical regression problem to achieve better predictive accuracy. The established CNN are designed for image classification, object

detection, semantic segmentation, instance segmentation, biomedical image segmentation, human pose estimation, generative adversarial network, super resolution, image restoration, video frame interpolation, video coding, but rare of them are oriented to a regression problem. The deeper the network usually results in higher predictive accuracy (see Figure 5.9) but the error rate will increase once the number of layers exceeds a certain point [405] as shown in Figure 5.11 (a). Given that the layer numbers can not grow infinitely, some functional modules that have proven successful in image classification problem can be introduced into the tailor-made CNN for a regression problem. For example, residual learning module (in ResNet18/50/101 network [380-382]); inception module (in GoogleNet network [388, 389]); dense module (in Densenet201 network [376, 377]) etc and the new modules are required. So, the tailor-made CNN for regression problem should be a modular structure to accommodate the typical functional modules.
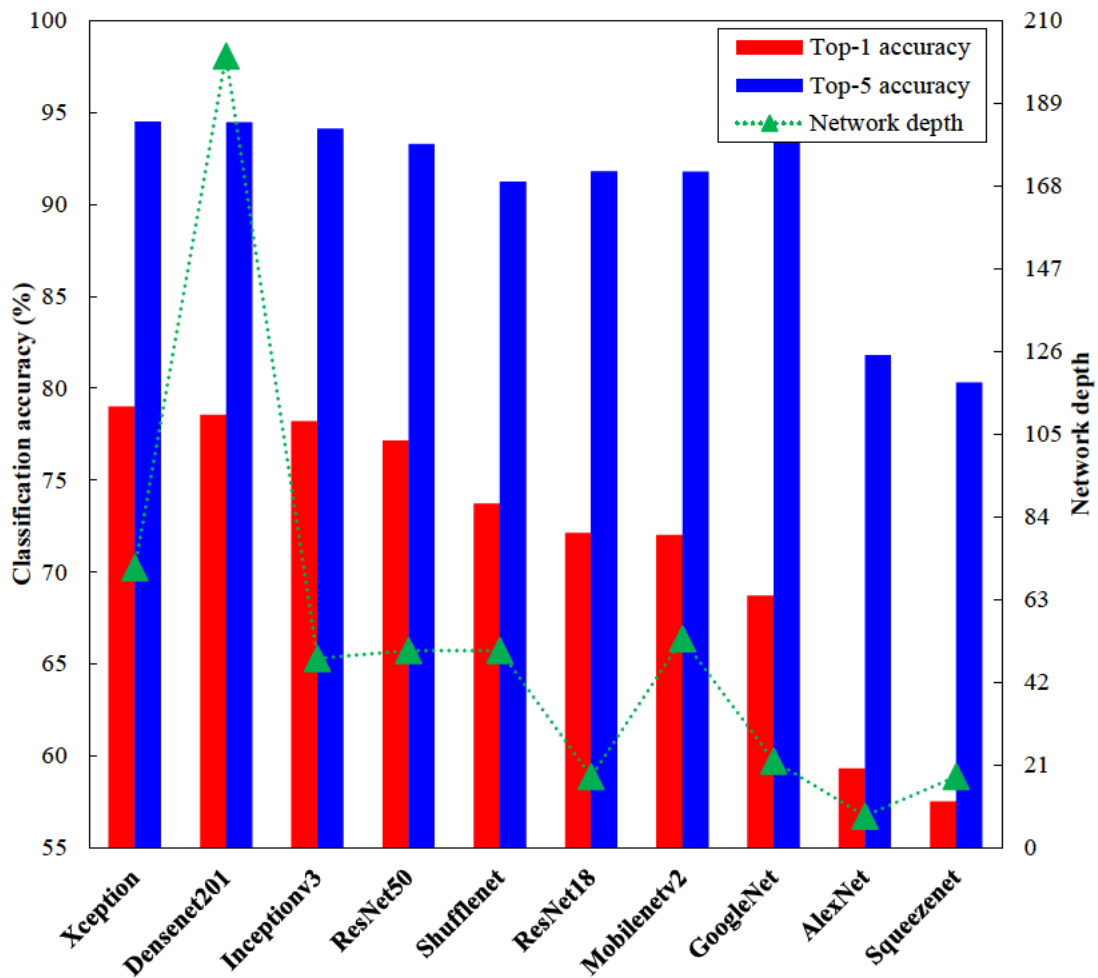
Figure 5.9. Accuracy and network depth of 10 classical CNN architectures for image classification training on ImageNet database [373].
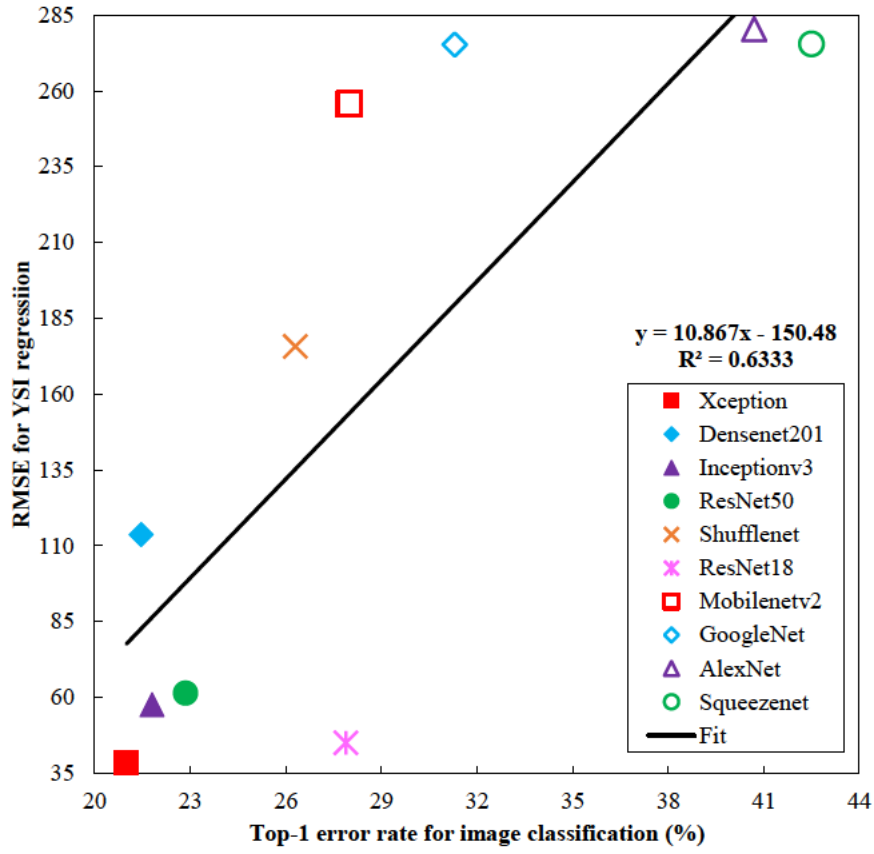
Figure 5.10. RMSE for YSI regression versus top-1 error rate for image classification of 10 classical CNN architectures.
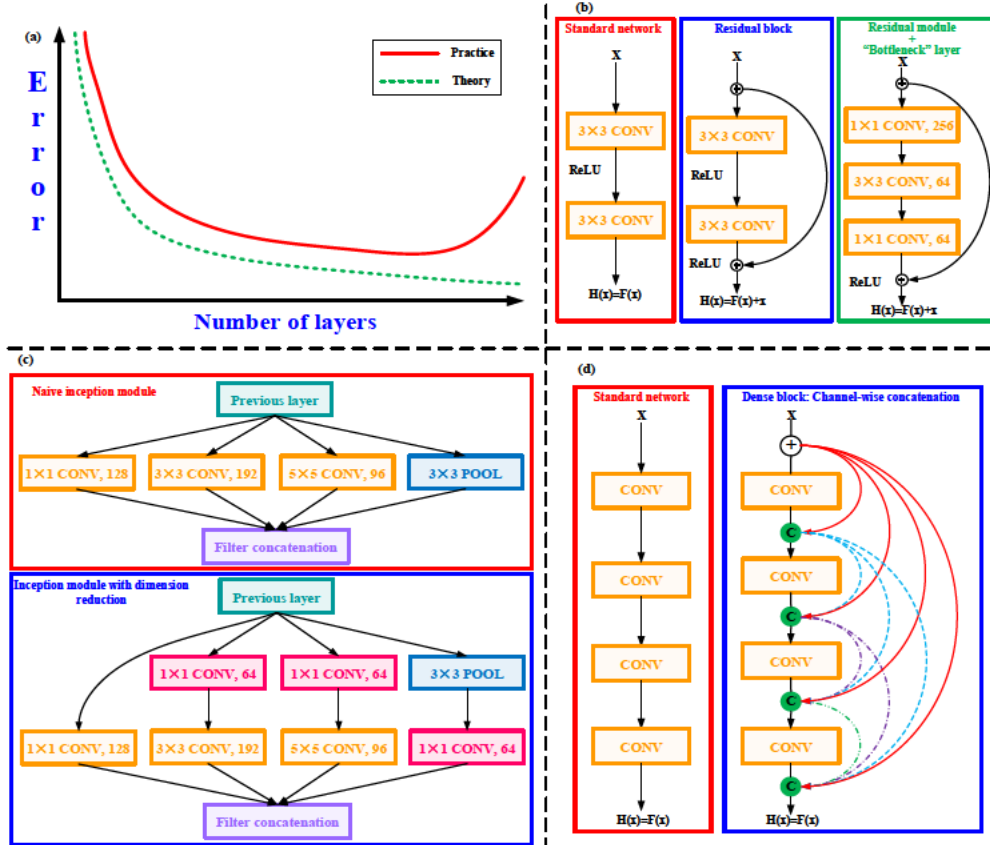
Figure 5.11. (a) Deep network performance versus number of layers; (b) residual learning module [380-382]; (c) inception module [378, 379]; (d) dense module [376, 377].

5.4 Conclusions

This work applies two technology roadmaps of ML and DL for YSI prediction and performs a systematic performance evaluation. The YSI dataset containing 444 pure compounds and 12 mixtures is set up for model training and validation. In the ML roadmap, the QSPR-UOB 3.0 functional classification system is developed to extract structural features and transforms them into a fuel molecular structure matrix. 19 ML algorithms are utilized to map the fuel molecular structure matrix and YSI matrix. Gaussian process regression obtains the best predictive accuracy and the R^2, RMSE are 0.9993 and 7.567 respectively. ML-QSPR model can apply to mixtures as well since the QSPR method decomposes the fuel molecules into component fragments. QSPR method eliminates the gap between pure compounds and mixtures. In the DL roadmap, a special design CNN of SDSeries38 is implemented for feature learning and regression. It is a standard series network architecture containing 1 image input layer, 9 feature learning modules and 1 regression module with 38 layers in total. The feature learning module is a stack of convolution layer, batch normalization layer, ReLU layer and max pooling layer and the regression module comprises a fully connected layer and a regression layer. The R^2 and RMSE of the DL-CNN model are 0.9953 and 19.58 respectively.

The ML-QSPR model achieves better predictive accuracy than the DL-CNN model of SDSeries38 and requires less computational resource for model training and prediction on new data. In the deep learning scope, the SDSeries38 network outperforms 10 classical in terms of predictive accuracy and model training/prediction speed. Most of the established CNN is developed for image classification and the network with higher top-1 accuracy for classification usually has better accuracy for a regression problem. Therefore, the network with higher top-1 accuracy is preferred when performing transfer learning from the image classification problem to the regression problem. Applying DL technology to regression problem and computer vision is in its infancy, some gaps need to be filled: (1) Tailor-made CNN architectures for regression problem are required to substantially improve model predictive capability. (2) Direct transfer learning the established CNN architectures from image classification to regression does not guarantee good performance and more innovative network architectures are required. (3) One of the specially-developed CNN for regression problem is to adopt a modular structure with typical functional modules such as residual module, inception module, dense module.

Chapter 6 High Throughput Fuel Screening by ML-QSPR and Chemical Kinetics for Internal Combustion Engines

6.1 Introduction

Substantive researches are carried out on the thermodynamic cycle/combustion mode optimization [254, 255, 257, 258] and clean-burning fuel components development (including petroleum-derived fuels and biofuels) [115, 259, 261, 264, 406]. Biofuels can improve the fuel quality, reduce the particulate emission and carbon footprint, more importantly, their molecular structure and hence the physicochemical properties provide an additional degree of freedom for property-oriented fuel design. A vast amount of biofuels are derived from biomass [407-410], but it is not feasible to identify the ideal molecules enabling efficient and clean combustion by performing fundamental combustion tests (shock tube, RCM, JSR, PFR, CVV) and engine experiments for all these compounds. Engine experiments are time-consuming and require a significant number of samples, so it is not practical to adopt engine tests for fuel screening especially for those expensive compounds.

The property-oriented fuel design concept integrates the fuel combustion/emission and fuel synthesis into consideration [268, 269]. It adopts engine combustion and emission requirements as design rules to determine the desirable physicochemical properties, and hence to identify the promising blendstock candidates from both fossil and bio-feedstock. Therefore, an efficient and precise model-based fuel screening approach is necessary to discover viable high-performance fuels for internal combustion engines. The recent progress of fuel screening approaches is summarized in Table 6.1 and these methods can broadly divide into two categories: (1) establishing a fuel property database by compiling the measured fuel property data; (2) constructing fuel property predictive models to compute the missing property data.

To realize the efficient and clean combustion process in the IC engine, the desirable fuel physicochemical properties should be clarified. The major physicochemical properties relevant to engine combustion/emission performance include 5 categories, 16 parameters: (1) volatility specification: melting point T_m, boiling point T_b,

vapor pressure (VP), enthalpy of vaporization ΔH_{vap}; (2) atomization specification: surface tension γ, kinematic viscosity ν; (3) energy density: lower heating value (LHV), liquid density ρ; (4) sooting tendency: yield sooting index (YSI); (5) ignitability: CN, RON, MON, ignition temperature (IT), flash point (FP), lower/upper flammability limits (LFL/UFL). The forecasting approaches of these 16 physicochemical properties by computer-aided molecular design (CAMD) technique are summarized in Table 3.3 and Table 4.1. QSPR and GCM (group contribution method) are the mainstream technologies for structural feature extraction. The regression models are developed by supervised machine learning algorithms including multiple linear regression (MLR), artificial neural networks (ANN), support vector machine (SVM), etc. A wide variety of QSPR/GCM schemes are developed to predict these 16 parameters, but none of them can apply to these 16 parameters simultaneously. Researchers have to apply different QSPR or GCM schemes to specific properties which cause additional difficulty to high-throughput (multiple fuel types and properties) fuel screening [411].

This work proposes a high-throughput fuel screening approach coupling quantitative structure property relationship (QSPR) and machine learning (ML) to identify the suitable fuel blendstocks with desirable properties enabling efficient and clean combustion. The fuel screening approach composes of Tier 1 fuel physicochemical property screening and Tier 2 chemical kinetic screening. The paper is structured as follows: Section 1 provides an overview of the objective, challenges and methods of fuel screening for internal combustion engine. Section 2 proposes the fuel screening approach by coupling ML-QSPR and chemical kinetics which consists of Tier 1 fuel physicochemical property screening and Tier 2 chemical kinetic screening. Two case studies showcase the fuel screening for spark ignition (SI) and compression ignition (CI) engines in section 3.1 and section 3.2 respectively. Section 3.3 presents the CI engine test results of the selected fuel components recommended by the fuel screening result. Section 4 presents the concluding remarks.

Table 6.1. Overview of fuel screening approaches and applications

Engine	Project	No. of compounds	Target properties	Chemical families	Method	Ref.
SI	Co-Optima	470	T_m, T_b, solubility, corrosivity, toxicity, safety, biodegradation, RON	n-Paraffins, iso-paraffins, olefins, aromatics, naphthenes, alcohols, ketones, fatty acid, esters, furans, ethers, multi-ring aromatics, aldehydes, fatty esters, carboxylic acids	Fuel properties database [230]	[266]
	Co-Optima	369	T_m, T_b, RON, MON, ΔH_{vap}, LHV, ρ, solubility, safety, biodegradation	Alkanes, alcohols, alkenes, esters, ethers, furans, ketones, polyfunctionals	Fuel properties database [230]	[412]
	Co-Optima	41	RON, OS	PRF, TPRF, and PRF40−ethanol blend	Gaussian process classifier	[413]
	TMFB	3,215	oxygen content, T_m, T_b, γ, ν, ΔH_{vap}, DCN	n-Paraffins, iso-paraffins, olefins, aromatics, naphthenes, alcohols, ethers, esters, ketones, aldehydes, and polyfunctionals	QSPR model	[191]
CI	Co-Optima	456	T_m, T_b, FP, T_{cloud}, lubricity, viscosity, conductivity, oxidation stability, solubility, toxicity, safety, corrosivity, biodegradation, CN, LHV, YSI	n-Paraffins, iso-paraffins, olefins, aromatics, naphthenes, alcohols, ketones, fatty acid, esters, furans, ethers, multi-ring Aromatics, aldehydes, fatty esters, carboxylic acids	Fuel properties database [230]	[414, 415]
	Co-Optima	36	T_m, T_b, FP, T_{cloud}, LHV, ν, ICN, water-solubility, YSI	Alkanes, ethers, alcohols, ketones, enones, esters	EPI Suite	[416]
	TMFB	3215	oxygen content, T_m, T_b, γ, ν, ΔH_{vap}, DCN	n-Paraffins, iso-paraffins, olefins, aromatics, naphthenes, alcohols, ethers, esters, ketones, aldehydes, and polyfunctionals	QSPR model	[191]
General[α]	Co-Optima	1977	RON, TSI, T_m	Hydrocarbons/oxygenates	BioCompoundM software	[417]
	NKRDPC	319895	ρ, T_m, T_b, T_c, P_c, V_c, ΔH_{vap}, ΔH_{fus}, LHV, specific impulse	Saturated hydrocarbon	Single-layer neural network	[418]

[α] General-purpose fuel property characterization tool.

6.2 Modeling approach

6.2.1 Methodological overview

To identify promising fuel blendstock candidates with desired properties enabling efficient and clean combustion, this work develops a high-throughput fuel screening approach for internal combustion engines by machine learning-quantitative structure property relationship (ML-QSPR) and chemical kinetics. The term "high throughout" means the proposed approach applies to multiple fuel types and fuel properties. The general scheme of the model-based fuel screening for internal combustion engines is shown in Figure 6.1. The high throughput fuel screening consists of Tier 1 physicochemical properties screening and Tier 2 chemical kinetic screening. Tier 1 and Tier 2 screenings aim at discovering the novel fuel candidates that meet the basic specification requirements and significantly lessening the workload of combustion research. The prioritized fuel candidates are then becoming the target molecules of the synthetic pathway analysis and catalytic production route design. Both fundamental combustion experiments (CVV, JSR, shock tube, RCM) and engine experiments examine the combustion and emission performance of the selected components. The fundamental combustion experiments provide valuable data for chemical kinetic mechanisms development, verification and understanding of the fuel reactivity, autoignition propensity, and reaction pathways. Engine experiments at diverse combustion modes and operating conditions enable to verify the potential of the chosen fuel molecules to achieve ultra-low pollutant emissions. The fundamental combustion tests and engine tests enrich the laboratory collections and literature data and it further nurtures the development of ML-QSPR predictive models.

Tier 1 screening examines the fuel physicochemical properties of melting point T_m, boiling point T_b, vapor pressure (VP), enthalpy of vaporization ΔH_{vap}, cetane number (CN), research octane number (RON), motor octane number (MON), ignition temperature (IT), flash point (FP), yield sooting index (YSI), liquid density (ρ), lower heating value (LHV), surface tension γ, lower/upper flammability limit (LFL/UFL). These properties are computed by the corresponding ML-QSPR predictive models. QSPR theory assumes that fuel physicochemical

properties are a sum result of the fuel molecular structure, hence, the compounds with similar molecular descriptors exhibit similar physicochemical properties. An ungraded QSPR-UOB 3.0 functional classification system is adopted to decompose molecules into structural fragments and record as fuel molecular structure matrix which is improved from QSPR-UOB 2.0 system [357]. ML algorithms are used to train the 15 predictive models correlating the fuel molecular structure matrix and fuel property matrix. UOB Fuel Property Database containing 23 chemical families, 1742 pure compounds and 465 mixtures is established for model training. Its data mainly derive from Co-Optimization of Fuels & Engines: fuel properties database [230], compendium of experimental cetane numbers [193], octane and cetane number data tabulation [208, 209], API data book [231], American petroleum institute research project 45 [232], yield sooting index database [321], CRC handbook of chemistry and physics [358]. The fuel properties screening criteria for SI and CI engines are set up to distinguish the molecules with desirable properties. Tier 2 fuel chemical kinetic screening examines ignition delay time, φ -sensitivity, laminar flame speed which represent the fuel reactivity, stratified charge sequential autoignition capacity, combustion rate and dilution ratio.
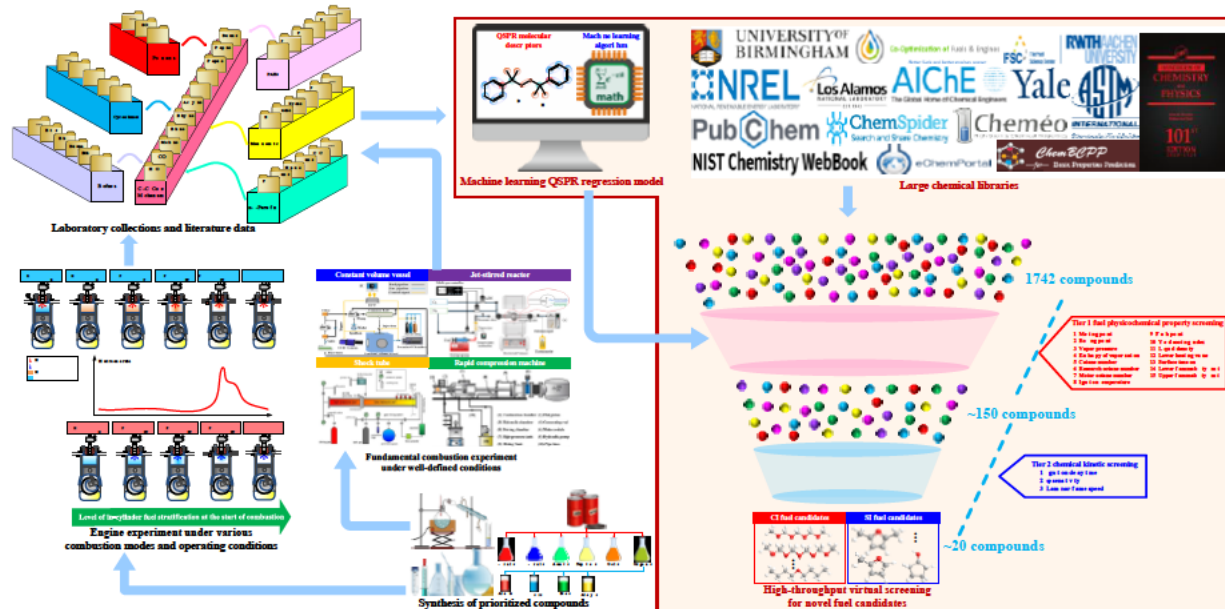
Figure 6.1. The workflow of virtual fuel screening by ML-QSPR and chemical kinetics.

6.2.2 Tier 1 fuel physicochemical property screening and Tier 2 chemical kinetic screening

The fuel screening approach composes of Tier 1 fuel physicochemical property screening and Tier 2 chemical kinetic screening. Tier 1 screening contains 5 categories: (1) volatility; (2) atomization; (3) energy density; (4) sooting propensity; (5) ignitability. The volatility is described by the T_m, T_b and ΔH_{vap}. The atomization performance is depicted by surface tension γ and dynamic viscosity μ, but the latter parameter does not participate in the current screening process. The 42 molecular descriptors of QSPR-UOB 3.0 system cannot fully capture the principle component of dynamic viscosity ($R^2<0.6$). The energy density is considered by LHV and ρ while the sooting propensity is evaluated by YSI. The ignitability is characterized by RON, OS for SI engines and CN for CI engines respectively. In Tier 1 screening, these 15 fuel properties are calculated by the ML-QSPR predictive models. The fuel specifications requirements recommended by Co-Optima and TMFB projects are summarized in Table 6.2.

Tier 2 screening consists of 3 parameters: (1) ignition delay times τ_{ig}; (2) φ-sensitivity; (3) laminar flame speed S_L. Boosted autoignition (knock) condition for SI engines is represented by φ=1.0, T_{init}=750~850K, P_{init}=50bar [419] which is also beyond RON (K <0) condition [420, 421]. The advanced autoignition condition for CI engines is represented by φ=0.3, T_{init}=1050~1200K, P_{init}=25bar [419]. The longer the ignition delay times, the better knock resistance performance in SI engines while the shorter the ignition delay times, the more readily autoignition in CI engines. φ-sensitivity quantifies the variation in how readily a fuel-air mixture autoignites as the equivalence ratio φ altering and the definition is shown in Eq. (1) [422]. The change of ignition delay times with the equivalence ratio allows creating higher φ-gradient within the combustion chamber for sequential autoignition. The regions with higher fuel concentration autoignite faster than the fuel-lean regions. The sequential autoignition can expand the engine speed-load range of kinetic controlled combustion by reducing the peak combustion pressure and pressure rise rate. Higher φ-sensitivity fuels can tune ignition timing via charge stratification and benefit the engine efficiency and noise. The laminar flame speed influences the premixed combustion properties (fuel-air mixture

reactivity, combustion rate, dilution tolerance) which is critical to both boosted SI combustion and advanced CI combustion strategies. The higher the laminar flame speed, the greater the combustion rate and dilution tolerance.

$$\eta = -\left.\frac{d\,log\left(\tau_{ig}\right)}{d\,log\left(\varphi\right)}\right|_{T,P} \tag{1}$$

where η, τ_{ig}, φ denote φ-sensitivity, ignition delay time and equivalence ratio respectively.

Table 6.2. Fuel specifications requirements recommended by Co-Optima and TMFB projects

Specifications	Fuel blendstocks for SI engine			Fuel blendstocks for CI engine	
	Co-Optima [266, 413]	TMFB$^\alpha$ [191]	TMFB$^\beta$ [191]	Co-Optima [414, 415, 423]	TMFB [191]
Oxygen content (wt.%)	N/A	≥10	≥10	N/A	≥10
T_b (°C)	20~165	60~120	60~150	<338	60~250
T_m (°C)	<-10	≤-20	≤0	<0	≤-20
T_{cloud} (°C)	N/A	N/A	N/A	<0	N/A
γ (mN/m)	N/A	≤30	≤35	N/A	≤30
ν @40°C (mm²/s)	N/A	≤2	≤5	1.9~4.1	≤4.5
ΔH_{vap} (kJ/kg)	N/A	≤60	≤60	N/A	N/A
FP (°C)	N/A	N/A	N/A	>52	N/A
Lubricity$^\delta$ (μm)	N/A	N/A	N/A	≤520	N/A
CN/DCN	N/A	≤20	≤25	≥40	≥40
RON	≥98	N/A	N/A	N/A	N/A
OS	≥8	N/A	N/A	N/A	N/A
LHV (MJ/kg)	N/A	N/A	N/A	>25	≥30
YSI	N/A	N/A	N/A	<200	N/A
ρ (kg/m³)	N/A	N/A	N/A	N/A	700

$^\alpha$ Pure-component biofuel candidates for SI engine; $^\beta$ Blending biofuel candidates for SI engine; $^\delta$ High Frequency Reciprocating Rig (HFRR) test by ASTM D975.

6.3 Results and discussion

Two case studies showcase the fuel blendstocks screening for SI and CI engines in section 6.3.1 and section 6.3.2 respectively. The ideal components for SI engines should have high antiknock property, low sooting tendency, good stratified charge combustion capability, high combustion rate and dilution tolerance. The ideal components for CI engines should have high reactivity and energy density; low sooting tendency, reasonable liquid breakup, atomization and vaporization performance.

6.3.1 Case study 1: fuel screening for SI engine

6.3.1.1 Tier 1 fuel physicochemical property screening for SI engine

Tier 1 fuel physicochemical property screening for SI engine consists of 5 categories, 10 properties filters and the flowchart is shown in Figure 6.2. The specifications constraints refer to Co-Optima [266, 413] and TMFB [191] projects which are compared in Table 6.2. Melting point constraint is equal to or less than -10°C to improve low-temperature fluidity and avoid low-temperature condensation. The lower and upper boundaries of the boiling point are 60°C and 150°C to ensure volatility at low temperatures. The enthalpy of vaporization is defined as the enthalpy change of one-mole liquid converting into gas phase at 25°C, 101 325 kPa and the upper boundary is constrained to 60kJ/mol. The surface tension and dynamic viscosity are restricted to 35mN/m and 1.5mPa·s which facilitate liquid droplet atomization and fuel-air mixing process to inhibit soot formation. Lower bounds are placed on the LHV ($\geq$1200kJ/mol, equivalent to ethanol of 1235kJ/mol) and liquid density (650 kg/m^3, equivalent to iso-octane of 687.8kg/m^3) to attain reasonable gravimetric energy densities. An upper bond of unified YSI is set as equal or less than 70 to limit the molecule sooting tendency and the blendstock candidates should be free of aromatic to meet this requirement. Fuel molecules with high autoignition resistance are desired to suppress end gas autoignition and stochastic knock in the combustion chamber. Given that the κ constant is a negative value for the modern boosted SI engines, the higher the RON and OS result in higher octane index (OI) [253]. To some extent, RON and OI (OI=RON-κ×OS) represent the lower and upper ends of the anti-knock capability, thus the

RON and OS should be equal or greater than 100 and 6 for high-performance SI fuels.

Tier 1 fuel screening for SI engine starts from 1074 chemical compounds in the UOB property database, the number of compounds satisfying the property constraint and those failing to satisfy the property requirements are demonstrated in solid and dotted line arrows respectively. The boiling point, melting point and RON are the three main property filters in this work as shown in Figure 6.2. 13 fuel types, 166 compounds are acquired from the Tier 1 fuel screening for SI engine as shown in Figure 6.3 (a) and the complete list is provided in the author's publication [424]. Saturated esters, acyclic ethers, unsaturated esters, polyfunctionals, alcohols account for 79.52% of Tier 1 fuel screening candidates. The remaining share comprises iso-alkanes, alkenes, cycloalkenes, aldehydes, furans, cycloalcohols, cycloketones. RON versus YSI of these compounds is plotted in Figure 6.3 (b) to identify those fuel types having strong autoignition resistance and low sooting tendency simultaneously. The saturated and unsaturated esters in the 166 compounds possess excellent antiknock performance and low sooting tendency. The represented candidates of these 13 fuel types are presented in Table 6.3 and they have the common characteristics of highly compact, branched molecular structures.

Hierarchical clustering analysis is implemented on the functional groups and molecules to identify the key functional groups of Tier 1 fuel screening candidates (166 compounds) for the SI engine and the heatmap is shown in Figure 6.4. A high-resolution, interactive version of the heatmap ("Tier1 SI fuel heatmap.json" file) with zooming capability is provided in author's publication [424]. Functional group types of 8~12 (see Figure 4.2) are the dominant factors for all these compounds and they are also the fuel reactivity descriptors in QSPR-UOB 3.0 scheme. Reducing the numbers of functional group types of 9, 10 and 11, 12 is to shorten the length of the straight-chain carbon backbone and the ring size of cyclic molecules respectively. Increasing the number of functional group type 8 is to increase the number of the branched-chain methyl group. Figure 6.4 also indicates that the acyclic ester group, aldehyde group and acyclic ester group (functional group type 16, 20, 21) benefit autoignition resistance and sooting tendency reduction. CCDB in non-aromatic, non-ring structure (e.g. alkenes) and cyclic

structure (e.g. furans) can improve the autoignition resistance but the side effect is deteriorating sooting emission.

The design rules of gasoline blendstock for boosted SI engine on the molecular level are summarized as: (1) constructing highly compact, branched molecular structure (necessary precondition); (2) introducing oxygen-containing functional groups, for example, ester group, ether group, hydroxyl group, ketone group, aldehyde group; (3) introducing unsaturated structure (e.g. CCDB) or cyclic structure (e.g. cycloalkenes, furans, cycloalcohols, cycloketones) and moving toward the central position of the molecule; (4) being free from the aromatic bond to reduce soot emissions.
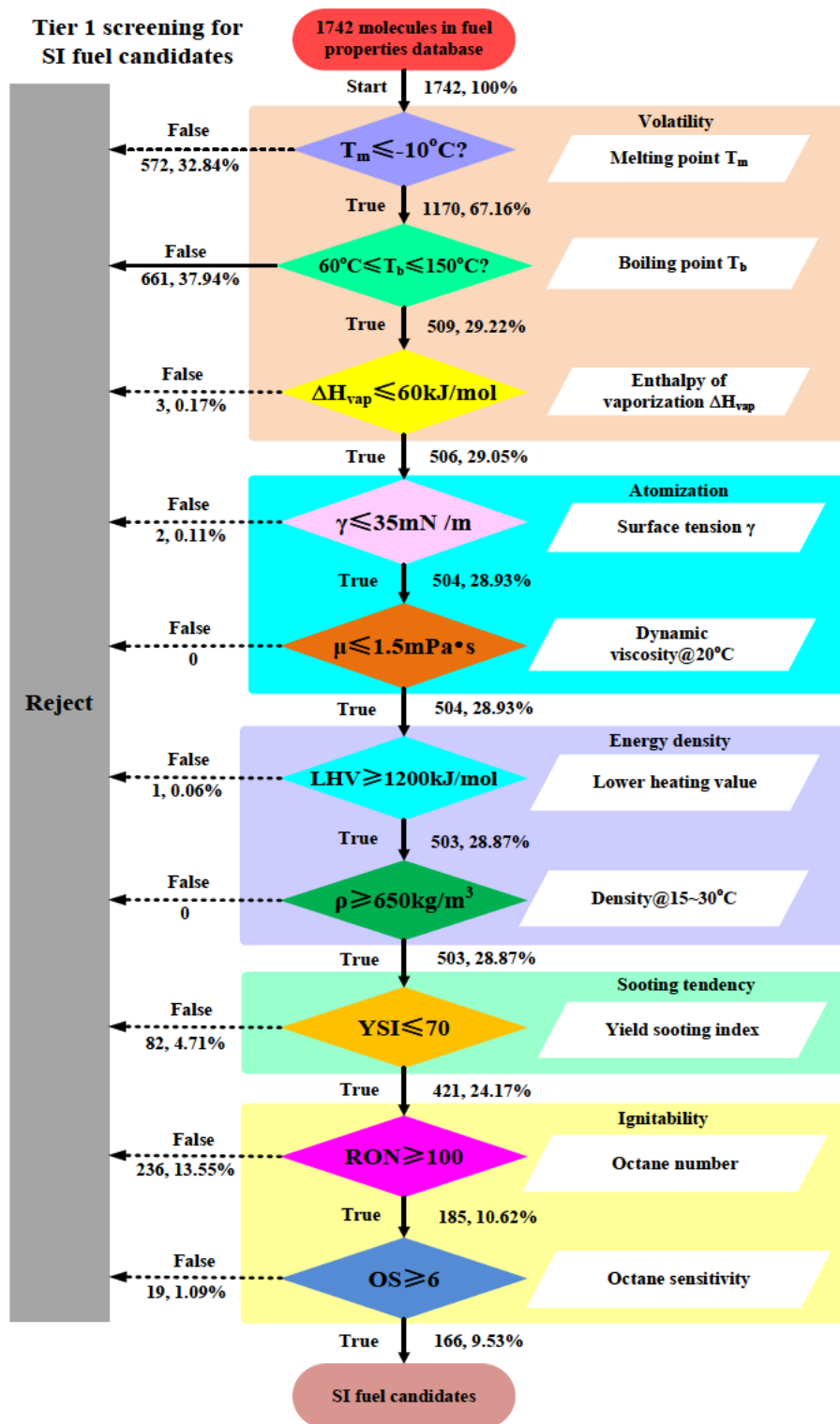
Figure 6.2. Tier 1 fuel physicochemical property screening for SI engine by ML-QSPR models.

Figure 6.3. Tier 1 fuel screening candidates for SI engine (166 compounds): (a) proportion of 13 fuel types; (b) RON versus YSI.

Figure 6.4. Hierarchical clustering of Tier 1 fuel screening candidates for SI engines (166 compounds), a high-resolution, interactive version of the heatmap ("Tier1 SI fuel heatmap.json" file) with zooming capability is provided in author's publication [424].
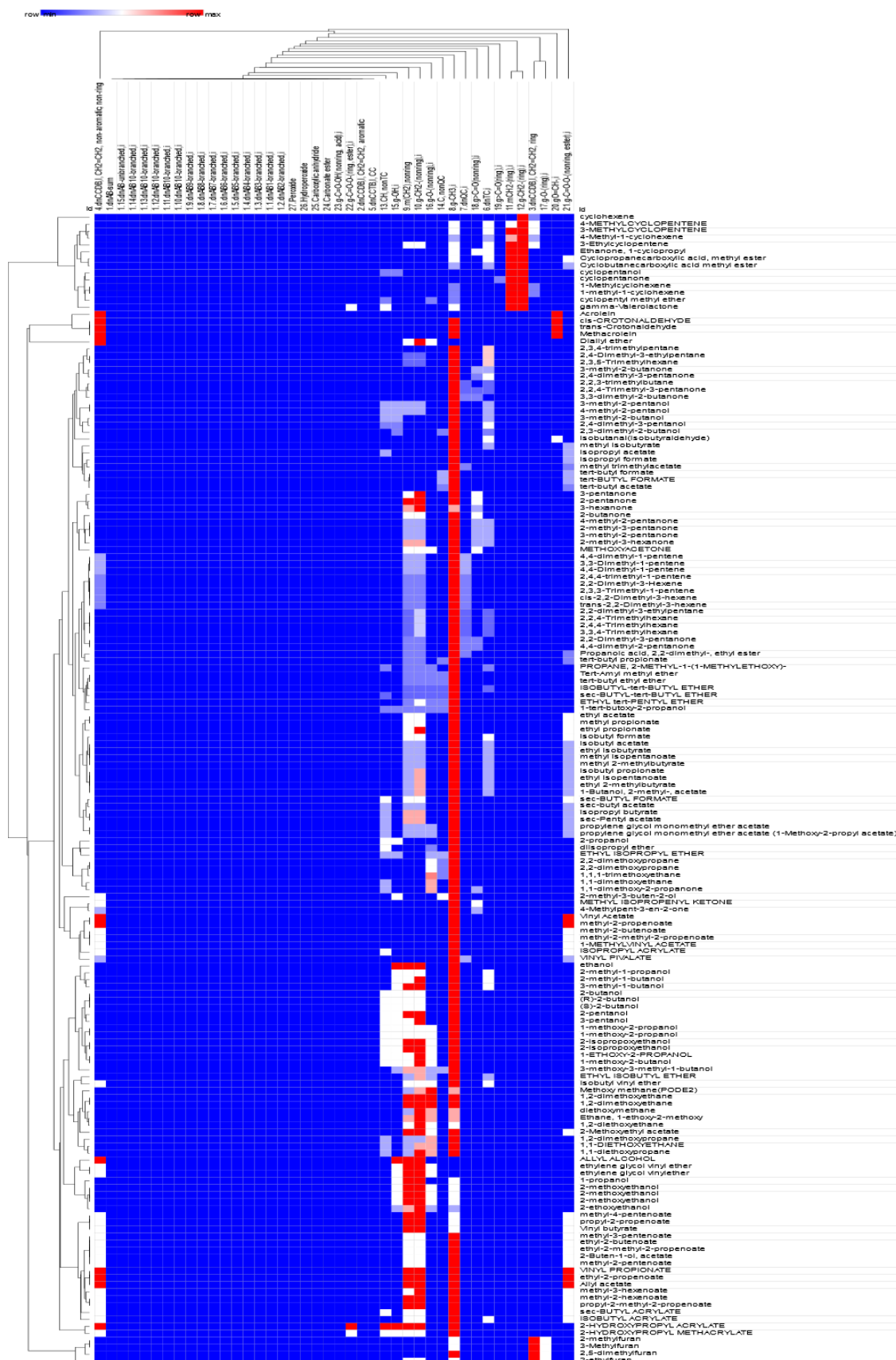
Table 6.3. Represented candidates of Tier 1 fuel screening for SI engines

Fuel candidates	Structure	Fuel type	CAS	RON	OS	YSI	Mechanism
2,2,3-Trimethylbutane		iso-Alkane	464-06-2	112.1	10.8	55.3	[425]
Diisobutylene		Alkene	107-39-1	106	19.5	68.5	[138, 426]
Cyclohexene		Cycloalkene	110-83-8	120.00*	37.50*	61.66	[427, 428]
Ethanol		Alcohol	64-17-5	111	15	10.3	[429]
Cyclopentanol		Cycloalcohol	96-41-3	103.15*	17.84*	36.01*	[430, 431]
iso-Butanal		Aldehyde	78-84-2	103.98*	9.75*	21.2	[432]
2-Butanone		Ketone	78-93-3	118	22.14*	17.6	[433, 434]
Cyclopentanone		Cycloketone	120-92-3	101	12	22	[435]
Methyl trimethylacetate		Saturated ester	598-98-1	123.97*	20.92*	30.8	N/A
Vinyl pivalate		Unsaturated ester	3377-92-2	123.24*	21.17*	47.85*	N/A
2,2-Dimethoxypropane		Acyclic ether	77-76-9	117.13*	26.06*	16.4	N/A
2,5-Dimethylfuran		Furan	625-86-5	119	30.63*	54.76	[436]
1,1-Dimethoxy-2-propanone		Polyfunctional	6342-56-9	116.78*	28.94*	14.8	N/A

* Predictive values calculated by the machine learning QSPR models in this work, experimental data not available.

6.3.1.2 Tier 2 chemical kinetic screening for SI engine

Tier 2 chemical kinetic screening examines 3 parameters: (1) ignition delay times τ_{ig}; (2) φ-sensitivity; (3) laminar flame speed S_L. It requires the detailed chemical kinetic mechanisms (including chemical kinetic mechanism, thermodynamic parameters, transport parameters) of the studied components. Ethanol, 2,5-dimethylfuran, 2-methylfuran cyclopentanone, diisobutylene, cyclopentanol are selected from the 166 fuel blendstock candidates. Toluene and iso-octane are also incorporated for comparison since they are the most common constituents of gasoline surrogates [111, 117]. The physicochemical properties and mechanism sources of these 8 candidates are provided in Table 6.4 and Table 6.5 respectively. Essential modifications (e.g. complement the missing thermodynamic/transport parameters, correct misnamed substances, adjust mechanism format to accommodate Chemkin Pro software) on the raw mechanisms are implemented and the modified mechanisms are provided in Supplementary Material.

The ignition delay time is test under φ=1.0, T_{init}=750~850K, P_{init}=50bar which represents the boosted autoignition (knock) condition for SI engine [419] as shown in Figure 6.5. The ignition delay times at 800K order from longest to shortest as: toluene (325.08ms) > ethanol (35.12ms) > 2,5-dimethylfuran (21.70ms) > cyclopentanone (20.52ms) > cyclopentanol (13.21ms) > diisobutylene (8.23ms) > iso-octane (4.49ms). No ignition is detected for 2-methylfuran at 800K and it autoignites at 860K.

The φ-sensitivity is studied under (a) φ=0.5, T_{init}=600~1400K, P_{init}=20bar; (b) φ=1.0, T_{init}=600~1400K, P_{init}=20bar; (c) φ=0.5, T_{init}=600~1400K, P_{init}=50bar; (d) φ=1.0, T_{init}=600~1400K, P_{init}=50bar (as shown in Figure 6.6) which cover the knock onset conditions of boosted SI engine. Cyclopentanol, iso-octane and diisobutylene achieve the greatest φ-sensitivity among the 8 selected fuel candidates. As the mixture initial temperature increases at the range of 1000K~1400K, the φ-sensitivity of iso-octane and diisobutylene decrease significantly from 0.62 and 0.74 to 0 while that of cyclopentanol stabilizes at the range 0.418~0.606. Cyclopentanol and cyclopentanone exhibit superior φ-sensitivity at a wide temperature regime. As the mixture initial pressure increases to 50bar, the

peak φ-sensitivity of cyclopentanol, iso-octane and diisobutylene decrease to 1.294 at 720K, 1.235 at 800K and 1.261 at 840K respectively. On the opposite, φ-sensitivity of cyclopentanone at P_{init}=50bar is greater than those at P_{init}=20bar. Overall, some conditions need to be addressed to organize the stratification combustion for boosted SI engine: (1) Construction of low-temperature combustion environment (e.g. EGR) is required as the φ-sensitivity peaks at the range of 640K~840K and it decreases significantly as increasing temperature. Cyclopentanol and cyclopentanone demonstrate good temperature adaptability on φ-sensitivity over a wide range of temperatures. (2) Appropriate boost pressure should be implemented since the φ-sensitivity usually decreases as increasing pressure. Cyclopentanone inhibits good pressure adaptability so it may partially offset the φ-sensitivity deterioration in boosted SI engine as gasoline additives.

The laminar flame speed is tested under φ=0.4~1.7, T_{init}=428K, P_{init}=1bar as shown in Figure 6.7 and the greater the flame speed representing higher combustion rate and dilution tolerance. The laminar flame speeds order from highest to lowest as: ethanol (85.46cm/s) > 2-methylfuran (83.45cm/s) > 2,5-dimethylfuran (70.90cm/s) > cyclopentanol (70.88cm/s) > toluene (70.52cm/s) > diisobutylene (69.45cm/s) > iso-octane (64.13cm/s).
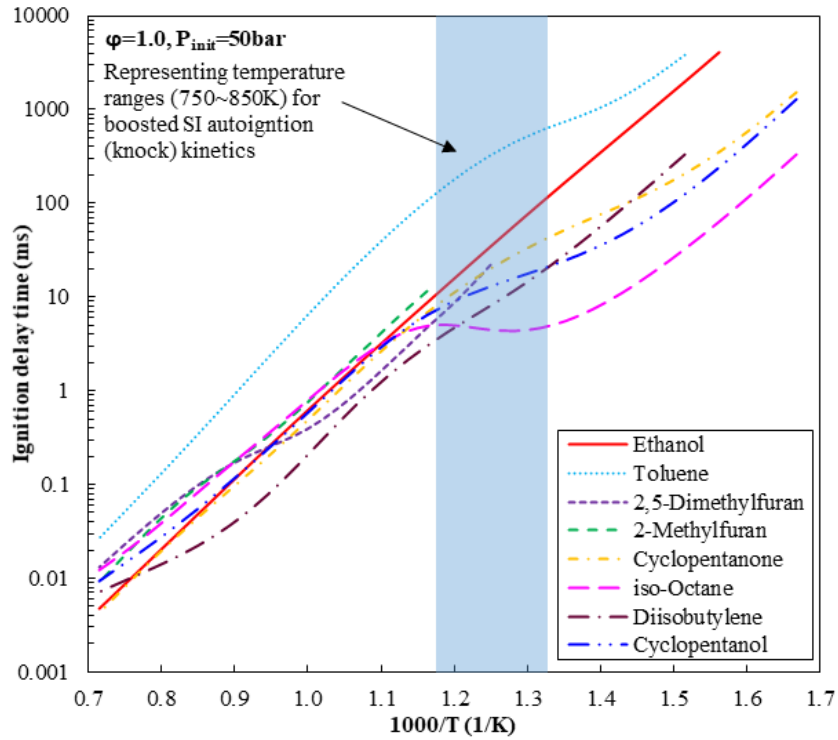
Figure 6.5. Ignition delay times of fuel (candidates for SI engines)-air mixture at φ =1.0, T_{init}=600~1400K, P_{init}=50bar.
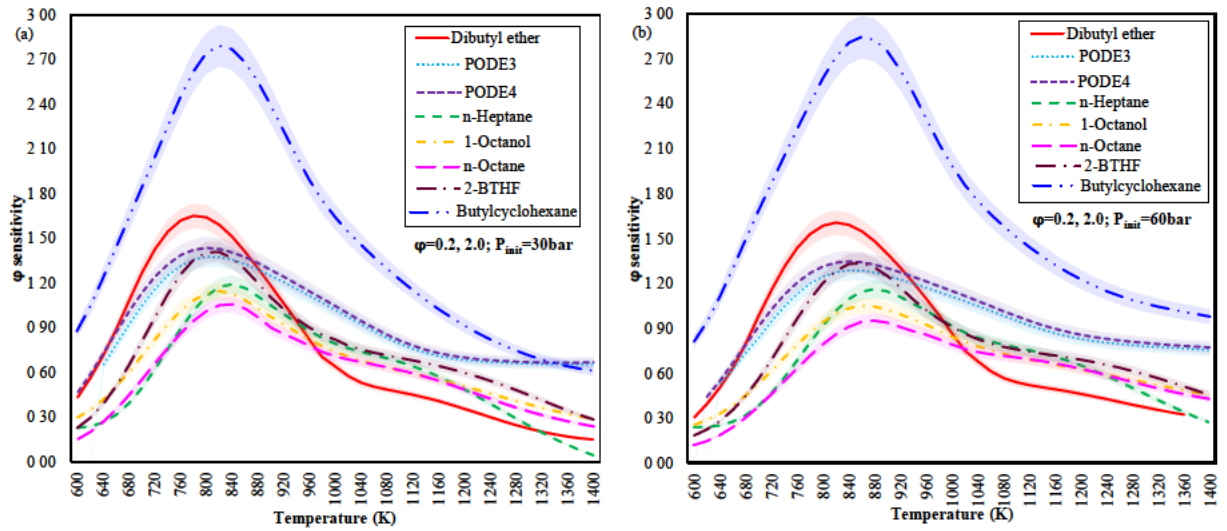
Figure 6.6. φ-sensitivity of fuel (candidates for SI engine)-air mixture at φ =0.5, 1.0, T_{init}=600~1400K, (a) P_{init}=20bar, (b) P_{init}=50bar; shaded region representing a 95% confidence interval.
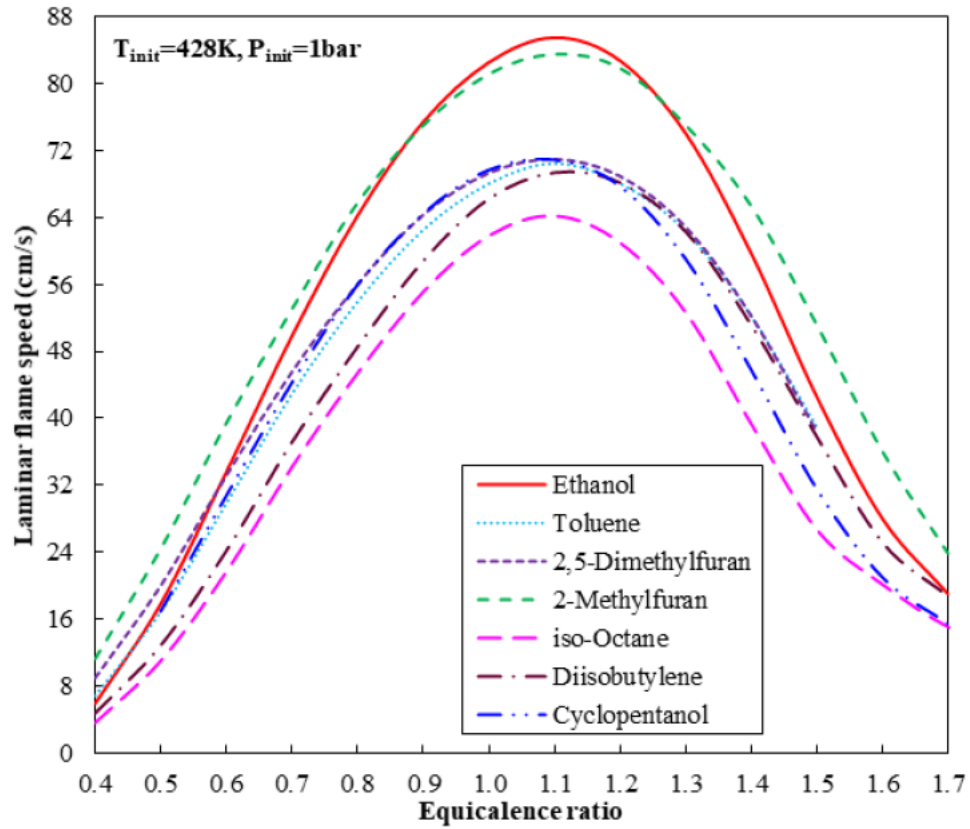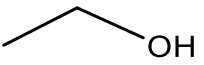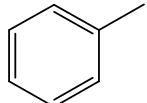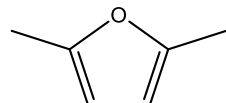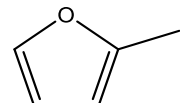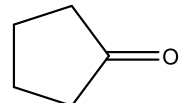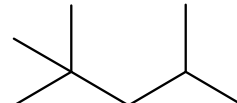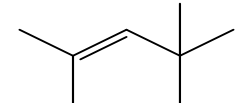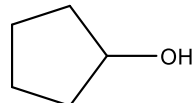
Figure 6.7. Laminar flame speed of fuel (candidates for SI engine)-air mixture at φ =0.4~1.7, T_{init} =428K, P_{init} =1bar.

Table 6.4. Physicochemical properties of 8 selected fuel blendstock candidates for SI engine

Fuel candidates	Ethanol	Toluene	2,5-Dimethylfuran	2-Methylfuran	Cyclopentanone	iso-Octane	Diisobutylene	Cyclopentanol
Structure								
Fuel type	Alcohol	Aromatic	Furan	Furan	Cycloketone	iso-Alkane	Alkene	Cycloalcohol
CAS	64-17-5	108-88-3	625-86-5	534-22-5	120-92-3	540-84-1	107-39-1	96-41-3
RON	111	120	119	103	101	100	106	103.15*
MON	96	104	86.13*	86	89	100	86.5	85.31*
OS	15	16	32.87	17	12	0	19.5	17.84
T_m (ºC)	-114.14	-95	-62.8	-91.2	-51.7	-107.36	-93.7	-17
T_b (ºC)	78.24	110.6	96	63.9	130.5	99.2	101.3	140.4
ΔH_{vap} (kJ/mol)	42.32	38.01	40.75*	36.48*	42.72	35.14	35.59	57.05
γ (dyne/cm)	21.99	27.93	34.74*	34.69*	18.73*	18.32	19.25	28.06*
LHV (kJ/mol)	-1235	-3734	-3290.95*	-2685.82*	-2698	-5065.3	-4937.3	-2890.36*
ρ (kg/m³)	789.3	862.3	888.3	913.2	948.7	687.8	715	948.8
YSI	10.3	170.9	54.75869963	52.69*	22	61.7	68.5	36.01*
IT (ºC)	363	482	421.98*	431.39*	431.10*	418	391	324.68*
FP (ºC)	13	4	7	-13.94*	26	-12	-6	51
VP (kPa)	7.87	2.38*	0.44*	0.47*	1.55	6.5	5.96	0.294
LFL (vol.%)	3.3	1.1	1.62*	1.72*	1.5	0.95	0.88	1.41*
UFL (vol.%)	19	7.1	12.09*	12.66*	9.84*	5.59*	6.77*	9.99*
This work	Recommend	N/A	Recommend	Recommend	Recommend	N/A	Recommend	Recommend
Co-Optima	[419]	N/A	[419]	[419]	[419]	N/A	[419]	N/A
TMFB/FSC	[192]	N/A	[191]	[191]	N/A	N/A	N/A	[430, 431]

* Predictive values calculated by the machine learning QSPR models in this work, experimental data not available.

Table 6.5. Detail chemical kinetic mechanisms of 8 selected blendstock candidates for SI engine

Fuel candidates	CAS	Species	Reactions	Application	Institute	Mechanism
Ethanol	64-17-5	113	710	τ_{ig}, S_L	NUI Galway	[429]
Toluene	108-88-3	530	2808	τ_{ig}	LLNL	[146, 437]
Toluene	108-88-3	679	1740	S_L	LLNL	[438]
2,5-Dimethylfuran	625-86-5	524	3143	τ_{ig}, S_L	Bielefeld University	[436]
2-Methylfuran	534-22-5	524	3143	τ_{ig}, S_L	Bielefeld University	[436]
Cyclopentanone	120-92-3	444	2269	τ_{ig}	LLNL	[435]
iso-Octane	540-84-1	874	3796	τ_{ig}	LLNL	[439]
iso-Octane	540-84-1	73	296	S_L	University of Wisconsin-Madison	[440]
Diisobutylene	107-39-1	897	3783	τ_{ig}	NUI Galway	[138]
Diisobutylene	107-39-1	178	758	S_L	Tsinghua University	[426]
Cyclopentanol	96-41-3	278	1475	τ_{ig}, S_L	RWTH Aachen University	[430, 431]

6.3.1.3 Performance assessment of 8 selected blendstock candidates for SI engine

The performance assessment composes of 5 basic components: (1) volatility; (2) atomization; (3) energy density; (4) sooting tendency; (5) ignitability and the weight of each component is equal to 20%. Volatility is characterized by T_m, T_b and ΔH_{vap} while atomization is described by surface tension γ. Energy density is indicated by LHV while YSI is used to mark sooting tendency. Ignitability is assessed by RON, OS, φ-sensitivity, S_L. Each of these parameters is normalized to a range of 0~20% and the score of a particular candidate is calculated by Eq. (6.2). The fuel candidate with a higher score reflects better overall performance. The fuel merit function Eq. (6.2) is tailor-made for current fuel screening application which provides a paradigm to evaluate the potential benefits among various fuel candidates. The users are encouraged to develop the customized criteria based on specific research requirements. For example, the fuel efficiency merit function for SI engine is proposed to evaluate the thermal efficiency benefits of various fuels by considering the impact of RON, OS, HOV, laminar flame speed and particulate matter index (PMI) [441]. The scores of the 8 selected fuel candidates in Tier 2 screening arrange from high to low as: diisobutylene > iso-octane > ethanol > cyclopentanone > cyclopentanol > 2,5-dimethylfuran > toluene > 2-methylfuran as shown in Figure 6.8. Diisobutylene obtains a good balance among atomization (γ =19.25dyne/cm), energy density (LHV=-4937.3kJ/mol) and sooting tendency (YSI =68.5). Ethanol has high volatility (T_m =-114.14°C, T_b =78.24oC, VP=7.87kPa) and good soot suppression capability (YSI=10.3) with the drawback of low energy density (LHV=-1235kJ/mol). Low sooting tendency (YSI=22) and good atomization performance (γ =18.73 dyne/cm) are the main advantages of cyclopentanone. Compared to cyclopentanone, cyclopentanol has a worse sooting tendency (YSI=36.01, predicted value) and atomization performance (γ =28.06 dyne/cm). 2,5-Dimethylfuran and 2-methylfuran are favorable components for blending into gasoline base fuel to enhance anti-knock quality.

$$Score = Volatility_{nirm} + Atomization_{norm} + Energy\ density_{norm} + Sooting\ tendency_{norm} + Ignitability_{norm}$$
$$= \frac{T_{m,norm} + T_{b,norm} + \Delta H_{vap,norm}}{3} + \gamma_{norm} + LHV_{norm} + YSI_{norm} + \frac{RON_{norm} + OS_{norm} + \eta_{norm} + S_{L,norm}}{4} \qquad (6.2)$$
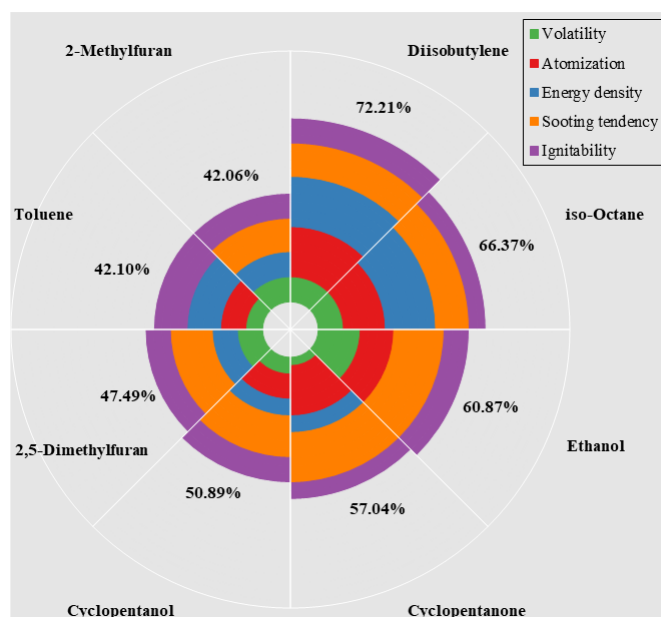
Figure 6.8. Performance comparison of volatility, atomization, energy density, sooting tendency and ignitability for 8 selected blendstock candidates for SI engine.

6.3.2 Case study 2: fuel screening for CI engine

6.3.2.1 Tier 1 fuel physicochemical property screening for CI engine

Tier 1 fuel physicochemical property screening for CI engine inspects volatility, atomization, energy density, sooting tendency, ignitability and the parameters constraints are depicted in Figure 6.9. The specifications constraints refer to the recommendation from Co-Optima [414, 415, 423] and TMFB [191] projects which are compared in Table 6.2. The upper bound of the melting point is -10°C to guarantee low-temperature fluidity. The boiling point is at the range of 60~250°C which follows the recommendation by TMFB project [191]. High enthalpy of vaporization benefits low-temperature combustion for NO_x reduction, so an upper limit is set as 75kJ/mol. The ceilings of surface tension and dynamic viscosity are 38mN/m and 2mPa·s which promote liquid droplet break up and homogeneous fuel-air mixture formation. As mentioned earlier, the filter of dynamic viscosity plays no role since corresponding machine learning QSPR model fails to reproduce the observed data accurately. The energy density requires the LHV and density are equal to or greater than 2700kJ/mol and 675kg/m³ respectively. Given that the YSI of diesel is around 215 [416], the YSI of the blendstock candidates is confined below 70 to reduce soot emission considerably by blending the candidates with diesel base fuel. Similar to Co-Optima [414, 415, 423] and TMFB [191] projects, a minimum CN of 40 is required for stable autoignition at a cold start. 10 fuel types, 129 compounds satisfy required criteria as shown in Figure 6.9 and the complete list is provided in author's publication [424]. Tier 1 fuel screening candidates for CI engine comprises 10 fuel types,129 compounds as shown in Figure 6.10 (a). Acyclic ethers, iso-alkanes, polyfunctionals make up 63.57% of Tier 1 fuel screening candidates. Alkenes, aldehydes, saturated esters, cycloalkanes, n-alkanes, other cyclic ethers, furans account for the rest 36.43%. CN versus YSI of these 129 compounds is plotted in Figure 6.10 (b) to identify those fuel types with high CN and low YSI simultaneously. Acyclic ethers and other cyclic ethers in the 129 compounds are readily autoignition and possess low sooting tendency. The represented candidates of these 10 fuel types are presented in Table 6.6 and they all have a long straight-chain structure in the molecules.

Hierarchical clustering analysis is conducted on the functional groups and the molecules to identify the key functional groups of Tier 1 fuel screening candidates (129 compounds) and the heatmap is shown in Figure 6.11. A high-resolution, interactive version of the heatmap ("Tier1 CI fuel heatmap.json" file) with zooming capability is provided in author's publication [424]. Reducing the number of methyl groups (functional group type 8) and increasing the number of methylene groups (functional group types 9~12) can boost CN. They are also the fuel reactivity descriptors in QSPR-UOB 3.0 scheme (see Figure 4.2). Adding ether group in acyclic or cyclic molecules can substantially improve CN while maintains modest YSI due to the introduction of the oxygen atom., so the color keys of functional groups 16, 17 demonstrate deep red color in the heat map. The design rules of diesel blendstock for CI engine on the molecular level are summarized as (1) increasing the number of methylene in series and reducing the number of side chains; (2) adding ether group and moving toward the center of the molecule to increase CN while maintaining modest YSI; (3) moving the CCDB, ester group, aldehyde group, ketone group, side chain and ring structure toward the edge of the molecule or even removed from a molecule; (4) avoiding the cyclic (e.g. cycloalkanes, clcloalkenes, cycloalcohols) and unsaturated (e.g. alkenes, unsaturated esters) structures; (5) discarding the aromatic structure.

Tier 1 screening for CI fuel candidates

1742 molecules in fuel properties database
Start — 1742, 100%

Volatility

$T_m \leqslant -10^{\circ}C$? — Melting point T_m
False — 572, 32.84%
True — 1170, 67.16%

$60^{\circ}C \leqslant T_b \leqslant 250^{\circ}C$? — Boiling point T_b
False — 198, 11.37%
True — 972, 55.80%

$\Delta H_{vap} \leqslant 75kJ/mol$ — Enthalpy of vaporization ΔH_{vap}
False — 28, 1.61%
True — 944, 54.19%

Atomization

$\gamma \leqslant 38mN/m$ — Surface tension γ
False — 4, 0.23%
True — 940, 53.96%

$\mu \leqslant 2mPa \cdot s$ — Dynamic viscosity@20°C
False — 0
True — 940, 53.96%

Energy density

$LHV \geqslant 2700kJ/mol$ — Lower heating value
False — 102, 5.86%
True — 838, 48.11%

$\rho \geqslant 675kg/m^3$ — Density@15~30°C
False — 3, 0.17%
True — 835, 47.93%

Sooting tendency

$YSI \leqslant 70$ — Yield sooting index
False — 298, 17.11%
True — 537, 30.83%

Ignitability

$CN \geqslant 40$ — Cetane number
False — 408, 23.42%
True — 129, 7.41%

CI fuel candidates

Reject

Figure 6.9. Tier 1 fuel physicochemical property screening for CI engine by ML-QSPR models.
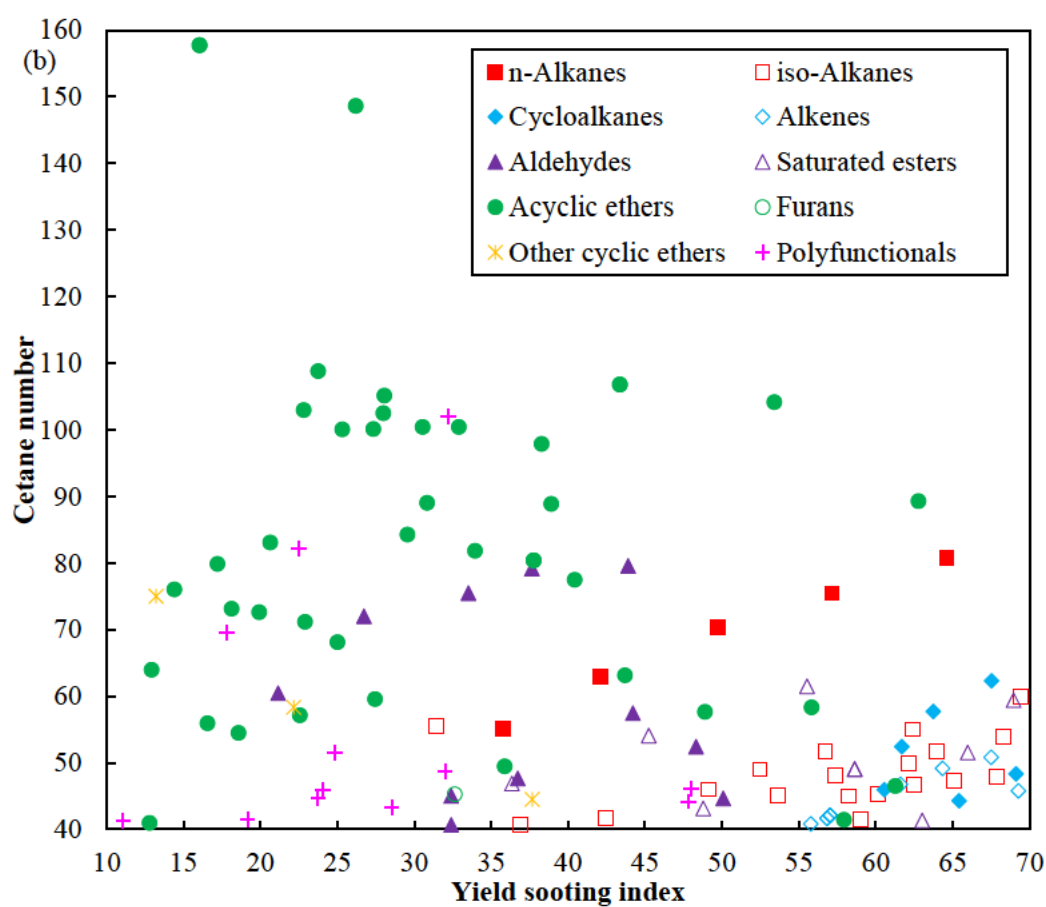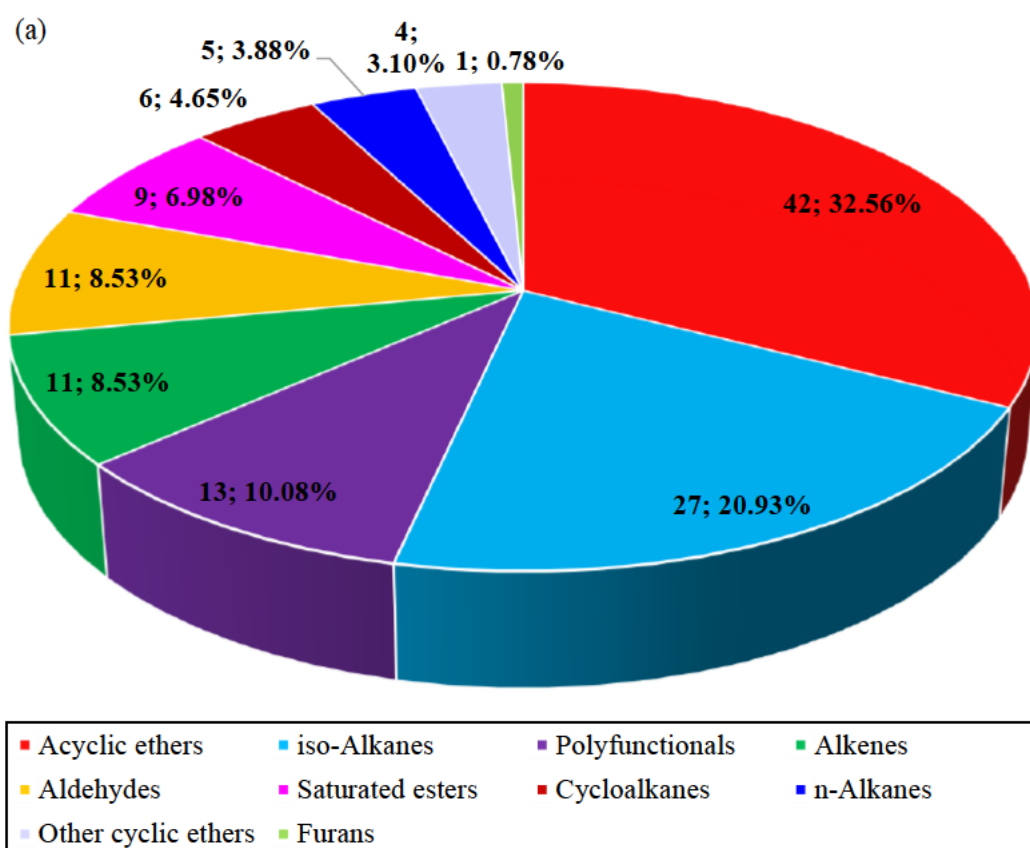
Figure 6.10. Tier 1 fuel screening candidates for CI engine (129 compounds): (a) proportion of 10 fuel types; (b) CN versus YSI.

Figure 6.11. Hierarchical clustering of Tier 1 fuel screening candidates for CI engine (129 compounds), a high-resolution, interactive version of the heatmap ("Tier1 CI fuel heatmap.json" file) with zooming capability is provided in author's publication [424].

Table 6.6. Represented candidates of Tier 1 fuel screening for CI engine

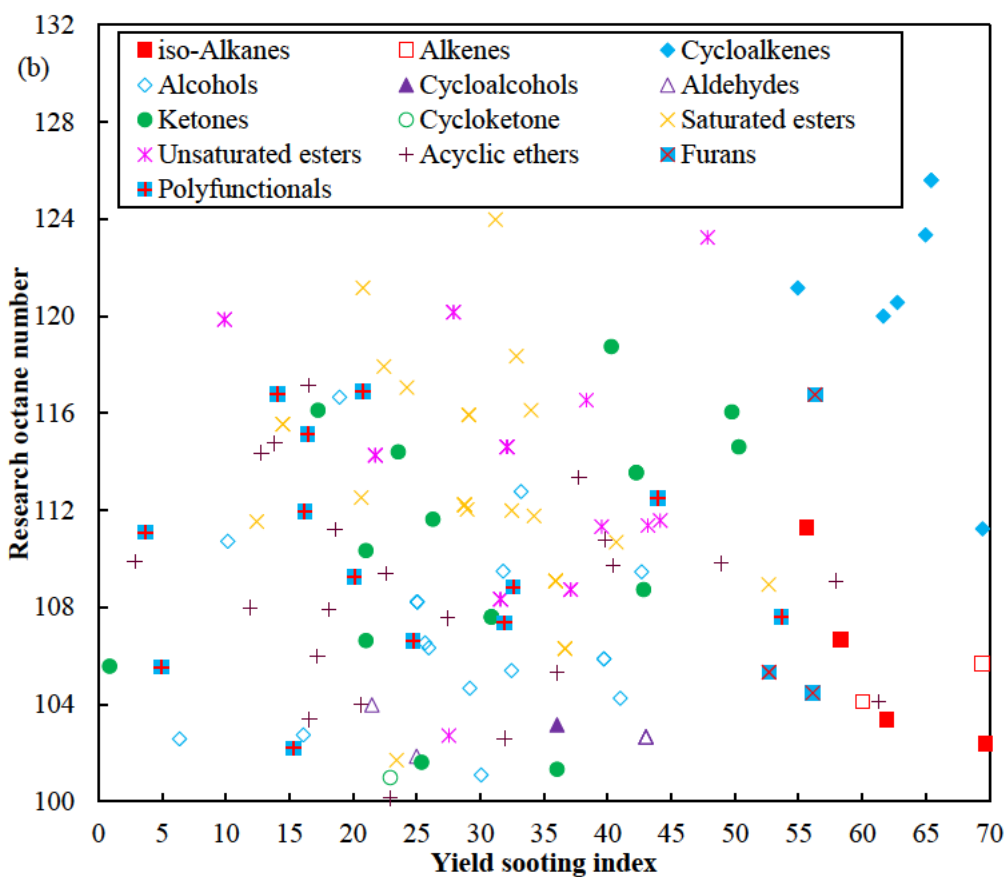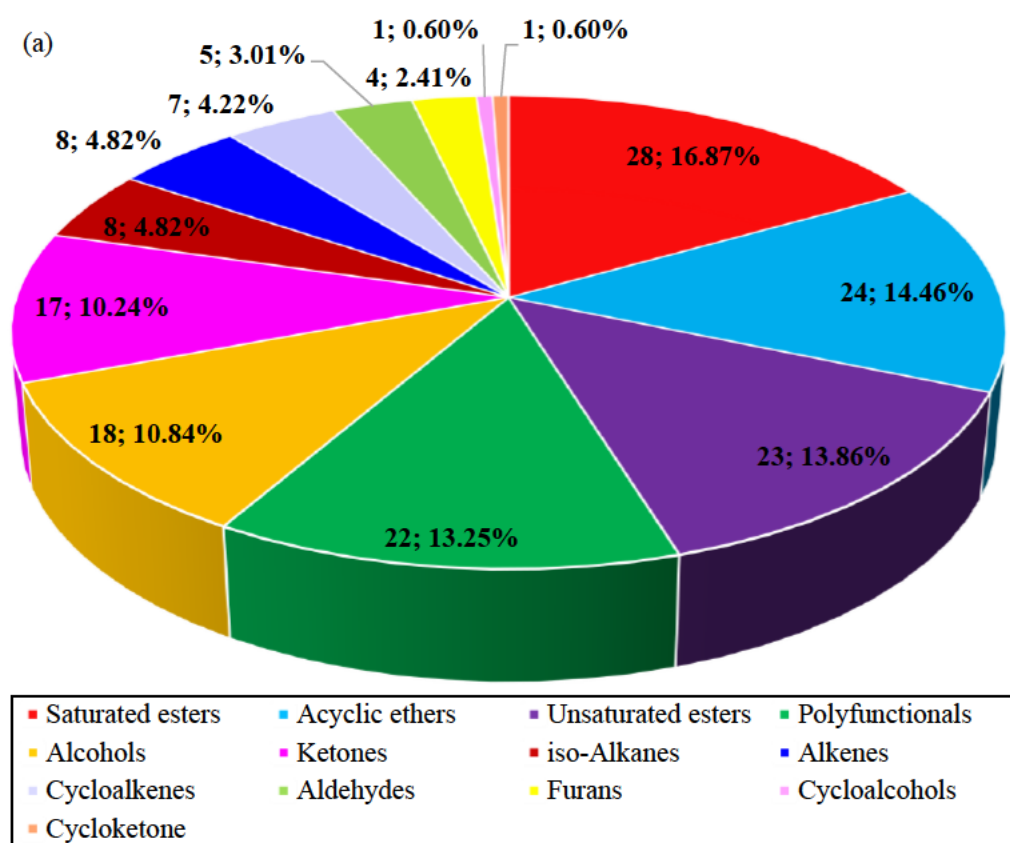| Fuel candidates	Structure	Fuel type	CAS	CN	YSI	Mechanism
n-Octane		n-Alkane	111-65-9	63.8	42.6	[91, 442]
2,2-Dimethyloctane		iso-Alkane	15869-87-1	59	31.44*	N/A
n-Butylcyclohexane		Cycloalkane	1678-93-9	47.6	76.8	[443, 444]
8-Methyl-1-nonene		Alkenes	26741-24-2	50.83*	67.51*	N/A
1-Octanal		Aldehyde	124-13-0	80.5	37.5	[445, 446]
Decyl acetate		Saturated ester	112-17-4	62	68.97	N/A
PODE3		Acyclic ether	13353-03-2	78	14.39*	[447]
2-Butyltetrahydrofuran		Furan	1004-29-1	45.5	32.61*	[448]
Ethyl tetrahydrofurfuryl ether		Other cyclic ether	62435-71-6	78.90	13.21*	N/A
Ethyl-3-ethoxypropionate		Polyfunctional	763-69-9	82.20*	22.51*	N/A

* Predictive values calculated by the ML-QSPR models in this work, experimental data not available.

6.3.2.2 Tier 2 chemical kinetic screening for CI engine

Dibutyl ether, PODE$_3$, n-heptane, n-octane, 2-butyltetrahydrofuran, butylcyclohexane are selected from the 129 fuel blendstock candidates. PODE$_4$ and 1-octanol are recommended by Kopernikus Project Power-to-X [447] and TMFB project [445, 449-451] respectively which provide high LHV of 3223.97kJ/mol and 4898.3kJ/mol. The physicochemical properties and mechanism sources of these 8 candidates are provided in Table 6.7 and Table 6.8 respectively. Modifications are conducted on the chemical kinetic mechanisms of 8 selected candidates for the CI engine and the modified mechanisms are provided in Data and Software Availability.

The ignition delay time is tested under $\varphi=0.3$, $T_{init}=1050K\sim1200K$, $P_{init}=25bar$ which represents the autoignition condition for advanced CI engine [419] as shown in Figure 6.12. The ignition times at 1060K order from longest to shortest as: n-octane (1.47ms) > dibutyl ether (1.38ms) > n-heptane (1.31ms) > 1-octanol (1.14ms) > butylcyclohexane (0.96ms) > 2-butyltetrahydrofuran (0.71ms) > PODE$_3$ (0.22ms) > PODE$_4$ (0.21ms).

The φ-sensitivity is studied under (a) $\varphi=0.2$, $T_{init}=600\sim1400K$, $P_{init}=30bar$; (b) $\varphi=2.0$, $T_{init}=600\sim1400K$, $P_{init}=30bar$; (c) $\varphi=0.2$, $T_{init\ t}=600\sim1400K$, $P_{init}=60bar$; (d) $\varphi=2.0$, $T_{init}=600\sim1400K$, $P_{init}=60bar$ (as shown in Figure 6.13) which cover the autoignition conditions in advanced CI engine. The peak φ-sensitivity of the 8 selected candidates at 30bar order from high to low as: butylcyclohexane (2.793) > dibutyl ether (1.652) > PODE$_4$ (1.441) > 2-butyltetrahydrofuran (1.404) > PODE$_3$ (1.380) > n-heptane (1.194) > 1-octanol (1.148) > n-octane (1.054) and the same order is observed at 60bar. Butylcyclohexane achieves the greatest φ-sensitivity over the range of 600~1400K among the 8 selected candidates. Similarly, PODE$_4$ and PODE$_3$ also exhibit high φ-sensitivity as increasing initial temperature. On the contrary, the φ-sensitivity of dibutyl ether peaks at 30bar occurs at 780K (1.652) but it decreases dramatically from 0.642 to 0.151 as initial temperature increases from 1000K to 1400K. Thus, dibutyl ether is more suitable for low temperature combustion compared to mixing controlled compression ignition.

The laminar flame speed is tested under $\varphi=0.4\sim1.7$, $T_{init}=428K$, $P_{init}=1bar$ as shown in Figure 6.14 and the fuel

candidates arrange from high to low as: PODE$_3$ (82.39cm/s) > PODE$_4$ (81.18cm/s) >dibutyl ether (76.35cm/s) > n-Heptane (73.57cm/s) > 1-octanol (70.86cm/s) > n-octane (74.09cm/s) > butylcyclohexane (67.67cm/s) > 2-butyltetrahydrofuran (60.60cm/s). Specially, the laminar flame speed of dimethyl ether is less temperature dependent than PODE$_3$, PODE$_4$ as φ decreasing from 0.7 to 0.4. Dibutyl ether possesses high combustion rate and dilution tolerance at fuel-lean conditions (φ=0.4~0.7) which particularly benefits low temperature combustion in CI engine.
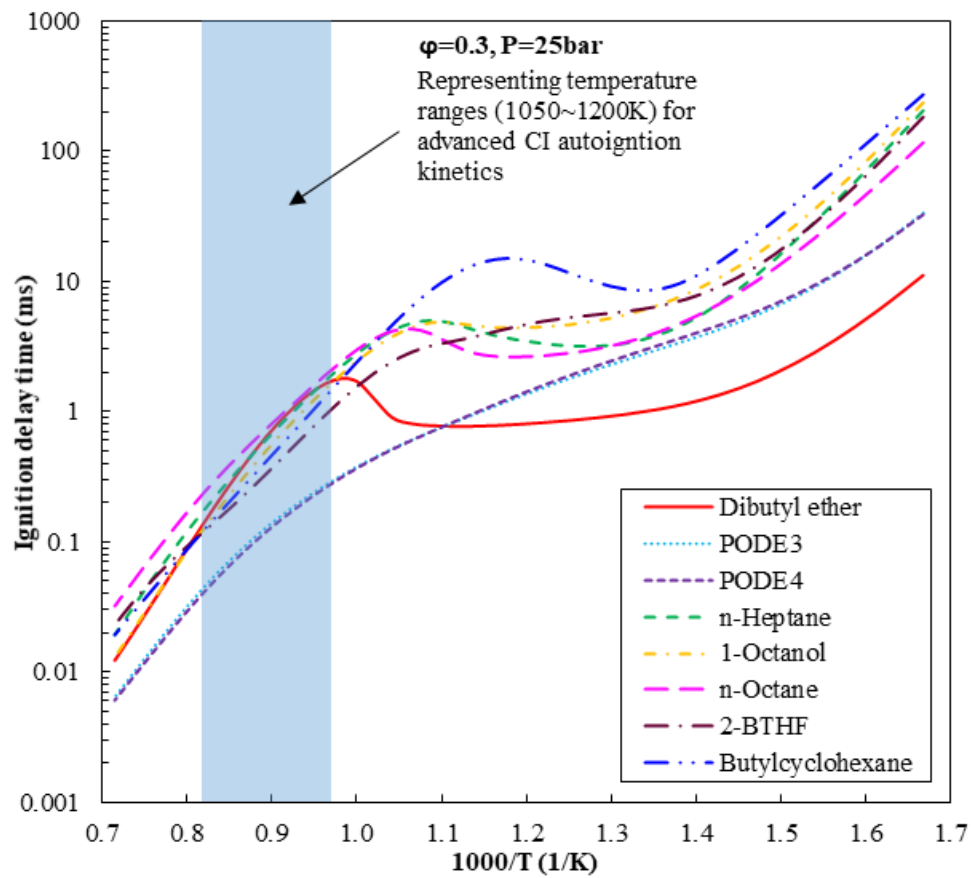
Figure 6.12. Ignition delay times of fuel (candidates for CI engine)-air mixture at φ=0.3, T$_{init}$=600~1400K, P$_{init}$=25bar.
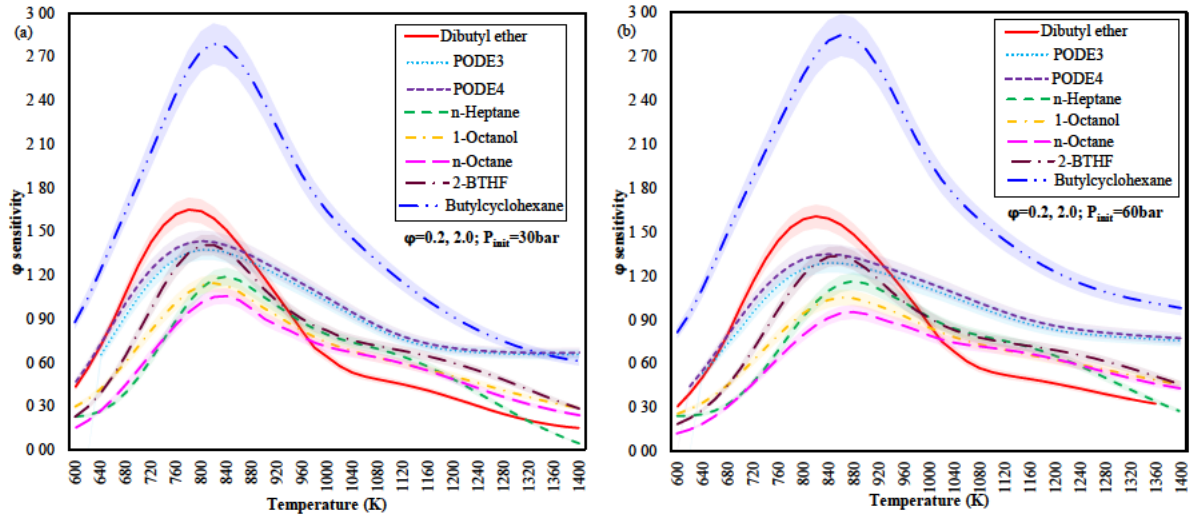
Figure 6.13. φ-sensitivity of fuel (candidates for CI engine)-air mixture at φ=0.2, 2.0, T_{init}=600~1400K, (a) P_{init}=30bar, (b) P_{init} =60bar; the shaded region representing a 95% confidence interval.
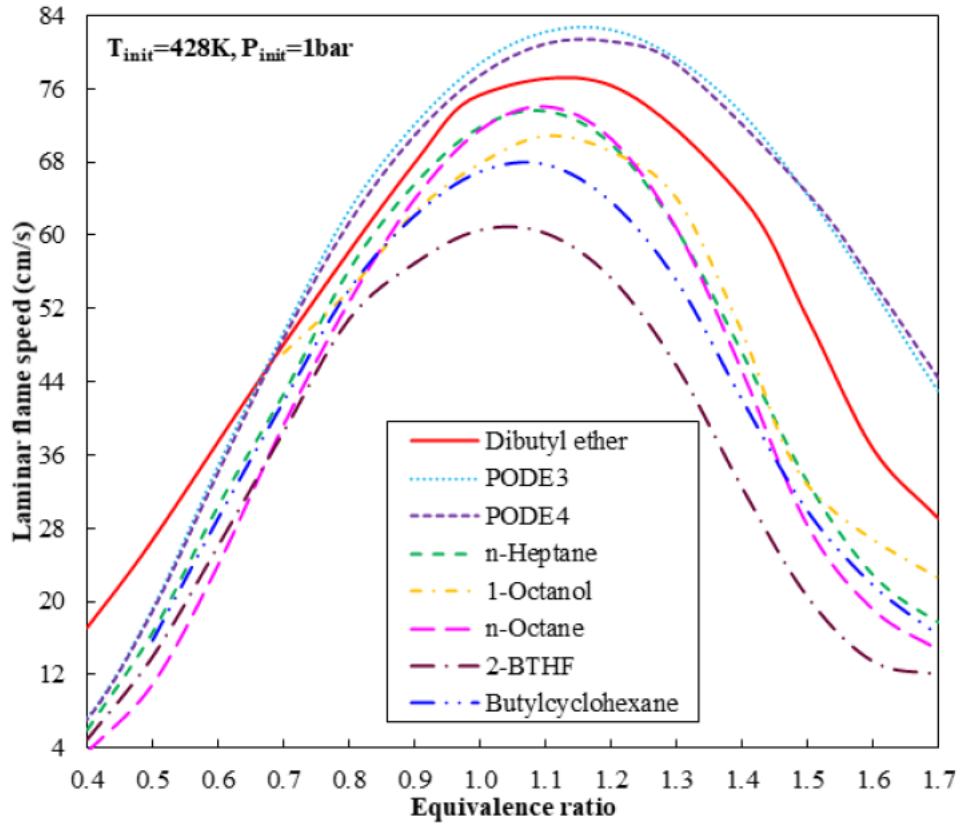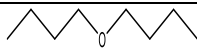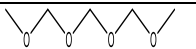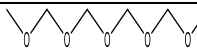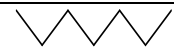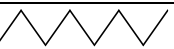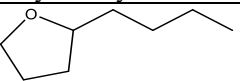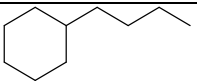
Figure 6.14. Laminar flame speed of fuel (candidates for CI engine)-air mixture at φ=0.4~1.7, T_{init} =428K, P_{init} =1bar.

Table 6.7. Physicochemical properties of 8 selected blendstock candidates for CI engine

Fuel candidates	Dibutyl ether	PODE$_3$	PODE$_4$	n-Heptane	1-Octanol	n-Octane	2-Butyltetrahydrofuran	Butylcyclohexane
Structure								
Fuel type	Acyclic ether	Acyclic ether	Acyclic ether	n-Alkane	Alcohol	n-Alkane	Furan	Cycloalkane
CAS	142-96-1	13353-03-2	13352-75-5	142-82-5	111-87-5	111-65-9	1004-29-1	1678-93-9
CN	95	78	90	56	39.1	63.8	45.5	47.6
T_m (ºC)	-96	-25.30*	2.81*	-90.549	-14.7	-56.73	-92.02*	-74.68
T_b (ºC)	141.6	183.10*	229.93*	98.38	194.7	125.62	163.05*	180.9
ΔH_{vap} (kJ/mol)	44.97	46.69*	56.42*	36.57	70.98	41.49	52.18*	49.36
γ (dyne/cm)	18.75*	35.20*	36.67*	19.78	27.43*	21.08	28.49*	26.51
LHV (kJ/mol)	-4946.9	-2741.27*	-3223.97*	-4464.7	-4898.3	-5074.2	-5385.0*	-6090.2
ρ (kg/m^3)	768.4	1015.1*	1049.86*	679.5	826.2	698.6	875*	790.2
YSI	38.7	14.39*	35.89*	36	41.1	42.6	32.61*	76.8
IT (ºC)	194	340.12*	356.74*	213	270	206	290.99*	246
FP (ºC)	25	83.54*	117.24*	-4	81	13	39.48*	48
VP (kPa)	0.898	0.507*	0.074*	6.09	0.01	1.86	5.462*	3.306*
LFL (vol.%)	0.9	1.15*	0.97*	1.05	0.84	0.96	1.40	0.79*
UFL (vol.%)	7.6	12.87*	11.82*	7	6.4	6.5	7.90*	5.5
This work	Recommend	Recommend	N/A	Recommend	N/A	Recommend	Recommend	Recommend
Co-Optima	N/A	N/A	N/A	N/A	N/A	N/A	N/A	[415]
TMFB/FSC	[449, 450, 452]	N/A	N/A	N/A	[445, 449-451]	N/A	[191, 448]	N/A
Power-to-X	N/A	[447]	[447]	N/A	N/A	N/A	N/A	N/A

* Predictive values calculated by the ML-QSPR model in this work, no experimental data available.

Table 6.8. Detail chemical kinetic mechanisms of 8 selected blendstock candidates for CI engine

Fuel candidates	CAS	Species	Reactions	Application	Institute	Mechanism
Dibutyl ether	142-96-1	436	2732	τ_{ig}, S_L	CNRS-INSIS	[453]
PODE$_3$	13353-03-2	322	1611	τ_{ig}, S_L	RWTH Aachen University	[447]
PODE$_4$	13352-75-5	322	1611	τ_{ig}, S_L	RWTH Aachen University	[447]
n-Heptane	142-82-5	1268	5336	τ_{ig}	NUI Galway	[454]
n-Heptane	142-82-5	73	296	S_L	University of Wisconsin-Madison	[440]
1-Octanol	111-87-5	1281	5510	τ_{ig}	RWTH Aachen University	[445]
1-Octanol	111-87-5	403	2374	S_L	Xi'an Jiaotong University	[446]
n-Octane	111-65-9	1147	4714	τ_{ig}	LLNL	[91]
n-Octane	111-65-9	80	194	S_L	Dalian University of Technology	[442]
2-Butyltetrahydrofuran	1004-29-1	424	1790	τ_{ig}, S_L	RWTH Aachen University	[448]
Butylcyclohexane	1678-93-9	397	2171	τ_{ig}	Shanghai Jiao Tong University	[443]
Butylcyclohexane	1678-93-9	348	2163	S_L	Stanford University	[444]

6.3.2.3 Performance assessment of 8 selected blendstock candidates for CI engine

The scores of the 8 selected fuel candidates in Tier 2 screening arrange from high to low as: Dibutyl ether > n-heptane > n-octane > 2-butyltetrahydrofuran > butylcyclohexane > 1-octanol > PODE$_3$ > PODE$_4$ as shown in Figure 17. Dibutyl ether possesses good volatility (T_m =-96°C, T_b =141.6°C, ΔH_{vap} =44.97kJ/mol), atomization (γ =18.75dyne/cm), ignitability (CN=95, φ-sensitivity at 60bar=1.60, S_L =75.31cm/s at φ=1.0) while maintaining modest energy density (LHV=-4946.9kJ/mol) and sooting tendency (YSI=38.7) among 8 selected candidates. n-Heptane possesses good atomization (γ =19.78dyne/cm) and sooting tendency (YSI=36) properties while n-octane provides good atomization (γ =21.08dyne/cm) and energy density (LHV=-5074.2kJ/mol). 2-Butyltetrahydrofuran and butylcyclohexane have the advantages of low sooting tendency (YSI=32.61) and high energy density (LHV=-6090.2kJ/mol) respectively for blending application into diesel base fuel. The CN of 1-octanol is 39.1 which fails to meet the ignitability requirement of this work (CN≥40). PODE$_3$ and PODE$_4$ are readily autoignition (CN: 78, 90) and ultra-low soot emission (YSI=14.39, 35.89), but their surface tensions are greater than the rest candidates which may deteriorate liquid atomization. They are favorable components for blending into diesel base fuel to improve ignitability and reduce sooting tendency.
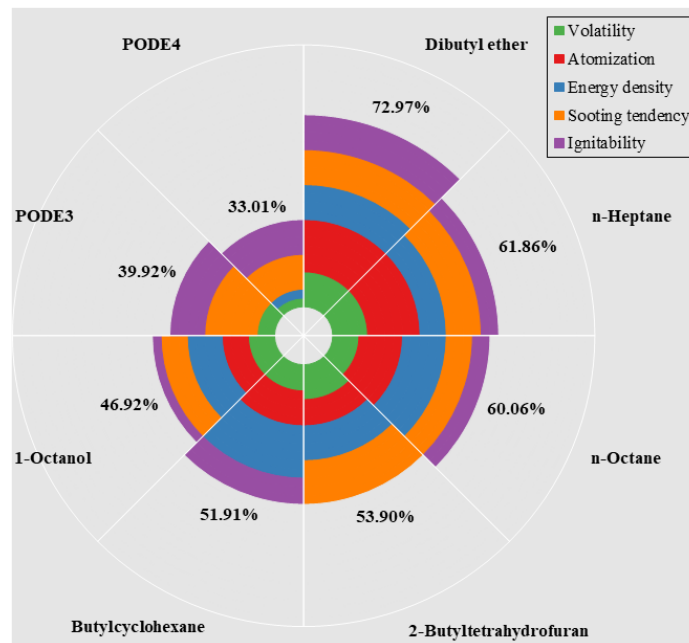
Figure 6.15. Performance comparison of volatility, atomization, energy density, sooting tendency and ignitability for 8 selected blendstock candidates for CI engine.

6.3.3 Engine results with tailor-made fuel of n-heptane-dibutyl ether-ethanol

6.3.3.1 Engine experimental program

This section performs the engine test to examine the combustion and emission characteristics of the fuels recommended by the fuel screening results. A single-cylinder compression ignition (CI) engine with a displacement of 532 cm^3 is applied as a test engine and the engine specifications are listed in Table 6.9. Correspondingly, the dibutyl ether and n-heptane are adopted as test fuels for the CI engine based on the results in section 6.3.2. The target-oriented fuel design [268, 269] advocates the ideal functional configuration of fuel components as "chemical ignition source−PM inhibitor−homogeneous charge" for efficient and clean combustion in CI engines. N-Heptane and dibutyl ether are regarded as a chemical ignition source and PM inhibitor respectively and they easily form homogeneous charge due to their low boiling points of 98.38°C, 141.6°C. The ethanol is introduced to adjust the CN of test fuel since the CN of dibutyl ether is as high as 95. The target CN of the test fuel is comparable to the reference diesel (provided by Shell) of 53.9. The ternary components fuel formulation by CN recommends the proportion of 50%n-heptane-40%dibutyl ether-10%ethanol (vol.%) and the fuel specifications are presented in Table 6.10. The engine test rig is illustrated in Figure 6.16 and the exhaust gas components are analyzed by MultiGas™ 2030 gas analyzer (see Table 6.11) [455, 456]. The O$_2$ concentration is required to calculate the specific emissions which is measured by AVL DiTEST Gas 1000 (measuring range: 0~25vol.%, resolution: 0.01 vol.%) [457]. The operating condition fix at the load of 2 bar IMEP, the speed of 1500 rpm and the experimental program are shown in Table 6.12. The configuration of pilot injection-main injection is adopted, the main injection timing varies as 5°CA, 8°CA, 11°CA BTDC and the pilot injection timing is 10°CA advanced the main injection. As the main injection timing moving from 5°CA BTDC to 11°CA BTDC, the fuel-air mixture tends to becomes more homogenous with increasing ignition delay times (IDT). The pilot injection duration accounts for 20 vol.% of the total injection duration to induce the cool flame ignition and the injection pressure keeps a constant of 550 bar. The pilot injection duration and main injection duration are determined

experimentally by achieving the preset load of 2 bar IMEP. The experimentally determined pilot injection duration, main injection duration for both diesel and 50%C7H16-40%DBE-10%EtOH at varied conditions are provided in Table 6.12. 50%n-heptane-40%dibutyl ether-10%ethanol has LHV of 39.86 MJ/kg, CN of 55.021, RON of 56.79, MON of 48.35, oxygen mass content of 8.94 mass%, density@20°C of 0.714g/cm^3. The fuel reactivity of CN, RON, MON are predicted by the ML-QSPR method which has been discussed in chapter 5. To reduce the data uncertainty, the in-cylinder combustion pressure profile is obtained by averaging the data of 1000 cycles. The test time of gas-phase emission measurement by MultiGas™ 2030 gas analyzer is set as 20 min to ensure the data reliability.
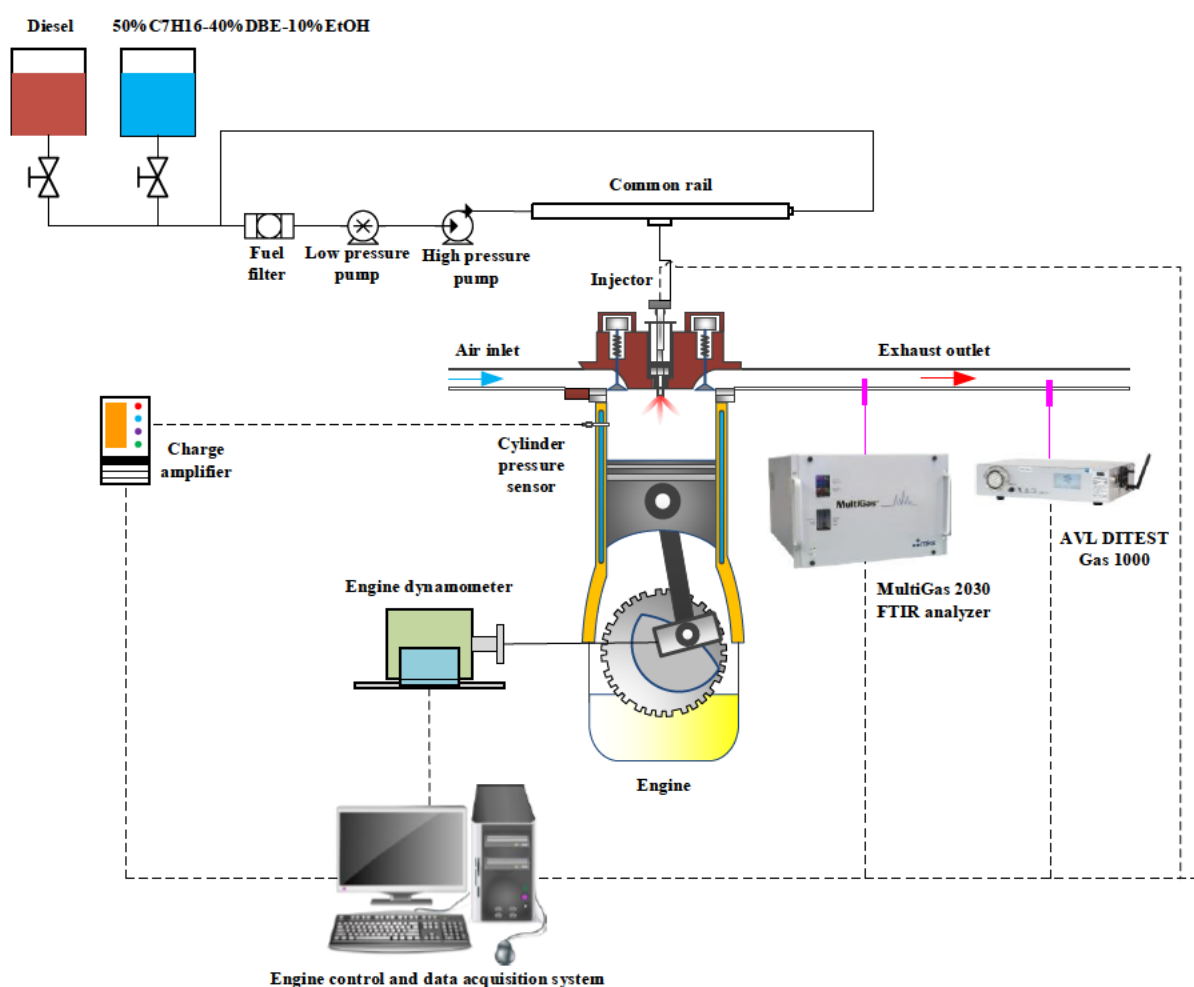
Figure 6.16. Schematic diagram of the engine test rig.

Table 6.9. Engine specifications [458].

Engine parameters	Specifications
Engine type	Compression ignition
Stroke type	Four-stroke
Number of cylinders	1
Cylinder bore x stroke (mm)	84 x 90
Connecting rod length (mm)	160
Compression ratio	16:1
Displacement (cm³)	499
Engine speed range (rpm)	900 ~2000
IMEP range (bar)	<7
Fuel injection pressure range (bar)	500~1500
Injection system	Common rail

Table 6.10. Test fuel specifications

Items	ULSD	n-Heptane	Dibutyl ether	Ethanol
CAS	N/A	142-82-5	142-96-1	64-17-5
Chemical formula	$C_{14}H_{26\ 09}$	C_7H_{16}	$C_8H_{18}O$	C_2H_6O
Molecular weight (g/mol)	194.44	100.21	130.23	46.07
Melting point (ºC)	N/A	-90.549	-96	-114.14
Boiling point (ºC)	N/A	98.38	141.6	78.24
Ignition temperature (ºC)	N/A	213	194	363
Flash point (ºC)	N/A	-4	25	13
Vapor pressure @25 °C (kPa)	N/A	6.09	0.898	7.87
Lower flammability limit	N/A	1.05	0.9	3.3
Upper flammability limit	N/A	7	7.6	19
Dynamic viscosity @20ºC (mPa*s)	2.0405	0.418	0.691	1.2
Surface tension @ 25ºC (dyne/cm)	N/A	19.78	18.75*	21.99
Heat of vaporization @ 25ºC (kJ/mol)	47.25	36.57	44.97	42.32
Density @20ºC (g/cm3)	0.8271	0.6795	0.7684	0.7893
Unified YSI	215.1 [416]	36	38.7	10.3
LHV (kJ/mol)	8302.6	4464.7	4946.9	1235
CN	53.9	56	95	12
RON	N/A	0	82.02*	111
MON	N/A	0	72.92*	96
OS	N/A	0	9.10*	15

*Predicted value is used since the measured value is not available.

Table 6.11. Specifications of MultiGas™ 2030 gas analyzer [456]

Items	Technical specifications
Gases and Vapors Measurable	Most molecules except for N_2, H_2, and O_2
Measurable Range (concentration)	100ppb~100%
Measurement Technique	FTIR Spectrometry
Spectral Resolution (cm^{-1})	0.5~16
Scan Speed (scan/sec @ 0.5cm-1)	1
Scan Time (sec)	1~300
Infrared Source	Silicon Carbide @ 1200°C
Reference Laser	Helium Neon (15798.2cm^{-1})
Max Purge Pressure (bar)	1.5
Spectrometer Purge Flow (L/min)	0.2
Optics Purge Flow (L/min)	0.2

Table 6.12. Experimental program

Fuel type	IMEP (bar)	Speed (rpm)	Injection pressure (bar)	Abbreviation	Pilot injection		Main injection	
					Timing (°CA BTDC)	Duration (ms)	Timing (°CA BTDC)	Duration (ms)
Diesel	2	1500	550	P15_M5	15	0.123	5	0.49
	2	1500	550	P18_M8	18	0.125	8	0.5
	2	1500	550	P21_M11	21	0.125	11	0.5
50%C7H16-40%DBE-10%EtOH	2	1500	550	P15_M5	15	0.133	5	0.53
	2	1500	550	P18_M8	18	0.135	8	0.54
	2	1500	550	P21_M11	21	0.14	11	0.56

6.3.3.2 Combustion character of n-heptane-dibutyl ether-ethanol

The injection profile, in-cylinder combustion pressure, heat release rate (HRR), mass fraction burn (MFB) profile of diesel and 50%C7H16-40%DBE-10%EtOH at the conditions of P21_M11, P18_M8, P15_M5 are plotted in Figure 6.17. The MFB05, MFB50, MFB95 for both fuels at various conditions are illustrated in Figure 6.18 which are the crank angles corresponding to 5%, 50%, 95% of total heat release. As the main injection timing advances from 5°CA BTDC to 11°CA BTDC, the MFB50 of 50%C7H16-40%DBE-10%EtOH moves forward from 6 °CA ATDC to 1°CA BTDC and that of diesel moves forward from 2°CA ATDC to 1°CA BTDC. The conversion from fuel chemical energy to the brake output work of IC engine contains 4 major energy losses of chemical energy loss in the exhaust, heat loss in the exhaust and heat transfer to the environment, pumping loss, mechanical loss by friction and accessories. The corresponding efficiency indices for these 4 energy losses are plotted in Figure 6.19. The brake thermal efficiency is the product of combustion efficiency, thermodynamic efficiency, gas exchange efficiency, mechanical efficiency as shown in Eq. (6.1). The formula to calculate these efficiency factors can be found in the textbook of internal combustion engine fundamentals [459].

$$\eta_{brake} = \eta_{combustion} \cdot \eta_{combustion} \cdot \eta_{gas\ exchange} \cdot \eta_{mechanical} \tag{6.1}$$

The MFB05 is regarded as the ignition initiation, the ignition delay time (IDT) is defined as the crank angle interval between the start of injection and MFB05. The combustion duration is defined as the crank angle interval between MFB05 and MFB95. The IDTs of 50%C7H16-40%DBE-10%EtOH keep constant as 8°CA as the main injection timing advances from 5 °CA BTDC to 11 °CA BTDC while those of diesel increases from 4°CA to 7°CA as shown in Figure 6.20 (a). The IDTs of 50%C7H16-40%DBE-10%EtOH are 1°CA~4 °CA longer than those of diesel at the studied conditions and it indicates that the CN of 50%C7H16-40%DBE-10%EtOH should be lower than diesel. The measured CN of the diesel is 53.9 and the predicted CN of 50%C7H16-40%DBE-10%EtOH by ML-QSPR method (see Chapter 5) is 55.021, thus the proposed method should further improve the predictive capability of ignition properties (CN, RON, MON) for fuel mixtures by expanding the training dataset. The

combustion duration of 50%C7H16-40%DBE-10%EtOH decreases from 11ºCA to 8 ºCA as the main injection timing advances from 5 ºCA BTDC to 11 ºCA BTDC. It indicates that 50%C7H16-40%DBE-10%EtOH tends to form a larger quantity of premixed fuel-air mixture as advancing injection timing.

The indicated thermal efficiency (ITE) of both fuels increases with advanced injection timing because the shortening combustion duration benefits the constant volume thermodynamic efficiency around the top dead center. Accordingly, the indicated specific fuel consumption (ISFC) of both fuels decreases with advanced injection timing. The ITE of 50%C7H16-40%DBE-10%EtOH increases as 23.97%, 25.22%, 26.68% at the main injection timing of 5ºCA, 8ºCA, 11ºCA BTDC which is 1.01%, 1.74%, 2.67% greater than diesel as shown in Figure 6.20 (b). The ISFC of 50%C7H16-40%DBE-10%EtOH decreases as 376.67 g/kWh, 358.09 g/kWh at the main injection timing of 5ºCA, 8ºCA BTDC which is 13.11 g/kWh and 2.35 g/kWh greater than diesel. This is because the LHV of 50%C7H16-40%DBE-10%EtOH (39.86 MJ/kg) is lower than diesel (43.11 MJ/kg) and a larger quantity is needed to achieve the target engine load. As the main injection timing advanced to 11ºCA BTDC, 50%C7H16-40%DBE-10%EtOH obtains lower ISFC of 338.44 g/kWh than diesel of 347.78 g/kWh because the enhanced ITE offsets the negative impact of modest LHV as shown in Figure 6.20 (b).
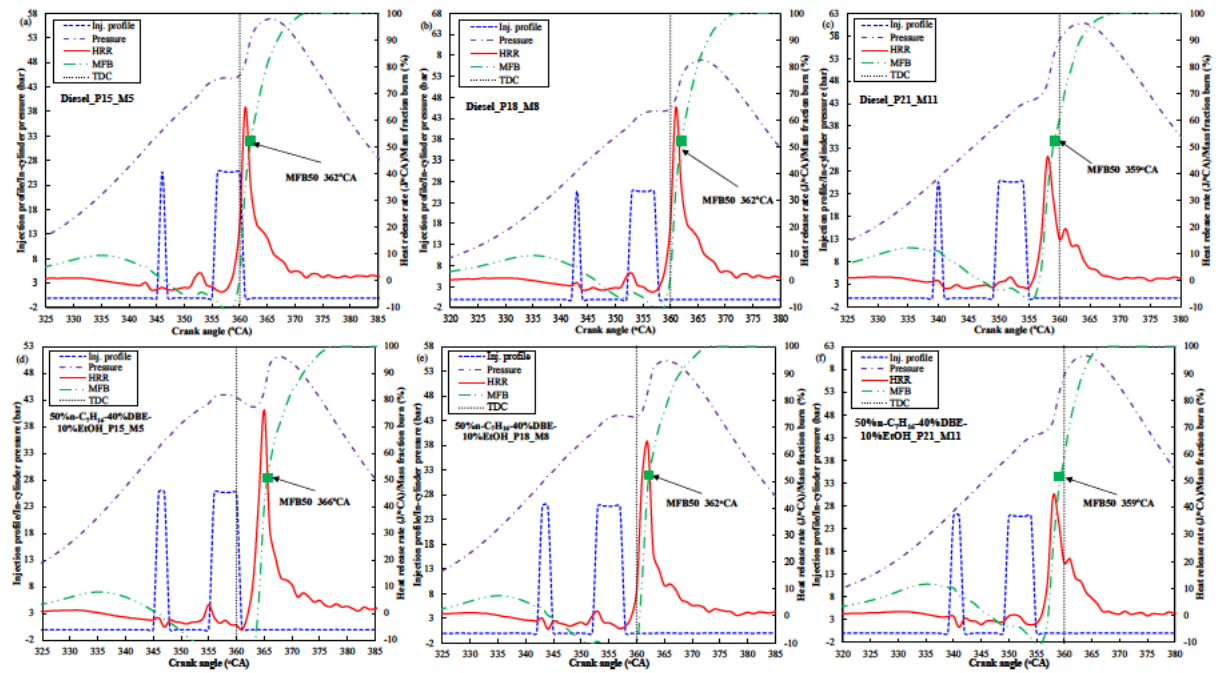
Figure 6.17. Injection profile, in-cylinder combustion pressure, HRR, MFB of diesel and 50%n-heptane-40%dibutyl ether-10%ethanol at the conditions of P21_M11, P18_M8, P15_M5.
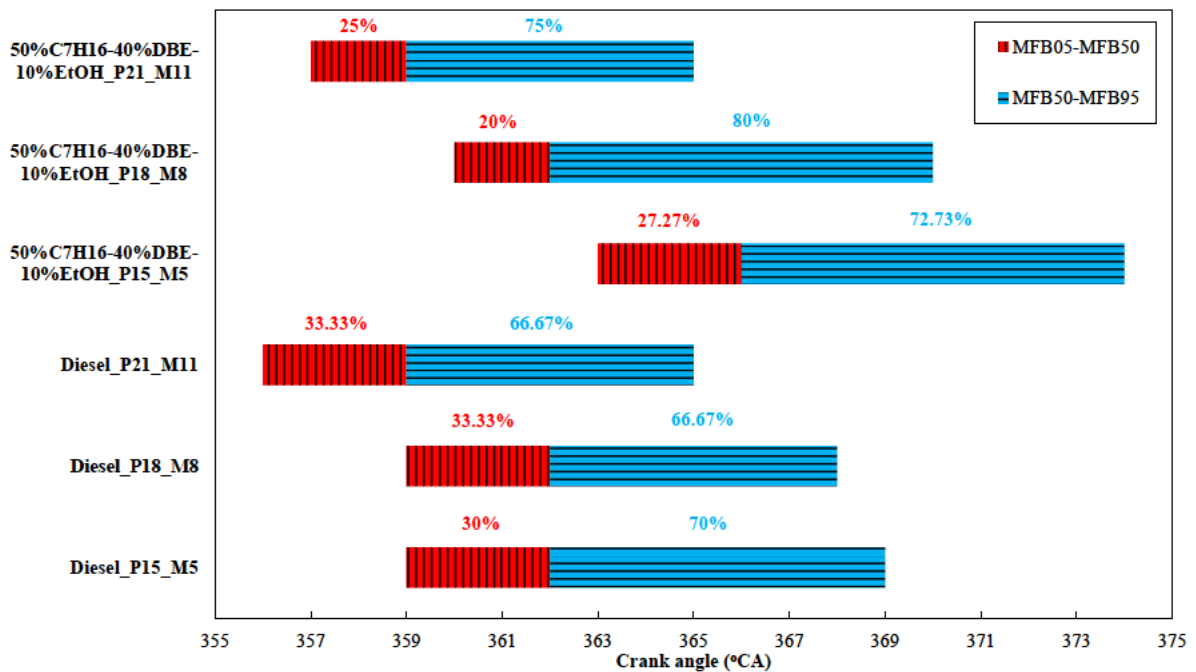
Figure 6.18. Combustion phasing (MFB05, MFB50, MFB95) of diesel and 50%n-heptane-40%dibutyl ether-10%ethanol at the conditions of P21_M11, P18_M8, P15_M5.
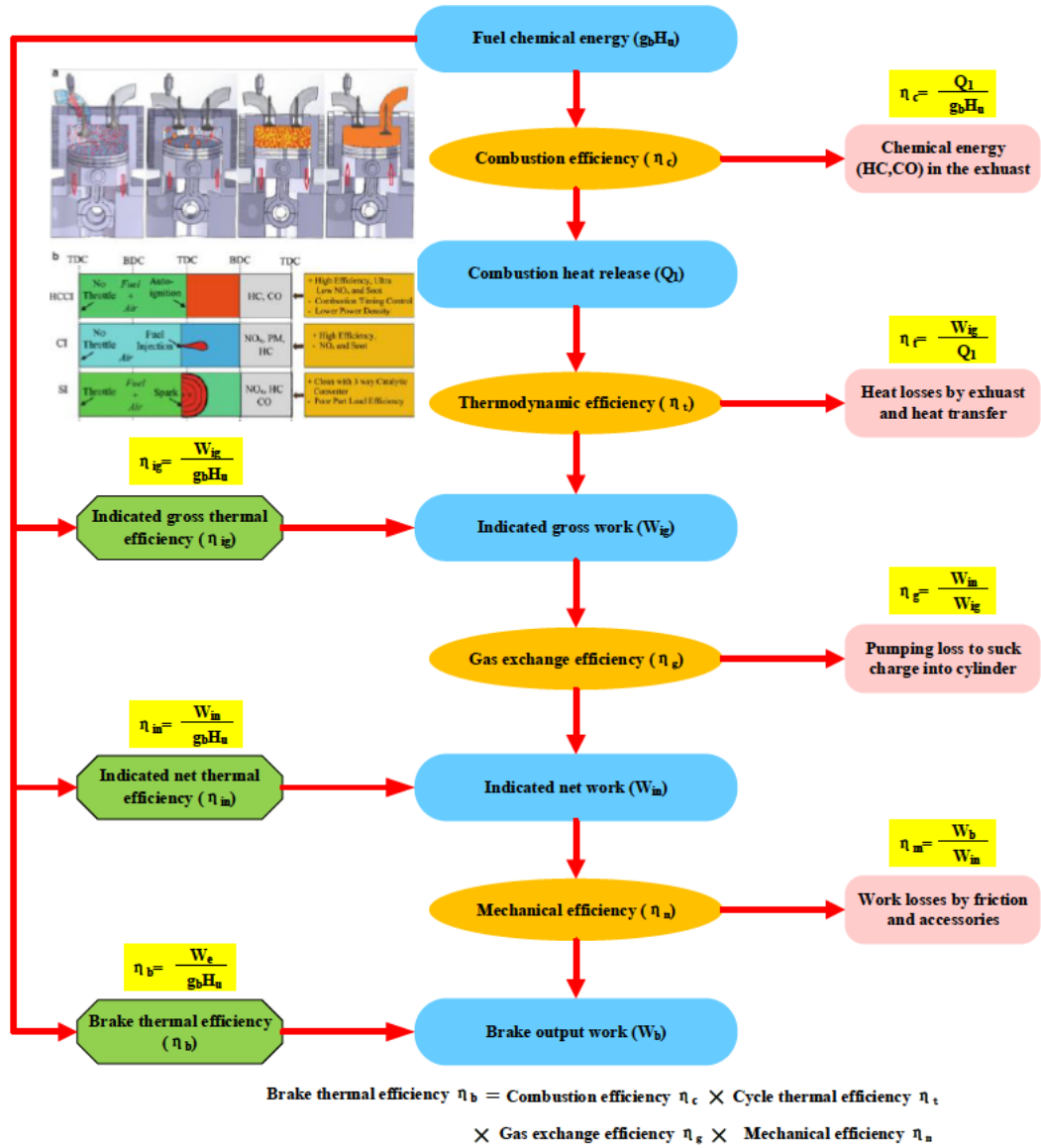
Figure 6.19. Overview of engine efficiency analysis [460].
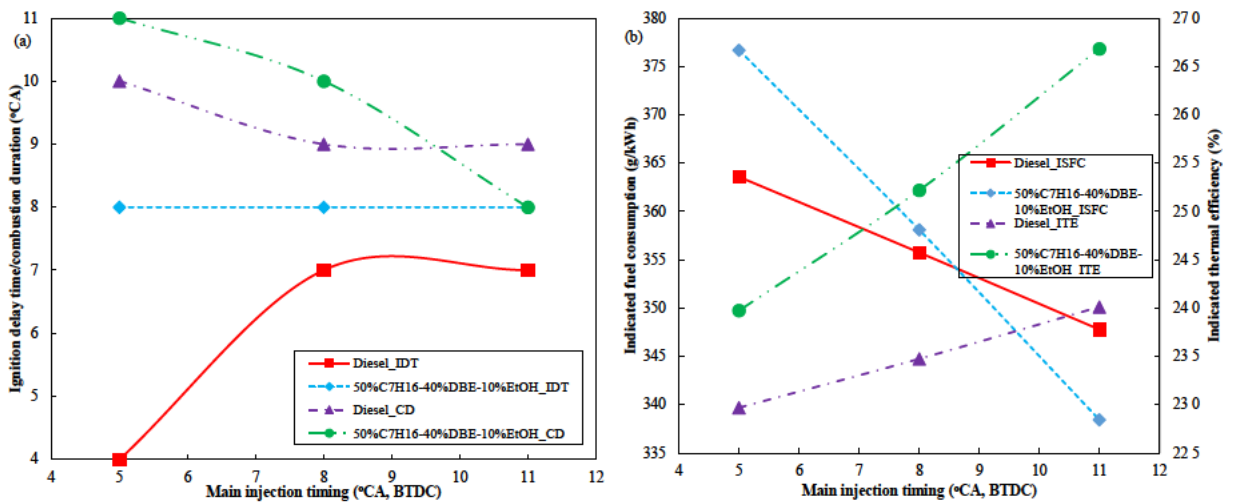
Figure 6.20. Ignition delay times, combustion duration, indicated specific fuel consumption, indicated thermal efficiency of diesel and 50%n-heptane-40%dibutyl ether-10%ethanol at the conditions of P21_M11, P18_M8, P15_M5.

6.3.3.3 Gas-phase emission character of n-heptane-dibutyl ether-ethanol

The CO_2, H_2O, NO_x, NO, CO, THC emissions of diesel and 50%C7H16-40%DBE-10%EtOH at the conditions of P21_M11, P18_M8, P15_M5 are compared in Figure 6.21. The THC determination accounts for the FID response factors by Eq. (6.2) which is recommended by MGRefsMaster Reference Calibrations [455, 461] and the diesel readings are C1 (number of carbon atoms) basis. The CO_2 emission from diesel is higher than 50%n-heptane-40%dibutyl ether-10%ethanol which indicates that the diesel obtains higher combustion completeness. The 50%n-heptane-40%dibutyl ether-10%ethanol obtains lower NO_x and NO emissions than diesel because the introduction of ethanol drops down the peak combustion temperature. The 50%n-heptane-40%dibutyl ether-10%ethanol emits comparable CO with diesel but produces a larger amount of unburned THC than diesel. The CH_4, C_2H_6, C_2H_2, C_2H_4, C_3H_6, CH_2O emissions of diesel and 50%C7H16-40%DBE-10%EtOH at the conditions of P21_M11, P18_M8, P15_M5 are compared in Figure 6.22. The THC produced from 50%C7H16-40%DBE-10%EtOH is greater than that of diesel, therefore, the methane, acetylene, ethylene, ethane, propylene, diesel emissions from 50%C7H16-40%DBE-10%EtOH are also greater than diesel-based Eq. (6.2). The gas-phase emission data in Figure 6.22 (a)~ Figure 6.22 (e) supports this conclusion. The 50%C7H16-40%DBE-10%EtOH produces a large quantity of formaldehyde emission than diesel under the studied conditions because its oxygen mass content accounts for 8.94%.

THC = (1.1×CH4)+(2.4×Acetylene)+(1.9×Ethylene)+(2×Ethane)+(2.85×Propylene)+(1.35×Diesel)   (6.2)
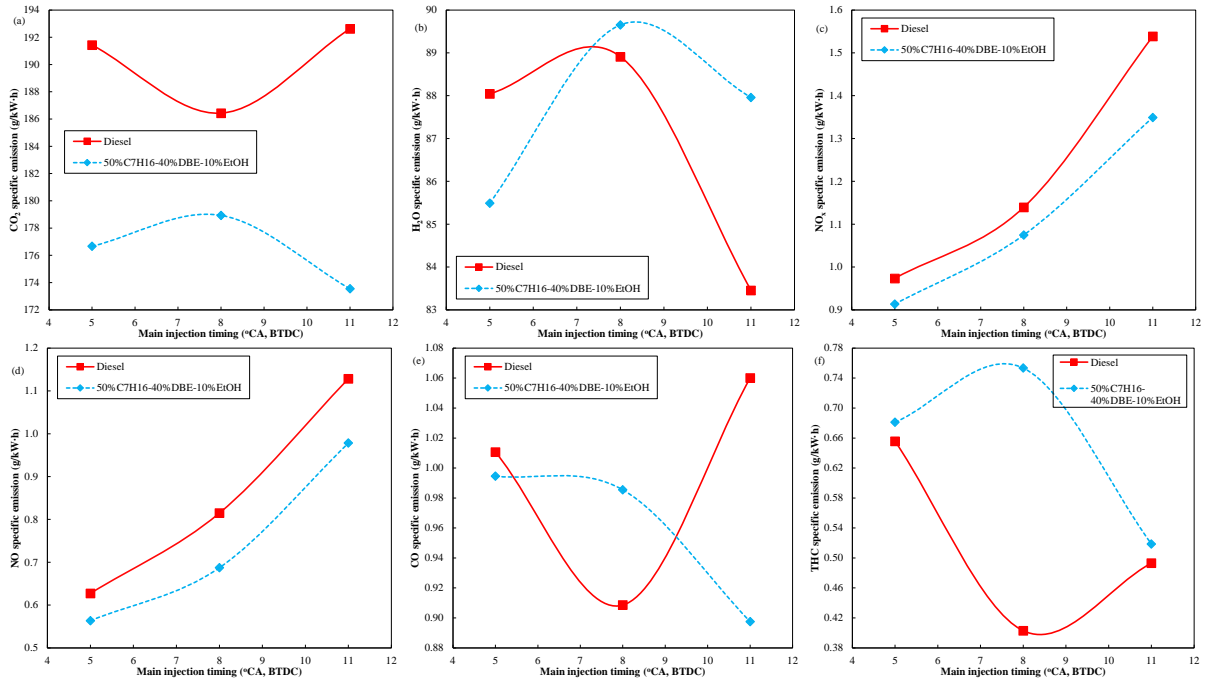
Figure 6.21. CO_2, H_2O, NO_x, NO, CO, THC emissions of diesel and 50%n-heptane-40%dibutyl ether-10%ethanol at the conditions of P21_M11, P18_M8, P15_M5.
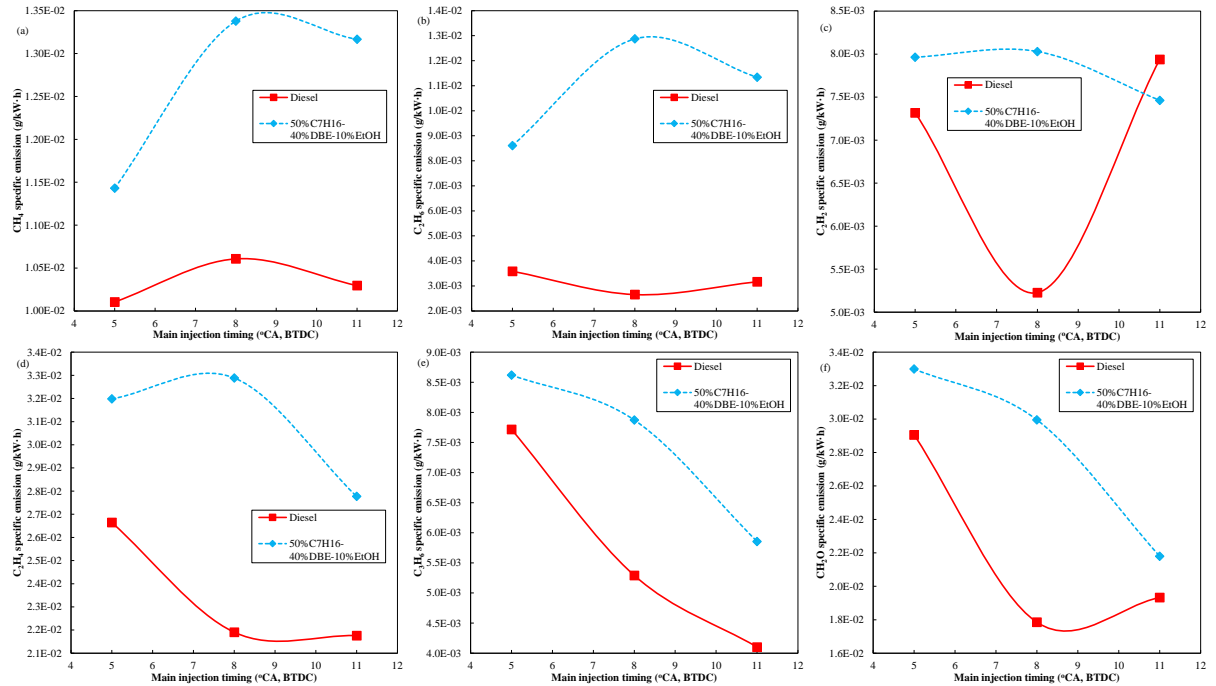
Figure 6.22. CH_4, C_2H_6, C_2H_2, C_2H_4, C_3H_6, CH_2O emissions of diesel and 50%n-heptane-40%dibutyl ether-10%ethanol at the conditions of P21_M11, P18_M8, P15_M5.

6.4 Conclusions

This work adopts machine learning-quantitative structure property relationship (ML-QSPR) and chemical kinetics to develop a high throughput fuel screening approach for internal combustion engines. The fuel screening approach can be used to discover promising transportation fuel blendstocks with drop-in diesel/gasoline base fuel potential which comprises Tier 1 fuel physicochemical property (volatility, atomization, energy density, sooting tendency, ignitability) screening and Tier 2 chemical kinetic screening. Two case studies toward SI and CI engines are showcased to identify fuel components enabling high efficiency and low pollutant emissions.

Tier 1 screening utilizes ML-QSPR to develop 15 fuel properties predictive models of melting point, boiling point, vapor pressure, enthalpy of vaporization, cetane number, research octane number, motor octane number, ignition temperature, flash point, yield sooting index, liquid density, lower heating value, surface tension, lower/upper flammability limit. QSPR-UOB 3.0 functional group classification system containing 42 molecular descriptors is established to convert the molecular feature into a fuel molecular structure matrix. UOB Fuel Property Database is constructed to store the fuel property data of 1742 pure compounds and 465 mixtures. 19 ML algorithms are applied to train the regression models connecting the fuel molecular structure matrix and fuel property matrix. 166 high antiknock and low sooting tendency compounds are identified for SI engines with the characteristics of (1) highly compact, branched molecular structure; (2) containing oxygenic functional groups; (3) containing unsaturated (e.g. CCDB) or cyclic structures and moving at the molecular center; (4) absence of aromatic structure. 129 high reactivity and low sooting tendency compounds are specified for CI engines with the characteristics of (1) lengthening chain length and reducing the number of primary C-H bond; (2) containing acyclic or cyclic ether group; (3) locating CCDB, ester group, aldehyde group, ketone group, branched-chain at the molecular edge; (4) free of cyclic, unsaturated, aromatic structures.

Tier 2 screening utilizes detailed chemical kinetic mechanisms to examine the ignition delay time, φ-sensitivity, laminar flames speed which describe the fuel reactivity, stratified charge sequential ignition capacity, combustion

rate and dilution tolerance respectively. It ranks 8 candidates for SI and CI engines based on the performance of volatility, atomization, energy density, sooting tendency, ignitability. The recommended orders are: (a) diisobutylene > iso-octane > ethanol > cyclopentanone > cyclopentanol > 2,5-dimethylfuran > toluene > 2-methylfuran for SI engines; (b) dibutyl ether > n-heptane > n-octane > 2-butyltetrahydrofuran > butylcyclohexane > 1-octanol > $PODE_3$ > $PODE_4$ for CI engines. Tier 2 chemical kinetic screening demands the availability of detailed chemical kinetic mechanisms for the studied molecules, therefore, a certain amount of Tier 1 fuel candidates free of detailed oxidation mechanisms cannot enter Tier 2 screening. An increasing number of well-validated mechanisms emerge as the progress of fuel combustion chemistry which will gradually bridge this knowledge gap and promote model-based fuel screening.

50%n-heptane-40%dibutyl ether-10%ethanol is formulated based on the results of Tier 1, Tier 2 fuel screening for CI engine and it is tested in a single cylinder, naturally aspirated CI engine. The 50%n-heptane-40%dibutyl ether-10%ethanol has two distinct benefits compared to diesel: (1) improving the indicated thermal efficiency by shortening combustion duration; (2) ethanol addition reducing the peak combustion temperature and NOx emission. The 50%n-heptane-40%dibutyl ether-10%ethanol contains 8.94 mass% and results in a greater amount of unburned THC and formaldehyde emission compare to diesel.

Chapter 7 Chemical Kinetic Study on Ignition and Flame Characteristics of Polyoxymethylene Dimethyl Ether 3 (PODE3)

7.1 Introduction

The high throughput virtual screening in chapter 6 recommends 129 compounds for CI engines, particularly the polyoxymethylene dimethyl ether 3 (PODE3) [462, 463] and diethoxymethane (DEM) [464, 465] are the promising carbon-neutral fuels with a high CN of 78 and 57.3 respectively. Thence, the kinetic properties of PODE3 and DEM are explored by detailed chemical kinetic mechanisms in chapter 7 and chapter 8 respectively. The synthesis and potential utilization of PODE3 are reviewed by Liu et al. [466] and Hackbarth et al. [467]. The application of PODE3 to IC engines and its environmental impacts are reviewed by Awad et al. [468]. An overview of the PODE3 production, PODE3 application as a pure fuel or fuel blends to IC engines, fuel spray and atomization characteristics, the benefit of PM reduction by utilizing PODE3, detailed and reduced chemical kinetic mechanisms, life cycle analysis of PODE3 are briefly summarized in Table 7.1.

For the category of PODE3 production, there are three fossil-free synthetic routes: (1) PODE3 synthesizes from trioxane and dimethoxymethane with the catalyst of methanesulfonic acid and Deloxan [469]. (2) PODE3 synthesizes from H_2 and CO_2 [470]. (3) PODE3 synthesizes from methanol derived from recycling CO_2 and H_2 [471]. For the category of PODE3 application as pure fuel in CI engine, it can adapt to both traditional MCCI (mixing controlled compression ignition) combustion mode [472, 473] and HCCI (homogeneous charge compression ignition) combustion mode [474]. But the PODE3 is not suitable to act as a "drop-in" diesel blendstock exceeding 33% mass content or as a pure fuel in CI engines because the PODE3 has poor compatibility with elastomers of the fueling supply system. Kass et al. [475] report that the volume expansion of elastomer exceeds 30% at 33% mass fraction of PODE3 with diesel. Fuorosilicone is the best elastomer material to resist volume swell among their studied materials, specially designed elastomers are needed to adopt PODE3 as diesel

blendstock in high-level bend proportion. Otherwise, the PODE3 blend fraction should maintain at a low level to ensure acceptable material compatibility with existing elastomers in-vehicle infrastructure. Therefore, the majority of the PODE3 engine application is adopted as fuel blendstocks as listed in Table 7.1. The regular use of PODE3 is blending with diesel [476-478] or diesel-biodiesel mixture [479] as an additive with the benefit of CN improvement and PM reduction. The pure PODE3 or PODE3-diesel mixture with high reactivity can blend with low reactivity fuels such as n-butanol [480-482], gasoline [483] for CI engines application. In addition to directly blending PODE3 with low reactivity fuels, RCCI (reactivity controlled compression ignition) combustion mode can be applied which injects low reactivity fuel (such as ethanol [484]) at the intake manifold and directly injects PODE3 into the cylinder. Addition PODE3 into diesel decreases spray tip penetration, Sauter mean diameter (SMD) and increases the spray projected area, spray cone angle [485]. The reduction of droplet SMD indicates that adding PODE3 can improve the spray atomization performance compared to pure diesel. The PODE3 soot reduction effect is quantified by PODE3-ethylene in laminar coflow diffusion flames [486] and burner-stabilized premixed flames [487]. PODE3 does not exist carbon-carbon bond and this chemical effect is the primary cause of soot formation reduction in PODE3-ethylene flames. Similar to ethanol [488], a small amount of PODE3 addition enhances soot formation, the soot volume fraction and particle size reduction benefit can be observed only when the PODE3 fraction exceeds about 10vol.% [486]. The social and environmental impacts are important reasons for PODE3 utilization in IC engine infrastructures because the total life cycle greenhouse gas (GHG) emission of PODE3 (17.76 g CO_2eq/MJ) derived from biomass is about 7.125 times less than diesel (126.54 g CO_2eq/MJ) in 20 year [489]. The well to wheel GHG emission of PODE3 (29 g CO_2eq/km) produced via power to liquid (PTL) technology can be 7.207 times less than diesel (209 g CO_2eq/km) [490].

This work studies the ignition delay times, adiabatic flame temperature (AFT) and premixed laminar flame speed of PODE3 under wide range conditions and compares with diesel surrogate of n-heptane. The heat release

process and rate-controlling reactions of the ignition process as well as flame temperature are revealed by the

chemical kinetic model. The dependence of ignition delay times and laminar flame speed on equivalence ratio,

temperature, pressure are quantified by the Arrhenius equation and machine learning (ML) regression model.

Table 7.1 Overview of the production and utilization of PODE3.

Items	Classification	Brief of research	Ref.
PODE3 production	Route 1	Start from dimethoxymethane and trioxane and uses methanesulfonic acid or Deloxan catalyst	[469]
	Route 2	Produce from hydrogen and carbon dioxide by combining established synthetic process	[470]
	Route 3	Synthesize by methanol derived from H2 and recycled CO2	[471]
Pure fuel in engine	HCCI mode	Combustion and emission characteristics under HCCI combustion mode	[474]
	MCCI mode	Combustion and emission characteristics under MCCI combustion mode	[472, 473]
Fuel blend in engine	PODE3-diesel	Combustion and emission characteristics of PODE3-diesel in CI engine	[476-478]
	PODE3-diesel-n-butanol	Combustion and emission characteristics of PODE3-diesel-n-butanol in CI engine	[480-482]
	PODE3-gasoline	Combustion and emission characteristics of PODE3-gasoline in CI engine	[483]
	PODE3-ethanol	Combustion and emission characteristics of PODE3-ethanol in CI engine	[484]
	PODE3-diesel-biodiesel	Combustion and emission characteristics of PODE3-diesel-biodiesel in CI engine	[479]
Elastomers compatibility	Elastomer materials	Fuorosilicone, neoprene, polyurethane, Epichlorohydrin/ethylene oxide (ECO), Blend of NBR/PVC (OZO), diene monomer (EPDM), silicone, styrene-butadiene rubber (SBR), acrylonitrile, butadiene rubbers (NBR)	[475]
Injection process	PODE3-diesel	Investigate the macroscopic spray and microscopic droplet characteristics of PODE3-diesel	[485]
Soot reduction	PODE3- ethylene	Investigate PODE3 soot reduction performance in PODE3-ethylene laminar co-flow diffusion flames	[486]
	PODE3- ethylene	Investigate PODE3 soot reduction performance in PODE3-ethylene burner-stabilized premixed flames	[487]
Detailed mechanism	Cai et al.	Describe PODE3 autoignition at the low/intermediate/high temperatures, 322 species,1611 reactions	[447]
	He et al.	Mimic the ignition characteristics at low temperature, 225 species, 1082 reactions	[491, 492]
	Sun et al.	Depict the high temperature flame chemistry of PODE3, 274 species, 1674 reactions	[493]
Reduced mechanism	Ren et al.	Predict PODE3 combustion and soot formation for engine application, 145 species, 585 reactions	[494]
	Lin et al.	Predict in-cylinder pressure and HRR for HCCI engine, 61 species and 190 reactions	[495]
	Huang et al.	PODE3/nature gas dual fuel mechanism for CFD simulation, 124 species, 650-reaction	[496]
Life cycle analysis	Biomass-derived PODE3	Evaluate the life cycle GHG emission of PODE3 (17.76 g CO2eq/MJ) and diesel (126.54 g CO2eq/MJ)	[490]
	PODE3	Compare the well to wheel GHG emission of PODE3 (29 g CO2eq/km) with diesel (209 g CO2eq/km)	[490]

7.2 Chemical kinetic model development and validation

The He et al. [491, 492] mechanism (225 species, 1082 reactions) describes PODE3 oxidation at the low to the intermediate temperature of 588K~1111K while the Sun et al. [493] mechanism (274 species, 1674 reactions) interprets the high temperature flame phenomenon at the temperature of 1500K or above. To predict the ignition and oxidation characteristics of PODE3 at a wide range temperature of 550~1250K, a modified PODE3 chemical kinetic mechanism is developed by merging the He et al. [491, 492] mechanism and the Sun et al. [493] mechanism. The He et al. [491, 492] mechanism is regarded as a master mechanism that is used as the primary source of species and reactions. The Sun et al. [493] mechanism is placed as a donor mechanism to act as the second source of species and reactions. For the overlapped elementary reactions in both mechanisms, the reaction rate constants from the master mechanism are set as default. To reduce the mechanism complexity and focus on the PODE3 oxidation behavior, the sub-mechanisms of decane, dodecane, hexadecane, diisobutylene, cyclohexane, methylcyclohexane, PAH are removed from the modified PODE3 mechanism. The premixed laminar flame speed is modeling by adopting the Sun et al. [493] mechanism to avoid interaction between the low temperature reactions and the high temperature chemistry which may reduce the convergence rate and predictive accuracy.

The modified PODE3 mechanism contains 271 species, 794 reactions and it is validated against the measured ignition delay times by RCM [491, 492] at varied conditions Tof φ=0.5, 1.0, 1.5 as shown in Figure 7.1 and the inputs used for Chemkin simulation are provided in Table 7.2. The predictive ignition delay times almost coincide with the measured values under fuel-lean, stoichiometric and fuel-rich conditions. The predictive ignition delay time profiles of φ=1.5 have an abnormal plateau at 1050K as shown in Figure 7.1 (c) which may be an incipient NTC behavior under fuel-rich, high dilution conditions. The PODE3 kinetic model also obtains a good agreement (R^2=0.9895) of premixed laminar flame speed between measured and simulated values at the condition of T_{init}=408K, P_{init}=1atm as shown in Figure 7.1. The model inputs of Chemkin simulation for premixed laminar flame speed are presented in Table 7.3. The Sun et al. [493] mechanism overestimates the flame speed at the full equivalence ratio from 0.7 to 1.6 and the maximal deviation reaches 9.58% at φ=0.7. The relative deviation decreases as increasing equivalence ratio and it drops below 5% at φ=1.0~1.6.
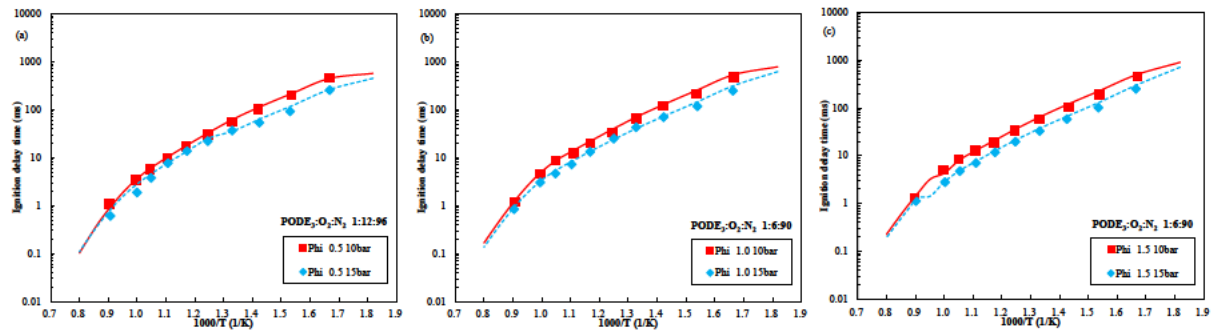
Figure 7.1. Ignition delay times of PODE3 at six different conditions, (a) $\varphi = 0.5$, PODE3:O_2:N_2=1:12:96, (b) φ =1.0, PODE3:O_2:N_2=1:6:90, (c) φ =1.5, PODE3:O_2:N_2=1:4:80; symbols, experimental data [491, 492]; lines, simulations.
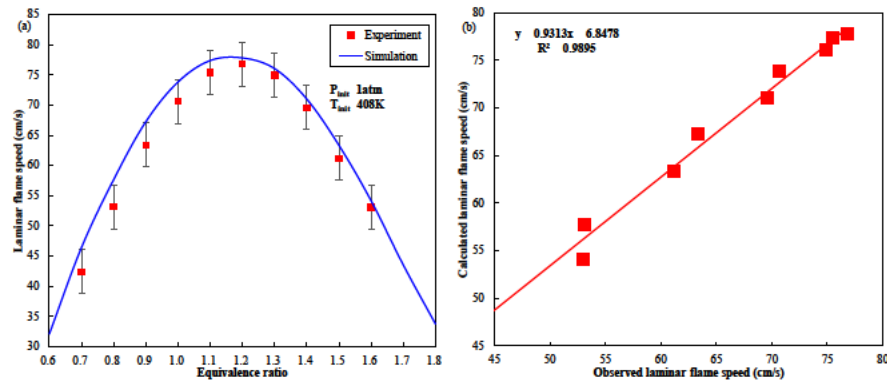
Figure 7.2. Premixed laminar flame speed of PODE3-air mixture at T_{init}=408K, P_{init}=1atm; symbols, experimental data [493]; lines, simulations.

Table 7.2. Inputs used for Chemkin simulation of ignition delay times for PODE3 in closed homogeneous batch reactor module

Item	Value
Problem type	Constrain pressure and solve energy equation
End time (s)	1
Temperature (K)	550~1250
Pressure (bar)	10/15
Volume (m^3)	Default
Heat loss (cal/s)	0
Surface temperature (°C)	Same as gas temperature
Equivalence ratio φ	0.5/1.0/1.5

Table 7.3. Inputs used for Chemkin simulation of premixed laminar flame speed for PODE3

Item	Value
Unburned gas temperature (K)	408
Pressure (atm)	1
Ambient temperature (K)	298
Maximum number of grid points allowed	350
Number of adaptive grid points	10
Adaptive grid control based on solution gradient	0.15
Adaptive grid control based on solution curvature	0.5
Starting axial position (mm)	0
Ending axial position (mm)	3
Initial stream inlet velocity (cm/s)	40
Equivalence ratio φ	0.6~1.8
Fuel mixture	PODE3
Oxidizer mixture	O_2/N_2 = 21/79 vol.%
Complete combustion products	CO_2/H_2O/N_2

7.3 Results and discussion

7.3.1 Ignition characteristics of PODE3

7.3.1.1 Ignition delay times, species evolution, dominant reactions and reaction pathway

The ignition delay times of PODE3 are compared with n-heptane at φ =0.25, 0.5, 1.0, 2.0, T_{init} = 550~1250K,

P_{init}=6.5, 13.5, 20, 45atm to evaluate their fuel reactivity as shown in Figure 7.3 and the model inputs for Chemkin

simulation is presented in Table 7.4. The studied conditions are recommended by Zhang et al. [454] to test the fuel

ignition characteristics covering low/intermediate/high temperature regimes by shock tube and rapid compression

machine at engine relevant conditions. The results of ignition delay time include 5 aspects: (1) The PODE3 is more

reactive than n-heptane at the full temperature range of 550~1250K given that it obtains lower ignition delay times.

(2) The fuel reactivity of both fuels enhance with increasing initial pressure and equivalence ratio. (3) PODE3 does

not exist an NTC behavior and exhibits a flat plateau at the intermediate temperature of 750~1050K while the n-

heptane demonstrates a distinct NTC behavior at 850~1050K. The dependence of ignition delay times on

temperature for PODE3 is monotonical and it favors the control of advanced low temperature combustion modes

for IC engines. (4) The flat plateau of PODE3 and NTC of n-heptane expand in temperature range and intensify in

the intensity with increasing equivalence ratio. (5) The flat plateau of PODE3 and NTC of n-heptane narrow in

temperature range and weaken in the intensity with increasing initial pressure.

The species evolution and temperature profile of PODE3 are compared with n-heptane at φ =1.0, T_{init}=600K,

900K, 1250K, P_{init}=20atm as shown in Figure 7.4. The formaldehyde (CH_2O) and hydroxyl radical (OH) are

regarded as the indicators of low temperature heat release (LTHR) and high temperature heat release (HTHR)

respectively. The LTHR and HTHR can be determined by the CH_2O and OH concentrations accompanied by

temperature profiles. The main findings are summarized below: (1) Both PODE3 and n-heptane demonstrate two-

stage heat release at low temperature regimes of 550~750K and 550~850K respectively. The transition from two-

stage heat release to single-stage heat release takes place as increasing temperature. (2) The LTHR intensity of

PODE3 is weaker than that of n-heptane. For example, the LTHR of PODE3 starts at 696K with the amplitude of 176K while the LTHR of n-heptane begins at 770K with the amplitude of 250.6K at φ =1.0, T_{init} =600K, P_{init} =20atm as shown in Figure 7.4 (a) and Figure 7.4 (d). (3) The active radical/species accumulation at LTHR of PODE3 is weaker than that of n-heptane. For example, the H2O2 accumulates 54.62% at LTHR for PODE3 while the n-heptane completes 98.53% H_2O_2 accumulation at LTHR at φ =1.0, T_{init} =600K, P_{init} =20atm as shown in Figure 7.4 (a) and Figure 7.4 (d).

The flat plateau in the PODE3 ignition delay times curve divides the full temperature regime into a low temperature ignition regime of 550~750K, mediate temperature ignition regime at 750~1050K, high temperature ignition regime of 1050~1250K. The simplified reaction pathways of PODE3 and n-heptane are illustrated in Figure 7.5. At low temperature, H atom abstraction from PODE3 on carbon position 1, position 2, position 3 forms PODE3a, PODE3b, PODE3c. 1st O_2 addition to PODE3x takes place to form PODE3xO2 and then isomerizes to form PODE3_OOH. 2nd O_2 addition to PODE3_OOH happens to produce PODE3_OOH_O2 which undergoes chain branching reaction to form ketohyroperoxide and hydroxyl radical. Ketohyroperoxide further decomposes into oxygenated species and hydroxyl radical as shown in Figure 7.5. As the reaction temperature increase to an intermediate temperature, the reaction rate of 2nd O_2 addition reduces and the PODE3_OOH occurs chain propagation reactions to produce cylic ether, ether, aldehyde. The macroscopic reactivity increase with increasing temperature is less impressive and thus a flat plateau appears in the ignition delay times profile. As the reaction temperature further increases to a high temperature regime, the reaction of 1st O_2 addition into PODE3x is inhibited and PODE3x undergoes high temperature decomposition to form R' and carbonyl radical. Meanwhile, the PODE3 molecule occurs high temperature decomposition to break the carbon-oxygen bond and form PODE2x+CH_3O or R+R'. Generally, the oxidation reaction pathway of PODE3 is similar to n-heptane because both of them contains: (1) 1st and 2nd O2 addition reactions, (2) low temperature chain branching reactions, (3) when 2nd O_2 addition is inhibited, the reactants form less reactive products by chain propagation reactions, (4) the fuel molecule takes

place high temperature decomposition as the reaction temperature further increases.

OH radical sensitivity analysis is performed to identify the ignition dominant reactions of PODE3 at φ=1.0, P_{init} =20atm, T_{init} =600K, 725K, 900K, 1075K, 1250K as shown in Figure 7.6 (a) ~ Figure 7.6 (e). The reactions with a positive sensitivity coefficient facilitate OH radical production and ignition propensity and vice versa. At T_{init} =600K, the reactions of 1st O_2 addition to PODE3x and 2nd O_2 addition to PODE3_OOH, promote OH radical production while the reactions of H atom abstraction from carbon position 1 of PODE3 by OH radical, PODE3b decomposition into COCOCOC*O and CH_3 inhibit the OH radical accumulation as shown in Figure 7.6 (a). At T_{init} =725K, the reactions of 2nd O_2 addition to PODE3_OOH, H atom abstract from PODE3 by OH, HO_2 radicals (except PODE3+OH=PODE3a+H_2O), H_2O_2 decomposition to form OH radicals favor OH production while the reactions of H atom abstraction from carbon position 1 of PODE3 by OH radical, PODE3b decomposition into COCOCOC*O and CH_3 inhibit the OH radical accumulation as shown in Figure 7.6 (b). At T_{init} =900K, the reactions of 2nd O_2 addition to PODE3_OOH, H_2O_2 decomposition to form OH radicals accelerate OH production while the reactions of PODE3_OOH=aldehyde+OH, CH_3OCH_2=CH_3+CH_2O suppress OH accumulation in Figure 7.6 (c). At T_{init} =1075K, the reaction sequence of CH_2O+HO_2=HCO+H_2O_2, H_2O_2(+M)=2OH(+M) is the primary OH source while the reaction sequence of CH_3OCH_2=CH_3+CH_2O, $2CH_3$(+M)=C_2H_6(+M) reduces the radical pool and thus restrain OH radical formation as shown in Figure 7.6 (d). At T_{init} =1075K, reaction sequence of CH_2O+HO_2=HCO+H_2O_2, H_2O_2(+M)=2OH(+M) and chain branching reaction of H+O_2=O+OH boost OH radical formation. On the contrary, the reaction sequence of CH_3OCH_2=CH_3+CH_2O, $2CH_3$(+M)=C_2H_6(+M) hinders OH radical accumulation as shown in Figure 7.6 (e).
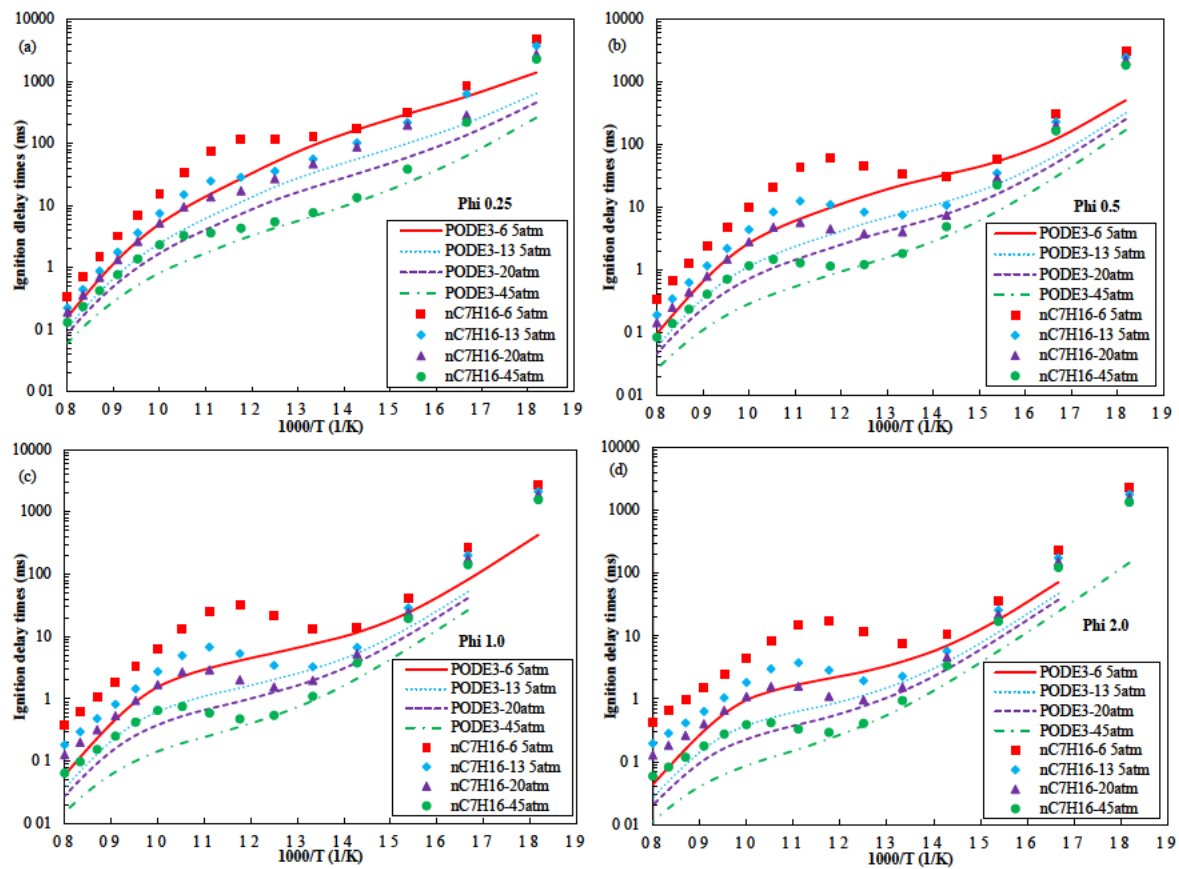
Figure 7.3. Ignition delay times of PODE3-air mixture and n-heptane-air mixture [493] at T_{init}=550~1250K, P_{init}=6.5, 13.5, 20, 45atm; (a) φ=0.25; (b) φ =0.5, (c) φ =1.0; (d) φ =2.0.
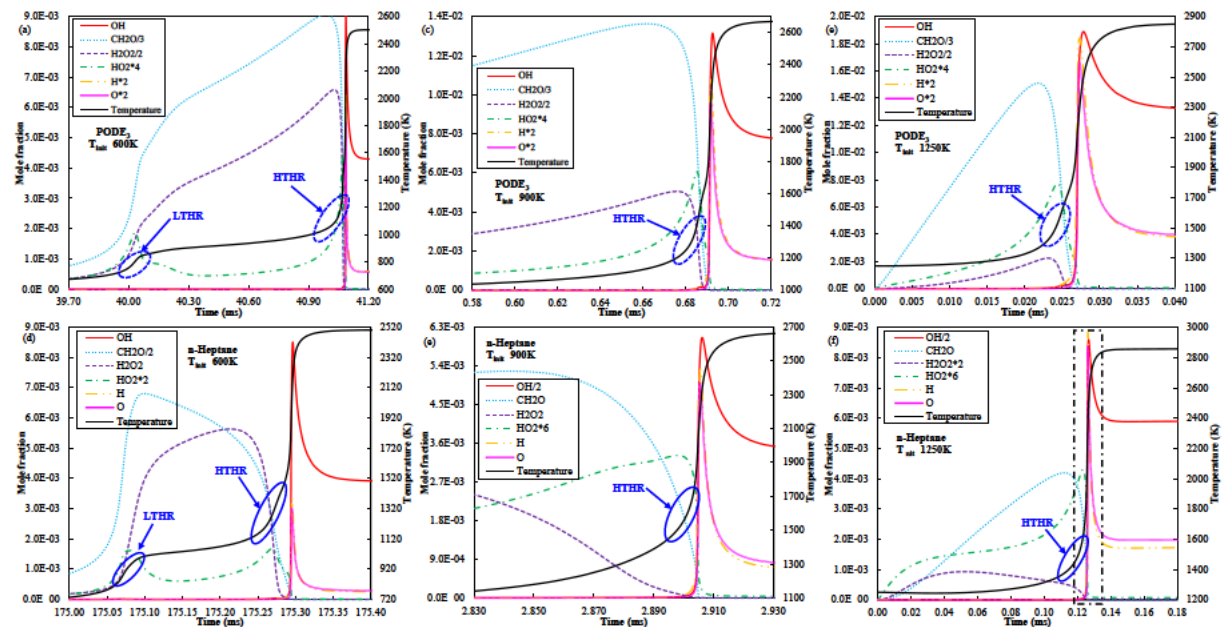
Figure 7.4. Species evolution of PODE3-air and n-heptane-air mixtures at φ =1.0, P_{init}=20atm, (a)/(d) T_{init}=600K, (b)/(e) T_{init} =900K, (c)/(f) T_{init} =1250K.
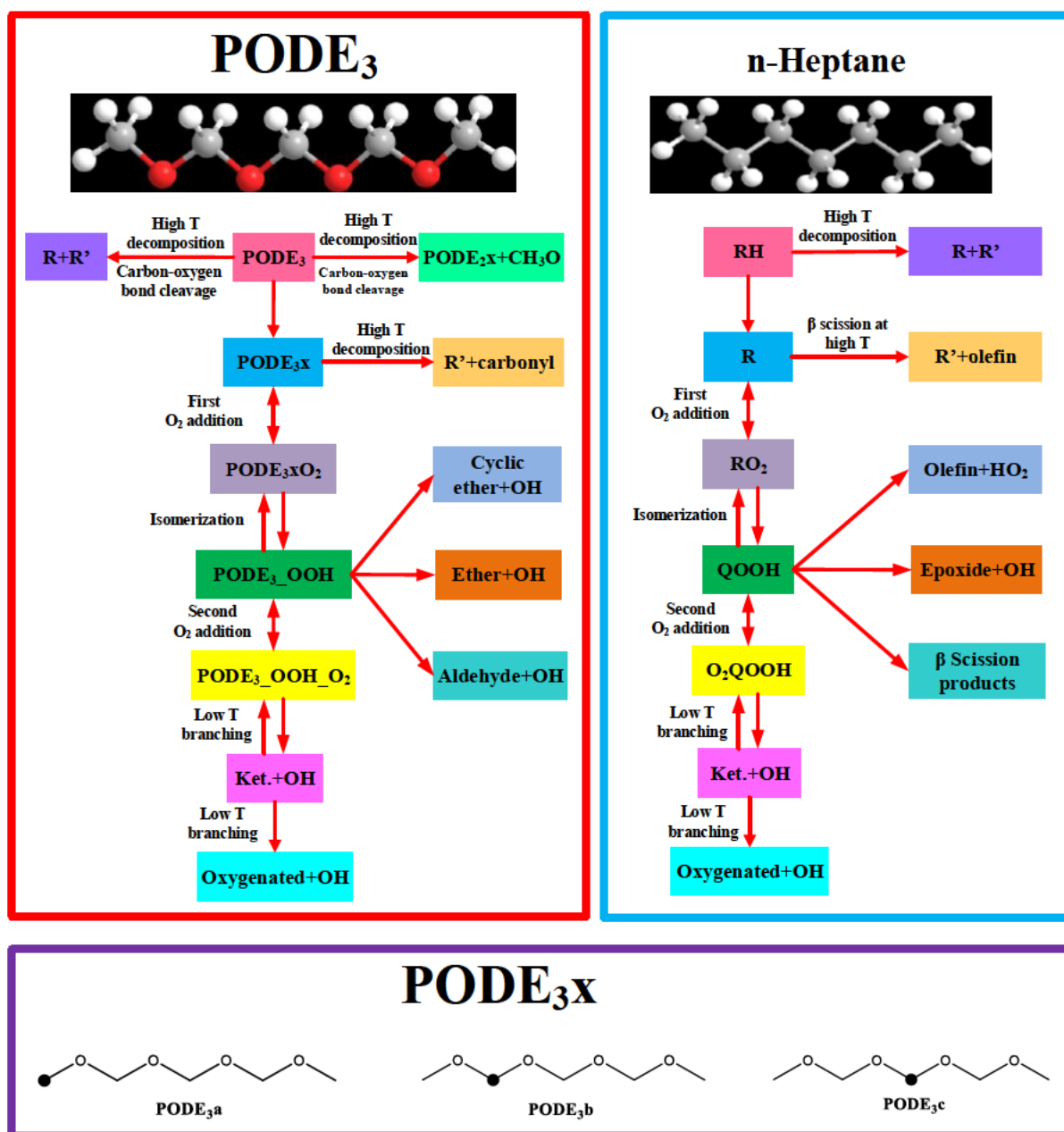
Figure 7.5. Simplified reaction pathways of PODE3, n-heptane [497] and the definition of PODE3x.

Figure 7.6. Sensitivity analysis on the OH species of PODE$_3$-air mixture at φ =1.0, P$_{init}$=20atm, (a) T$_{init}$=600K, (b) T$_{init}$ =725K, (c) T$_{init}$ =900K, (d) T$_{init}$ =1075K, (e) T$_{init}$ =1250K.

Table 7.4. Inputs used for Chemkin simulation of ignition delay times for PODE3 and n-heptane in closed homogeneous batch reactor module

Item	Value
Problem type	Constrain pressure and solve energy equation
End time (s)	10
Temperature (K)	550~1250
Pressure (atm)	6.5/13.5/20/45
Volume (m^3)	Default
Heat loss (cal/s)	0
Surface temperature (°C)	Same as gas temperature
Equivalence ratio φ	0.25/0.5/1.0/2.0
Fuel mixture	PODE3 or n-heptane
Oxidizer mixture	O$_2$/N$_2$ = 21/79 vol.%
Complete combustion products	CO$_2$/H$_2$O/N$_2$

7.3.1.2 Quantification and rank the sensitive factors of ignition delay times

This section aims at quantifying the dependence of ignition delay times on equivalence ratio, temperature, pressure and the number of samples for PODE3 and n-heptane are 234 and 240 as shown in Table 7.5. The normality of the ignition delay times data is tested by the Kolmogorov-Smirnov test and Shapiro-Wilk test simultaneously. The significances for both fuels are far below 0.05 for both tests which indicate that the ignition delay times data disobey normal distribution. Therefore, the non-parametric measure of the Spearman test is adopted to quantify the correlation coefficients instead of the Pearson test which demands dependent variables obey normal distribution. The Spearman correlation coefficients of ignition delay times for PODE3 and n-heptane are shown in Figure 7.7. The factor influence on ignition delay times of PODE3 order from high to low as: temperature>>equivalence ratio>pressure. The order for n-heptane is slightly different from PODE3 as: temperature>> pressure > equivalence ratio. Overall, the temperature is of the utmost importance of ignition delay times for PODE3 and n-heptane.
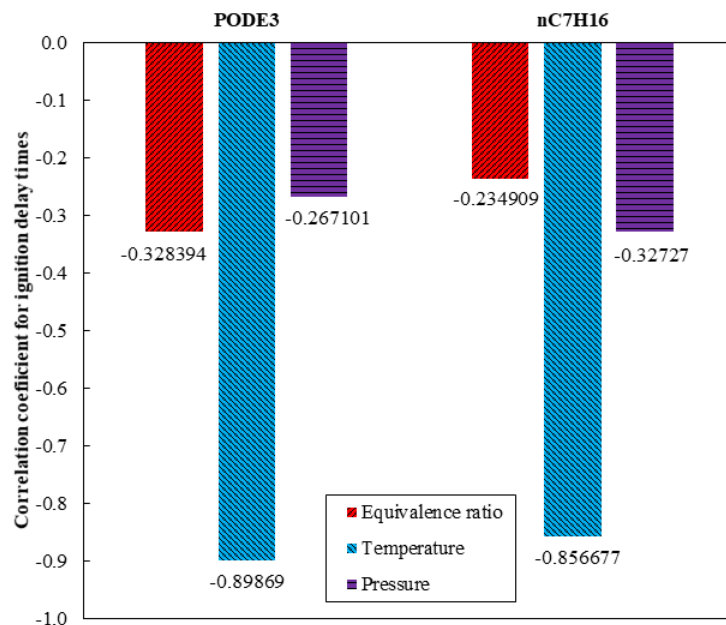
Figure 7.7. Spearman correlation coefficients of ignition delay times for PODE3 and n-heptane.

Table 7.5. Normality test ignition delay times for PODE3 and n-heptane

Item	Kolmogorov-Smirnov* test (PODE$_3$/n-heptane)			Shapiro-Wilk test (PODE$_3$/n-heptane)		
	Statistics	D_f	S_{ig}	Statistics	D_f	S_{ig}
τ_{ig}	0.394/0.406	234/240	6.8891E-105/5.2972E-114	0.280/0.333	234/240	3.0575E-29/1.0933E-28

* Lilliefors significance correction

7.3.1.3 Regression between ignition delay times and equivalence ratio, temperature, pressure

This section aims at fitting an Arrhenius manner formula to describe the dependence of ignition delay time on the equivalence ratio, temperature, pressure as shown in Eq. (7.1) [498]. The coefficients of A, E_a, m, n are computed by iterative optimization to minimize the residual of ignition delay times. The Arrhenius expressions of ignition delay times for PODE3 and n-heptane are provided in Eq. (7.2) and Eq. (7.3) respectively.

$$\tau_{ig} = \frac{1}{\kappa} = A \cdot exp\left(\frac{E_a}{8.314 \cdot T}\right) \cdot p^m \cdot \varphi^n \tag{7.1}$$

where κ, A, E_a, m, n represent global reaction rate of one step mechanism, pre-exponential factor, apparent activation energy, pressure exponent and equivalence ratio exponent respectively. φ, T, p denote the equivalence ratio, mixture initial temperature and mixture initial pressure.

$$\tau_{ig\ PODE_3} = \frac{1}{\kappa_{PODE_3}} = 0.015 \cdot exp\left(\frac{53216.492}{8.314 \cdot T}\right) \cdot p^{-0.96} \cdot \varphi^{-1.143} \tag{7.2}$$

$$\tau_{ig\ nC_7H_{16}} = \frac{1}{\kappa_{nC_7H_{16}}} = 9.881E-08 \cdot exp\left(\frac{112889.509}{8.314 \cdot T}\right) \cdot p^{-0.341} \cdot \varphi^{-0.342} \tag{7.3}$$

where $\varphi \in (0.25, 2.0)$, $T \in (550K, 1250K)$, $p \in (6.5atm, 45atm)$.
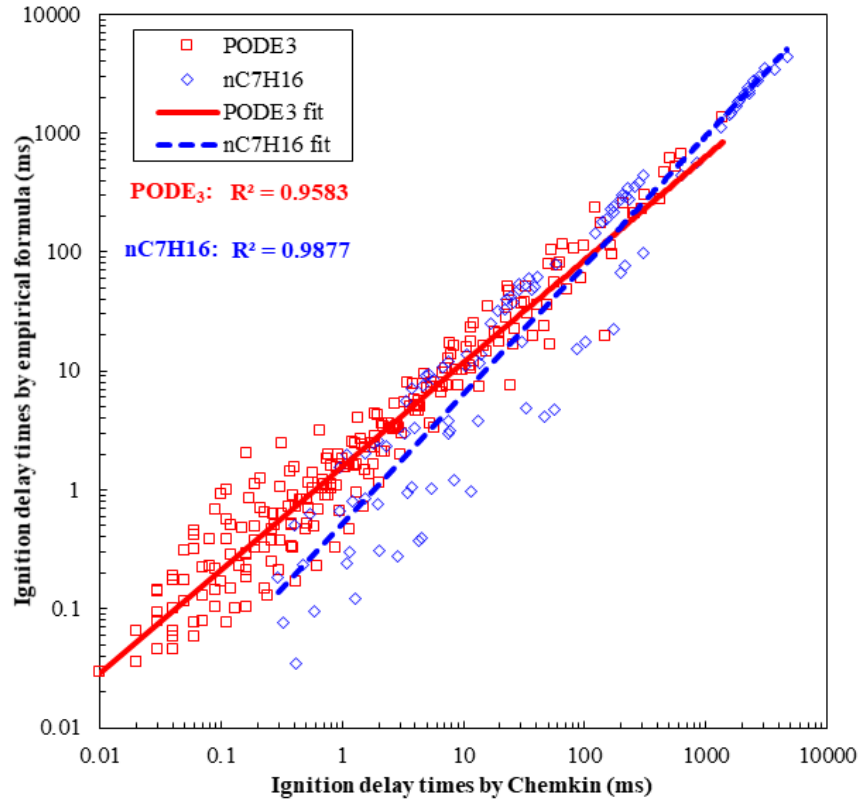
Figure 7.8. Comparison of the ignition delay times between Chemkin predicted values and Arrhenius formula predicted values for PODE3 and n-heptane at φ=0.25/0.5/1.0/2.0, T_{init}=550−1250K, P_{init}=6.5/13.5/20/45atm.

7.3.2 Adiabatic flame temperature (AFT) of PODE3

The adiabatic flame temperature (AFT) is the maximum combustion temperature that can be obtained by the particular gas phase combustibles under specific conditions excluding heat losses, chemical kinetic and mass transportation restrictions [499]. The AFT of PODE3 are compared with n-heptane at $\varphi=0.2\sim2.0$; T_{init}= 298, 358, 398, 458, 518K; P_{init}= 1, 5, 10, 20, 40atm shown in Figure 7.9. The AFT is computed by the chemical and phase equilibrium calculation module of Chemkin software at the condition of constant enthalpy and constant pressure. The test conditions are referred to the research by Zhang et al. [454]. The key findings are summarized below: (1) The AFT of both fuels reach its peaks at $\varphi=1.0$ and they decrease gradually as the mixtures become leaner or richer. (2) The AFT of both fuels increase with initial temperature. (3) The AFT of both fuels are independent of initial pressure. (4) The AFT of PODE3 are lower than those of n-heptane at $\varphi=0.9\sim1.1$ and the PODE3 obtains higher AFT than those of n-heptane at $\varphi=0.2\sim0.9$ and $\varphi=1.1\sim2.0$. At the fuel-lean condition ($\varphi=0.2\sim0.9$), the fuel quantity is the reaction rate-controlling factor. Given that there is a fixed amount of fuel, the PODE3 requires less amount of air than n-heptane to reach an identical equivalence ratio because it contains oxygen atoms in the fuel molecule. Thus, PODE3 reaches higher AFT than n-heptane at fuel-lean condition. At the fuel-rich condition ($\varphi=1.1\sim2.0$), the oxygen quantity is the reaction rate-controlling factor. At a fixed amount of oxygen molecules, larger quantity of PODE3 is needed than n-heptane to reach an identical equivalence ratio. Thence, the AFT of PODE3 are higher than those of n-heptane at fuel-rich condition.
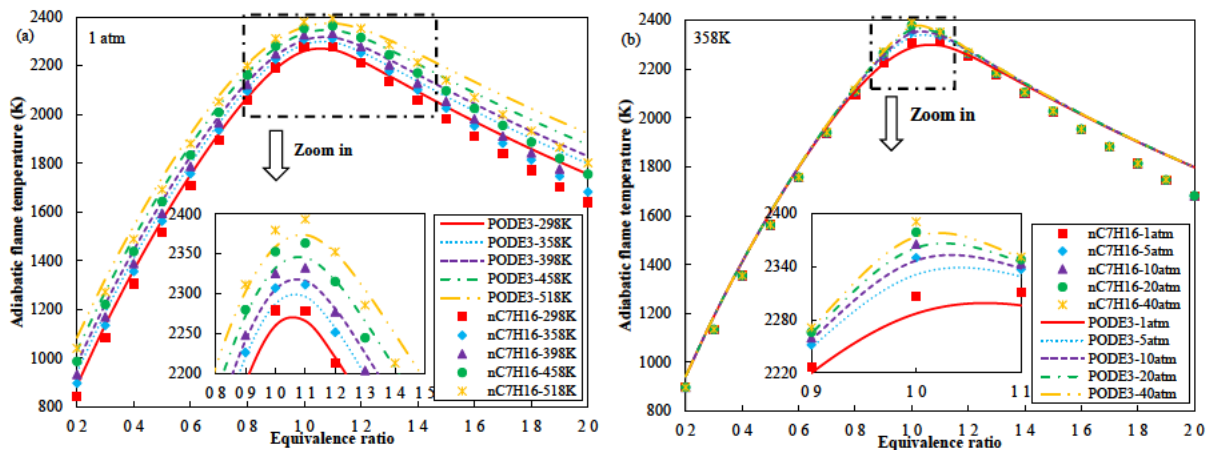
Figure 7.9. Dependence of adiabatic flame temperature on (a) initial temperature; (b) initial pressure for PODE3-air mixture and n-heptane-air mixture at $\varphi=0.2\sim2.0$; T_{init}= 298 358, 398, 458, 518K; P_{init}= 1, 5, 10, 20, 40atm.

7.3.3 Premixed laminar flame speed of PODE3

7.3.3.1 Premixed laminar flame speed and corresponding dominant reactions

The premixed laminar flame speed and flame flux (also known as flame mass burning rate) of PODE3 are compared with n-heptane at t φ=0.2~2.0; T_{init}= 298 358, 398, 458, 518K; P_{init}= 1, 5, 10, 20, 40atm as shown in Figure 7.10. The model inputs for the Chemkin simulation of laminar flame speed are presented in Table 7.6. The flame speed and flame flux are expressed by Eq. (7.4), Eq. (7.5) and the test conditions are the same as section 7.3.2. The key findings are summarized as below: (1) The PODE3 has a higher laminar flame speed than n-heptane at the same condition. (2) The laminar flame speeds of both fuels increase with initial temperature. (3) The laminar flame speeds of both fuels decrease with increasing initial pressure while the laminar flame fluxes increase with pressure.

The global reaction rate, adiabatic flame temperature through the flame zone, and thermal diffusivity of the unburned gas are the three principal influence factors of premixed laminar flame speed as expressed in Eq. (7.4) [500]. The result of section 7.3.1.3 indicates that PODE3 has a higher global reaction rate than n-heptane due to higher pre-exponential factor and smaller activation energy as shown in Eq. (7.2) and Eq. (7.3). The result of section 7.3.2 indicates that the PODE3 obtains higher AFT than n-heptane at the fuel-lean condition (φ=0.2~0.9) and the fuel-rich condition (φ=1.1~2.0). At φ=0.9~1.1, the AFT of PODE3 is about 5~20°C lower than n-heptane at the studied conditions but they are broadly comparable as shown in Figure 7.9. The thermal diffusivity of PODE3 is lower than that of n-heptane as shown in Figure 7.11. Given that the laminar flame speed of PODE3 is higher than n-heptane which must be attributed to the higher global reaction rate and higher AFT of PODE3 than n-heptane.

$$S_L = \frac{\lambda}{\rho \cdot C_p} \cdot \frac{T_f - T_i}{T_i - T_0} \cdot \frac{1}{\delta_r} = \sqrt{\frac{\lambda}{\rho \cdot C_p} \cdot \frac{T_f - T_i}{T_i - T_0} \cdot \frac{d\varepsilon}{dt}} \propto \sqrt{\alpha \cdot T_f \cdot k} \propto e^{\frac{-E_a}{2R_u T}} \propto p^{\frac{n-2}{2}}$$

(7.4)

where S_L, λ, ρ, C_p are the premixed laminar flame speed, the thermal conductivity of unburned combustible, density and constant pressure specific heat capacity of unburned combustible. T_0, T_i, T_f

represent the unburned gas temperature, combustible ignition temperature and flame temperature. δ_r, ε, α denote the length of the reaction zone, reaction progress variable and thermal diffusivity. Both $\dfrac{d\varepsilon}{dt}$ and k are the reaction rate. E_a, R_u, n are the activation energy, universal gas constant and the order of the chemical reaction.

$$\dot{m} = \rho \cdot S_L \tag{7.5}$$

where $\dot{m}$ denotes the flame mass burning rate [501], also known as laminar flame flux [502].

Increasing the initial temperature has a positive effect on the laminar flame speed while increasing the initial pressure has a negative effect which are expressed as $S_L \propto e^{\frac{-E_a}{2R_uT}}$ and $S_L \propto p^{\frac{n-2}{2}}$ as shown in Eq. (7.4). The dependence of laminar flame speed on temperature and pressure for both PODE3 and n-heptane are quantified as Eq. (7.6) ~ Eq. (7.9). Eq. (7.6) ~ Eq. (7.9) reveal two findings: (1) the laminar flame speed of PODE3 is less temperature-dependent than n-heptane. (2) The global reaction orders of PODE3 and n-heptane are 1.250 and 1.222 respectively which indicates that PODE3 is more reactive than n-heptane. It further supports the conclusion in section 7.3.1.

$$S_{L,\,PODE_3} \propto e^{\frac{-E_a}{2R_uT}} = e^{\frac{m}{T}} = e^{\frac{-678.516}{T}} \tag{7.6}$$

$$S_{L,\,nC_7H_{16}} \propto e^{\frac{-E_a}{2R_uT}} = e^{\frac{m}{T}} = e^{\frac{-793.542}{T}} \tag{7.7}$$

$$S_{L,\,PODE_3} \propto p^{\frac{n-2}{2}} = p^{\frac{1.250-2}{2}} = p^{-0.375} \tag{7.8}$$

$$S_{L,\,nC_7H_{16}} \propto p^{\frac{n-2}{2}} = p^{\frac{1.222-2}{2}} = p^{-0.389} \tag{7.9}$$

Flame temperature is the dominant factor to influence the laminar flame speed, the flame temperature sensitivity analysis is conducted at $\varphi = 1.0$, $T_{init} = 298/358/398/458/518K$, $P_{init} = 1/5/10/20/40$ atm to identify the key reactions affecting laminar flame speed as shown in Figure 7.12 and Figure 7.13. The reactions with positive sensitivity coefficients favor flame temperature elevation and vice versa. $H+O_2=O+OH$ is the chain branching reaction at high temperature and $CO+OH=CO_2+H$ is the principal exothermic reaction, both of them favor the flame temperature elevation and laminar flame speed.
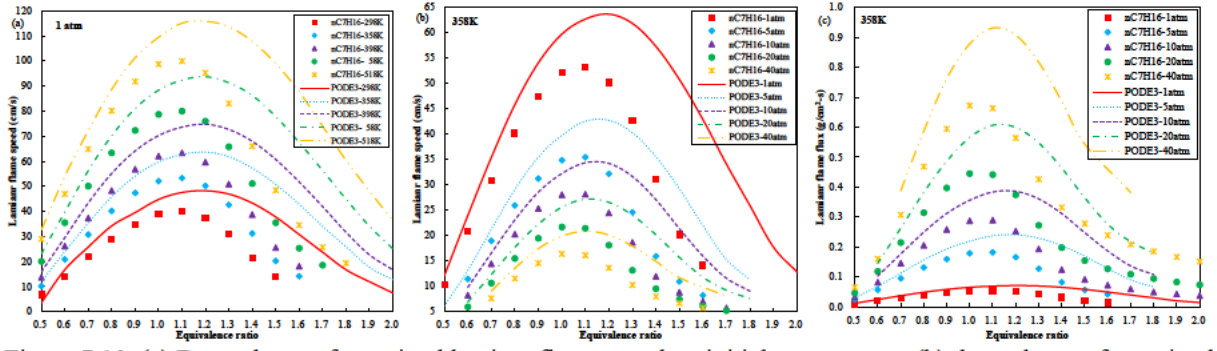
Figure 7.10. (a) Dependence of premixed laminar flame speed on initial temperature; (b) dependence of premixed laminar flame speed on initial pressure; (c) dependence of premixed laminar flame flux on initial pressure for PODE3-air and n-heptane-air [454].
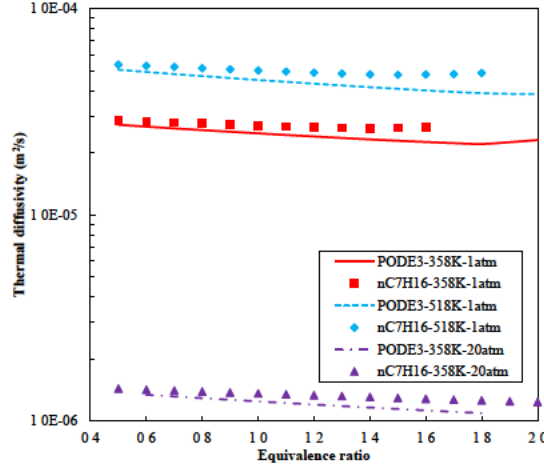
Figure 7.11. Thermal diffusivity of PODE3 and n-heptane at (a) 358K, 1atm; (b) 518K, 1atm; (c) 358K, 20atm.
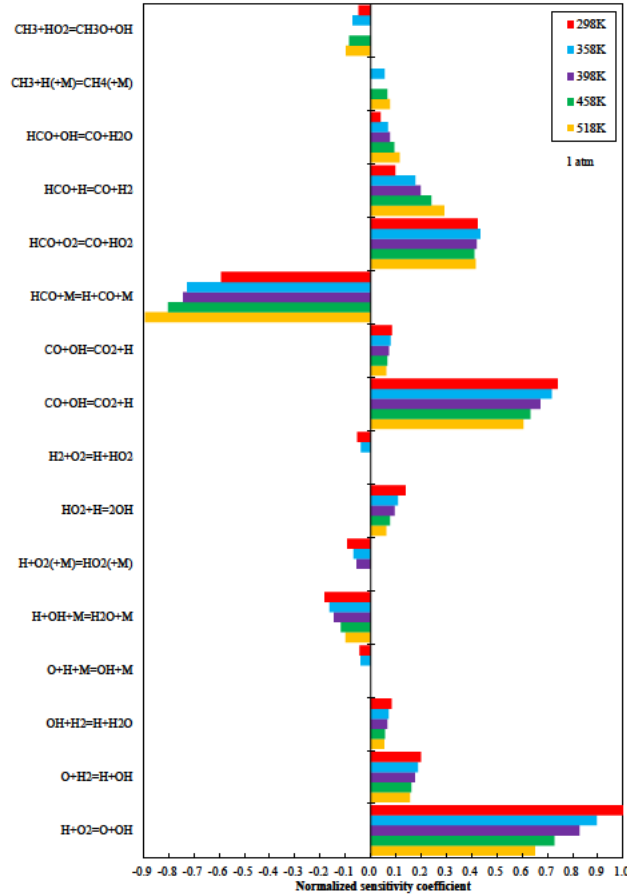
Figure 7.12. Sensitivity analysis on the flame temperature of PODE3-air mixture at φ=1.0, T_{init}=298/358/398/458/518K, P_{init}=1atm.
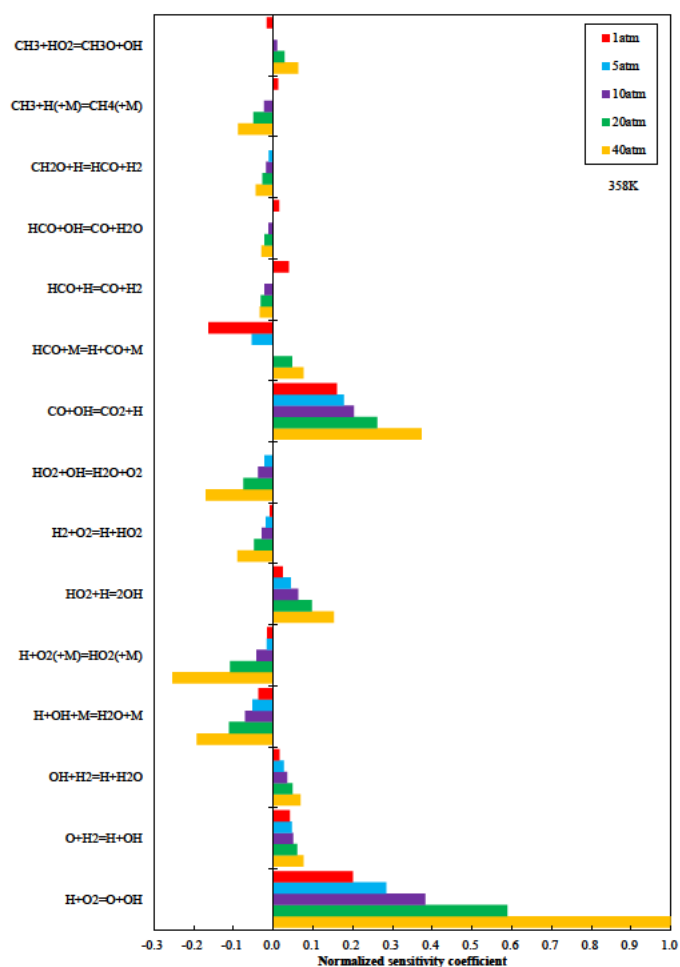
Figure 7.13. Sensitivity analysis on the flame temperature of PODE3-air mixture at $\varphi=1.0$, $T_{init}=358K$, $P_{init}=1/5/10/20/40atm$.

Table 7.6. Inputs used for Chemkin simulation of premixed laminar flame speed for PODE3 and n-heptane

Item	Value
Unburned gas temperature (K)	298/358/398/458/518
Pressure (atm)	1/5/10/20/40
Ambient temperature (K)	298
Maximum number of grid points allowed	1000
Number of adaptive grid points	10
Adaptive grid control based on solution gradient	0.1
Adaptive grid control based on solution curvature	0.5
Starting axial position (mm)	0
Ending axial position (mm)	3
Initial stream inlet velocity (cm/s)	40
Equivalence ratio φ	0.5~2.0
Fuel mixture	PODE3/n-heptane
Oxidizer mixture	$O_2/N_2 = 21/79$ vol.%
Complete combustion products	$CO_2/H_2O/N_2$

7.3.3.2 Regression between laminar flame speed and equivalence ratio, temperature, pressure

Unlike the relationship between ignition delay times and equivalence ratio, temperature, pressure, the Arrhenius type formula cannot accurately describe the dependence of laminar flame speed on equivalence ratio, temperature, pressure. To predict the laminar flame sped at particular conditions, a machine learning algorithm is utilized to build the regression model between laminar flame speed and equivalence ratio, temperature, pressure. for PODE3 and n-heptane. The models are accessible in Data and Software Availability section. The model functions and predictive errors are listed in Table 7.7. The machine learning models achieve high predictive accuracy with R2 of 0.999978 for PODE3 and 0.999998 for n-heptane as shown in Figure 7.14 (a). The PODE3 (RMSE=0.127) has greater RMSE than n-heptane (RMSE=0.03) which further supports that the n-heptane regression model obtains better predictive accuracy as shown in Table 7.7. The machine learning regression model for PODE3 has larger predictive residuals at the range of 0~50 cm/s than at the range of 50~120 cm/s as shown in Figure 7.14 (b).
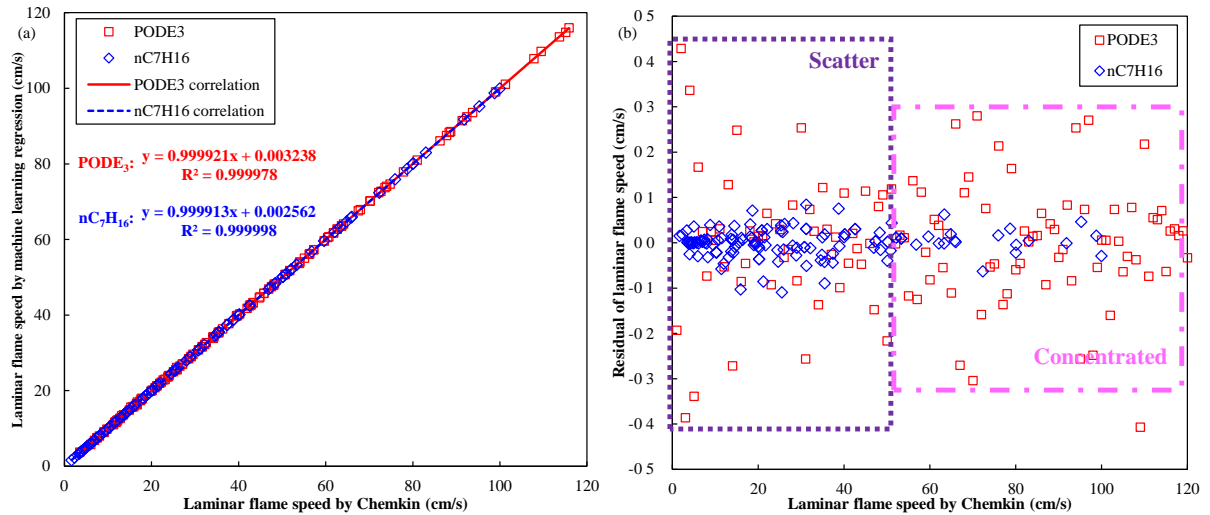
Figure 7.14. (a) Comparison of laminar flame speed between Chemkin computation and machine learning regression model at $\varphi=0.5\sim2.0$, $T_{init}=298\sim518K$, $P_{init}=1\sim40atm$; (b) Residual of the predicted laminar flame speed.

Table 7.7. Model function and regression metrics of the machine learning regression model for PODE3 and n-heptane

Property	Training dataset	Model type	Kernel function	Basic function	R^2	MAE	MSE	RMSE
Laminar flame speed	PODE3 (146)	Gaussian process regression	Matern52	Linear	0.999978	0.087	0.016	0.127
	n-Heptane (134)	Gaussian process regression	Matern32	None	0.999998	0.021	0.001	0.03

7.4 Conclusions

This work comprehensively examines the ignition characteristics, adiabatic flame temperature, premixed laminar flame speed of PODE3 at engine relevant conditions and compares with the commonly used diesel surrogate of n-heptane. A modified PODE3 with 271 species, 794 reactions is developed to predict the low/intermediate/high temperature oxidation characteristics by merging the He et al. [491, 492] mechanism (225 species, 1082 reactions) and Sun et al. [493] mechanism (274 species, 1674 reactions). The proposed PODE3 mechanism removes sub-mechanisms of decane, dodecane, hexadecane, diisobutylene, cyclohexane, methylcyclohexane, PAH existing in the He et al. [491, 492] mechanism. Thus current PODE3 mechanism is faster and requires less computational resource than He et al. [491, 492] mechanism with comparable predictive accuracy. The major conclusions are summarized as below:

(1) PODE3 has a higher global reaction rate than n-heptane and thus results in lower ignition delay times. The ignition delay times of PODE3 and n-heptane can be expressed as

$$\tau_{ig\ PODE_3} = \frac{1}{\kappa_{PODE_3}} = 0.015 \cdot exp\left(\frac{53216.492}{8.314 \cdot T}\right) \cdot p^{-0.96} \cdot \varphi^{-1.143}$$

and

$$\tau_{ig\ nC_7H_{16}} = \frac{1}{\kappa_{nC_7H_{16}}} = 9.881E - 08 \cdot exp\left(\frac{112889.509}{8.314 \cdot T}\right) \cdot p^{-0.341} \cdot \varphi^{-0.342}$$

respectively. Temperature is the dominant factor of ignition delay times.

(2) The PODE3 may not suitable to act as the chemical ignition source of fuel mixtures because of insufficient temperature increasing amplitude and active radical accumulation during low temperature heat release (LTHR) compared to n-heptane.

(3) PODE3 has higher AFT than n-heptane at fuel-lean condition (φ=0.2~0.9) and fuel-rich condition (φ=1.1~2.0), but it is 5~20°C lower than n-heptane at φ=0.9~1.1.

(4) PODE3 has a higher laminar flame speed than n-heptane at t φ=0.2~2.0; T_{init}= 298~518K; P_{init}= 1~40 atm because of higher global reaction rate and AFT. The high laminar flame speed of PODE3 enhances the dilution tolerance and enables an application to low temperature combustion modes with a high dilution ratio.

Chapter 8 Chemical Kinetic Modeling of Diethoxymethane Oxidation: A Carbon-Neutral Fuel

8.1 Introduction

United Nations (UN) adopts Paris Agreement in 2015 in response to the threat of climate change by limit global warming to well below 2°C, preferably to 1.5°C compared to the pre-industrial era [503]. To achieves this long-term temperature rise target, involved parties/countries commit to reaching greenhouse gas emissions (GHG) as soon as possible and ultimately arrive at global climate neutrality by mid-21-century. Carbon neutrality targets, low-/zero- carbon solutions and corresponding markets have been constructed in numerous countries, regions, and corporations since the agreement entered into force. Massively actions have been implemented in energy and transport and profound impacts have been achieved in R&D and business.

In the energy sector, a concept of "Power to X" (also known as energy to X) is proposed to utilize renewable energy (such as sunlight, wind energy) abundant molecules (CO_2, H_2O, N_2) to produce fuels as an energy carrier and chemicals. Especially the fuels synthesized from renewable feedstocks and produced by sustainable energy/electricity are called carbon-neutral fuels which have zero net GHG emissions and carbon footprint. Biofuels derived from biomass (such as lignocellulose, algae, etc.) [264, 407-410] or waste plastics [407, 504-506] also belong to carbon-neutral fuels. An overview of the research and innovation initiatives toward carbon neutrality is presented in Table 8.1. SUNERGY project [507] converts renewable energy (sunlight energy, wind energy) and abundant molecules (CO_2, H_2O, N_2) into fossil-free fuels and chemicals for climate neutrality. SUNRISE project [508, 509] and Energy-X project [510, 511] are the sub-initiatives of SUNERGY project [507]. Solar Fuels Networks project [512] utilizes solar energy and abundant resource (CO_2, H_2O) to produce carbon-neutral fuels. Power-to-X (Kopernikus project, Germany) project [513, 514] transforms renewable electricity into fuels, gases, heat, chemicals, plastics, cosmetics. Power-to-X (European Commission) project [515] evaluates the technological competitiveness of power-to-gas and power-to-liquid and compares them with other low-carbon techniques by

2050. Co-optimization of fuels and engines [267, 415, 419, 516, 517] regards fuels as dynamic design variants to optimize engine combustion and emission performance. Fuel science center [264, 518, 519] converts renewable electricity and alternative carbon feedstock into liquid fuels as energy carriers for carbon-neutral and ultra-low pollutant emission propulsion systems. Renewable fuels are biofuels derived from inedible biomass and e-fuels from carbon capture and storage (CCS).

The sustainable carbon cycle is illustrated in Figure 8.1 which contains carbon-neutral fuels production from renewable electricity and abundant molecules (CO_2, H_2O) and carbon-neutral fuels utilization in power and transportation sectors [520]. Synthesis gas of CO and H2 is the hub of the sustainable carbon cycle, which is the platform chemical for the target carbon-neutral fuels. Therein, the H2 is generated by water splitting electrolysis through hydrogen evolution reaction and oxygen evolution in alkaline and acidic media [521, 522]. The renewable electricity for water electrolysis is supplied by solar panels and wind turbines. The CO can be prepared by CO_2 electrocatalysis or photocatalysis process [523, 524] while the CO_2 is obtained by carbon capture and storage (CCS) technology [525, 526]. The CO2 is captured from biomass, ambient air, power plant emissions and the technologies reaching demonstration or commercial scale include post-combustion amines, pre-combustion natural gas processing, membranes polymeric, pre-combustion integrated gasification combined cycle (IGCC)+CCS, post-combustion adsorption, bioenergy with CCS (BECCS) industry, direct air capture (DAC), oxy-combustion coal power plant [525]. The demonstration or commercial level carbon storage technologies compose of saline formations, CO_2- enhanced oil recovery (EOR), CO_2- enhanced gas recovery (EGR), depleted oil & gas fields [525].

Straight chain ethers usually have satisfactory cetane numbers and soot reduction benefits due to high oxidation reactivity and existing oxygen atoms, thus they are used as diesel blendstocks to adjust the fuel reactivity and oxygen mass content. An overview of the C3, C5, C7, C9, C11 alkanes and corresponding ethers with 1~5 O-atom substituents is provided in Table 8.2 and those with citations indicate detailed chemical kinetic mechanisms are

available. Diethoxymethane (DEM) is getting more and more attention due to its good reactivity (CN: 57.3 [193]) and low sooting tendency (YSI: 18.5 [321, 370]). DEM is one of the 129 compounds for CI engines recommended by the virtual fuel screening in chapter 6. This chapter intends to propose a detailed kinetic mechanism including low and high temperature chemistry and explore the ignition delay time and laminar flame speed. There are two published DEM mechanisms ahead of this work, one interprets the DEM high temperature fuel-rich flame at sub-atmospheric pressure [527] and the other emulates the high temperature ignition delay times from fuel-lean to fuel-rich conditions [528]. But the scopes of these mechanisms are limited to high temperature (>1100K) oxidation, they are unable to describe the low and mediate temperature oxidation conditions due to the lack of a low temperature reaction scheme. To address this deficiency, this work proposes a detailed chemical kinetic mechanism of diethoxymethane (DEM) including low and high temperature reaction schemes to describe the ignition process at full temperature regime and high temperature premixed laminar flame.
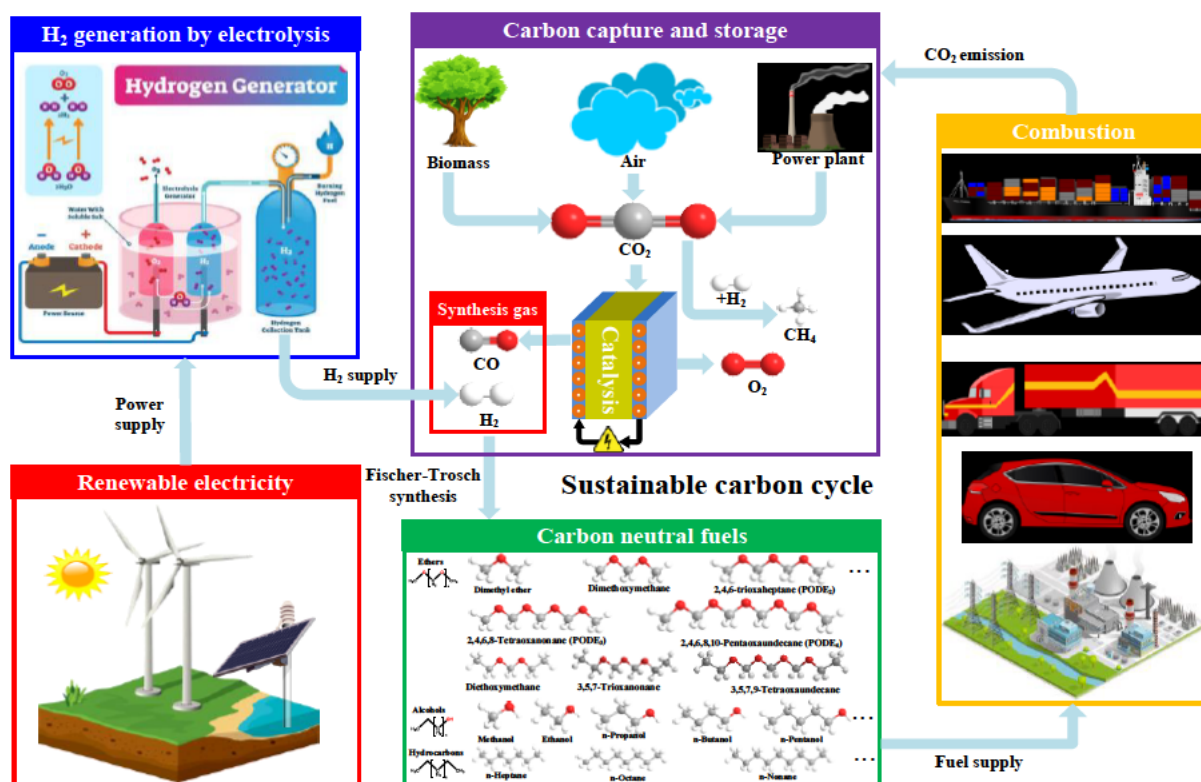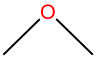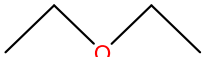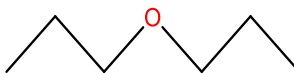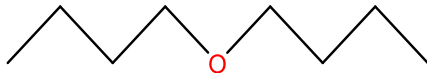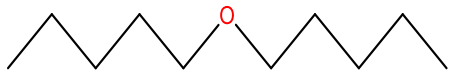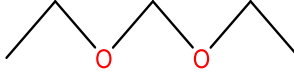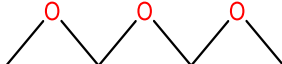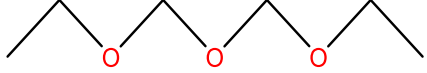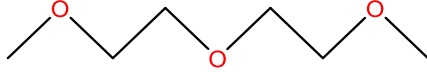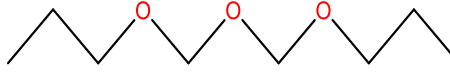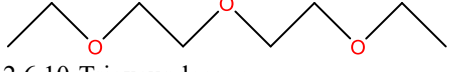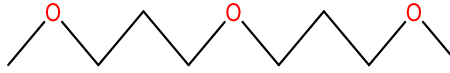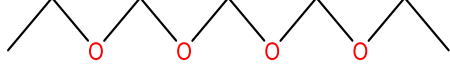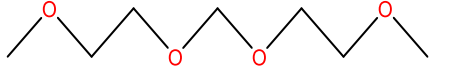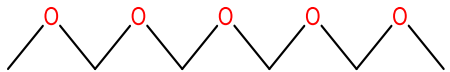
Figure 8.1. Carbon neutral fuels production and utilization in a sustainable carbon cycle.

Table 8.1 Overview of the research and innovation initiatives toward carbon neutrality by carbon-neutral fuels and biofuels.

Project	Grant agency	Duration	Details	Ref.
SUNERGY	European Commission	2019-Present	Utilize renewable energy (sunlight, wind) and abundant molecules (CO_2, H_2O, N_2) to produce fuels and chemicals.	[507]
SUNRISE	European Commission	2019-2020	Sub-initiative under SUNERGY project.	[508, 509]
Energy-X	European Commission	2021-2028	Sub-initiative under SUNERGY project.	[510, 511]
Solar Fuels Networks	University of Liverpool	2018-Present	Produce renewable fuels from abundant resources (CO_2, H_2O) by solar energy.	[512]
Power-to-X	Kopernikus project	2016-2019	Convert and store renewable energy into other substantial energy carriers and chemical products.	[513, 514]
Power-to-X	European Commission	2019	Evaluate the utilization of low-carbon electricity to convert power into gas or liquid energy carrier.	[515]
Co-optimization of fuels and engines	NREL	2016-Present	Design fuel blendstock enabled high-efficiency, low-emission engine operation	[267, 415, 419, 516, 517]
Fuel science center	RWTH Aachen University	2007-Present	Develop liquid fuels from renewable electricity and alternative carbon feedstock	[264, 518, 519]

Table 8.2. Overview of C3, C5, C7, C9, C11 alkanes and corresponding ethers with 1~5 O-atom substituents.

No.	3[α]	5	7	9	11
0[β]	n-Propane [529-531][γ]	n-Pentane [530-533]	n-Heptane [91, 454, 497]	n-Nonane [91]	n-Undecane [91]
1	Dimethyl ether [534]	Diethyl ether [535, 536]	Dipropyl ether [537]	Dibutyl ether [452, 453]	Dipentyl ether
2	N/A	Dimethoxymethane [534, 538, 539]	Diethoxymethane, this work	Dipropoxymethane	Dibutoxymethane
3	N/A	N/A	2,4,6-trioxaheptane (PODE2) [447]	3,5,7-Trioxanonane 2,5,8-Trioxanonane	Propoxymethyl ether 3,6,9-Trioxaundecane 2,6,10-Trioxaundecane
4	N/A	N/A	N/A	2,4,6,8-Tetraoxanonane (PODE3) [447, 492, 493, 540]	3,5,7,9-Tetraoxaundecane 2,5,7,10-Tetraoxaundecane
5	N/A	N/A	N/A	N/A	2,4,6,8,10-Pentaoxaundecane (PODE4) [447]

[α] Number in the first row is the total number of C-atoms and O-atoms.

[β] Number in the first column is the total number of O-atoms.

[γ] The newly-published comprehensive mechanisms covering low and high temperature reactions are provided in the reference and those without reference denote comprehensive mechanisms are unavailable so far.

8.2 Chemical kinetic mechanism formulation

8.2.1 Mechanism development and naming of species

Two sets of DEM mechanisms are built which are low-high temperature mechanism (735 species, 3488 reactions) and high temperature mechanism (333 species, 1661 reactions) respectively. The DEM low-high temperature mechanism describes the autoignition at low, intermediate, high temperature regimes from 550~1250K while the DEM high temperature mechanism interprets the high temperature flame phenomenon. The DEM chemical kinetic mechanism development follows a hierarchical structure including C0-C4 core mechanism, n-/i-paraffins sub-mechanism, monoaromatics sub-mechanism, polycyclic aromatic hydrocarbon (PAH) sub-mechanism, alcohols sub-mechanism, dimethoxymethane (DMM) sub-mechanism, diethoxymenthane sub-mechanism as shown in Figure 8.2.

The DEM low-high temperature mechanism integrates the Aramco 3.0 mechanism [541], DMM mechanism [538] and DEM sub-mechanism developed in this work. Aramco 3.0 mechanism [541] act as the base mechanism containing C0-C4 species (H_2, CO, methane, ethane, ethylene, acetylene, formaldehyde, acetaldehyde, propane, propene, butane, etc.), n-/i-paraffins (n-pentane, iso-pentane, neo-pentane, n-hexane, C7-C8 paraffins, etc.), monoaromatics (benzene, phenol, cyclopentadiene, fulvene), PAH (indene, naphthalene, fluorene, etc.), alcohols (methanol, ethanol, n-propanol, iso-propanol, etc.). DMM mechanism [538] is adopted to connect the DEM sub-mechanism and Aramco 3.0 mechanism [541] by providing the intermediate products. The DEM high temperature mechanism incorporates PODE2-4 mechanism [447] and DEM sub-mechanism. The PODE2-4 mechanism [447] acts as the base mechanism which contains the C0-C4 core mechanism to describe the reactions of small molecules and DMM sub-mechanism. In other words, the PODE2-4 mechanism [447] in DEM high temperature mechanism has the same function as Aramco 3.0 mechanism [541] and DMM mechanism [538] in DEM low-high temperature mechanism.

The molecular structure of DEM is similar to n-heptane as shown in Figure 8.3, DEM can be regarded as replacing the carbon atoms at position 3 of n-heptane with oxygen atoms. The physicochemical properties of DEM and n-heptane are compared in Table 8.3. Analogy method is adopted to determine the oxidation reaction pathway and rate constants of elementary reactions and n-heptane is regarded as the analogical object. The comparison of

simplified oxidation reaction pathways between DEM and n-heptane are presented in Figure 8.3. H-atom abstraction from n-heptane forms 4 alkyl radicals of C7H15-1, C7H15-2, C7H15-3, C7H15-4 while H-atom abstraction from DEM produces 3 DEMx radicals of DEM1, DEM2, DEM3 because oxygen atom occupies position 3 as shown in Figure 8.4. Similar to n-heptane, the DEM low temperature oxidation has twice O_2 addition process: the 1st O_2 addition to DEMx to form DEMxO2 and then isomerizes to form QOOH. 2nd O_2 addition occurs on QOOH to produce O2QOOH. The DEM low temperature chain branching reactions trigger by the reactions of O2QOOH = carbonylhydroperoxide + OH, carbonylhydroperoxide = oxygenated radical + OH. Similar to n-heptane, the 2nd O_2 addition reaction is suppressed and the QOOH decomposes into cyclic ethers, acyclic ethers, β-scission products at intermediate temperature. As the reaction temperature further increasing, the DEM molecule and DEMx radicals take place high temperature decomposition by carbon-oxygen bond breakage. The representative species of the DEM molecule, DEMx radical, DEMxO2 radical, β-scission products, R2OOH, R2O, QOOH, O2QOOH, carbonylhydroperoxide, oxygenated radicals in the DEM low-high temperature mechanism are compared with the analogical objects in the n-heptane mechanism [454] as shown in Table 8.3.
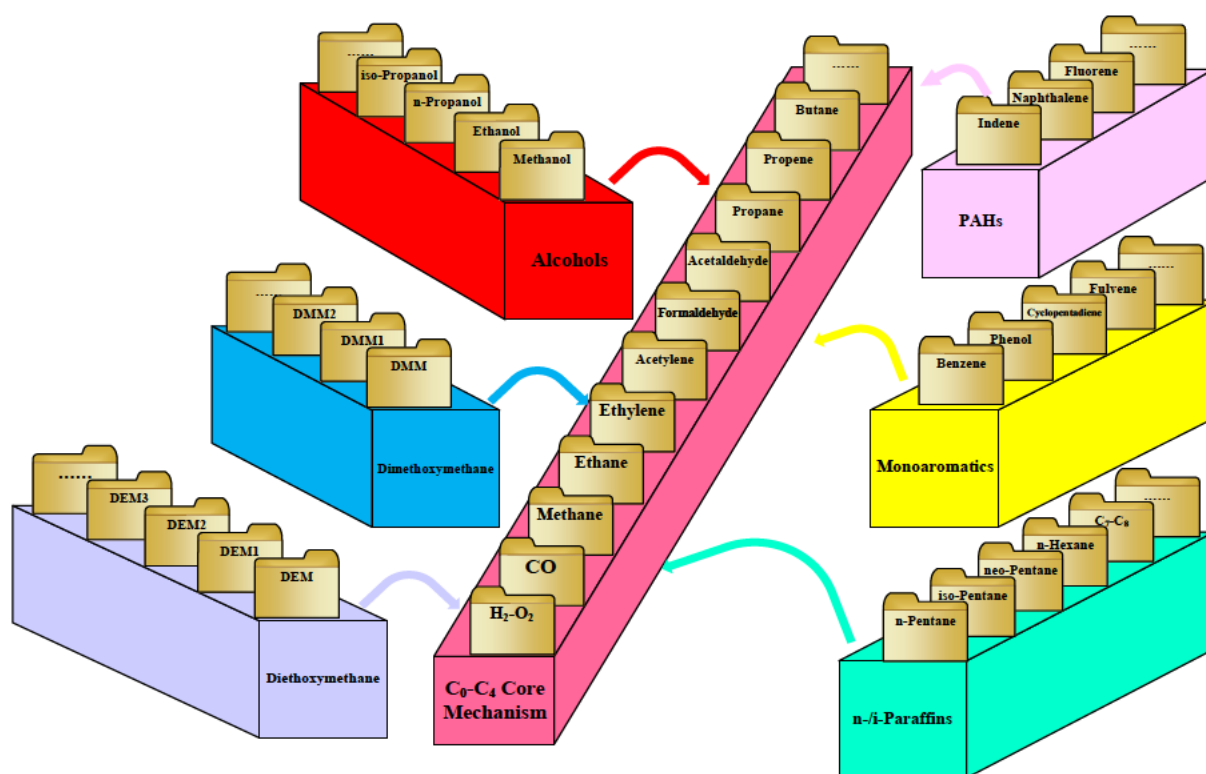
Figure 8.2. The hierarchical/modular structure and overall interrelationships between component libraries in the detailed DEM mechanism.
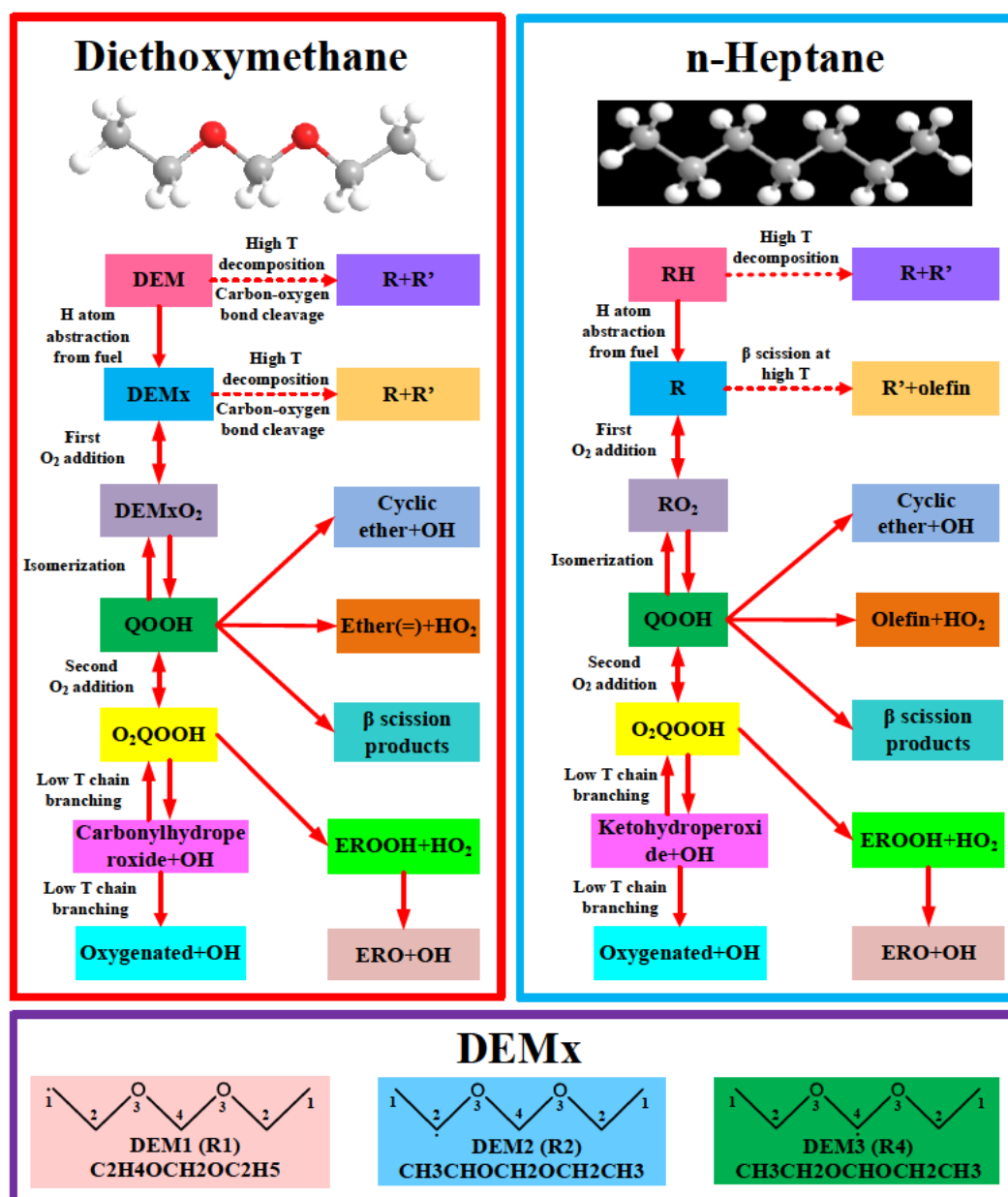
Figure 8.3. Comparison of simplified oxidation reaction pathways between DEM and n-heptane [454].
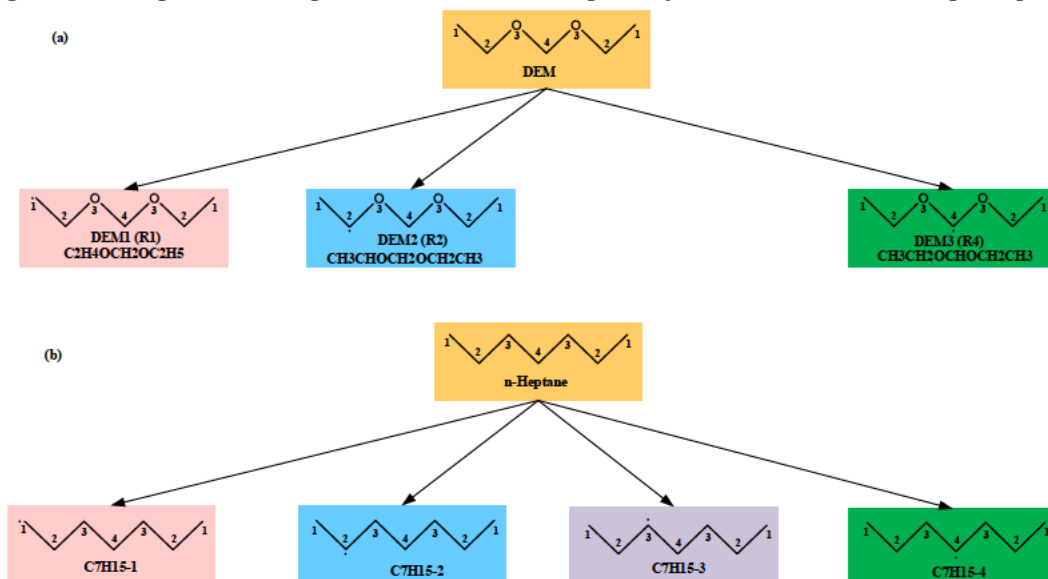
Figure 8.4. Molecular structure of (a) DEM, (b) n-heptane and H-atom abstraction products from fuel molecules.

Table 8.3. Physicochemical properties of DEM.

Name	Diethoxymethane (Ethylal)	n-Heptane
CAS	462-95-3	142-82-5
Formula	C5H12O2	C7H16
Molecular structure		
Molecular weight	104.15	100.20
Boiling point (ºC)	88.05 [542]	98.38 [542]
Melting Point (ºC)	-66.5 [543]	-90.5[543]
Flash point (ºC)	-5 [544]	-4 [544]
Enthalpy of vaporization (kJ/mol)	35.74 [542]	36.57 [542]
Cetane number	57.3 [193]	56 [193]
Lower heating value (MJ/kg)	29.7 [545]	44.925 [546]
YSI	18.5 [321, 370]	36 [321, 370]

Table 8.4. Representative species in DEM mechanism and analogy species in n-heptane mechanism [454].

Name (DEM)	Molecular structure	Name (n-heptanne)	Molecular structure
DEM		nC7H16	
DEM2/ CH3CHOCH2OCH2CH3		C7H15-2	
DEM2O2/ CH3COOHOCH2OC2H5		C7H15-2O2	
C2H5OCH2OCHCH2		C7H14-1	
C2H5OCOCOOHHCH3		C7H15OOH-2	
C2H5OCOCOHCH3		C7H15-2O	
CH3CHOCOCOOHHCH3		C7H14OOH 2-6	
CH3COOOCOCOOHCHH3		C7H14OOH 2-5O2	
CH3COOCOCOOHHCHH3		C7KET26	
CH3COOCOCOHCH3		C7KET26O	

8.2.2 Reaction classes and detailed oxidation reaction pathway

The key low temperature oxidation reaction pathway of DEM is presented in Figure 1 which includes the H-atom abstraction from fuel molecules, the 1st O_2 addition, DEMxO2 isomerization to form QOOH, the 2nd O_2 addition, low temperature chain branching reactions. The DEM low-high temperature mechanism composes of 13 high temperature reaction classes and 24 low temperature reaction classes as listed in Table 8.5 and Table 8.6 respectively. The reaction classes of hydrocarbons and oxygenated fuels are referred to [91, 497, 547, 548] and [447, 538] respectively. The reaction classes constitute a reaction network to interpret from fuel molecules consumption to final products. Each reaction class corresponds to a set of elementary reactions and the reaction rate constants are specified by the rate rules of particular reaction classes [447]. The rate constants K of H-atom abstraction from DEM and n-heptane by O_2, OH, C_2H_5 are compared in Figure 8.6 and the corresponding pre-exponential factor A, temperature exponent n and activation energy E_a are listed in Table 8.7. The rate constants of H-atom abstraction from DEM are higher than those of n-heptane. The H-atom abstraction reaction rates order from high to low as: from position 2 carbon atim>from position 4 carbon atom>from position 1 carbon atom at most conditions. The reaction class 24 describes the $R\dot{O}_2$ radical isomerization via transition state ring structure to form QOOH and the internal H-atom abstraction of DEM1O2, DEM2O2, DEM3O2 are presented in Figure 8.7. The intermediate transition state structure considered in the DEM low-high temperature mechanism ranges from 5 to 10 member ring.
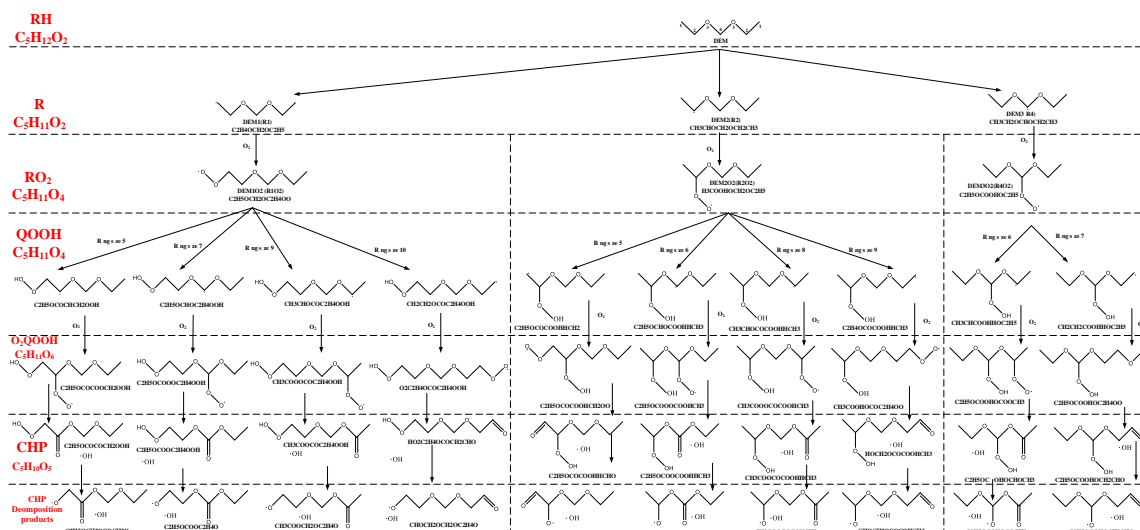
Figure 8.5. The key low temperature oxidation reaction pathway of DEM.
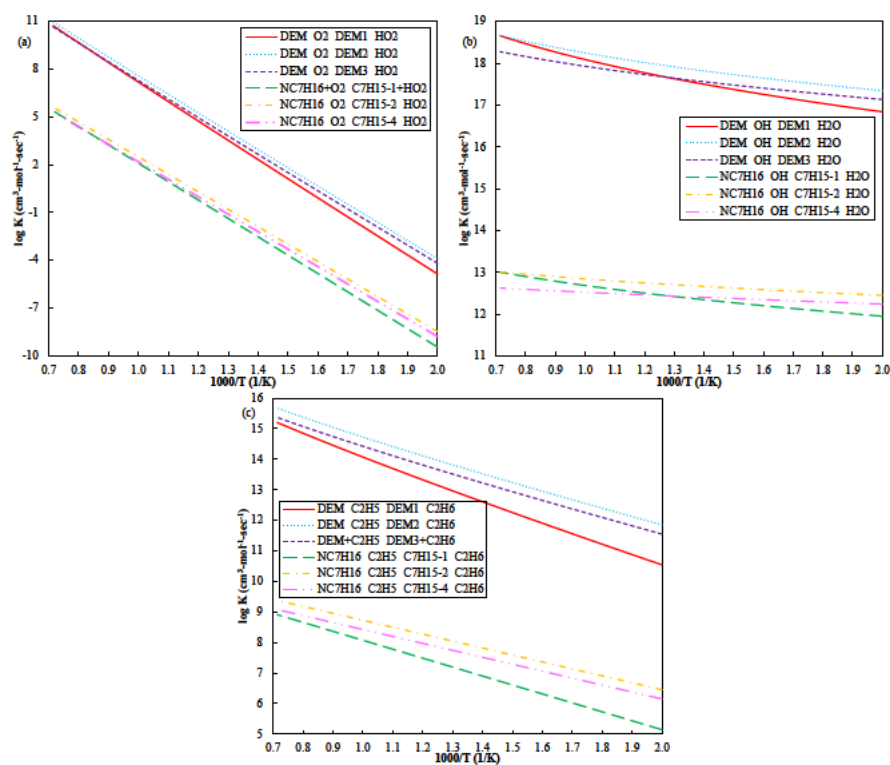
Figure 8.6. Rate constants of H atom abstraction from DEM and n-heptane [454].
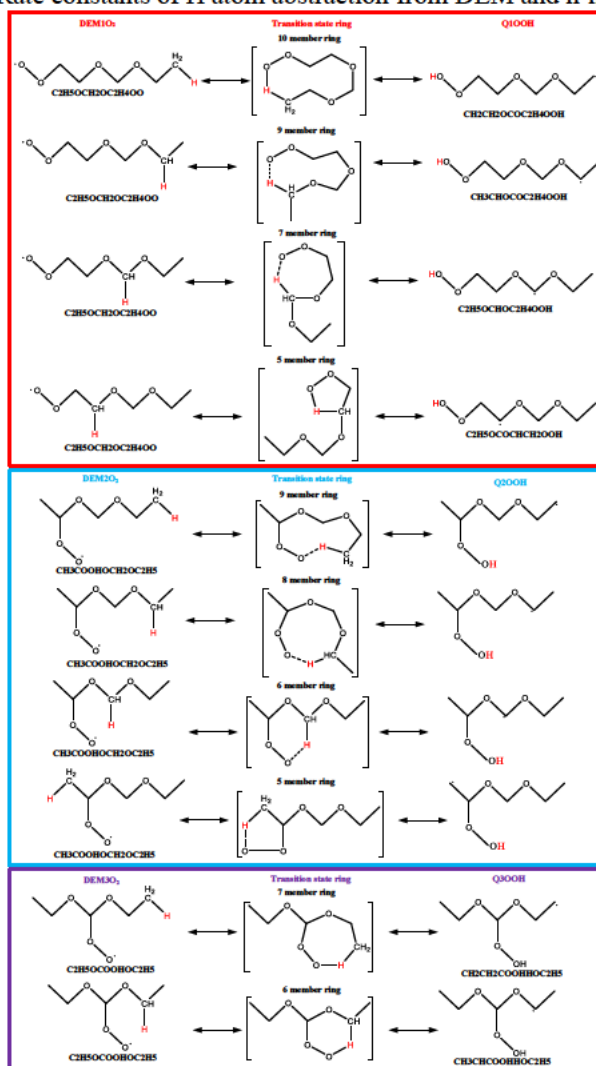
Figure 8.7. Internal H-atom abstraction of $R\dot{O}_2$ radicals via transition state ring structure, reaction class 24: $R\dot{O}_2 \Longleftrightarrow QOOH$.

Table 8.5. High temperature reaction classes considered in the DEM mechanism.

Class ID	High temperature reaction classes	Rate constant source					
		Aramco [541]	3.0 DMM [538]	n-Heptane [454]	DEM [527, 549]	DEM [528]	This work
1	Unimolecular fuel decomposition: Fuel<=> $\dot{R}$ + $\ddot{R}$	√	√	×	√	√	√
2	Fuel decomposition: RH<=> $\dot{R}$ + $\dot{H}$	×	√	√	×	×	√
3	H-atom abstraction from the fuel by O_2/ $\dot{H}$ / $\dot{O}$ / $\dot{O}H$ / $H\dot{O}_2$ / $\dot{C}H_3$ / $CH_3\dot{O}$ / $CH_3\dot{O}_2$ / $\dot{O}_2CHO$ / $\dot{C}_2H_3$ / $\dot{C}_2H_5$ (primary/secondary carbon sites): RH+ $\dot{X}$ <=> $\dot{R}$ +XH	√	√	√	√	√	√
4	$\dot{R}$ radical decomposition	√	×	×	√	√	√
5	$\dot{R}$ radical isomerization	√	√	×	×	√	√
6	C–O: β-scission of $\dot{R}$ radicals	×	√	×	×	×	√
7	O–C–O: β-scission of $\dot{R}$ radicals	×	√	×	×	×	√
8	H-atom abstraction reactions from alkenes	√	×	×	×	×	√
9	Addition of radical species O and OH to alkenes	√	×	×	×	×	√
10	Reactions of alkenyl radicals with $H\dot{O}_2$, CH3O2, and C2H5O2	√	×	×	×	×	√
11	Alkenyl radical decomposition	√	×	×	×	×	√
12	Alkene decomposition	√	×	×	×	×	√
13	Retroene decomposition reactions	√	×	×	×	×	√

Table 8.6. Low temperature reaction classes considered in the DEM mechanism.

Class ID	Low-temperature reaction classes	Rate constant source					
		Aramco 3.0 [541]	DMM [538]	n-Heptane [454]	DEM [527, 549]	DEM [528]	This work
14	Addition of O_2 to $\dot{R}$ radicals: $\dot{R} + O_2 <=> R\dot{O}_2$	√	√	√	×	×	√
15	$\dot{R} + O_2 <=> E + H\dot{O}_2$ (E: radicals contain carbon-carbon double bond)	×	×	√	×	×	√
16	$\dot{R} + R'\dot{O}_2 <=> R\dot{O} + R'\dot{O}$ (including $\dot{R} + H\dot{O}_2 <=> R\dot{O} + OH$)	√	√	√	×	×	√
17	$R\dot{O}_2 + Fuel <=> ROOH + \dot{R}'$	×	×	√	×	×	√
18	$R\dot{O}_2 + R'\dot{O}_2 <=> R\dot{O} + R'\dot{O} + O_2$ (including $R\dot{O}_2 + CH_3\dot{O}_2 <=> R\dot{O} + CH_3\dot{O} + O_2$)	√	×	√	×	×	√
19	$R\dot{O}_2 + H\dot{O}_2 <=> ROOH + O_2$	√	×	√	×	×	√
20	$R\dot{O}_2 + H_2O_2 <=> ROOH + H\dot{O}_2$	√	×	√	×	×	√
21	$ROOH <=> R\dot{O} + \dot{O}H$	×	×	√	×	×	√
22	$R\dot{O}$ decomposition	√	√	√	×	×	√
23	$R\dot{O}_2 <=> E + H\dot{O}_2$	×	×	√	×	×	√
24	$R\dot{O}_2$ radical isomerization: $R\dot{O}_2 <=> QOOH$	√	√	√	×	×	√
25	$R\dot{O}_2$ concerted eliminations: $R\dot{O}_2 <=> alkene + H\dot{O}_2$	√	×	×	√	×	√
26	QOOH = cyclic ether + OH (cyclic ether formation)	√	√	√	×	×	√
27	QOOH decomposition (β-scission products)	√	√	√	×	×	√
28	QOOH = alkene + HO_2 (radical site beta to OOH group)	√	×	√	×	×	√
29	QOOH = ether with carbon-carbon double bond + HO_2	×	×	×	×	×	√
30	QOOH = alkene + carbonyl + OH (radical site gamma to OOH group)	√	×	√	×	×	√
31	Addition of O_2 to $\dot{Q}OOH$: $\dot{Q}OOH + O_2 <=> \dot{O}_2QOOH$	√	√	√	×	×	√
32	$\dot{O}_2QOOH$ isomerization to form carbonylhydroperoxide and $\dot{O}H$	√	√	√	×	×	√
33	Carbonylhydroperoxide decomposition to form oxygenated radicals and $\dot{O}H$	√	√	√	×	×	√
34	Cyclic ether reactions with $\dot{O}H$ and $H\dot{O}_2$	√	√	√	×	×	√
35	Decomposition of large carbonyl species and carbonyl radicals	√	×	√	×	×	√
36	$\dot{O}_2QOOH <=> EROOH + H\dot{O}_2$ (EROOH: olefinic hydroperoxy)	×	×	√	×	×	√
37	EROOH decomposition	×	×	√	×	×	√

Table 8.7. Elementary reactions of H-atom abstract from fuel and rate constant for DEM and n-heptane [454].

DEM reactions	A ($cm^3mol^{-1}s^{-1}K^{-n}$)	n	E_a (cal/mole)	n-Heptane reactions	A ($cm^3mol^{-1}s^{-1}K^{-n}$)	n	E_a (cal/mole)
DEM+O2=DEM1+HO2	4.20E+13	1.7	52800	NC7H16+O2=C7H15-1+HO2	4.20E+13	0	52800
DEM+O2=DEM2+HO2	2.80E+13	1.7	50160	NC7H16+O2=C7H15-2+HO2	2.80E+13	0	50160
DEM+O2=DEM3+HO2	1.40E+13	1.7	50160	NC7H16+O2=C7H15-4+HO2	1.40E+13	0	50160
DEM+OH=DEM1+H2O	5.46E+07	3.513	868.4	NC7H16+OH=C7H15-1+H2O	2.73E+07	1.813	868.4
DEM+OH=DEM2+H2O	2.82E+10	2.635	504.7	NC7H16+OH=C7H15-2+H2O	1.41E+10	0.935	504.7
DEM+OH=DEM3+H2O	1.12E+12	2.02	846.5	NC7H16+OH=C7H15-4+H2O	5.62E+11	0.32	846.5
DEM+C2H5=DEM1+C2H6	1.00E+11	2	13400	NC7H16+C2H5=C7H15-1+C2H6	1.00E+11	0	13400
DEM+C2H5=DEM2+C2H6	1.00E+11	2	10400	NC7H16+C2H5=C7H15-2+C2H6	1.00E+11	0	10400
DEM+C2H5=DEM3+C2H6	5.00E+10	2	10400	NC7H16+C2H5=C7H15-4+C2H6	5.00E+10	0	10400

8.2.3 Thermochemical and transport data

The species thermodynamic data (enthalpy, entropy, and specific heat capacity fitting to two temperature ranges) and transport data (geometrical configuration, Lennard-Jones potential well depth, Lennard-Jones collision diameter, dipole moment, polarizability) in current DEM mechanisms are mainly derived from Aramco 3.0 mechanism [541], DMM mechanism [538], n-heptane mechanism [454], DEM high temperature mechanism [527, 528, 549], 2-methylalkanes (C_7 to C_{20}) mechanism [91], dibutyl ether mechanism [453] and PODE$_{2-4}$ mechanism [447]. Those thermodynamic data of those species not presented in the above mechanisms are computed by group additivity theory [195, 550, 551]. The transport data of Lennard-Jones potential well depth and collision diameter for unreported stable species are determined by empirical correlation proposed by Tee et al. [552]. Yaws' handbook [553]. The required critical pressure, critical temperature and acentric factor are retrieved from AP1700 Material property calculation and inquiry platform [554] and NIST Chemistry WebBook [542]. The transport data of unreported radicals come out of those radicals with identical formulas and similar molecular structures or parental species.

8.3 Results and discussion

8.3.1 Ignition delay time validation

DEM/O_2/N_2 mixture ignition delay times measured by shock tube (ST) and rapid compression machine (RCM) at φ=1.0, T_{init}=500~1400K, P_{init}=30bar are reported by Lehrheuer et al. [545]. The simulated ignition delay times by DEM low-high temperature mechanism (hereafter referred to as UOB mechanism) are compared with measured values as shown in Figure 8.8 and the model inputs for Chemkin simulation are listed in Table 8.8. The present DEM mechanism accurately reproduces the low, intermediate, high temperature ignition delay times. The DEM ignition delay times increase as the dilution ratio of N_2/O_2 increasing from 3.76 to 12. The DEM-air mixture (N_2/O_2=3.76) exhibits a plateau at an intermediate temperature regime of 640~960K. The slope of ignition delay times at intermediate temperature increases as the dilution ratio of N_2/O_2 increasing from 3.76 to 12. A sensitivity analysis on OH species respective to reaction pre-exponential factor is performed at φ=1.0, T_{init}=500/800/1200K, P_{init}=30bar, N_2/O_2=3.76 as shown in Figure 8.9 (a), Figure 8.9 (b), Figure 8.9 (c). At T_{init}=500K, the reaction of Eq. (8.1) (H-atom abstraction from position 2 carbon atom of DEM by O_2) and reaction sequence of Eq. (8.2) ~ Eq. (8.3) facilitate low temperature ignition while the reaction of Eq. (8.4) (H-atom abstraction from position 4 carbon atom of DEM by HO_2 radical) suppresses OH radical formation. At T_{init}=800K, the H-atom abstraction from position 2, 4 carbon atoms of DEM by CH_3, C_2H_5 radicals restrain OH radical production. At T_{init}=1200K, the Eq. (8.5) ~ Eq. (8.7) are the chain branching reaction sequence while the Eq. (8.8) (H-atom abstraction from position 4 carbon atom of DEM by C_2H_5 radical).

$$DEM+O2=CH3CHOCH2OCH2CH3+HO2 \tag{8.1}$$

$$C2H5O2+CH2O=C2H5O2H+HCO \tag{8.2}$$

$$C2H5O2H=C2H5O+OH \tag{8.3}$$

$$DEM+HO2=CH3CH2OCHOCH2CH3+H2O2 \tag{8.4}$$

$$CH3CHO+HO2=CH3CO+H2O2 \tag{8.5}$$

$$CH2O+HO2=HCO+H2O2 \tag{8.6}$$

$$H2O2(+M)=2OH(+M) \tag{8.7}$$

$$DEM+C2H5=CH3CH2OCHOCH2CH3+C2H6 \tag{8.8}$$

The high temperature ignition delay times of DEM measured by ST at φ=0.5/1.0/2.0, T_{init}=1000~1400K, P_{init}=2/4/10bar are studied by Zhang et al. [528] and they also proposed the DEM high temperature mechanism (hereafter referred to as SCU mechanism). This UOB mechanism and the SCU mechanism are simulated at the studied conditions and compare with ST experimental data as shown in Figure 8.10 and the model inputs for Chemkin simulation are presented in Table 8.8. At P_{init}=2/4 bar, the discrepancy of UOB mechanism decreases as equivalence ratio increases from 0.5 to 2.0 which is opposite to SCU mechanism as shown in Figure 8.10 (a), Figure 8.10 (b). At φ=1.0, the predictive accuracy of UOB mechanism improves as the mixture initial pressure increases from 2 bar to 10 bar as shown in Figure 8.10 (c). Generally, both the UOB mechanism and the SCU mechanism are capable of capturing the dependence of ignition delay time on equivalence ratio, temperature, pressure at the high temperature regime of 1000~1400K.
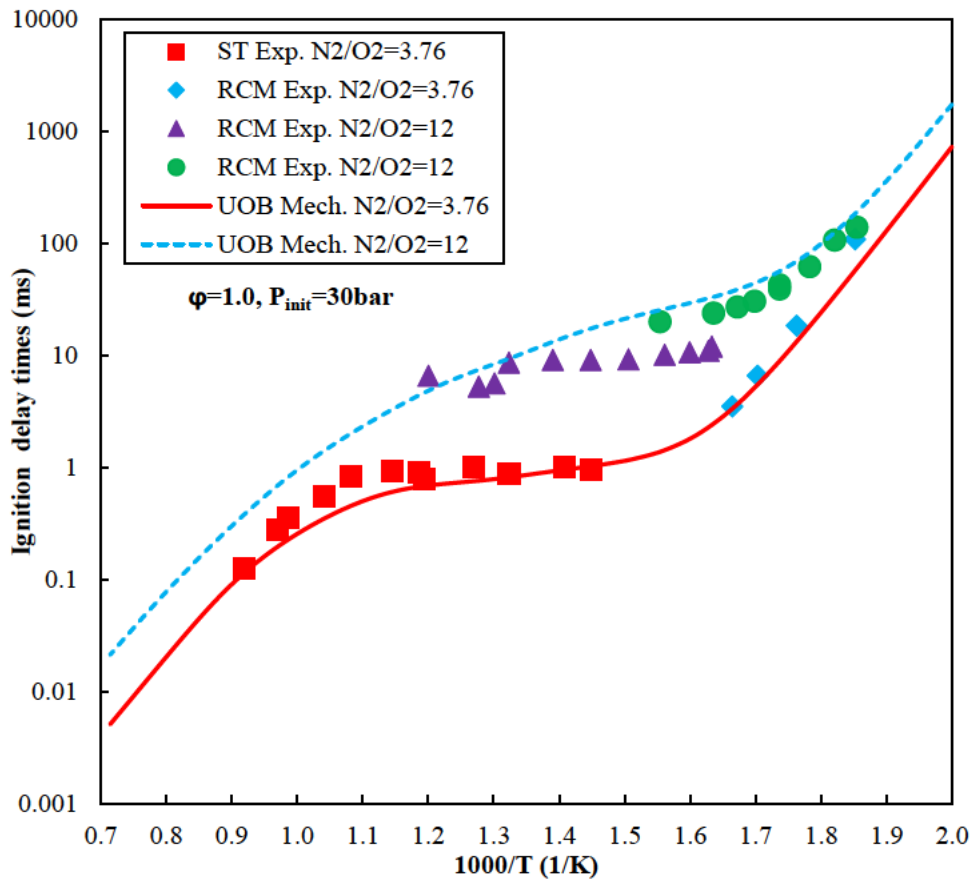
Figure 8.8. Experimental (symbols [545]) and modeling results (lines) of DEM/O_2/N_2 mixture ignition delay times at φ=1.0, T_{init}=500~1400K, P_{init}=30bar.
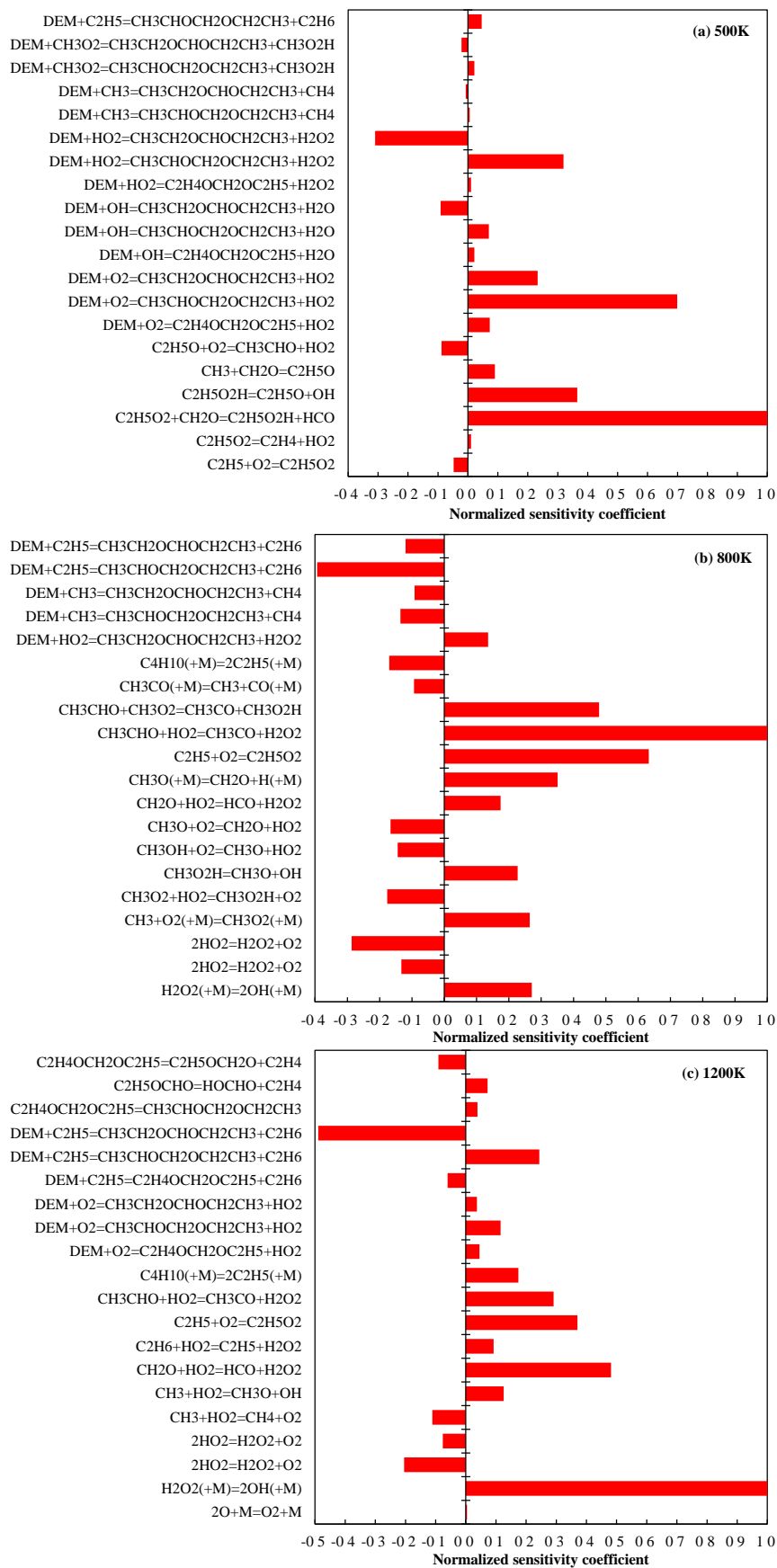
Figure 8.9. Sensitivity analysis on OH species respective to reaction A-factors for DEM O_2/N_2 mixture oxidation at $\varphi=1.0$, $T_{init}=500/800/1200K$, $P_{init}=30bar$, $N_2/O_2=3.76$.
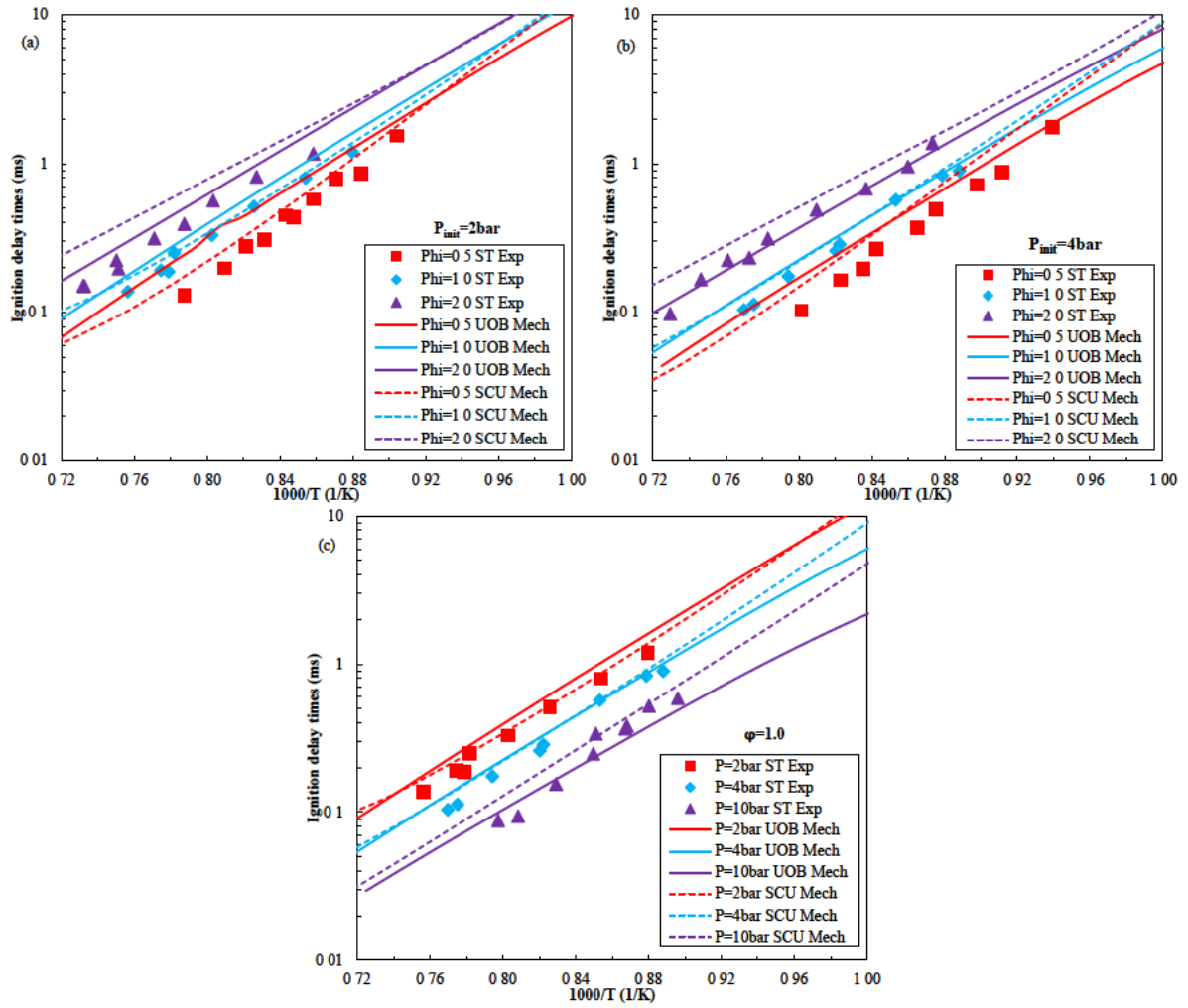
Figure 8.10. Experimental (symbols [528]) and modeling results (lines) of DEM/O_2/Ar mixture ignition delay times at φ=0.5/1.0/2.0, T_{init}=1000~1400K, P_{init}=2/4/10bar.

Table 8.8. Boundary conditions of DEM oxidation for ignition delay time experiments/simulations

	Compositions	φ	T_{init} (K)	P_{init} (bar)
Lehrheuer et al. [545]	DEM:O_2:N_2=1:7:26.32	1.0	500~1400	30
	DEM:O_2:N_2=1:7:84	1.0	500~1400	30
Zhang et al. [528]	DEM:O_2:Ar=1:14:85	0.5	1000~1400	2/4
	DEM:O_2:Ar=1:7:92	1.0	1000~1400	2/4/10
	DEM:O_2:Ar=1:3.5:95.5	2.0	1000~1400	2/4

8.3.2 Laminar flame speed validation

The premixed laminar flame speed of the DEM-air mixture is measured at φ=0.5~2.0, T_{init}= 1atm, P_{init}=1.01/2.50bar by Kopp et al. [555]. The developed DEM high temperature mechanism (hereafter referred to as UOB mechanism) emulates the studied flame phenomena at the studied condition and compared with the observed values as shown in Figure 8.11. A good agreement between measured and predicted values is obtained at the studied conditions and the UOB mechanism slightly overestimates the laminar flame speed of 2.27 ~ 2.85 cm/s at φ=0.9~1.2, P_{init}=1.01bar. as shown in Figure 8.11. The sensitivity analysis of flame temperature respective to reaction pre-exponential factors is performed at φ=0.8, 1.1, 1.5 as shown in Figure 8.12 (a) ~ Figure 8.12 (c). The reaction sequence of Eq. (8.9) ~ Eq. (8.10) favors flame temperature elevation while Eq. (8.11) ~ Eq. (8.12) are endothermic reactions.

$$H+O2=O+OH \tag{8.9}$$

$$CO+OH=CO2+H \tag{8.10}$$

$$H2O+M=H+OH+M \tag{8.11}$$

$$CH3+H(+M)=CH4(+M) \tag{8.12}$$
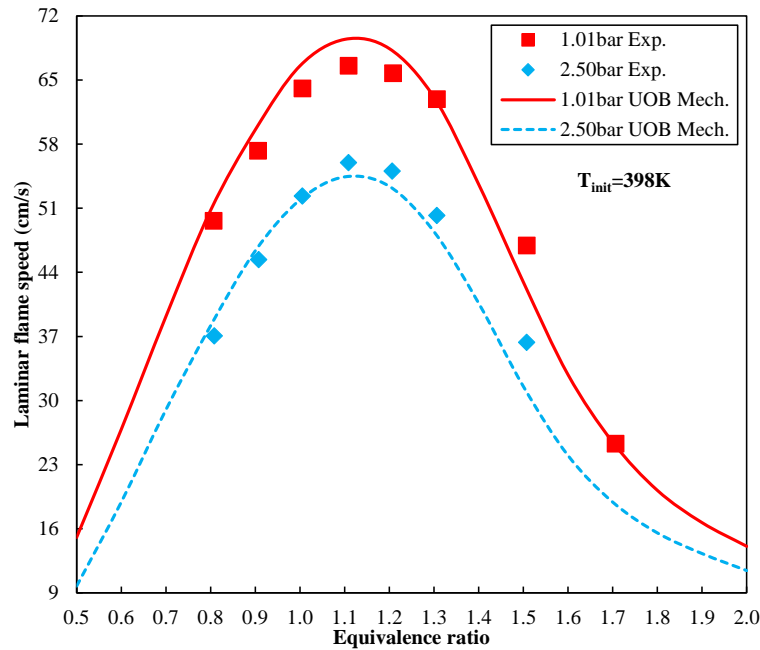
Figure 8.11. Experimental (symbols [555]) and modeling results (lines) for the laminar flame speed of DEM in air at φ=0.5~2.0, T_{init}= 1atm, P_{init}=1.01/2.50bar.
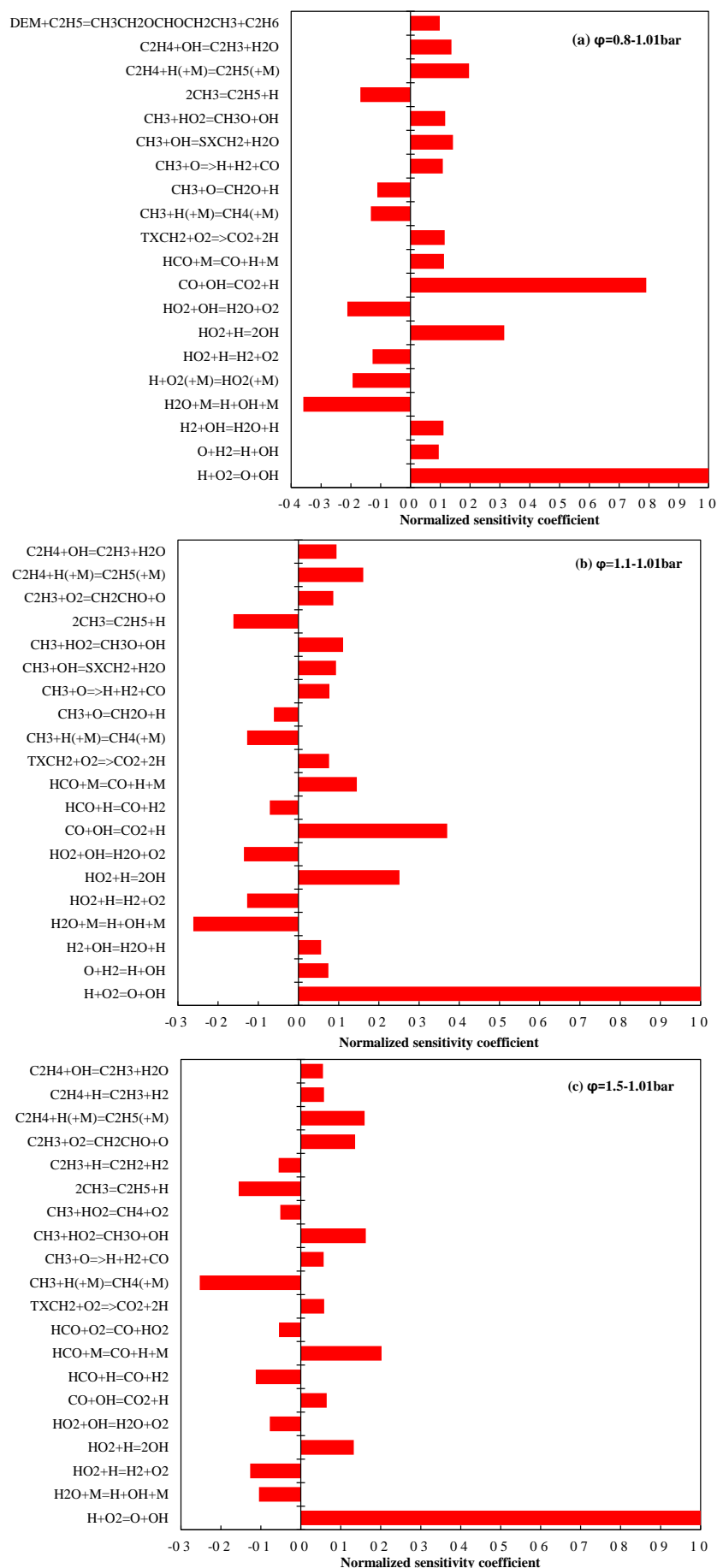
Figure 8.12. Sensitivity analysis on flame temperature respective to reaction A-factors for DEM oxidation at T_{init}=398K, P_{init}=1.01bar, (a) φ=0.8, (b) φ=1.1, (c) φ=1.5.

8.3.3 Comparison of ignition delay times between DEM and n-heptane

The ignition delay times of DEM-air mixture and n-heptane-air mixture [454] are compared at φ=0.25/0.5/1.0/2.0, T_{init}= 500~1400K, P_{init}= 13.5/20/38/55bar as shown in Figure 8.13. The major findings are summarized as below: (1) DEM has stronger low temperature reactivity than n-heptane at 500~660K. (2) There is no observed NTC behavior for DEM and a flat plateau is observed at the intermediate temperature of 640~920K. On the contrary, the n-heptane exhibits obvious NTC behavior at 720~900K. (3) DEM has higher high temperature reactivity than n-heptane at 1000~1250K. The species evolution of DEM-air mixture is compared with the n-heptane-air mixture at φ=1.0, T_{init}=560/800/1200K, P_{init}= 38bar as shown in Figure 8.14. The results indicate that both DEM and n-heptane demonstrate low temperature heat release (LTHR) and high temperature heat release (HTHR) at 560K. DEM rapidly accumulates formaldehyde at low temperatures compared to n-heptane. At 800K, DEM transits from two-stage heat release to single-stage heat release which is different from n-heptane. Both fuels exhibit HTHR at 1200K.
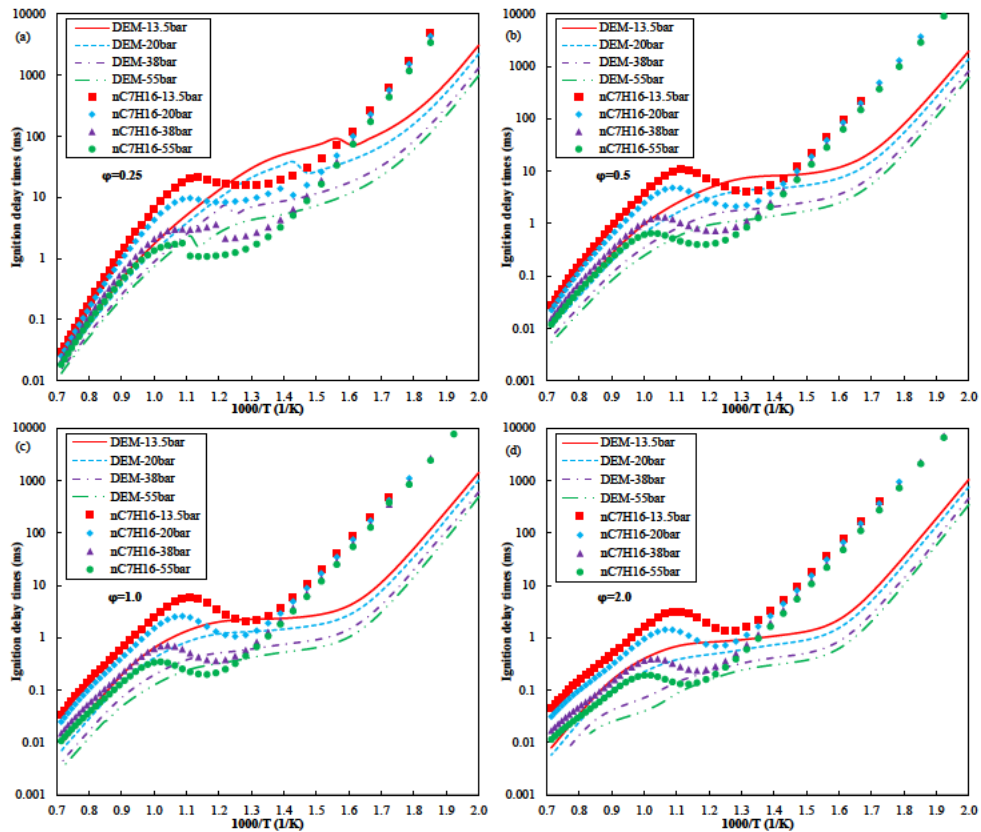
Figure 8.13. Ignition delay times of DEM-air mixture and n-heptane-air mixture [454] at T_{init}= 500~1400K, P_{init}= 13.5/20/38/55bar, (a) φ=0.25, (b) φ=0.5, (c) φ=1.0, (d) φ=2.0.
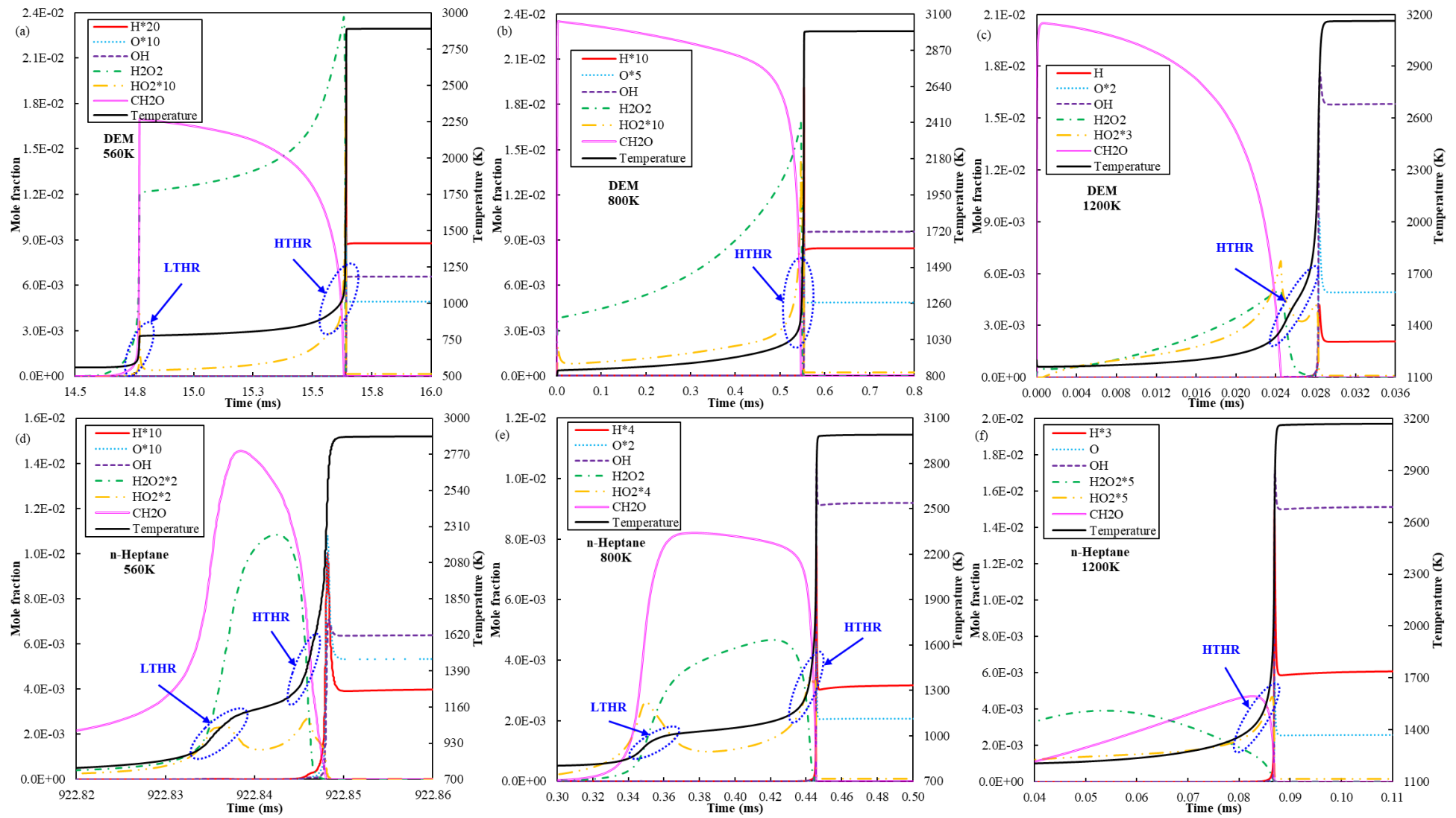
Figure 8.14. Species evolution of DEM-air mixture and n-heptane-air mixture [454] at $\varphi=1.0$, $P_{init}= 38bar$, (a)/(d) $T_{init}= 560K$, (b)/(e) $T_{init}= 800K$, (c)/(f) $T_{init}= 1200K$.

8.4 Conclusions

Two sets of DEM detail chemical kinetic mechanisms are developed: one is the low-high temperature mechanism to describe the autoignition behavior at the full temperature regime; the other is the high temperature mechanism to interpret the high temperature flame phenomenon. The DEM mechanism is assembled as a hierarchical structure containing C0-C4 core mechanism, n-/i-paraffins sub-mechanism, monoaromatics sub-mechanism, PAHs sub-mechanism, alcohols sub-mechanism, dimethoxymethane (DMM) sub-mechanism, DEM sub-mechanism. The analogy method is used to construct the DEM sub-mechanism and the n-heptane is the analogical object. The DEM oxidation reaction pathway is designed which contains 13 low temperature reaction classes and 24 high temperature reaction classes. The pre-exponential factor, temperature exponent, activation energy of particular elementary reactions are specified by the rate rule of specific reaction classes. The proposed mechanism is validated against ignition delay time, laminar flame speed and satisfactory agreement is obtained between the predictive and measured values. The proposed mechanism enables to compute ignition delay time at engine relevant conditions and premixed laminar flame speed at ambient to mediate pressure. But it is still far from perfect, there are some knowledge gaps in DEM research as listed below:

(1) A jet stir reactor experiment to test DEM oxidation 500~1200K is needed to obtain species evolution data for mechanism development.

(2) Due to a lack of speciation data, the rate constants of elementary reaction of DEM sub-mechanism are semi-empirical. Quantum chemistry technology (like Gaussian 03 or ReaxFF) tool is needed to determine the reaction rate constant of DEM sub-mechanism.

Chapter 9 Conclusions and Suggestions for Future Research

9.1 Study findings and conclusions

This work showcases the application of artificial intelligence and chemical kinetics to property-oriented fuel design for the internal combustion engine. There are two main lines of this research: (1) Application of machine learning (ML) and deep learning (DL) technology into fuel molecular structure to establish the surrogate formulation tool and fuel screening tool. The fuel screening tool consists of 15 fuel physicochemical properties regression models for multiple fuel types based on ML. As an attempt, DL technology is applied to build the YSI regression model which automatically extracts the molecular structure features and performs regression operation. The advantages and disadvantages of ML and DL are examined and revealed. (2) Application of chemical kinetics to investigate the ignition behavior, flame phenomenon, oxidation reaction pathways, rate-controlling reactions of the target molecules selected by the fuel screening tool. There are seven key achievements of surrogate formulation method, fuel ignition quality prediction models, fuel physicochemical properties regression models, application of ML & DL to YSI prediction, fuel screening tool, PODE3 ignition and flame characteristics, detail chemical kinetic mechanism of diethoxymethane which interpret in section 9.1.1 ~ section 9.1.7 respectively.

9.1.1 Group contribution method (GCM) for surrogate formulation

This work develops a GCM to formulate surrogates toward practical fuels such as diesel, biodiesel, gasoline, aviation jet fuels, etc. with accurate fuel compositions. The GCM contains two sub-systems: (1) GCM-UOB 1.0 functional group classification system with 22 molecular descriptors which decomposes the target fuels and surrogate components into typical structural fragments. (2) Chemical fragments regression model which minimizes the functional group discrepancy between the target fuel and surrogate fuels. The GCM obtains a good agreement of the studied fuel physicochemical properties including ignition delay times, premixed laminar flame speed, species evolution, liquid phase density, sound speed, kinematic viscosity for POSF 4658 jet fuel, rapeseed methyl ester (RME) biodiesel, diesel, FACE C gasoline. ML technology cannot directly apply to the chemical engineering

problem without tabular data input. GCM-UOB 1.0 system converts the fuel molecules into molecular structure matrix which digitalizes the chemical structural features and it is the foundation of ML implementation into fuel ignition quality prediction (chapter 3), 15 fuel physicochemical properties prediction (chapter 4), yield sooting index (YSI) prediction (chapter 5), fuel screen tool utilization (chapter 6).

9.1.2 Machine learning-quantitative structure property relationship (ML-QSPR) method for fuel ignition quality (CN/RON/MON) prediction

This work proposes a ML-QSPR method to predict the fuel ignition quality of CN, RON, MON for pure compounds and mixtures. The ML-QSPR method comprises 2 sub-systems: (1) QSPR-UOB 2.0 functional classification system containing 32 molecular descriptors. It upgrades from GCM-UOB 1.0 system by adding 10 functional group fragments to recognize the aromatics with one benzene ring, two fused benzene rings, three fused benzene rings. (2) Machine learning (ML) based regression model correlating the fuel molecular structure matrix and ignition quality matrix. Gaussian process regression algorithm is approved as the best ML algorithm. High predictive accuracy of CN, RON, MON are obtained and their R^2 reach 0.9911, 0.9874, 0.9731. The QSPR theory assumes that the fuel physicochemical properties are the results of the molecular structure and thus the variation of the chemical structure varies macroscopic properties. The QSPR method is constructed on the atomic level rather than molecular level, thus the ML-QSPR method can apply to fuel compounds beyond the model training dataset and fuel mixtures.

9.1.3 Machine learning-quantitative structure property relationship (ML-QSPR) method for 15 fuel physicochemical properties prediction

This works extends the application of the ML-QSPR method from fuel ignition quality of CN, RON, MON to 15 fuel physicochemical properties of melting point, boiling point, vapor pressure, heat of vaporization, cetane number (CN), research octane number (RON), motor octane number (MON), ignition temperature (IT), flash point (FP), yield sooting index (YSI), liquid density, lower heating value (LHV), surface tension, lower flammability

limit (LFL), upper flammability limit (UFL). The ML-QSPR method contains two sub-systems: (1) QSPR-UOB 3.0 functional group classification system with 44 molecular descriptors. It upgrades from QSPR-UOB 2.0 system by complementing 5 functional group fragments to depict the aromatics with 3~5 fused benzene rings and adding ester group in ring structure, carbonate ester, carboxylic anhydride, hydroperoxide and peroxide to distinguish ester group in ring structure or non-ring, carbonate esters, carboxylic anhydrides, hydroperoxides and peroxides. Its function is to convert the fuel molecules into a fuel molecular structure matrix as the structural features. (2) ML-based regression models connecting the fuel molecular structure matrix and fuel property matrix. The R^2 of CN, RON, MON, melting point, boiling point, heat of vaporization, surface tension, LHV, liquid density, YSI, IT, FP, VP, LFL, UFL are 0.9898, 0.9884, 0.9758, 0.9653, 0.9484, 0.9968, 0.9898, 0.9959, 0.9946, 0.9993, 0.9603, 0.9798, 0.9972, 0.9935, 0.9486 and they reach reasonable overall predictive accuracy. These regression models of 15 fuel physicochemical properties constitute the Tier 1 fuel physicochemical property screening of the fuel screening tool.

9.1.4 Application of machine learning (ML) and deep learning (DL) to YSI (yield sooting index) prediction

The DL technology especially the convolution neural network (CNN) is usually used in computer vision problems of image classification and object detection. This work applies ML and DL to regression problem and takes YSI as a case study. In the ML technology roadmap, the ML-QSPR method is adopted for YSI prediction. In the DL technology roadmap, a standard series network of SDSeries38 is developed to automatically extract the chemical structure feature and perform regression operations. The SDSeries38 network is tailor-made for regression problem which contains 9 feature learning modules and 1 regression module. Each feature learning module is a stack of convolution layer, batch normalization layer, rectified linear unit (ReLU) layer, max pooling layer. The regression layer consists of a fully connected layer and a regression layer. The ML-QSPR method (R^2=0.9993, RMSE=7.567) outperforms SDSeries38 network (R^2=0.9953, RMSE=13.352) in predictive accuracy.

The SDSeries38 network demands higher computational resources and consumes greater time for model training and prediction than the ML-QSPR method. In addition, the SDSeries38 network outperforms the pretrained CNN of AlexNet, Densenet201, GoogleNet, Inceptionv3, Mobilenetv2, ResNet18, Resnet50, Shufflenet, Squeezenet, Xception in terms of predictive accuracy and time consumption. The results indicate that tailor-made CNN is required for regression problem to improve predictive accuracy and efficient computation. Direct transfer learning of pretrained CNN from image classification problem to regression problem usually cannot obtain satisfactory predictive accuracy.

9.1.5 High throughput fuel screening for IC engines by ML-QSPR and chemical kinetics

This work develops a high throughput fuel screening tool for IC engines containing Tier 1 fuel physicochemical property screening and Tier 2 chemical kinetic screening. The term "high throughput" denotes this tool applies to 24 fuel types and 15 fuel physicochemical properties. Tier 1 fuel physicochemical property screening includes 5 categories and 16 parameters: (1) Volatility specification: melting point T_m, boiling point T_b, vapor pressure (VP), enthalpy of vaporization ΔH_{vap}. (2) Atomization specification: surface tension γ, dynamic viscosity. (3) Energy density: lower heating value (LHV), liquid density. (4) Sooting tendency: yield sooting index (YSI). (5) Ignitability: CN, RON, MON, ignition temperature (IT), flash point (FP), lower/upper flammability limits (LFL/UFL). Tier 2 chemical kinetic screening includes: (1) ignition delay time; (2) φ sensitivity; (3) laminar flame speed. Take CI engines as an example, the Tier 1 screening criteria list as below: (a) melting point: $T_m \leq$ -10°C; (b) boiling point: 60 °C $\leq T_b \leq$ -250°C; (c) enthalpy of vaporization: $\Delta H_{vap} \leq$ 75kJ/mol; (d) surface tension: $\gamma \leq$ 38mn/m; (e) dynamic viscosity: $\mu \leq$ 2mpa·s; (f) lower heating value: LHV $\geq$ 2700kJ/mol; (g) liquid density: $\rho \geq$ 675kg/m³; (h) yield sooting index: YSI$\leq$ 70; (i) cetane number: CN$\geq$ 40.

The Tier 1 screening selects 129 compounds from UOB Fuel Property Database containing 1797 pure compounds. Tier 2 screening demands high reactivity (namely low ignition delay time), combustion with high φ-gradient (namely high φ-sensitivity), high dilution tolerance (high premixed laminar flame speed). The scores of

the 8 selected fuel candidates in Tier 2 screening arrange from high to low as: dibutyl ether > n-heptane > n-octane > 2-butyltetrahydrofuran > butylcyclohexane > 1-octanol > PODE3 > PODE4. 50%n-heptane-40%dibutyl ether-10%ethanol is formulated as a test fuel and applies into CI engine test and the results indicate that it improves the indicated thermal efficiency (ITE) and reduces NO_x formation by shortening the combustion duration and decreasing the peak combustion temperature.

9.1.6 Ignition and flame characteristics of polyoxymethylene dimethyl ether 3 (PODE3)

PODE3 is one of the recommended candidates for CI engines by the high throughput fuel screening tool and its ignition and flame characteristics are compared with n-heptane by chemical kinetic modeling in this work. The main conclusions are summarized as below:

(1) PODE3 is more reactive than n-heptane and results in lower ignition delay times in the full temperature regime. The ignition delay times of PODE3 and n-heptane are expressions in Arrhenius format as

$$\tau_{ig\ PODE_3} = \frac{1}{\kappa_{PODE_3}} = A \cdot exp\left(\frac{E_a}{8.314 \cdot T}\right) \cdot p^m \cdot \varphi^n = 0.015 \cdot exp\left(\frac{53216.492}{8.314 \cdot T}\right) \cdot p^{-0.96} \cdot \varphi^{-1.143} \qquad \text{and}$$

$$\tau_{ig\ nC_7H_{16}} = \frac{1}{\kappa_{nC_7H_{16}}} = A \cdot exp\left(\frac{E_a}{8.314 \cdot T}\right) \cdot p^m \cdot \varphi^n = 9.881E-08 \cdot exp\left(\frac{112889.509}{8.314 \cdot T}\right) \cdot p^{-0.341} \cdot \varphi^{-0.342}$$

respectively. The higher pre-exponential factor and lower activation energy of PODE3 than n-heptane indicate that the PODE3 exceeds n-heptane in fuel reactivity.

(2) Even though PODE3 exists high overall oxidation reactivity, but it is not suitable to act as a chemical ignition source in the fuel mixture due to insufficient temperature rise and active radical accumulation during low temperature heat release (LTHR).

(3) PODE3 reaches higher adiabatic flame temperature (AFT) than n-heptane at most conditions except at near stoichiometry of φ=0.9~1.1 and the AFT difference at φ=0.9~1.1 is as minor as 5~20°C.

(4) PODE3 has higher premixed laminar flame speed than n-heptane due to higher global reactivity and AFT which benefits the application of PODE3 into low temperature combustion (LTC) mode due to better dilution tolerance.

9.1.7 Chemical kinetic modeling of diethoxymethane (DEM) oxidation

Similar to PODE3, DEM is one of the recommended candidates for CI engines by the high throughput fuel screening tool and both of them are carbon-neutral fuels. This work proposes a detailed chemical kinetic mechanism of DEM with low and high temperature chemistry to describe the ignition behavior at a full temperature regime. The DEM mechanism development is based on a hierarchical structure containing C0-C4 core mechanism, n-/i-paraffins sub-mechanism, monoaromatics sub-mechanism, PAHs sub-mechanism, alcohols sub-mechanism, dimethoxymethane sub-mechanism, diethoxymethane sub-mechanism. The DEM reaction pathway is clarified as below: At low temperature of 550~640K, H atom abstraction from DEM molecule takes place to form DEMx. The 1st O_2 addition occurs on DEMx to produce DEMxO2 and then undergoes isomerization to form QOOH. The 2nd O_2 happens on QOOH to form O2QOOH. The low temperature chain branching reaction takes place as O2QOOH decomposes to carbonylhydroperoxide and hydroxyl radical, then it furthers decomposes to oxygenated radical and hydroxyl radical. At the intermediate temperature of 640~960K, the 2nd O_2 addition reaction is interrupted and QOOH decomposes to form cyclic ethers, acyclic ethers and β-scission products. At a high temperature of 960~1250K, the 1st O_2 addition reaction is suppressed and the DEM molecule, DEMx radicals decompose by breaking the C-O bond. The DEM mechanism containing high temperature chemistry is also provided to emulate the high temperature flame phenomenon.

9.2 Suggestions for future research

9.2.1 Promote the integration of artificial intelligence (AI) techniques and fuel research

As discussed in section 1.2.2, the goal of the Fuel Genome Project is to address the forward problem of fuel property prediction and the inverse problems of molecule design, retrosynthesis, reaction condition design. This work proposes two solutions to address the forward problem: (1) At first, the QSPR theory is used to develop the molecules as molecular descriptors and convert the molecules into a numerical representation. Then adopt ML algorithms are adopted to correlate the structural feature and the property feature and establish the regression model. This solution has been applied to 15 physicochemical properties and a good agreement between the observed and forecasting results is obtained as shown in chapter 3 and chapter 4. (2) The CNN is directly applied to a set of molecule graphs of interest to learn the molecular feature automatically and perform the regression operation. This technical pathway has been applied to the YSI prediction as discussed in chapter 5. In summary, this work along with other published literature has made great efforts and obtained some exciting achievements on the forward problem of property prediction.

The objective and technology roadmap of property prediction and molecule design are presented in Figure 9.1. The molecular descriptor scheme needs to be upgraded to simultaneously address the forward problem of property prediction and the inverse problem of molecule design. The improved molecular descriptor scheme should meet the biuniqueness requirement namely each fuel molecule has a unique numerical representation and each numerical representation corresponds to a unique fuel molecule. In other words, the new molecular descriptor scheme should issue a unique digital identity to the fuel molecules. The simplified molecular-input line-entry (SMILE) system is a commonly used molecular descriptor that meets the biuniqueness requirement for the Fuel Genome Project. A new challenge is to convert the line notations of SMILE system into a numerical representation.

Regarding the 1st inverse problem of molecule design, there are two technical routes of ML-QSPR and DL-CNN to address this problem as shown in Figure 9.1. The principle and developed procedure of ML-QSPR and DL-CNN have been discussed in chapter 4 and chapter 5 in detail, they won't be covered here. The key

improvement and difference in the future work is to adopt the new molecular descriptor scheme with the characteristic of biuniqueness. On the contrary, there is a huge knowledge gap in the inverse problems to be filled by the researchers. Regarding the 1st inverse problem of molecule design, there is no available program or software available to design the molecular structure to achieve desired properties. Based on the experience from Material Genome Initiative and Polymer Genome, there are two technical pathways to address this problem as shown in Figure 9.1: One solution is to use an RNN encoder to create the continuous molecular representation latent space and then use RNN decoder to decode the structural information of the point in the latent space [32, 556]. The surrogate model conducts a gradient-based optimization in the continuous latent space to find out the highest latent representation with desired properties. Then the new representation is decoded into SMILE strings and corresponding molecule which will be validated experimentally. The other solution is to integrate genetic algorithms and ML algorithms to produce multiple generations of molecules, then a high-throughput screening is conducted to identify those molecules with desired properties [26, 557]. Even though the computational toolkits developed by Material Genome Initiative and Polymer Genome cannot directly transfer to the fuel field, but these examples inspire the technological roadmap for fuel molecule design to address desired properties.

Regarding the 2nd and 3rd inverse problems of retrosynthesis and reaction condition design, there are a few programs and software available for retrosynthetic analysis as summarized in section 1.3.4, but only the RetSynth tool is specially developed to handle the fossil and bio-based fuels. The tailor-made retrosynthesis tool for fuel molecules is needed to rapidly identify the viable reaction pathway to produce fuels of interest.
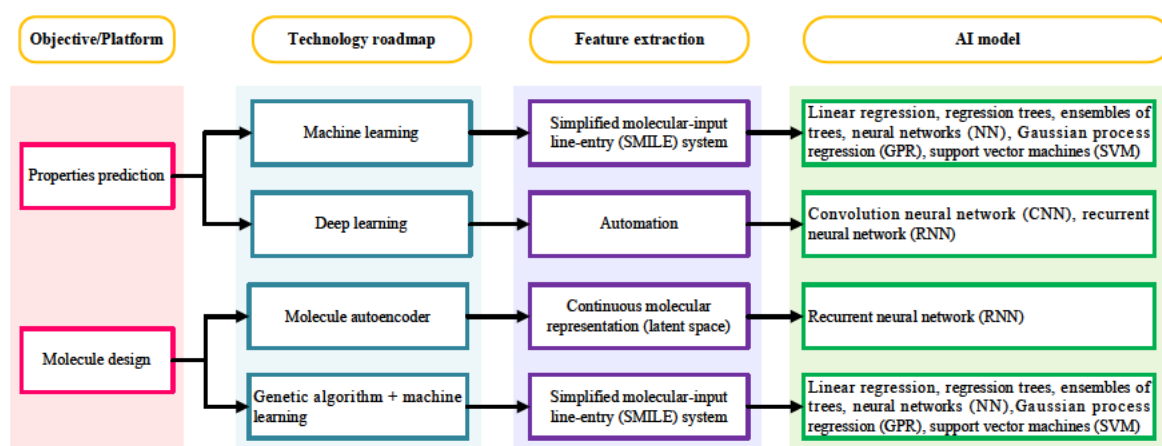
Figure 9.1. Overview of the objective and technology roadmap of the Fuel Genome Project.

9.2.2 Develop high throughput fuel screening tool applicable to fuel mixtures and expand the fuel physicochemical property database

The high throughput fuel screening tool developed in chapter 6 is based on the 15 fuel properties regression models proposed in chapter 4. Its scope is limited to pure compounds because there are no property data available for mixtures except for CN, RON, MON. But the market fuels are a blend of components and their properties need to be optimized by formulating the fuel components rather than just predicting the properties of individual components. This is a critical issue of high-grade fuel formulation to address highly advanced emission control requirements. The fuel properties as a function of blending ratio are nonlinear, the best-known example is the effect of blending ethanol into mineral gasoline on dry vapor pressure equivalent (DVPE). The principle of QSPR guarantees it can apply to fuel mixture and thus the Fuel Property Database needs to incorporate the mixture properties.

9.2.3 Calibrate the rate constants of elementary reactions in diethoxymethane sub-mechanism by jet-stirred reactor experiment and quantum chemistry

This DEM mechanism contains both high temperature and low temperature mechanisms, so it enables to compute ignition delay time at engine relevant conditions and premixed laminar flame speed at ambient to mediate pressure. But it is still far from perfect, there are some knowledge gaps in DEM research as listed:

(1) Jet stirred reactor experiment of DEM oxidation at 500~1200K is needed to obtain species evolution data for mechanism development.

(2) Due to a lack of speciation data, the rate constants of elementary reaction in DEM sub-mechanism are semi-empirical. Quantum chemistry technology (like Gaussian 03 or ReaxFF) tool is needed to determine the reaction rate constant of DEM sub-mechanism. But for other sub-mechanisms are not needed, because they are derived from well-validated mechanisms based on a hierarchical structure.

Appendix

Appendix A Measured and predicted values of T_b, LHV for hydrocarbons, alcohols, ethers, aldehydes, ketones, esters

Class	Name	CAS	Formula	Tb (°C)$^\alpha$	Tb (°C)$^\beta$	LHV (kJ/mol) $^\alpha$	LHV (kJ/mol) $^\beta$
n-Alkanes	Methane	74-82-8	CH4	-161.5	-91.8	-803	-820
n-Alkanes	Ethane	74-84-0	C2H6	-88.6	-60.5	-1429	-1425
n-Alkanes	Propane	74-98-6	C3H8	-42.1	-28.0	-2043	-2033
n-Alkanes	n-Butane	106-97-8	C4H10	-0.5	4.8	-2657	-2642
n-Alkanes	n-Pentane	109-66-0	C5H12	36.1	37.2	-3245	-3250
n-Alkanes	n-Hexane	110-54-3	C6H14	68.7	68.5	-3855	-3859
n-Alkanes	n-Heptane	142-82-5	C7H16	98.4	98.4	-4465	-4467
n-Alkanes	n-Octane	111-65-9	C8H18	125.6	126.4	-5074	-5076
n-Alkanes	n-Nonane	111-84-2	C9H20	150.8	152.4	-5685	-5684
n-Alkanes	n-Decane	124-18-5	C10H22	174.1	176.2	-6294	-6293
n-Alkanes	n-Undecane	1120-21-4	C11H24	195.9	198.0	-6904	-6901
n-Alkanes	n-Dodecane	112-40-3	C12H26	216.3	218.0	-7514	-7510
n-Alkanes	n-Tridecane	629-50-5	C13H28	235.4	236.5	-8123	-8118
n-Alkanes	n-Tetradecane	629-59-4	C14H30	253.5	253.9	-8733	-8727
n-Alkanes	n-Pentadecane	629-62-9	C15H32	270.6	270.3	-9342	-9335
n-Alkanes	n-Hexadecane	544-76-3	C16H34	286.9	286.2	-9952	-9944
n-Alkanes	n-Heptadecane	629-78-7	C17H36	303.0	301.6	-10600	-10552
n-Alkanes	n-Octadecane	593-45-3	C18H38	316.0	316.8	-11200	-11161
n-Alkanes	n-Nonadecane	629-92-5	C19H40	330.0	331.8	-11800	-11769
n-Alkanes	n-Eicosane	112-95-8	C20H42	344.1	346.6	-12400	-12378
n-Alkanes	n-Heneicosane	629-94-7	C21H44	359.0	360.9	-13000	-12986
n-Alkanes	n-Docosane	629-97-0	C22H46	369.0	374.8	-13600	-13595
n-Alkanes	n-Tricosane	638-67-5	C23H48	451.0	388.0	-14200	-14203
n-Alkanes	n-Tetracosane	646-31-1	C24H50	391.0	400.3	-14800	-14812
n-Alkanes	n-Pentacosane	629-99-2	C25H52	401.9	411.7	-15400	-15420
n-Alkanes	n-Hexacosane	630-01-3	C26H54	415.0	422.1	-16000	-16029
n-Alkanes	n-Heptacosane	593-49-7	C27H56	442.0	431.4	-16600	-16637
n-Alkanes	n-Octacosane	630-02-4	C28H58	432.0	439.6	-17200	-17246
n-Alkanes	n-Nonacosane	630-03-5	C29H60	443.0	446.7	-17800	-17855
n-Alkanes	n-Triacontane	638-68-6	C30H62	451.0	452.7	-18400	-18463
n-Alkanes	Dotriacontane	544-85-4	C32H66	470.0	461.7	-19600	-19680
n-Alkanes	n-Hexatriacontane	630-06-8	C36H74	N/A	469.6	-22100	-22114
2-Methylalkanes	2-Methylpropane	75-28-5	C4H10	-11.7	4.9	-2649	-2638
2-Methylalkanes	2-Methylbutane	78-78-4	C5H12	27.8	34.7	-3240	-3247
2-Methylalkanes	2-Methylpentane	107-83-5	C6H14	60.2	64.3	-3849	-3855
2-Methylalkanes	2-Methylhexane	591-76-4	C7H16	90.0	92.9	-4460	-4464

2-Methylalkanes	2-Methylheptane	592-27-8	C8H18	117.6	120.2	-5069	-5072
2-Methylalkanes	2-Methyloctane	3221-61-2	C9H20	143.0	145.8	-5679	-5681
2-Methylalkanes	2-Methylnonane	871-83-0	C10H22	167.0	169.3	-6289	-6289
2-Methylalkanes	2-Methyldecane	6975-98-0	C11H24	189.2	190.7	N/A	-6898
2-Methylalkanes	2-Methylundecane	7045-71-8	C12H26	211.0	210.0	N/A	-7506
3-Methylalkane	3-Methylpentane	96-14-0	C6H14	63.3	58.2	-3851	-3858
3-Methylalkane	3-Methylhexane	589-34-4	C7H16	92.0	87.1	-4463	-4467
3-Methylalkane	3-Methylheptane	589-81-1	C8H18	118.9	114.9	-5072	-5075
3-Methylalkane	3-Methyloctane	2216-33-3	C9H20	144.0	141.1	-5681	-5684
3-Methylalkane	3-Methylnonane	5911-04-6	C10H22	168.0	165.3	-6291	-6292
3-Methylalkane	3-Methyldecane	13151-34-3	C11H24	N/A	187.4	N/A	-6901
3-Methylalkane	3-Methylundecane	1002-43-3	C12H26	212.0	207.4	-7511	-7509
2,2-Dimetylalkanes	2,2-Dimethylpropane	463-82-1	C5H12	9.5	32.1	-3250	-3246
2,2-Dimetylalkanes	2,2-dimethylbutane	75-83-2	C6H14	49.7	58.8	-3841	-3854
2,2-Dimetylalkanes	2,2-dimethylpentane	590-35-2	C7H16	79.2	85.3	-4451	-4463
2,2-Dimetylalkanes	2,2-dimethylhexane	590-73-8	C8H18	106.8	111.2	-5063	-5071
2,2-Dimetylalkanes	2,2-Dimethylheptane	1071-26-7	C9H20	133.0	136.0	-5672	-5680
2,2-Dimetylalkanes	2,2-dimethyloctane	15869-87-1	C10H22	154.0	159.6	-6281	-6288
2,2-Dimetylalkanes	2,2-Dimethylnonane	17302-14-6	C11H24	N/A	181.7	N/A	-6897
2,2-Dimetylalkanes	2,2-Dimethyldecane	17302-37-3	C12H26	N/A	202.2	N/A	-7505
2-Methylalkenes	2-Methylpropene	115-11-7	C4H8	-7.0	7.4	-2524	-2518
2-Methylalkenes	2-Methyl-1-butene	563-46-2	C5H10	31.1	37.5	-3116	-3127
2-Methylalkenes	2-methyl-1-pentene	763-29-1	C6H12	62.1	67.2	-3722	-3736
2-Methylalkenes	2-Methyl-1-hexene	6094-02-6	C7H14	92.0	96.0	-4335	-4344
2-Methylalkenes	2-methyl-1-heptene	15870-10-7	C8H16	119.0	123.5	-4948	-4953
2-Methylalkenes	2-Methyl-1-octene	4588-18-5	C9H18	144.0	149.2	-5530	-5561
2-Methylalkenes	2-Methyl-1-nonene	2980-71-	C10H20	167.0	172.8	-6165	-6170

		4					
2-Methylalkenes	2-Methyl-1-decene	13151-27-4	C11H22	N/A	194.2	N/A	-6778
2-Methylalkenes	2-Methyl-1-undecene	18516-37-5	C12H24	N/A	213.3	N/A	-7387
1-Alkenes	Ethene	74-85-1	C2H4	-103.8	-55.2	-1323	-1308
1-Alkenes	Propene	115-07-1	C3H6	-47.6	-25.7	-1926	-1913
1-Alkenes	1-Butene	106-98-9	C4H8	-6.3	4.8	-2541	-2522
1-Alkenes	1-Pentene	109-67-1	C5H10	30.0	35.3	-3130	-3130
1-Alkenes	1-hexene	592-41-6	C6H12	63.4	65.1	-3740	-3739
1-Alkenes	1-heptene	592-76-7	C7H14	94.0	93.7	-4350	-4347
1-Alkenes	1-octene	111-66-0	C8H16	121.3	120.8	-4961	-4956
1-Alkenes	1-nonene	124-11-8	C9H18	146.9	145.9	-5572	-5564
1-Alkenes	1-decene	872-05-9	C10H20	171.0	169.0	-6181	-6173
1-Alkenes	1-Undecene	821-95-4	C11H22	192.7	189.8	-6788	-6781
1-Alkenes	1-dodecene	112-41-4	C12H24	213.4	208.5	-7398	-7390
1-Alkynes	Acetylene	74-86-2	C2H2	-84.7	-41.6	-1257	-1231
1-Alkynes	Methylacetylene	74-99-7	C3H4	-23.2	-7.5	-1849	-1836
1-Alkynes	Ethylacetylene	107-00-6	C4H6	8.1	18.2	-2465	-2444
1-Alkynes	1-pentyne	627-19-0	C5H8	39.9	44.6	-3051	-3053
1-Alkynes	1-Hexyne	693-02-7	C6H10	71.2	71.4	-3661	-3661
1-Alkynes	1-Heptyne	628-71-7	C7H12	99.8	98.0	-4272	-4270
1-Alkynes	1-octyne	629-05-0	C8H14	126.2	124.2	-4882	-4879
1-Alkynes	1-Nonyne	3452-09-3	C9H16	150.8	149.7	-5493	-5487
1-Alkynes	1-Decyne	764-93-2	C10H18	174.0	174.2	-6104	-6096
1-Alkynes	1-Undecyne	2243-98-3	C11H20	N/A	197.4	N/A	-6704
1-Alkynes	1-Dodecyne	765-03-7	C12H22	N/A	219.3	N/A	-7313
Cycloalkanes	Cyclopropane	75-19-4	C3H6	-31.0	-18.5	-1959	-1915
Cycloalkanes	Cyclobutane	287-23-0	C4H8	12.5	20.4	-2568	-2482
Cycloalkanes	cyclopentane	287-92-3	C5H10	49.2	56.7	-3071	-3048
Cycloalkanes	cyclohexane	110-82-7	C6H12	80.7	90.0	-3656	-3615
Cycloalkanes	cycloheptane	291-64-5	C7H14	118.8	120.9	-4290	-4182
Cycloalkanes	cyclooctane	292-64-8	C8H16	151.1	149.6	-4913	-4749
Cycloalkanes	Cyclononane	293-55-0	C9H18	173.0	176.0	-5536	-5316
Cycloalkanes	Cyclodecane	293-96-9	C10H20	202.3	199.3	N/A	-5882
Cycloalkanes	Cycloundecane	294-41-7	C11H22	N/A	219.1	N/A	-6449
Cycloalkanes	Cyclododecane	294-62-2	C12H24	N/A	234.9	N/A	-7016
Alkylbenzenes	Benzene	71-43-2	C6H6	80.1	95.4	-3136	-3053
Alkylbenzenes	Toluene	108-88-3	C7H8	110.6	118.2	-3734	-3739
Alkylbenzenes	Ethylbenzene	100-41-4	C8H10	136.2	140.4	-4345	-4347
Alkylbenzenes	n-Propylbenzene	103-65-1	C9H12	159.2	162.6	-4954	-4956
Alkylbenzenes	n-Butylbenzene	104-51-8	C10H14	183.3	184.5	-5564	-5564
Alkylbenzenes	n-Pentylbenzene	538-68-1	C11H16	203.0	205.8	-6174	-6173

Alkylbenzenes	n-Hexylbenzene	1077-16-3	C12H18	226.0	226.4	-6784	-6782
1-Alkanols	Methanol	67-56-1	CH4O	64.5	44.6	-638	-629
1-Alkanols	Ethanol	64-17-5	C2H6O	78.2	70.3	-1235	-1238
1-Alkanols	1-Propanol	71-23-8	C3H8O	97.0	95.5	-1844	-1846
1-Alkanols	1-Butanol	71-36-3	C4H10O	117.6	119.9	-2456	-2455
1-Alkanols	1-Pentanol	71-41-0	C5H12O	137.6	143.2	-3061	-3063
1-Alkanols	1-Hexanol	111-27-3	C6H14O	156.9	165.0	-3677	-3672
1-Alkanols	1-Heptanol	111-70-6	C7H16O	178.0	185.4	-4286	-4280
1-Alkanols	1-Octanol	111-87-5	C8H18O	194.7	204.1	-4898	-4889
1-Alkanols	1-Nonanol	143-08-8	C9H20O	213.7	221.4	-5501	-5497
1-Alkanols	1-Decanol	112-30-1	C10H22O	229.0	237.2	-6116	-6106
1-Alkanols	1-Undecanol	112-42-5	C11H24O	246.0	251.9	-6726	-6714
1-Alkanols	1-Dodecanol	112-53-8	C12H26O	264.1	265.7	-7338	-7323
1-Alkanols	1-Tridecanol	112-70-9	C13H28O	287.0	278.7	-7901	-7931
1-Alkanols	1-Tetradecanol	112-72-1	C14H30O	295.8	291.2	-8491	-8540
1-Alkanols	1-Pentadecanol	629-76-5	C15H32O	318.0	303.5	-9114	-9148
1-Alkanols	1-Hexadecanol	36653-82-4	C16H34O	N/A	315.6	-9724	-9757
1-Alkanols	1-Heptadecanol	1454-85-9	C17H36O	324.0	327.6	-10300	-10365
1-Alkanols	1-Octadecanol	112-92-5	C18H36O	351.0	339.5	-10900	-10974
1-Alkanols	1-Nonadecanol	1454-84-8	C19H40O	345.0	351.3	-11500	-11583
1-Alkanols	1-Eicosanol	629-96-9	C20H42O	356.0	362.9	-12100	-12191
1-Alkanols	1-Docosanol	661-19-8	C22H46O	N/A	385.1	N/A	-13408
1-Alkanols	1-Hexacosanol	506-52-5	C26H54O	N/A	422.3	N/A	-15842
2-Alkanols	2-Propanol	67-63-0	C3H8O	82.2	82.2	-1830	-1833
2-Alkanols	2-Butanol	78-92-2	C4H10O	99.4	103.7	-2441	-2442
2-Alkanols	2-Pentanol	6032-29-7	C5H12O	119.1	124.4	-3052	-3050
2-Alkanols	2-Hexanol	626-93-7	C6H14O	N/A	144.2	-3666	-3659
2-Alkanols	2-Heptanol	543-49-7	C7H16O	159.0	162.7	-4330	-4267
2-Alkanols	2-Octanol	123-96-6	C8H18O	179.0	179.7	-4880	-4876
2-Alkanols	2-Nonanol	628-99-9	C9H20O	193.5	195.1	-5490	-5485
2-Alkanols	2-Decanol	1120-06-5	C10H22O	N/A	209.0	N/A	-6093
2-Alkanols	2-Undecanol	1653-30-1	C11H24O	231.0	221.3	N/A	-6702
2-Alkanols	2-Dodecanol	10203-28-8	C12H26O	249.0	232.2	N/A	-7310
Acyclic ethers	Dimethyl ether	115-10-6	C2H6O	-24.8	-22.7	-1328	-1296
Acyclic ethers	Diethyl ether	60-29-7	C4H10O	34.4	31.1	-2504	-2516
Acyclic ethers	Dipropyl ether	111-43-3	C6H14O	90.1	84.2	-3725	-3736
Acyclic ethers	Dibutyl ether	142-96-1	C8H18O	141.6	134.5	-4947	-4956

Acyclic ethers	Di-n-pentyl ether	693-65-2	C10H22O	187.0	180.3	-6170	-6175
Acyclic ethers	Di-n-hexyl ether	112-58-3	C12H26O	220.0	220.7	-7384	-7395
Aldehydes	Formaldehyde	50-00-0	CH2O	-19.1	8.2	-527	-478
Aldehydes	Acetaldehyde	75-07-0	C2H4O	20.8	30.5	-1105	-1083
Aldehydes	Propanal	123-38-6	C3H6O	48.0	54.4	-1686	-1691
Aldehydes	Butanal	123-72-8	C4H8O	74.8	77.3	-2304	-2300
Aldehydes	Pentanal	110-62-3	C5H10O	103.0	98.6	-2910	-2908
Aldehydes	Hexanal	66-25-1	C6H12O	129.6	118.0	-3520	-3517
Aldehydes	Heptanal	111-71-7	C7H14O	153.0	135.2	-4136	-4125
Aldehydes	Octanal	124-13-0	C8H16O	174.0	150.2	-4740	-4734
Aldehydes	Nonanal	124-19-6	C9H18O	195.0	163.0	-5350	-5343
Aldehydes	Decanal	112-31-2	C10H20O	212.0	173.8	-5959	-5951
Aldehydes	Undecanal	112-44-7	C11H22O	117.0	182.9	-6570	-6560
Aldehydes	Dodecanal	112-54-9	C12H24O	185.0	190.5	-7179	-7168
2-Ketones	2-Propanone	67-64-1	C3H6O	56.1	54.8	-1659	-1663
2-Ketones	2-Butanone	78-93-3	C4H8O	79.6	80.7	-2268	-2272
2-Ketones	2-Pentanone	107-87-9	C5H10O	102.2	106.1	-2880	-2880
2-Ketones	2-Hexanone	591-78-6	C6H12O	127.6	130.7	-3490	-3489
2-Ketones	2-Heptanone	110-43-0	C7H14O	151.0	154.1	-4100	-4097
2-Ketones	2-Octanone	111-13-7	C8H16O	173.0	176.0	-4698	-4706
2-Ketones	2-Nonanone	821-55-6	C9H18O	194.0	196.2	-5321	-5314
2-Ketones	2-Decanone	693-54-9	C10H20O	211.0	214.5	N/A	-5923
2-Ketones	2-Undecanone	112-12-9	C11H22O	233.1	230.9	N/A	-6531
2-Ketones	2-Dodecanone	6175-49-1	C12H24O	N/A	245.3	N/A	-7140
Methyl esters	Methyl formate	107-31-3	C2H4O2	31.6	25.6	-892	-862
Methyl esters	Methyl acetate	79-20-9	C3H6O2	56.7	52.0	-1461	-1467
Methyl esters	Methyl propionate	554-12-1	C4H8O2	78.6	77.9	-2078	-2076
Methyl esters	Methyl butanoate	623-42-7	C5H10O2	101.9	103.1	-2550	-2684
Methyl esters	Methyl pentanoate	624-24-8	C6H12O2	127.4	127.1	N/A	-3293
Methyl esters	Methyl hexanoate	106-70-7	C7H14O2	151.0	149.6	N/A	-3901
Methyl esters	Methyl heptanoate	106-73-0	C8H16O2	169.7	170.0	N/A	-4510
Methyl esters	Methyl octanoate	111-11-5	C9H18O2	194.1	188.4	N/A	-5118
Methyl esters	Methyl nonanoate	1731-84-6	C10H20O2	N/A	204.7	N/A	-5727
Methyl esters	Methyl decanoate	110-42-9	C11H22O2	233.0	218.9	-6349	-6335
Methyl esters	Methyl undecanoate	1731-86-8	C12H24O2	N/A	231.4	N/A	-6944

$^\alpha$ Measured values; $^\beta$ predicted values.

Data and Software Availability

The research data and software generated during the current study have been archived in the google cloud and they are available from the author Runzhao Li (Email: upon reasonable request and with permission of Prof. Athanasios Tsolakis (Email: and Dr. Jose Martin Herreros (Email: Most of them are publicly available in the author's publications, please refer to the author publications page.

References

[1] Dempsey AB, Curran SJ, Wagner RM. A perspective on the range of gasoline compression ignition combustion strategies for high engine efficiency and low NOx and soot emissions: Effects of in-cylinder fuel stratification. International Journal of Engine Research 2016;17(8):897-917.

[2] Wissink ML, Curran SJ, Roberts G, Musculus MPB, Mounaïm-Rousselle C. Isolating the effects of reactivity stratification in reactivity-controlled compression ignition with iso-octane and n-heptane on a light-duty multi-cylinder engine. International Journal of Engine Research 2017;19(9):907-26.

[3] Moses-DeBusk M, Curran SJ, Lewis SA, Connatser RM, Storey JME. Impacts of Air-Fuel Stratification in ACI Combustion on Particulate Matter and Gaseous Emissions. Emission Control Science and Technology 2019;5(3):225-37.

[4] Kokjohn SL, Hanson RM, Splitter DA, Reitz RD. Fuel reactivity controlled compression ignition (RCCI): a pathway to controlled high-efficiency clean combustion. International Journal of Engine Research 2011;12(3):209-26.

[5] Gasteiger J, Engel T. Chemoinformatics: A Textbook. Weinheim: Wiley; 2003.

[6] Panerati J, Schnellmann MA, Patience C, Beltrame G, Patience GS. Experimental methods in chemical engineering: Artificial neural networks–ANNs. The Canadian Journal of Chemical Engineering 2019;97(9):2372-82.

[7] Nucleic Acids https://courses.lumenlearning.com/wmopen-nmbiology1/chapter/nucleic-acids/.

[8] Polymer https://www.sciencenewsforstudents.org/article/scientists-say-polymer.

[9] What is the Human Genome Project? National Institutes of Health (NIH): National Human Genome Research Institute:https://www.genome.gov/human-genome-project/What.

[10] International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. Nature 2004;431(7011):931-45.

[11] Human Genome Project Timeline of Events. National Institutes of Health (NIH): National Human Genome Research Institute:https://www.genome.gov/human-genome-project/Timeline-of-Events.

[12] Fine R, Fivenson E. Lessons from the Human Genome Project. 2019:https://sitn.hms.harvard.edu/flash/2019/lessons-from-the-human-genome-project/.

[13] MGI white paper: Materials Genome Initiative for Global Competitiveness. National Science and Technology Council 2011:https://www.mgi.gov/sites/default/files/documents/materials_genome_initiative-final.pdf.

[14] Materials Genome Initiative Strategic Plan. 2014:https://www.nist.gov/system/files/documents/2017/05/09/MGI-StrategicPlan-4.pdf.

[15] Liu Y, Esan OC, Pan Z, An L. Machine learning for advanced energy materials. Energy and AI 2021;3.

[16] Liu Y, Zhao T, Ju W, Shi S. Materials discovery and design using machine learning. Journal of Materiomics 2017;3(3):159-77.

[17] de Pablo JJ, Jackson NE, Webb MA, Chen L-Q, Moore JE, Morgan D, et al. New frontiers for the materials genome initiative. npj Computational Materials 2019;5(1).

[18] Materials Genome Initiative.https://www.mgi.gov/.

[19] Materials Innovation Ecosystem https://research.gatech.edu/materials/MGI.

[20] Chen L, Pilania G, Batra R, Huan TD, Kim C, Kuenneth C, et al. Polymer informatics: Current status and critical next steps. Materials Science and Engineering: R: Reports 2021;144.

[21] Batra R, Song L, Ramprasad R. Emerging materials intelligence ecosystems propelled by machine learning. Nature Reviews Materials 2020.

[22] Mannodi-Kanakkithodi A, Chandrasekaran A, Kim C, Huan TD, Pilania G, Botu V, et al. Scoping the polymer genome: A roadmap for rational polymer dielectrics design and beyond. Materials Today

2018;21(7):785-96.

[23] Kim C, Chandrasekaran A, Huan TD, Das D, Ramprasad R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. The Journal of Physical Chemistry C 2018;122(31):17575-85.

[24] Huan TD, Mannodi-Kanakkithodi A, Kim C, Sharma V, Pilania G, Ramprasad R. A polymer dataset for accelerated property prediction and design. Sci Data 2016;3:160012.

[25] Doan Tran H, Kim C, Chen L, Chandrasekaran A, Batra R, Venkatram S, et al. Machine-learning predictions of polymer properties with Polymer Genome. Journal of Applied Physics 2020;128(17).

[26] Kim C, Batra R, Chen L, Tran H, Ramprasad R. Polymer design using genetic algorithm and machine learning. Computational Materials Science 2021;186.

[27] Polymer Genome: An informatics platform for polymer property prediction and design using machine learning.https://www.polymergenome.org/.

[28] Dehmer M, Varmuza K, Bonchev D, Emmert-Streib F. Statistical Modelling of Molecular Descriptors in QSAR-QSPR. 2012.

[29] Nieto-Draghi C, Fayet G, Creton B, Rozanska X, Rotureau P, de Hemptinne J-C, et al. A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes. Chem Rev 2015;115(24):13093-164.

[30] Engel T, Gasteiger J. Applied Chemoinformatics: Achievements and Future Opportunities. Weinheim: Wiley; 2018.

[31] Randić M. Molecular bonding profiles. Journal of Mathematical Chemistry 1996;19(3):375-92.

[32] Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: Generative models for matter engineering. Science 2018;361(6400):360-5.

[33] Schütt KT, Chmiela S, Lilienfeld OAv, Tkatchenko A, Tsuda K, Müller K-R. Machine Learning Meets Quantum Physics. Switzerland: Springer; 2020.

[34] Moriwaki H, Tian YS, Kawashita N, Takagi T. Mordred: a molecular descriptor calculator. J Cheminform 2018;10(1):4.

[35] Hinselmann G. BlueDesc - Molecular Descriptor Calculator. University of Tübingen 2003:http://www.ra.cs.uni-tuebingen.de/software/bluedesc/.

[36] ChemDes. 2015:http://www.scbdd.com/chemdes/.

[37] Dong J, Cao DS, Miao HY, Liu S, Deng BC, Yun YH, et al. ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. J Cheminform 2015;7:60.

[38] Cao DS, Xu QS, Hu QN, Liang YZ. ChemoPy: freely available python package for computational biology and chemoinformatics. Bioinformatics 2013;29(8):1092-4.

[39] ChemoPy.https://github.com/ifyoungnet/Chemopy.

[40] Cinfony https://cinfony.github.io/.

[41] Dragon 7.https://chm.kode-solutions.net/products_dragon.php.

[42] Mordred.https://github.com/mordred-descriptor/mordred.

[43] PaDEL http://www.yapcwsoft.com/dd/padeldescriptor/.

[44] Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 2011;32(7):1466-74.

[45] PyDPI. Central South University 2012:https://pypi.org/project/pydpi/.

[46] Cao DS, Liang YZ, Yan J, Tan GS, Xu QS, Liu S. PyDPI: freely available python package for chemoinformatics, bioinformatics, and chemogenomics studies. J Chem Inf Model 2013;53(11):3086-96.

[47] Rcpi.https://bioconductor.org/packages/release/bioc/html/Rcpi.html.

[48] Cao DS, Xiao N, Xu QS, Chen AF. Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds and their interactions. Bioinformatics 2015;31(2):279-81.

[49] Deep Learning and Traditional Machine Learning: Choosing the Right Approach. MathWorks:https://uk mathworks.com/campaigns/offers/deep-learning-vs-machine-learning-algorithm.html?s_tid=mo_dl_pers_vid_ebook&elqCampaignId=6443.

[50] Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. Chem Rev 2019;119(18):10520-94.

[51] Lu H, Guo L, Azimi M, Huang K. Oil and Gas 4.0 era: A systematic review and outlook. Computers in Industry 2019;111:68-90.

[52] Statistics and Machine Learning Toolbox User's Guide. MathWorks, Inc 2020:https://uk mathworks.com/help/pdf_doc/stats/stats.pdf.

[53] Kim P. MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence. New York: Apress; 2017.

[54] Turner RE. Sparse Gaussian Process Approximations. Gaussian Process Summer School 2017 2017:http://gpss.cc/gpss17/slides/gp-approx-new.pdf.

[55] Fitting the Multiple Linear Regression Model. Statistics Knowledge Portal:https://www.jmp.com/en_ch/statistics-knowledge-portal/what-is-multiple-regression/fitting-multiple-regression-model html.

[56] Molecular design software. WIKIPEDIA:https://en.wikipedia.org/wiki/Molecular_design_software.

[57] AiZynthFinder https://github.com/MolecularAI/aizynthfinder.

[58] Genheden S, Thakkar A, Chadimova V, Reymond JL, Engkvist O, Bjerrum E. AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. J Cheminform 2020;12(1):70.

[59] ASKCOS web version. Machine Learning for Pharmaceutical Discovery and Synthesis Consortium:https://askcos mit.edu/.

[60] ASKCOS software package.https://github.com/ASKCOS/ASKCOS.

[61] Chemical.AI https://free.chemical.ai/.

[62] IBM RXN https://rxn res.ibm.com/.

[63] ICSYNTH https://www.deepmatter.io/products/icsynth/.

[64] Bøgevig A, Federsel H-J, Huerta F, Hutchings MG, Kraut H, Langer T, et al. Route Design in the 21st Century: The ICSYNTH Software Tool as an Idea Generator for Synthesis Prediction. Organic Process Research & Development 2015;19(2):357-68.

[65] Molecule.one.https://molecule.one/.

[66] RetSynth Tool. Sandia Corporation 2019:https://github.com/sandialabs/RetSynth.

[67] Law J, Zsoldos Z, Simon A, Reid D, Liu Y, Khew SY, et al. Route Designer: a retrosynthetic analysis tool utilizing automated retrosynthetic rule generation. J Chem Inf Model 2009;49(3):593-602.

[68] SciFindern-Retrosynthesis.https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder/retrosynthesis-planning.

[69] Spaya https://spaya.ai/.

[70] SYLVIA.https://www.mn-am.com/products/sylvia.

[71] SYLVIA program manual. Molecular Networks GmbH 2016:https://www mn-am.com/files/docs/sylvia/sylvia_manual.pdf.

[72] Szymkuc S, Gajewska EP, Klucznik T, Molga K, Dittwald P, Startek M, et al. Computer-Assisted Synthetic Planning: The End of the Beginning. Angew Chem Int Ed Engl 2016;55(20):5904-37.

[73] Klucznik T, Mikulak-Klucznik B, McCormack MP, Lima H, Szymkuć S, Bhowmick M, et al. Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. Chem 2018;4(3):522-32.

[74] Wang Z, Zhao W, Hao G, Song B. Mapping the resources and approaches facilitating computer-aided synthesis planning. Organic Chemistry Frontiers 2021;8(4):812-24.

[75] Krishnasamy A, Gupta SK, Reitz RD. Prospective fuels for diesel low temperature combustion engine applications: A critical review. International Journal of Engine Research 2020.

[76] Miller JA, Sivaramakrishnan R, Tao Y, Goldsmith CF, Burke MP, Jasper AW, et al. Combustion chemistry in the twenty-first century: Developing theory-informed chemical kinetics models. Progress in Energy and Combustion Science 2021;83.

[77] Masri AR. Challenges for turbulent combustion. Proceedings of the Combustion Institute 2020.

[78] Dallüge J, Beens J, Brinkman UAT. Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool. Journal of Chromatography A 2003;1000(1-2):69-108.

[79] Bertsch W. Two-Dimensional Gas Chromatography. Concepts, Instrumentation, and Applications - Part 1: Fundamentals, Conventional Two-Dimensional Gas Chromatography, Selected Applications. Journal of High Resolution Chromatography 1999;22(12):647-65.

[80] Bertsch W. Two-Dimensional Gas Chromatography. Concepts, Instrumentation, and Applications - Part 2: Comprehensive Two-Dimensional Gas Chromatography. Journal of High Resolution Chromatography 2000;23(3):167-81.

[81] Mueller CJ, Cannella WJ, Bays JT, Bruno TJ, DeFabio K, Dettman HD, et al. Diesel Surrogate Fuels for Engine Testing and Chemical-Kinetic Modeling: Compositions and Properties. Energy Fuels 2016;30(2):1445-61.

[82] CRC Report FACE-1-Chemical and Physical Properties of FACE Research Diesel Fuels. Coordinating Research Council (CRC) 2010:http://crcsite.wpengine.com/wp-content/uploads/2019/05/FACE-1-Chem-and-Phys-Props-of-FACE-Research-Diesel-Fuels-1.pdf.

[83] Report CDEV-2013-2065-RT: Detailed Hydrocarbon Analysis of FACE Diesel Fuels Using Comprehensive Two-Dimensional Gas Chromatography. Coordinating Research Council (CRC) 2013:http://crcsite.wpengine.com/wp-content/uploads/2019/05/GCxGC-analysis-of-FACE-fuels-RG-v4-0-Nov3.pdf.

[84] CRC Report No. AVFL-24: FACE Gasolines and Blends with Ethanol-Detailed Characterization of Physical and Chemical Properties. Coordinating Research Council (CRC) 2014:http://crcsite.wpengine.com/wp-content/uploads/2019/05/AVFL-24-FACE-Gasolines-Report-071414.pdf.

[85] ASTM D6730-19 Standard Test Method for Determination of Individual Components in Spark Ignition Engine Fuels by 100-Metre Capillary (with Precolumn) High-Resolution Gas Chromatography. American Society for Testing and Materials (ASTM) international; 2019.

[86] ASTM D6729-14 Determination of Individual Components in Spark Ignition Engine Fuels by 100 Metre Capillary High Resolution Gas Chromatography. American Society for Testing and Materials (ASTM) international; 2014.

[87] Wu Z, Mao Y, Raza M, Zhu J, Feng Y, Wang S, et al. Surrogate fuels for RP-3 kerosene formulated by emulating molecular structures, functional groups, physical and chemical properties. Combustion and Flame 2019;208:388-401.

[88] Faravelli T, Manenti F, Ranzi E. Mathematical Modelling of Gas-Phase Complex Reaction Systems: Pyrolysis and Combustion. Amsterdam, Netherlands: Elsevier; 2019.

[89] Curran HJ. Developing detailed chemical kinetic mechanisms for fuel combustion. Proceedings of the Combustion Institute 2019;37(1):57-81.

[90] Lu T, Law CK. Toward accommodating realistic fuel chemistry in large-scale computations. Progress in Energy and Combustion Science 2009;35(2):192-215.

[91] Sarathy SM, Westbrook CK, Mehl M, Pitz WJ, Togbe C, Dagaut P, et al. Comprehensive chemical kinetic modeling of the oxidation of 2-methylalkanes from C7 to C20. Combustion and Flame 2011;158(12):2338-57.

[92] Sarathy SM, Yeung C, Westbrook CK, Pitz WJ, Mehl M, Thomson MJ. An experimental and kinetic modeling study of n-octane and 2-methylheptane in an opposed-flow diffusion flame. Combustion and Flame 2011;158(7):1277-87.

[93] Puduppakkam KV, Liang L, Naik CV, Meeks E, Bunting BG. Combustion and Emissions Modeling of a Gasoline HCCI Engine Using Model Fuels. SAE Technical Paper 2009-01-0669 2009.

[94] Naik CV, Puduppakkam K, Wang C, Kottalam J, Liang L, Hodgson D, et al. Applying Detailed Kinetics to Realistic Engine Simulation: the Surrogate Blend Optimizer and Mechanism Reduction Strategies. SAE International Journal of Engines 2010;3(1):241-59.

[95] Puduppakkam KV, Naik CV, Wang C, Meeks E. Validation Studies of a Detailed Kinetics Mechanism for Diesel and Gasoline Surrogate Fuels. SAE Technical Paper 2010-01-0545 2010.

[96] Su X, Ra Y, Reitz RD. A Surrogate Fuel Formulation Approach for Real Transportation Fuels with Application to Multi-Dimensional Engine Simulations. SAE International Journal of Fuels and Lubricants 2014;7(1):236-49.

[97] Mueller CJ, Cannella WJ, Bruno TJ, Bunting B, Dettman HD, Franz JA, et al. Methodology for Formulating Diesel Surrogate Fuels with Accurate Compositional, Ignition-Quality, and Volatility Characteristics. Energy & Fuels 2012;26(6):3284-303.

[98] Krishnasamy A, Reitz RD, Willems W, Kurtz E. Surrogate Diesel Fuel Models for Low Temperature Combustion. 2013;1.

[99] Yu J, Wang Z, Zhuo X, Wang W, Gou X. Surrogate Definition and Chemical Kinetic Modeling for Two Different Jet Aviation Fuels. Energy & Fuels 2016.

[100] Yu J, Ju Y, Gou X. Surrogate fuel formulation for oxygenated and hydrocarbon fuels by using the molecular structures and functional groups. Fuel 2016;166:211-8.

[101] Ilieş BD, Khandavilli M, Li Y, Kukkadapu G, Wagnon SW, Abdul Jameel AG, et al. Probing the Chemical Kinetics of Minimalist Functional Group Gasoline Surrogates. Energy & Fuels 2021;35(4):3315-32.

[102] Abdul Jameel AG, Naser N, Emwas A-H, Sarathy SM. Surrogate formulation for diesel and jet fuels using the minimalist functional group (MFG) approach. Proceedings of the Combustion Institute 2019;37(4):4663-71.

[103] Abdul Jameel AG, Naser N, Issayev G, Touitou J, Ghosh MK, Emwas A-H, et al. A minimalist functional group (MFG) approach for surrogate fuel formulation. Combustion and Flame 2018;192:250-71.

[104] Al-Esawi N, Al Qubeissi M. A new approach to formulation of complex fuel surrogates. Fuel 2021;283.

[105] Li A, Yu L, Lu X, Huang Z, Zhu L. Experimental and Modeling Study on Autoignition of a Biodiesel/n-Heptane Mixture and Related Surrogate in a Heated Rapid Compression Machine. Energy & Fuels 2019;33(5):4552-63.

[106] Li A, Ji W, Huang Z, Zhu L. Predictions of oxidation and autoignition of large methyl ester with small molecule fuels. Fuel 2019;251:162-74.

[107] Li A, Zhu L, Mao Y, Zhai J, Han D, Lu X, et al. Surrogate formulation methodology for biodiesel based on chemical deconstruction in consideration of molecular structure and engine combustion factors. Combustion and Flame 2019;199:152-67.

[108] Li A, Zhang Z, Cheng X, Lu X, Zhu L, Huang Z. Development and validation of surrogates for RP-3 jet fuel based on chemical deconstruction methodology. Fuel 2020;267.

[109] Violi A, Yan S, Eddings EG, Sarofim AF, Granata S, Faravelli T, et al. Experimental formulation and kinetic model for JP-8 surrogate mixtures. Combustion Science and Technology 2002;174(11-12):399-417.

[110] Kim D, Violi A. On the importance of species selection for the formulation of fuel surrogates. Proceedings of the Combustion Institute 2020.

[111] Sarathy SM, Farooq A, Kalghatgi GT. Recent progress in gasoline surrogate fuels. Progress in Energy and Combustion Science 2018;65:67-108.

[112] Pitz WJ, Cernansky NP, Dryer FL, Egolfopoulos FN, Farrell JT, Friend DG, et al. Development of an Experimental Database and Chemical Kinetic Models for Surrogate Gasoline. SAE Technical Paper 2007-01-0175 2007.

[113] Battin-Leclerc F. Detailed chemical kinetic models for the low-temperature combustion of hydrocarbons with application to gasoline and diesel fuel surrogates. Progress in Energy and Combustion Science 2008;34(4):440-98.

[114] Simmie JM. Detailed chemical kinetic models for the combustion of hydrocarbon fuels. Progress in Energy and Combustion Science 2003;29(6):599-634.

[115] Boot MD, Tian M, Hensen EJM, Mani Sarathy S. Impact of fuel molecular structure on auto-ignition behavior – Design rules for future high performance gasolines. Progress in Energy and Combustion Science 2017;60:1-25.

[116] Zhen X, Wang Y, Liu D. An overview of the chemical reaction mechanisms for gasoline surrogate fuels. Applied Thermal Engineering 2017;124:1257-68.

[117] Piehl JA, Zyada A, Bravo L, Samimi-Abianeh O. Review of Oxidation of Gasoline Surrogates and Its Components. Journal of Combustion 2018;2018:1-27.

[118] Dagaut P, Cathonnet M. The ignition, oxidation, and combustion of kerosene: A review of experimental and kinetic modeling. Progress in Energy and Combustion Science 2006;32(1):48-92.

[119] Edwards T, Colket M, Cernansky N, Dryer F, Egolfopoulos F, Friend D, et al. Development of an experimental database and kinetic models for surrogate jet fuels. 45th AIAA Aerospace Sciences Meeting and Exhibit 2012.

[120] Colket M, Heyne J, Rumizen M, Gupta M, Edwards T, Roquemore WM, et al. Overview of the National Jet Fuels Combustion Program. AIAA Journal 2017;55(4):1087-104.

[121] Pitz WJ, Mueller CJ. Recent progress in the development of diesel surrogate fuels. Progress in Energy and Combustion Science 2011;37(3):330-50.

[122] Farrell JT, Cernansky NP, Law FLDaCK, Friend DG, Hergart CA, McDavid RM, et al. Development of an Experimental Database and Kinetic Models for Surrogate Diesel Fuels. SAE Technical Paper 2007-01-0201 2007.

[123] Szymkowicz PG, Benajes J. Development of a Diesel Surrogate Fuel Library. Fuel 2018;222:21-34.

[124] Dooley S, Won SH, Chaos M, Heyne J, Ju Y, Dryer FL, et al. A jet fuel surrogate formulated by real fuel properties. Combustion and Flame 2010;157(12):2333-9.

[125] Dooley S, Won SH, Heyne J, Farouk TI, Ju Y, Dryer FL, et al. The experimental evaluation of a methodology for surrogate fuel formulation to emulate gas phase combustion kinetic phenomena. Combustion and Flame 2012;159(4):1444-66.

[126] Mehl M, Chen JY, Pitz WJ, Sarathy SM, Westbrook CK. An Approach for Formulating Surrogates for Gasoline with Application toward a Reduced Surrogate Mechanism for CFD Engine Modeling. Energy & Fuels 2011;25(11):5215-23.

[127] Zhang C, Wang R, Hui X, Xue X, Lin Y, Sung C-J. Nonlinear Threshold Sooting Index Prediction Method for Surrogate Formulation Emulating Sooting Characteristics: A Case Study Using RP-3 Jet Fuels. Energy & Fuels 2020;34(8):9990-9.

[128] Del Pecchia M, Fontanesi S. A methodology to formulate multicomponent fuel surrogates to model flame propagation and ignition delay. Fuel 2020;279.

[129] Del Pecchia M, Pessina V, Berni F, d'Adamo A, Fontanesi S. Gasoline-ethanol blend formulation to mimic laminar flame speed and auto-ignition quality in automotive engines. Fuel 2020;264.

[130] Al Rashidi MJ, Thion S, Togbe C, Dayma G, Mehl M, Dagaut P, et al. Elucidating reactivity regimes in cyclopentane oxidation: Jet stirred reactor experiments, computational chemistry, and kinetic modeling. Proceedings of the Combustion Institute 2017;36(1):469-77.

[131] Silke EJ, Pitz WJ, Westbrook CK, Ribaucour M. Detailed chemical kinetic modeling of cyclohexane oxidation. J Phys Chem A 2007;111(19):3761-75.

[132] H. Wang, E. Dames, B. Sirjean, D. A. Sheen, R. Tango, A. Violi, et al. A high-temperature chemical kinetic model of n-alkane (up to n-dodecane), cyclohexane, and methyl-, ethyl-, n-propyl and n-butyl-cyclohexane oxidation at high temperatures, JetSurF version 2.0. 2010:http://web.stanford.edu/group/haiwanglab/JetSurF/JetSurF2.0/index html.

[133] Weber BW, Pitz WJ, Mehl M, Silke EJ, Davis AC, Sung C-J. Experiments and modeling of the autoignition of methylcyclohexane at high pressure. Combustion and Flame 2014;161(8):1972-83.

[134] Pitz WJ, Naik CV, Mhaoldúin TN, Westbrook CK, Curran HJ, Orme JP, et al. Modeling and experimental investigation of methylcyclohexane ignition in a rapid compression machine. Proceedings of the Combustion Institute 2007;31(1):267-75.

[135] Eldeeb MA, Jouzdani S, Wang Z, Sarathy SM, Akih-Kumgeh B. Experimental and Chemical Kinetic Modeling Study of Dimethylcyclohexane Oxidation and Pyrolysis. Energy & Fuels 2016;30(10):8648-57.

[136] Dagaut P, Ristori A, Frassoldati A, Faravelli T, Dayma G, Ranzi E. Experimental and semi-detailed kinetic modeling study of decalin oxidation and pyrolysis over a wide range of conditions. Proceedings of the Combustion Institute 2013;34:289-96.

[137] Westbrook CK, Pitz WJ, Mehl M, Glaude PA, Herbinet O, Bax S, et al. Experimental and Kinetic Modeling Study of 2-Methyl-2-Butene: Allylic Hydrocarbon Kinetics. J Phys Chem A 2015;119(28):7462-80.

[138] Metcalfe WK, Pitz WJ, Curran HJ, Simmie JM, Westbrook CK. The development of a detailed chemical kinetic mechanism for diisobutylene and comparison to shock tube ignition times. Proceedings of the Combustion Institute 2007;31:377-84.

[139] Andrae JCG. Development of a detailed kinetic model for gasoline surrogate fuels. Fuel 2008;87(10-11):2013-22.

[140] Khaled F, Badra J, Farooq A. A shock tube study of C4–C6 straight chain alkenes + OH reactions. Proceedings of the Combustion Institute 2017;36(1):289-98.

[141] Westbrook CK, Naik CV, Herbinet O, Pitz WJ, Mehl M, Sarathy SM, et al. Detailed chemical kinetic reaction mechanisms for soy and rapeseed biodiesel fuels. Combustion and Flame 2011;158(4):742-55.

[142] Sarathy SM, Vranckx S, Yasunaga K, Mehl M, Oßwald P, Metcalfe WK, et al. A comprehensive chemical kinetic combustion model for the four butanol isomers. Combustion and Flame 2012;159(6):2028-55.

[143] Kaiser EW, Wailington TJ, Hurley MD, Platz J, Curran HJ, Pitz WJ, et al. Experimental and modeling study of premixed atmospheric-pressure dimethyl ether-air flames. Journal of Physical Chemistry A 2000;104(35):8194-206.

[144] Fischer SL, Dryer FL, Curran HJ. The reaction kinetics of dimethyl ether. I: High-temperature pyrolysis and oxidation in flow reactors. Int J Chem Kinet 2000;32(12):713-40.

[145] Curran HJ, Fischer SL, Dryer FL. The reaction kinetics of dimethyl ether. II: Low-temperature oxidation in flow reactors. Int J Chem Kinet 2000;32(12):741-59.

[146] Nakamura H, Darcy D, Mehl M, Tobin CJ, Metcalfe WK, Pitz WJ, et al. An experimental and modeling study of shock tube and rapid compression machine ignition of n-butylbenzene/air mixtures. Combustion and Flame 2014;161(1):49-64.

[147] Pei Y, Mehl M, Liu W, Lu T, Pitz WJ, Som S. A Multicomponent Blend as a Diesel Fuel Surrogate for Compression Ignition Engine Applications. Journal of Engineering for Gas Turbines and Power 2015;137(11).

[148] Malewicki T, Gudiyella S, Brezinsky K. Experimental and modeling study on the oxidation of Jet A and the n-dodecane/iso-octane/n-propylbenzene/1,3,5-trimethylbenzene surrogate fuel. Combustion and Flame 2013;160(1):17-30.

[149] Marinov NM, Pitz WJ, Westbrook CK, Vincitore AM, Castaldi MJ, Senkan SM, et al. Aromatic and

polycyclic aromatic hydrocarbon formation in a laminar premixed n-butane flame. Combustion and Flame 1998;114(1-2):192-213.

[150] Raj A, Prada IDC, Amer AA, Chung SH. A reaction mechanism for gasoline surrogate fuels for large polycyclic aromatic hydrocarbons. Combustion and Flame 2012;159(2):500-15.

[151] Li YY, Zhang LD, Wang ZD, Ye LL, Cai JH, Cheng ZJ, et al. Experimental and kinetic modeling study of tetralin pyrolysis at low pressure. Proceedings of the Combustion Institute 2013;34:1739-48.

[152] Narayanaswamy K, Blanquart G, Pitsch H. A consistent chemical mechanism for oxidation of substituted aromatic species. Combustion and Flame 2010;157(10):1879-98.

[153] ASTM D5291 - 16 Standard Test Methods for Instrumental Determination of Carbon, Hydrogen, and Nitrogen in Petroleum Products and Lubricants. American Society for Testing and Materials (ASTM) international 2016.

[154] Japanwala S, Chung KH, Dettman HD, Gray MR. Quality of distillates from repeated recycle of residue. Energy & Fuels 2002;16(2):477-84.

[155] Won SH, Dooley S, Veloo PS, Wang H, Oehlschlaeger MA, Dryer FL, et al. The combustion properties of 2,6,10-trimethyl dodecane and a chemical functional group analysis. Combustion and Flame 2014;161(3):826-34.

[156] Abdul Jameel AG, Naser N, Emwas A-H, Dooley S, Sarathy SM. Predicting Fuel Ignition Quality Using 1H NMR Spectroscopy and Multiple Linear Regression. Energy & Fuels 2016;30(11):9819-35.

[157] Dahmen M, Marquardt W. A Novel Group Contribution Method for the Prediction of the Derived Cetane Number of Oxygenated Hydrocarbons. Energy & Fuels 2015;29(9):5781-801.

[158] Define and solve a problem by using Solver. Microsoft:https://support microsoft.com/en-us/office/define-and-solve-a-problem-by-using-solver-5d1a388f-079d-43ac-a7eb-f63e45925040.

[159] EXCEL SOLVER ONLINE HELP. Frontline Systems, Inc:https://www.solver.com/excel-solver-online-help.

[160] ASTM D6550-15 Standard Test Method for Determination of Olefin Content of Gasolines by Supercritical-Fluid Chromatography. American Society for Testing and Materials (ASTM) international; 2015.

[161] ASTM D5580-15 Standard Test Method for Determination of Benzene, Toluene, Ethylbenzene, p/mXylene, o-Xylene, C9 and Heavier Aromatics, and Total Aromatics in Finished Gasoline by Gas Chromatography1. American Society for Testing and Materials (ASTM) international; 2015.

[162] ASTM D5599-18 Standard Test Method for Determination of Oxygenates in Gasoline by Gas Chromatography and Oxygen Selective Flame Ionization Detection. American Society for Testing and Materials (ASTM) international; 2018.

[163] ASTM D4815-15b Standard Test Method for Determination of MTBE, ETBE, TAME, DIPE, tertiary-Amyl Alcohol and C1 to C4 Alcohols in Gasoline by Gas Chromatography1. American Society for Testing and Materials (ASTM) international; 2015.

[164] ASTM E2997-16 Standard Test Method for Analysis of Biodiesel Products by Gas ChromatographyMass Spectrometry. 2016.

[165] Herbinet O, Pitz WJ, Westbrook CK. Detailed chemical kinetic mechanism for the oxidation of biodiesel fuels blend surrogate. Combustion and Flame 2010;157(5):893-908.

[166] The CRECK Modeling Group. http://creckmodelingchempolimiit/menu-kinetics/menu-kinetics-detailed-mechanisms.

[167] Sarathy SM, Kukkadapu G, Mehl M, Javed T, Ahmed A, Naser N, et al. Compositional effects on the ignition of FACE gasolines. Combustion and Flame 2016;169:171-93.

[168] Widegren JA, Bruno TJ. Thermal Decomposition Kinetics of the Aviation Turbine Fuel Jet A. Ind Eng Chem Res 2008;47(13):4342-8.

[169] Huber ML, Lemmon EW, Bruno TJ. Surrogate Mixture Models for the Thermophysical Properties of

Aviation Fuel Jet-A. Energy & Fuels 2010;24(6):3565-71.

[170] Lemmon EW, Huber ML, McLinden MO. NIST Reference Fluid Thermodynamic and Transport Properties—REFPROP Version 9.1. National Institute of Standards and Technology (NIST) 2013:https://www.nist.gov/system/files/documents/srd/REFPROP9.PDF.

[171] Dagaut P, Gaïl S, Sahasrabudhe M. Rapeseed oil methyl ester oxidation over extended ranges of pressure, temperature, and equivalence ratio: Experimental and modeling kinetic study. Proceedings of the Combustion Institute 2007;31(2):2955-61.

[172] Merkisz J, Fuć P, Lijewski P, Kozak M. Rapeseed Oil Methyl Esters (RME) as Fuel for Urban Transport. Alternative Fuels, Technical and Environmental Conditions. 2016.

[173] NIST Mass Spectrometry Data Center. Biomolecular Measurement Division (BMD):https://chemdata.nist.gov/.

[174] Qian Y, Yu L, Li Z, Zhang Y, Xu L, Zhou Q, et al. A new methodology for diesel surrogate fuel formulation: Bridging fuel fundamental properties and real engine combustion characteristics. Energy 2018;148:424-47.

[175] Yu L, Mao Y, Li A, Wang S, Qiu Y, Qian Y, et al. Experimental and modeling validation of a large diesel surrogate: Autoignition in heated rapid compression machine and oxidation in flow reactor. Combustion and Flame 2019;202:195-207.

[176] Yu L, Mao Y, Qiu Y, Wang S, Li H, Tao W, et al. Experimental and modeling study of the autoignition characteristics of commercial diesel under engine-relevant conditions. Proceedings of the Combustion Institute 2019;37(4):4805-12.

[177] Mati K, Ristori A, Gaïl S, Pengloan G, Dagaut P. The oxidation of a diesel fuel at 1–10atm: Experimental study in a JSR and detailed chemical kinetic modeling. Proceedings of the Combustion Institute 2007;31(2):2939-46.

[178] Taylor J, Sluder S, Wright K. Fuels for Advanced Combustion (FACE). Diesel Engine Efficiency and Emissions Research (DEER) Conference 2007:https://www.energy.gov/sites/prod/files/2014/03/f9/deer07_taylor.pdf.

[179] Fuels for Advanced Combustion Engines (FACE) Working Group Mission Statement. 2016:http://crcsite.wpengine.com/wp-content/uploads/2019/05/FACE-Revised-Mission-Statement_Jan6.pdf.

[180] Sarathy SM, Kukkadapu G, Mehl M, Wang W, Javed T, Park S, et al. Ignition of alkane-rich FACE gasoline fuels and their surrogate mixtures. Proceedings of the Combustion Institute 2015;35(1):249-57.

[181] Goldsborough SS, Hochgreb S, Vanhove G, Wooldridge MS, Curran HJ, Sung C-J. Advances in rapid compression machine studies of low- and intermediate-temperature autoignition phenomena. Progress in Energy and Combustion Science 2017;63:1-78.

[182] Yu R, Liu J, Ma B. The dependence of NTC behavior on the equivalence ratio and nitrogen fraction in cool flame region. Fuel 2020;271.

[183] Kaczmarek D, Shaqiri S, Atakan B, Kasper T. The influence of pressure and equivalence ratio on the NTC behavior of methane. Proceedings of the Combustion Institute 2020.

[184] Chen B, Togbé C, Wang Z, Dagaut P, Sarathy SM. Jet-stirred reactor oxidation of alkane-rich FACE gasoline fuels. Proceedings of the Combustion Institute 2017;36(1):517-24.

[185] ASTM D613-18a Standard Test Method for Cetane Number of Diesel Fuel Oil. American Society for Testing and Materials (ASTM) international; 2018.

[186] ASTM D6890-18 Standard Test Method for Determination of Ignition Delay and Derived Cetane Number (DCN) of Diesel Fuel Oils by Combustion in a Constant Volume Chamber. American Society for Testing and Materials (ASTM) international; 2018.

[187] ASTM D7668-17 Standard Test Method for Determination of Derived Cetane Number (DCN) of Diesel Fuel Oils—Ignition Delay and Combustion Delay Using a Constant Volume Combustion Chamber Method.

American Society for Testing and Materials (ASTM) international; 2017.

[188] ASTM D8183-18 Standard Test Method for Determination of Indicated Cetane Number (ICN) of Diesel Fuel Oils using a Constant Volume Combustion Chamber—Reference Fuels Calibration Method. American Society for Testing and Materials (ASTM) international; 2018.

[189] ASTM D2699-19 Standard Test Method for Research Octane Number of Spark-Ignition Engine Fuel. American Society for Testing and Materials (ASTM) international; 2019.

[190] ASTM D2700-19 Standard Test Method for Motor Octane Number of Spark-Ignition Engine Fuel. American Society for Testing and Materials (ASTM) international; 2019.

[191] Dahmen M, Marquardt W. Model-Based Design of Tailor-Made Biofuels. Energy & Fuels 2016;30(2):1109-34.

[192] Dahmen M, Marquardt W. Model-Based Formulation of Biofuel Blends by Simultaneous Product and Pathway Design. Energy & Fuels 2017;31(4):4096-121.

[193] Yanowitz J, Ratcliff MA, McCormick RL, Taylor JD, Murphy MJ. Compendium of Experimental Cetane Numbers. 2017.

[194] 2,4,6,8-Tetraoxanonane. AmBeed 2021:https://www.ambeed.com/products/13353-03-2.html.

[195] Benson SW. Thermochemical kinetics: methods for the estimation of thermochemical data and rate parameters. 2nd ed. New York: John Wiley & Sons; 1976.

[196] Benson SW, Buss JH. Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties. The Journal of Chemical Physics 1958;29(3):546-72.

[197] Benson SW, Cruickshank FR, Golden DM, Haugen GR, O'Neal HE, Rodgers AS, et al. Additivity rules for the estimation of thermochemical properties. Chem Rev 1969;69(3):279-324.

[198] Joback KG, Reid RC. Estimation of Pure-Component Properties from Group-Contributions. Chemical Engineering Communications 1987;57(1-6):233-43.

[199] Dearden JC, Rotureau P, Fayet G. QSPR prediction of physico-chemical properties for REACH. SAR QSAR Environ Res 2013;24(4):279-318.

[200] Le T, Epa VC, Burden FR, Winkler DA. Quantitative structure-property relationship modeling of diverse materials properties. Chem Rev 2012;112(5):2889-919.

[201] Katritzky AR, Kuanar M, Slavov S, Hall CD, Karelson M, Kahn I, et al. Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. Chem Rev 2010;110(10):5714-89.

[202] Bowden JN, Johnston AA, Russell JA. Octane-Cetane Relationship. SOUTHWEST RESEARCH INSTITUTE 1974.

[203] Kalghatgi GT. Auto-Ignition Quality of Practical Fuels and Implications for Fuel Requirements of Future SI and HCCI Engines. SAE Technical Paper 2005-01-0239 2005.

[204] Janecek D, Rothamer D, Ghandhi J. Investigation of cetane number and octane number correlation under homogenous-charge compression-ignition engine operation. Proceedings of the Combustion Institute 2017;36(3):3651-7.

[205] III TWR, Matheaus AC. Fuel Requirements for HCCI Engine Operation. SAE Technical Paper 2003-01-1813 2003.

[206] Inagaki K, Fuyuto T, Nishikawa K, Nakakita K, Sakata I. Dual-Fuel PCI Combustion Controlled by In-Cylinder Stratification of Ignitability. SAE Technical Paper 2006-01-0028 2006.

[207] Becker K. The Influence of an Ignition Accelerator on the Ignition Quality and Anti-Knock Properties of Light Hydrocarbons in the Diesel Engine. SAE Technical Paper 760163 1976.

[208] Kubic WL. A Group Contribution Method for Estimating Cetane and Octane Numbers. Los Alamos National Laboratory Report No LA-UR-16-25529 2016:https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-16-25529.

[209] Kubic WL, Jenkins RW, Moore CM, Semelsberger TA, Sutton AD. Artificial Neural Network Based Group Contribution Method for Estimating Cetane and Octane Numbers of Hydrocarbons and Oxygenated Organic Compounds. Ind Eng Chem Res 2017;56(42):12236-45.

[210] DeFries TH, Kastrup RV, Indritz D. Prediction of cetane number by group additivity and carbon-13 Nuclear Magnetic Resonance. Ind Eng Chem Res 1987;26(2):188-93.

[211] Yang H, Fairbridge C, Ring Z. Neural Network Prediction of Cetane Numbers for Isoparaffins and Diesel Fuel. Petroleum Science and Technology 2001;19(5-6):573-86.

[212] Saldana DA, Starck L, Mougin P, Rousseau B, Pidol L, Jeuland N, et al. Flash Point and Cetane Number Predictions for Fuel Compounds Using Quantitative Structure Property Relationship (QSPR) Methods. Energy & Fuels 2011;25(9):3900-8.

[213] Creton B, Dartiguelongue C, de Bruin T, Toulhoat H. Prediction of the Cetane Number of Diesel Compounds Using the Quantitative Structure Property Relationship. Energy & Fuels 2010;24(10):5396-403.

[214] Guan C, Zhai J, Han D. Cetane number prediction for hydrocarbons from molecular structural descriptors based on active subspace methodology. Fuel 2019;249:1-7.

[215] Baghban A, Adelizadeh M. On the determination of cetane number of hydrocarbons and oxygenates using Adaptive Neuro Fuzzy Inference System optimized with evolutionary algorithms. Fuel 2018;230:344-54.

[216] Kessler T, Dorian G, Mack JH. Application of a Rectified Linear Unit (Relu) Based Artificial Neural Network to Cetane Number Predictions. Proceedings of the Asme Internal Combustion Engine Fall Technical Conference, 2017, Vol 1 2017.

[217] Guo Z, Lim KH, Chen M, Thio BJR, Loo BLW. Predicting cetane numbers of hydrocarbons and oxygenates from highly accessible descriptors by using artificial neural networks. Fuel 2017;207:344-51.

[218] Kessler T, Sacia ER, Bell AT, Mack JH. Predicting the Cetane Number of Furanic Biofuel Candidates Using an Improved Artificial Neural Network Based on Molecular Structure. Proceedings of the Asme Internal Combustion Engine Fall Technical Conference, 2016 2016.

[219] Smolenskii EA, Bavykin VM, Ryzhov AN, Slovokhotova OL, Chuvaeva IV, Lapidus AL. Cetane numbers of hydrocarbons: calculations using optimal topological indices. Russ Chem B+ 2008;57(3):461-7.

[220] Yang H, Ring Z, Briker Y, McLean N, Friesen W, Fairbridge C. Neural network prediction of cetane number and density of diesel fuel from its chemical composition determined by LC and GC-MS. Fuel 2002;81(1):65-74.

[221] Liu Z, Zhang L, Elkamel A, Liang D, Zhao S, Xu C, et al. Multiobjective Feature Selection Approach to Quantitative Structure Property Relationship Models for Predicting the Octane Number of Compounds Found in Gasoline. Energy & Fuels 2017;31(6):5828-39.

[222] Al-Fahemi JH, Albis NA, Gad EAM. QSPR Models for Octane Number Prediction. Journal of Theoretical Chemistry 2014;2014:1-6.

[223] Hosoya H. Chemical meaning of octane number analyzed by topological indices. Croat Chem Acta 2002;75(2):433-45.

[224] Smolenskii EA, Ryzhov AN, Bavykin VM, Myshenkova TN, Lapidus AL. Octane numbers (ONs) of hydrocarbons: a QSPR study using optimal topological indices for the topological equivalents of the ONs. Russ Chem B+ 2007;56(9):1681-93.

[225] Westbrook CK, Sjöberg M, Cernansky NP. A new chemical kinetic method of determining RON and MON values for single component and multicomponent mixtures of engine fuels. Combustion and Flame 2018;195:50-62.

[226] Singh E, Badra J, Mehl M, Sarathy SM. Chemical Kinetic Insights into the Octane Number and Octane Sensitivity of Gasoline Surrogate Mixtures. Energy & Fuels 2017;31(2):1945-60.

[227] Badra JA, Bokhumseen N, Mulla N, Sarathy SM, Farooq A, Kalghatgi G, et al. A methodology to relate

octane numbers of binary and ternary n-heptane, iso-octane and toluene mixtures with simulated ignition delay times. Fuel 2015;160:458-69.

[228] Abdul Jameel AG, Van Oudenhoven V, Emwas A-H, Sarathy SM. Predicting Octane Number Using Nuclear Magnetic Resonance Spectroscopy and Artificial Neural Networks. Energy & Fuels 2018;32(5):6309-29.

[229] Daly SR, Niemeyer KE, Cannella WJ, Hagen CL. Predicting fuel research octane number using Fourier-transform infrared absorption spectra of neat hydrocarbons. Fuel 2016;183:359-65.

[230] Co-Optimization of Fuels & Engines: Fuel Properties Database. National Renewable Energy Laboratory:https://www.nrel.gov/transportation/fuels-properties-database/.

[231] API Data Book. http://wwwepconcom/api-data-bookhtml.

[232] Knocking Characteristics of Pure Hydrocarbons. American Petroleum Institute Research Project 45 1958.

[233] Lapidus AL, Bavykin VM, Smolenskii EA, Chuvaeva IV. Cetane numbers of hydrocarbons as a function of their molecular structure. Doklady Chemistry 2008;420(2):150-5.

[234] Ogawa H, Nishimoto H, Morita A, Shibata G. Predicted diesel ignitability index based on the molecular structures of hydrocarbons. International Journal of Engine Research 2016;17(7):766-75.

[235] Won SH, Haas FM, Dooley S, Dryer FL. Chemical functional group descriptor for jet fuel surrogate. 10th US National Combustion Meeting 2017.

[236] Carpenter DO. Impact of Cycloalkanes on Ignition Propensity Measured as Derived Cetane Number in Multi-Component Surrogate Mixtures. Master's thesis of University of South Carolina 2019.

[237] Wang Y, Cao Y, Wei W, Davidson DF, Hanson RK. A new method of estimating derived cetane number for hydrocarbon fuels. Fuel 2019;241:319-26.

[238] Kalghatgi G, Babiker H, Badra J. A Simple Method to Predict Knock Using Toluene, N-Heptane and Iso-Octane Blends (TPRF) as Gasoline Surrogates. SAE International Journal of Engines 2015;8(2):505-19.

[239] Morgan N, Smallbone A, Bhave A, Kraft M, Cracknell R, Kalghatgi G. Mapping surrogate gasoline compositions into RON/MON space. Combustion and Flame 2010;157(6):1122-31.

[240] Naser N, Yang SY, Kalghatgi G, Chung SH. Relating the octane numbers of fuels to ignition delay times measured in an ignition quality tester (IQT). Fuel 2017;187:117-27.

[241] Foong TM, Morganti KJ, Brear MJ, da Silva G, Yang Y, Dryer FL. The octane numbers of ethanol blended with gasoline and its surrogates. Fuel 2014;115:727-39.

[242] Choose Regression Model Options. MATLAB Help Center 2019:https://uk mathworks.com/help/stats/choose-regression-model-options.html#bvmnwhd-1.

[243] Abou-Rachid H, Bonneviot L, Xu G, Kaliaguine S. On the correlation between kinetic rate constants in the auto-ignition process of some oxygenates and their cetane number: a quantum chemical study. Journal of Molecular Structure: THEOCHEM 2003;621(3):293-304.

[244] Albahri TA. Structural Group Contribution Method for Predicting the Octane Number of Pure Hydrocarbon Liquids. Ind Eng Chem Res 2003;42(3):657-62.

[245] Worldwide Fuel Charter, 6th Edition. Organisation Internationale des Constructeurs d'Automobiles (OICA); 2019.

[246] Worldwide Fuel Charter, 5th Edition. Organisation Internationale des Constructeurs d'Automobiles (OICA); 2013.

[247] Gschwend D, Soltic P, Wokaun A, Vogel F. Review and Performance Evaluation of Fifty Alternative Liquid Fuels for Spark-Ignition Engines. Energy & Fuels 2019;33(3):2186-96.

[248] Motor Gasolines Technical Review. Chevron Corporation; 2009.

[249] Diesel Fuels Technical Review. Chevron Corporation; 2007.

[250] Aviation Fuels Technical Review. Chevron Corporation; 2006.

[251] Richards P. Automotive Fuels Reference Book, 3rd Edition. Warrendale, Pennsylvania: SAE International; 2014.

[252] Herbinet O, Pitz WJ, Westbrook CK. Detailed chemical kinetic oxidation mechanism for a biodiesel surrogate. Combustion and Flame 2008;154(3):507-28.

[253] Kalghatgi G. Fuel/Engine Interactions. SAE International; 2013.

[254] Agarwal AK, Singh AP, Maurya RK. Evolution, challenges and path forward for low temperature combustion engines. Progress in Energy and Combustion Science 2017;61:1-56.

[255] Saxena S, Bedoya ID. Fundamental phenomena affecting low temperature combustion and HCCI engines, high load limits and strategies for extending these limits. Progress in Energy and Combustion Science 2013;39(5):457-88.

[256] Yao M, Zheng Z, Liu H. Progress and recent trends in homogeneous charge compression ignition (HCCI) engines. Progress in Energy and Combustion Science 2009;35(5):398-437.

[257] Musculus MPB, Miles PC, Pickett LM. Conceptual models for partially premixed low-temperature diesel combustion. Progress in Energy and Combustion Science 2013;39(2-3):246-83.

[258] Reitz RD, Duraisamy G. Review of high efficiency and clean reactivity controlled compression ignition (RCCI) combustion in internal combustion engines. Progress in Energy and Combustion Science 2015;46:12-71.

[259] Lu X, Han D, Huang Z. Fuel design and management for the control of advanced compression-ignition combustion modes. Progress in Energy and Combustion Science 2011;37(6):741-83.

[260] Kohse-Hoinghaus K. Combustion in the future: The importance of chemistry. Proc Combust Inst 2020.

[261] Huo X, Huq NA, Stunkel J, Cleveland NS, Starace AK, Settle AE, et al. Tailoring diesel bioblendstock from integrated catalytic upgrading of carboxylic acids: a "fuel property first" approach. Green Chemistry 2019;21(21):5813-27.

[262] Coniglio L, Bennadji H, Glaude PA, Herbinet O, Billaud F. Combustion chemical kinetics of biodiesel and related compounds (methyl and ethyl esters): Experiments and modeling – Advances and future refinements. Progress in Energy and Combustion Science 2013;39(4):340-82.

[263] Lai JYW, Lin KC, Violi A. Biodiesel combustion: Advances in chemical kinetic modeling. Progress in Energy and Combustion Science 2011;37(1):1-14.

[264] Leitner W, Klankermayer J, Pischinger S, Pitsch H, Kohse-Hoinghaus K. Advanced Biofuels and Beyond: Chemistry Solutions for Propulsion and Production. Angew Chem Int Ed Engl 2017;56(20):5412-52.

[265] Hashim H, Narayanasamy M, Yunus NA, Shiun LJ, Muis ZA, Ho WS. A cleaner and greener fuel: Biofuel blend formulation and emission assessment. Journal of Cleaner Production 2017;146:208-17.

[266] Farrell JT, Holladay J, Wagner R. Fuel Blendstocks with Potential to Optimize Future Gasoline Engine Performance. National Renewable Energy Laboratory 2018:https://www.nrel.gov/docs/fy18osti/69009.pdf.

[267] Farrell J, Holladay J, Wagner R. Co-Optimization of Fuels & Engines FY16 Year in Review. 2016:https://www.nrel.gov/docs/fy17osti/67595.pdf.

[268] Li R, Liu Z, Han Y, Cai Y, Wang X, Zheng J, et al. Target-Oriented Fuel Design for the Homogeneous Charge Autoignition Combustion Mode: A Case Study of a n-Heptane–PODE3–Ethanol Mixture. 1. A Pathway To Increase the Combustion Efficiency and Reduce Pollutant Emissions. Energy & Fuels 2018;33(1):16-30.

[269] Li R, Liu Z, Han Y, Cai Y, Wang X, Zheng J, et al. Target-Oriented Fuel Design for the Homogeneous Charge Autoignition Combustion Mode: A Case Study of a n-Heptane–PODE3–Ethanol Mixture. 2. Identification of a Functional Configuration of Fuel Components. Energy & Fuels 2018;33(1):31-49.

[270] Katritzky AR, Maran U, Karelson M, Lobanov VS. Prediction of Melting Points for the Substituted Benzenes: A QSPR Approach. Journal of Chemical Information and Computer Sciences 1997;37(5):913-9.

[271] Coley CW, Barzilay R, Green WH, Jaakkola TS, Jensen KF. Convolutional Embedding of Attributed Molecular Graphs for Physical Property Prediction. J Chem Inf Model 2017;57(8):1757-72.

[272] Bhat AU, Merchant SS, Bhagwat SS. Prediction of Melting Points of Organic Compounds Using Extreme Learning Machines. Ind Eng Chem Res 2008;47(3):920-5.

[273] Godavarthy SS, Robinson RL, Gasem KAM. An Improved Structure−Property Model for Predicting Melting-Point Temperatures. Ind Eng Chem Res 2006;45(14):5117-26.

[274] Saldana DA, Starck L, Mougin P, Rousseau B, Creton B. On the rational formulation of alternative fuels: melting point and net heat of combustion predictions for fuel compounds using machine learning methods. SAR QSAR Environ Res 2013;24(4):259-77.

[275] Wang Q, Ma P, Neng S. Position Group Contribution Method for Estimation of Melting Point of Organic Compounds. Chinese Journal of Chemical Engineering 2009;17(3):468-72.

[276] Boiling Point and Melting Point Prediction for Aliphatic, Non-Hydrogen-Bonding Compounds.

[277] Zhao L, Yalkowsky SH. A Combined Group Contribution and Molecular Geometry Approach for Predicting Melting Points of Aliphatic Compounds. Ind Eng Chem Res 1999;38(9):3581-4.

[278] Tsuchiya Y, Hasegawa H, Iwatsubo T. Prediction of the melting point of n-alkanes using the molecular dynamics method. The Journal of Chemical Physics 2001;114(5):2484-8.

[279] Zhang J-h, Liu Z-m, Liu W-r. QSPR study for prediction of boiling points of 2475 organic compounds using stochastic gradient boosting. Journal of Chemometrics 2014;28(3):161-7.

[280] Roubehie Fissa M, Lahiouel Y, Khaouane L, Hanini S. QSPR estimation models of normal boiling point and relative liquid density of pure hydrocarbons using MLR and MLP-ANN methods. J Mol Graph Model 2019;87:109-20.

[281] Espinosa G, Yaffe D, Cohen Y, Arenas A, Giralt F. Neural network based quantitative structural property relations (QSPRs) for predicting boiling points of aliphatic hydrocarbons. J Chem Inf Comput Sci 2000;40(3):859-79.

[282] Dai YM, Zhu ZP, Cao Z, Zhang YF, Zeng JL, Li X. Prediction of boiling points of organic compounds by QSPR tools. J Mol Graph Model 2013;44:113-9.

[283] Katritzky AR, Lobanov VS, Karelson M. Normal Boiling Points for Organic Compounds: Correlation and Prediction by a Quantitative Structure−Property Relationship. Journal of Chemical Information and Computer Sciences 1998;38(1):28-41.

[284] Katritzky AR, Mu L, Lobanov VS, Karelson M. Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics. The Journal of Physical Chemistry 1996;100(24):10400-7.

[285] Jin L, Bai P. QSPR study on normal boiling point of acyclic oxygen containing organic compounds by radial basis function artificial neural network. Chemometrics and Intelligent Laboratory Systems 2016;157:127-32.

[286] Jin L, Bai P. Prediction of the normal boiling point of oxygen containing organic compounds using quantitative structure–property relationship strategy. Fluid Phase Equilibria 2016;427:194-201.

[287] OSAGHI B, SAFA F. QSPR study on the boiling points of aliphatic esters using the atom-type-based AI topological indices. Revue Roumaine de Chimie 2019;64(2):183-9.

[288] Ericksen D, Wilding WV, Oscarson JL, Rowley RL. Use of the DIPPR Database for Development of QSPR Correlations: Normal Boiling Point. Journal of Chemical & Engineering Data 2002;47(5):1293-302.

[289] Yan X, Lan T, Jia Q, Yan F, Wang Q. A norm indexes-based QSPR model for predicting the standard vaporization enthalpy and formation enthalpy of organic compounds. Fluid Phase Equilibria 2020;507.

[290] Jia Q, Yan X, Lan T, Yan F, Wang Q. Norm indexes for predicting enthalpy of vaporization of organic compounds at the boiling point. Journal of Molecular Liquids 2019;282:484-8.

[291] Gharagheizi F. Determination of normal boiling vaporization enthalpy using a new molecular-based model. Fluid Phase Equilibria 2012;317:43-51.

[292] Homer J, Generalis SC, Robson JH. Artificial neural networks for the prediction of liquid viscosity, density,

heat of vaporization, boiling point and Pitzer's acentric factor Part I. Hydrocarbons. Physical Chemistry Chemical Physics 1999;1(17):4075-81.

[293]   Gharagheizi F, Babaie O, Mazdeyasna S. Prediction of Vaporization Enthalpy of Pure Compounds using a Group Contribution-Based Method. Ind Eng Chem Res 2011;50(10):6503-7.

[294]   Kolská Z, Růžička V, Gani R. Estimation of the Enthalpy of Vaporization and the Entropy of Vaporization for Pure Organic Compounds at 298.15 K and at Normal Boiling Temperature by a Group Contribution Method. Ind Eng Chem Res 2005;44(22):8436-54.

[295]   Tu C-H, Liu C-P. Group-contribution estimation of the enthalpy of vaporization of organic compounds. Fluid Phase Equilibria 1996;121(1-2):45-65.

[296]   Kondor A, Járvás G, Kontos J, Dallos A. Temperature dependent surface tension estimation using COSMO-RS sigma moments. Chemical Engineering Research and Design 2014;92(12):2867-72.

[297]   Delgado EJ, Diaz GA. A molecular structure based model for predicting surface tension of organic compounds. SAR QSAR Environ Res 2006;17(5):483-96.

[298]   Knotts TA, Wilding WV, Oscarson JL, Rowley RL. Use of the DIPPR Database for Development of QSPR Correlations: Surface Tension†. Journal of Chemical & Engineering Data 2001;46(5):1007-12.

[299]   Dai Y-m, Liu Y-n, Li X, Cao Z, Zhu Z-p, Yang D-w. Estimation of surface tension of organic compounds using quantitative structure-property relationship. Journal of Central South University 2012;19(1):93-100.

[300]   Wang J, Du H, Liu H, Yao X, Hu Z, Fan B. Prediction of surface tension for common compounds based on novel methods using heuristic method and support vector machine. Talanta 2007;73(1):147-56.

[301]   Pazuki GR, Nikookar M, Sahranavard L. Prediction of Surface Tension of Pure Hydrocarbons by An Artificial Neural Network System. Petroleum Science and Technology 2011;29(22):2384-96.

[302]   Farzi R, Esmaeilzadeh F. Prediction of surface tension of pure hydrocarbons using Esmaeilzadeh-Roshanfekr equation of state and group contribution method. Fluid Phase Equilibria 2016;427:353-61.

[303]   Gharagheizi F, Eslamimanesh A, Mohammadi AH, Richon D. Use of Artificial Neural Network-Group Contribution Method to Determine Surface Tension of Pure Compounds. Journal of Chemical & Engineering Data 2011;56(5):2587-601.

[304]   Saldana DA, Starck L, Mougin P, Rousseau B, Ferrando N, Creton B. Prediction of Density and Viscosity of Biofuel Compounds Using Machine Learning Methods. Energy & Fuels 2012;26(4):2416-26.

[305]   Katritzky AR, Chen K, Wang Y, Karelson M, Lucic B, Trinajstic N, et al. Prediction of liquid viscosity for organic compounds by a quantitative structure-property relationship. Journal of Physical Organic Chemistry 2000;13(1):80-6.

[306]   Ivanciuc O, Ivanciuc T, Filip PA, Cabrol-Bass D. Estimation of the Liquid Viscosity of Organic Compounds with a Quantitative Structure−Property Model. Journal of Chemical Information and Computer Sciences 1999;39(3):515-24.

[307]   Gharagheizi F, Mirkhani SA, Keshavarz MH, Farahani N, Tumba K. A molecular-based model for prediction of liquid viscosity of pure organic compounds: A quantitative structure property relationship (QSPR) approach. Journal of the Taiwan Institute of Chemical Engineers 2013;44(3):359-64.

[308]   Cai G, Liu Z, Zhang L, Zhao S, Xu C. Quantitative Structure–Property Relationship Model for Hydrocarbon Liquid Viscosity Prediction. Energy & Fuels 2018;32(3):3290-8.

[309]   Rajappan R, Shingade PD, Natarajan R, Jayaraman VK. Quantitative Structure−Property Relationship (QSPR) Prediction of Liquid Viscosities of Pure Organic Compounds Employing Random Forest Regression. Ind Eng Chem Res 2009;48(21):9708-12.

[310]   Pan Y, Jiang JC, Wang R, Jiang JJ. Predicting the net heat of combustion of organic compounds from molecular structures based on ant colony optimization. Journal of Loss Prevention in the Process Industries 2011;24(1):85-9.

[311]   Gharagheizi F. A simple equation for prediction of net heat of combustion of pure chemicals. Chemometrics

and Intelligent Laboratory Systems 2008;91(2):177-80.

[312] Albahri TA. Accurate prediction of the standard net heat of combustion from molecular structure. Journal of Loss Prevention in the Process Industries 2014;32:377-86.

[313] Frutiger J, Marcarie C, Abildskov J, Sin G. A Comprehensive Methodology for Development, Parameter Estimation, and Uncertainty Analysis of Group Contribution Based Property Models—An Application to the Heat of Combustion. Journal of Chemical & Engineering Data 2015;61(1):602-13.

[314] Albahri TA. Method for predicting the standard net heat of combustion for pure hydrocarbons from their molecular structure. Energy Conversion and Management 2013;76:1143-9.

[315] Gharagheizi F, Mirkhani SA, Tofangchi Mahyari A-R. Prediction of Standard Enthalpy of Combustion of Pure Compounds Using a Very Accurate Group-Contribution-Based Method. Energy & Fuels 2011;25(6):2651-4.

[316] Karelson M, Perkson A. QSPR prediction of densities of organic liquids. Computers & Chemistry 1999;23(1):49-59.

[317] Ihmels EC, Gmehling J. Extension and Revision of the Group Contribution Method GCVOL for the Prediction of Pure Compound Liquid Densities. Ind Eng Chem Res 2003;42(2):408-12.

[318] Stefanis E, Constantinou L, Tsivintzelis I, Panayiotou C. New Group-Contribution Method for Predicting Temperature-Dependent Properties of Pure Organic Compounds. International Journal of Thermophysics 2005;26(5):1369-88.

[319] St John PC, Kairys P, Das DD, McEnally CS, Pfefferle LD, Robichaud DJ, et al. A Quantitative Model for the Prediction of Sooting Tendency from Molecular Structure. Energy & Fuels 2017;31(9):9983-90.

[320] Kessler T, Schwartz T, Wong H-W, Mack JH. Screening Compounds for Fast Pyrolysis and Catalytic Biofuel Upgrading Using Artificial Neural Networks. *ASME 2019 Internal Combustion Engine Division Fall Technical Conference.* 2019.

[321] Das DD, St. John PC, McEnally CS, Kim S, Pfefferle LD. Measuring and predicting sooting tendencies of oxygenates, alkanes, alkenes, cycloalkanes, and aromatics on a unified scale. Combustion and Flame 2018;190:349-64.

[322] Gao Z, Zou X, Huang Z, Zhu L. Predicting sooting tendencies of oxygenated hydrocarbon fuels with machine learning algorithms. Fuel 2019;242:438-46.

[323] Kwon H, Shabnam S, van Duin ACT, Xuan Y. Numerical simulations of yield-based sooting tendencies of aromatic fuels using ReaxFF molecular dynamics. Fuel 2020;262.

[324] Borhani TNG, Afzali A, Bagheri M. QSPR estimation of the auto-ignition temperature for pure hydrocarbons. Process Safety and Environmental Protection 2016;103:115-25.

[325] Pan Y, Jiang J, Wang R, Cao H, Cui Y. Predicting the auto-ignition temperatures of organic compounds from molecular structure using support vector machine. J Hazard Mater 2009;164(2-3):1242-9.

[326] Pan Y, Jiang J, Wang R, Cao H. Advantages of support vector machine in QSPR studies for predicting auto-ignition temperatures of organic compounds. Chemometrics and Intelligent Laboratory Systems 2008;92(2):169-78.

[327] Pan Y, Jiang J, Wang R, Cao H, Zhao J. Prediction of auto-ignition temperatures of hydrocarbons by neural network based on atom-type electrotopological-state indices. J Hazard Mater 2008;157(2-3):510-7.

[328] Gharagheizi F. An accurate model for prediction of autoignition temperature of pure compounds. Journal of Hazardous Materials 2011;189(1-2):211-21.

[329] Frutiger J, Marcarie C, Abildskov J, Sin G. Group-contribution based property estimation and uncertainty analysis for flammability-related properties. J Hazard Mater 2016;318:783-93.

[330] Lazzús JA. Autoignition Temperature Prediction Using an Artificial Neural Network with Particle Swarm Optimization. International Journal of Thermophysics 2011;32(5):957-73.

[331] Keshavarz MH, Gharagheizi F, Ghanbarzadeh M. A simple correlation for prediction of autoignition

temperature of various classes of hydrocarbons. Journal of the Iranian Chemical Society 2012;10(3):545-57.

[332] Katritzky AR, Stoyanova-Slavova IB, Dobchev DA, Karelson M. QSPR modeling of flash points: an update. J Mol Graph Model 2007;26(2):529-36.

[333] Katritzky AR, Petrukhin R, Jain R, Karelson M. QSPR analysis of flash points. J Chem Inf Comput Sci 2001;41(6):1521-30.

[334] Tetteh J, Suzuki T, Metcalfe E, Howells S. Quantitative Structure−Property Relationships for the Estimation of Boiling Point and Flash Point Using a Radial Basis Function Neural Network. Journal of Chemical Information and Computer Sciences 1999;39(3):491-507.

[335] Mirshahvalad H, Ghasemiasl R, Raoufi N, Malekzadeh Dirin M. A Neural Network QSPR Model for Accurate Prediction of Flash Point of Pure Hydrocarbons. Mol Inform 2019;38(4):e1800094.

[336] Bagheri M, Bagheri M, Heidari F, Fazeli A. Nonlinear molecular based modeling of the flash point for application in inherently safer design. Journal of Loss Prevention in the Process Industries 2012;25(1):40-51.

[337] Alibakshi A. Strategies to develop robust neural network models: Prediction of flash point as a case study. Anal Chim Acta 2018;1026:69-76.

[338] Gharagheizi F, Ilani-Kashkouli P, Mirkhani SA, Mohammadi AH. Computation of Upper Flash Point of Chemical Compounds Using a Chemical Structure-Based Model. Ind Eng Chem Res 2012;51(13):5103-7.

[339] Rowley JR, Rowley RL, Wilding WV. Estimation of the flash point of pure organic chemicals from structural contributions. Process Safety Progress 2010;29(4):353-8.

[340] Santos SM, Nascimento DC, Costa MC, Neto AMB, Fregolente LV. Flash point prediction: Reviewing empirical models for hydrocarbons, petroleum fraction, biodiesel, and blends. Fuel 2020;263.

[341] Gharagheizi F, Eslamimanesh A, Ilani-Kashkouli P, Mohammadi AH, Richon D. QSPR molecular approach for representation/prediction of very large vapor pressure dataset. Chemical Engineering Science 2012;76:99-107.

[342] Yaffe D, Cohen Y. Neural network based temperature-dependent quantitative structure property relations (QSPRs) for predicting vapor pressure of hydrocarbons. J Chem Inf Comput Sci 2001;41(2):463-77.

[343] Beck B, Breindl A, Clark T. QM/NN QSPR models with error estimation: vapor pressure and logP. J Chem Inf Comput Sci 2000;40(4):1046-51.

[344] Katritzky AR, Slavov SH, Dobchev DA, Karelson M. Rapid QSPR model development technique for prediction of vapor pressure of organic compounds. Computers & Chemical Engineering 2007;31(9):1123-30.

[345] Asher WE, Pankow JF. Vapor pressure prediction for alkenoic and aromatic organic compounds by a UNIFAC-based group contribution method. Atmospheric Environment 2006;40(19):3588-600.

[346] Pan Y, Jiang J, Wang R, Cao H, Cui Y. A novel QSPR model for prediction of lower flammability limits of organic compounds based on support vector machine. J Hazard Mater 2009;168(2-3):962-9.

[347] Bagheri M, Rajabi M, Mirbagheri M, Amin M. BPSO-MLR and ANFIS based modeling of lower flammability limit. Journal of Loss Prevention in the Process Industries 2012;25(2):373-82.

[348] Chen C-C, Lai C-P, Guo Y-C. A novel model for predicting lower flammability limits using Quantitative Structure Activity Relationship approach. Journal of Loss Prevention in the Process Industries 2017;49:240-7.

[349] Pan Y, Ji X, Ding L, Jiang J. Prediction of Lower Flammability Limits for Binary Hydrocarbon Gases by Quantitative Structure-A Property Relationship Approach. Molecules 2019;24(4).

[350] Gharagheizi F. A new group contribution-based model for estimation of lower flammability limit of pure compounds. J Hazard Mater 2009;170(2-3):595-604.

[351] Lazzús JA. Neural network/particle swarm method to predict flammability limits in air of organic

compounds. Thermochimica Acta 2011;512(1-2):150-6.

[352] Albahri TA. Prediction of the lower flammability limit percent in air of pure compounds from their molecular structures. Fire Safety Journal 2013;59:188-201.

[353] Gharagheizi F. Prediction of upper flammability limit percent of pure compounds from their molecular structures. J Hazard Mater 2009;167(1-3):507-10.

[354] Pan Y, Jiang J, Ding X, Wang R, Jiang J. Prediction of flammability characteristics of pure hydrocarbons from molecular structures. AIChE Journal 2009:NA-NA.

[355] Pan Y, Jiang J, Wang R, Cao H, Cui Y. Prediction of the Upper Flammability Limits of Organic Compounds from Molecular Structures. Ind Eng Chem Res 2009;48(10):5064-9.

[356] Gharagheizi F. Chemical Structure-Based Model for Estimation of the Upper Flammability Limit of Pure Compounds. Energy & Fuels 2010;24(7):3867-71.

[357] Li R, Herreros JM, Tsolakis A, Yang W. Machine learning regression based group contribution method for cetane and octane numbers prediction of pure fuel compounds and mixtures. Fuel 2020;280.

[358] Rumble JR. CRC Handbook of Chemistry and Physics 101st Edition. CRC Press 2020.

[359] Albahri TA, Alashwak DA. Modeling of pure compounds surface tension using QSPR. Fluid Phase Equilibria 2013;355:87-91.

[360] DIPPR 801 Database. American Institute of Chemical Engineers (AIChE) 2019:https://www.aiche.org/dippr.

[361] Das DD, McEnally CS, Kwan TA, Zimmerman JB, Cannella WJ, Mueller CJ, et al. Sooting tendencies of diesel fuels, jet fuels, and their surrogates in diffusion flames. Fuel 2017;197:445-58.

[362] ASTM D1322-19 Standard Test Method for Smoke Point of Kerosene and Aviation Turbine Fuel. American Society for Testing and Materials (ASTM) 2019.

[363] Calcote HF, Manos DM. Effect of molecular structure on incipient soot formation. Combustion and Flame 1983;49(1-3):289-304.

[364] Barrientos EJ, Lapuerta M, Boehman AL. Group additivity in soot formation for the example of C-5 oxygenated hydrocarbon fuels. Combustion and Flame 2013;160(8):1484-98.

[365] Crossley SP, Alvarez WE, Resasco DE. Novel Micropyrolyis Index (MPI) to Estimate the Sooting Tendency of Fuels. Energy & Fuels 2008;22(4):2455-64.

[366] McEnally CS, Pfefferle LD. Sooting tendencies of oxygenated hydrocarbons in laboratory-scale flames. Environ Sci Technol 2011;45(6):2498-503.

[367] Schweidtmann AM, Rittig JG, König A, Grohe M, Mitsos A, Dahmen M. Graph Neural Networks for Prediction of Fuel Ignition Quality. Energy & Fuels 2020;34(9):11395-407.

[368] McEnally CS, Pfefferle LD, Das DD. Yield Sooting Index Database Volume 1. 2016:https://doi.org/10.7910/DVN/9FTNCK.

[369] Das DD, McEnally CS, Pfefferle LD. Sooting tendencies of unsaturated esters in nonpremixed flames. Combustion and Flame 2015;162(4):1489-97.

[370] McEnally CS, Das DD, Pfefferle LD. Yield Sooting Index Database Volume 2: Sooting Tendencies of a Wide Range of Fuel Compounds on a Unified Scale. HARVARD Dataverse 2017:https://doi.org/10.7910/DVN/7HGFT8.

[371] Cai G, Liu Z, Zhang L, Shi Q, Zhao S, Xu C. Systematic performance evaluation of gasoline molecules based on quantitative structure-property relationship models. Chemical Engineering Science 2021;229.

[372] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision 2015;115(3):211-52.

[373] ImageNet. 2016:http://www.image-net.org.

[374] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017.

[375] Tsang S-H. Review: Xception — With Depthwise Separable Convolution, Better Than Inception-v3 (Image Classification). Medium 2018:https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568.

[376] Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017:2261-9.

[377] Tsang S-H. Review: DenseNet — Dense Convolutional Network (Image Classification). Medium 2018:https://towardsdatascience.com/review-densenet-image-classification-b6631a8ef803.

[378] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016:2818-26.

[379] Tsang S-H. Review: Inception-v3 — 1st Runner Up (Image Classification) in ILSVRC 2015. Medium 2018:https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c.

[380] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016:770-8.

[381] Tsang S-H. Review: ResNet-38 — Wider or Deeper ResNet? (Image Classification & Semantic Segmentation). Medium 2019:https://sh-tsang.medium.com/resnet-38-wider-or-deeper-resnet-image-classification-semantic-segmentation-f297f2f73437.

[382] Wu Z, Shen C, van den Hengel A. Wider or Deeper: Revisiting the ResNet Model for Visual Recognition. Pattern Recognition 2019;90:119-33.

[383] Redmon J. Darknet: Open Source Neural Networks in C https://pjreddie.com/darknet/.

[384] Redmon J. ImageNet Classification.https://pjreddie.com/darknet/imagenet/.

[385] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proceedings of the AAAI Conference on Artificial Intelligence 2017;31(1):4278-84.

[386] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM 2017;60(6):84-90.

[387] Tsang S-H. Review: AlexNet, CaffeNet — Winner of ILSVRC 2012 (Image Classification). Medium 2018:https://medium.com/coinmonks/paper-review-of-alexnet-caffenet-winner-in-ilsvrc-2012-image-classification-b93598314160.

[388] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper With Convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015:1-9.

[389] Tsang S-H. Review: GoogLeNet (Inception v1)— Winner of ILSVRC 2014 (Image Classification). Medium 2018:https://medium.com/coinmonks/paper-review-of-googlenet-inception-v1-winner-of-ilsvlc-2014-image-classification-c2b3565a64e7.

[390] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. International Conference on Learning Representations 2015(1-14).

[391] Tsang S-H. Review: VGGNet — 1st Runner-Up (Image Classification), Winner (Localization) in ILSVRC 2014. 2018.

[392] Zhang X, Zhou X, Lin M, Sun J. ShuffleNet-An Extremely Efficient Convolutional Neural Network for Mobile Devices. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition 2018.

[393] Ma N, Zhang X, Zheng H-T, Sun J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. Computer Vision – ECCV 2018 2018:122-38.

[394] Tsang S-H. Review: ShuffleNet V1 — Light Weight Model (Image Classification). Medium 2019:https://towardsdatascience.com/review-shufflenet-v1-light-weight-model-image-classification-5b253dfe982f.

[395] Tsang S-H. Reading: ShuffleNet V2 — Practical Guidelines for Efficient CNN Architecture Design (Image

Classification). Medium 2020:https://sh-tsang medium.com/reading-shufflenet-v2-practical-guidelines-for-e-fficient-cnn-architecture-design-image-287b05abc08a.

[396] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition 2018:4510-20.

[397] Tsang S-H. Review: MobileNetV2 — Light Weight Model (Image Classification). Medium 2019:https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification-8febb490e61c.

[398] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. 2016:1-13.

[399] Tsang S-H. Review: SqueezeNet (Image Classification). Medium 2019:https://towardsdatascience.com/review-squeezenet-image-classification-e7414825581a.

[400] Mingxing Tan QL. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning 2019;97:6105-14.

[401] Tsang S-H. [Paper] EfficientNet-Rethinking Model Scaling for Convolutional Neural Networks (Image Classification). Medium 2020:https://sh-tsang.medium.com/efficientnet-rethinking-model-scaling-for-convolutional-neural-networks-image-classification-ef67b0f14a4d.

[402] Practical deep learning examples with MATLAB. MathWorks, Inc 2019:https://uk mathworks.com/campaigns/offers/deep-learning-examples-with-matlab html.

[403] ChemDraw Professional. PerkinElmer:https://www.perkinelmer.com.cn/product/chemdraw-professional-chemdrawpro.

[404] Beale MH, Hagan MT, Demuth HB. Deep Learning Toolbox User's Guide. MathWorks, Inc 2020:https://uk mathworks.com/help/pdf_doc/deeplearning/nnet_ug.pdf.

[405] Ramo K. Hands-On Java Deep Learning for Computer Vision. Packt; 2019.

[406] Kohse-Höinghaus K. A new era for combustion research. Pure and Applied Chemistry 2019;91(2):271-88.

[407] Wang Z, Burra KG, Lei T, Gupta AK. Co-pyrolysis of waste plastic and solid biomass for synergistic production of biofuels and chemicals-A review. Progress in Energy and Combustion Science 2021;84.

[408] Ren J, Cao J-P, Zhao X-Y, Liu Y-L. Fundamentals and applications of char in biomass tar reforming. Fuel Processing Technology 2021;216.

[409] Abou Rjeily M, Gennequin C, Pron H, Abi-Aad E, Randrianalisoa JH. Pyrolysis-catalytic upgrading of bio-oil and pyrolysis-catalytic steam reforming of biogas: a review. Environmental Chemistry Letters 2021.

[410] Mohan D, Pittman CU, Steele PH. Pyrolysis of Wood/Biomass for Bio-oil: A Critical Review. Energy & Fuels 2006;20(3):848-89.

[411] Katritzky AR, Kuanar M, Slavov S, Hall CD, Karelson M, Kahn I, et al. Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction. Chem Rev 2010;110(10):5714-89.

[412] McCormick RL, Fioroni G, Fouts L, Christensen E, Yanowitz J, Polikarpov E, et al. Selection Criteria and Screening of Potential Biomass-Derived Streams as Fuel Blendstocks for Advanced Spark-Ignition Engines. SAE International Journal of Fuels and Lubricants 2017;10(2):442-60.

[413] Lunderman S, Fioroni GM, McCormick RL, Nimlos MR, Rahimi MJ, Grout RW. Screening Fuels for Autoignition with Small-Volume Experiments and Gaussian Process Classification. Energy & Fuels 2018;32(9):9581-91.

[414] Fioroni G, Fouts L, Luecke J, Vardon D, Huq N, Christensen E, et al. Screening of Potential Biomass-Derived Streams as Fuel Blendstocks for Mixing Controlled Compression Ignition Combustion. *SAE Technical Paper Series.* 2019.

[415] Wagner R, Gaspar D, Bryan P, McCormick R. Co-Optimization of Fuels & Engines FY19 Year in Review. 2019:https://www.energy.gov/sites/prod/files/2020/06/f75/beto-co-optima-fy19-yir-report-june-.pdf.

[416] Huq NA, Huo X, Hafenstine GR, Tifft SM, Stunkel J, Christensen ED, et al. Performance-advantaged ether

diesel bioblendstock production by a priori design. Proc Natl Acad Sci U S A 2019;116(52):26421-30.

[417] Whitmore LS, Davis RW, McCormick RL, Gladden JM, Simmons BA, George A, et al. BioCompoundML: A General Biofuel Property Screening Tool for Biological Molecules Using Random Forest Classifiers. Energy & Fuels 2016;30(10):8410-8.

[418] Li G, Hu Z, Hou F, Li X, Wang L, Zhang X. Machine learning enabled high-throughput screening of hydrocarbon molecules for the design of next generation fuels. Fuel 2020;265.

[419] Farrell J, Wagner R, Gaspar D, Moen C. Co-Optimization of Fuels & Engines FY18 Year in Review. 2018:https://www.energy.gov/sites/prod/files/2019/06/f64/Co-Optima_YIR8_FINAL_LOWRES%20190619_0.pdf.

[420] Szybist JP, Splitter DA. Understanding chemistry-specific fuel differences at a constant RON in a boosted SI engine. Fuel 2018;217:370-81.

[421] Szybist JP, Splitter DA. Pressure and temperature effects on fuels with varying octane sensitivity at high load in SI engines. Combustion and Flame 2017;177:49-66.

[422] Messerly RA, Rahimi MJ, St. John PC, Luecke JH, Park J-W, Huq NA, et al. Towards quantitative prediction of ignition-delay-time sensitivity on fuel-to-air equivalence ratio. Combustion and Flame 2020;214:103-15.

[423] Gaspar D. Co-Optima Overview. Pacific Northwest National Laboratory 2019:https://www.energy.gov/sites/prod/files/2019/04/f61/Co-Optima%20Overview_NL0029892a_1.pdf.

[424] Li R, Herreros JM, Tsolakis A, Yang W. Integrated machine learning-quantitative structure property relationship (ML-QSPR) and chemical kinetics for high throughput fuel screening toward internal combustion engine. Fuel 2022;307.

[425] Atef N, Issayev G, Mohamed SY, Najjar A, Wang Z, Wang J-Y, et al. Chemical kinetic study of triptane (2,2,3-trimethylbutane) as an anti-knock additive. Combustion and Flame 2019;210:399-412.

[426] Ren S, Kokjohn SL, Wang Z, Liu H, Wang B, Wang J. A multi-component wide distillation fuel (covering gasoline, jet fuel and diesel fuel) mechanism for combustion and PAH prediction. Fuel 2017;208:447-68.

[427] Dayma G, Glaude PA, Fournet R, Battin-Leclerc F. Experimental and modeling study of the oxidation of cyclohexene. Int J Chem Kinet 2003;35(7):273-85.

[428] Ribaucour M, Lemaire O, Minetti R. Low-temperature oxidation and autoignition of cyclohexene: A modeling study. Proceedings of the Combustion Institute 2002;29(1):1303-10.

[429] Mittal G, Burke SM, Davies VA, Parajuli B, Metcalfe WK, Curran HJ. Autoignition of ethanol in a rapid compression machine. Combustion and Flame 2014;161(5):1164-71.

[430] Cai L, Kröger L, Döntgen M, Leonhard K, Narayanaswamy K, Sarathy SM, et al. Exploring the combustion chemistry of a novel lignocellulose-derived biofuel: cyclopentanol. Part I: quantum chemistry calculation and kinetic modeling. Combustion and Flame 2019;210:490-501.

[431] Cai L, Minwegen H, Kruse S, Daniel Büttgen R, Hesse R, Ramalingam A, et al. Exploring the combustion chemistry of a novel lignocellulose-derived biofuel: cyclopentanol. Part II: experiment, model validation, and functional group analysis. Combustion and Flame 2019;210:134-44.

[432] Zhang J, Pan L, Zhang Z, Mo J, Huang Z. Shock Tube and Kinetic Modeling Study of Isobutanal Oxidation. Energy & Fuels 2013;27(5):2804-10.

[433] Hemken C, Burke U, Lam K-Y, Davidson DF, Hanson RK, Heufer KA, et al. Toward a better understanding of 2-butanone oxidation: Detailed species measurements and kinetic modeling. Combustion and Flame 2017;184:195-207.

[434] Hemken C, Burke U, Graf I, Ruwe L, Park S, Sarathy SM, et al. A laminar flame investigation of 2-butanone, and the combustion-related intermediates formed through its oxidation. Proceedings of the Combustion Institute 2017;36(1):1175-83.

[435] Zhang K, Lokachari N, Ninnemann E, Khanniche S, Green WH, Curran HJ, et al. An experimental,

theoretical, and modeling study of the ignition behavior of cyclopentanone. Proceedings of the Combustion Institute 2019;37(1):657-65.

[436] Tran L-S, Wang Z, Carstensen H-H, Hemken C, Battin-Leclerc F, Kohse-Höinghaus K. Comparative experimental and modeling study of the low- to moderate-temperature oxidation chemistry of 2,5-dimethylfuran, 2-methylfuran, and furan. Combustion and Flame 2017;181:251-69.

[437] Darcy D, Nakamura H, Tobin CJ, Mehl M, Metcalfe WK, Pitz WJ, et al. A high-pressure rapid compression machine study of n-propylbenzene ignition. Combustion and Flame 2014;161(1):65-74.

[438] Mehl M, Pitz WJ, Westbrook CK, Curran HJ. Kinetic modeling of gasoline surrogate components and mixtures under engine conditions. Proceedings of the Combustion Institute 2011;33(1):193-200.

[439] Mehl M, Curran HJ, Pitz WJ, Westbrook CK. Chemical kinetic modeling of component mixtures relevant to gasoline. 4th European Combustion Meeting, Vienna, Austria 2009:https://combustion.llnl.gov/content/assets/docs/combustion/LLNL_CONF_410968.pdf.

[440] Wang H, Yao M, Reitz RD. Development of a Reduced Primary Reference Fuel Mechanism for Internal Combustion Engine Combustion Simulations. Energy & Fuels 2013;27(12):7843-53.

[441] Farrell JT, Zigler BT, Ratcliff MA, Miles P, Kolodziej C, Sjoberg M, et al. Efficiency Merit Function for Spark-Ignition Engines. National Renewable Energy Laboratory 2018.

[442] Chang Y, Jia M, Liu Y, Li Y, Xie M, Yin H. Application of a Decoupling Methodology for Development of Skeletal Oxidation Mechanisms for Heavy n-Alkanes from n-Octane to n-Hexadecane. Energy & Fuels 2013;27(6):3467-79.

[443] Mao Y, Wang S, Wu Z, Qiu Y, Yu L, Ruan C, et al. An experimental and kinetic modeling study of n-butylcyclohexane over low-to-high temperature ranges. Combustion and Flame 2019;206:83-97.

[444] Wang H, Dames E, Sirjean B, Sheen DA, Tango R, Violi A, et al. A high-temperature chemical kinetic model of n-alkane (up to n-dodecane), cyclohexane, and methyl-, ethyl-, n-propyl and n-butyl-cyclohexane oxidation at high temperatures, JetSurF version 2.0. 2010:http://web.stanford.edu/group/haiwanglab/JetSurF/JetSurF2.0/index.html.

[445] Cai L, Uygun Y, Togbé C, Pitsch H, Olivier H, Dagaut P, et al. An experimental and modeling study of n-octanol combustion. Proceedings of the Combustion Institute 2015;35(1):419-27.

[446] Li Q, Liu H, Zhang Y, Yan Z, Deng F, Huang Z. Experimental and kinetic modeling study of laminar flame characteristics of higher mixed alcohols. Fuel Processing Technology 2019;188:30-42.

[447] Cai L, Jacobs S, Langer R, vom Lehn F, Heufer KA, Pitsch H. Auto-ignition of oxymethylene ethers (OMEn, n = 2–4) as promising synthetic e-fuels from renewable electricity: shock tube experiments and automatic mechanism generation. Fuel 2020;264.

[448] Cai L, Minwegen H, Beeckmann J, Burke U, Tripathi R, Ramalingam A, et al. Experimental and numerical study of a novel biofuel: 2-Butyltetrahydrofuran. Combustion and Flame 2017;178:257-67.

[449] Kerschgens B, Cai L, Pitsch H, Heuser B, Pischinger S. Di-n-buthylether, n-octanol, and n-octane as fuel candidates for diesel engine combustion. Combustion and Flame 2016;163:66-78.

[450] Heuser B, Mauermann P, Wankhade R, Kremer F, Pischinger S. Combustion and emission behavior of linear C8-oxygenates. International Journal of Engine Research 2015;16(5):627-38.

[451] Hoppe F, Heuser B, Thewes M, Kremer F, Pischinger S, Dahmen M, et al. Tailor-made fuels for future engine concepts. International Journal of Engine Research 2015;17(1):16-27.

[452] Cai L, Sudholt A, Lee DJ, Egolfopoulos FN, Pitsch H, Westbrook CK, et al. Chemical kinetic study of a novel lignocellulosic biofuel: Di-n-butyl ether oxidation in a laminar flow reactor and flames. Combustion and Flame 2014;161(3):798-809.

[453] Thion S, Togbé C, Serinyel Z, Dayma G, Dagaut P. A chemical kinetic study of the oxidation of dibutyl-ether in a jet-stirred reactor. Combustion and Flame 2017;185:4-15.

[454] Zhang K, Banyon C, Bugler J, Curran HJ, Rodriguez A, Herbinet O, et al. An updated experimental and

kinetic modeling study of n-heptane oxidation. Combustion and Flame 2016;172:116-35.

[455] MGRefsMaster Reference Calibrations Version 2008. MKS Instruments, Inc 2008.

[456] MultiGas™ 2030. MKS Instruments, Inc 2017:https://www.mksinst.com/p/MULTIGAS-2030.

[457] AVL DiTEST Gas 1000 user manual. AVL DiTEST GmbH 2020:https://www.avlditest.com/index.php/en/downloads-en html.

[458] Serhan N, Tsolakis A, Martos FJ. Effect of propylene glycol ether fuelling on the different physico-chemical properties of the emitted particulate matters: Implications of the soot reactivity. Fuel 2018;219:1-11.

[459] Heywood JB. Internal Combustion Engine Fundamentals, 2nd Edition. New York: McGraw-Hill; 2018.

[460] Jääskeläinen H. Engine Efficiency. 2020:https://dieselnet.com/tech/engine_efficiency.php.

[461] Reschke GD. Optimization of a Flame Ionization Detector for Determination of Hydrocarbon in Diluted Automotive Exhausts. SAE Technical Paper 770141 1977.

[462] Lin Q, Tay KL, Zhao F, Yang W. Enabling robust simulation of polyoxymethylene dimethyl ether 3 (PODE3) combustion in engines. International Journal of Engine Research 2021.

[463] Lin Q, Tay KL, Yu W, Zong Y, Yang W, Rivellini L-H, et al. Polyoxymethylene dimethyl ether 3 (PODE3) as an alternative fuel to reduce aerosol pollution. Journal of Cleaner Production 2021;285.

[464] Jacobs S, Döntgen M, Alquaity ABS, Hesse R, Kruse S, Beeckmann J, et al. A Comprehensive Experimental and Kinetic Modeling Study of the Combustion Chemistry of Diethoxymethane. Energy & Fuels 2021.

[465] Li R, Herreros JM, Tsolakis A, Yang W. Chemical kinetic modeling of diethoxymethane oxidation: A carbon–neutral fuel. Fuel 2021;291.

[466] Liu H, Wang Z, Li Y, Zheng Y, He T, Wang J. Recent progress in the application in compression ignition engines and the synthesis technologies of polyoxymethylene dimethyl ethers. Applied Energy 2019;233-234:599-611.

[467] Hackbarth K, Haltenort P, Arnold U, Sauer J. Recent Progress in the Production, Application and Evaluation of Oxymethylene Ethers. Chemie Ingenieur Technik 2018;90(10):1520-8.

[468] Awad OI, Ma X, Kamil M, Ali OM, Ma Y, Shuai S. Overview of polyoxymethylene dimethyl ether additive as an eco-friendly fuel for an internal combustion engine: Current application and environmental impacts. Sci Total Environ 2020;715:136849.

[469] Klokic S, Hochegger M, Schober S, Mittelbach M. Investigations on an efficient and environmentally benign poly(oxymethylene) dimethyl ether (OME3-5) fuel synthesis. Renewable Energy 2020;147:2151-9.

[470] Burre J, Bongartz D, Mitsos A. Production of Oxymethylene Dimethyl Ethers from Hydrogen and Carbon Dioxide—Part II: Modeling and Analysis for OME3–5. Ind Eng Chem Res 2019;58(14):5567-78.

[471] Ouda M, Yarce G, White RJ, Hadrich M, Himmel D, Schaadt A, et al. Poly(oxymethylene) dimethyl ether synthesis – a combined chemical equilibrium investigation towards an increasingly efficient and potentially sustainable synthetic route. Reaction Chemistry & Engineering 2017;2(1):50-9.

[472] Barro C, Parravicini M, Boulouchos K. Neat polyoxymethylene dimethyl ether in a diesel engine; part 1: Detailed combustion analysis. Fuel 2019;256.

[473] Barro C, Parravicini M, Boulouchos K, Liati A. Neat polyoxymethylene dimethyl ether in a diesel engine; part 2: Exhaust emission analysis. Fuel 2018;234:1414-21.

[474] Jing Z, Zhang C, Cai P, Li Y, Yin P, Wang J. An experimental and numerical study of polyoxymethylene dimethyl ethers on a homogeneous charge compression ignition engine. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 2020:1-16.

[475] Kass M, Wissink M, Janke C, Connatser R, Curran S. Compatibility of Elastomers with Polyoxymethylene Dimethyl Ethers and Blends with Diesel. *SAE Technical Paper Series.* 2020.

[476] Lin Q, Tay KL, Yu W, Yang W, Wang Z. Effects of polyoxymethylene dimethyl ether 3 (PODE3) addition and injection pressure on combustion performance and particle size distributions in a diesel engine. Fuel 2021;283.

[477] Chen H, He J, Chen Z, Geng L. A comparative study of combustion and emission characteristics of dual-fuel engine fueled with diesel/methanol and diesel–polyoxymethylene dimethyl ether blend/methanol. Process Safety and Environmental Protection 2021;147:714-22.

[478] Liu H, Wang Z, Zhang J, Wang J, Shuai S. Study on combustion and emission characteristics of Polyoxymethylene Dimethyl Ethers/diesel blends in light-duty and heavy-duty diesel engines. Applied Energy 2017;185:1393-402.

[479] Chen H, He J, Hua H. Investigation on Combustion and Emission Performance of a Common Rail Diesel Engine Fueled with Diesel/Biodiesel/Polyoxymethylene Dimethyl Ethers Blends. Energy & Fuels 2017;31(11):11710-22.

[480] Huang H, Li Z, Teng W, Zhou C, Huang R, Liu H, et al. Influence of n-butanol-diesel-PODE3-4 fuels coupled pilot injection strategy on combustion and emission characteristics of diesel engine. Fuel 2019;236:313-24.

[481] Huang H, Li Z, Teng W, Huang R, Liu Q, Wang Y. Effects of EGR rates on combustion and emission characteristics in a diesel engine with n-butanol/PODE3-4/diesel blends. Applied Thermal Engineering 2019;146:212-22.

[482] Huang H, Liu Q, Teng W, Pan M, Liu C, Wang Q. Improvement of combustion performance and emissions in diesel engines by fueling n -butanol/diesel/PODE 3–4 mixtures. Applied Energy 2017.

[483] Chen H, Su X, Li J, Zhong X. Effects of gasoline and polyoxymethylene dimethyl ethers blending in diesel on the combustion and emission of a common rail diesel engine. Energy 2019;171:981-99.

[484] Liu J, Yang J, Sun P, Gao W, Yang C, Fang J. Compound combustion and pollutant emissions characteristics of a common-rail engine with ethanol homogeneous charge and polyoxymethylene dimethyl ethers injection. Applied Energy 2019;239:1154-62.

[485] Li D, Gao Y, Liu S, Ma Z, Wei Y. Effect of polyoxymethylene dimethyl ethers addition on spray and atomization characteristics using a common rail diesel injection system. Fuel 2016;186:235-47.

[486] Tan YR, Salamanca M, Pascazio L, Akroyd J, Kraft M. The effect of poly(oxymethylene) dimethyl ethers (PODE3) on soot formation in ethylene/PODE3 laminar coflow diffusion flames. Fuel 2021;283.

[487] Ferraro F, Russo C, Schmitz R, Hasse C, Sirignano M. Experimental and numerical study on the effect of oxymethylene ether-3 (OME3) on soot particle formation. Fuel 2021;286.

[488] Yan F, Xu L, Wang Y, Park S, Sarathy SM, Chung SH. On the opposing effects of methanol and ethanol addition on PAH and soot formation in ethylene counterflow diffusion flames. Combustion and Flame 2019;202:228-42.

[489] Mahbub N, Oyedun AO, Kumar A, Oestreich D, Arnold U, Sauer J. A life cycle assessment of oxymethylene ether synthesis from biomass-derived syngas as a diesel additive. Journal of Cleaner Production 2017;165:1249-62.

[490] Hank C, Lazar L, Mantei F, Ouda M, White RJ, Smolinka T, et al. Comparative well-to-wheel life cycle assessment of OME3–5 synfuel production via the power-to-liquid pathway. Sustainable Energy & Fuels 2019;3(11):3219-33.

[491] He T, Liu H-y, Wang Y, Wang B, Liu H, Wang Z. Development of Surrogate Model for Oxygenated Wide-Distillation Fuel with Polyoxymethylene Dimethyl Ether. SAE International Journal of Fuels and Lubricants 2017;10(3).

[492] He T, Wang Z, You X, Liu H, Wang Y, Li X, et al. A chemical kinetic mechanism for the low- and intermediate-temperature combustion of Polyoxymethylene Dimethyl Ether 3 (PODE 3). Fuel 2018;212:223-35.

[493] Sun W, Wang G, Li S, Zhang R, Yang B, Yang J, et al. Speciation and the laminar burning velocities of poly(oxymethylene) dimethyl ether 3 (POMDME 3) flames: An experimental and modeling study. Proceedings of the Combustion Institute 2017;36(1):1269-78.

[494] Ren S, Wang Z, Li B, Liu H, Wang J. Development of a reduced polyoxymethylene dimethyl ethers (PODEn) mechanism for engine applications. Fuel 2019;238:208-24.

[495] Lin Q, Tay KL, Zhou D, Yang W. Development of a compact and robust Polyoxymethylene Dimethyl Ether 3 reaction mechanism for internal combustion engines. Energy Conversion and Management 2019;185:35-43.

[496] Huang H, Chen Y, Zhu J, Chen Y, Lv D, Zhu Z, et al. Construction of a Reduced PODE3/Nature Gas Dual-Fuel Mechanism under Enginelike Conditions. Energy & Fuels 2019;33(4):3504-17.

[497] Curran HJ, Gaffuri P, Pitz WJ, Westbrook CK. A comprehensive modeling study of n-heptane oxidation. Combustion and Flame 1998;114(1-2):149-77.

[498] Assanis DN, Filipi ZS, Fiveland SB, Syrimis M. A Predictive Ignition Delay Correlation Under Steady-State and Transient Operation of a Direct Injection Diesel Engine. Journal of Engineering for Gas Turbines and Power 2003;125(2).

[499] ANSYS Chemkin Tutorials Manual. ANSYS Reaction Design 2016.

[500] Kuo KK-y. Principles of Combustion. Hoboken, New Jersy: John Wiley & Sons; 2005.

[501] Glassman I, Yetter R. Combustion. 4th ed. New York: Elsevier Inc; 2008.

[502] Law CK. Combustion Physics. United States of America: Cambridge University Press; 2010.

[503] Paris Agreement. UNITED NATIONS 2015:https://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf.

[504] Antelava A, Jablonska N, Constantinou A, Manos G, Salaudeen SA, Dutta A, et al. Energy Potential of Plastic Waste Valorization: A Short Comparative Assessment of Pyrolysis versus Gasification. Energy & Fuels 2021;35(5):3558-71.

[505] C M, T S, M C. A review on conversion techniques of liquid fuel from waste plastic materials. International Journal of Energy Research 2017;41(11):1534-52.

[506] Ma C, Yu J, Wang B, Song Z, Xiang J, Hu S, et al. Chemical recycling of brominated flame retarded plastics from e-waste for clean fuels production: A review. Renewable and Sustainable Energy Reviews 2016;61:433-50.

[507] SUNERGY.https://www.sunergy-initiative.eu/.

[508] SUNRISE https://sunriseaction.com/.

[509] Faber C, Allahverdiyeva-Rinne Y, Artero V, Baraton L, Barbieri A, Bercegol H, et al. SUNRISE Technological Roadmap. 2020:https://sunriseaction.com/wp-content/uploads/2020/06/Roadmap_finalversion_web.pdf.

[510] Energy-X https://www.energy-x.eu/.

[511] Nørskov JK, Latimer A, Dickens CF. Research needs towards sustainable production of fuels and chemicals. ENERGY-X 2019:https://www.energy-x.eu/research-needs-report/.

[512] Solar Fuels Network.https://www.solarfuelsnetwork.com/.

[513] Kopernikus project P2X https://www.kopernikus-projekte.de/en/projects/p2x.

[514] Deutz S, Bongartz D, Heuser B, Kätelhön A, Schulze Langenhorst L, Omari A, et al. Cleaner production of cleaner fuels: wind-to-wheel – environmental assessment of CO2-based oxymethylene ether as a drop-in fuel. Energy & Environmental Science 2018;11(2):331-43.

[515] Bossmann T, Fournié L, Humberset L, Khallouf P. The role and potential of Power-to-X in 2050. 2019.

[516] CO-OPTIMIZATION OF FUELS & ENGINES. US Department of Energy:https://www.energy.gov/eere/bioenergy/co-optimization-fuels-engines.

[517] Farrell J, Wagner R, Holladay J, Moen C. Co-Optimization of Fuels & Engines FY17 Year in Review. 2017:https://www.energy.gov/sites/prod/files/2018/04/f50/Co-Optima_YIR7_FINAL_Web_180417_0.pdf.

[518] The Fuel Science Center-Adaptive Conversion Systems for Renewable Energy and Carbon Sources. RWTH

Aachen University:https://www.fuelcenter.rwth-aachen.de/cms/Fuelcenter/~siun/Forschung/lidx/1/.

[519] Deshmukh AY, Davidovic M, Grenga T, Schumacher L, Kirsch V, Palmer J, et al. Bio-hybrid Fuels: From Molecular Structure to Combustion and Emissions. Conference: 8th International Conference - "Fuel Science-From Production to Propulsion" 2020:https://www.researchgate.net/publication/342617004_Bio-hybrid_Fuels_From_Molecular_Structure_to_Combustion_and_Emissions.

[520] Muradov NZ, Veziroğlu TN. Carbon-Neutral Fuels and Energy Carriers. Boca Raton: CRC Press; 2011.

[521] Wang S, Lu A, Zhong CJ. Hydrogen production from water electrolysis: role of catalysts. Nano Converg 2021;8(1):4.

[522] Wu H, Feng C, Zhang L, Zhang J, Wilkinson DP. Non-noble Metal Electrocatalysts for the Hydrogen Evolution Reaction in Water Electrolysis. Electrochemical Energy Reviews 2021.

[523] Garba MD, Usman M, Khan S, Shehzad F, Galadima A, Ehsan MF, et al. CO_2 towards fuels: A review of catalytic conversion of carbon dioxide to hydrocarbons. Journal of Environmental Chemical Engineering 2021;9(2).

[524] Tu W, Zhou Y, Zou Z. Photocatalytic conversion of CO(2) into renewable hydrocarbon fuels: state-of-the-art accomplishment, challenges, and prospects. Adv Mater 2014;26(27):4607-26.

[525] Bui M, Adjiman CS, Bardow A, Anthony EJ, Boston A, Brown S, et al. Carbon capture and storage (CCS): the way forward. Energy & Environmental Science 2018;11(5):1062-176.

[526] Garcia-Garcia G, Fernandez MC, Armstrong K, Woolass S, Styring P. Analytical Review of Life-Cycle Environmental Impacts of Carbon Capture and Utilization Technologies. ChemSusChem 2021;14(4):995-1015.

[527] Dias V, Vandooren J. Experimental and modeling studies of C2H4/O2/Ar, C2H4/methylal/O2/Ar and C2H4/ethylal/O2/Ar rich flames and the effect of oxygenated additives. Combustion and Flame 2011;158(5):848-59.

[528] Zhang C, He J, Li Y, Li X, Li P. Ignition delay times and chemical kinetics of diethoxymethane/O2/Ar mixtures. Fuel 2015;154:346-51.

[529] Hashemi H, Christensen JM, Harding LB, Klippenstein SJ, Glarborg P. High-pressure oxidation of propane. Proceedings of the Combustion Institute 2019;37(1):461-8.

[530] Baigmohammadi M, Patel V, Nagaraja S, Ramalingam A, Martinez S, Panigrahy S, et al. Comprehensive Experimental and Simulation Study of the Ignition Delay Time Characteristics of Binary Blended Methane, Ethane, and Ethylene over a Wide Range of Temperature, Pressure, Equivalence Ratio, and Dilution. Energy & Fuels 2020;34(7):8808-23.

[531] Mohamed AAE-S, Panigrahy S, Sahu AB, Bourque G, Curran H. An experimental and kinetic modeling study of the auto-ignition of natural gas blends containing C1–C7 alkanes. Proceedings of the Combustion Institute 2020.

[532] Bugler J, Rodriguez A, Herbinet O, Battin-Leclerc F, Togbé C, Dayma G, et al. An experimental and modelling study of n-pentane oxidation in two jet-stirred reactors: The importance of pressure-dependent kinetics and new reaction pathways. Proceedings of the Combustion Institute 2017;36(1):441-8.

[533] Bugler J, Marks B, Mathieu O, Archuleta R, Camou A, Grégoire C, et al. An ignition delay time and chemical kinetic modeling study of the pentane isomers. Combustion and Flame 2016;163:138-56.

[534] Shrestha KP, Eckart S, Elbaz AM, Giri BR, Fritsche C, Seidel L, et al. A comprehensive kinetic model for dimethyl ether and dimethoxymethane oxidation and NO interaction utilizing experimental laminar flame speed measurements at elevated pressure and temperature. Combustion and Flame 2020;218:57-74.

[535] vom Lehn F, Cai L, Pitsch H. Impact of thermochemistry on optimized kinetic model predictions: Auto-ignition of diethyl ether. Combustion and Flame 2019;210:454-66.

[536] Sakai Y, Herzler J, Werler M, Schulz C, Fikri M. A quantum chemical and kinetics modeling study on the autoignition mechanism of diethyl ether. Proceedings of the Combustion Institute 2017;36(1):195-202.

[537] Serinyel Z, Lailliau M, Dayma G, Dagaut P. A high pressure oxidation study of di-n-propyl ether. Fuel 2020;263.

[538] Jacobs S, Döntgen M, Alquaity ABS, Kopp WA, Kröger LC, Burke U, et al. Detailed kinetic modeling of dimethoxymethane. Part II: Experimental and theoretical study of the kinetics and reaction mechanism. Combustion and Flame 2019;205:522-33.

[539] Vermeire FH, Carstensen H-H, Herbinet O, Battin-Leclerc F, Marin GB, Van Geem KM. Experimental and modeling study of the pyrolysis and combustion of dimethoxymethane. Combustion and Flame 2018;190:270-83.

[540] Li R, Herreros JM, Tsolakis A, Yang W. Chemical kinetic study on ignition and flame characteristic of polyoxymethylene dimethyl ether 3 (PODE3). Fuel 2020;279.

[541] Zhou C-W, Li Y, Burke U, Banyon C, Somers KP, Ding S, et al. An experimental and chemical kinetic modeling study of 1,3-butadiene combustion: Ignition delay time and laminar flame speed measurements. Combustion and Flame 2018;197:423-38.

[542] NIST Chemistry WebBook https://webbook nist.gov/chemistry/.

[543] PubChem. National Institutes of Health:https://pubchem ncbi.nlm nih.gov/.

[544] ChemSpider. Royal Society of Chemistry:http://www.chemspider.com/.

[545] Lehrheuer B, Hoppe F, Heufer KA, Jacobs S, Minwegen H, Klankermayer J, et al. Diethoxymethane as tailor-made fuel for gasoline controlled autoignition. Proceedings of the Combustion Institute 2019;37(4):4691-8.

[546] McAllister S, Chen J-Y, Fernandez-Pello AC. Fundamentals of Combustion Processes. New York: Springer; 2011.

[547] Curran H. A comprehensive modeling study of iso-octane oxidation. Combustion and Flame 2002;129(3):253-80.

[548] Westbrook CK, Mehl M, Pitz WJ, Kukkadapu G, Wagnon S, Zhang K. Multi-fuel surrogate chemical kinetic mechanisms for real world applications. Phys Chem Chem Phys 2018;20(16):10588-606.

[549] Dias V, Vandooren J, Jeanmart H. Experimental and Modeling Study of Propanal/H2/O2/Ar Flames at Low Pressure. Combustion Science and Technology 2016;188(4-5):556-70.

[550] Lay TH, Bozzelli JW, Dean AM, Ritter ER. Hydrogen Atom Bond Increments for Calculation of Thermodynamic Properties of Hydrocarbon Radical Species. The Journal of Physical Chemistry 1995;99(39):14514-27.

[551] Glaude PA, Pitz WJ, Thomson MJ. Chemical kinetic modeling of dimethyl carbonate in an opposed-flow diffusion flame. Proceedings of the Combustion Institute 2005;30(1):1111-8.

[552] Tee LS, Gotoh S, Stewart WE. Molecular Parameters for Normal Fluids. Lennard-Jones 12-6 Potential. Industrial & Engineering Chemistry Fundamentals 1966;5(3):356-63.

[553] Yaws CL. Yaws' handbook of thermodynamic and physical properties of chemical compounds : physical, thermodynamic and transport properties for 5,000 organic chemical compounds. Knovel; 2003.

[554] AP1700 Material property calculation and inquiry platform.http://www.ap1700.com/.

[555] Kopp WA, Kröger LC, Cai L, Hesse R, Kruse S, Alquaity ABS, et al. Combustion Chemistry of Diethoxymethane (DEM): Experiments, Theory and Modeling. The 6th Fuel Science-From Production to Propulsion 2018.

[556] Gomez-Bombarelli R, Wei JN, Duvenaud D, Hernandez-Lobato JM, Sanchez-Lengeling B, Sheberla D, et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. ACS Cent Sci 2018;4(2):268-76.

[557] Ren F, Ward L, Williams T, Laws KJ, Wolverton C, Hattrick-Simpers J, et al. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. Sci Adv 2018;4(4):eaaq1566.